METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

In Silico Methods for Predicting Drug Toxicity

Edited by

Emilio Benfenati

Mario Negri Inst for Pharm Rsrch, IRCCS, Milan, Italy

Humana Press

Editor
Emilio Benfenati
Mario Negri Inst for Pharm Rsrch
IRCCS
Milan, Italy

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-3607-6 ISBN 978-1-4939-3609-0 (eBook)
DOI 10.1007/978-1-4939-3609-0

Library of Congress Control Number: 2016937006

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Humana Press is a brand of Springer
Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

The use of in silico methods for pharmaceutical substances is quite consolidated in certain areas and, at the same time, is facing new perspectives in others. The whole area is complex for the high number of models, which apply different approaches, such as quantitative structure-activity relationships, models to evaluate binding to receptors, ADME methods, and tools for system biology. Today the challenge is even more difficult than in the past, since we discovered that pharmaceuticals reach the environment and affect the living systems. Thus, a wise planning of the new drugs and suitable treatment systems of the waste should take into consideration not only the toxic effects towards human beings but also the environmental effects.

[AU1] When we look at models for mutagenicity, organ-specific toxicity, reprotoxicity, and repeated dose toxicity, we can say that for some properties the models have a long tradition, while for many other properties the research is in the initial phase, and in some cases the current performance of the models is not sufficient; indeed, the reliability of the models is not homogeneous.

The scenario is evolving not only for the different general approaches and for the number of applications but also for the impact of the debate on in silico models, which is ongoing in other sectors. In Europe, the REACH regulation established some criteria for the use of in silico models, while in the USA the initiative Tox21 is challenging the traditional way to conduct toxicological screening. Other initiatives are offering new perspectives within different industrial sectors, and different points of view may arise from experiences achieved within major pharmaceutical companies, consultants, and centers offering access to internet-based resources.

On the basis of this complex series of factors, this book aims to present the theory and the applications, the common standards and the perspectives, giving voice to contributions from the different stakeholders. Several contributions derive from academia and research institutes in pharmacology, others from regulatory bodies, industry, and consultants of pharmaceutical companies. Contributions are also derived from several parts of the world, since the in silico modeling studies are conducted all around the world.

Besides a general introduction, the book is divided in three main parts. In Part I, there are contributions relative to sophisticated models addressing the binding to receptors, pharmacokinetics and adsorption, metabolism, distribution, and excretion. These are general processes, and the reader can see the approaches that are used.

In Part II, the book goes through a series of models for specific toxicological and ecotoxicological endpoints. Each endpoint offers different approaches, depending on the specific property and on the level of maturity of the tools.

Finally, Part III of the book offers a broad view of the main initiatives and new perspectives which will very likely improve our way of modeling pharmaceuticals. The direct experience of some of the key stakeholders provides the personal experience, with useful, practical insights.

In this book, we combine the theoretical, advanced research with the practical application of the tools. It is important to understand the theoretical basis, but it is also important to know how to correctly use the tools. We should know what each tool offers, to better

exploit what is available, but we have to know where the limitations are and avoid misuse of the tools, generating false expectations and false interpretation of the results. The book contains a step-by-step discussion showing how to extract all available information from the models, used alone or combined, but also indicating the uncertainty of the results with useful case studies. Tens of models are introduced, and tens of practical case studies explain how to use the programs and interpret the results, because modern programs do not simply have the calculated value as output.

[AU2] Since the in silico methods are evolving, we will give voice of the new perspectives and initiatives around the world, which are attempting to change the classical way to make studies in toxicology. Thus, it is important to be prepared to understand changes in the paradigms which are anticipated.

Computational toxicology is a fascinating area, but also a complex one, and the best way to solve the complex phenomena generating toxicity is to use a battery of tools. These tools will be more and more integrated. In silico tools offer the advantage of incorporating data and knowledge from different fields, such as chemistry, biology, -omics, and pharmacology. The beauty of this approach is that the computational methods define through the number and algorithms the ideal way to establish a dialogue between different scientific domains. This approach is transparent and allows for maintaining all the features associated with the original data, including for instance the information on the uncertainty and vari-

[AU3] ability. The main limitation of this approach is maybe based on our limitation to think in a complex way and to exploit the best of what technology can offer.

Milan, Italy ***Emilio Benfenati***

Contents

Contributors

Tara S. Barton-Maclaren • *Existing Substances Risk Assessment Bureau, Health Canada, Ottawa, ON, Canada*

Arianna Bassan • *S-IN Soluzioni Informatiche SRL, Vicenza, Italy*

Emilio Benfenati • *Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Milano, Italy*

Frederic Y. Bois • *DRC/VIVA/METO, INERIS, Verneuil en Halatte, France*

Alessandro Brigo • *Roche Pharma Research & Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland*

Lorenza Broccardo • *S-IN Soluzioni Informatiche SRL, Vicenza, Italy*

Céline Brochot • *DRC/VIVA/METO, INERIS, Verneuil en Halatte, France*

Julien Burton • *Systems Toxicology Unit and EURL ECVAM, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra, Italy*

Angelo Carotti • *Dipartimento di Farmacia – Scienze del Farmaco, Università degli Studi di Bari "Aldo Moro", Bari, Italy*

Minjun Chen • *Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, USA*

Antonella Ciancetta • *Molecular Modeling Section (MMS), Dipartimento di Scienze del Farmaco, Università Padova, Padova, Italy*

Antonia Diukendjieva • *Department of QSAR & Molecular Modeling, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria*

Elena Fioravanzo • *S-IN Soluzioni Informatiche SRL, Vicenza, Italy*

Matteo Floris • *Dipartimento di Scienze Biomediche, Università degli Studi di Sassari, Sassari, Italy*

Robert Fraczkiewicz • *Simulations Plus, Inc., Lancaster, CA, USA*

Jayeeta Ghosh • *Simulations Plus, Inc., Lancaster, CA, USA*

Giuseppina Gini • *DEIB, Politecnico di Milano, Milano, Italy*

Azadi Golbamaki • *Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, RCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Milano, Italy*

Vijay Gombar • *Simulations Plus, Inc., Lancaster, CA, USA*

Dennie G.A.J. Hebels • *Department of Toxicogenomics, MERLN Institute, Maastricht University, Maastricht, The Netherlands; Department of Cell Biology-Inspired Tissue Engineering, MERLN Institute, Maastricht University, Maastricht, The Netherlands*

Ralf Herwig • *Bioinformatics Group, Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany*

Mark Hewitt • *School of Pharmacy, Faculty of Science and Engineering, University of Wolverhampton, Wolverhampton, UK*

Huixiao Hong • *Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, USA*

Danyel G.J. Jennen • *Department of Toxicogenomics, Maastricht University, Maastricht, The Netherlands*

Richard S. Judson • *National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA*
Supratik Kar • *Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India*
Jos C.S. Kleinjans • *Department of Toxicogenomics, Maastricht University, Maastricht, The Netherlands*
Simona Kovarich • *S-IN Soluzioni Informatiche SRL, Vicenza, Italy*
Sunil Kulkarni • *Existing Substances Risk Assessment Bureau, Health Canada, Ottawa, ON, Canada*
Michael S. Lawless • *Simulations Plus, Inc., Lancaster, CA, USA*
Serena Manganelli • *Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Milano, Italy*
Giuseppe Felice Mangiatordi • *Dipartimento di Farmacia – Scienze del Farmaco, Università degli Studi di Bari "Aldo Moro", Bari, Italy*
Kamel Mansouri • *ORISE, National Center for Computational Toxicology, Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA*
Marco Marzo • *Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Milano, Italy*
Enrico Mombelli • *INERIS, Verneuil en Halatte, France*
Stefano Moro • *Molecular Modeling Section (MMS), Dipartimento di Scienze del Farmaco, Università Padova, Padova, Italy*
Wolfgang Muster • *Roche Pharma Research & Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland*
Glenn J. Myatt • *Leadscope, Inc., Columbus, OH, USA*
Hui Wen Ng • *Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, USA*
Orazio Nicolotti • *Dipartimento di Farmacia – Scienze del Farmaco, Università degli Studi di Bari "Aldo Moro", Bari, Italy*
Ettore Novellino • *Dipartimento di Farmacia, Università degli Studi di Napoli "Federico II", Napoli, Italy*
Manuela Pavan • *S-IN Soluzioni Informatiche SRL, Vicenza, Italy*
Fabiola Pizzo • *Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Milano, Italy*
Katarzyna Przybylak • *School of Pharmacy and Biomolecular Science, Liverpool John Moores University, Liverpool, UK*
Donald P. Quigley • *Leadscope, Inc., Columbus, OH, USA*
Giuseppa Raitano • *Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Milano, Italy*
Axel Rasche • *Bioinformatics Group, Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany*

Alessandra Roncaglioni • *Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, IRCCS - Istituto di Ricerche Farmacologiche "Mario Negri", Milano, Italy*
Kunal Roy • *Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India*
Martin Smieško • *Molecular Modeling, Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland*
Mattia Sturlese • *Molecular Modeling Section (MMS), Dipartimento di Scienze del Farmaco, Università Padova, Padova, Italy*
Weida Tong • *Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR, USA*
Ivanka Tsakovska • *Department of QSAR & Molecular Modelling, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria*
Angelo Vedani • *Molecular Modeling, Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland*
Mathieu Vinken • *Department of In Vitro Toxicology and Dermato-Cosmetology, Vrije Universiteit Brussel, Brussels, Belgium*
Marvin Waldman • *Simulations Plus, Inc., Lancaster, CA, USA*
Gerard J.P. van Westen • *Division of Medicinal Chemistry, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands*
Andrew P. Worth • *Systems Toxicology Unit and EURL ECVAM, Institute for Health and Consumer Protection, Joint Research Centre, European Commission, Ispra, Italy*
Chihae Yang • *Molecular Networks GmbH, Erlangen, Germany; Altamira LLC, Columbus, OH, USA*

Author Queries

Chapter No.: FM 0002663039

Queries	Details Required	Author's Response
AU1	Please check if edit to the sentence "When we look at models…" is okay	
AU2	Should "voice of" be changed to "voice to"?	
AU3	Should "maybe" be deleted in the sentence "The main limitation of…"?	

Part I

Modeling a Pharmaceutical in the Human Body

Chapter 1

QSAR Methods

Giuseppina Gini

Abstract

In this chapter, we introduce the basis of computational chemistry and discuss how computational methods have been extended to some biological properties and toxicology, in particular. Since about 20 years, chemical experimentation is more and more replaced by modeling and virtual experimentation, using a large core of mathematics, chemistry, physics, and algorithms. Then we see how animal experiments, aimed at providing a standardized result about a biological property, can be mimicked by new in silico methods. Our emphasis here is on toxicology and on predicting properties through chemical structures. Two main streams of such models are available: models that consider the whole molecular structure to predict a value, namely QSAR (Quantitative Structure Activity Relationships), and models that find relevant substructures to predict a class, namely SAR. The term in silico discovery is applied to chemical design, to computational toxicology, and to drug discovery. We discuss how the experimental practice in biological science is moving more and more toward modeling and simulation. Such virtual experiments confirm hypotheses, provide data for regulation, and help in designing new chemicals.

Key words Computer models, Toxicity prediction, SAR and QSAR

1 Starting from Chemistry

"All science is computer science." When a New York Times article published on March 25, 2001 used this sentence in the title, the general public was aware that the introduction of computers has changed the way that experimental sciences has been carried out so far. Chemistry together with physics is the best example of such a new way of making science.

A new discipline, *chemoinformatics* has been in existence for the past two decades [1, 2]. Many of the activities performed in chemoinformatics are information retrieval [3], aimed at searching for new molecules of interest when a single molecule has been identified as being relevant. However, chemoinformatics is more than "chemical information"; it requires strong algorithmic development.

It is useful to remember that models of atoms were defined by analogy with different systems; Thomson in 1897 modeled the

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_1,

atom as a sphere of positive electricity with negative particles; Rutherford in 1909 adapted the solar system model with a dense positively charged nucleus surrounded by negative electrons.

Finally in the 1920s the electron cloud model was defined; in this model an atom consists of a dense nucleus composed of protons and neutrons surrounded by electrons. A molecule is an electrically neutral group of two or more atoms held together by covalent bonds, sharing electrons. The valence model naturally transforms a molecule into a graph, where the nodes are atoms and the edges are bonds. This graph representation is usually called 2D chemical structure.

The graph theory, whose basic definition has been established back in eighteenth century, initially evolved through chemistry. Two scientists in particular, Alexander C. Brown and James J. Sylvester, developed the molecular representation as nodes (atoms, indicated by their name) and bonds. The edges are assigned weights according to the bond: single, double, triple, or aromatic where electrons are delocalized. Today hydrogens are implicitly represented in the graph since they are assumed to fill the unused valences [4].

A common representation of the graph is the adjacency matrix, a square matrix with dimension N equal to the number of atoms. Each position (i, j) in the matrix specifies the absence (0 value) or the presence of a bond connecting the atoms i and j, filled with 1, 2, 3 to indicate simple, double or triple bond, 4 for amide bond, and 5 for aromatic bond. The diagonal elements are always zero. An example of a matrix representation is in Fig. 1.

This is only one of the possible representations of a molecule. Structure Data Format (SDF) files represent the molecule into two blocks: the atom block and the bond block. A database record entry in Simplified Molecular Input Line Entry Specification (SMILES) [5] is very popular. This is a short string representation of the molecular structure, in a context free language expressing the graph visit in a depth first style, listing bonds and atoms encountered, and adding parentheses for branches. Hydrogens are left out. Table 1 shows some examples of different representations for molecules.

	C1	*C2*	*C3*	*C4*	*C5*
C1	0	1	0	0	0
C2	1	0	1	0	1
C3	0	1	0	1	0
C4	0	0	1	0	0
C5	0	1	0	0	0

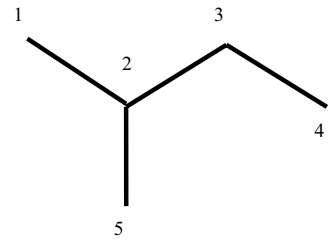


Fig. 1 Adjacency matrix of 2-methylbutane pictured after hydrogen elimination; only the five carbon atoms are considered and numbered

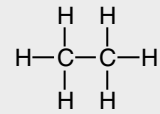
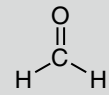
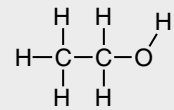
The SMILES notation suffers the lack of a unique representation, since a molecule can be encoded beginning anywhere. In Table 1 we see for ethanol four SMILES string, all correct. Therefore, a method of encoding a molecule was quickly developed that provided an invariant SMILES representation, called canonical SMILES [6].

Recent developments in line notations are the InChI (International Chemical Identifier) codes, supported by the International Union of Pure and Applied Chemistry (IUPAC), which can uniquely describe a molecule, at different levels of detail, but is not intended for human readability [7].

What about the real shape of molecules? They are 3D objects and as such they should be represented. Let us take again as an example the 2-methylbutane molecule, illustrated as a simple drawing in Fig. 1. Its formula, SMILES, and 3D conformation are illustrated in Fig. 2a–c.

Defining the 3D shape of a molecule will take us to the basic methods of computational chemistry.

Table 1
Examples of molecules with their SMILES code

SMILES	Name	Formula	Graph
CC	Ethane	CH_3CH_3	
C=O	Formaldehyde	CH_2O	
O=C			
CCO	Ethanol	CH_3CH_2OH	
OCC			
C(C)O			
C(O)C			

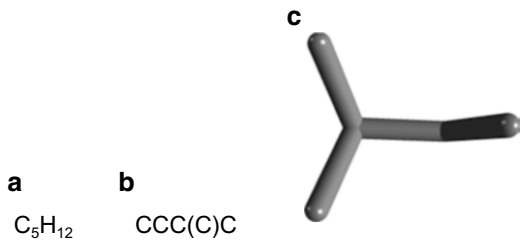


Fig. 2 2-Methylbutane: (**a**) chemical formula, (**b**) SMILES, and (**c**) 3D conformer (from NIH PubChem)

2 Computational Chemistry

Computational chemistry is a branch of chemistry that uses computers to assist in solving chemical problems, studying the electronic structure of solids, liquids, and designing new materials and drugs. It uses the results of theoretical chemistry, incorporated into programs, to calculate the structures and properties of molecules. The methods cover both static and dynamic situations: accurate methods—*ab initio* methods and less accurate methods—called *semiempirical.*

It all happened in about 50 years.

1. In the early 1950s, the first semiempirical atomic orbital calculations.
2. The first *ab initio* calculations in 1956 at MIT.
3. Nobel prize for Chemistry, in 1998, assigned to John Pople and Walter Kohn, for Computational Chemistry.
4. Nobel prize for Chemistry assigned in 2013 to chemistry assigned to Martin Karplus, Michael Levitt, and Arieh Warshel for their development of multiscale models for complex chemical systems.

Computational chemistry is a way to move away from the traditional approach of solving scientific problems using only direct experimentation, but it does not remove experimentation. Experiments produce new data and facts. The role of theory is to situate all the new data into a framework, based on mathematical rules.

Computational chemistry uses theories to produce new facts in a manner similar to the real experiments. It is now possible to simulate in the computer an experiment before running it.

In modeling chemical processes, two variables are important, namely time and temperature. It is necessary to make dynamic simulations and to model the force fields that exist between atoms and explain the bonds breaking. This task usually requires solving the quantum mechanics equations.

A hierarchy of simulation levels provides different levels of details. The study of the fundamental properties without the introduction of empirical parameters is the so-called *ab initio* methods. Those computations consider the electronic and structural properties of the molecule at the absolute zero temperature. They are computationally expensive, so the size of the molecules is limited to a few hundred atoms. When the *ab initio* methods cannot be used, it is possible to introduce empirical parameters to obtain the so-called molecular dynamics methods.

Today the applications of quantum mechanics to chemistry are widely used. The most notable is the Gaussian software, developed at the Carnegie Mellon University of Pittsburgh (PA). This program gained a large popularity since the Nobel Prize for chemistry in 1998 was assigned to Pople, one of the inventors of Gaussian.

2.1 Molecular Simulation

Using computers to calculate the intermolecular forces, it is possible to compute a detailed "history" of the molecules. Analyzing this history, by the methods of statistical mechanics, affords a detailed description of the behavior of matter [8]. Three techniques are available:

- Molecular Dynamics (MD) Simulation. In this technique, the forces between molecules are calculated explicitly and the motion of the molecules is computed using a numerical integration method. The starting conditions are the positions of the atoms (from a known crystal structure) and their velocities (randomly generated). Following Newton's equations, from the initial positions, velocities and forces, it is possible to calculate the positions and velocities of the atoms at a small time interval later. From these new positions the forces are recalculated and another step in time made. Following an equilibration period of many thousands of time steps, during which the system "settles down" to the desired temperature and pressure a production period begins where the history of the molecules is stored for later analysis.
- Monte Carlo (MC) Simulation. Monte Carlo simulation resembles the Molecular Dynamics method in that it also generates a history of the molecules in a system, which is subsequently used to calculate the bulk properties of the system by means of statistical mechanics. However, the procedure for moving the atoms employs small random moves used in conjunction with a sampling algorithm to confine the random walk to thermodynamically meaningful configurations.
- Molecular Mechanics (MM) Modeling. MM is a method for predicting the structures of complex molecules, based on the energy minimization of its potential energy function, obtained empirically, by experiment, or by the methods of quantum chemistry. The energy minimization method is an advanced algorithm to optimize the speed of convergence. The methods main advantage is its computational cheapness.

Any of those methods is necessary to optimize the 3D structure of the molecule before constructing models that use the 3D shape instead of the graph representation of the molecules.

3 Biological Models for Toxicology

Since the nineteenth century the practice of animal experimentation was established in physiology, microbiology, and surgery. The explosion in molecular biology in the second half of the twentieth century increased the importance of in vivo models [9].

All models have their limitations, their prediction can be poor, and their transferability to the real phenomena they model can be unsatisfactory. So extrapolating data from animal models to the environment or to human health depends on the degree to which the animal model is an appropriate reflection of the condition under investigation.

These limitations are, however, an intrinsic part of all modeling approaches. Most of the questions about animal models are ethical more than scientific; in public health, the use of animal models is imposed by strict regulations and is unlikely that any health authority will approve novel drugs without supporting animal data.

3.1 Bioassays for Toxicity

Toxicity is the degree to which a substance can damage an organism. Toxicity is a property of concern for every chemical substance. Theophrastus Phillipus von Hohenheim (1493–1541) Paracelsus wrote: "All things are poison and nothing is without poison; only the dose makes a thing not a poison."

The relationship between dose and its effects on the exposed organism is of high significance in toxicology. The process of using animal testing to assess toxicity of chemicals has been defined in the following way:

- Toxicity can be measured by its effects on the target.
- Because individuals have different levels of response to the same dose of a toxin, a population-level measure of toxicity is often used which relates the probabilities of an outcome for a given individual in a population. Example is LD_{50}: the dose that causes the death of 50 % of the population.
- When the dose is individuated, multiply it for a "safety factor," to account for uncertainty in the data and for differences between species. For example, use 10 if data are from mammals or 100 if data come from other animals.

This process is based on assumptions that usually are very crude and presents many open issues. For instance, it is more difficult to determine the toxicity of chemical mixtures (gasoline, cigarette smoke, waste) since the percentages of the chemicals can vary and the combination of the effects is not just a summation of them.

Perhaps the most common continuous measure of biological activity is the $\log(IC_{50})$ (inhibitory concentration), which measures the concentration of a particular compound necessary to induce a 50 % inhibition of the biological activity under investigation. Similarly the median lethal dose, LD_{50}, is the dose required to kill half the members of a tested population after a specified test duration. It has been created by J.W. Trevan in 1927 and is usually expressed in milligrams per kilogram of body weight. LD_{50} is not the lethal dose for all subjects, only for half of them.

The dose–response relationship describes the change in effect on an organism caused by differing levels of doses to a chemical after a certain exposure time. A dose–response curve is a *x*–*y* graph relating the dose to the response of the organism.

- The measured dose is plotted on the *X* axis and the response is plotted on the *Y* axis.
- The response is a physiological or biochemical response.
- LD_{50} is used in human toxicology; IC_{50}—inhibition concentration and its dual EC_{50}—effect concentration are used in pharmacology.

Usually the logarithm of the dose is plotted on the *X* axis and in such cases the curve is typically sigmoidal, with the steepest portion in the middle. In Fig. 3, we see an example of the dose–response curve for LD_{50}.

Today also in vitro testing is available. It is the scientific analysis of the effects of a chemical on cultured bacteria or mammalian cells. Experiments using in vitro systems are useful in the early phases of medical studies where the screening of large number of potential therapeutic candidates may be necessary, or in making fast tests for possible pollutants. However, in vitro systems are nonphysiological and have important limitations. It is known that their results poorly correlate with the results of in vivo. However, there are substantial advantages in using in vitro systems to advance mechanistic understanding of toxicant activities and the use of human cells to define human-specific toxic effects.

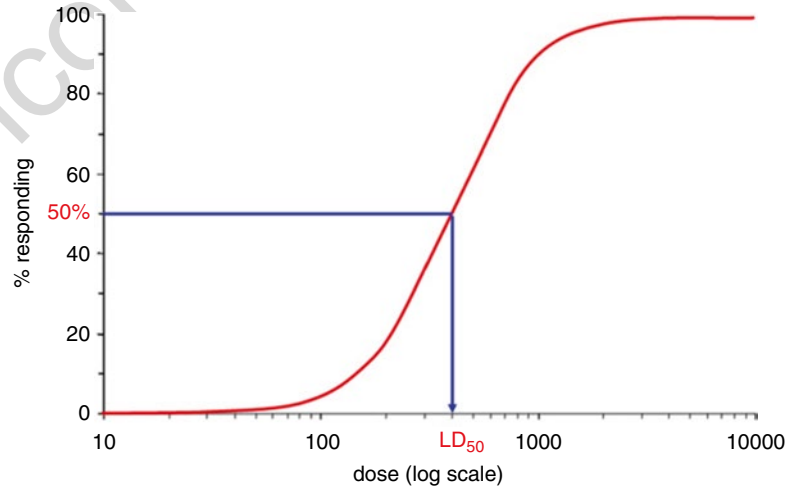


Fig. 3 A curve for log LD_{50}. *Source*: http://www.dropdata.org/RPU/pesticide_activity.htm

4 In Silico Methods

Animal testing refers to the use of nonhuman animals in experiments. Worldwide it is estimated that the number of vertebrate animals annually used for animal experiments is in the order of tens of millions. In toxicity, animal tests are called in vivo models; they give doses for some species and are used to extrapolate data to human health or to the environment. As we said above, the extrapolation of data from species to species is not obvious. For instance, the lethal doses for rats and for mice are sometimes very different.

How to construct a model that relates a chemical structure to the effect was investigated even before computers were available. The term in silico today covers the methods devoted to this end; in silico refers to the fact that computers are used and computers have silicon in their hardware. The most-known in silico methods are the QSAR (Quantitative Structure Activity Relationships) methods, based on the assertion that the molecular structure is responsible for all the activities [10–12].

4.1 QSAR

From quantitative data, we can build a QSAR model that seeks to correlate our particular response variable of interest with molecular descriptors that have been computed or even measured from the molecules themselves. What we today refer to as QSAR methods were first pioneered by Corwin Hansch [13] in the 1940s, who analyzed congeneric series of compound and formulated the QSAR equation:

$$Log1/C = ap + bs + c\mathrm{Es} + \mathrm{const}$$

where

C = effect concentration
p = octanol–water partition coefficient
s = Hammett substituent constant (electronic)
Es = Taft's substituent constant

Log P octanol–water partition coefficient, is the ratio of concentrations of a compound in the two phases of a mixture of two immiscible solvents at equilibrium. It is a measure of the difference in solubility of the compound in these two solvents. Normally one of the solvents is water while the second is hydrophobic such as octanol. With high octanol/water partition coefficient the chemical substance is hydrophobic and preferentially distributed to hydrophobic compartments such as cell membrane, while hydrophilic are found in hydrophilic compartments such as blood serum. Log P values today are predicted in most of the cases [14].

The definitions of the chemical structure and of the function remain a challenge today, but relating structure to property is widely adopted in drug discovery and in risk assessment.

Sometimes the QSAR methods take more specific names as: QSPR (quantitative structure property relationship) or QSTR (quantitative structure toxicity relationship). QSPR are used for physicochemical properties, as the boiling point, the solubility, log P [15].

They all correlate a dependent variable (the effect or response) with a set of independent variables (usually calculated properties or descriptors). They are statistical models and can be applied to predict the responses for unseen data points entirely in silico. It is possible to compute them from a model, not from an experiment.

4.1.1 Molecular Descriptors

The generation of informative data from molecular structures is of high importance in chemoinformatics since it is often used in statistical analyses of the molecules. There are many possible approaches to calculate molecular descriptors [16] that represent local or global salient characteristics of the molecule. Different classes of descriptors are:

- Constitutional descriptors, depending on the number and type of atoms, bonds, and functional groups.
- Geometrical descriptors that give molecular surface area and volume, moments of inertia, shadow area projections, and gravitational indices.
- Topological Indices, based on the topology of molecular graph [4]. Only the structural information is used in generating the description. Examples are the Wiener index (the sum of the number of bonds between all nodes in a molecular graph) and the Randic index (the branching of a molecule).
- Physicochemical descriptors attempt to estimate the physical properties of molecules. Examples are molecular weight, hydrogen bond acceptors, hydrogen bond donors, and partition coefficients, as log P. The calculation of log P predicts the logarithm of the partition coefficient between octanol and water and indicates the general lipophilicity (or hydrophobicity) of the substance.
- Electrostatic descriptors, such as partial atomic charges and others depending on the possibility to form hydrogen bonds.
- Quantum chemical descriptors, related to the molecular orbital and their properties.
- Fingerprints are instead binary strings coding the presence/absence of structures of interest, which are previously listed according to knowledge of which chemical entities can be relevant. Since substructure searching requires a time-consuming subgraph isomorphism algorithm, a sub-

structure screening rapid method was developed to create the structure-key fingerprints. The fingerprint is a binary string encoding a molecule, where the 1 or 0 in a position means that the substructure of this position in the dictionary is present or not. The dictionaries depend on the property under investigation.

4.1.2 Model Construction

The selection of descriptors to use follows the build-up method (adding one at a time) or the build-down method (removing one at a time). Also optimization methods based for instance on Genetic Algorithms can be applied.

Whatever method is then chosen [16] to develop predictive models, it is important to take heed of the model quality statistics and ensure a correct modeling methodology is used such as testing the model against an external and unseen test set to ensure it is not overfitting to the training set. Model extrapolation is another concern that frequently occurs when models are applied outside the space from which the models were generated. Again, numerous model statistics are available that can indicate if new data points, from which responses are to be predicted, can be applied to the model [17].

Two types of supervised learning methods are applied widely in building models chemoinformatics and toxicology: classification and regression. Classification methods assign new objects, in our case molecules, to two or more classes—most frequently either biologically active or inactive. Regression methods attempt to use continuous data, such as a measured biological response variable, to correlate molecules with that data so as to predict a continuous numeric value for new and unseen molecules using the generated model.

The most-often used methods for classification are Partial Least Squares, Linear Discriminant Analysis, Naive Bayesian Classifier, Decision Trees, Recursive Partitioning, and Support Vector Machines, whereas, for regression modeling, Multiple Linear Regression, Partial Least Squares, Support Vector Machines, and Artificial Neural Networks [18] are often used.

4.1.3 Model Acceptability

In many cases, published QSAR models implement the leave-one-out cross-validation procedure and compute the cross-validated determination coefficient R^2, called q^2. If y_{pi} and y_i are the predicted and observed property values, *ypim* and *yim*, respectively, are the average values of the predicted and observed property values, the determination coefficient is defined as

$$R^2 = 1 \quad \left(SUM\left(y_{pi} - y_i \right)^2 / SUM\left(y_{pi}^m - y_i^m \right)^2 \right) \qquad (1)$$

A high value of q^2 (for instance, $q^2 > 0.5$) is considered as an indicator or even as the ultimate proof that the model is highly predictive.

A high q^2 is the necessary condition for a model to have a high predictive power; however, it is not a sufficient condition. Beside the wide accepted criteria of checking q^2, some additional, stricter conditions are often used [19]. Indeed different parameters and values have been proposed. Regardless to the absolute values, we have to remember that the statistical performance of any model is related to the uncertainty and variability of the original data used to build the model.

4.1.4 Model Interpretation

Model interpretation is considered important since people would find it useful to understand the models from known basic principles. A low number of descriptors used and their role in a simple equation are often considered as necessary to accept a QSAR result.

There is generally a trade-off between prediction quality and interpretation quality. Interpretable models are generally desired in situations where the model is expected to provide information about the problem domain and how best to navigate through chemistry space allowing the medicinal chemist to make informed decisions. However, these models tend to suffer in terms of prediction quality as they become more interpretable. The reverse is true with predictive models in that their interpretation suffers as they become more predictive. Models that are highly predictive tend to use molecular descriptors that are not readily interpretable by the chemist. However, predictive models are generally not intended to provide transparency, but predictions that are more reliable and can therefore be used as high-throughput models. If interpretability is of concern, other methods are available, more or less as a kind of expert systems, or SAR.

However, both SAR and QSAR are predictive statistical models and as such they suffer the problems of the statistical learning theory, the theoretical framework about inference, that deals about how to gain knowledge from a set of data so to make prediction.

Learning from data assumes the statistical nature of the phenomena that generate data; it needs to observe a phenomenon, construct a model, and make predictions using the model. It is well known that it is always possible to find a function that fits the data. However, such function could be very bad in predicting new data, in particular if data are noisy. Among the many functions that can accomplish the task of inducing a model, we need to quantify their characteristics, as performance and simplicity. Simplicity has no unique definition; in statistics people prefer models with few free parameters, in physics models with few constants, in QSAR models with interpretable descriptors. The definition of any property depends on the specific phenomenon under study, so the "no free lunch theorem" expresses the limitations of all our inductive methods. The "no free lunch theorem" is a popular name to indicate the practical results of theorems demonstrated by Wolpert and Macready [20] and stating that any two models are equivalent when their performance is averaged across all possible problems.

The practical indication from this theorem is that we need assumptions on the phenomenon to study, otherwise there is no better algorithm. In other terms, data cannot replace knowledge. In practice, we should accurately describe which method we have successfully used and which priors explain its success. As far as priors hold, the learning method used is successful to get predictions.

4.2 SAR

SAR (Structure–Activity Relationships) typically makes use of rules created by experts to produce models that relates subgroups of the molecule atoms to a biological property. The SAR approach consists in detecting particular structural fragments of molecule already known to be responsible for the toxic property under investigation.

In the mutagenicity/carcinogenicity domain, the key contribution in the definition of such toxicophores comes from [21], who compiled a list of 19 Structural Alerts (SA) for DNA reactivity. Practically SAs are rules that state the condition of mutagenicity by the presence or the absence of peculiar chemical substructures. It is important mentioning that SAs are sound hypotheses that derive from chemical properties and have a sort of mechanistic interpretation; however, their presence alone is not a definitive method to prove the property under investigation, since the substituents present in some cases are able to change the classification.

A few examples exist of automatic construction of such SAR systems. The structure of chemicals is explicitly taken into account by some graph-mining approaches, which mine large datasets for frequent substructures. On the other hand, human experts usually estimate toxicity through the detection of particular structural fragments, already known to be responsible for the toxic property under investigation. In the literature, such fragments are referred to as SAs and are derived by human experts from knowledge of the biochemical mechanism of action; these mechanisms are quite studied for genotoxicity but in general are still poorly understood and largely unknown.

To this end, an automatic method for SA extraction is SARpy (SAR in python), a new ad hoc approach to automatically generate SAR models by finding the relevant fragments; it means that it can extract a set of rules directly from data without any a priori knowledge [22]. Briefly, the algorithm generates substructures of arbitrary complexity and automatically selects the fragments to become SAs on the basis of their prediction performance on a training set. The rule set extracted for each model is then applied to the new molecule/s for prediction. The model tags the compound as toxic when one or more SAs for the specific toxicity endpoint are present in the molecular structure and as nontoxic if no SA is found by the model. Moreover the user can ask SARpy to also generate rules

related to nontoxic substances and use them to better assign molecules to the nontoxic class.

Given a training set of molecular structures expressed in the SMILES notation, with their experimental activity binary labels, SARpy generates every substructure in the set and mines correlations between the incidence of a particular molecular substructure and the activity of the molecules that contain it. This is done in three steps starting just from the SMILES:

- Fragmentation: this recursive algorithm considers every combination of bond breakages working directly on the SMILES string. This fast procedure is capable of computing every substructure of the molecular input set.
- Evaluation: each substructure is validated as potential SA on the training set; it is a complete match against the training structures, aimed at assessing the predictive power of each fragment.
- Rule set extraction: from the huge set of substructures collected, a reduced set of rules is extracted in the form: "IF contains <SA> THEN <apply activity label>".

The input and output to SARpy are expressed as SMILES. The output rules can be used as a predictive model simply by calling them.

4.3 QSAR and SAR Today

QSAR models can be generated using a wide variety of statistical methods and a large choice of molecular descriptors. The obtained QSAR model is usually a nonlinear relation between descriptors values and the property. If the main aim of QSAR is simply prediction, the attention should be focused on the quality of the model and not on its interpretation. Moreover it is dangerous to attempt to interpret statistical models, since correlation does not imply causality. On this basis, we can differentiate predictive QSARs, focused on prediction accuracy, from descriptive QSARs, focused on interpretability. If interpretability is an issue, SAR models are usually developed.

If the main aim of SAR and QSAR is simply prediction, the attention should be focused on model quality and not on its interpretation [10]. Regarding the interpretability of QSAR models, Livingstone [12] states: "The need for interpretability depends on the application, since a validated mathematical model relating a target property to chemical features may, in some cases, be all accurate estimates of the chemicals activity." Descriptive QSAR, however, is highly appreciated by stakeholders to characterize the toxic risk of chemicals. Structural rules are expressions that correlate local characteristics of the molecule to a risk and usually can be explained in terms of reactivity or activation of biological pathways.

In Table 2, we see the number of results obtained searching the web for the terms so far introduced.

Table 2
The number of results obtained by Google search on the terms—April 2015

Classifiers	3,600,000
Predictive modeling	2,480,000
toxicity testing method	1,650,000
Adverse outcome pathway	681,000
QSAR	635,000
In silico testing	514,000
SAR toxicity[a]	458,000
3D QSAR	445,000
2D QSAR	358,000

[a]The pure SAR acronym refers to many more technical methods

The old QSAR paradigm considered only congeneric compounds in the hypotheses that:

- Compounds in the series must be closely related.
- Same mode of action is supposed.
- Basic biological activities are investigated.
- Linear relations are constructed.

Basically today there has been a shift from some of the characteristics introduced at the beginning of this chapter toward a more complex situation. New QSAR and SAR methods, developed in the last decade, are aimed at:

- Heterogeneous compound sets.
- Mixed modes of action.
- Complex biological endpoints.
- Large number of properties.
- Non linear modeling.

4.4 Consensus Models

The development of computer programs able to contain in explicit form the knowledge about a given domain was the basis of the development of "Expert Systems" in the 1970s [23]. Soon expert systems moved from the initial rule-based representation to the modern modeling and interpretation systems. The starting "Machine Learning" community developed in the same years a way to make use of data in absence of knowledge which led to the development of Inductive Trees, well exemplified by C4.5 [24] and after by the commercial system CART.

Using different representations to reach a common agreement or a problem solution led to the idea of using computationally different methods on different problem representations, so to make use of their relative strengths. Examples are the hybrid neural and symbolic learning systems and the neuro-fuzzy system that combines connectionist and symbolic features in form of fuzzy rules. While the neural representation offers the advantage of homogeneity, distribution, and parallelization and of working with incomplete and noisy data, the symbolic representation brings the advantages of human interpretation and knowledge abstraction [25]. Independently a similar evolution in the Pattern Recognition community proposed to combine classifiers. In this area, most of the intuitions started with a seminal work about bagging classifiers [26], which opened the way to ensemble systems. Combining the predictions of a set of classifiers has shown to be an effective way to create composite classifiers that are more accurate than any of the component classifiers. There are many methods for combining the predictions given by component classifiers, as voting, combination, ensemble, and mixture of experts [27].

In the literature, we can find at least two main streams, namely "ensembles" of highly correct classifiers that disagree as much as possible, and "mixture of experts," built on the idea to train individual networks on a subtask, and then combine their predictions with a "gating" function that depends on the input. Basic combinations as majority vote or average of continuous outputs are sometimes effective. In this case, the classifiers are developed in parallel and they result combined. Finally, it is possible to use a sequential approach, so to train the final classifier using the outputs of the input classifiers as new features. In QSAR literature, they are simply called consensus models and are not yet fully exploited. Examples in QSAR are in [28, 29]; some ensemble QSAR models are also available in VEGA (http://www.vega-qsar.eu).

Why ensembles works and why they outperform single classifiers can be explained considering the error in classifiers. Usually the error is expressed [30] as:

$$\text{Error} = \text{noise} + \text{bias}^2 + \text{variance} \qquad (2)$$

where bias is the expected error of the classifier due to the fact that the classifier is not perfect; variance is the expected error due to the particular training set used, and *noise* is irreducible.

We observe that models with too few parameters can perform poorly, but the same applies to models with too many parameters. A model which is too simple, or too inflexible, will have a large bias, while a model which has too much flexibility will have high variance. Usually, the bias is a decreasing function of the complexity of the model, while variance is an increasing function of the complexity, as illustrated in Fig. 4. The concepts of bias and

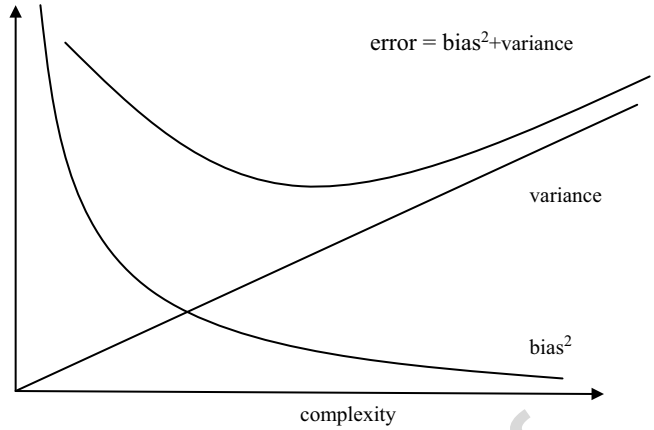


Fig. 4 The error function for different complexities of the model

variance are of help in understanding the balance between the conflicting requirements of fitting our training set accurately to obtain a good predictor. We seek a predictor sufficiently insensitive to the noise on the training data, to reduce variance, but flexible enough to approximate our model function and so minimize bias. There is a trade-off between the two components of the error and balancing them is an important part of error reduction [31, 32].

5 From Animal Models to Human and Environment Protection

Toxicity testing typically involves studying adverse health outcomes in animals administered with doses of toxicants, with subsequent extrapolation to expected human responses. The system is expensive, time consuming, low throughput, and often provides results of limited predictive value for human health. The toxicity testing methods are largely the same for industrial chemicals, pesticides, and drugs and have led to a backlog of tens of thousand chemicals to which humans are potentially exposed but whose potential toxicity remains largely unknown.

This potential risk has urged national and international organizations in making a plan for assessing the toxicity of those chemicals. In USA, for instance, EPA (Environmental Protection Agency) routinely uses predictive QSAR based on existent animal testing to authorize new chemicals. Recently in the USA, a new toxicity testing plan, "Human Toxome Project," has been launched which will make extensive experimentation using predictive, high-throughput cell-based assays (of human organs) to evaluate perturbations in key pathways of toxicity. There is no consensus about this concept of "toxicity pathway" (*see* Chapter 14) that in the opinion of many should be instead "disruption of biological pathways." The target of the project is to gain more information directly from human data, so

to check in a future, with specific experiments, the most important pathways. In the European Union, the REACH legislation for industrial chemical has been introduced together with specific regulations for cosmetics, pesticide, food additives. REACH is accepting, still with restrictions, QSAR models as well as read across [33].

The subject about regulations for human and environmental protection is out of the scope of this chapter. We only mention that different regulations apply for

- Air pollutants.
- Industrial products (e.g., REACH).
- Food.
- Drinking water.
- Cosmetics and detergents.
- Pesticides.
- Drugs.

There is only limited international agreement on the regulations and doses. In a separate chapter, we will address the issues related to the international regulations. Another chapter will be devoted to the experience in USA regulation.

Of the many open problems in assessing toxicology using in silico models we discuss about a few points. The first is the causal or mechanistic value of the QSAR equation. The QSAR for LS50, for instance, does not have a simple interpretation in term of logic sentences. This is why recent work in modeling pathways has started. Another point is about ethical issues. It is really needed to make experiments on animal? This will take us to the last point: how good a predictive model can be?

5.1 Mechanism or Causality

Hume argued that causality cannot be perceived and instead we can only perceive correlation. And indeed the basic biological experiments aim at finding a correlation (positive or negative) between some features and effect.

Discovering causal relationships in toxicology is a challenging topic. More recently studies address the so-called adverse outcome pathway (AOP) with the aim to identify the workflow from the molecular initiating event to the final outcome, as will be illustrated in Chapter 14.

Biologists want to understand why the effect can be explained in terms of metabolism, transformation substances, etc. This is often with the vague terms of "mode of action" or "mechanistic interpretation." Unfortunately there is no unique definition of mode of action: in some cases this is an observed behavior as narcosis, in other it is a supposed chemical transformation. This is more complex than considering the organic chemical transformations since they happen in an organism where different biological pathways are usually supposed.

Inferring causality from data through Bayesian Networks is today an active area of research and hopefully some answers could be automatically found using those tools [34].

5.2 Ethical Issues

Toxicity testing typically involves studying adverse health outcomes in animals subjected to high doses of toxicants with subsequent extrapolation to expected human responses at lower doses. The system is expensive, time consuming, low throughput, and often provides results of limited predictive value for human health.

Conversely each year a huge number of new substances are synthetized and possibly sent to the market. It is really necessary to test all of them on animals? Even more, it is necessary to synthetize them or would it be better to in silico assess their properties before making them, using a proactive strategy?

The Declaration of Bologna, in 1999, called the 3 R (for Reduce, Refine, and Replace), proposed a manifesto to develop alternative methods that could save millions of animals. In this scenario, the ethical issues, however, are advocated also by authorities that have to protect humans and see the animals as a more ethical use than that of humans.

The stakeholders in the toxicity assessment are:

- Scientists and producers: they want modeling of the process, discovery of properties. In other words, build knowledge and translate it rapidly in products and drugs.
- Regulators and standardization organizations: they want to be convinced by some general rule (mechanism of action). In other words, reduce the risk of erroneous evaluations. Be fast and conservative in decisions taking.
- Public, media, and opinion makers; they wants to be protected against risk at 100 %. Part of the population is strongly against the use of animal models.

As we may understand, good and validated in silico models can attract agreement from multiple actors. We have to note that for QSAR systems the output of the model, despite its good predictive value, is not sufficient; documentation enabling the user to accept or not the prediction is necessary. Some European projects, as CALEIDOS and PROSIL, are working in this direction. In the following chapters, we will address how to accept and interpret the results of a model for a large series of endpoints.

6 Conclusions

Alongside classical methods as in vivo and in vitro experiments, the use of computational tools is gaining more and more interest in the scientific community that is necessary, though it is obviously

desirable to attempt some explanation of the mechanism in chemical terms, but it is often not necessary, "per se." On this basis, we can differentiate predictive (Q)SARs, focused on prediction accuracy, from descriptive (Q)SARs, focused on descriptor interpretability. The usage of predictive QSAR models is growing, since they provide fast and reliable assessments for the benefit of the industrial world, both as accompaniment or replacement of existing techniques. For regulatory purposes, it is important to obtain satisfactory accuracy on new chemical families not well studied. In this area, it is important to develop models that can take advantage of statistical analysis on great numbers and can be further refined using cooperative methods to improve or confirm the results and give more insights into the domain [35].

7 Notes

As we have seen there are many models, many techniques, and also many reasons to build up a model. The intended use of a model can greatly affect its development. The research community working for pharmaceutical compounds with in silico methods is suing these methods to identify new active compounds, so the framework which is considered has to avoid false positives. Industry uses confidential data, often large sets of them or at least large sets of structures. The framework considered by the community of researchers and users of in silico models for toxicological endpoints is quite the opposite. The data at the basis of the models are quite limited (*see* the chapters below) with the exception of the availability of thousands of data for the Ames test. Regulators want to avoid false negatives. Regulators want to see all the documentation at the basis of the model, so the use of confidential data may represent a problem. It is important to consider the different purposes to build up the model and to be consistent with the intended use.

Acknowledgements

We kindly acknowledge the EU Life + projects CALEIDOS and PROSIL.

References

1. Brown N (2009) Chemoinformatics—an introduction for computer scientists. ACM Comput Surv 41(2), 8
2. Gasteiger J, Engel T (eds) (2003) Chemoinformatics: a textbook. Wiley-VCH, Weinheim, Germany
3. Willett P, Barnard JM, Downs GM (1998) Chemical similarity searching. J Chem Inf Comput Sci 38:983–996
4. Balaban AT (1985) Applications of graph theory in chemistry. J Chem Inf Comput Sci 25:334–343

5. Weininger D (1988) Smiles a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comput Sci 28:31–36
6. Weininger D, Weininger A, Weininger JL (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. J Chem Inf Comput Sci 29:97–101
7. Adam D (2002) Chemists synthesize a single naming system. Nature 417:369
8. Schlick T (2002) Molecular modeling and simulation: an interdisciplinary guide. Springer-Verlag, New York
9. Chow PHK, Ng RTH, Ogden BE (eds) (2008) Using animal model in biomedical research, 1st edn. World Scientific Publishing, Singapore
10. Benfenati E, Gini G (1997) Computational predictive programs (expert systems) in toxicology. Toxicology 119:213–225
11. Hartung T (2009) Toxicology for the twenty-first century. Nature 460(9):208–212
12. Livingstone DJ (2000) The characterization of chemical structures using molecular properties. A survey. J Chem Inf Comput Sci 40:195–209
13. Hansch C, Malony PP, Fujita T, Muir RM (1962) Correlation of biological activity of phenoxyacetic acids with hammett substituent constants with partition coefficients. Nature 194:178–180
14. Ghose AK, Crippen GM (1986) Atomic physicochemical parameters for three-dimensional structure directed quantitative structure-activity relationships. I Partition coefficients as a measure of hydrophobicity. J Comput Chem 7:565–577
15. Kubinyi H (2002) From narcosis to hyperspace: the history of QSAR. Quant Struct-Act Relat 21:348–356
16. Karelson M (2000) Molecular descriptors in QSAR/QSPR. Wiley-VCH, Weinheim, Germany
17. Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning: data mining, inference, and prediction. Springer-Verlag, New York, NY
18. Gini G, Katritzky A (eds) (1999) Predictive toxicology of chemicals: experiences and impact of artificial intelligence tools. Proc. AAAI Spring symposium on predictive toxicology, report SS-99-01. AAAI Press, Menlo Park, CA
19. Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model 20:269–276
20. Wolpert DH, Macready WG (1997) No free lunch theorems for optimization. IEEE Trans Evol Comput 1:67
21. Ashby J (1985) Fundamental SAs to potential carcinogenicity or noncarcinogenicity. Environ Mutagen 7:919–921
22. Ferrari T, Cattaneo D, Gini G, Golbamaki N, Manganaro A, Benfenati E (2013) Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. SAR QSAR Environ Res 24(5):365–383
23. Jackson P (1999) Introduction to expert systems, 3rd edn. Addison Wesley Longman, Boston, MA
24. Quinlan JR (1993) C4.5: programs for machine learning. Morgan Kauffman, San Francisco, CA
25. Neagu C-D, Gini G (2003) Neuro-Fuzzy knowledge integration applied to toxicity prediction, chapter 12. In: Jain R, Abraham A, Faucher C, Jan van der Zwaag B (eds) Innovations in knowledge engineering. Advanced Knowledge International Pty Ltd, Magill, South Australia, pp 311–342
26. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
27. Kittler JM, Hatef R, Duin R, Matas J (1998) On combining classifiers. IEEE Trans Pattern Anal Mach Intell 20(3):226–239
28. Gini G, Lorenzini M, Benfenati E, Brambilla R, Malvé L (2001) Mixing a symbolic and a subsymbolic expert to improve carcinogenicity prediction of aromatic compounds, LNCS 2096. Springer-Verlag, New York
29. Gini G, Craciun M, Koening C, Benfenati E (2004) Combining unsupervised and supervised artificial neural networks to predict aquatic toxicity. J Chem Inf Comput Sci 44(6):1897–1902
30. Friedman J (1997) On bias, variance, 0/1 loss and the curse of dimensionality. Data Min Knowl Discov 1:55–77
31. Gini G, Benfenati E (2007) e-modelling: foundations and cases for applying AI to life sciences. Int J Artif Intell Tools 16(2):243–268
32. Breiman L (2001) Statistical modelling: the two cultures. Stat Sci 16(3):199–231
33. Gini G, Franchi AM, Manganaro A, Golbamaki A, Benfenati E (2014) ToxRead: a tool to assist in read across and its use to assess mutagenicity of chemicals. SAR QSAR Environ Res 25(12):1–13
34. Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P (2012) Causal inference using graphical models with the R Package pcalg. J Stat Softw 47(11):1–26
35. Benfenati E, Gini G, Hoffmann S, Luttik R (2010) Comparing in vivo, in vitro and in silico methods and integrated strategies for chemical assessment: problems and prospects. ATLA Altern Lab Anim 38(2):153–166

Chapter 2

In Silico 3D Modeling of Binding Activities

Stefano Moro, Mattia Sturlese, Antonella Ciancetta, and Matteo Floris

Abstract

In silico three-dimensional (3D) molecular modeling tools based upon the receptor/enzyme–ligand docking simulation in protein crystal structures and/or homology modeling of receptors have been reliably used in pharmacological research and development for decades. Molecular docking methodologies are helpful for revealing facets of activation and inactivation, thus improving mechanistic understanding and predicting molecular ligand binding activity, and they can have a high level of accuracy, and have also been explored and applied in chemical risk assessment. This computational approach is, however, only applicable for chemical hazard identification situations where the specific target receptor for a given chemical is known and the crystal structure/homology model of the receptor is available.

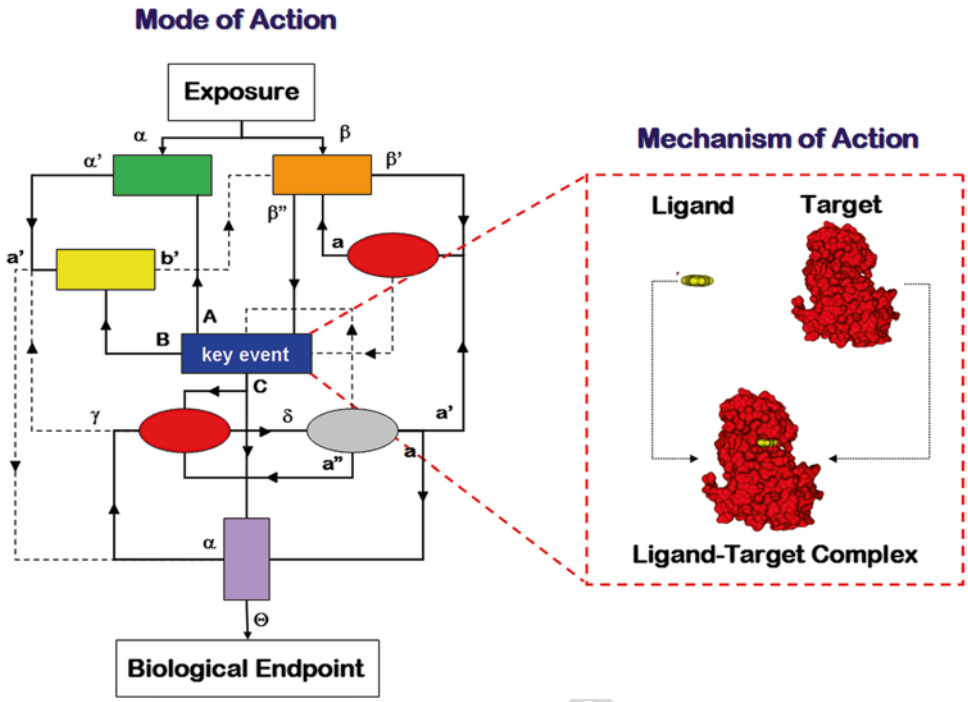
Key words Molecular modeling, Molecular docking, Scoring function, Binding affinity prediction, Chemical risk assessment

1 Introduction

Developing and evaluating predictive strategies to elucidate the mode of biological impact of environmental chemicals is a major objective of the concerted efforts of any computational toxicology program. The biological activity of any chemical compounds is based on its appropriate recognition by specific biological target, for example an enzyme or a receptor. We can define "mechanism of action" of a chemical compound as the detailed molecular description of key events in the induction of a biological response. The mechanism of action of a chemical compound is related to its "mode of action" that we can define as the description of key events and processes, starting with interaction of an agent with the cell through functional and anatomical changes, resulting in a health endpoint (Fig. 1) [1].

In principle, the rationalization of the receptor/enzyme–ligand interaction could follow a three-step process. First, an enzyme or receptor involved in a physiopathological process needs to be undoubtedly identified. Second, the structure of the enzyme or receptor needs to be solved. Finally, the structure of the ligand that binds the enzyme or the receptor must be known.

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_2, © Springer Science+Business Media New York 2016



[AU2] **Fig. 1** Flow chart elucidating the differences between "mode of action" and "mechanism of action" concepts

There is no shortage of information about the first step. Good structural information is still lacking for many interesting enzymes and receptors but, in the last two decades, an increasing number of fundamental biological targets have been solved. The genes of many enzymes and receptors have been cloned, so making it possible to obtain them in sufficient amount to experimentally determine their structure by X-ray crystallography or by nuclear magnetic resonance (NMR). The Protein Data Bank (PDB) is the open access repository where all solved structures of biopolymers are deposited [2]. Obviously, even with a good three-dimensional structure for the biological target, it is not trivial to understand where and how tightly a ligand can bind to it. A number of factors combine to make this problem an extremely challenging one:

- Will a particular ligand fit in an active/recognition site?
- What holds it in?
- How tightly do these ligands bind?
- How can a different molecule fit in the same active/recognition site?

The aim of this chapter is to describe molecular docking technologies as a potential valuable tool to identify or describe the "mechanism of action" guiding selection of test species and protocols to experimentally characterize its "mode of action" for relevant endpoints in risk assessments.

2 Methods

2.1 Molecular Docking Methodologies

Molecular docking is a computational technique aimed at the prediction of the most favorable ligand–target spatial configuration and an estimate of the corresponding complex interaction energy, although as stated at the beginning accurate scoring methods remain still elusive (Fig. 2) [3]. Docking methodologies are helpful for revealing facets of activation and inactivation, thus improving mechanistic understanding and predicting molecular ligand binding activity, and they can have a high level of accuracy, and have also been explored and applied in chemical risk assessment [4–6].

In the first step, a conformational search algorithm explores the possible ligand conformations (poses) inside the target binding pocket. In the second step, a scoring function is applied to evaluate and select the most favorable pose. In many programs, the two parts are coupled and the scoring function drives the ligand poses generation. Docking is often used to mine a database of compounds for those most likely to be active, with a ranking of the ligand molecules by the docking score, a process usually referred to as (structure-based) virtual screening [3]. Due to various possible errors in the docking or scoring process, a visual inspection of the "best" scoring hits and final selection is always needed.

2.2 Conformational Search Algorithm

Docking a ligand into a binding site needs to compute several degrees of freedom. These are the six degrees of translational and rotational freedom of one body relative to another and then the conformational degrees of freedom of the ligand and of the protein.

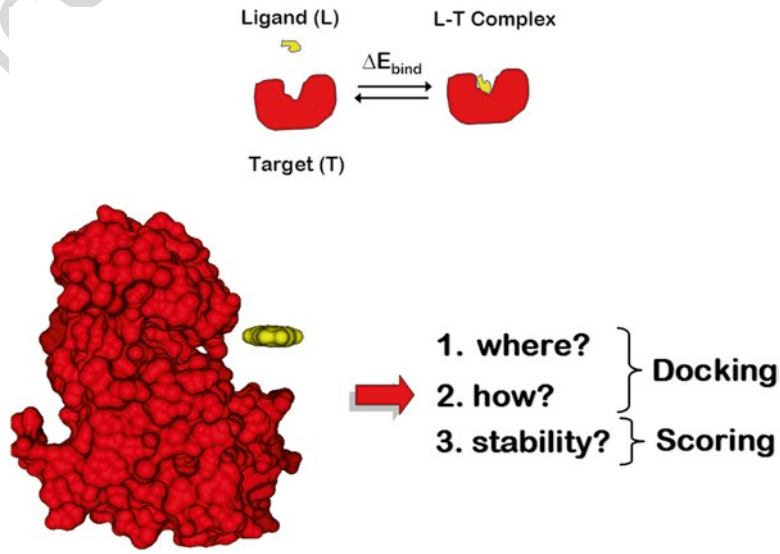


Fig. 2 Molecular docking key concepts

In the rigid docking algorithms only six degrees of freedom of the small organic molecules are considered, corresponding to translation and rotation, with both ligand and protein treated as rigid bodies. Today the standard is the semi-flexible docking, where the conformational flexibility of ligands is also taken in account while the protein is kept rigid (Fig. 3) [3]. A systematic search of all the rotatable bonds of a drug-like molecule is not efficient from a computational view point because the number of the possible combinations of the rotamers increases exponentially with the number of rotatable bonds. The search algorithms address this problem and aim to explore the conformational space of the ligands inside the protein active site in an efficient and fast fashion. In the approaches based on systematic methods, the result is exactly reproducible and the conformational space is somehow reduced and simplified [3].

Protein flexibility can be included in the protocol using "on-the-fly" generation of side-chain conformations while the protein site points are being generated or by using multiple protein conformations [3]. Such algorithms where the protein conformational space is also in part explored are called flexible docking methods. In the attempt to minimize the high computational cost generally only conformations that are close to the experimentally determined target structure are evaluated [3]. The less computationally demanding possibility is to include amino acids side-chain flexibility exploiting rotamer libraries.

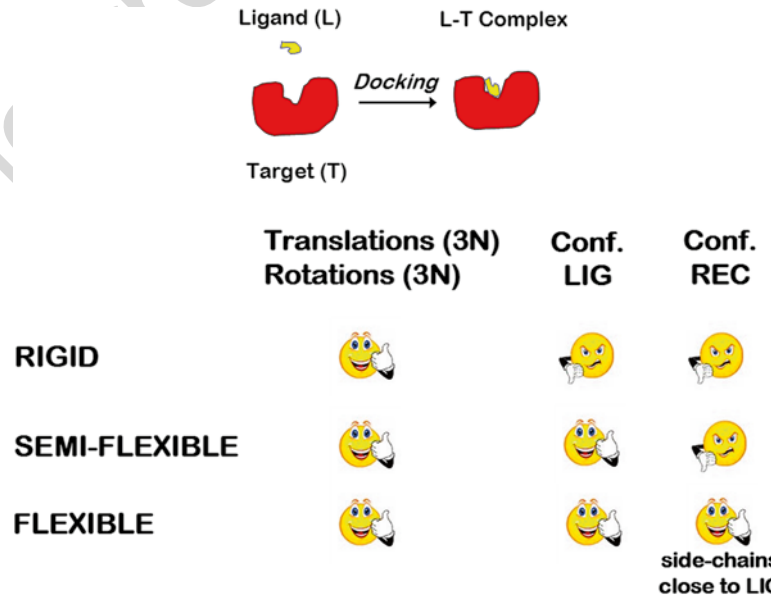


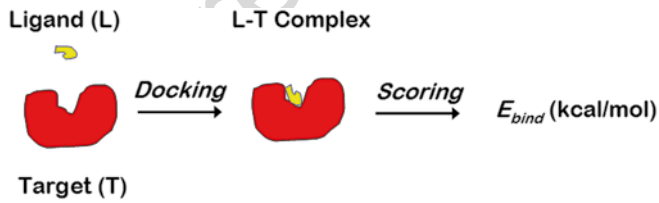
Fig. 3 Conventional classification of docking protocols referring to the exploration of ligand and/or target conformational space

2.3 Scoring Functions

Energy scoring functions are mathematical functions used to estimate the binding energy of a ligand to the protein target active pocket. Unfortunately, scoring is the weakest step in docking methodologies. In fact, in the majority of the cases, it is unable to accurately reproduce the experimental binding data [3].

Common scoring functions used in the molecular docking software simplify dramatically the thermodynamics of the binding event. The principal parts of ligand–protein interactions are taken in account to estimate in a fast way the most important energy contributions. Electrostatic and steric energy terms are generally included together with an explicit evaluation of the hydrogen bonding interaction [3]. An internal energy part could also be included, while entropy and desolvation effects are neglected. The scoring process can also be a multistep procedure composed by a first fast analysis followed by a more accurate and computational demanding rescoring phase.

Scoring functions can be grouped in three families: molecular mechanics force field, empirical, and knowledge-based scoring functions (Fig. 4) [3]. In molecular mechanics, the energy includes intra-molecular and inter-molecular contributions. Molecules are represented using force field-specific atom and bond types with atom-centered partial charges. Bond energy derives from a bond stretching, bond angle, torsion angle, and improper torsion angle energy terms. The electrostatic energy is estimated using the Coulomb equation, while for the van der



1. Molecular mechanics force field-based scoring function:

$$E_{bind} = \sum_{i=1}^{lig}\sum_{j=1}^{rec}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332\frac{q_i q_j}{\varepsilon_0 r_{ij}}\right)$$

2. Empirical scoring function:

$$E_{bind} = E_0 + E_{hb}\sum_{hb} f(\Delta R, \Delta\alpha) + E_{ionic}\sum_{ionic} f(\Delta R, \Delta\alpha) + E_{lipo}\sum_{lipo}\left|A_{lipo}\right| + E_{rot}NROT$$

3. Knowledge-based scoring function:

$$E_{bind} = \gamma\sum_{i=1}^{lig}\sum_{j=1}^{rec}E_{ij}(r) + (1-\gamma)\times\left[\sum_{i=1}^{lig}E_i(SASA, SASA_0) + \sum_{j=1}^{rec}E_j(SASA, SASA_0)\right]$$

Fig. 4 Conventional classification of the most popular scoring functions

Waals contribution the Lennard–Jones energy term is used. The AMBER [7] and OPLS [8] force fields are well parameterized for protein and small organic molecules, but the disadvantage is that they are more computationally demanding than the knowledge-based and empirical scoring functions.

Empirical scoring functions approximate the binding energy as a sum of uncorrelated energy terms. Coefficients are obtained from a regression analysis of a set of ligands with known experimental binding energy to the target and with available X-ray structures of the complex. They have the role to compensate for possible error of the energy terms used; examples are ChemScore [9], the Piecewise Linear Potential (PLP) [10], and X-Score [11]. Their accuracy depends on how well the ligand and receptor were represented in the training data used to fit the coefficients. They can be optimized for particular tasks, like binding mode prediction, ranking of a particular set of inhibitors or to study a particular target.

Knowledge-based scoring functions are composed of multiple weighted molecular features related to ligand–receptor binding modes. The features are often atom–atom distances between protein and ligand in the complex, but also the number of intermolecular hydrogen bonds or atom–atom contact energies. A large number of X-ray diffraction crystals of protein–ligand complexes are used as a knowledge base. A putative protein–ligand complex can be assessed on the basis of how similar its features are to those in the knowledge base. These contributions are summed over all pairs of atoms in the complex and the resulting score is converted into a pseudo-energy function estimating the binding affinity. The coefficients of the features can be fitted using a linear regression analysis, but also other non-linear statistical approaches can be used, like neural network, Bayesian modeling, or machine learning technique like Random Forest analysis. Examples are PMF [12], DrugScore [13], LUDI [14], and RF-Score [15]. Disadvantages with this class of scoring functions are difficulties in the evaluation of the chemical–physical meaning of the score and the risk of errors when trying to predict ligands not included in the training set [3] (*see* **Note 1**).

2.4 Physic-Based Post-Processing Scoring Methods

In the last step of the computational protocol, when the most promising ligands have been selected it is possible to further evaluate their interaction with the target with more demanding computational approaches. For example, the top ranked compounds from a virtual screening study can be rescored before the final selection is done. It is also possible to apply these techniques in a project in optimization phase to the most promising derivatives of the lead compound. There are different high-quality methods based on a rigorous physical framework, however they still have to be further evaluated to better understand the potential and limits. Additional improvements are still needed to correctly model the high complexity of the ligand binding event [16, 17].

Post-processing methods consider only the bound and unbound states of the ligand–protein complex without taking in account the intermediate states [17, 18]. This simplification sensibly reduces the computational cost compared to other physic-based methods. The free energy of binding is therefore estimated as follows:

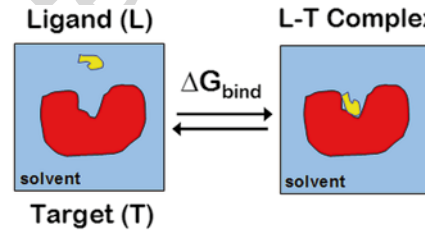
$$G_{\text{bind}} = G_{\text{complex}} - \left(G_{\text{ligand}} + G_{\text{protein}}\right)$$

The most popular post-processing method is probably the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) method [19]. In this approach, the individual energy terms are decomposed in a gas phase component calculated using the force field and a solvation energy term (Fig. 5). For ligands also an entropic contribution is included:

$$G_{\text{ligand}} = G_{\text{gas}} + G_{\text{solvation}} - \text{TS}_{\text{ligand}}$$

The electrostatic contribution to the free energy of solvation is evaluated using an implicit solvent model: the Poisson–Boltzmann equation in MM-PBSA or the generalized Born equation in MM-GBSA [20]. The hydrophobic contribution to the free energy of solvation is taken in account evaluating the solvent accessible surface area (SASA) of the molecule.

$$G_{\text{solvation}} = G_{\text{PB/GB}} + G_{\text{SASA}}$$



$$\Delta G_{bind} = \Delta G_{L-T} - \Delta G_L - \Delta G_T$$

Approximate ΔG_{bind} as:

$$\Delta G_{bind} \approx \overline{G}_{L-T} - \overline{G}_L - \overline{G}_T$$

where $\bar{G}_X$ is the calculated average free energy:

$$\overline{G}_X = \overline{E}_{MM} + G_{Solv} - TS_{MM}$$

where $\bar{E}_{MM}$ is the average potential energy:

$$\overline{E}_{MM} = \overline{E}_{bond} + \overline{E}_{angle} + \overline{E}_{tors} + \overline{E}_{vdW} + \overline{E}_{elec}$$

G_{Solv} is the calculated solvation free energy

- TS_{MM} is the solute entropy, which can be estimated by using normal-mode analysis

Fig. 5 The MM-PBSA approach represents the post-processing method to evaluate free energies of binding or to calculate absolute free energies of molecules in solution

Only in some studies the vibrational entropy is taken in account for the ligand using normal mode analysis [20] because of the high computational cost and the risk of producing large errors. The final free energy of binding is estimated comparing the energy terms of the ligand and protein alone with the complex. The approach has been used to calculate absolute and relative binding affinities with error frequently of 1 or 5 kcal/mol. Protein flexibility is taken in account using molecular dynamics (MD) simulation or a faster energy minimization protocol. MM-GBSA using simply energy minimization can evaluate one ligand per minute. Still too slow to be applied in virtual screening studies, but order of magnitude faster than MM-PBSA using molecular dynamics with accuracy sometimes comparable or even higher [20].

The linear interaction energy (LIE) method developed by Åqvist represents a plausible compromise between accuracy and computational speed in determining the free energy of binding [21]. The LIE approach is based on the assumption that the inhibitor free energy of binding to a macromolecule is linearly correlated to several energy terms that can be calculated using a molecular mechanic force field. In the original version, the LIE binding free energy is approximated using the following equation (Fig. 6) [21]:

$$G_{\text{bind}} = \alpha \; G^{\text{vdw}} + \beta \; G^{\text{el}} + \gamma$$

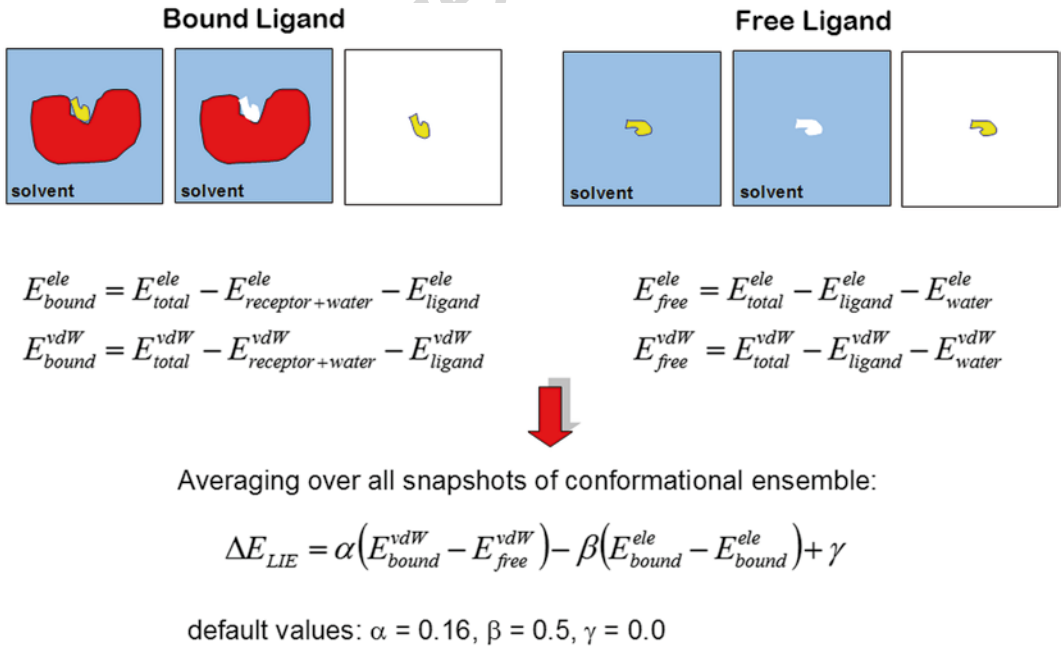


Fig. 6 Schematic depiction of procedure for the estimation of the ligand binding free energy by the LIE approach

where the van der Waals (G^{vdw}) and the electrostatic (G^{el}) interaction energy of the ligand with its surrounding environment are evaluated for the bound and unbound state. The Δ in the equation denotes that the difference between these two states is calculated. For both states, the averages of these energy contributions are computed on a population of conformations sampled by a molecular dynamics or a Monte Carlo procedure. Using a training set of molecules with known activity, a semi-empirical energy model is built by fitting the energy terms to the experimental free energy of binding. The LIE method assumes that the intra-molecular strain, entropy, and desolvation effects are embedded in this linear response, and can be cancelled out by the empirically determined scaling parameters. The constant term γ can be substituted with a third energy term containing the difference in solvent-accessible surface area of the ligand, scaled by an empirical coefficient [21]. The molecular dynamics sampling method can be substituted by simple energy minimization with a sensible decrease in the calculation times [22]. The LIE method demonstrated to result in accurate predictions of relative and absolute free energy of binding with error around 1–5 kcal/mol [22, 23].

The post-processing methods seem so far the best approaches to bridge the gap between simple docking scoring methods and more rigorous free-energy calculations to improve accuracy at a practicable computational cost.

2.5 Environmental Risk Assessment: Docking-Based Key Study

Recently, Walkers and collaborators reported a nice example of the potentiality of molecular docking to assist ecotoxicity testing in environmental risk assessment of drugs [4]. The aim of Walker's work was to evaluate whether molecular docking offers a potential tool to predict the effects of pharmaceutical compounds on non-target organisms (Fig. 7). In particular, three highly prescribed drugs such as Diclofenac, Ibuprofen, and Levonorgestrel which frequently pollute freshwater environments were selected as examples. Their primary drug targets are cyclooxygenase 2 (COX2) and progesterone receptor (PR). Molecular docking experiments were performed using these drugs and their primary drug target homologs for *Danio rerio*, *Salmo salar*, *Oncorhynchus mykiss*, *Xenopus tropicalis*, *Xenopus laevis*, and *Daphnia pulex*. The results show that fish and frog COX2 enzymes are likely to bind Diclofenac and Ibuprofen in the same way as humans but that *D. pulex* would not. Binding will probably lead to inhibition of COX function and reduced prostaglandin production. Levonorgestrel was found to bind in the same binding pocket of the progesterone receptor in frogs and fish as the human form. This suggests implications for the fecundity of fish and frogs which are exposed to Levonorgestrel. This study can be considered an interesting example in which molecular docking may provide a valuable support to anticipate the ecotoxicity profile of a drug by guiding selection of test species and protocols for relevant chronic test endpoints in environmental risk assessments.

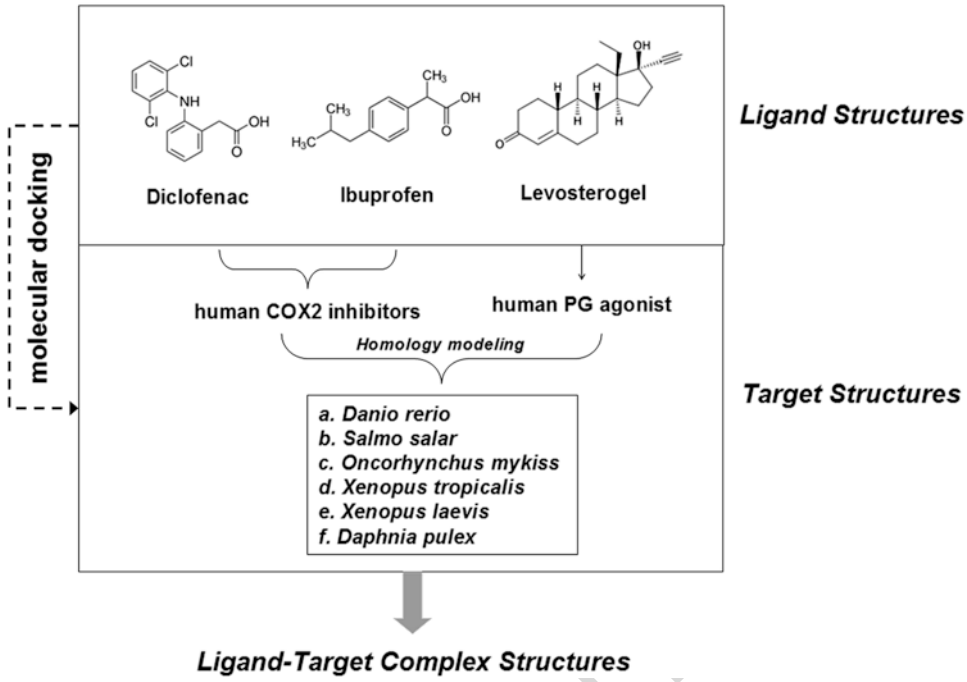


Fig. 7 Workflow of the docking-based key study for environmental risk assessment reported by Walker and collaborators [4]

3 Note

1. Modern sophisticated docking methods allow a fast evaluation of a large number of ligand poses corresponding to different conformations and orientations of the small organic molecule in the protein target. Recently also receptor flexibility has started to be considered [3]. Although the sensible improvement in the speed of calculation and efficacy of the conformational search algorithm, several limits are still challenging the predicting capability of these approaches especially affecting the scoring functions.

   The calibration of the scoring functions is generally based on the data available in the X-ray crystal structures of small organic molecules in complex with proteins. These crystal structures could include important uncertainties as a result of the subjective interpretation of the experimental electron-density map concerning in particular: (1) the identity of the isoelectric nitrogen and oxygen of the side chains of asparagine and glutamine, (2) the position of whole flexible residues, like lysine and glutamate, especially at the protein surface, or of mobile loops, (3) also ligand atoms can be ambiguous, for example the position of pyridine nitrogen of asymmetrical substituted pyridine,

(4) the identification and location of water molecules, that are often isoelectronic to common buffer constituents in crystallization media, (5) the influence of crystallization media can affect crystal morphology, but also the ligand and the active site conformation, (6) since hydrogen are not experimentally observed the ionization and tautomeric states cannot be determined and could be difficult to evaluate. The degree of confidence in the position of a particular atom or residue can be assessed using the temperature factors and examining the structure together with the electron-density map.

The data set used to calibrate the scoring function tends to be unbalanced as a consequence of the smaller number in the X-ray crystal structures of low-affinity (Ki > 1 mM) ligands compared to the high-affinity molecules. As a consequence the effects of unfavorable geometries of ligands in the protein pocket are not considered. The dipole moment of the ligand and the molecular electrostatic potential of the protein are often not included in the scoring functions. Residual flexibility of the protein or ligands is also not considered and entropic effects are often neglected. Some docking algorithms try to approximate such important contribution analyzing the number of rotatable bonds affected by the binding event. The desolvation event is roughly evaluated by the area of the interacting hydrophobic surfaces. Generally, an inadequate evaluation of the desolvation effect can result in an overestimation of the affinity of polar compounds. When protein flexibility is considered in the search algorithm especially with a minimization step the risk is that locally introduced strain is dissipated by other part of the protein to such an extent to become unrecognizable by the scoring function. Non-classical types of interactions are often neglected or not accurately evaluated: cation–π interactions, charge transfer interactions, hydrogen bonding to π-systems, halogen bonding, orthogonal dipolar alignment, dipolar antiperiplanar interactions, π-stacking, π edge-to-face contacts, and hydrogen bonding involving CH groups.

The limits of the scoring functions are the direct consequence of our incomplete understanding of the energetic contributions of individual interactions. Formulating rules is possible only within certain boundaries especially if we consider that molecular interactions behave in a highly non-additive fashion [3]. The link between thermodynamics and geometry of ligand–protein complexes still remains elusive.

Additionally, as clearly stated by Tirado-Rives and Jorgensen [24], the "window of activity" is very tiny. Thus, the free energy difference between the best ligand that one might reasonably expect to identify using virtual screening (potency, ≈50 nM) and the experimental detection limit (potency, ≈100 μM) is only about 4.5 kcal/mol. The free energy contributions due to

conformational factors alone for typical drug-like ligands (which are usually neglected in most scoring functions) can be as large as this.

In conclusion, molecular docking of potential environmental chemicals to putative macromolecular targets for toxicity provides a measure of their capacity to interact and hence is an aid in the (pre)screening process for specific modes of toxicity. These results provide a rationale for developing further, more complete testing strategies. However, because of the greater diversity of chemical space and binding affinity domains being considered and the differences in the strategic application of the results (the need to minimize false negatives), these molecular modeling strategies require additional considerations when assessing chemical hazards.

Acknowledgements

[AU5] This work was funded by the University of Padova, Italy, and the Italian Ministry for University and Research, Rome, Italy (MIUR, PRIN2008: protocol number 200834TC4L_002). S.M. participates in the European COST Action CM1207 (GLISTEN).

References

1. Simon T (2014) Environmental risk assessment: a toxicological approach. CRC, Boca Raton, p 139
2. Berman AM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data bank. Nucleic Acids Res 28:235–242
3. Bortolato A, Fanton M, Mason JS, Moro S (2013) Molecular docking methodologies. Methods Mol Biol 924:339–360
4. Walker SD, McEldowney S (2013) Molecular docking: a potential tool to aid ecotoxicity testing in environmental risk assessment of pharmaceuticals. Chemosphere 93:2568–2577
5. Zou X, Zhou X, Lin Z, Deng Z, Yin D (2013) A docking-based receptor library of antibiotics and its novel application in predicting chronic mixture toxicity for environmental risk assessment. Environ Monit Assess 185:4513–4527
6. Vedani A, Dobler M, Smieško M (2012) VirtualToxLab—a platform for estimating the toxic potential of drugs, chemicals and natural products. Toxicol Appl Pharmacol 261: 142–153
7. Case DA, Cheatham TE 3rd, Darden T, Gohlke H, Luo R, Merz KM Jr et al (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688
8. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236
9. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 11:425–445
10. Verkhivker GM (2004) Computational analysis of ligand binding dynamics at the intermolecular hot spots with the aid of simulated tempering and binding free energy calculations. J Mol Graph Model 22:335–348
11. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. J Comput Aided Mol Des 16:11–26
12. Muegge I (2006) PMF scoring revisited. J Med Chem 49:5895–5902
13. Velec HF, Gohlke H, Klebe G (2005) DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. J Med Chem 48:6296–6303

14. Bohm HJ (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 8:243–256
15. Ballester PJ, Mitchell JB (2010) A machine learning approach to predicting protein-ligand binding affinity with applications to molecular docking. Bioinformatics 26: 1169–1175
16. Huang N, Jacobson MP (2007) Physics-based methods for studying protein-ligand interactions. Curr Opin Drug Discov Devel 10: 325–331
17. Sabbadin D, Ciancetta A, Moro S (2014) Bridging molecular docking to membrane molecular dynamics to investigate GPCR-ligand recognition: the human A2A adenosine receptor as a key study. J Chem Inf Model 54:169–183
18. Sabbadin D, Moro S (2014) Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. J Chem Inf Model 54:372–376
19. Gouda H, Kuntz ID, Case DA, Kollman PA (2003) Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PBSA and thermodynamic integration methods. Biopolymers 68:16–34
20. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 10:449–461
21. Åqvist J, Medina C, Samuelsson JE (1994) A new method for predicting binding affinity in computer-aided drug design. Protein Eng 7:385–391
22. Åqvist J, Luzhkov VB, Brandsdal BO (2002) Ligand binding affinities from MD simulations. Acc Chem Res 35:358–365
23. Bortolato A, Moro S (2007) In silico binding free energy predictability by using the linear interaction energy (LIE) method: bromobenzimidazole CK2 inhibitors as a case study. J Chem Inf Model 47:572–582
24. Tirado-Rives J, Jorgensen WL (2006) Contribution of conformer focusing to the uncertainty in predicting free energies for protein-ligand binding. J Med Chem 49:5880–5884

Author Queries

Chapter No.: 2 0002663016

Queries	Details Required	Author's Response
AU1	Please approve edits made here: " We can define "mechanism of action" of a ....health endpoint"	
AU2	Please approve amended Figure 1 caption	
AU3	Can the sentence " Still too slow to be applied in virtual screening studies, but order of …" be changed to " MM-GBSA is still too slow to be applied in virtual screening studies, but has the order of .." for better readability	
AU4	Please approve edits made to the text under Acknowledgments section	

Chapter 3

Modeling Pharmacokinetics

Frederic Y. Bois and Céline Brochot

Abstract

Pharmacokinetics is the study of the fate of xenobiotics in a living organism. Physiologically based pharmacokinetic (PBPK) models provide realistic descriptions of xenobiotics' absorption, distribution, metabolism, and excretion processes. They model the body as a set of homogeneous compartments representing organs, and their parameters refer to anatomical, physiological, biochemical, and physicochemical entities. They offer a quantitative mechanistic framework to understand and simulate the time-course of the concentration of a substance in various organs and body fluids. These models are well suited for performing extrapolations inherent to toxicology and pharmacology (e.g., between species or doses) and for integrating data obtained from various sources (e.g., in vitro or in vivo experiments, structure–activity models). In this chapter, we describe the practical development and basic use of a PBPK model from model building to model simulations, through implementation with an easily accessible free software.

Key words 1,3-Butadiene, PBPK, Monte Carlo simulations, Numerical integration, *R* software

1 Introduction

The therapeutic or toxic effects of chemical substances not only depend on interactions with biomolecules at the cellular level, but also on the amount of the active substance reaching target cells (i.e., where the effects arise). Therefore, conceptually, two phases can be distinguished in the time course of such effects: the absorption, transport, and elimination of substances into, in, and out of the body including target tissues (pharmacokinetics), and their action on these targets (pharmacodynamics). Schematically, pharmacokinetics (or toxicokinetics for toxic molecules) can be defined as the action of the body on substances, and pharmacodynamics as the action of substances on the body. Pharmacokinetic and pharmacodynamics first aim at a qualitative understanding of the underlying biology. They also use mathematical models to analyze and extrapolate measurements of various biomarkers of exposure, susceptibility or effect, in order to quantitatively predict effects. This chapter focuses on toxicokinetic models and in particular on physiologically based pharmacokinetic (PBPK) models.

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_3, © Springer Science+Business Media New York 2016

Toxicokinetic models aim to link an external exposure to an internal dosimetry in humans (e.g., concentration in blood, urine, or in tissues) by describing the process of absorption, distribution, metabolism, and excretion (ADME) that undergoes a substance in living organisms. A class of toxicokinetic models, the physiologically based pharmacokinetic (PBPK) models, bases the description on the ADME processes on the physiology and the anatomy of individuals, and the biochemistry of the compounds. A PBPK model subdivides the body in compartments representing organs connected through a fluid, usually blood. Model parameters correspond to physiological and biochemical entities specific to the body and compounds, such as organ volumes, tissue blood flows, affinities of the compounds for the tissues, or the metabolic clearance.

The first works in pharmacokinetic modeling were based on physiological descriptions of the body [1–6]. However, at the time, the corresponding mathematical models were too complex to be solved. Research and applications then focused on simpler one-, two-, or three-compartment models [7], which proved to be adequate for describing and interpolating concentration–time profiles of many drugs in blood or other biological matrices. However, for substances with complex kinetics, or when inter-species extrapolations were required, simple models were insufficient and research continued on physiological models [8–12].

Over the years, the ever-increasing computing capabilities and the advent of statistical approaches applicable to uncertainty and population variability modeling have turned PBPK models into well-developed tools for safety assessment of chemical substances [13]. A significant advance has been the development of quantitative structure–properties models for the chemical-dependent parameters of PBPK models (e.g., tissue affinities) [14, 15]. Those developments are still ongoing and have led to large generic models which can give quick, even if approximate, answers to pharmacokinetic questions, solely on the basis of a chemical's formula and limited data [16–18].

The mechanistic basis of PBPK models is particularly well adapted to toxicological risk assessment [19, 20] and also in the pharmaceutical industry for the development of new therapeutic substances [21], in particular for dealing with extrapolations inherent to these domains (in vitro to in vivo, laboratory animals to human populations, various exposure or dosing schemes, etc.). PBPK models can be applied in two different steps of the risk assessment framework. First, these models can be used to better characterize the relationship between the exposure dose and the adverse effects by modeling the internal exposure in the target tissues (i.e., where the toxic effects arise) [22]. Secondly, PBPK models can be used in the exposure assessment to estimate the external exposure using human biomonitoring data, like the concentrations of chemicals in blood or urine [23, 24]. These predictions can then

be compared to existing exposure guidance or reference values such as tolerable daily intakes [25].

To provide a general overview of the basis and applications of PBPK modeling, the first section of this chapter describes the development of a PBPK model (model formulation, parameter estimation). We then propose to illustrate the different steps with 1,3-butadiene, a volatile organic compound that is carcinogenic to humans (group 1 in the IARC classification).

2 Development of a PBPK Model

In this section, we present the steps to follow in developing a PBPK model. Recently, the International Programme on Chemical Safety provided guidance on the characterization and application of PBPK models in risk assessment [20]. The guidance aimed to propose a standardized framework to review and critically evaluate the available toxicological data, and describe thoroughly the development of the model, i.e., structure, equations, parameter estimation, model evaluation, and validation. The ICRP framework also aimed to harmonize good modeling practices between risk assessors and model developers [26–28].

2.1 Principles and Model Equations

A PBPK model represents the organism of interest—human, rat, mouse, etc.—as a set of compartments, each corresponding to an organ, group of organs or tissues (e.g., adipose tissue, bone, brain, gut) having similar blood perfusion rate (or permeability) and affinity for the substance of interest. Transport of molecules between those compartments by blood, lymph, or diffusion, and further absorption, distribution, metabolism, or excretion (ADME) processes are described by mathematical equations (formally differential equations) whose structure is governed by physiology (e.g., blood flow in exit of gut goes to liver) [29, 30]. As such, PBPK modeling is an integrated approach to understand and predict the pharmacokinetic behavior of chemical substances in the body.

Drug distribution into a tissue can be rate-limited by either perfusion or permeability. Perfusion-rate-limited kinetics apply when the tissue membranes present no barrier to diffusion. Blood flow, assuming that the drug is transported mainly by blood, as is often the case, is then the limiting factor to distribution in the various cells of the body. That is usually true for small lipophilic drugs. A simple perfusion-limited PBPK model is depicted in Fig. 1. It includes the liver, well-perfused tissues (lumping brain, kidneys, and other viscera), poorly perfused tissues (muscles and skin), and fat. The organs have been grouped into those compartments under the criteria of blood perfusion rate and lipid content. Under such criteria, the liver should be lumped with the well-perfused tissues, but is left separate here as it is supposed to be the site of

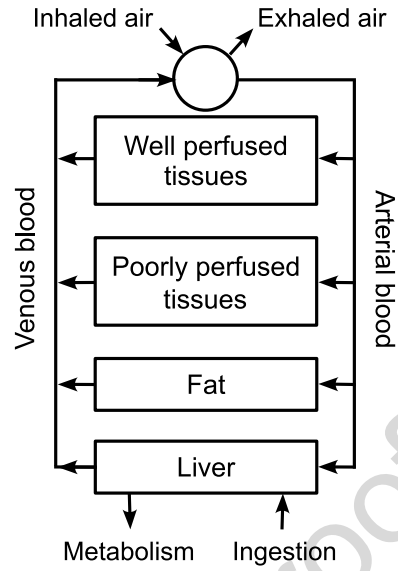


Fig. 1 Schematic representation of a simple, perfusion-limited, PBPK model. The model equations are detailed in Subheading 2 of the text

metabolism, a target effect site, and a port of entry for oral absorption (assuming that the gut is a passive absorption site which feeds into the liver via the portal vein). Bone can be excluded from the model if the substance of interest does not distribute to it. The substance is brought to each of these compartments via arterial blood. Under perfusion limitation, the instantaneous rate of entry for the quantity of drug in a compartment is simply equal to the (blood) volumetric flow rate through the organ times the incoming blood concentration. At the organ exit, the substance's venous blood concentration is assumed to be in equilibrium with the compartment concentration, with an equilibrium ratio named "partition coefficient" or "affinity constant" [30]. In the following we will note Qi the quantity of substance in compartment i, Ci the corresponding concentration, Vi the volume of compartment i, Fi the blood flow to that compartment, and PCi the corresponding tissue over blood partition coefficient. Note that all differentials are written for quantities, rather than concentrations because molecules are transported. Arguably, they are proportional to differentials for concentrations, but only if volumes are constant (and they may not be). For consistency, we strongly suggest you work with quantities. The rate of change of the quantity of substance in the poorly perfused compartment, for example, can therefore be described by the following differential equation:

$$\frac{\partial Q_{\mathrm{pp}}}{\partial t} = F_{\mathrm{pp}} \times \left(C_{\mathrm{art}} - \frac{Q_{\mathrm{pp}}}{P_{\mathrm{pp}} V_{\mathrm{pp}}} \right) \tag{1}$$

where Q_{pp} is the quantity of substance at any given time in the poorly perfused compartment, F_{pp} the blood volumetric flow rate through that group of organs, C_{art} the substance's arterial blood concentration, P_{pp} the poorly perfused tissues over blood partition coefficient, and V_{pp} the volume of the poorly perfused compartment. Since Q_{pp} kinetics are governed by a differential equation, it is part of the so-called "state variables" of the model. The tissue over blood partition coefficient P_{pp} measures the relative affinity of the substance for the tissue compared to blood. It is easy to check that, at equilibrium,

$$\frac{\partial Q_{pp}}{\partial t} = 0 \Rightarrow C_{art} - \frac{Q_{pp}}{P_{pp} V_{pp}} = 0 \Rightarrow P_{pp} = \frac{C_{pp}}{C_{art}} \quad (2)$$

if we denote by C_{pp} the concentration of the substance in the poorly perfused compartment. Similarly, for the well-perfused and the fat compartments we can write the following equations for the two state variables Q_{wp}, and Q_{fat}, respectively:

$$\frac{\partial Q_{wp}}{\partial t} = F_{wp} \times \left(C_{art} - \frac{Q_{wp}}{P_{wp} V_{wp}} \right) \quad ((3))$$

$$\frac{\partial Q_{fat}}{\partial t} = F_{fat} \times \left(C_{fat} - \frac{Q_{fat}}{P_{fat} V_{fat}} \right) \quad (4)$$

The equation for the last state variable, Q_{liv} (for the liver) is a bit more complex, with a term for metabolic clearance, with first-order rate constant k_{met}, and a term corresponding to the oral ingestion rate of the compound (quantity absorbed per unit time), R_{ing} which corresponds to the administration rate if gut absorption is complete, or to a fraction of it otherwise:

$$\frac{\partial Q_{liv}}{\partial t} = F_{liv} \left(C_{art} - \frac{Q_{liv}}{P_{liv} V_{liv}} \right) - k_{met} Q_{liv} + R_{ing} \quad (5)$$

Obviously, this is a minimal model for metabolism, and much more complex terms may be used for saturable metabolism, binding to blood proteins, multiple enzymes, metabolic interactions, extrahepatic metabolism, etc. If the substance is volatile, and if accumulation in the lung tissue itself is neglected, the arterial blood concentration C_{art} can be computed as follows, assuming instantaneous equilibrium between blood and air in the lung:

$$C_{art} = \frac{F_{pul}(1 - r_{ds}) C_{inh} + F_{tot} C_{ven}}{F_{pul}(1 - r_{ds}) / P_a + F_{tot}} \quad (6)$$

where F_{tot} is the blood flow to the lung, F_{pul} the pulmonary ventilation rate, r_{ds} the fraction of dead space (upper airways' volume unavailable for blood-air exchange) in the lung, P_a the blood over

air partition coefficient, and C_{inh} is the concentration inhaled. Equation 6 can be derived from a simple balance of mass exchanges between blood and air under equilibrium conditions. C_{ven} is the concentration of compound in venous blood and can be obtained as the sum of compound concentrations in venous blood at the organ exits weighted by corresponding blood flows:

$$C_{ven} = \frac{\sum_{x\in\{pp,wp,fat,liv\}} \left(\frac{F_x Q_x}{P_x V_x} \right)}{F_{pp} + F_{wp} + F_{fat} + F_{liv}} \tag{7}$$

Finally, the substance's concentration in exhaled air, C_{exh}, can be obtained under the same equilibrium conditions as for Eq. 6:

$$C_{exh} = (1 - r_{ds}) \frac{C_{art}}{P_a} + r_{ds} C_{inh} \tag{8}$$

Note that C_{art}, C_{ven}, and C_{exh}, are not specified by differential equations, but by algebraic equations. Those three variables are not fundamental in our model and could be expressed using only parameters and state variables. They are just (very) convenient "output variables" that we may want to record during simulation and that facilitate model writing.

The above model assumes that all the substance present in blood is available for exchange with tissues. This may not be true if a fraction of the substance is bound, for example to proteins, in blood or tissues. In that case it is often assumed that binding/unbinding is rapid compared to the other processes. The equations are then written in terms of unbound quantities and the rapid equilibrium assumption is used to keep track of the balance bound/unbound quantity in each organ or tissue [30].

Diffusion across vascular barriers or cellular membranes can be slower than perfusion. This condition is likely to be met by large polar molecules. In that case, to account for diffusion limitation, a vascular sub-compartment is usually added to each organ or tissue of interest. Diffusion between that vascular sub-compartment and the rest of the tissue is modeled using the Fick's law. A diffusion barrier can also exist between the extracellular and intracellular compartments. Consequently, PBPK models exhibit very different degrees of complexity, depending on the number of compartments used and their eventual subdivisions [31].

2.2 Parameter Estimation

A PBPK model needs a considerable amount of information to parameterize. At the system level, we find substance-independent anatomical (e.g., organ volume), physiological (e.g., cardiac output), and some biochemical parameters (e.g., enzyme concentrations). All those are generic, in the sense that they do not depend on the substance(s) of interest, and are relatively well documented in

humans and laboratory animals [29, 32–36]. They can be assigned once for ever, at least in first approximation, for an "average" individual in a given species at a given time.

There are also, inevitably, substance-specific parameters which reflect the specific interactions between the body and the substance of interest. In many cases, values for those parameters are not readily available. However, such parameters often depend, at least in part, on the physicochemical characteristics of molecule studied (e.g., partition coefficients depend on lipophilicity, passive renal clearance depends on molecular weight). In that case, they can be estimated, for example by quantitative structure–activity relationships (QSARs) [37, 38], also referred to as quantitative structure–property relationships (QSPRs) when "simple" parameter values are predicted. Molecular simulation (quantum chemistry) models can also be used [39, 40], in particular for the difficult problem of metabolic parameters' estimation. QSARs are statistical models (often a regression) relating one or more parameters describing chemical structure (predictors) to a quantitative measure of a property or activity (here a parameter value in a PBPK model) [15, 41–44]. However, when predictive structure–property models are not available (as is often the case with metabolism, for example), the parameters have to be measured in vitro (for an extensive review *see* [45, 46]) or estimated from in vivo experiments and are much more difficult to obtain.

However, using average parameter values does not correctly reflect the range of responses expected in a human population, nor the uncertainty about the value obtained by QSARs, in vitro experiments or in vivo estimation [47]. Inter-individual variability in PK can have direct consequences on efficacy and toxicity, especially for substances with a narrow therapeutic window. Therefore, simulation of inter-individual variability should be an integral part of the prediction of PK in humans. The mechanistic framework of PBPK models provides the capacity of predicting inter-individual variability in PK when the required information is adequately incorporated. To that effect, two modeling strategies have been developed in parallel: The first approach has been mostly used for data-rich substances. It couples a pharmacokinetic model to describe time-series measurements at the individual level and a multilevel (random effect) statistical model to extract a posteriori estimates of variability from a group of subjects [48, 49]. In a Bayesian context, a PBPK model can be used at the individual level, and allows easy inclusion of many subject-specific covariates [50]. The second approach also takes advantage of the predictive capacity of PBPK models but simply assigns a priori distributions to the model parameters (e.g., metabolic parameters, blood flows, organ volumes) and forms distributions of model predictions by Monte Carlo simulations [51].

2.3 Solving the Model Equations

Many software programs can actually be used to build and simulate a PBPK model. Some are very general simulation platforms—*R* [52], GNU MCSim [53, 54], Octave [55], Scilab [56], Matlab® [57], Mathematica® [58], to name a few. Those platforms usually propose some PBPK-specific packages or functionalities that ease model development. An alternative is to use specialized software (e.g., PK-Sim® [59], Simcyp® [60], GastroPlus® [61], Merlin-expo [62]), which has often an attractive interface. However, in that case the model equations cannot usually be modified and only the parameter values can be changed or assigned pre-set values or distributions.

2.4 Evaluation of the Model

The evaluation (checking) of the model is an integral part of its development to objectively demonstrate the reliability and relevance of the model. Model evaluation is often associated with a defined purpose, such as a measure of internal dosimetry relevant to the mode of action of the substance (e.g., the area under the curve or maximal concentration in the target tissues during critical time windows). The objective here is to establish confidence in the predictive capabilities of the model for a few key variables. A common way to evaluate a model's predictability is to confront its predictions to an independent data set, i.e., that has not been used for model development. That is called cross-validation in statistical jargon. For example, the evaluation step could check that the model is able to reproduce the peaks and troughs of tissue concentrations under repeated exposure scenarios. Model evaluation is not limited to a confrontation between model predictions and data, but also requires checking the plausibility of the model structure, its parameterization and the mathematical correctness of equations (e.g., the conservation of mass, organ volumes, and blood flows). Because of their mechanistic description of ADME processes, PBPK model structures and parameter values must be in accordance with biological reality. Parameter values inconsistent with physiological and biological knowledge limit the use of the model for extrapolation to other exposure scenarios, and ultimately need to be corrected by the acquisition of new data, for example.

2.5 Model Validation and Validity Domain

Most models are valid only on a defined domain. That is true even for the most fundamental models in physics. The term "validation" is rarely used in the context of toxicokinetic modeling as it is almost impossible to validate in all generality a model of the whole body. Actually, it is not done because it is bound to fail. It would require experimental data for all state variables (time evolution of concentration in all compartments) and model parameters under innumerable exposure scenarios. In that context, to be useful, the validation process should first define a validity domain. For example, we should not expect PBPK models to give accurate descriptions of within-organ differences in concentrations (organs are

described as homogeneous "boxes"). There is actually an avenue of research for improved organ descriptions. As far as time scale is concerned, we are doing pretty well for long-term [17], but for inhalation at the lung level in particular, PBPK models are not suitable for time scales lower than a couple of minutes (the cyclicity of breathing is not described). Metabolism and the description of metabolites distribution is a deeper problem, as it branches on the open-ended field of systems biology [63]. In that area the domain of validity becomes harder to define and is usually much smaller than that of the parent molecule. The model's domain of validity should be documented, to the extent possible, and even more carefully as we venture into original and exotic applications. Fortunately, the assumptions consciously made during model development usually help in delineating the domain of validity.

3 A PBPK Model for 1,3-Butadiene

In this section, we propose to apply the model development process presented above to the development of a PBPK model for 1,3-butadiene, a volatile organic compound. First, some background information on 1,3-butadiene will be provided to fulfill some requirements of the guidance defined by the International Programme on Chemical Safety [20]. Because the aim here is not to run a risk assessment on butadiene, most sections of the guidance will be omitted (e.g., the comparison with the default approaches).

3.1 Setting Up Background

An extensive literature exists on 1,3-butadiene human uses, exposures, toxicokinetics, and mode of action, *see* for example [64, 65].

1,3-Butadiene (CAS No. 106-99-0) is a colorless gas under normal conditions. It is used for production of synthetic rubber, thermoplastic resins and other plastics, and is also found in cigarette smoke and combustion engine fumes. It enters the environment from engine exhaust emissions, biomass combustion, and from industrial on-site uses. The highest atmospheric concentrations have been measured in cities and close to industrial sources. The general population is exposed to 1,3-butadiene primarily through ambient and indoor air. Tobacco smoke may contribute significant amounts of 1,3-butadiene at the individual level. It is a known carcinogen, acting through its metabolites [65].

1,3-Butadiene metabolism is a complex series of oxidation and reduction steps [65]. Briefly, the first step in the metabolic conversion of butadiene is the cytochrome P450-mediated oxidation to 1,2-epoxy-3-butene (EB). EB may subsequently be exhaled, conjugated with glutathione, further oxidized to 1,2:3,4-diepoxybutane (DEB), or hydrolyzed to 3-butene-1,2-diol (BDD). DEB can then be hydrolyzed to 3,4-epoxy-1,2-butanediol (EBD) or conjugated with glutathione. BDD can be further oxidized to EBD.

EBD can be hydrolyzed or conjugated with glutathione. The metabolism for 1,3-butadiene to EB is the rate-limiting step for the formation of all its toxic epoxy metabolites. It makes sense, given the above, to define the cumulated amount of 1,3-butadiene metabolites formed in the body as the measure of its internal dose for cancer risk assessment purposes.

3.2 Model Development and Evaluation

3.2.1 Software Choice

In our butadiene example, we will use the *R* software and its package deSolve. We will assume that the reader has a minimal working of knowledge of *R* and has *R* and deSolve installed. *R* is freely available for the major operating systems (Unix/Linux, Windows, Mac OS) and deSolve provides excellent functions for integrating differential equations. *R* is easy to use, but not particularly fast. If you need to run many simulations (say several thousands or more) you should code your model in C language, compile it, and have deSolve call your compiled code (see the deSolve manual for that). An even faster alternative (if you need to do Bayesian model calibration, for example) is to use *GNU MCSim*. You can actually use *GNU MCSim* to develop C code for deSolve.

3.2.2 Defining the Model Structure and Equations

Our research group has previously developed and published a PBPK model for 1,3-butadiene on the basis of data collected on 133 human volunteers during controlled low dose exposures. We used it for various studies and as an example of Bayesian PBPK analysis [66–68]. That model (*see* Fig. 2) is a minimal description of butadiene distribution and metabolism in the human body after inhalation. Three compartments lump together tissues with similar perfusion rate (blood flow per unit of tissue mass): the "well-perfused" compartment regroups the liver, brain, lungs,

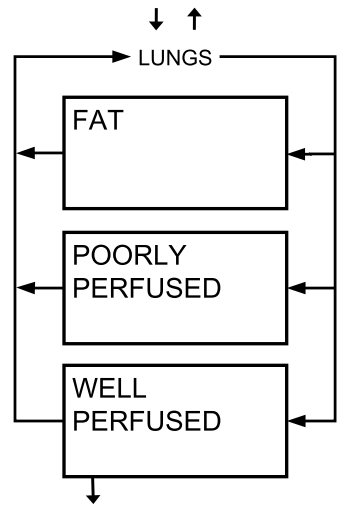


Fig. 2 Representation of the PBPK model used for 1,3-butadiene. The model equations and parameters are detailed in Subheading 3 of the text

kidneys, and other viscera; the "poorly perfused" compartment lumps muscles and skin; and the third is "fat" tissues. Butadiene can be metabolized into an epoxide in the liver, kidneys, and lung, which are part of the well-perfused compartment. Our model will therefore include four essential "state" variables, which will each have a governing differential equation: the quantities of butadiene in the fat, in the well-perfused compartment, in the poorly perfused compartment, and the quantity metabolized. Actually, the latter is a "terminal" state variable which depends on the others state variables and has no dependent. We could dispense with it if we did not want to compute and output it. That would save computation time, which grows approximately with the square of the number of state variables of the model.

In an *R* script code for use with deSolve, we first need to define the model state variable and assign them initial values (values they will take at the start of a simulation, those are called "boundary conditions" in technical jargon). The syntax is quite simple (the full script is given in Appendix):

```
      y = c("Q_fat" = 0,   # Quantity of butadiene in fat (mg)
    "Q_wp"  = 0,   # ~       in well-perfused (mg)
    "Q_pp"  = 0,   # ~       in poorly-perfused (mg)
    "Q_met" = 0)   # ~       metabolized (mg)
```

That requests the creation of y as a vector of four named components, all initialized here at the value zero (i.e., we assume no previous exposure to butadiene, or no significant levels of butadiene in the body in case of a previous exposure). The portions of lines starting with the pound sign (#) are simply comments for the reader and are ignored by the software. We have chosen milligrams as the unit for butadiene quantities and it is useful to indicate it here. In *R* indentation and spacing do not matter and we strive for readability.

We then need to define similarly, as a named vector, the model parameters:

```
      parameters = c(
    "BDM"    = 73,        # Body mass (kg)
    "Height" = 1.6,       # Body height (m)
    "Age"    = 40,        # in years
    "Sex"    = 1,         # code 1 is male, 2 is female
    "Flow_pul"      = 5,    # Pulmonary ventilation rate (L/min)
    "Pct_Deadspace" = 0.7,  # Fraction of pulmonary deadspace
    "Vent_Perf"     = 1.14, # Ventilation over perfusion ratio
    "Pct_LBDM_wp"   = 0.2,  # wp tissue as fraction of lean mass
    "Pct_Flow_fat"  = 0.1,  # Fraction of cardiac output to fat
    "Pct_Flow_pp"   = 0.35, # ~                    to pp
    "PC_art" = 2,         # Blood/air partition coefficient
    "PC_fat" = 22,        # Fat/blood ~
    "PC_wp"  = 0.8,       # wp/blood  ~
    "PC_pp"  = 0.8,       # pp/blood  ~
```

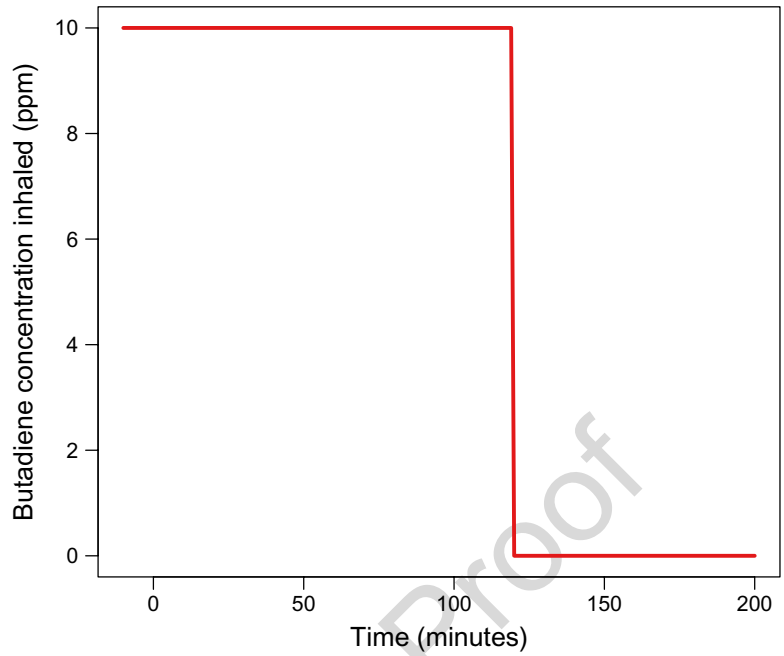


Fig. 3 Plot of the time–concentration profile of butadiene inhaled generated by the function $C_{inh}(t)$ of the example script. $C_{inh}(t)$ is used as a forcing function for the model simulations


```
"Kmetwp" = 0.25)     # Rate constant for metabolism (1/min)
```

We will see next how those parameters are used in the model equations, but you notice already that they are not exactly, except for the partition coefficients and metabolic rate constant, the parameters used in Eqs. 1, and 3–5. They are in fact scaling coefficients used to model parameter correlations in an actual subject.

Before we get to the model core equations, we need to define the value of the concentration of butadiene in inhaled air. This is an "input" to the model and we will allow it to change with time, so it is a dynamic boundary condition to the model (deSolve uses the term "forcing function"). We use here a convenient feature of *R*, defining C_{inh} as an approximating function.

```
C_inh = approxfun(x = c(0, 100), y = c(10, 0),
                  method = "constant", f = 0, rule = 2)
```

The instruction above defines a function of time $C_{inh}(t)$, right continuous (option $f = 0$) and constant by segments (the option method = "linear" would yield a function linear by segments). At times 0 and 100 (x values), it takes values y 10 and then 0, respectively. Before time zero and after time 100, $C_{inh}(t)$ will take the closest y value defined (option rule = 2). Figure 3 shows the behavior of the function $C_{inh}(t)$ so defined.

Formally you do not necessarily need such an input function in your model. C_{inh} could simply be a constant, or no input could be used if you were to model just the elimination of butadiene out of

body following exposure. Indeed, the initial values of the state variables would have to be non-null in that case.

Now we need to define a function that will compute the derivatives at the core of the model, as a function of time t—used for example when parameters are time varying, or for computing $C_{inh}(t)$, of the current state variable values y, and of the parameters. Here is the (simplified) code of that function which we called "bd.model" (intermediate calculations have been deleted for clarity, we will see them later):

```
bd.model = function(t, y, parameters) { # function header
 # function body:
 with (as.list(y), {
  with (as.list(parameters), {
  # … (part of the code omitted for now)
  # Time derivatives for quantities
  dQ_fat = Flow_fat * (C_art - Cout_fat)
  dQ_wp  = Flow_wp  * (C_art - Cout_wp) - dQmet_wp
  dQ_pp  = Flow_pp  * (C_art - Cout_pp)
  dQ_met = dQmet_wp;
     return(list(c(dQ_fat, dQ_wp, dQ_pp, dQ_met),    # derivatives
           c("C_ven" = C_ven, "C_art" = C_art)))      # extra outputs
  }) # end with parameters
 }) # end with y
} # end of function bd.model()
```

The first two "with" nested blocks (they extend up to the end of the function) are an obscure but useful feature of *R*. Remember that y and "parameters" are arrays with named components. In *R*, you should refer to their individual components by writing for example "parameters["PC_fat"]" for the fat over blood partition coefficient. That can become clumsy and the "with" statements allow you to simplify the notation and call simply "PC_fat".

The most important part of the "bd.model" function is the calculation of the derivatives. As you can see they are given an arbitrary name and computed similarly to the equations given above (e.g., Eq. 1). Obviously we need to have defined the temporary variables "Cout_fat", "Cout_wp", and "dQmet_wp" but they are part of the omitted code and we will see them next. Finally, the function needs to return (as a list, that is imposed by deSolve) the derivatives computed and eventually the output variables we might be interested in (in our case, for example C_{ven} and C_{art}).

The code we omitted for clarity was simply intermediate calculations. First some obvious conversion factors:

```
# Define some useful constants
 MW_bu = 54.0914    # butadiene molecular weight (in grams)
```

```
ppm_per_mM = 24450 # ppm to mM under normal conditions
# Conversions from/to ppm
ppm_per_mg_per_l = ppm_per_mM / MW_bu
mg_per_l_per_ppm = 1 / ppm_per_mg_per_l
```

The following instructions scale the compartment volumes to body mass. The equation for the fraction of fat is taken from [69]. That way, the volumes correlate as they should to body mass or lean body mass:

```
# Calculate fraction of body fat
Pct_BDM_fat = (1.2 * BDM / (Height * Height) - 10.8 *(2 - Sex) +
                0.23 * Age - 5.4) * 0.01
# Actual volumes, 10% of body mass (bones…) receive no butadiene
Eff_V_fat = Pct_BDM_fat * BDM
Eff_V_wp = Pct_LBDM_wp * BDM * (1 - Pct_BDM_fat)
Eff_V_pp = 0.9 * BDM - Eff_V_fat - Eff_V_wp
```

The blood flows are scaled similarly to maintain adequate perfusion per unit mass:

```
# Calculate alveolar flow from total pulmonary flow
Flow_alv = Flow_pul * (1 - Pct_Deadspace)
# Calculate total blood flow from Flow_alv and the V/P ratio
Flow_tot = Flow_alv / Vent_Perf
# Calculate actual blood flows from total flow and percent flows
Flow_fat = Pct_Flow_fat * Flow_tot
Flow_pp = Pct_Flow_pp * Flow_tot
Flow_wp = Flow_tot * (1 - Pct_Flow_pp - Pct_Flow_fat)
```

We have now everything needed to compute concentrations at time t in the various compartments or at their exit:

```
# Calculate the concentrations
C_fat = Q_fat / Eff_V_fat
C_wp = Q_wp / Eff_V_wp
C_pp = Q_pp / Eff_V_pp
# Venous blood concentrations at the organ exit
Cout_fat = C_fat / PC_fat
Cout_wp = C_wp / PC_wp
Cout_pp = C_pp / PC_pp
```

The next two lines are typical computational tricks. The right-hand sides will be used several times in the subsequent calculations. It is faster, and more readable to define them as temporary variables:

```
# Sum of Flow * Concentration for all compartments
dQ_ven = Flow_fat * Cout_fat + Flow_wp * Cout_wp +
         Flow_pp * Cout_pp
```

```
                # Quantity metabolized in liver (included in
well-perfused)
        dQmet_wp = Kmetwp * Q_wp
        C_inh.current = C_inh(t) # to avoid calling C_inh() twice
```

The last series of intermediate computations obtain C_{art}—as in Eq. 6, with a unit conversion for $C_{inh}(t)$, C_{ven} as in Eq. 7 (those two will be defined as outputs in the function's return statement), the alveolar air concentration C_{alv}, and finally the exhaled air concentration C_{exh}:

```
        # Arterial blood concentration
        # Convert input given in ppm to mg/l to match other units
        C_art = (Flow_alv * C_inh.current * mg_per_l_per_ppm +
dQ_ven) /
                (Flow_tot + Flow_alv / PC_art)
        # Venous blood concentration (mg/L)
        C_ven = dQ_ven / Flow_tot
        # Alveolar air concentration (mg/L)
        C_alv = C_art / PC_art
        # Exhaled air concentration (ppm!)
        if (C_alv <= 0) {
          C_exh = 10E-30 # avoid round off errors
        } else {
         C_exh = (1 - Pct_Deadspace) * C_alv * ppm_per_mg_per_l
+
                Pct_Deadspace * C_inh.current
        }
```

The calculation of C_{exh} just above is an example of computational trick to avoid rounding errors (useful if you later want to take the log of C_{exh}, you want to avoid values like -7×10^{-16} for example). It also illustrates one idiosyncrasy of *R*: spacing and disposition do not matter *except* that "} else {" must be on the same line.

3.2.3 Running the Model

The *R* script we detailed above is almost ready to perform simulations. We just need to define the output times (times at which we will want to look at the results, here a sequence from zero to 1440 min, every 10 min), load the deSolve library (so far we have only used standard *R* functions) and call the integration routine "ode", storing its results in the variable "results":

```
        # Define the computation output times (minutes)
        times = seq(from=0, to=1440, by=10)
        # Call the ODE solver
        library(deSolve)
        results = ode(times = times, func = bd.model, y = Y, parms
= parms)
```

By default, deSolve uses the *lsode* integration routine for stiff systems [70]. This is a very efficient solver, but you have the choice of several integrators (*see* the deSolve manual for details).

The content of results can be looked at, saved to a file, further manipulated or simply plotted:

```
# results is basically a table
results
# Plot the results of the simulation
plot(results)
```

Figure 4 shows the plot obtained (just for the four butadiene quantities state variables). That is in essence all it takes to write and simulate a PBPK model.

3.2.4 Running Monte Carlo Simulations

Running Monte Carlo simulations in *R*, for uncertainty or sensitivity analyses [49], is rather easy. *R* is fundamentally a statistical software and is well equipped for random numbers generation. The skeleton for a Monte Carlo simulation script is simply a loop of *n* iterations:

```
for (iteration in 1:1000) { # 1000 Monte Carlo simulations
# Sample randomly some parameters
...
# Reduce output times eventually
times = c(0, 1440)
```

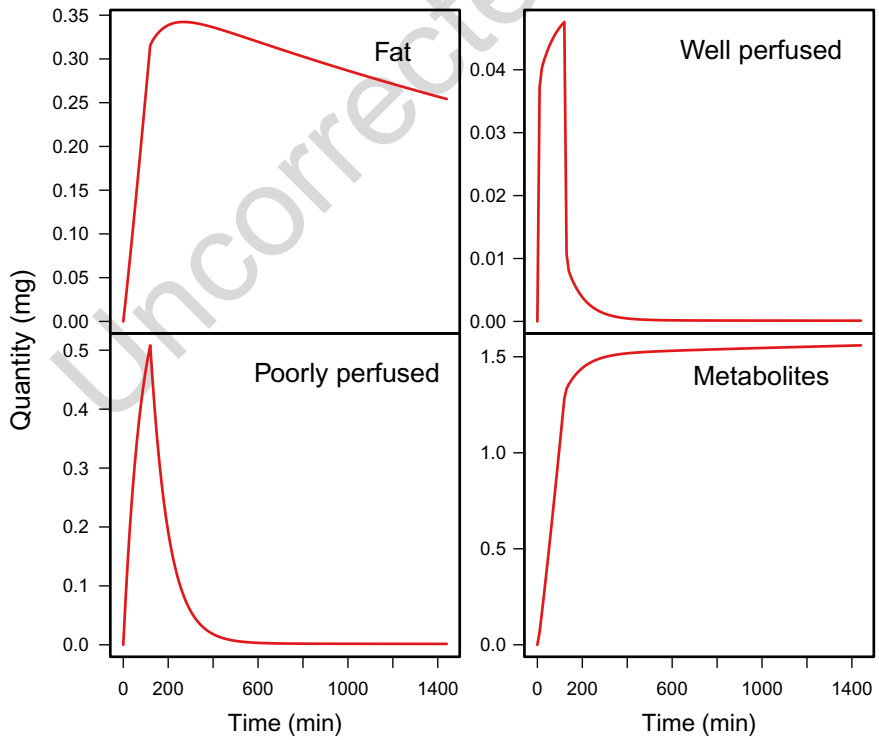



Fig. 4 Simulated time courses of the quantities of butadiene in the compartments of the sample PBPK model. Inhalation exposure was specified as shown in Fig. 3

```
  # Integrate
 tmp = ode(times = times, func = bd.model, y = y,
          parms = parameters)
  # Accumulate results in a table
 …
        } # end Monte Carlo loop
```

Here too the ellipsis (…) refers to pieces of code we will detail below. The full script is given in Appendix). The calculations inside the "for" loop are performed a thousand time. At each iteration, new parameter values are randomly sampled. For example, if we choose to sample only four parameters (we could sample all) from normal distributions, the code would look like:

```
  # Sample randomly some parameters
  parameters["BDM"]      = rnorm(1, 73,   7.3)
  parameters["Flow_pul"] = rnorm(1, 5,    0.5)
  parameters["PC_art"]   = rnorm(1, 2,    0.2)
  parameters["Kmetwp"]   = rnorm(1, 0.25, 0.025)
```

For each parameter, one normal random variable is drawn with a mean set to the value used in the simple script above, and a standard deviation equal to 10 % of the mean. When doing Monte Carlo simulations, you usually do not want to look at the distributions of state or output variables at thousands of different times (that is heavy). Here we decided to look at them only at time 1440 min, so we reset the times array. Note that the starting time (here zero) still needs to be defined among the times. The integrator is then called and its results stored in the "tmp" table. But that is only one set of results in a thousand and we need to accumulate those results. The following few lines of code show how to keep only the results obtained at time 1440 (line 2 or the tmp table) but without the output time (which is always 1440) (the "-1" in "tmp[2,-1]" removes the first column). It is also very useful to store the sampled parameter values:

```
      if (iteration == 1) { # initialize
       results = tmp[2,-1]
                  sampled.parms = c(parameters["BDM"],
parameters["Flow_pul"],
                   parameters["PC_art"], parameters["Kmetwp"])
      } else { # accumulate
       results = rbind(results, tmp[2,-1])
       sampled.parms = rbind(sampled.parms,
                                        c(parameters["BDM"],
parameters["Flow_pul"],
                  parameters["PC_art"], parameters["Kmetwp"]))
      }
```

When the Monte Carlo loop is finished we probably want to save the accumulated results in a file (unless the simulations are very fast to compute):

```
        # Save the results
```

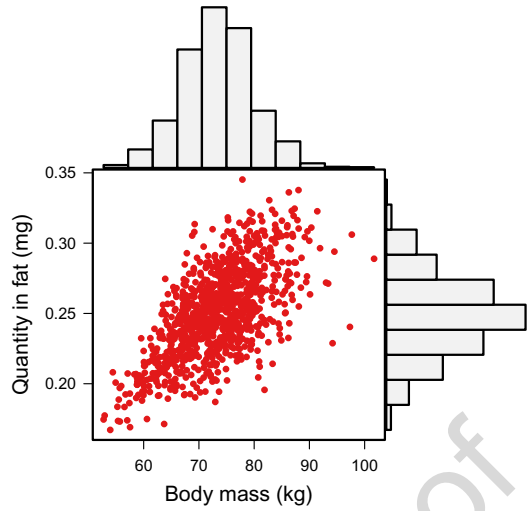


Fig. 5 Illustration of the PBPK model Monte Carlo simulation results. The dot plot shows the quantity of butadiene in fat after 1 day as a function of sampled body mass. The random sampling of other parameters explains the dispersion of the results, however the quantity in fat is clearly sensitive to body mass. The marginal histograms show the distributions of the sampled values for body mass and of the predicted quantities of butadiene in fat. A sizeable uncertainty affects those results

```
save(sampled.parms, results, file="MTC.dat.xz",
compress = "xz")
# use load(file="MTC.dat.xz") to read them back in
```

Finally, such large amounts of information are best handled with statistical and graphical methods. Figure 5 shows a nicer version of the three simple plots which would be produced by the following lines:

```
# Plot the results
hist(sampled.parms[,1])
hist(results[,1])
plot(sampled.parms[,1], results[,1])
```

Figure 5 shows the relationship between the Monte Carlo sampled body mass values and the resulting prediction for the quantity in fat after a day. You can observe an obvious and expected correlation between the two (butadiene storage in fat increases with the fat compartment volume which in turn increases with body mass). The increase in butadiene storage is roughly proportional to body mass, so that is a sensitive parameter. The relationship is not perfect because three other parameters were sampled. We can that way study the sensitivity of any model prediction, at any time, with respect to any model parameter [49]. The plot also shows the marginal distributions of body masses and butadiene quantities in fat. The uncertainty attached to predictions is about ±50 %. That type of histogram can give an idea of the reliability of any model prediction.

A thousand Monte Carlo simulations took us a few minutes on a laptop computer. A thousand is actually a small number if you want to accurately characterize upper or lower percentiles of the resulting distributions. If computation time becomes an issue you can divide it by a factor 10 if you compile your model in C—*GNU MCSim* [53, 54] can actually produce a C code compatible with deSolve without having to learn the C language. A factor 100 can be gained if you work only with *GNU MCSim*.

4 Conclusion

PBPK modeling is more and more used in research, development, and regulation [71, 72]. Obviously, the precision and accuracy of PBPK model will be only as good as those of the QSAR predictions or in vitro data used to set their parameters. Quality assurance of those components is therefore an important issue [26, 73], and we have seen that in several areas (metabolism in particular), research work is still needed. As to the models themselves, their validity will probably be easier to check if they are generic and with a stable and well-documented structure [74]. This requirement, however, runs somewhat contrary to the next challenge: Coupling PBPK models to predictive pharmacology or toxicity models, both at the cellular level and at the organ level [75]. We hope however, that this step-by-step introduction to PBPK model development and simulation will help the reader in his/her first steps into that exciting area.

Acknowledgment

This work was partly supported by the European Commission, 7th FP project 4-FUN (grant agreement 308440) and the French Ministry for the Environment (Programme 190 toxicologie).

5 Appendix

R script for the butadiene PBPK model:

```
    #=============================================
==================
    # Butadiene human PBPK model
    # Define and initialize the state variables
    y = c("Q_fat" = 0,   # Quantity of butadiene in fat (mg)
          "Q_wp"  = 0,   # ~       in well-perfused (mg)
          "Q_pp"  = 0,   # ~       in poorly-perfused (mg)
          "Q_met" = 0)   # ~       metabolized (mg)
    # Define the model parameters
    # Units:
```

```
# Volumes: liter
# Time:    minute
# Flows:   liter / minute
parameters = c(
  "BDM"    = 73,         # Body mass (kg)
  "Height" = 1.6,        # Body height (m)
  "Age"    = 40,         # in years
  "Sex"    = 1,          # code 1 is male, 2 is female
  "Flow_pul"      = 5,   # Pulmonary ventilation rate (L/min)
  "Pct_Deadspace" = 0.7, # Fraction of pulmonary deadspace
  "Vent_Perf"     = 1.14, # Ventilation over perfusion ratio
  "Pct_LBDM_wp"   = 0.2, # wp tissue as fraction of lean mass
  "Pct_Flow_fat"  = 0.1, # Fraction of cardiac output to fat
  "Pct_Flow_pp"   = 0.35, # ~                    to pp
  "PC_art" = 2,          # Blood/air partition coefficient
  "PC_fat" = 22,         # Fat/blood ~
  "PC_wp"  = 0.8,        # wp/blood ~
  "PC_pp"  = 0.8,        # pp/blood ~
  "Kmetwp" = 0.25)       # Rate constant for metabolism (1/min)
# The input air concentration (in parts per million) can vary with time
C_inh = approxfun(x = c(0,120), y = c(10,0), method="constant", f=0, rule=2)
# Check the input concentration profile just defined
plot(C_inh(1:300), xlab = "Time (min)",
     ylab = "Butadiene air concentration (ppm)", type = "l")
# Define the model equations
bd.model = function(t, y, parameters) {
 with (as.list(y), {
 with (as.list(parameters), {
 # Define some useful constants
 MW_bu = 54.0914   # butadiene molecular weight (in grams)
   ppm_per_mM = 24450 # ppm to mM under normal conditions
 # Conversions from/to ppm
 ppm_per_mg_per_l = ppm_per_mM / MW_bu
 mg_per_l_per_ppm = 1 / ppm_per_mg_per_l
 # Calculate Flow_alv from total pulmonary flow
 Flow_alv = Flow_pul * (1 - Pct_Deadspace)
 # Calculate total blood flow from Flow_alv and the V/P ratio
 Flow_tot = Flow_alv / Vent_Perf
 # Calculate fraction of body fat
  Pct_BDM_fat = (1.2 * BDM / (Height * Height) - 10.8 *(2 - Sex) +
                 0.23 * Age - 5.4) * 0.01
  # Actual volumes, 10% of body mass (bones…) get no butadiene
```

```
    Eff_V_fat = Pct_BDM_fat * BDM
    Eff_V_wp  = Pct_LBDM_wp  * BDM * (1 - Pct_BDM_fat)
    Eff_V_pp  = 0.9 * BDM - Eff_V_fat - Eff_V_wp
    # Calculate actual blood flows from total flow and percent
flows
    Flow_fat = Pct_Flow_fat * Flow_tot
    Flow_pp  = Pct_Flow_pp  * Flow_tot
    Flow_wp  = Flow_tot * (1 - Pct_Flow_pp - Pct_Flow_fat)
    # Calculate the concentrations
    C_fat = Q_fat / Eff_V_fat
    C_wp  = Q_wp  / Eff_V_wp
    C_pp  = Q_pp  / Eff_V_pp
    # Venous blood concentrations at the organ exit
    Cout_fat = C_fat / PC_fat
    Cout_wp  = C_wp  / PC_wp
    Cout_pp  = C_pp  / PC_pp
    # Sum of Flow * Concentration for all compartments
    dQ_ven = Flow_fat * Cout_fat + Flow_wp * Cout_wp +
Flow_pp * Cout_pp
    C_inh.current = C_inh(t) # to avoid calling C_inh() twice
    # Arterial blood concentration
    # Convert input given in ppm to mg/l to match other units
    C_art = (Flow_alv * C_inh.current * mg_per_l_per_ppm +
dQ_ven) /
          (Flow_tot + Flow_alv / PC_art)
    # Venous blood concentration (mg/L)
    C_ven = dQ_ven / Flow_tot
    # Alveolar air concentration (mg/L)
    C_alv = C_art / PC_art
    # Exhaled air concentration (ppm!)
    if (C_alv <= 0) {
     C_exh = 10E-30 # avoid round off errors
    } else {
     C_exh = (1 - Pct_Deadspace) * C_alv * ppm_per_mg_per_l +
          Pct_Deadspace * C_inh.current
    }
    # Quantity metabolized in liver (included in well-perfused)
    dQmet_wp = Kmetwp * Q_wp
    # Differentials for quantities
    dQ_fat = Flow_fat * (C_art - Cout_fat)
    dQ_wp  = Flow_wp  * (C_art - Cout_wp) - dQmet_wp
    dQ_pp  = Flow_pp  * (C_art - Cout_pp)
    dQ_met = dQmet_wp
    # The function bd.model must return at least the derivatives
    list(c(dQ_fat, dQ_wp, dQ_pp, dQ_met),     # derivatives
      c("C_ven" = C_ven, "C_art" = C_art)) # extra outputs
    }) # end with parameters
  }) # end with y
```

```
} # end bd.model
# Define the computation output times
times = seq(from=0, to=1440, by=10)
# Call the ODE solver
library(deSolve)
results = ode(times = times, func = bd.model, y = y, parms =
parameters)
# results is basically a table
results
# Plot the results of the simulation
plot(results)
# End
# End Simple Simulation.
#=============================================
==================
#=============================================
==================
# Monte Carlo simulations
# We assume that a simple simulation has already been run, so
that
# y, parameters, C_inh, and bd.model have all been defined
and that
# deSolve has been loaded.
for (iteration in 1:1000) { # 1000 Monte Carlo simulations…
 # Sample randomly some parameters
 parameters["BDM"]     = rnorm(1, 73,   7.3)
 parameters["Flow_pul"] = rnorm(1, 5,    0.5)
 parameters["PC_art"]   = rnorm(1, 2,    0.2)
 parameters["Kmetwp"]   = rnorm(1, 0.25, 0.025)
 # Reduce output times eventually. We only care about time
1440,
 # but time zero still needs to be specified
 times = c(0, 1440)
 # Integrate
 tmp = ode(times = times, func = bd.model, y = y, parms =
parameters)
 if (iteration == 1) { # initialize
  results = tmp[2,-1]
            sampled.parms      =      c(parameters["BDM"],
parameters["Flow_pul"],
             parameters["PC_art"], parameters["Kmetwp"])
 } else { # accumulate
  results = rbind(results, tmp[2,-1])
  sampled.parms = rbind(sampled.parms,
                                      c(parameters["BDM"],
parameters["Flow_pul"],
            parameters["PC_art"], parameters["Kmetwp"]))
 }
```

```
    } # end Monte Carlo loop
    # Save the results, specially if they took a long time to
compute
    save(sampled.parms, results, file="MTC.dat.xz", compress
= "xz")
    # use load(file="MTC.dat.xz") to read them back in
    # Plot the results
    hist(sampled.parms[,1])
    hist(results[,1])
    plot(sampled.parms[,1], results[,1])
    # End Monte Carlo Simulations.
    #=========================================
==================
```

References

1. Haggard HW (1924) The absorption, distribution, and elimination of ethyl ether. I. The amount of ether absorbed in relation to the concentration inhaled and its fate in the body. J Biol Chem 59:737–751
2. Haggard HW (1924) The absorption, distribution, and elimination of ethyl ether. II. Analysis of the mechanism of absorption and elimination of such a gas or vapor as ethyl ether. J Biol Chem 59:753–770
3. Haggard HW (1924) The absorption, distribution, and elimination of ethyl ether. III. The relation of the concentration of ether, or any similar volatile substance, in the central nervous system to the concentration in the arterial blood, and the buffer action of the body. J Biol Chem 59:771–781
4. Haggard HW (1924) The absorption, distribution, and elimination of ethyl ether. IV. The anesthetic tension of ether and the physiological response to various concentrations. J Biol Chem 59:783–793
5. Haggard HW (1924) The absorption, distribution, and elimination of ethyl ether. V. The importance of the volume of breathing during the induction and termination of ether anesthesia. J Biol Chem 59:795–802
6. Teorell T (1937) Kinetics of distribution of substances administered to the body. Arch Int Pharmacodyn Ther 57:205–240
7. Gibaldi M, Perrier D (1982) Pharmacokinetics. Marcel Dekker, New York
8. Bischoff KB, Dedrick RL, Zaharko DS et al (1971) Methotrexate pharmacokinetics. J Pharm Sci 60:1128–1133
9. Dedrick RL, Forrester DD, Cannon JN et al (1973) Pharmacokinetics of 1-beta-D-arabinofuranosylcytosine (ARA-C) deamination in several species. Biochem Pharmacol 22: 2405–2417
10. Gerlowski LE, Jain RK (1983) Physiologically based pharmacokinetic modeling: principles and applications. J Pharm Sci 72:1103–1127
11. Droz PO, Guillemin MP (1983) Human styrene exposure—V. Development of a model for biological monitoring. Int Arch Occup Environ Health 53:19–36
12. Lutz RJ, Dedrick RL, Tuey D et al (1984) Comparison of the pharmacokinetics of several polychlorinated biphenyls in mouse, rat, dog, and monkey by means of a physiological pharmacokinetic model. Drug Metab Dispos 12:527–535
13. Rowland M, Peck C, Tucker G (2011) Physiologically-based pharmacokinetics in drug development and regulatory science. Annu Rev Pharmacol Toxicol 51:45–73
14. Nestorov I, Aarons L, Rowland M (1998) Quantitative structure-pharmacokinetics relationships: II. A mechanistically based model to evaluate the relationship between tissue distribution parameters and compound lipophilicity. J Pharmacokinet Biopharm 26:521–545
15. Peyret T, Poulin P, Krishnan K (2010) A unified algorithm for predicting partition coefficients for PBPK modeling of drugs and environmental chemicals. Toxicol Appl Pharmacol 249:197–207
16. Jones HM, Parrott N, Jorga K et al (2006) A novel strategy for physiologically based predictions of human pharmacokinetics. Clin Pharmacokinet 45:511–542
17. Beaudouin R, Micallef S, Brochot C (2010) A stochastic whole-body physiologically based pharmacokinetic model to assess the impact of inter-individual variability on tissue dosimetry

over the human lifespan. Regul Toxicol Pharmacol 57:103–116

18. Clewell HJ III, Gentry PR, Covington TR et al (2004) Evaluation of the potential impact of age- and gender-specific pharmacokinetic differences on tissue dosimetry. Toxicol Sci 79: 381–393
19. United States Environmental Protection Agency (US EPA) (2006) Approaches for the application of Physiologically Based Pharmacokinetic (PBPK) models and supporting data in risk assessment (final report). United States Environmental Protection Agency, Washington, DC
20. International Programme on Chemical Safety (IPCS) (2010) Characterization and application of physiologically based pharmacokinetic models in risk assessment, World Health Organization. International Programme on Chemical Safety, Geneva
21. Peters SA (2011) Physiologically Based Pharmacokinetic (pbpk) modeling and simulations: principles, methods, and applications in the pharmaceutical industry. Wiley, Hoboken, NJ
22. Andersen ME (1995) What do we mean by … dose? Inhal Toxicol 7:909–915
23. Clewell HJ, Tan YM, Campbell JL et al (2008) Quantitative interpretation of human biomonitoring data. Toxicol Appl Pharmacol 231:122–133
24. Ulaszewska MM, Ciffroy P, Tahraoui F et al (2012) Interpreting PCB levels in breast milk using a physiologically based pharmacokinetic model to reconstruct the dynamic exposure of Italian women. J Expo Sci Environ Epidemiol 22:601–609
25. Zeman FA, Boudet C, Tack K et al (2013) Exposure assessment of phthalates in French pregnant women: results of the ELFE pilot study. Int J Hyg Environ Health 216:271–279
26. Loizou G, Spendiff M, Barton HA et al (2008) Development of good modelling practice for physiologically based pharmacokinetic models for use in risk assessment: the first steps. Regul Toxicol Pharmacol 50:400–411
27. Barton HA, Chiu WA, Setzer W et al (2007) Characterizing uncertainty and variability in physiologically-based pharmacokinetic (PBPK) models: state of the science and needs for research and implementation. Toxicol Sci 99:395–402
28. Andersen ME (1995) Development of physiologically based pharmacokinetic and physiologically based pharmacodynamic models for applications in toxicology and risk assessment. Toxicol Lett 79:35–44
29. Clewell RA, Clewell HJ (2008) Development and specification of physiologically based pharmacokinetic models for use in risk assessment. Regul Toxicol Pharmacol 50:129–143
30. Campbell JL Jr, Clewell RA, Gentry PR et al (2012) Physiologically based pharmacokinetic/toxicokinetic modeling. In: Reisfeld B, Mayeno AN (eds) Computational toxicology, Methods in molecular biology series. Humana, New York, pp 439–499
31. Thompson MD, Beard DA (2011) Development of appropriate equations for physiologically based pharmacokinetic modeling of permeability-limited and flow-limited transport. J Pharmacokinet Pharmacodyn 38:405–421
32. International Commission on Radiological Protection (ICRP) (1975) Report of the Task Group on Reference Man—a report prepared by a Task Group of Committee 2 of the International Commission on Radiological Protection. Pergamon, Oxford
33. International Commission on Radiological Protection (ICRP) (2002) Basic anatomical and physiological data for use in radiological protection: reference values. ICRP Publication 89. Ann ICRP 32(3–4):5–265
34. Davies B, Morris T (1993) Physiological parameters in laboratory animals and humans. Pharm Res 10:1093–1095
35. Brown RP, Delp MD, Lindstedt ST et al (1997) Physiological parameter values for physiologically based pharmacokinetic models. Toxicol Ind Health 14:407–484
36. Young JF, Branham WS, Sheenan DM et al (1997) Physiological "constants" for PBPK models for pregnancy. J Toxicol Environ Health 52:385–401
37. Ekins S, Waller CL, Swaan PW et al (2000) Progress in predicting human ADME parameters in silico. J Pharmacol Toxicol Methods 44:251–272
38. Poulin P, Haddad S (2012) Advancing prediction of tissue distribution and volume of distribution of highly lipophilic compounds from a simplified tissue-composition-based model as a mechanistic animal alternative method. J Pharm Sci 101:2250–2261
39. Butina D, Segall MD, Frankcombe K (2002) Predicting ADME properties in silico: methods and models. Drug Discov Today 7:S83–S88
40. Buch I, Giorgino T, De Fabritiis G (2011) Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. Proc Natl Acad Sci U S A 108(25): 10184–10189
41. Fiserova-Bergerova V, Diaz ML (1986) Determination and prediction of tissue-gas

partition coefficients. Int Arch Occup Environ Health 58:75–87

42. Gargas ML, Seybold PG, Andersen ME (1988) Modeling the tissue solubilities and metabolic rate constant (Vmax) of halogenated methanes, ethanes, and ethylenes. Toxicol Lett 43:235–256
43. Poulin P, Theil FP (2000) A priori prediction of tissue: plasma partition coefficients of drugs to facilitate the use of physiologically-based pharmacokinetic models in drug discovery. J Pharm Sci 89:16–35
44. Poulin P, Ekins S, Theil FP (2011) A hybrid approach to advancing quantitative prediction of tissue distribution of basic drugs in human. Toxicol Appl Pharmacol 250:194–212
45. Adler S, Basketter D, Creton S et al (2011) Alternative (non-animal) methods for cosmetics testing: current status and future prospects—2010. Arch Toxicol 85:367–485
46. Bal-Price A, Jennings P (eds) (2014) In vitro toxicology systems. Humana, New York
47. Bois FY, Jamei M (2012) Population-based pharmacokinetic modeling and simulation. In: Lyubimov AV (ed) Encyclopedia of drug metabolism and interactions. Wiley, Hoboken, pp 1–27
48. Yuh L, Beal S, Davidian M et al (1994) Population pharmacokinetic/pharmacodynamic methodology and applications: a bibliography. Biometrics 50:566–575
49. Bernillon P, Bois FY (2000) Statistical issues in toxicokinetic modeling: a Bayesian perspective. Environ Health Perspect 108(Suppl 5):883–893
50. Gelman A, Bois FY, Jiang J (1996) Physiological pharmacokinetic analysis using population modeling and informative prior distributions. J Am Stat Assoc 91:1400–1412
51. Bois FY, Jamei M, Clewell HJ (2010) PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. Toxicology 278:256–267
52. R Development Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
53. Bois FY (2009) GNU MCSim: Bayesian statistical inference for SBML-coded systems biology models. Bioinformatics 25:1453–1454
54. Anonymous. GNU MCSim. http://www.gnu.org/software/mcsim/
55. Anonymous. GNU Octave. http://www.gnu.org/software/octave/
56. Anonymous Scilab. http://www.scilab.org/
57. Anonymous. MATLAB and Simulink for technical computing—MathWorks. http://mathworks.com/
58. Anonymous. Wolfram mathematica: definitive system for modern technical computing. http://www.wolfram.com/mathematica/
59. Anonymous. Bayer technology services: PK-Sim®. http://www.systems-biology.com/products/pk-sim.html
60. Anonymous. Simcyp—population based pharmacokinetic modelling and simulation. http://www.simcyp.com/
61. Anonymous. Simulations Plus, Inc|Modeling & Simulation Software|Consulting Services for Pharmaceutical Research. http://www.simulations-plus.com/
62. Anonymous. MERLIN-Expo|Exposure Assessment Software. http://merlin-expo.4funproject.eu/
63. Bois FY (2009) Physiologically-based modelling and prediction of drug interactions. Basic Clin Pharmacol Toxicol 106:154–161
64. United States Environmental Protection Agency (US EPA) (2002) Health Assessment of 1,3-Butadiene. United States Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, Washington Office, Washington, DC
65. Kirman CR, Albertini RJ, Sweeney LM et al (2010) 1,3-Butadiene: I. Review of metabolism and the implications to human health risk assessment. Crit Rev Toxicol 40:1–11
66. Bois FY, Smith T, Gelman A et al (1999) Optimal design for a study of butadiene toxicokinetics in humans. Toxicol Sci 49: 213–224
67. Mezzetti M, Ibrahim JG, Bois FY et al (2003) A Bayesian compartmental model for the evaluation of 1,3-butadiene metabolism. J R Stat Soc Ser C 52:291–305
68. Bois F (2012) Bayesian inference. In: Reisfeld B, Mayeno AN (eds) Computational toxicology, vol II. Humana, New York, pp 597–636
69. Deurenberg P, Weststrate JA, Seidell JC (1991) Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. Br J Nutr 65:105–141
70. Byrne GD, Hindmarsh AC (1987) Stiff ODE solvers: a review of current and coming attractions. J Comput Phys 70:1–62
71. Caldwell JC, Evans MV, Krishnan K (2012) Cutting edge PBPK models and analyses: providing the basis for future modeling efforts and bridges to emerging toxicology paradigms. J Toxicol 2012:852384
72. Zhao P, Zhang L, Grillo JA et al (2011) Applications of Physiologically Based Pharmacokinetic (PBPK) modeling and simu-

lation during regulatory review. Clin Pharmacol Ther 89:259–267

73. Zhao P, Rowland M, Huang S-M (2012) Best practice in the use of physiologically based pharmacokinetic modeling and simulation to address clinical pharmacology regulatory questions. Clin Pharmacol Ther 92:17–20
74. Rostami-Hodjegan A, Tamai I, Pang KS (2012) Physiologically based pharmacokinetic (PBPK) modeling: it is here to stay! Biopharm Drug Dispos 33:47–50
75. Geenen S, Yates JW, Kenna JG et al (2013) Multiscale modelling approach combining a kinetic model of glutathione metabolism with PBPK models of paracetamol and the potential glutathione-depletion biomarkers ophthalmic acid and 5-oxoproline in humans and rats. Integr Biol 5:877–888

Chapter 4

Modeling ADMET

Jayeeta Ghosh, Michael S. Lawless, Marvin Waldman, Vijay Gombar, and Robert Fraczkiewicz

Abstract

Drug discovery and development is a costly and time-consuming endeavor (Calcoen et al. Nat Rev Drug Discov 14(3):161–162, 2015; The truly staggering cost of inventing new drugs. Forbes. http://www.forbes.com/sites/matthewherper/2012/02/10/the-truly-staggering-cost-of-inventing-new-drugs/, 2012; Scannell et al. Nat Rev Drug Discov 11(3):191–200, 2012). Over the last two decades, computational tools and in silico models to predict ADMET (Adsorption, Distribution, Metabolism, Excretion, and Toxicity) profiles of molecules have been incorporated into the drug discovery process mainly in an effort to avoid late-stage failures due to poor pharmacokinetics and toxicity. It is now widely recognized that ADMET issues should be addressed as early as possible in drug discovery. Here, we describe in detail how ADMET models can be developed and applied using a commercially available package, ADMET Predictor™ 7.2 (ADMET Predictor v7.2. Simulations Plus, Inc., Lancaster, CA, USA).

Key words ADMET, Adsorption, Distribution, Metabolism

1 Introduction

ADMET profiling of molecules consists of two steps: the first involves building quantitative structure–property relationship (QSPR) models for desired ADMET endpoints and the second step involves using those QSPR models to predict the modeled endpoints for compounds of interest. Since these models require merely drawing a molecule's structure for making predictions, even virtual chemical libraries can be scored or ranked based on ADMET liabilities. This in silico profiling can help progress only those molecules along the discovery chain that are less likely to fail later in the drug discovery process. This may may positively impact the very high costs [1–3] of drug discovery and development. QSPR models can also be used to guide structural modifications to improve ADMET properties [4].

In addition to the in silico models developed inside the firewalls of many companies, a number of free and commercial software

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_4,

packages are available for building and applying ADMET models. A representative, but not exhaustive, list of such software packages is given in Tables 1 and 2. Similar lists are available elsewhere [5].

In this chapter, we describe how ADMET Predictor can be used for estimating crucial physicochemical and biological properties for large numbers of compounds during virtual library screening in early drug discovery. The simplest application of ADMET Predictor is to profile the library for a single property, but profiling only on one property carries with it the dangers of "one-dimensional

Table 1
Free predictive ADMET software

ADMET software	Predicted ADMET properties	Link
CAESAR/VEGA	Bioconcentration factor, skin sensitization, mutagenicity, developmental toxicity, carcinogenicity, aquatic toxicity	http://www.caesar-project.eu/index.php?page=links http://www.caesar-project.eu/ http://www.vega-qsar-eu/
Chem Prop (OSIRIS EDITION)	Solubility, log *P*, air/water, octanol/air, melting point, boiling point, vapor pressure, soil sorption, human toxicology	http://www.ufz.de/index.php?en=6738
EPI Suite™	Melting point, boiling point, vapor pressure, water solubility, log *P*, p*K*a, aquatic toxicity	http://www.epq.gov/oppt/exposure/pubs/episuitedl.htm
Lazar	Mutagenicity, repeated dose toxicity, carcinogenicity, fathead minnow acute toxicity	http://lazar.in-silico.de
OpenTox platform: ToxPredict	log *P*, p*K*a, reproductive toxicity, carcinogenicity	http://apps.ideaconsult.net:8080/ToxPredict
OSIRIS property explorer	Solubility, cLog *P*, toxicity risk assessment, mutagenicity, reproductive toxicity, carcinogenicity	http://www.organic-chemistry.org/prog/peo/
SMARTCyp	Metabolism	http://www.farma.ku.dk/smartcyp/index.php
T.E.S.T.	Boiling point, flash point, surface tension, viscosity, density, water solubility, thermal conductivity, vapor pressure, melting point, mutagenicity, acute toxicity, aquatic toxicity	http://www.epa.gov/
ToxTree	Skin irritation, eye irritation, mutagenicity, carcinogenicity	http://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxtree
VCCLAB	Solubility, log *P*	http://www.vcclab.org/

Table 2
Commercial predictive ADMET software

ADMET software	Predicted ADMET properties	Link
ACD/Percepta	Abraham solvation parameters, aqueous solubility, boiling point, log *D*, log *P*, p*K*a, blood–brain barrier permeation, CYP inhibitors and substrates, distribution, max. recommended daily dose, oral bioavailability, P-gp specificity, PK, regioselectivity of metabolism, acute toxicity, aquatic toxicity, endocrine system disruption, mutagenicity, health effects, hERG inhibition, skin irritation	http://www.acdlabs.com/products/percepta/percepta/predictors.php
ARChem SPARC	p*K*a, log *D*, boiling point, vapor pressure, diffusion coefficient (air and water), solubility, log *P*, Henry's constant	http://archemcalc.com/
BioByte Bio-Loom	CLOGP and CMR	http://www.biobyte.com/bb/prod/bioloom.html
Biovia Discovery Studio and TOPKAT	log *P*, solubility, human intestinal absorption, blood–brain barrier penetration, CYP2D6 binding, hepatotoxicity, mutagenicity, carcinogenicity, developmental toxicity, rat oral LD50, rat inhalation toxicity, rat chronic LOAEL, skin irritation, eye irritation, skin sensitization, fathead minnow LC50, daphnia magna EC50	http://acceleys.com/products/discovery-studio/qsar-admet-and-predictive-toxicology.html
BMDRC PreADMET	Caco-2, MDCK, blood–brain barrier, human intestinal absorption, plasma protein binding, skin permeability, Ames test, and rodent carcinogenicity	http://preadment.bmdrc.org/
ChemAxon Calculator Plugins	log *P*, log *D*, p*K*a, solubility	http://chemaxon.com/products/calculator-plugins
ChemDBsoft SLIPPER	log *D*, solubility, permeability, fraction absorbed	http://www.chemdsoft.com/slipper-logp-logc-logd-logsw-fa.html
Compudrug Pallas System	log *P*, log *D*, p*K*a, metabolism, toxicity	http://www.compudrug.com/pallas_system
Fujitsu ADMEWORKS	Solubility, log *P*, Lead-likeness, human intestinal absorption, blood–brain barrier, P-gp transporter, 2D6 Km, 3A4 Ki, 3A4 Km, 3A4 inhibition, carcinogenicity, mutagenicity, chromosomal aberration, hERG inhibition, skin sensitization, biodegradability, bioconcentration factor	http://www.fqs.pl/Chemistry_Materials_Life_Science/products/admeworks_predictor
Genexplain PASS	Overview of biological activities, pharmacotherapeutic effects, toxicity, metabolism, gene regulation, and transport-related activities	http://www.genexplain.com/pass

(continued)

Table 2
(continued)

ADMET software	Predicted ADMET properties	Link
Leadscope	Genetic toxicity, carcinogenicity, reproductive toxicity, developmental toxicity, neurotoxicity, adverse hepatobiliary effects, adverse urinary tract effects, adverse cardiological effects	http://www.leadscope.com
Lhasa Ltd. Derek, Meteor and Sarah	Irritation, skin sensitization, mutagenicity, genotoxicity, teratogenicity, reproductive toxicity, carcinogenicity, metabolism	http://www.lhasalimited.org/products/
MoKa, VolSurf, MetaSite	log *P*, log *D*, p*K*a, solubility, metabolism	http://www.moldiscovery.com/
MolCode Toolbox	Water solubility, log *P*, eye irritation, skin sensitization, mutagenicity, repeated dose toxicity, reproductive toxicity, biodegradability	http://reachqsar.com/
Multicase CASE Ultra	Mutagenicity, hepatotoxicity, skin and eye toxicity, fetal survival, transporters, renal toxicity, teratogenicity, carcinogenicity, acute toxicity, reproductive toxicity	http://www.multicase.com/case-ultra-models
Optibrium StarDrop™	log *P*, log *D* at pH 7.4, aqueous solubility, PBS solubility at pH 7.4, human intestinal absorption, blood–brain barrier penetration, CYP2C9 and 2D6 affinity, P-gp transport, hERG IC50, plasma protein binding, P450 metabolism	http://www.optibrium.com/stardrop-features.php
Schrodinger QikProp	log *P*, log *D*, solubility, blood–brain barrier, Caco-2 and MDCK permeability, hERG, human serum albumin, water/gas partition coefficient	http://www.schrodinger.com/QikProp/
Simulations Plus' ADMET Predictor	p*K*a, log *P*, human jejunal permeability, skin permeability, solubility, blood–brain barrier penetration, volume of distribution, plasma protein binding, P-gp substrate and inhibition, OATP1B1 inhibition, CYP metabolism and inhibition, UGT substrate, estrogen and androgen receptor binding, skin sensitivity, respiratory sensitivity, aquatic toxicity, hERG inhibition, rat and mouse carcinogenicity, acute rat toxicity, and mutagenicity	http://www.simulations-plus.com/Products.aspx?grpID=1&cID=11&pID=13
TerraQSAR™	*Daphnia magna*, estrogen receptor binding, fathead minnow LC50, log *P*, mouse and rat oral LD50, rat and mouse intravenous LD50, skin irritation	http://www.terrabase-inc.com/

thinking." Therefore, ADMET Predictor offers functionality to rank order compounds using one of the default ADMET Risk™ filters, which combine predictions from numerous ADMET models that have been parameterized against a focused subset of the World Drug Index. Along with the numerical risk score, ADMET Predictor assigns alphanumeric risk codes that indicate the predicted ADMET issues associated with a compound. In the last section, we present case studies exemplifying application of ADMET Predictor to the prediction of metabolites.

2 Materials

2.1 Software Requirements

1. ADMET Predictor—to calculate descriptors, build a model, and score desired chemical structures with the developed model.

2.2 Optional Software

1. MedChem Designer™—to draw chemical structures.
2. MedChem Studio™—to organize data in a spreadsheet and perform scaffold analysis.
3. Microsoft Excel™—to prepare tab-delimited input files.

2.3 Experimental Requirements (Model Building Only)

1. Sufficient quantity of high-quality experimental data: chemical structures with corresponding measured values of the property of interest.

3 Methods

The main objective of using in silico ADME models in the drug discovery screening process is to either prioritize or deprioritize molecules for synthesis and/or further experimentation. ADMET Predictor provides models to screen compounds for more than 140 physicochemical, biopharmaceutical, toxicity, and metabolism properties. The program also calculates more than 300 carefully selected molecular descriptors with which users can build and install new models from their own assay data. If the property of interest is part of Simulations Plus' models built into ADMET Predictor, then one should follow the protocol that is described in Subheading 3.1 to use the built-in models. If the property of interest is not part of the built-in models, then the protocol described in Subheading 3.2 should be followed to build new models.

3.1 Using ADMET Predictor's Built-In Models

ADMET Predictor can read chemical structures in the following formats:

- SMILES strings (2D predictive models only).
- CTFile formats (formerly known as ISIS™ file formats) [6].

3.1.1 Preparing the Input File

- Molecular Structure Format (MOL)
- Structure Data Format (SDF)
- Reaction Data Format (RDF)

In the following section, we describe a detailed protocol to create one of these input files that can be read into ADMET Predictor.

3.1.1.1 SMILES (.smi) File

SMILES is an acronym for Simplified Molecular Input Line Entry System [7, 8] and is used to encode a particular valence representation of a molecular structure as a linear string. The instructions for creating SMILES strings for molecules can be found on the Daylight website [9]. As an example, the molecular structure of Diazepam with its SMILES representation is shown in Fig. 1.

The structures in SMILES format are input into ADMET Predictor in a tab-delimited ASCII text file with extension .smi. The file contains one compound per line (record) and the number of fields in each line must be equal. The first field must be the SMILES string followed by compound identifier and optional property values. An optional header can be added as the first line. However, if the file contains data columns in addition to the SMILES string and identifier, the header line is required. The number of header words must equal the number of data columns in subsequent lines. The first word must be "SMILES." Comments or empty lines are *NOT* allowed in the file and missing data are indicated by a single ~ (tilde) character.

A sample .smi file is shown below:

SMILES Name MeltingPoint CAS No. ExlogP ExSol

CC(=O)Nc1ccc(OCC)cc1 4-Ethoxyacetanilide 407.5 000062-44-2 1.58 0.766

O=C(N)C Acetamide 342.5 000060-35-5 -1.26 ~

O=C(O)C Acetic acid~000064-19-7 -0.17 ~

O1CC(O)C(O)C(O)C1O Arabinose 431000147-81-9 -2.32 550

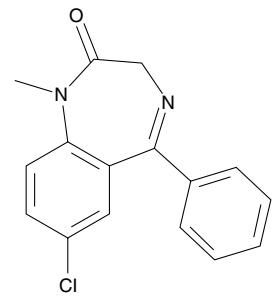


Fig. 1 Molecular structure of Diazepam, which has the SMILES: =C1N(c2c(C(c3ccccc3)=NC1)cc(Cl)cc2)C

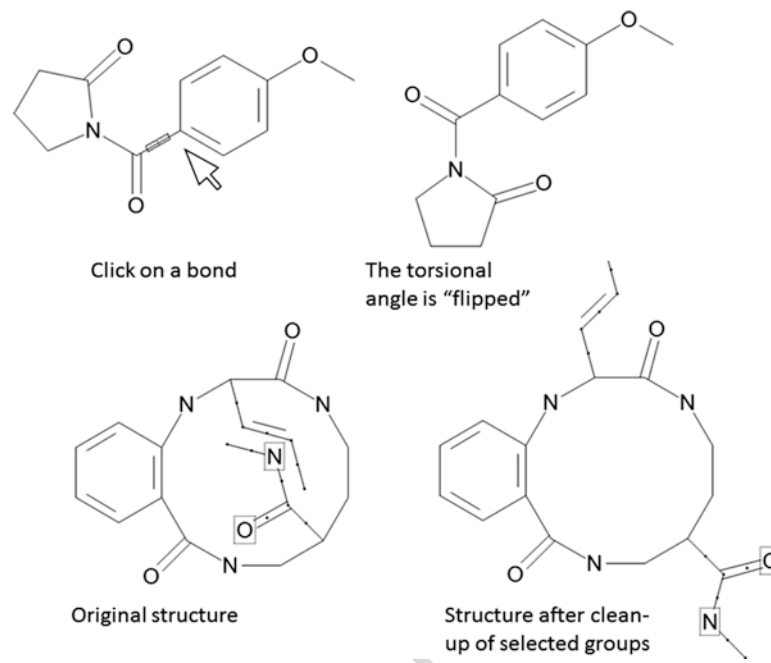


Fig. 2 Drawing and manipulating structures

The .smi input file can be quickly and easily prepared in an Excel spreadsheet. The SMILES strings can be generally extracted from supplementary information included with published articles or translated from structure drawings using either MedChem Designer (please see below) or other structure-drawing programs. The SMILES string for many compounds can also be found by searching online databases such as ChEMBL, PubChem, and ChemSpider.

New compounds may be easily added to the spreadsheet by drawing them using MedChem Designer, a free chemical structure-drawing program [10]. Multiple structures may be drawn and manipulated on the canvas (*see* Fig. 2). The optional Optical Structure Recognition (OSR) tool allows you to extract chemical structures from displayed images in Word documents, PDF files, PowerPoint slides, web pages, etc., by simply positioning the transparent window capture tool over the image of interest and clicking the Convert Image button (*see* Fig. 3). The chemical structure will be retrieved automatically and displayed in MedChem Designer. Any errors in the automatic conversion to a chemical structure can be easily corrected using MedChem Designer's chemical editing capabilities. (Note: paid license to ADMET Predictor or MedChem Studio is required for OSR capability.)

3.1.1.2 CTFile Formats: RDF, SDF, and MOL File

In addition to SMILES, ADMET Predictor can read MOLfiles, RDfiles, and SDfiles (.mol, .rdf, and .sdf extensions). Briefly, the MOL file contains two- or three-dimensional atomic coordinates and a bond connectivity table. SDF and RDF files are assemblages

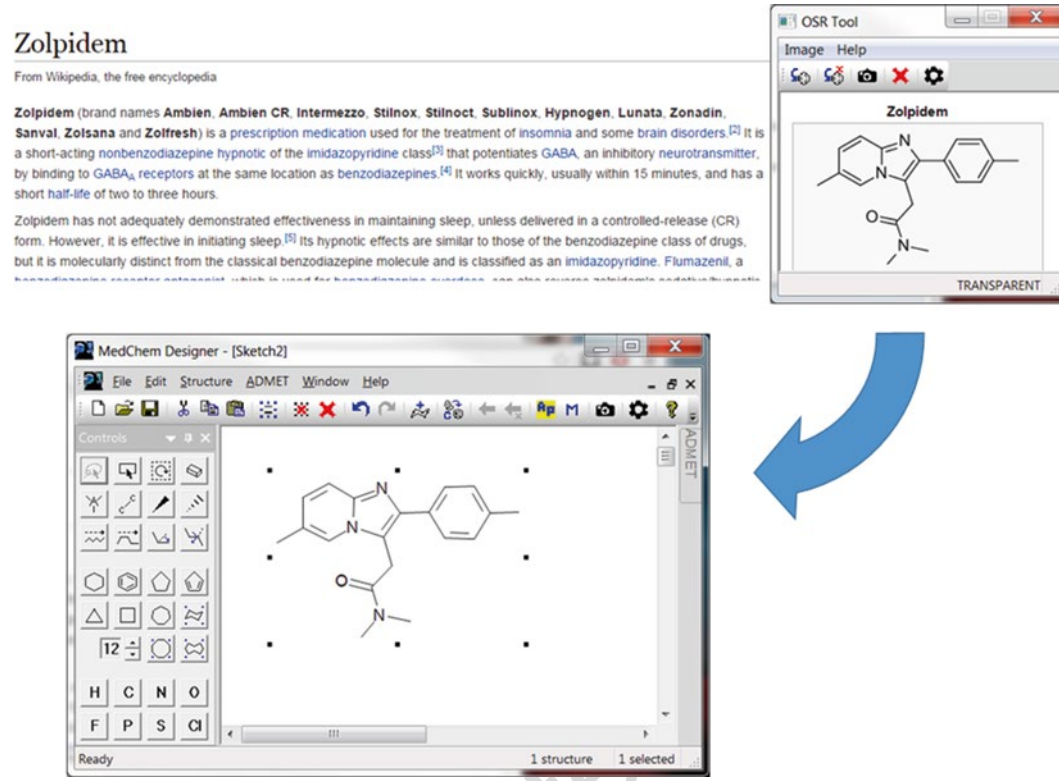


Fig. 3 Optical Structure Recognition (OSR) tool in MedChem Designer™

of multiple MOL file records and supplemental property data. Missing supplemental property data fields are allowed for an individual molecule record. ADMET Predictor identifies data column headers in its initial scan of the input file.

MOL files (file name extension .mol) are also easily produced by molecular drawing programs such as ChemDraw® [11] and Accelrys Draw [12]. The simplest way to create an RDF or SDF file compatible with ADMET Predictor is to export data from a commercial chemical database such as Biovia Insight [13] or using workflow programs such as Pipeline Pilot [14]. Most molecular databases contain 2D coordinates with a connectivity table to enable two-dimensional molecular depictions. This is the minimal amount of structural information needed to generate 2D descriptors. In contrast, some databases contain 3D atomic coordinates obtained either from X-ray crystallography or generated in silico by a molecular structure prediction program. The "Open 2D" menu option in ADMET Predictor will process both 2D and 3D SDF/RDF formats because in this mode ADMET Predictor uses only the list of atoms and connectivity tables contained in the SDF/RDF files and ignores actual coordinate information. In contrast, the "Open 3D" option will produce meaningful results only for 3D SDF/RDF files since the program uses the supplied 3D

coordinates to calculate 3D molecular descriptors which are used as inputs for calculating ADMET properties. Users should consult their chemical database administrator or software user manual for specific information regarding how to export or generate SDF or RDF files within their scientific environment.

3.1.2 Running ADMET Predictor Models

1. Start ADMET Predictor. From the main menu, click File > Open 2D structures. Now select your input file created in Subheading 3.1.1. To predict properties, select "Calculate ADMET Properties" from the "Calculate" menu. The "Run Options" dialog box opens. You can either keep the default run options or change them according to your needs. For example, you can change the pH (default = 7.4) at which you want to calculate pH-dependent properties, or you can include aliphatic hydroxyl groups in pK_a calculations (default = off).
2. Click the "Calculate" button to accept your selected options. The calculated results are displayed under the "Molecular Data" tab in a molecular record spreadsheet. ADMET Predictor property models are displayed in columns colored gray and green. Molecular descriptors are displayed in blue columns. Columns shaded in pink indicate user input data such as experimental results or other descriptors. Scroll to the right to see all columns.
3. Left click on the structure column to open the "Structure Visualization" window to display atomic descriptors mapped on the atoms (*see* Fig. 4). Pressing [Shift] + left click on a structure will open the structure in MedChem Designer for editing or metabolite generation.
4. Predicted properties are organized in different modules.
   (a) The Physicochemical and Biopharmaceutical Module contains 28 properties such as pK_a, log P, log D, various permeability models, blood–brain barrier, fraction protein unbound, fraction unbound in human liver microsomes, and transporter inhibition. For a full list of properties please see the ADMET Predictor manual.
   (b) The Metabolism Module predicts kinetic constants (Michaelis [15] constant (Km), maximum metabolic rate (Vmax), and intrinsic clearance (CLint)), and inhibition flag for five major human cytochrome P450 enzymes: 1A2, 2C9, 2C19, 2D6, and 3A4. The model for overall Human Liver Microsomal intrinsic clearance is a separate model. In addition, substrate classification models predict whether a given molecule is a substrate for one of nine CYPs, while the regioselectivity models predict the sites of metabolism for each molecule classified as a CYP substrate. These models make separate predictions for each of nine human

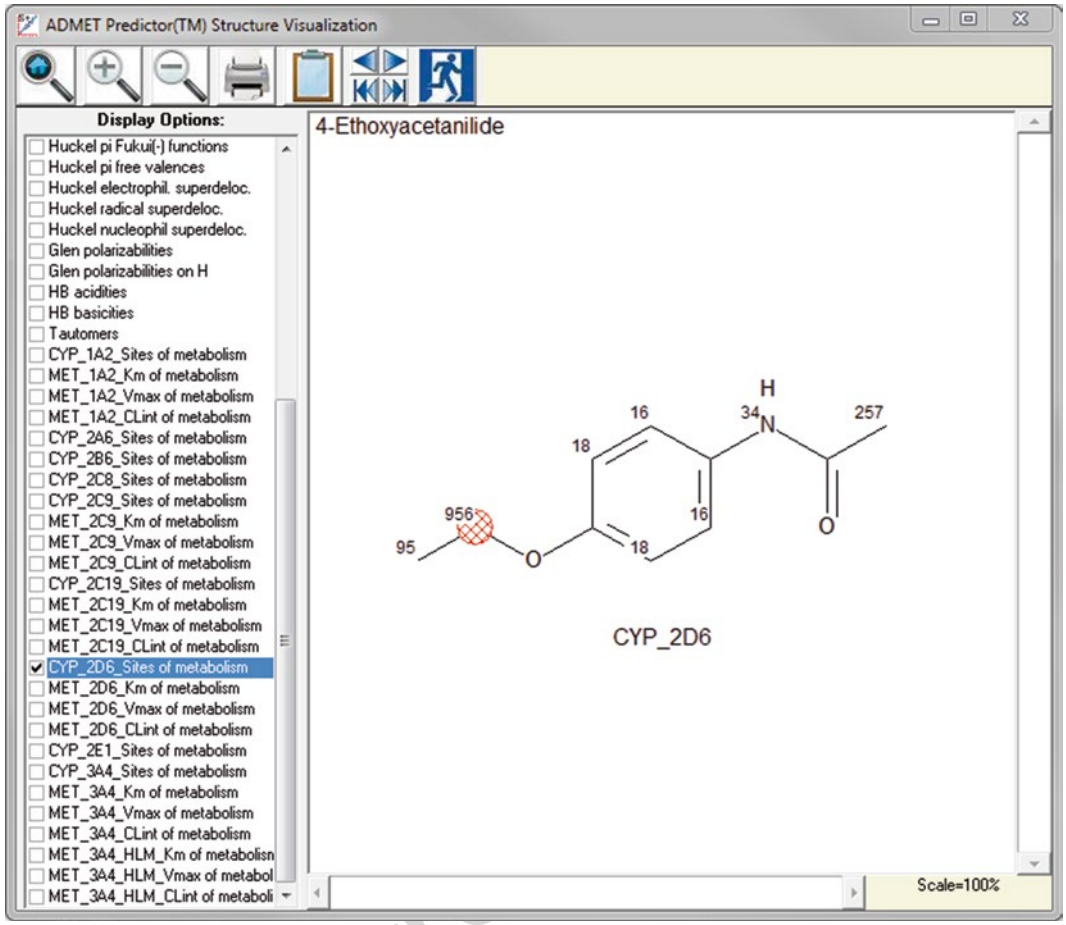


Fig. 4 Structure visualization window showing primary site of CYP 2D6 metabolic attack displayed on the compound structure

cytochrome P450 enzymes: 1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, and 3A4. Another model classifies the molecule as a substrate/nonsubstrate for the most important class of phase II enzyme, human uridine 5'-diphosphate-glucuronosyltransferase (UGT).

(c) The Toxicity Module predicts endocrine toxicity, maximum recommended therapeutic dose, aquatic toxicity, carcinogenicity, genotoxicity (Ames mutagenicity, chromosomal aberrations), cardiac and hepatic toxicity, environmental toxicities such as bioconcentration factor, and others. For a full list of toxicity models please see the ADMET Predictor manual.

(d) The Simulation Module features a special set of predictive models that are mechanistic rather than statistical. These predictions are based on a simplified GastroPlus™ [16] simulation of the pharmacokinetics of an orally administered drug

at several default dose levels (1, 10, 100, and 1000 mg). They use various ADMET model results as inputs along with dose and predefined human physiology, and solve a deterministic, region-dependent system of differential equations. The output is an estimate of fraction absorbed and the optimal dose that yields a targeted pharmacokinetic parameter such as an effective blood plasma concentration.

5. Check all the relevant properties for your project and compare predictions against experimental values, if present. If a required property is not part of ADMET Predictor's built-in models but sufficient experimental data are available, then proceed to the following model-building protocol to build a new QSPR model.

3.2 Build and Use Your Own ADMET Models

3.2.1 QSPR Model Building Process

1. ADMET Modeler™ is an integrated module of ADMET Predictor that automates the process of building high-quality QSPR models.
2. Once the experimental data have been properly curated, one should prepare an input file for ADMET Predictor as described above. Here one must pay special attention to data accuracy (chemical structure, units of reported values, assay protocol, measured endpoint, etc.). Failure to use correct structures, and/or inaccurate, non-uniform experimental data will result in models of little value.
3. Open the ADMET Predictor main window.
4. Open an input file and calculate ADMET properties along with descriptors.
5. Select the Basic Modeler Settings tab and select Dependent Variable from the drop-down menu (*see* Fig. 5).
6. From the "Dependent Variable" drop-down menu, select the variable name representing the experimental data for which you wish to build a model.
7. Open the "Adv. Modeler Settings" tab.
8. Select Test Set allocation percent (set at, e.g., 20 %) and the method to divide training and test sets from the Test Set sub-tab (Kohonen [17] is the default).
9. Next, click the Descriptors icon to open the Descriptors sub-tab. In the Descriptor Number Reduction window, select a value for Minimum Representation (default is 4) and select the Sensitivity Analysis method (default is Truncated Linear Analysis).
10. On the Kohonen sub-tab, confirm that the default value of Automatic is selected.
11. Click the ANNE icon to open the ANNE sub-tab. Use Automatic settings of the network architecture (Network neurons and Network inputs (descriptors)).

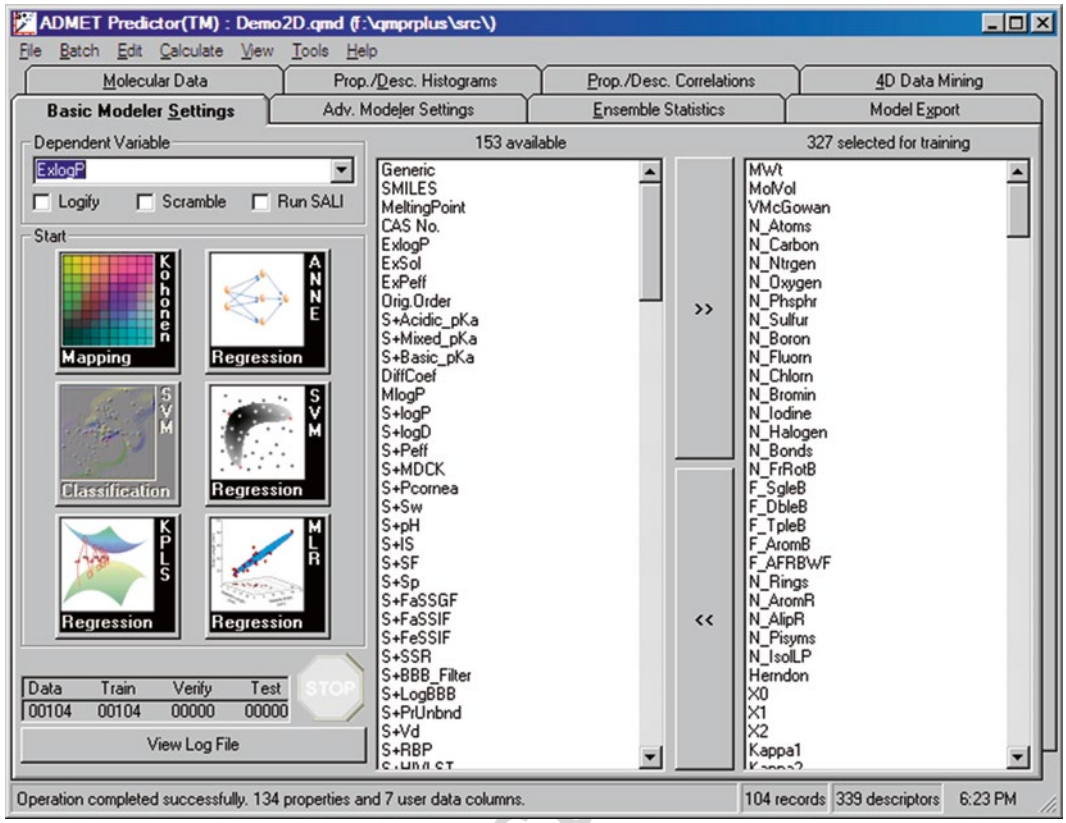


Fig. 5 ADMET Predictor main window, Basic Modeler Settings tab

12. Return to the Basic Modeler Settings tab, click the desired modeling method button, e.g., ANNE Regression or ANNE Classification icon for regression and classification models, respectively.
13. As model training begins, the Ensemble Statistics tab will open and messages related to this process are displayed in the status bar. You may click the Stop icon at any time to stop the training process.
14. Wait until "Modeler training is complete" is displayed in the status bar, and then proceed to find the best ensemble model (*see* Fig. 6).
15. To find the best ensemble, click the Find Best Ensemble button on the right. If you check the Show Adv Settings box, a new window with various criteria for finding the best ensemble will pop up. Try different options or simply accept the default values to find the best ensemble, as shown in Fig. 7.
16. To view model performance, click on the best ensemble (green cell) on the grid. The Performance Viewer window opens showing the plot of experimental vs predicted values if you built a regression model, or a contingency plot for classification models, as shown in Fig. 8.

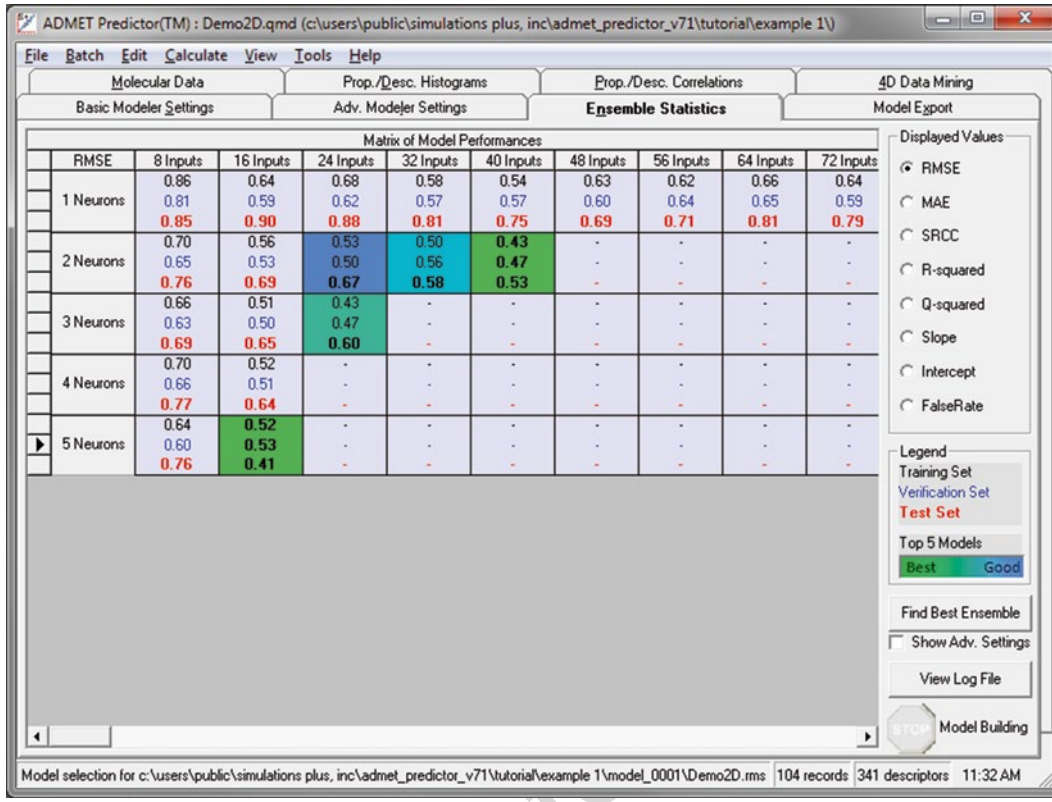


Fig. 6 Best ensemble highlighted in *bright green*

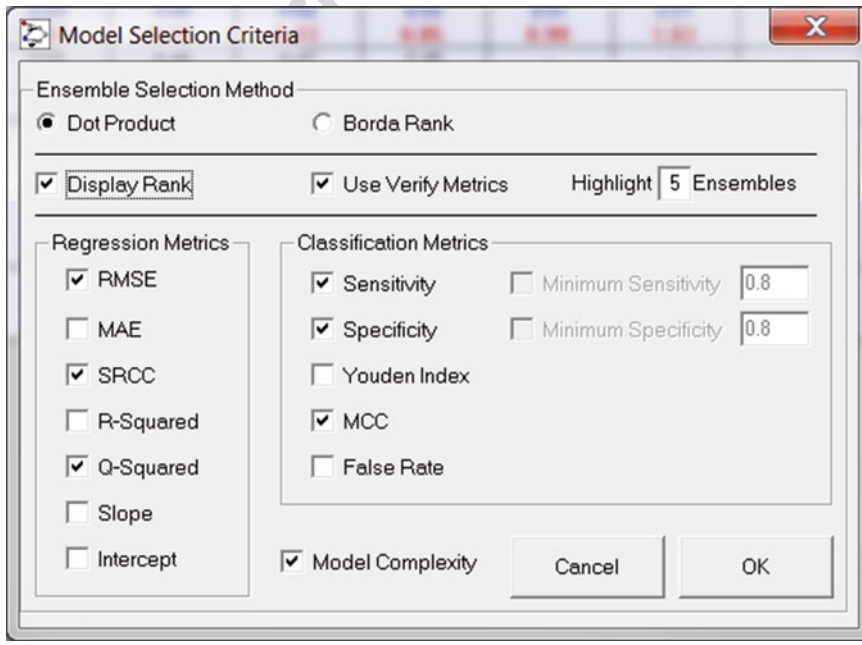


Fig. 7 Model Selection Criteria dialog box

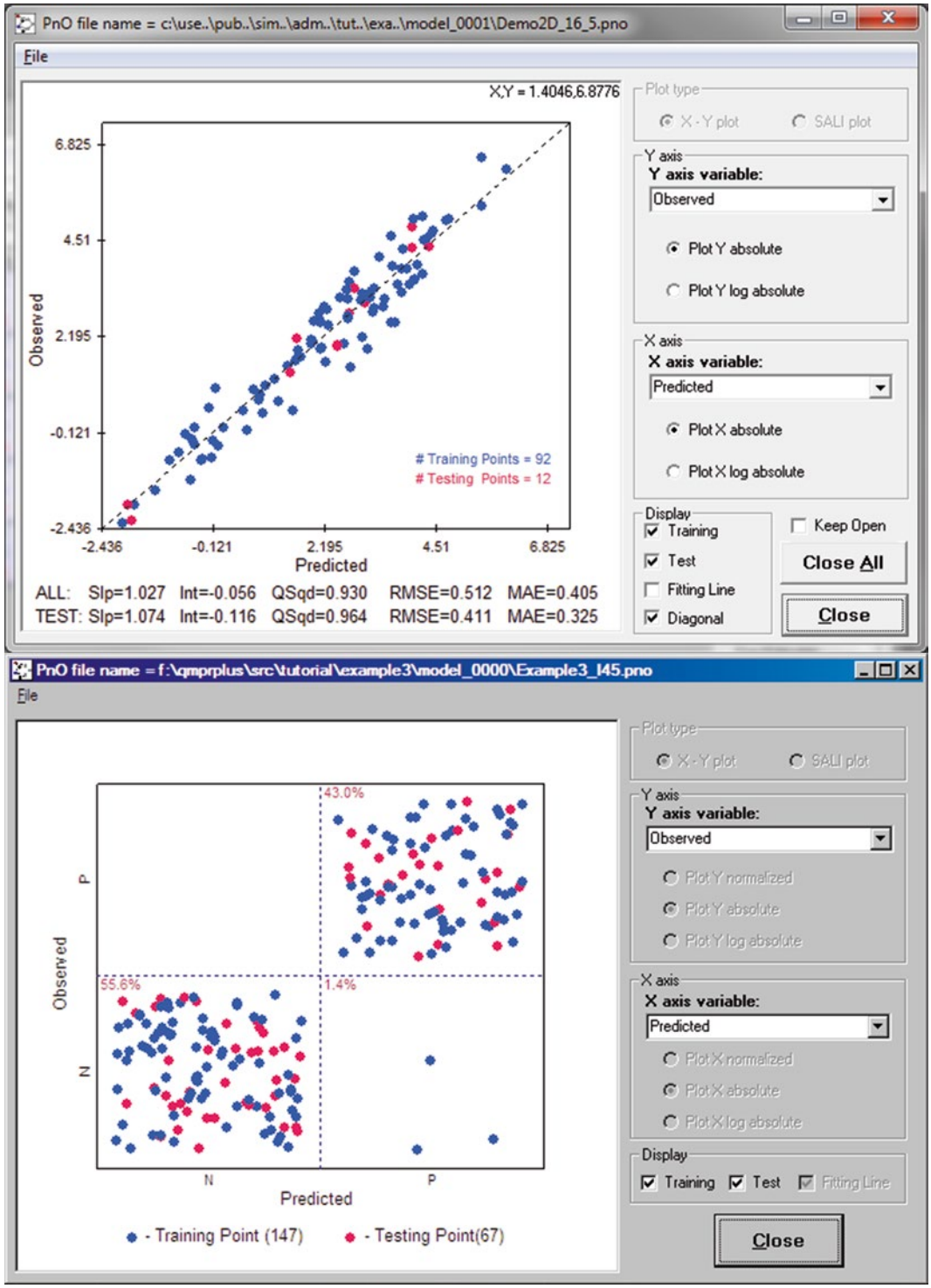


Fig. 8 Performance Viewer window for regression (*upper graph*) and classification (*lower graph*) models

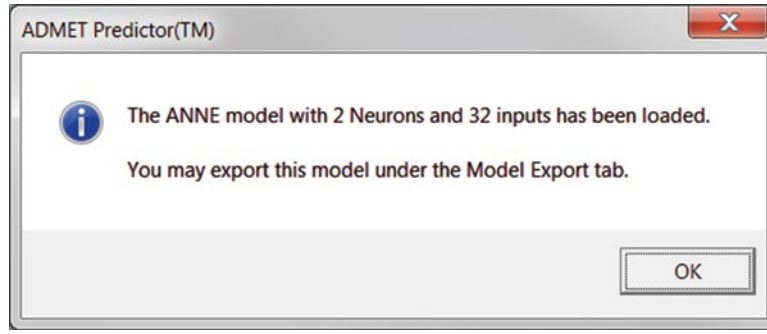


Fig. 9 Model loaded and ready for export message

17. Now try changing the model-building options (in the Adv. Modeler Settings tab), such as test set selection method, sensitivity analysis algorithm, number of neurons, and number of inputs to build alternative models.
18. To make the final model selected in **step 17** a part of the ADMET Predictor models, holding the [ctrl] key, click on the cell that displays the best ensemble model (the one highlighted in green). A message seeking confirmation to load the chosen ANNE model with "*n*" neurons and "*i*" inputs is displayed (*see* Fig. 9).
19. Click OK to confirm. Next, export the model using the Model Export tab (*see* Fig. 10). It shows a performance table of the individual member networks that make up the ANNE ensemble. Optionally, you may deselect some of these networks (with a left mouse click) before model export, resulting in a smaller ensemble, although this is usually not recommended unless certain models are significant outliers from the majority (very high RMSE for regression models, or very high false rate for classification models).
20. On the Model Export tab, click Export Current Ensemble as New Model. The Add New Model dialog box will be displayed (*see* Fig. 11). Fill in this form with appropriate entries (model name, type, location of model files, etc.) and click on Append as Last Model. The Add New Model dialog box closes. The model is displayed in the last row of the Model Editor window and is automatically selected. Click OK to save the newly added model.
21. If you now return to the Molecular Data tab and click on the User Models tab at the bottom, you should be able to see your new model which can be used for predictions on untested compounds.

3.3 Application of QSAR Models in Virtual Screening

1. Prepare a structure file for compounds to be screened (as described in Subheading 3.1.1).
2. Open the prepared structure file (.smi or .sdf or .mol) and calculate properties as described in Subheading 3.1.2.

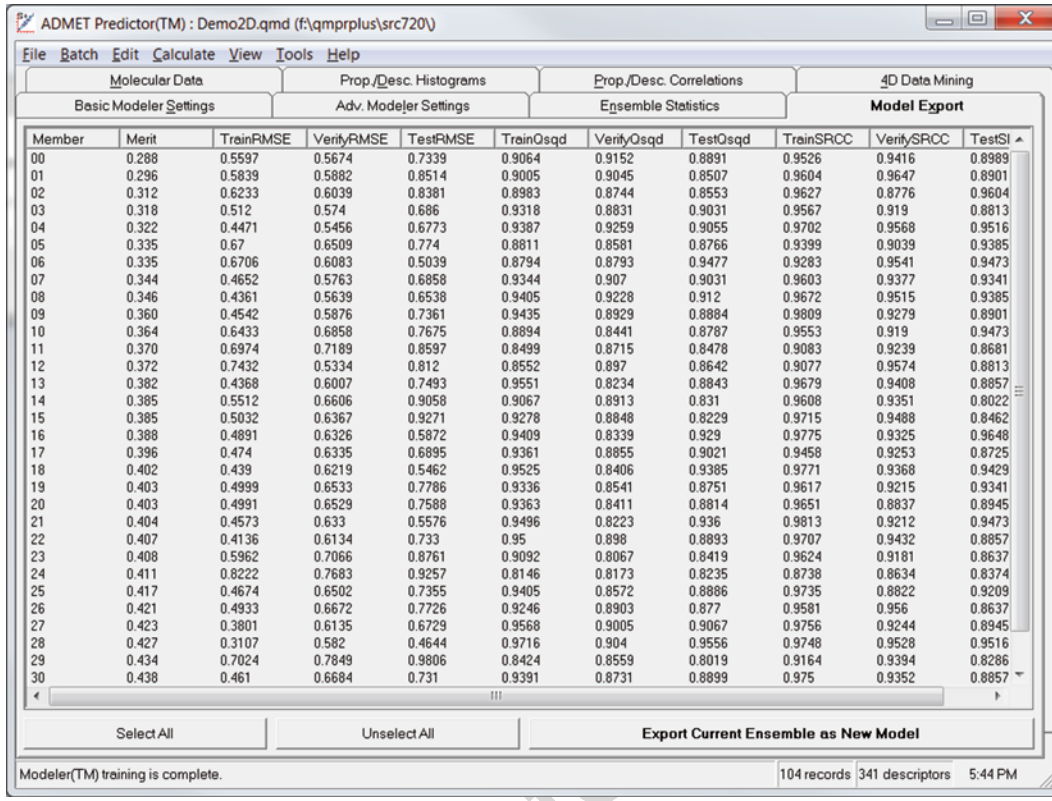


Fig. 10 Model Export tab

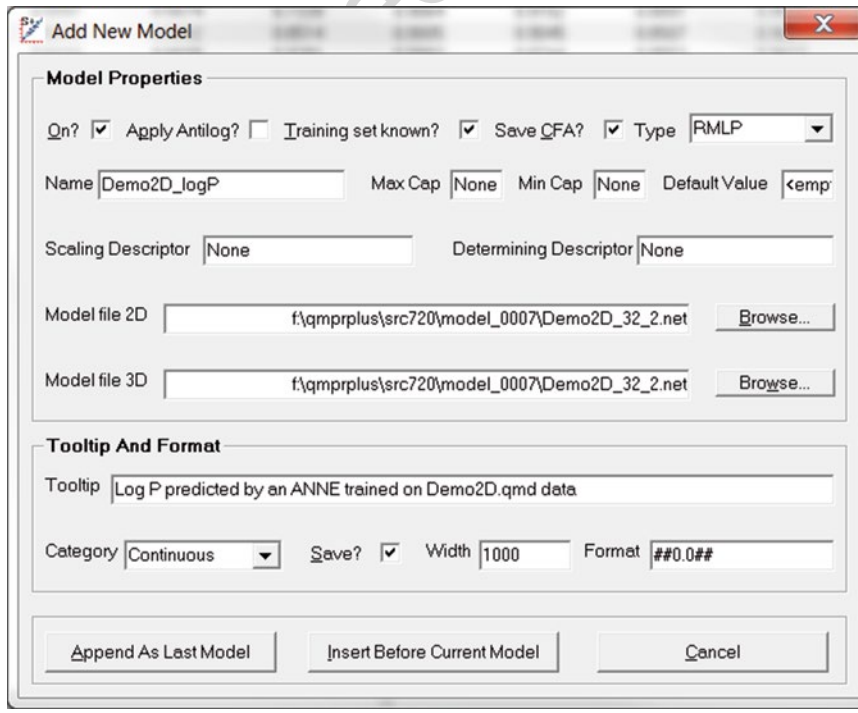


Fig. 11 Add New Model dialog box

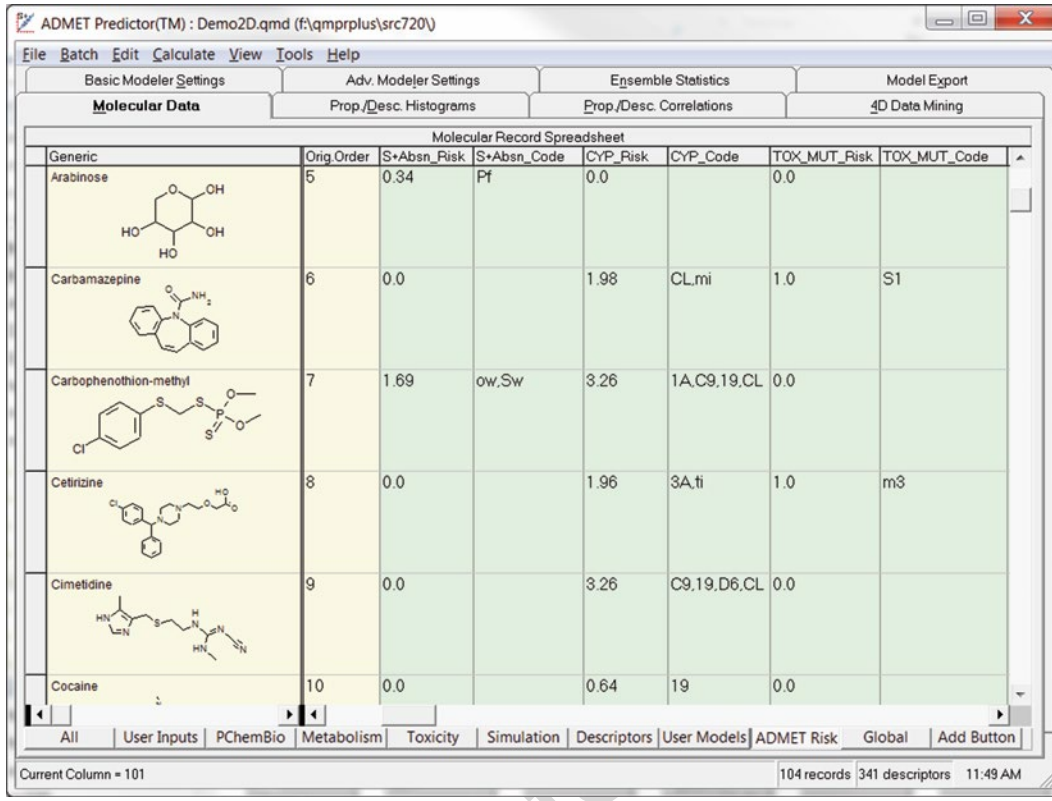


Fig. 12 ADMET Risk tab showing different risks and codes

3. Click on the ADMET Risk tab (Fig. 12) on the ADMET Predictor spreadsheet. You will see four pairs of columns; one each for absorption (Absn), metabolism (CYP), toxicity (TOX), and overall ADMET (ADMET) risks likely to be associated with a compound. The numerical score for each risk type is derived from a set of rules (inspired by the Lipinski's Rule of 5, but more extensive) based on predicted values from multiple prebuilt models in ADMET Predictor. An alphanumeric code, if present, indicates the violated rule(s), i.e., the type of ADMET liability(ies) likely to be associated with the molecule. These computed risk scores, having been carefully validated with a focused subset of over 2000 molecules from the World Drug Index, have been found to be accurate to allow rank ordering of a large number of compounds.
4. The S+Absn_Risk model includes eight rules based on descriptors and predicted properties as part of the Physicochemical and Biopharmaceutical Module, which are size, flexibility, hydrogen bond donors and acceptors, charge, lipophilicity, permeability, and solubility.
5. The S+Absn_Risk score can be between 0 and 8; for about 90 % of the focused subset of WDI [18], the S+Absn_Risk is

below 3.5. The S+Absn_Code indicates the type of likely liability for the compound based on the rule(s) violated out of these eight rules. The CYP_Risk model is comprised of seven individual rules, based on CYP clearance, Km, and Ki (inhibition constant) predictions from ADMET Predictor. The CYP_Code column indicates the specific CYP liability for each compound that violates any of the seven rules. The CYP_Risk is less than 1 for 85 % of the focused WDI.

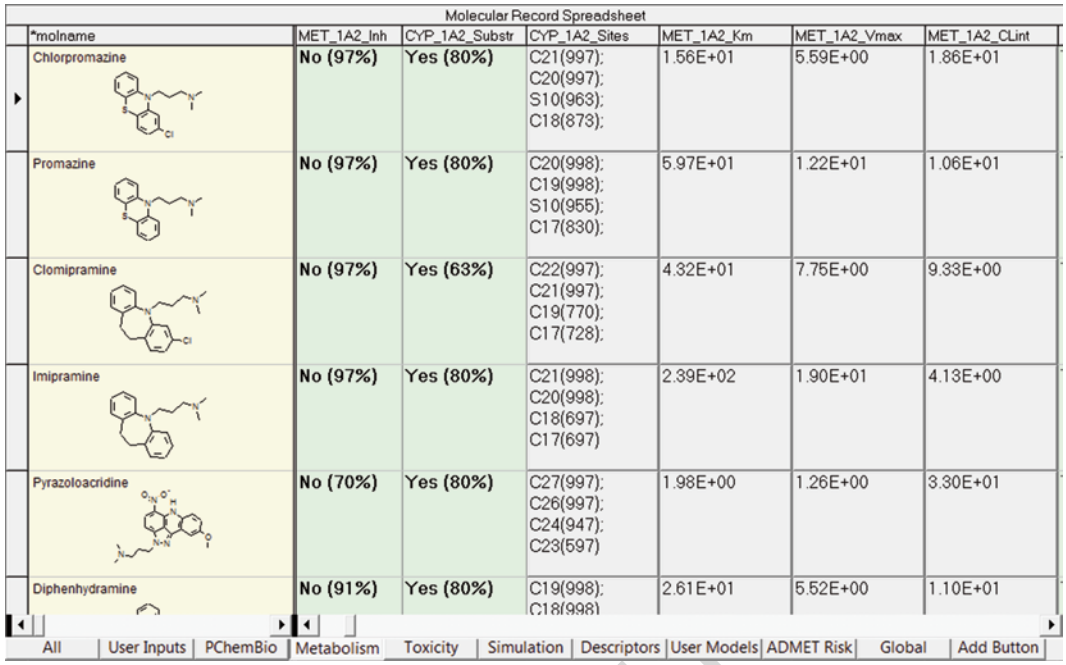
6. The TOX_Risk model consists of seven rules. This risk indicates the likelihood of a compound to have acute toxicity in rats or mice, hERG toxicity, hepatotoxicity, and mutagenicity. For about 90 % of the focused subset of WDI, the TOX_Risk is below 3.3. The TOX_Code, if present, indicates the likely toxicity liability for each compound.
7. Finally, the global ADMET_Risk model combines all the above risks plus two additional rules based on fraction unbound in plasma and steady state volume of distribution. There are 24 different rules that contribute to the default ADMET_Risk score; which is below 6.5 for about 90 % of the focused WDI. ADMET Risk can be edited to add your own rules or to modify the default rules.

3.3.1 A Case Study Using Metabolism Models

In this section, we will discuss the Metabolism module in greater detail and provide an interesting case study using our regioselectivity (site of metabolism) models. The prediction of sites and products of metabolism for xenobiotic and endogenous compounds is an important aspect of research in the development and use of pharmaceuticals. Toxicity or side effects due to metabolites can be detrimental and may play a major role in the withdrawal of a new drug from the market as well as contributing to the high attrition rates in the development of new chemical entities. Metabolites can also be beneficial, adding to the therapeutic efficacy. For prodrugs, it is the metabolite that is the active moiety.

1. Prepare and open a structure file for which predictions are desired (see detailed description of how to create input files in Subheading 3.1.1).
2. Calculate the properties as described in Subheading 3.1.2.
3. Scroll spreadsheet columns to the range occupied by the Metabolism module or click the Metabolism tab (*see* Fig. 13).

Metabolism models have been described in Subheading 3.1.2. Some models, indicated by the gray background of the spreadsheet cells, offer a deeper, atomic level of detail. Click one of the gray cells to reveal sites of potential metabolic attack by a particular CYP and, if applicable, the rates of site-specific attacks mapped onto a molecular structure. Chlorpromazine, a substrate of human CYP1A2, offers an interesting example of predictive


Molecular Record Spreadsheet

*molname	MET_1A2_Inh	CYP_1A2_Substr	CYP_1A2_Sites	MET_1A2_Km	MET_1A2_Vmax	MET_1A2_CLint
Chlorpromazine	No (97%)	Yes (80%)	C21(997); C20(997); S10(963); C18(873);	1.56E+01	5.59E+00	1.86E+01
Promazine	No (97%)	Yes (80%)	C20(998); C19(998); S10(955); C17(830);	5.97E+01	1.22E+01	1.06E+01
Clomipramine	No (97%)	Yes (63%)	C22(997); C21(997); C19(770); C17(728);	4.32E+01	7.75E+00	9.33E+00
Imipramine	No (97%)	Yes (80%)	C21(998); C20(998); C18(697); C17(697)	2.39E+02	1.90E+01	4.13E+00
Pyrazoloacridine	No (70%)	Yes (80%)	C27(997); C26(997); C24(947); C23(597)	1.98E+00	1.26E+00	3.30E+01
Diphenhydramine	No (91%)	Yes (80%)	C19(998); C18(998)	2.61E+01	5.52E+00	1.10E+01

All | User Inputs | PChemBio | Metabolism | Toxicity | Simulation | Descriptors | User Models | ADMET Risk | Global | Add Button

Fig. 13 Predictions of metabolic properties related to the human CYP1A2 P450 enzyme

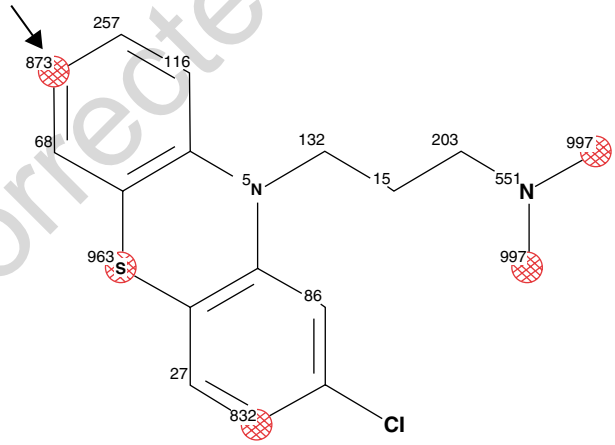


Fig. 14 Propensity of individual atoms of chlorpromazine toward metabolic attack by human CYP1A2. *Arrow* indicates site known in 2009; see text for details

regioselectivity. Figure 14 reveals atomic propensity scores for the *first* step of metabolic oxidation by this enzyme. These scores, ranging between 0 and 1000, should *not* be confused with probabilities. They may be compared in a relative sense for atoms within the same molecule but such comparisons are not applicable between different molecules. In this sense, the scores are *not* transferable. The five atoms carrying scores above the model's

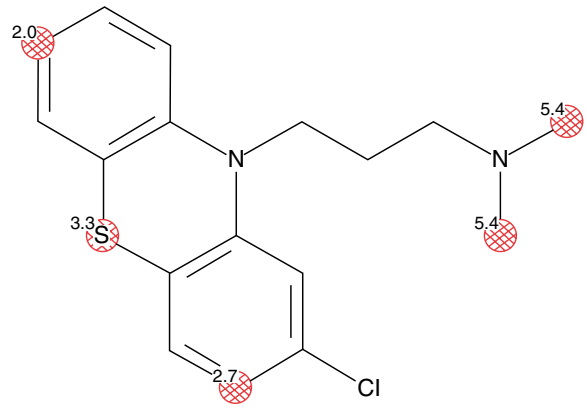


Fig. 15 Predicted intrinsic clearances in μL/min/mg microsomal protein for individual atoms of chlorpromazine for the complete metabolic oxidation by human CYP1A2

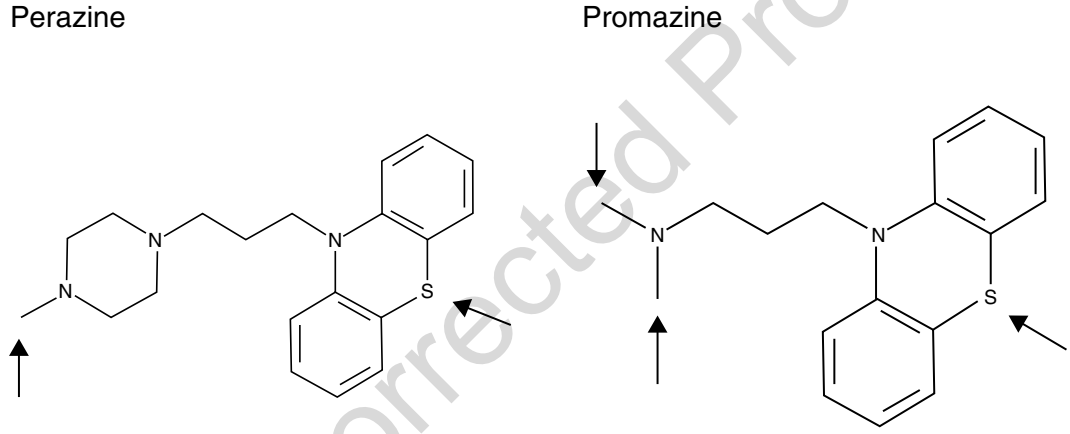


Fig. 16 Observed sites of perazine and promazine metabolism by CYP1A2

threshold are highlighted by red cross-hatched circles. Figure 15 shows that predicted intrinsic clearances (CLint) for these atoms indicate no large differences in the atomic rates of oxidation by 1A2.

We chose chlorpromazine as a case study, since as of 2009, the only reported metabolite (of which we were aware at the time—other metabolites had been reported but the enzymes producing them had not been determined) resulted from hydroxylation of C-7, identified by the arrow in Fig. 14, mediated by 1A2 and 2D6 [19]. However, the 1A2 dataset has two similar structures, perazine and promazine, shown in Fig. 16 with arrows showing their reported CYP1A2-mediated sites of metabolism [20, 21].

Based on these results, it seems reasonable for the 1A2 model to predict sites of metabolism for the sulfur and *N*-methyl carbons of chlorpromazine as well. Such assignments, as of 2009, would be considered "false" positives! However, in late 2010, both

N-demethylation and sulfoxidation of chlorpromazine were reported as metabolites formed by CYP1A2 [22]. Later, we found two more articles confirming sulfoxidation of chlorpromazine [23] and the formation of its [18] 3-hydroxy metabolite [24]. Thus, "false" positives in the model in 2009 became true positives in 2010! Out of four, only one site (with relatively low score) thus remains as a tentatively false-positive prediction for 1A2, perhaps pending future verification.

References

1. Calcoen D, Elias L, Yu X (2015) What does it take to produce a breakthrough drug? Nat Rev Drug Discov 14(3):161–162
2. The truly staggering cost of inventing new drugs (2012) Forbes. http://www.forbes.com/sites/matthewherper/2012/02/10/the-truly-staggering-cost-of-inventing-new-drugs/
3. Scannell JW, Blanckley A, Boldon H, Warrington B (2012) Diagnosing the decline in pharmaceutical R&D efficiency. Nat Rev Drug Discov 11(3):191–200
4. ADMET Predictor v7.2, Simulations Plus, Inc.: Lancaster, CA, USA
5. Directory of computer-aided Drug Design tools. (2014) http://www.click2drug.org/directory_ADMET.html
6. Dalby A, Nourse JG, Hounshell WD, Gushurst AKI, Grier DL, Leland BA, Laufer J (1991) Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. J Chem Inf Comput Sci 32(3):245–260
7. Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J Chem Inf Comp Sci 28:31–36
8. Weininger D, Weininger A, Weininger JL (1989) SMILES. 2. Algorithm for generation of unique SMILES notation. J Chem Inf Comp Sci 29:97–101
9. SMILES (2007) Daylight Chemical Information Systems. Inc, www.daylight.com/smiles/index.html
10. MedChem Designer v3.0, Simulations Plus, Inc.: Lancaster, CA, USA
11. ChemDraw v15.0, PerkinElmer: Waltham, MA, USA
12. Accelrys Draw v4.1, Biovia: San Diego, CA, USA
13. Biovia Insight v1.0, Biovia: San Diego, CA, USA
14. Pipeline Pilot v9.0, Biovia: San Diego, CA, USA
15. Michaelis L, Menten M (1913) Die Kinetik der Invertinwirkung. Biochem Z 49:333–339
16. GastroPlus v9.0, Simulations Plus, Inc.: Lancaster, CA, USA
17. Kohonen T (1982) Self-Organized Formation of Topologically Correct Feature Maps. Biol Cybern 43:59–69
18. World Drug Index v2007/04, Thomson Scientific: London, UK
19. Yoshii K, Kobayashi K, Tsumuji M, Tani M, Shimada N, Chiba K (2000) Identification of human cytochrome P450 isoforms involved in the 7-hydroxylation of chlorpromazine by human liver microsomes. Life Sci 67:175–184
20. Wójcikowski J, Pichard-Garcia L, Maurel P, Daniel WA (2003) Contribution of human cytochrome P-450 isoforms to the metabolism of the simplest phenothiazine neuroleptic promazine. Br J Pharmacol 138(8):1465–1474
21. Wójcikowski J, Pichard-Garcia L, Maurel P, Daniel WA (2004) The metabolism of the piperazine-type phenothiazine neuroleptic perazine by the human cytochrome P-450 isoenzymes. Eur Neuropsychopharmacol 14(3):199–208
22. Wojcikowski J, Boksa J, Daniel WA (2010) Main contribution of the cytochrome P450 isoenzyme 1A2 (CYP1A2) to N-demethylation and 5-sulfoxidation of the phenothiazine neuroleptic chlorpromazine in human liver—A comparison with other phenothiazines. Biochem Pharmacol 80:1252–1259
23. Cashman JR, Yang Z, Yang L, Wrighton SA (1993) Stereo- and regioselective N- and S-oxidation of tertiary amines and sulfides in the presence of adult human liver microsomes. Drug Metab Dispos 21(3):492–501
24. Morel E, Lloyd KG, Dahl SG (1987) Anti-apomorphine effects of phenothiazine drug metabolites. Psychopharmacology (Berl) 92(1):68–72

Part II

The Applications of In Silico Models for the Different Endpoints

Chapter 5

In Silico Prediction of Chemically Induced Mutagenicity: How to Use QSAR Models and Interpret Their Results

Enrico Mombelli, Giuseppa Raitano, and Emilio Benfenati

Abstract

Information on genotoxicity is an essential piece of information gathering for a comprehensive toxicological characterization of chemicals. Several QSAR models that can predict Ames genotoxicity are freely available for download from the Internet and they can provide relevant information for the toxicological profiling of chemicals. Indeed, they can be straightforwardly used for predicting the presence or absence of genotoxic hazards associated with the interactions of chemicals with DNA.

Nevertheless, and despite the ease of use of these models, the scientific challenge is to assess the reliability of information that can be obtained from these tools. This chapter provides instructions on how to use freely available QSAR models and on how to interpret their predictions.

Key words Mutagenicity, Ames test, QSAR, Predictive reliability, Structural alerts

1 Introduction

The assessment of information on mutagenicity represents an important component for the evaluation of the toxicological characteristics of chemicals [1]. For instance, in the field of drug discovery the detection of mutagenic potential of a chemical can result in the rejection of a promising chemotype owing to the deleterious consequences that the introduction of gene mutations can elicit. In addition, the characterization of genotoxicity is required for the regulatory qualification of impurities in drug substances [2] and it is a mandatory requirement for all the different tonnage bands defined by the overarching REACH regulation [3].

Mutagenic effects caused by chemical agents can be detected by the Ames test that was devised by Bruce Ames during the 1970s [4]. This test is still commonly in use in many toxicological laboratories around the world because of its good interlaboratory reproducibility, aptitude at testing different agents, cost-effectiveness, and structure–activity analysis [5]. The remarkable juxtaposition of these attributes has brought the Ames test to the forefront

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_5, © Springer Science+Business Media New York 2016

of modern toxicology. Indeed, this test is a paradigm for the development of nowadays in vitro toxicology and it has been nicknamed "the stethoscope of genetic toxicology for the twenty-first century" [5] given that testing strategies for carcinogenicity rely on the Ames test as an essential first-tier assay [5, 6].

This test is based upon the ability of *Salmonella typhimurium* and *Escherichia coli* auxotrophic strains to recover the ability to synthesize an essential amino acid (histidine for *S. typhimurium* and tryptophan for *E. coli*) as a consequence of the mutagenic effect of chemicals to which they are exposed. The design of the experimental protocol enables the detection of bacterial colonies that can grow in the absence of essential amino acids as a result of a back mutation that restores their biosynthetic capabilities. The detection of this back mutation to wild type has the potential to identify point mutations that are caused by the substitution, addition, or deletion of one or few DNA base pairs. At least five bacterial strains should be used when testing a chemical [7], including strains that are sensitive to oxidizing mutagens, cross-linking agents and hydrazines (*E. coli* WP2 or *S. typhimurium* TA102, *see* **Note 1**).

Anyhow, it is important to note that, as stated in the OECD guideline [7], mammalian tests may be more appropriate when evaluating certain classes of drugs. For example, the Ames test is not the most appropriate choice for chemicals displaying a high bactericidal activity such as certain antibiotics, topoisomerase inhibitors, and some nucleoside analogs.

The interlaboratory reproducibility of the Ames test is estimated at 85–90 % [8, 9] and these percentages represent the upper limit of predictive performance that can be expected from QSAR models for the same endpoint. Indeed, these models are derived from data obtained by means of the same protocol. In other words, these findings mean that 10–15 % of the chemicals that were experimentally tested gave different results when analyzed in different laboratories. Therefore, this experimental uncertainty in terms of false negative or positive predictions is transposed into the semi-empirical QSAR models that cannot be expected to be more reliable than their experimental counterpart.

Consequently, one key issue that should be given attention when judging the reliability of a QSAR model predicting Ames genotoxicity is whether or not this model predicts with a reliability that is comparable to the reproducibility of the test. It is worth mentioning that this comparison has to be critically assessed as a function of the number and chemotypes of the chemicals that compose the external test set that was adopted in order to validate the model. For example, if the external test set does not include all the chemotypes that are covered by the training set, the estimated predictive performance of the model will only be representative of a subset of chemical structures.

One final word of caution should be added with respect to models whose alleged performance is much higher than the experimental test they are meant to replace. This special situation could indicate a potential overfitting of the model and its lack of ability to provide reliable prediction for new cases (i.e. molecules that are not included in its training set).

The theory of electrophilic reactivity by Miller and Miller [10] adequately describes the molecular mechanisms that control the genotoxicity of chemicals as detected by the Ames test. Indeed, this theory has proved to be in agreement with the observations ever since it was formulated in the late 1970s. According to this theory, the vast majority of known chemical carcinogens are also genotoxic since they are (or are metabolized to) reactive electrophiles that react with nucleic acids. The (Q)SAR models described in this chapter (*see* Subheading 2.6) conform to this theory by identifying structural fragments that trigger electrophilic reactions as formalized by E-state values and fragments (e.g. CAESAR) and by structural alerts (SA) validated by experts (e.g. Toxtree SA) or automatically extracted by learning algorithms (e.g. SARpy).

Because of the complementary nature of these tools, this chapter illustrates the practical application of models covering the three main categories of in silico tools for the prediction of the mutagenic potential of chemicals: (Q)SAR models that are based on numerical descriptors (e.g. partition coefficients, topological descriptors, functional group counts), rule-based expert systems that are based on structural alerts (molecular fragments that are associated with the occurrence of adverse outcomes), and hybrid models combining these two approaches. Models based on all these approaches are implemented within the freely available VEGA platform (version 1.0.8): CAESAR, SARpy, and ToxTree-VEGA (TT-VEGA) (*see* **Note 2**). A brief description of the models is given in the following paragraphs and more detailed information can be found in the literature therein cited.

2 Materials

2.1 Performance Characterization of (Q)SAR Models

The performance of models predicting the presence and absence of toxicological hazards is usually described by Cooper statistics [11] that characterize the predictive capabilities of diagnostic tests: sensitivity, specificity, and accuracy (or concordance). Sensitivity is the ability to identify a chemical that presents a toxicological hazard as toxic; specificity is the ability to correctly identify chemicals that do not present toxicological hazards as safe; and accuracy describes the overall concordance between predicted and experimental values. Their mathematical definitions are the following:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Concordance} = \frac{TP + TN}{TP + FN + TN + FP}$$

where TP = number of true positive predictions, TN = number of true negative predictions, FP = number of false positive predictions, FN = number of false negative predictions.

In the presence of skewed data sets (e.g. a data set including a majority of non-mutagenic chemicals), Cooper statistics are not fully reliable. It is therefore more appropriate to compute the Matthews correlation coefficient (MCC) which is defined as follows:

$$MCC = \frac{TPTN + FPFN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

The MCC ranges from −1 to +1. A MCC value of +1 represents a total agreement between experimental results and predictions; a value of 0 no better than random prediction, and a value of −1 indicates a total disagreement between predicted and observed values.

2.2 Software Requirement

TTVEGA, CAESAR, and SARpy models are embedded within the standalone software application VEGA (v. 1.0.8) that allows for a secure in-house execution of the three models without the need to send information to any external server [12]. VEGA can be also used for batch processing of multiple chemical structures. The software application can be freely downloaded for the VEGA website [12] and it can be installed and used on any operative system supporting JAVA.

2.3 Optional Software

Any software application that allows to draw chemical structures and convert them into two types of chemical file formats supported by VEGA: "Simplified Molecular Input Line Entry specification" (SMILES) [13] or "Structure Data Format" (SDF) can be used in order to generate input structures. Several chemical drawing programs can perform this task: VEGA ZZ [14], ACD/ChemSketch [15], MarvinSketch [16], and the OECD QSAR Toolbox [17] (for SMILES formats only).

This list is not exhaustive and these applications are subjected to different software licenses and terms and conditions of use.

2.4 VEGA: The Workflow

VEGA has a simple workflow which is schematically depicted in Fig. 1. Basically, a user types or pastes a SMILES string in the blank space at the top of the user interface and then adds it to a working list of molecules to be analyzed. Once that a SMILES string is added at the working list, it is possible to highlight it and visually

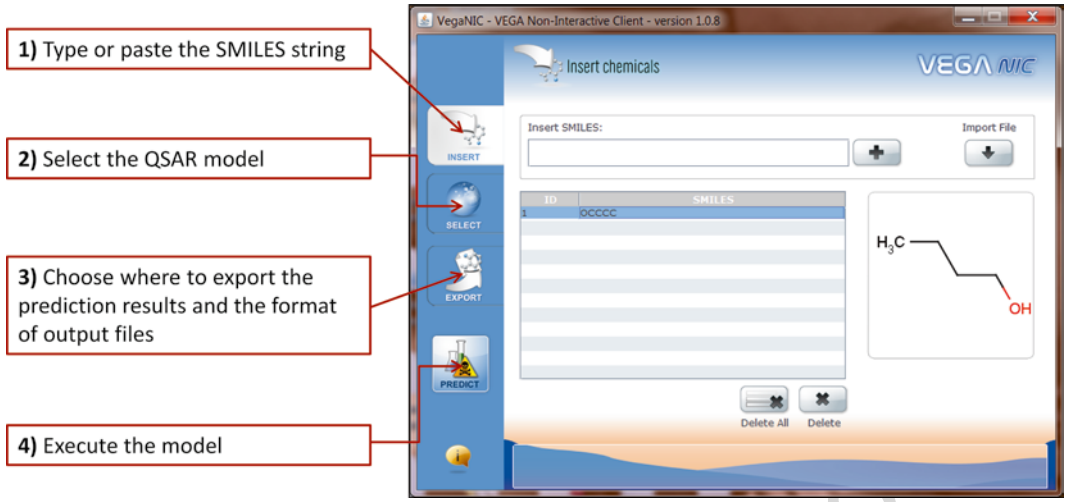


Fig. 1 Workflow of VEGA. *The SMILES string corresponding to 1-butanol was used as input structure*

check the two-dimensional structure encoded by the text line. This checkpoint is crucial. Indeed, several structural inaccuracies can take place at this stage and compromise the reliability of the predictions [18].

If needed, users can also input multiple molecules at once ("import File" button at the top right of the user interface). In this case the file contains a list of SMILES codes saved in "txt" or "smi" format.

Thanks to the "Select" button it is then possible to choose the model(s) of interest, to specify the desired output format (PDF or csv files), and to indicate where the prediction reports should be saved ("Export" button). Finally, the selected model(s) can be executed by clicking on the "Predict" button.

2.5 Applicability Domain

All the models that will be described in the following paragraphs adopt the VEGA definition of applicability domain [19] (*see* **Note 3**). According to this definition, the degree of membership of a query chemical to the applicability domain of the model is described by an Applicability Domain Index (ADI) with values that range from 0 (no membership) to 1 (full membership). Chemicals characterized by ADI values that are less than 0.7 are to be regarded as potentially not belonging to the AD. ADI values that are within the range 0.7–0.9 represent a critical region since the query chemical could be out of the applicability domain. Finally, ADI values that are greater than or equal to 0.9 indicate chemicals that should be regarded as belonging to the applicability domain of the model.

These reference values represent a general guideline and they should be interpreted in the light of a thorough inspection of the sub-indexes that compose the ADI: the similarity index,

the concordance index, the accuracy index, and the atom-centered fragments index. If, as in the case of the CAESAR model, the chemical structures are characterized by numerical descriptors the ADI takes also into account a check of the ranges in descriptor values (*see* **Note 4**).

These critical factors should always be analyzed when interpreting results and they will be described in the following paragraphs.

2.5.1 Similarity Index

This index takes into account the degree of similarity between the query chemical and the three most similar chemicals. Values close to 1 indicate that the chemotype of the query chemical is well represented by the training set of the model (*see* **Note 5**). On the other hand, lower values could indicate that the prediction is an extrapolation since the query chemicals is located in regions of the chemical space that are scarcely populated. In this case the prediction cannot be supported by the evaluation of similar chemicals. This does not mean that the prediction is wrong. It means that the user should gather further elements to support the model results. In particular, additional models should be run to get support.

2.5.2 Concordance Index

This index provides information on the concordance between the predicted value for the query chemical and the experimental values of the three most similar chemicals. Values that are close to zero may indicate an unreliable prediction and the possible identification of a region in the chemical space whose structure–toxicity behavior is not adequately described by the model. Therefore, a careful inspection of chemicals that give rise to conflicting predictions is requested. Indeed, one or more structural analogs can be characterized by experimental values that are at odds with the prediction for the target compound.

For instance, a visual inspection may easily identify the presence of a specific toxic SA within the structure of the structural analog(s).

Consequently, two compounds that are similar from a chemical point of view may differ for the presence/absence of structural alerts, and this fact can explain differences in their property values.

If the user does not recognize SA, it is possible to run VEGA on the similar compound with the conflicting value; VEGA will list the SA, which can then be compared.

2.5.3 Accuracy Index

When assessing the reliability of predictions, it is important to understand how well a model predicts the toxicity in the region of the chemical space where the query chemical is located. This index informs on such a local reliability by taking into account the classification accuracy of the three most similar chemicals. Low values for this index should warn about a lack of predictive accuracy. In this case, additional models should be run, to see if they have better accuracy.

2.5.4 Atom-Centered Fragments (ACF) Index

This index takes into account the presence of one or more fragments that are not found in the training set, or that are rare fragments. An index value equal to 1 implies that all atom-centered fragments of the target compound were found in the training set. On the other hand, a value that is less than 0.7 implies that a prominent number of atom-centered fragments of the target compound have not been found in the compounds of the training set or are rare fragments of the training set.

Also in this case, it is recommended to run additional models, because each model can bring new information as a function of its own training set.

2.5.5 Model Descriptors Range Index

Computed only for the CAESAR model, this index checks if the descriptors calculated for the predicted compound are inside the range of descriptors of the training and test set. The index has value 1 if all descriptors are inside the range, 0 if at least one descriptor is out of the range.

2.6 Models Description

To compare the performance of three VEGA models, we applied them to the same evaluation set. This data set counts more than 6000 compounds evenly distributed between mutagens and non-mutagens and was used within the European LIFE project ANTARES for the evaluation of different QSAR models [20].

In the next paragraphs, for each model we report the statistical values referred to the entire evaluation set (6064 compounds) and to the molecules belonging to the applicability domain that are out of its training set.

2.6.1 Benigni-Bossa Mutagenicity (TT-VEGA)

TT-VEGA (version 1.0.0-DEV) is based on a series of rules defined by Benigni and Bossa that detects mutagenic chemicals [21]. This rulebase was originally implemented within the Toxtree application freely distributed by the European Joint Research Center [21].

Toxicity Data Source

Data were extracted from the ISSCAN database [22] and includes 730 compounds, 350 of which are mutagenic.

Description of the Model

Toxtree is a rule-based system that includes alerts for genotoxic carcinogenicity and non-genotoxic carcinogenicity. Genotoxic carcinogenicity alerts can be considered as a valuable tool for the detection of compounds that yield positive results during an Ames test. The version of Toxtree implemented within the VEGA platform offers the same compilation of rules as the original version [21]. This model offers a compilation of SA that refers mainly to knowledge on the mechanism of action for genotoxic carcinogenicity (i.e. they are also pertinent for mutagenic activity in bacteria). The SAs detecting non-genotoxic carcinogens are not to be taken into account when applying this model since non-genotoxic carcinogens cannot, by definition, be detected by the Ames test.

Model Statistics

- **Global performance** (calculated on 6064 compounds):
- Accuracy = 0.75, Specificity = 0.65, Sensitivity = 0.83, MCC = 0.49.
- **Performance in ADI out of training** (calculated on 1419 compounds with ADI > 0.9):
- Accuracy = 0.87, Specificity = 0.75, Sensitivity = 0.94, MCC = 0.72.

Interpretation of the Output

TT-VEGA classifies query chemicals as mutagenic when one or more SAs are detected within their molecular structure or as a non-mutagenic if no SA is identified.

2.6.2 Mutagenicity Model (CAESAR) (Version 2.1.12)

Toxicity Data Source

The CAESAR model [23] was developed on the basis of 4204 chemicals (2348 mutagenic and 1856 non-mutagenic) extracted from the Bursi data set [24]. This initial set was then split into training set (3367 chemicals, 80 % of the entire data set) and external test set (837 chemicals, 20 % of the entire data set) [24].

Description of the Model

The algorithm of the model is described in Ferrari and Gini [23]. CAESAR-VEGA automatically calculates chemical descriptors for the chemicals of interests and contains a subset of Toxtree rules (*see* previous paragraph) to enhance the sensitivity of the model. The model integrates two complementary predictive approaches in series (statistical and rule-based): a support vector machine (SVM) algorithm coupled to two sets of structural alerts rules aimed at reducing the false negative rate. In order not to inflate the false positive rate a chemical which is identified as negative during the first two steps (SVM output and first SA filter) and positive by the second set of rules is flagged as a suspicious mutagenic chemical.

If the user wants only the results of the statistical model, (s)he can check if the model identifies SA and discard this approach.

Model Statistics

- **Global performance** (calculated on 6064 compounds):
- Accuracy = 0.81, Specificity = 0.69, Sensitivity = 0.91, MCC = 0.63.
- **Performance in ADI out of training** (calculated on 942 compounds with ADI > 0.9):
- Accuracy = 0.79, Specificity = 0.61, Sensitivity = 0.93, MCC = 0.57.

During this evaluation, compounds predicted as suspicious mutagens were considered as mutagens.

Interpretation of the Output

CAESAR-VEGA classifies chemicals as mutagenic, non-mutagenic, and suspicious mutagenic. Suspicious chemicals are associated with higher predictive uncertainty.

2.6.3 Mutagenicity SARpy Model (Version 1.0.6—DEV)

The data set employed for rule extraction was retrieved from the CAESAR model for Ames mutagenicity (*see* previous paragraph). This model and VEGA CAESAR share the same training set.

Toxicity Data Source

Description of the Model

SARpy (SAR in python) is a QSAR method that identifies relevant fragments and extracts a set of rules directly from data without any a priori knowledge [25]. The algorithm generates substructures; relevant SAs are automatically selected on the basis of their prediction performance for a training set. The application of this modeling approach to the CAESAR data set extended the previous work [25] by extracting two sets of rules: one for mutagenicity (112 rules) and the other for non-mutagenicity (93 rules) (*see* **Note 6**).

The SARpy application is available through a graphic interface or through the VEGA platform.

Model Statistics

- **Global performance** (calculated on 6064 compounds):
- Accuracy = 0.77, Specificity = 0.71, Sensitivity = 0.82, MCC = 0.54.
- **Performance in ADI out of training** (calculated on 880 compounds with ADI > 0.9):
- Accuracy = 0.81, Specificity = 0.67, Sensitivity = 0.92, MCC = 0.62.

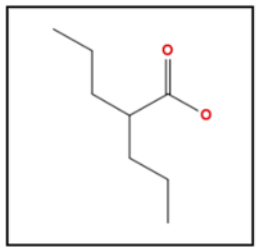
Interpretation of the Output

If the target compound matches one or more mutagenicity rules, the prediction will be "mutagenic"; if the target compound matches one or more non-mutagenicity rules (or no rules), the prediction will be "non-mutagenic."

3 Methods

A critical assessment of predictions is the most demanding aspect related to the interpretation of the output of (Q)SAR models. VEGA facilitates the interpretability of (Q)SAR predictions by breaking down several critical aspects of the applicability domain as described in Subheading 2.5. Nevertheless, possible misinterpretations can still take place and the following examples will provide further insights into the analysis of (Q)SAR results.

The first two examples illustrate predictions characterized by a clear output which is concordant across all VEGA models. On the contrary, the last example is more challenging and it will advise the reader about complex cases. The purpose of this section is to provide an insight into the critical assessment of QSAR predictions and to highlight relevant aspects that should be taken into account when analyzing (Q)SAR outputs.



Systematic Name: (2-Propylvaleric acid)
CAS Registry Number: 99-66-1
VEGA SMILES: O=C(O)C(CCC)CCC
Experimental activity: Non-mutagenic in Ames test

Fig. 2 Valproic acid structure, chemical information, and experimental activity [26]

3.1 Case Study: Valproic Acid (Fig. 2)

- **CAESAR results**: *Prediction is non-mutagenic and the result appears reliable.*

The CAESAR model does not identify any SA linked to mutagenic activity.

Similarity values for the six most similar compounds are very high (ranging from 0.989 to 0.903). Furthermore, experimental and predicted toxicities agree for all the similar molecules that were found in the training set. Indeed, predicted and experimental toxicities systematically designate non-mutagenic chemicals (*see* **Note 7**).

On the basis of this information and in particular thanks to a visual inspection of the first three similar compounds, the predicted substance is considered into the applicability domain of the model (ADI = 0.978) (*see* Fig. 3).

- **SARpy results**: *Prediction is non-mutagenic and the result appears reliable.*

The model finds within the structure of the query chemical only SAs for non-mutagenicity ("Inactive" rules) (*see* Fig. 4).

Also in this case, the query chemical falls into the applicability domain (ADI = 0.978) and the predicted and experimental toxicities for the most similar compounds are the same. This behavior is not completely surprising since CAESAR and SARpy are based on the same training set. Nevertheless, this result corroborates the prediction computed by CAESAR by assessing toxicities according to a complementary analysis executed by a different algorithm.

- **TT-VEGA results**: *Prediction is non-mutagenic and the result appears reliable.*

Similarly to what described for the CAESAR model, Toxtree does not find any SA for mutagenicity.

The most similar compounds shown in the output are different from those of CAESAR and SARpy since the corresponding training sets are different. These structural analogs are characterized by lower similarity values (ranging from 0.823 to 0.773) and this lower degree of similarity is reflected by the ADI (0.906). This degree of overall similarity combined with a lack of identification of SA substantiates the validity of the prediction (*see* **Note 8**).

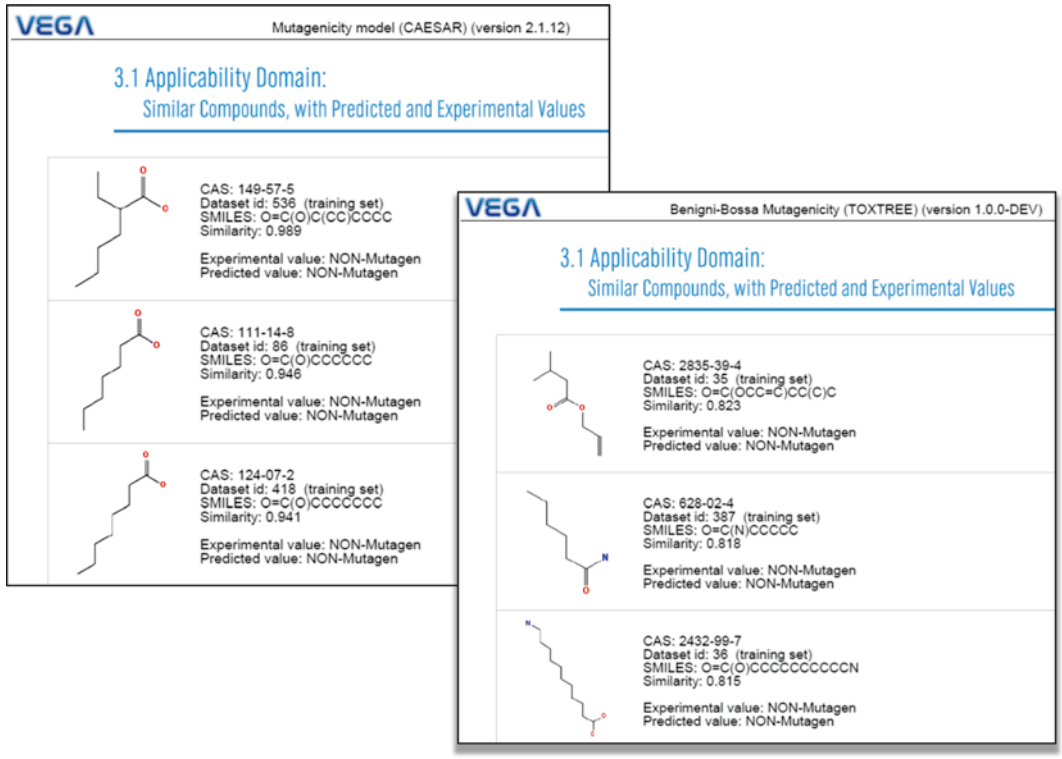


Fig. 3 A particular of the three on six most similar compounds that are shown in the pdf outputs of the models for Valproic acid. SARpy and CAESAR display the same molecules

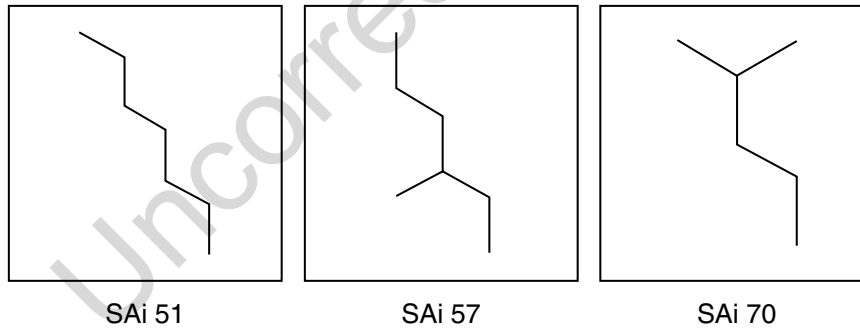


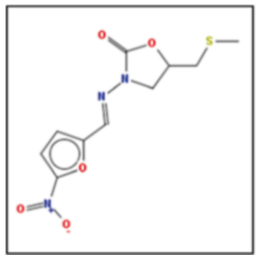
Fig. 4 Inactive SAs identified by SARpy in Valproic acid molecule

- **Overall evaluation**: *for this case there is agreement between the three models, and each model is corroborated by the high ADI value.*

3.2 Case Study: Nifuratel (See Fig. 5)

- **CAESAR results:** *Prediction is mutagenic and the result appears reliable.*

The model identifies one fragment related to mutagenic activity included within the Benigni-Bossa rulebase [21]: Nitro aromatic, SA27 (*see* Fig. 6).



Systematic Name: (2-Oxazolidinone, 5-((methylthio)methyl)-3-(((5-nitro-2-furanyl) methylene)amino)-
CAS Registry Number : 4936-47-4
VEGA SMILES: O=C2OC(CN2(N=Cc1oc(cc1)[N+](=O)[O-]))CSC
Experimental activity: Mutagenic in Ames test

Fig. 5 Nifuratel structure, chemical information, and experimental activity [27]

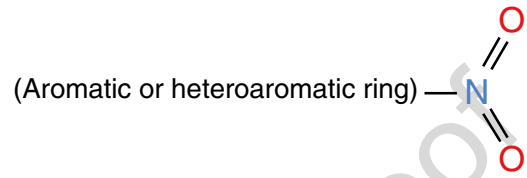


Fig. 6 Nitro aromatic structural alert no. 27

In addition to the six most similar molecules found in the training set, the model shows the three most similar compounds having the same fragment (*see* Fig. 7).

The similarity index is high, 0.9. The concordance for similar molecules and the accuracy index are both equal to 1.

For these reasons the predicted substance is considered into the applicability domain (ADI = 0.948).

- **SARpy results:** *Prediction is mutagenic and the result appears reliable.*

In this case the identified fragments are four and all linked to mutagenic activity (*see* Fig. 8).

SARpy also shows the most similar compounds that are characterized by the presence of the identified fragments. In this case predictions and experimental values agree for all the structural analogs.

This prediction is characterized by the same ADI (and his subindexes) as the prediction computed by the CAESAR model.

- **Toxtree results**: *Prediction is mutagenic and the result appears reliable.*

As explained in Subheading 2.6, CAESAR contains a subset of Toxtree rules and both models identify the same nitro aromatic fragment that plays a key role in supporting the prediction.

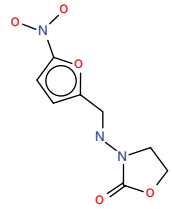
The ADI value (0.933) is slightly lower than what observed for CAESAR and SARpy; this is related only to the index of similarity (0.871) while the other indices are all excellent.

- **Overall evaluation**: *all models agree, and there are good examples of similar compounds suggesting the predictions.*

Fragment found: Nitro aromatic

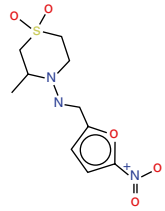
Nitro aromatic (Benigni/Bossa structural alert no. 27).

Following, the most similar compounds from the model's dataset having the same fragment.



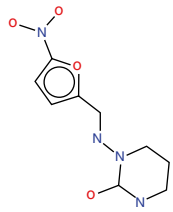
CAS: 67-45-8
Dataset id: 1522 (training set)
SMILES: O=C2OCCN2(N=Cc1oc(cc1)[N+](=O)[O-])
Similarity: 0.922

Experimental value: Mutagen
Predicted value: Mutagen



CAS: 23256-30-6
Dataset id: 2794 (training set)
SMILES: O=[N+]([O-])c2oc(C=NN1CCS(=O)(=O)CC1C)cc2
Similarity: 0.899

Experimental value: Mutagen
Predicted value: Mutagen



CAS: 75888-03-8
Dataset id: 3089 (training set)
SMILES: O=C2NCCCN2(N=Cc1oc(cc1)[N+](=O)[O-])
Similarity: 0.882

Experimental value: Mutagen
Predicted value: Mutagen

Fig. 7 Part of the CAESAR output in Nifuratel prediction: three of the most similar compounds within training set that have the same SA27 fragment found in the target

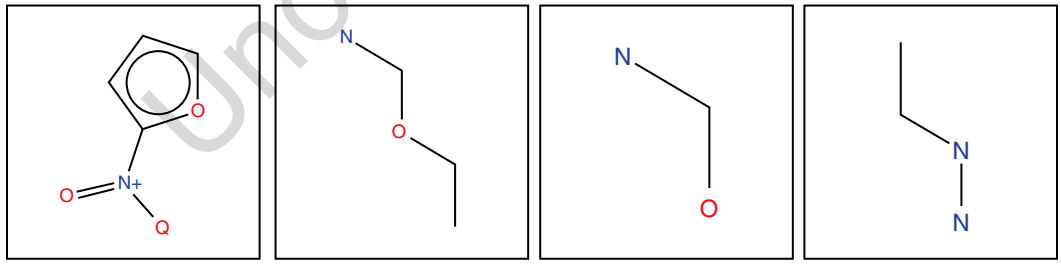
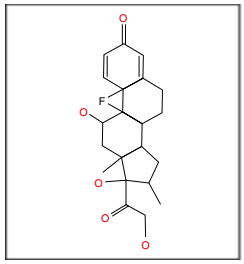


Fig. 8 Four active fragments identified by SARpy

3.3 Case Study: Dexamethasone (See Fig. 9).

Unlike the previous examples, in this case the output is equivocal because the prediction models are in disagreement and show very low values of ADI.

- **CAESAR results**: *Prediction is non-mutagenic but the result may not be reliable.*



Systematic Name: Pregna-1,4-diene-3,20-dione, 9-fluoro-11,17,21-trihydroxy-16-methyl-, (11beta,16alpha)-
CAS Registry Number: 50-02-2
VEGA SMILES:
O=C1C=CC3(C(=C1)CCC2C4CC(C)C(O)(C(=O)CO)C4(C)(CC(O)C23(F)))(C)
Experimental activity : Non mutagenic in Ames test

Fig. 9 Dexamethasone structure, chemical information, and experimental activity [28]

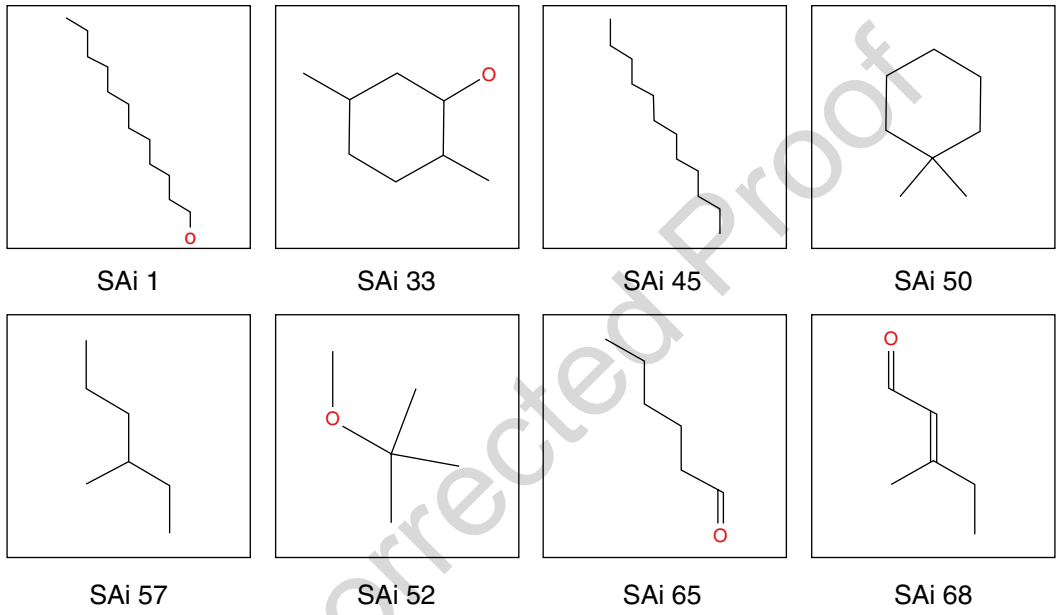


Fig. 10 Examples of inactive fragments identified by SARpy

Although similarity, concordance, and accuracy indices are high (respectively 0.875, 1 and 1), ADI is equal to 0.795, therefore Dexamethasone could be out of the applicability domain of the model. This lack of reliability is caused by a low (0.85) value of the ACF index.

- **SARpy results**: *Prediction is non-mutagenic but the result may not be reliable.*

The model identifies nine inactive fragments. Some of these fragments are depicted in Fig. 10.

Even if the values of similarity and ACF indexes are the same than what observed when using CAESAR, the ADI (0.721) value is lower because the accuracy does not reach the minimal recommended threshold (0.676) (*see* Fig. 11).

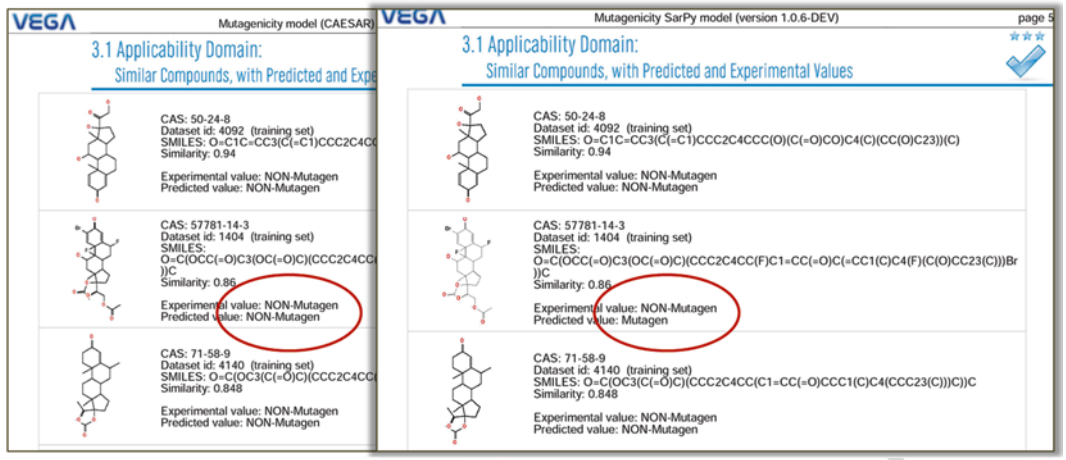


Fig. 11 The *red circle* indicates the different predictions of CAESAR and SARpy for the second most similar compound. Since the prediction computed by SARpy does not match the experimental activity, its accuracy is lower than what observed when using CAESAR

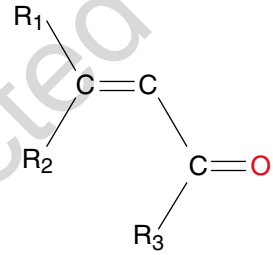


Fig. 12 SA10 (α, β unsaturated carbonyl)

- **TT-VEGA results**: *Prediction is mutagenic but the result may not be reliable.*

The model identifies the presence of the SA10 as a cause of mutagenicity of the target compound (*see* Fig. 12).

Conversely to what observed for Nifuratel, the predictions yielded by CAESAR and TT-VEGA are in disagreement since CAESAR does not contain the SA10 fragment in its subset of rules (see above).

The unreliability of the TT-VEGA prediction is highlighted by the poor value of its ADI (0) that is determined by low values of the concordance, accuracy, and ACF indices (0, 0, and 0.6 respectively).

Indeed, even if the prediction yielded by TT-VEGA is characterized by a similarity index which is greater (0.922) than the corresponding index of CAESAR and SARpy, the experimental and the predicted values are in disagreement for all the similar compounds in the output.

Difficult cases such as this example could benefit from tools such as ToxRead (*see* Chapter 13) that can perform read-across analysis while providing p-values calculated by using the Fisher's test and accuracies for each structural alert. In this case ToxRead could provide an insight into the analysis of the SA10 fragment by showing its low accuracy (0.49) and *p*-value (0.015).

On the contrary, the nine fragments identified by SARpy have accuracies ranging from 0.7 to 0.9 and *p*-values $<10^{-6}$.

The examples detailed in the previous paragraphs highlight the fact that a thorough analysis of all the factors that influence the predictive accuracy of a model should be taken into account instead of simply relying on the final prediction. Several potential pitfalls can be prevented by analyzing all the sub-indices that compose the ADI and by a visual inspection of the input molecule versus all the identified structural analogs. Particularly, the pertinence of such a visual inspection can be corroborated by the recognition of SA within the query chemical and/or its structural analogs.

It is also important to point out that QSAR and read-across predictions are not mutually exclusive and that such a synergy can potentially provide relevant information in difficult cases that are characterized by fuzzy QSAR predictions (e.g. the case of Dexamethasone). Indeed, an expert can always compare the results computed by a model with its own read-across prediction on the basis of the identified analogs. These concepts will be discussed in Chapter 13.

4 Notes

1. The predictive models discussed in this chapter do not predict for a specific *S. typhimurium* strain. On the other hand, ADMET predictor (Absorption, Distribution, Metabolism, Elimination, and Toxicity of chemical substances), a commercial tool, includes ten different models for different strains of *S. typhimurium* with and without microsomal activation [29]. We notice that the performance of the "general" mutagenicity models was superior compared to the strain-specific models, when tested in a large set of compounds [20].
2. There are several commercial or freely available software programs that can predict mutagenic hazards. In addition to the VEGA platform, other examples of free models are T.E.S.T.

(Toxicity Estimation Software Tool) [30] and Toxtree (Estimation of Toxic Hazard—A Decision Tree Approach) by Ideaconsult Ltd. [31].

3. VEGA calculates the applicability domain through a program which is different from the (Q)SAR model predicting the value of interest.
4. The ADI measurement within VEGA is composed of a series of sub-indices which vary depending on the (Q)SAR model.
5. For the models embedded within the VEGA platform, the expression "training set" refers to the set of molecules used during the calibration of the models and their internal validation. The membership of the most similar structural analogs of the query chemical (training or test set) is specified in the output provided by the software.
6. The output format is different for TEST. In this case the output shows the most similar structural analogs of the query chemical that are found in the test set and, if prompted by the user, it also shows the most similar compounds identified in the training set.
7. SARpy adopts SAs but these fragments are not based on "a priori" knowledge of the biochemical mechanism of action like for the rules-based systems (such as Toxtree and DEREK); it is more correct to refer to SARpy as a statistical model, which is highly transparent and communicates the extracted knowledge by means of rules.
8. Another major difference between SARpy and the rule-based models is that SARpy shows rules associated with lack of toxicity. These fragments are most frequently present in the non-mutagenic compounds of the training set. However, considering the SA for mutagenicity there are strong similarities with rule-based models.
9. The evaluation on the similar compounds carried out by using VEGA can be regarded as a kind of read-across approach. The user may also apply VEGA for read across, without considering the prediction done by the model.
10. Please notice that each model in VEGA has its own data set. Also the ADI is based on this data set. It may be that the same chemical is characterized by conflicting properties value (mutagenic or non-mutagenic) depending on the data set.

Acknowledgements

We acknowledge the LIFE + CALEIDOS project.

References

1. Moore MM, Myers MB, Heflich RH (2000) Mutagenesis and genetic toxicology. In: Williams PL, James RC, Roberts SM (eds) Principles of toxicology: environmental and industrial applications, 2nd edn. Wiley-Interscience, New York, pp 239–264
2. Sutter A, Amberg A, Boyer S et al (2013) Use of *in silico* systems and expert knowledge for structure-based assessment of potential mutagenic impurities. Regul Toxicol Pharmacol 67:39–52
3. Regulation (EC) No. 1907/2006 of the European Parliament and of the Council, of December 18, 2006 concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission
4. Ames BN, McCann J, Yamasaki E (1975) Methods for detecting carcinogens and mutagens with the salmonella/mammalian-microsome mutagenicity test. Mutat Res 31: 347–364
5. Claxton LD, Umbuzeiro GA, DeMarini DM (2010) The *Salmonella* mutagenicity assay: the stethoscope of genetic toxicology for the 21st century. Environ Health Perspect 118: 1515–1522
6. Benigni R, Bossa C, Battistelli CL et al (2013) IARC classes 1 and 2 carcinogens are successfully identified by an alternative strategy that detects DNA-reactivity and cell transformation ability of chemicals. Mutat Res 758:56–61
7. OECD (1997) Test No. 471: bacterial reverse mutation test, OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
8. Piegorsch W, Zeiger E (1991) Measuring intra-assay agreement for the Ames Salmonella assay. In: Hotorn L (ed) Statistical methods in toxicology, lecture notes in medical informatics, vol 43. Springer-Verlag, Berlin, pp 35–41
9. Sushko I, Novotarskyi S, Körner R et al (2010) Applicability domains for classification problems: benchmarking of distance to models for Ames mutagenicity set. J Chem Inf Model 50:2094–2111
10. Miller JA, Miller EC (1977) Ultimate carcinogens as reactive mutagenic electrophiles. In: Hiatt HH, Watson JD, Winston JA (eds) Origins of human cancer, mechanisms of carcinogenesis, Book B. Cold Spring Harbor Laboratory, New York, pp 605–627
11. Cooper JA, Saracci R, Cole P (1979) Describing the validity of carcinogen screening tests. Br J Cancer 39:87–89
12. Virtual models for evaluating the properties of chemicals within a global architecture. http://www.vega-qsar.eu/download.html. Accessed 19 Jun 2015
13. SMILES, Simplified Molecular Input Line Entry System. http://www.daylight.com/smiles/index.html. Accessed 19 Jun 2015
14. Drug Design Laboratory. http://nova.disfarm.unimi.it/cms/index.php?Software_projects. Accessed 19 Jun 2015
15. ACD/ChemSketch for Academic and Personal Use. http://www.acdlabs.com/resources/freeware/chemsketch/. Accessed 19 Jun 2015
16. Marvin, intuitive applications and API for chemical sketching, visualization and data exploration. http://www.chemaxon.com/products/marvin/. Accessed 19 Jun 2015
17. The OECD QSAR Toolbox. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm. Accessed 19 Jun 2015
18. Mombelli E, Devillers J (2010) Evaluation of the OECD (Q)SAR Application Toolbox and Toxtree for predicting and profiling the carcinogenic potential of chemicals. SAR QSAR Environ Res 21:731–752
19. Benfenati E, Pardoe S, Martin T (2013) Using toxicological evidence from QSAR models in practice. ALTEX 30:19–40
20. Bakhtyari NG, Raitano G, Benfenati E et al (2013) Comparison of *in silico* models for prediction of mutagenicity. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 31:45–66
21. Benigni R, Bossa C, Jeliazkova N et al (2008) Benigni/Bossa rulebase for mutagenicity and carcinogenicity—a module of toxtree. JRC scientific and technical reports. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_23241_EN.pdf. Accessed 19 Jun 2015
22. Istituto Superiore di Sanità. "Chemical carcinogens structures and experimental data" (ISSCAN). http://www.epa.gov/ncct/dsstox/sdf_isscan_external.html. Accessed 19 Jun 2015
23. Ferrari T, Gini G (2010) An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chem Cent J 4:S2
24. Kazius J, McGuire R, Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem 48:312–330

25. Ferrari T, Cattaneo D, Gini G et al (2013) Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. SAR QSAR Environ Res 24: 365–383
26. Hansen K, Mika S, Schroeter T et al (2009) Benchmark data set for in silico predictions of Ames mutagenicity. J Chem Inf Model 49:2077–2081
27. Byeon WH, Hyun HH, Lee SY (1976) Mutagenicity of nitro furan nitroimidazol and nitrothiazole derivatives on salmonella microsome system. Kor J Microbiol 14:151–158
28. Singh H, Singh JR, Dhillon VS et al (1994) In vitro and in vivo genotoxicity evaluation of hormonal drugs II. Dexamethasone. Mutat Res 308:89–97
29. Simulations Plus, Inc. http://www.simulations-plus.com. Accessed 19 Jun 2015
30. Quantitative structure activity relationship, Toxicity Estimation Software Tool (TEST). http://www.epa.gov/nrmrl/std/qsar/qsar.html. Accessed 19 Jun 2015
31. Toxtree—toxic hazard estimation by decision tree approach. http://toxtree.sourceforge.net. Accessed 19 Jun 2015

Chapter 6

In Silico Methods for Carcinogenicity Assessment

Azadi Golbamaki and Emilio Benfenati

Abstract

Screening compounds for potential carcinogenicity is of major importance for prevention of environmentally induced cancers. A large sequence of alternative predictive models, ranging from short-term biological assays (e.g. mutagenicity tests) to theoretical models, have been attempted in this field. Theoretical approaches such as (Q)SAR are highly desirable for identifying carcinogens, since they actively promote the replacement, reduction, and refinement of animal tests. This chapter reports and describes some of the most noted (Q)SAR models based on the human expert knowledge and statistically approach, aiming at predicting the carcinogenicity of chemicals. Additionally, the performance of the selected models has been evaluated and the results are interpreted in details by applying these prediction models to some pharmaceutical molecules.

Key words Carcinogenicity, Structural alerts, Genotoxicity, Non-genotoxicity, QSAR, In silico, Toxtree, SARpy, Applicability domain index

1 Introduction

The study of the chemical carcinogenesis mechanisms and determining the safety of the existing and the new chemicals are of increasing importance and necessity to protect human health. From the point of view of mechanism of action, the carcinogens are classified into: (a) genotoxic carcinogens, which cause damage to DNA—many known mutagens are in this category, and often mutation is one of the first steps in the development of cancer [1]; and (b) epigenetic or non-genotoxic carcinogens that do not bind covalently to DNA, do not directly cause DNA damage, and are usually negative in the standard mutagenicity assays [2]. The unifying feature of all genotoxic carcinogens is that they are either electrophiles or can be activated to electrophilic reactive intermediates. This fact has been originally proposed by the Miller's [3, 4]. On the contrary, non-genotoxic carcinogens act through a large variety of different and specific mechanisms.

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_6, © Springer Science+Business Media New York 2016

The mechanisms of action and the metabolic fate of a large number of carcinogens have been already investigated. These studies shed light on the structural features that were frequently present in carcinogenic compounds. Several chemical functional groups and structural alerts (SAs) were identified by researchers through analysis of the results of experimental (veterinary laboratory) carcinogenicity tests. These compounds were mainly genotoxic carcinogens as supported by the specific results from tests for genotoxicity (Ames test [5], Micronucleus assay [6], etc.). Diversely, the recognition of SAs for non-genotoxic carcinogens is far behind, because no unifying theory provides scientific support. A number of SAs and characteristics of several types of non-genotoxic carcinogens have been summarized by Woo et al. [2] (*see* **Notes 1** and **2**).

The long-term carcinogenesis bioassays using animal testing methods have played a central role in assessment of chemical's carcinogenicity, however, for ethical and practical reasons their use is dramatically declining, and the genotoxicity short-term tests have taken the pivotal role in the pre-screening of carcinogenicity. The need to reduce animal testing, time, and cost in the process of assessment of carcinogenicity of chemicals had lead to an increased use of in silico methods as toxicological risk assessment tools. Among the in silico methods, the use of (Q)SAR models is supported by several legislative authorities (REACH [7]) upon fulfillment of the required characteristics of a (Q)SAR model according to the indications reported by different legislations. This goes hand in hand with the progress made in the field of the computational predictive models to date.

(Q)SARs are often incorporated into expert systems. An expert system is any formalized system that is mostly computer-based, and that can be used to make predictions based on prior information [8].

There are many (Q)SAR models published in the literature for predicting genotoxicity and carcinogenicity. The most commonly modeled endpoint for genotoxicity is the Ames test mutagenicity. The application of the Ames test to large numbers of chemicals has shown that this test has a high predictivity for chemical carcinogens (around 80 %) [9]. Most models are classifiers that predict a chemical compound as genotoxic (and thus carcinogenic) or not. Since the recognition of non-genotoxic carcinogenicity SAs is not extended compared to genotoxic SAs, few models are available for identifying non-genotoxic carcinogens [10]. While the SAs for genotoxic carcinogens have been identified to a high extent and used widely within predictive models for genotoxicity, the SAs for identifying non-genotoxic carcinogens are still a concern for the investigators. Benigni et al. (Toxtree 2.6.0) have recently enhanced the set of non-genotoxic SAs that captures carcinogens [9]. This list can provide a considerable insight to the possible variety of

mechanism of actions underlying the non-genotoxic carcinogenicity. Hence, the approaches for (Q)SAR analysis and identification of SAs for non-genotoxic carcinogens differ accordingly to their specific mechanism of action of these chemicals (interaction with proteins, DNA replication enzymes, etc.) (*see* **Note 1**). A number of SAs and characteristics of several types of non-genotoxic carcinogens have been summarized and discussed by Woo et al. [2].

However, statistical-based models will provide predictions that are based on the knowledge acquired from the training set that had been used to develop the model. In fact, these models are suitable in predicting both genotoxic and non-genotoxic carcinogens. For unknown non-genotoxic SAs, the statistical-based models can fill the information gap. In other words, these models may provide insight into the recognition of the missing information in the SAs list developed by human experts by investigation through experimental results mostly based on the Ames test.

In the context of prediction of carcinogenicity by (Q)SAR models, it is essential to integrate results from both expert systems and statistical-based models. This approach will considerably improve the prediction performance of (Q)SARs.

There are several commercial and non-commercial expert systems for predicting genotoxicity and carcinogenicity [11, 12]. Freely available models include VEGA-CAESAR [13], SARpy [14], Toxtree [15], OncoLogic [16], OECDE Toolbox [17], and lazar [18]. Alternately, MultiCASE [19], TOPKAT [20], HazardExpert [21], and DEREK [22, 23] are some of the most common commercial expert system.

Expert systems are based on three main modeling approaches which are rule-based, statistical-based, or hybrid methods [24]. Rule-based methods codify the human rules which identify certain potential molecular fragments responsible for carcinogenicity. Statistical models extract the information from a set of chemicals by using data mining methods [25].

Rule-based systems combine toxicological knowledge, expert judgment, and fuzzy logic. OncoLogic, DEREK, HazardExpert as well as implemented modules in Toxtree and the OECD Toolbox are rule-based systems.

Statistical-based systems use a variety of statistical, rule-induction, artificial intelligence, and pattern recognition techniques to build models from different databases used as training sets. For example, MultiCASE and TOPKAT are commercial statistical-based models while lazar and VEGA-CAESAR are statistical-based and publicly available. Additionally, most of the models published in the literature but not implemented are statistical-based (*see* **Note 2**).

A description of some of the most common non-commercial (Q)SAR models is provided below. Three case studies are given in this chapter to illustrate the use and the performance of a number of these models.

2 QSAR Models for Carcinogenicity

2.1 VEGA-CAESAR (Version 1.1.0)

CAESAR is a model implemented in the VEGA platform [26]. This model uses a statistical-based approach to generate categorical carcinogenicity models. CAESAR is based on the counter-propagation artificial neural network (CP ANN) algorithm. Artificial neural networks (ANNs) as a statistical approach appear to be suitable and promising for prediction of carcinogenicity for dissimilar data sets of chemicals. One of the main advantages of ANNs is that non-linear relationships can be modeled without any assumptions about the form of the model.

2.2 Toxtree (Version 2.6.0)

Toxtree is a standalone expert rule-based SAR program. This application is a classifier that places chemicals into categories and predicts various kinds of toxic effect by applying decision tree approaches, including the Begnini-Bossa rule-base for mutagenicity and carcinogenicity [27]. The Toxtree module applies human expert rules developed by Begnini and Bossa to identify SAs for mutagenicity and carcinogenicity that may be present in a chemical structure. Carcinogenic SAs are functional groups or substructures that are mechanistically and/or statistically associated with the induction of cancer. Begnini-Bossa SAs for the prediction of mutagenicity and carcinogenicity are highly correlated with Ames mutagenicity. The Begnini-Bossa system contains a list of SAs for the evaluation of carcinogenicity. Structural features represented in the system are easy to understand and interpretable since they have a mechanistic foundation. Toxtree offers additional QSAR models for aromatic amines and alpha, beta-unsaturated aldehydes. The Toxtree output contains "structural alert for genotoxic carcinogenicity" that shows the presence or absence of a SA for Salmonella mutagenicity, and "structural alert for non-genotoxic carcinogenicity" that indicates the presence or absence of a non-genotoxic (epigenetic) SA.

2.3 SARpy (Version 1.0)

SARpy is a desktop software based on a statistical modeling approach. Through a data mining method, SARpy extracts relevant fragments (molecular substructures) from the analysis of the correlation between the structure, written with simplified molecular input line entry system (SMILES) format, and the endpoint. Using SARpy, and a data set of chemicals with valid experimental results (binary categorical data), users can develop new classification models. SARpy is able to extract both "ACTIVE" (e.g. carcinogenic) and "INACTIVE" (e.g. non-carcinogenic) fragments from chemical structures. In order to discover new carcinogenic SA, we combined three different carcinogenesis databases as a training set and by the aid of SARpy, developed a new carcinogenicity model which consists of a rule set or a collection of SMARTS with their likelihood ratio values in the mentioned training set.

The data gathered for the development of this new rule set are carcinogenicity data collections based on studies on different species. In particular, the data in the training set are a combination of: (1) the carcinogenicity data set (rat) of the EU-funded ANTARES project [28]; (2) the long-term carcinogenicity bioassay on rodents (rat and mouse) ISSCAN data set [29]; and (3) the carcinogenicity (rat and mouse) data set provided by Kirkland et al. [30]. The data set (1680 chemicals together with their carcinogenicity data) built as described above was used as the training set for the extraction of rules. SARpy extracted more than 100 rules from which by applying a human expert judgment we selected 130 rules. The human expert selection aimed to delete the alerts that produced a high number of false negative or false positive predictions. The performance of this model, as tested on the test set obtained from eChemPortal inventory (258 compounds), was as follows: accuracy = 0.67, sensitivity = 0.62, specificity = 0.70.

2.4 OncoLogic™ (Version 8.0)

OncoLogic™ [31] is a desktop computer program released by the U.S. Environmental Protection Agency (EPA) [32] that evaluates the likelihood that a chemical may cause cancer. OncoLogic™ predicts cancer-causing potential by: applying the rules of structure–activity relationship (SAR) analysis, mimicking the decision logic of human experts, and incorporating knowledge of how chemicals cause cancer in animals and humans. This version of the software has a new CAS/name look-up feature under the "Organics SAR" module for approximately 1500 chemicals for which available cancer data can be used directly to create a chemical report. This removes the need to draw the chemical structure for these substances as was necessary in the previous versions of the software.

2.5 Lazar

Lazy structure–activity relationships (lazar) [18] is a standalone program with k-nearest-neighbor approach which can predict chemical endpoints from a training set based on structural fragments. It uses a SMILES file and precomputed fragments with occurrences as well as target class information for each compound as training input. It also features regression, in which case the target activities consist of continuous values. Lazar uses activity-specific similarity (i.e. each fragment contributes with its significance for the target activity) that is the basis for predictions and confidence index for every single prediction.

3 Case Studies

3.1 Case Study 1

An example of Toxtree (v.2.6.0) carcinogenicity prediction.

As it is explained in the Toxtree user manual [33] for estimating carcinogenicity with Toxtree, the following steps should be taken: After launching Toxtree in Windows™ platform, first,

the chemical structures for analysis may be submitted by inserting directly the SMILES, or by using an interactive 2D graphical editor, or in a batch mode by using CSV, TXT, or SDF file formats. Second, among the list of decision tree modules the user may select "carcinogenicity (genotox and non-genotox) and mutagenicity rule-base by ISS" [27] option from the Method menu. Finally, in order to apply the active decision tree on the current compound, the Estimate button should be pressed. If one or more genotoxic or non-genotoxic SA are found in the molecular structure, the name and the identification number of that SA are indicated in the graphical user interface, and the chemical is predicted as carcinogen. Otherwise, the prediction result will be non-carcinogen. Figure 1 shows an example of classification result visualization.

Captafol is an antibacterial drug and fungicide and is categorized as a carcinogen in the Carcinogenic Potency Database (CPDB) [34]. Toxtree v. 2.6.0 finds a SA for genotoxic carcinogenicity (QSA8_gen.Aliphatic halogens) and a SA for non-genotoxic carcinogenicity (QSA50_nogen.dicarboximid) in this chemical structure. By clicking on the name of these two SAs, they become highlighted and the user can see their position in the chemical structure (Fig. 2). The classification results can be saved as a file (CSV, SDF, or TXT format), together with the list of applied SAs.

3.2 Case Study 2

2-Amino-5-nitrothiazole or aminonitrothiazole is an antiprotozoal drug. Antiprotozoal agent is a class of pharmaceuticals used in the treatment of protozoan infection. Figure 3 shows the chemical structure and Table 1 shows the carcinogenicity test summary report as published by the CPDB [34]. Based on the experimental results of TD_{50} on rat species, this chemical is considered as a carcinogen.

VEGA-CAESAR (v. 1.1.0), lazar, Toxtree (v. 2.6.0), and the SARpy (v. 1.0) model predicted this chemical correctly as carcinogen. Figure 4 shows two genotoxic SAs found in the chemical structure of 2-amino-5-nitrothiazole: "SA_27: Nitro-aromatic" and "SA_28: primary aromatic amine, hydroxyl amine and its derived esters". VEGA-CAESAR returned applicability domain (AD) index of 0.5 for the prediction of this drug, and the explanation is "the predicted compound is outside the AD of the model." The "measured activity" of lazar given in the output is "Experimental result(s) from the training data set," so the chemical is inside the AD of the program. Toxtree and SARpy do not report any AD index in their predictions.

Performing prediction with the model constructed by means of SARpy for this chemical, an additional fragment is recognized as responsible for the carcinogenicity property. Figure 5 shows the SA found by this model. Overall, based on these multiple predictions, we can see that there is agreement, even though each model has a different level of reliability.

As a conclusion, all evidences point toward a carcinogenic effect.

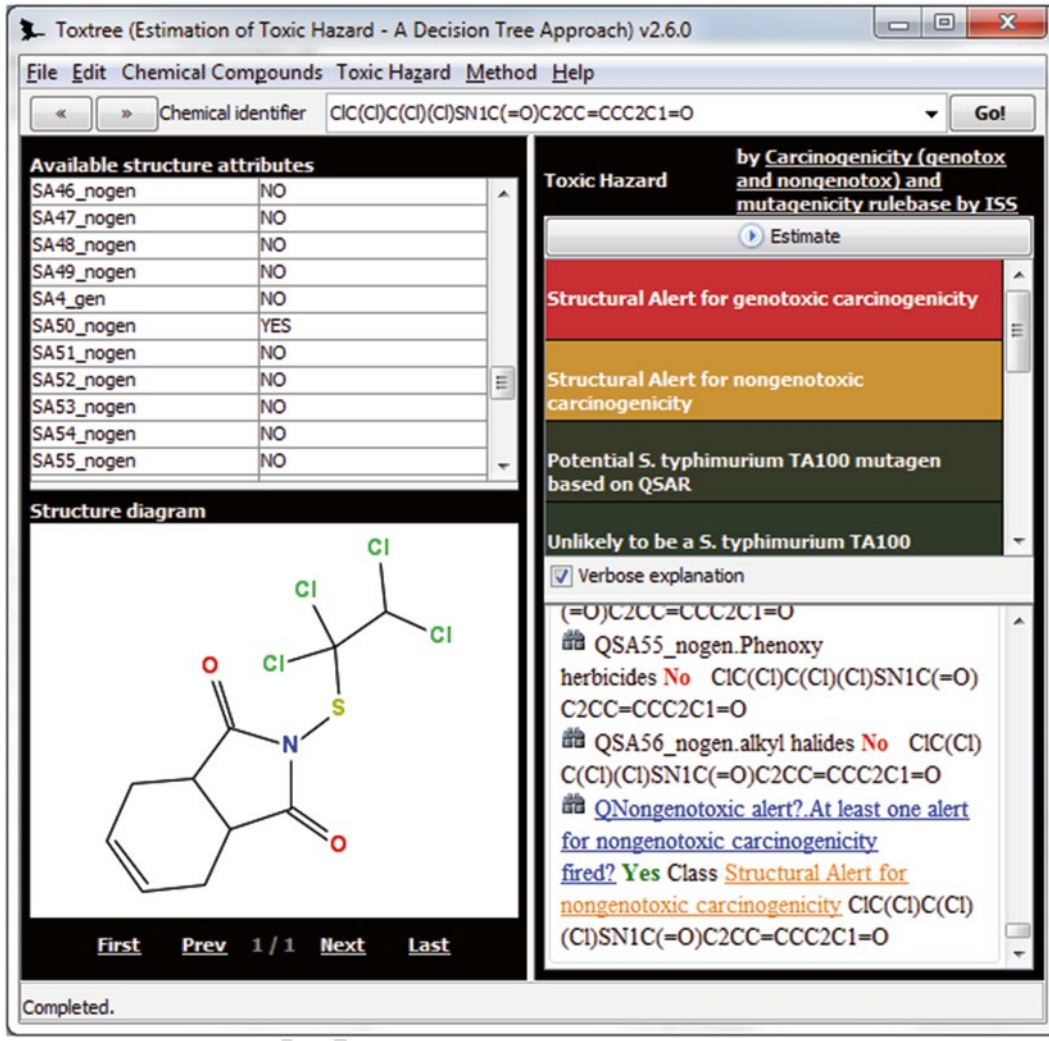


Fig. 1 Toxtree v. 2.6.0 mutagenicity and carcinogenicity prediction for Captafol

3.3 Case Study 3

Bemitradine is an antihypertensive, vasodilator agent, and a diuretic. Figure 6 shows the chemical structure and Table 2 shows the carcinogenicity test summary report as published by the CPDB. Based on the experimental results of TD_{50} on rat species, this chemical is considered as carcinogen.

Toxtree (v. 2.6.0) and SARpy (v. 1.0) model predicted this chemical correctly as carcinogen; conversely, VEGA-CAESAR (v. 1.1.0) and lazar prediction for this chemical was non-carcinogen. Figure 7 shows the genotoxic SA found in the chemical structure, whereas the model constructed by means of SARpy matched another fragment to the molecular structure as responsible for the carcinogenicity property. Figure 8 shows the SA found by the SARpy model. Toxtree and SARpy do not have any

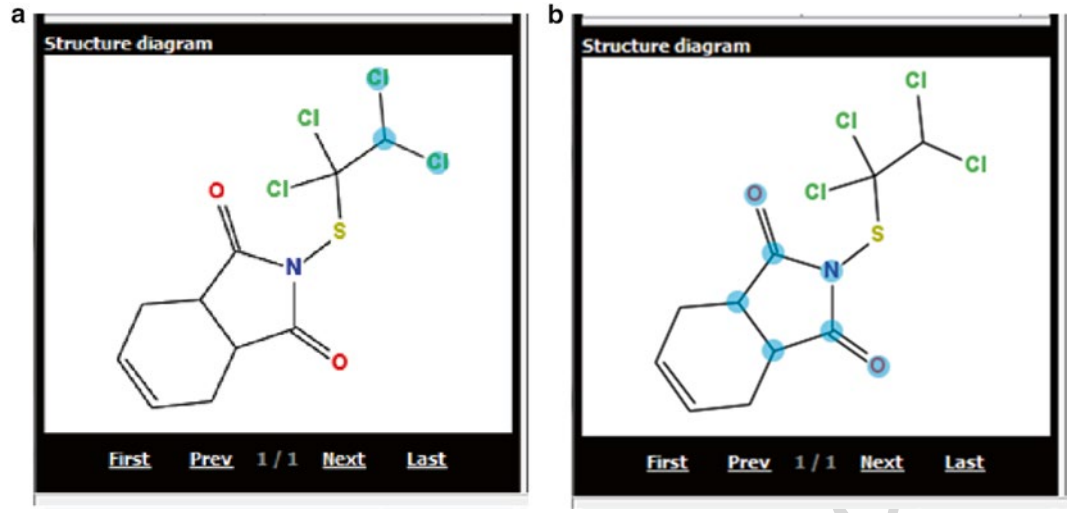


Fig. 2 Genotoxic and non-genotoxic structure alerts found by Toxtree 2.6.0 for Captafol; (**a**) QSA8_gen.Aliphatic halogens; (**b**) QSA50_nogen.dicarboximid are highlighted in the molecular structure

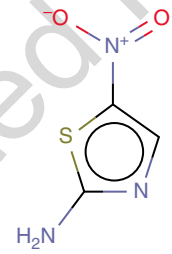


Fig. 3 2-Amino-5-nitrothiazole, with CAS number: 121-66-4 and SMILES: O=[N+]([O–])c1cnc(N)s1

t1.1 **Table 1**

t1.2 **Cancer test summary reported in the CPDB for 2-amino-5-nitrothiazole**

t1.3	Rat target sites		Mouse target sites		TD50 (mg/kg/day)	
t1.4	Male	Female	Male	Female	Rat	Mouse
t1.5	No positive	kid lun mgl[a]	No positive	No positive	44.6	No positive

t1.6 *akid* kidney, *lun* lung, *mgl* mammary gland

277 AD index along with their prediction results. The AD index of
278 VEGA-CAESAR for this chemical is equal to zero and in the
279 prediction output file it is reported that the predicted compound
280 is outside the AD of the model. The lazar confidence index for its
281 prediction is 0.02.

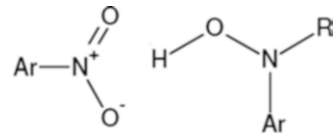


Fig. 4 Genotoxic structure alerts found by Toxtree in the molecular structure of 2-amino-5-nitrothiazole; SA_27: Nitro-aromatic is shown on the *left* side, while SA_28: primary aromatic amine, hydroxyl amine and its derived esters is shown on the *right*, where Ar stands for any aromatic/heteroaromatic ring and R stands for any atom/group

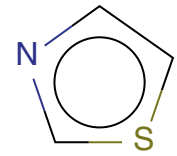


Fig. 5 Carcinogenicity structure alert found by the SARpy model for which the chemical is predicted as carcinogen

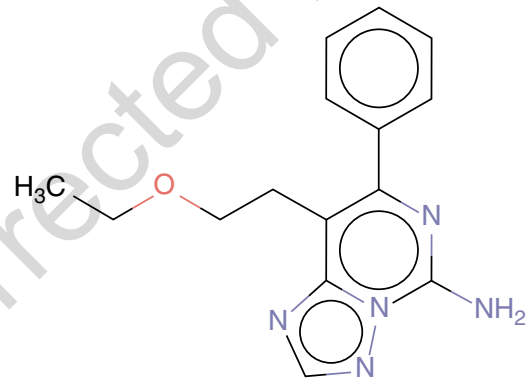


Fig. 6 Bemitradine chemical structure with CAS number: 88133-11-3 and SMILES: n2cnn3c(nc(c1ccccc1)c(c23)CCOCC)N

t2.1 **Table 2**

t2.2 **Cancer test summary reported in the CPDB for Bemitradine**

t2.3	Rat target sites		Mouse target sites		TD50 (mg/kg/day)	
t2.4	**Male**	**Female**	**Male**	**Female**	**Rat**	**Mouse**
t2.5	liv	liv mgl[a]	No test	No test	548 m	No test

t2.6 [a]liv liver, *mgl* mammary gland

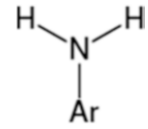


Fig. 7 QSA28_gen. Primary aromatic amine, hydroxyl amine, and its derived structure alert found by Toxtree in the molecular structure of Bemitradine

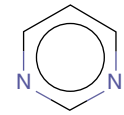


Fig. 8 Carcinogenicity structure alert found by the SARpy model for which the chemical is predicted as carcinogen

Toxtree (v. 2.6.0) prediction for this chemical was: "Negative for non-genotoxic carcinogenicity and positive for genotoxic carcinogenicity." The SA recognized by Toxtree in the molecular structure is "QSA28_gen. Primary aromatic amine, hydroxyl amine and its derived esters (with restrictions)."

However, there are two restrictions to this rule. In fact, if the following conditions are true then the compound is predicted as non-carcinogen:

- Chemicals with *ortho*-disubstitution, or with an ortho carboxylic acid substituent are excluded.
- Chemicals with a sulfonic acid group ($-SO_3H$) on the same ring of the amino group are excluded.

and in this case study, none of them are applied.

Overall, on the basis of the results of the different models and the low confidence value of lazar and the fact that it is out of AD of VEGA-CAESAR, of course one cannot exclude the possible carcinogenic effect. On the contrary, there are elements to support the toxic effect which cannot be ruled out by the presence of some results going in the opposite direction. Thus, the overall assessment should go for carcinogenicity, but with a higher uncertainty, compared to the results for the case study 1.

3.4 Case Study 4

Amobarbital (formerly known as amylobarbitone or sodium amytal) is a drug that is a barbiturate derivative (*see* Fig. 9). It has sedative-hypnotic properties. On the basis of CPDB it is classified as a non-carcinogen (*see* Table 3). Toxtree (v. 2.6.0), lazar, VEGA-CAESAR (v. 1.1.0), and the SARpy (v. 1.0) model predicted this molecular structure correctly (i.e. non-carcinogen) as confirmed by the experimental result. In addition, the VEGA-CAESAR prediction result has a reliability feature that for this compound: "the predicted compound is into the Applicability Domain of the model."

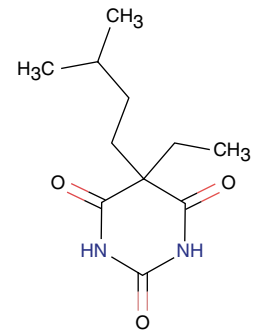


Fig. 9 Amobarbital with CAS number: 57-43-2 and SMILES: CCC1(CCC(C)C)C(=O)NC(=O)NC1=O

t3.1 **Table 3**
t3.2 **Cancer test summary reported in the CPDB for Amobarbital**

t3.3	Rat target sites		Mouse target sites		TD50 (mg/kg/day)	
t3.4	Male	Female	Male	Female	Rat	Mouse
t3.5	No positive	No test	No test	No test	No positive	No test

In fact, the model has the experimental value of this compound. The AD index of this chemical in the VEGA-CAESAR prediction is equal to 1 (*see* **Note 3**). The lazar reported this chemical as an already existing chemical inside its training set, so we consider it inside its AD. As it is mentioned above, Toxtree and SARpy do not have any AD index along with their prediction results.

As a conclusion, all the prediction results of the above-mentioned models indicate the non-carcinogenic effect of the compound, which are concordant with the experimental value.

4 Notes

1. The different sources of the data used within the different models should always be considered. The CAESAR model is closely related to the rat carcinogenicity, while other models tend to balance results from different studies. There may be differences between the carcinogenicity in animals and in humans [32].
2. It should be noted that the data available for building carcinogenicity models derive studies which identified in several cases effects on different organs (i.e. test for hepatocarcinogenicity, polmonarcarcinogenicity). Therefore, building organ-specific

carcinogenicity may be the best approach in order to obtain models with higher prediction performance. Nevertheless, the number of experimental results on organ-specific carcinogenicity is at the time limited making them inadequate for building a (Q)SAR model with high performance.

3. VEGA provides the experimental result of the target compound, if available. The experimental value prevails on the predicted one, and thus the AD index is 1. The predicted value of the target compound is also given in the summary page.

References

1. Arcos JC (1995) Chemical induction of cancer: modulation and combination effects. An inventory of the many factors which influence carcinogenesis. Springer Science & Business Media, New York
2. Woo Y, Lai D (2003) Mechanisms of action of chemical carcinogens, and their role in Structure-Activity Relationships (SAR) analysis and risk assessment. In: Benigni R (ed) Quantitative Structure-Activity Relationship (QSAR) models of mutagens and carcinogens. CRC, Boca Raton, FL, pp 41–80
3. Miller J, Miller E (1977) Origins of human cancer. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 605–627
4. Miller EC, Miller JA (1981) Searches for ultimate chemical carcinogens and their reactions with cellular macromolecules. Cancer 47(10):2327–2345
5. Ames BN (1979) Identifying environmental chemicals causing mutations and cancer. Science 204(4393):587–593
6. OECD Test No. 487: In Vitro Mammalian Cell Micronucleus Test. OECD Publishing
7. Benfenati E (2013) Theory, guidance and applications on QSAR and REACH. ORCHESTRA consortium
8. Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MT, Judson PN, Payne MP, Richard AM, Tichy M (1997) The development and validation of expert systems for predicting toxicity. Workshop (ECVAM Workshop 24). p 2
9. Benigni R, Bossa C, Tcheremenskaia O, Giuliani A (2010) Alternatives to the carcinogenicity bioassay: *in silico* methods, and the in vitro and in vivo mutagenicity assays. Expert Opin Drug Metab Toxicol 6(7):809–819. doi: 10.1517/17425255.2010.486400
10. Toropov AA, Toropova AP, Benfenati E (2009) Additive SMILES-based carcinogenicity models: probabilistic principles in the search for robust predictions. Int J Mol Sci 10(7): 3106–3127
11. Rositsa S, Mojca FG, Andrew W (2010) Review of QSAR models and software tools for predicting genotoxicity and carcinogenicity. JRC scientific and technical reports-EUR 24427 EN. Publications Office of the European Union, Luxembourg
12. Benfenati E, Benigni R, Demarini DM, Helma C, Kirkland D, Martin TM, Mazzatorta P, Ouédraogo-Arras G, Richard AM, Schilter B, Schoonen WG, Snyder RD, Yang C (2009) Predictive models for carcinogenicity and mutagenicity: frameworks, state-of-the-art, and perspectives. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 27(2):57–90. doi:10.1080/10590500902885593
13. Ferrari T, Gini G, Benfenati E (2009) Support vector machines in the prediction of mutagenicity of chemical compounds. Proc NAFIPS 2009, June 14–17, Cincinnati, USA. pp 1–6
14. Ferrari T, Gini G, Bakhtyari NG, Benfenati E (2011) Mining toxicity structural alerts from SMILES: a new way to derive structure activity relationships. Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium. pp120–127.doi:10.1109/cidm.2011.5949444
15. Toxtree v, Ideaconsult Ltd, Sofia BAahtsn
16. Woo Y-T, Lai DY, Argus MF, Arcos JC (1995) Development of structure-activity relationship rules for predicting carcinogenic potential of chemicals. Toxicol Lett 79(1):219–228
17. OECD (2010) OECD Toolbox. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm. Accessed 20 June 2015
18. Helma C (2006) Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Mol Divers 10(2):147–158
19. Klopman G, Rosenkranz HS (1994) Approaches to SAR in carcinogenesis and

mutagenesis. Prediction of carcinogenicity/mutagenicity using MULTI-CASE. Mutat Res 305(1):33–46

20. Enslein K, Gombar VK, Blake BW (1994) Use of SAR in computer-assisted prediction of carcinogenicity and mutagenicity of chemicals by the TOPKAT program. Mutat Res 305(1): 47–61
21. Smithing MP, Darvas F (1992) HazardExpert: an expert system for predicting chemical toxicity. ACS Symposium series American Chemical Society
22. Sanderson D, Earnshaw C (1991) Computer prediction of possible toxic action from chemical structure; the DEREK system. Hum Exp Toxicol 10(4):261–273
23. Ridings J, Barratt M, Cary R, Earnshaw C, Eggington C, Ellis M, Judson P, Langowski J, Marchant C, Payne M (1996) Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. Toxicology 106(1):267–279
24. Helma C (2005) *In silico* predictive toxicology: the state-of-the-art and strategies to predict human health effects. Curr Opin Drug Discov Dev 8(1):27–31
25. Helma C, Cramer T, Kramer S, De Raedt L (2004) Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. J Chem Inf Comput Sci 44(4):1402–1411. doi:10.1021/ci034254q
26. VEGA hwv-qe
27. Benigni R (2008) The Benigni/Bossa rulebase for mutagenicity and carcinogenicity—a module of Toxtree. JRC Scientific and Technical Reports, pp 1–78
28. ANTARES (2012) http://www.antares-life.eu/. Accessed 20 June 2015
29. Chemical carcinogens structures and experimental data (ISSCAN) (2008) ISS. http://www.epa.gov/ncct/dsstox/sdf_isscan_external.html. Accessed 20 June 2015
30. Kirkland D, Aardema M, Henderson L, Müller L (2005) Evaluation of the ability of a battery of three in vitro genotoxicity tests to discriminate rodent carcinogens and non-carcinogens: I. Sensitivity, specificity and relative predictivity. Mutat Res 584(1):1–256
31. Mayer J, Cheeseman MA, Twaroski ML (2008) Structure-activity relationship analysis tools: validation and applicability in predicting carcinogens. Regul Toxicol Pharmacol 50(1):50–58. doi:10.1016/j.yrtph.2007.09.005
32. EPA Environmental Protection Agency. http://www.epa.gov/nrmrl/std/qsar/qsar.html. Accessed 20 June 2015
33. IdeaConsult (2009) Toxtree user manual. IdeaConsult. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/Toxtree_user_manual.pdf. Accessed 20 June 2015
34. The Carcinogenic Potency Database (CPDB) B, CA (USA). http://potency.berkeley.edu

Chapter 7

VirtualToxLab: Exploring the Toxic Potential of Rejuvenating Substances Found in Traditional Medicines

Martin Smieško and Angelo Vedani

Abstract

Docking and quantifying the binding of small molecules to the 3D structure of a macromolecular bioregulator by computational techniques is a typical task in R&D aimed at the design and optimization of medically or otherwise active compounds. Much less known is the fact that these methods can be successfully applied for the purpose of toxicity prediction—for example, detecting a compound's potential binding to so-called "off-targets" already at the preclinical stage. In this chapter, we provide an overview of such a computational approach, discuss its strengths and weaknesses, and include a case study—focused on natural compounds present in traditional medicines.

Key words Protein-mediated toxicity, Molecular docking, Scoring, Toxicity prediction, Binding mode, Binding affinity, Pharmacokinetic properties

1 Introduction

Toxicology and computational chemistry are two disciplines whose synergistic combination has not been explored all too often in the past, but an ever growing importance has been witnessed. Their combination follows a concept established in rational drug design, where computational chemistry and molecular modeling are used for predicting the pharmacological activity of a small molecule—mechanistically triggered by its binding at the desired target. Analogously in toxicology, computational methods could be employed for identifying compounds leading to undesired effects as a result of their binding to relevant macromolecular targets other than the primary bioregulator—the so-called "off targets."

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_7, © Springer Science+Business Media New York 2016

2 General Concept

2.1 Pharmacokinetic Properties

Before envisioning the computational evaluation of a compound's ability to bind to a protein target, its availability at the site of action needs to be addressed. From the possible entry point into the human organism (e.g. transdermal, by ingestion or inhalation), the oral route has been studied in most detail [1, 2], particularly in pharmaceutical R&D, because it is the most convenient (comfortable) way of administration for the prospective patient to be treated. Knowledge gathered on the oral absorption and availability of small drug molecules is of equal importance for toxicology, because compounds associated with a harmful potential might easily reach the gastrointestinal tract (GIT) by ingestion, either intendedly (e.g. through food ingredients and additives, colorants, drugs) or unintendedly (as an undesired contaminant of any of the former).

Exploring the pharmacokinetic properties of a compound may provide hints on a compound's specificity. In drug-design studies, it has been observed that an increasing lipophilicity of a molecule (i.e. by adding lipophilic substituents to it) might assist in improving its binding affinity, but may thereby jeopardize its specificity and decrease the ligand efficiency. Therefore, extremely lipophilic compounds (featuring a large, positive $\log P$ value) would show a non-specific interaction pattern—i.e. possibly affecting multiple targets and accumulate in adipose tissues of the body where they could persist for a prolonged period of time and possibly causing chronic adverse effects. On the other hand, hydrophilic compounds are readily filtered in the kidneys, leading to a fast clearance from the organism and, consequently, lowering the chance of triggering adverse effects.

Obtaining the most common pharmacokinetic characteristics of a given compound is quite straightforward. According to the widely accepted "Lipinski's rule of five" [1], a compound would be likely absorbed from the GIT if its molecular weight is lower than 500, the number of hydrogen bond donors and acceptors is lower than 5 and 10, respectively, and the compound's water–octanol partition coefficient ($\log P$) is lower than 5. The values of the first three descriptors can be calculated by analyzing the compound's 2D structure, while for the $\log P$ value, many trained models exist [3–7], capable to estimate the actual value by interpolation. Lipinski's rules can be augmented with two additional rules (postulated by Veber et al.) [2] limiting the number of rotatable bonds to less than 10 and polar molecular surface area to 140 Å^2.

2.2 Toxicity and Ligand Binding to Off-Targets

Toxicity and adverse effects stem from a typically non-covalent interaction (for toxicity triggered by covalently bound, i.e. reactive chemical species, please refer to ref. 8) of a small molecule with a bioregulator (receptor, enzyme, ion channel, DNA). Such an

interaction can be quite unspecific, e.g. a highly lipophilic compound may be accommodated by any (at least partially) hydrophobic macromolecular cavity in order to "escape" from its interaction-wise unfavorable aqueous environment. Here, the compound's binding would mainly be driven by desolvation effects (releasing unfavorable solvent molecules from hydrophobic cavities within a protein is beneficial for the overall binding) and weak dispersion interactions lacking any strictly preferred spatial arrangement (surface-to-surface interaction). On the other hand, a specific interaction of a small molecule with the protein target, e.g. displaying a high degree of both shape and volume complementarity to one unique protein binding site (or an allosteric or enzyme active site) with a well-defined and a thermodynamically and kinetically stable binding mode, would in addition to hydrophobic contacts likely include also several directional interactions such as salt-bridges and hydrogen bonds. In both cases, the compound's binding to a protein may be considered as an interference with the finely tuned system of hormones, feed-back effectors, and endogenous compounds (e.g. displacing a hormone or natural substrate from the binding or active site, or transport protein, inhibiting or activating an ion channel) that would eventually perturb the physiological homeostasis within the organism and which would possibly manifest itself as adverse effects or toxicity. The impact of such effects in vivo cannot (yet) be computationally quantified with a desirable accuracy; however, the dose–response relationship would suggest that (at the given target) the more affine a compound, the more severe adverse effects or even toxicity are to be expected.

Exploring compound's potential for protein-mediated toxicity using computational methods relies on identifying a specific non-covalent binding mode of the evaluated molecule at the macromolecular target, a concept widely used for drug design and known as molecular docking, and employing a scoring function to estimate (quantify) the binding energy.

2.3 Molecular Docking for Identifying Off-Target Binding

Molecular docking is the most convenient alternative to experimental methods directly determining the compound's binding (e.g. in vitro assay, crystallography). Its main advantage is that it can be used also for analyzing hypothetical compounds, i.e. those not yet chemically synthesized, which allows for early screening and decision making, thus saving time and resources. The key prerequisite for application of molecular docking approach is availability of a 3D structure of the target macromolecule. This can be experimentally determined using any of the standard techniques (NMR, X-ray crystallography) or built computationally using structural information of related proteins or similar structural subunits in homology modeling. In any case, the 3D structure, especially in the vicinity of the binding site, must be as detailed as possible with well-resolved positions of atoms in amino acids,

cofactors, ligands as well as solvent (water) molecules, so that the spatial arrangement of crucial stabilizing interactions—particularly energetically prominent H-bond networks involving also water molecules—could be unambiguously determined. Ideally, several 3D structures of the target macromolecule are at hand, with different bound ligands, which can serve as templates for pre-orienting and pre-positioning of structures being docked and at the same time provide information on target's local (e.g. amino acid side-chains flexibility) and global (e.g. backbone, loop, or large-unit rearrangement) flexibility.

When aiming at the prediction of toxicity, molecular docking is quite challenging because of typically low similarity to be expected, in terms of size, shape, and chemical composition. In addition, no template structure (bound small molecule, similar to the one of interest) might be available. In the process of lead optimization, solving the crystal structure of a lead compound bound to the target protein is therefore of utmost interest. Based on that structure, novel derivatives, typically featuring only conservative structural modifications triggering small changes in the host structure (e.g. introducing H-bond donors/acceptors or lipophilic moieties to match the binding-pocket character better), are thought to be straightforward to obtain. This implies that the new ligand's conformation and its orientation within the binding site remains identical or, at least, similar. This fact allows to largely reduce the degrees of freedom to be explored in docking. It also simplifies the pose generation, so that even a rigid-docking protocol (keeping the macromolecule fixed) can produce reasonable results. However, when docking a compound dissimilar to any of the templates, as much structural information as possible, e.g. protein and ligand conformation, thermal displacement (*B*) factors, binding site shape and volume, pharmacophore assumptions, structural and displaceable solvent molecules, must be extracted from all known ligand–target structures and productively combined in order to rationalize the generation of binding modes, simultaneously decreasing the computational complexity and speeding up the docking. Random-searching algorithms (i.e. randomly modifying the ligand's and protein's conformation along with rotation and translation of the ligand) have theoretically a potential to identify all feasible binding modes, but due to complexity of the mathematical solution would need an enormous amount of computational time for an exhaustive sampling and therefore find only a limited use. Even thoroughly rationalized docking techniques require a rather computationally expensive geometry refinement to produce poses with reasonable interaction patterns and therefore molecular docking for predicting the off-target binding cannot be generally classified as a high-throughput method.

Binding modes generated by molecular docking allow for a mechanistic interpretation of interaction at atomic level and are of

great value for further evaluation. For example, the human androgen receptor can be viewed as an anti-target as any interference with it could trigger endocrine disruption, but a compound binding to the androgen receptor identified by the docking procedure having a novel non-steroidal scaffold could serve as a basis for development of novel anabolic (agonist) or anticancer (antagonist) drugs. In case the molecule docked to the off-target would be for example a promising drug candidate, with its binding mode in hand one could modify its structure at a site that would (e.g. sterically) hinder binding to an off-target and that would be still tolerated at the desired (original) target. Such a modification might save the compound from being discarded from the development pipeline because of risk of adverse effects and even improve its selectivity and safety. In case the tested molecule would be a natural compound binding to a pharmacologically relevant target, the binding mode could indicate sites where such a structure could be simplified (e.g. removing of functional groups not involved in a favorable interaction with the target) or extended (e.g. adding a lipophilic group filling an otherwise empty part of the binding pocket) by methods of the synthetic chemistry in order to obtain a novel ligand.

2.4 Scoring Poses

While the main task of molecular docking is to identify binding modes with the most favorable ligand–target interaction energy, the scoring procedure is used to put obtained binding modes into context of a complete thermodynamic cycle, whose equilibrium is defined by the difference of free energy of the ligand and target in the unbound state and after they form a non-covalent complex. Therefore a typical scoring function, besides including enthalpic terms (electrostatic, van der Waals, H-bonding, and metal interactions), should account also for entropic terms, e.g. desolvation costs of both ligand as well as binding site at the target macromolecule, contributions stemming from solvent displacement, and penalties associated with the loss of degrees of freedom of the bound ligand and interacting amino acids in the target molecule. Entropic contributions may be calculated with a satisfactory accuracy without knowing more about dynamic properties of the interacting partners, therefore such terms are frequently approximated by summing up averaged contributions, e.g. averaged gain per displaced solvent molecules or immobilized rotatable bond, or by using empirical values [9, 10].

A scoring function might be trained in order to reproduce as closely as possible experimentally determined binding affinities of a set of compounds. However, training automatically reduces the applicability domain of a scoring function to a set of compounds similar to those in the training set. As mentioned above, the off-target binding is usually examined for compounds substantially different from those used for training (e.g. experimental binding

affinities of a set of congeneric compounds from a classical medicinal chemistry lead optimization were used to train the scoring function, but a structurally dissimilar agrochemical is being evaluated), prediction based on a trained scoring function would therefore be extrapolated and very uncertain.

Despite rapid development in the field and growing complexity, there is (up-to-date) no scoring function available that would produce satisfactory results for a whole range of biologically relevant targets. Therefore, further analyses reaching beyond simple scoring, e.g. inspection of the dynamic stability of binding modes using molecular dynamics (MD) simulations or the consensus scoring employing conceptually different techniques, are highly recommended.

2.5 VirtualToxLab

The *VirtualToxLab* is an in silico technology for estimating the toxic potential—endocrine and metabolic disruption, some aspects of carcinogenicity and cardiotoxicity—of drugs, chemicals, and natural products [11]. The technology is based on an automated protocol that simulates and quantifies the binding of small molecules toward a series of currently 16 proteins, known or suspected to trigger adverse effects: ten nuclear receptors (androgen, estrogen α, estrogen β, glucocorticoid, liver X, mineralocorticoid, peroxisome proliferator-activated receptor γ, progesterone, thyroid α, thyroid β), four members of the cytochrome P450 enzyme family (1A2, 2C9, 2D6, 3A4), a cytosolic transcription factor (aryl hydrocarbon receptor) and a potassium ion channel (hERG). The toxic potential of a compound—its ability to trigger adverse effects—is derived from its computed binding affinities toward these very proteins (reference). The computationally demanding simulations are executed in client–server mode on a Linux cluster of the University of Basel. The graphical-user interface supports all computer platforms, allows building and uploading molecular structures, inspecting and downloading the results and, most important, rationalizing any prediction at the atomic level by interactively analyzing the binding mode of a compound with its target protein(s) in real-time 3D/4D. Access to the *VirtualToxLab* is available free of charge for universities, governmental agencies, regulatory bodies, and non-profit organizations.

3 Estimating the Toxic Potential of Compounds from Traditional Medicines

3.1 Rejuvenation Compounds and Traditional Medicine

We performed a study exploring compounds occurring in rejuvenating or anti-aging preparations present in various traditional medicines. The latter enjoy a large popularity especially on the Asian and African continent and whether explicable or not, are used in the maintenance of health as well as in the prevention, diagnosis, improvement, or treatment of physical and mental

illnesses. Such herbal and fungal preparations contain highly species-specific secondary metabolites—compounds which might help in fighting the various symptoms of aging, such as overall weakness and decreased metabolism, reduced immunity, cognition, fertility, or muscle strength, decline in memory functions or loss of skin elasticity. Some of these symptoms can be associated with an age-related natural ligand (hormone) depletion followed by insufficient activation of associated bioregulators. For example, a low testosterone level would prevent from the androgen receptor activation and result in decreased transcription of AR-regulated genes for muscle growth. The VirtualToxLab with its target portfolio covering several nuclear receptors seems to be the right tool for screening of potential rejuvenating compounds.

The use of preparations (or single compounds isolated therefrom) recommended by traditional medicines is sometimes documented by medicinal studies—for example, antioxidants (vitamins, flavonoids) have been shown to scavenge free radical thus preventing DNA and protein from being damaged by such reactive chemical species [12], but frequently little or no evidence exists, which poses potential risks (side effects, toxicity) of "blind" usage of not properly explored and standardized preparations. On the other hand, a substantial number of modern drugs has been inspired by natural (and traditional) medicines, therefore screening such compounds by modern techniques (including in silico methods) may lead to beneficial discoveries and perhaps new drugs.

From the safety point of view, all chemical entities including natural compounds (or products of plant or animal origin containing secondary metabolites), which might occur within the human gastrointestinal tract (intended or unintended, e.g. trough food contaminants with agricultural origin) should be characterized and analyzed to the extent that we apply for pharmaceuticals.

3.2 Compound Identification and Modeling

Scientific (Pubmed, ScienceDirect) as well as general purpose (Google) electronic search engines were used along with keywords: "rejuvenat*", "anti-ag(e)ing", "traditional", "medicine" to retrieve information about biological organisms and their secondary metabolites that could be associated with supposed or described biological effects. In matching publications from peer-reviewed journals, names and structure formulas of 35 unambiguously characterized secondary metabolites from seven plant and three mushroom species were identified (Table 1). Compounds with already known beneficial properties (e.g. flavonoid antioxidants, vitamins), well-researched (e.g. cardioglycosides), or acting at a different target organism (e.g. anti-infectives) were excluded from our analysis. If available, the 3D structures of the underlying compounds were retrieved from the Cambridge Structure Database (CSD) [13]. Using small-molecule crystal structure geometries as input structures when dealing with natural compounds featuring extremely

t1.1 **Table 1**
t1.2 **Summary of pharmacokinetic parameters for analyzed compounds**

t1.3 t1.4 Organism	Compound	Rule of five violations	MW	Log $P_{o/w}$	PSA (Å^2)	Rot. bonds
t1.5 t1.6 *Withania somnifera*	Anaferine	0	224	1.3	51	4
	Anahygrine	0	224	1.4	44	4
	Cuscohygrine	0	224	1.5	34	4
	Isopelletierine	0	141	0.7	41	2
	Withaferin A	0	471	2.6	112	3
	Withanone	0	471	3.0	104	3
t1.7	14β-Hydroxywithanone	0	487	2.0	116	2
t1.8	Withadienolide	0	487	1.9	126	2
t1.9	Withanolide A	0	471	2.6	101	2
t1.10	Withasomnine	0	184	2.4	19	1
t1.11 t1.12 *Ginkgo biloba*	Ginkgolide A	0	408	1.2	143	1
	Ginkgolide B	0	424	0.5	169	1
	Ginkgolide C	1	440	0.2	186	1
	Ginkgolide J	0	424	0.2	170	1
t1.13	Ginkgolide P	0	424	−0.3	159	2
t1.14	Bilobalide	0	326	0.2	142	1
t1.15 *Pueraria mirifica*	Miroestrol	0	358	0.6	116	0
	Deoxymiroestrol	0	342	1.7	96	0
t1.16	Isomiroestrol	0	358	1.2	118	0
t1.17 *Panax ginseng*	Panaxadiol	1	461	5.5	41	1
	Falcarinol	1	244	5.8	23	11
	Panaxicol	0	278	3.6	69	12
t1.18	Panaxatriol	1	477	4.6	58	1
t1.19	Protopanaxadiol	1	461	5.4	56	4
t1.20 *Centenella asiatica*	Asiatic acid	0	489	4.8	104	2
t1.21 t1.22 *Rosmarinus communis*	Carnosic acid	0	332	4.4	74	2
t1.23 t1.24 *Hypericum perforatum*	Hyperforin	2	537	6.3	68	11

(continued)

Table 1
(continued)

Organism	Compound	Rule of five violations	MW	Log $P_{o/w}$	PSA (Å²)	Rot. bonds
Ganoderma lucidum	Ganoderol A	1	439	7.6	46	5
	Ganoderol B	1	441	7.4	40	5
	(*R*)-Ganodone	0	328	3.0	111	4
	(*S*)-Ganodone	0	328	3.0	110	4
	Lucidone	0	403	2.7	106	1
	Ganoderenic acid A	1	515	3.3	149	5
Tremella fuciformis	Oosporein	0	306	−0.2	186	1
Phellinus linteus	Hispidin	0	246	1.1	105	2

complex ring systems (e.g. multiply fused and/or spiro) would seem to be appropriate as this facilitates identifying the correct ring puckering as well as correct assignment of asymmetric centers in the molecule.

The calculation of descriptors related to pharmacokinetics was performed using the Schrodinger's QikProp program (rule-of-five violations, molecular weight [MW], polar surface area [PSA]) [14] and the VCC Lab AlogPs algorithm (Log $P_{o/w}$) [7]. Finally, all structures were submitted to the VirtualToxLab for an automated simulation of the binding mode(s) and estimation of the associated affinities toward all 16 targets (cf. above). For selected ligand–target complexes, molecular dynamics simulations using the Desmond software [15] were performed to examine the dynamic stability of intermolecular interactions.

4 Results and Interpretation

4.1 Pharmacokinetic Properties

The values for the pharmacokinetic descriptors are summarized in Table 1 with favorable properties highlighted in green, potentially problematic in orange and unfavorable ones in red. With a few exceptions (e.g. panaxicol, falcarinol, and hyperforin), the studied compounds are quite rigid, lipophilic, and of low-molecular weight, thus fulfilling most of criteria defined by the Lipinski's rule-of-five. This suggests that they could be absorbed from the gastrointestinal tract after oral intake and, therefore, would be available in the systemic circulation. As a consequence of the very low PSA (<90 Å²), some of the compounds (e.g. withasomnine, carnosic acid) could even cross the blood–brain barrier and interact

with bioregulators in the central nervous system. Secondary metabolites from *Ginkgo biloba*, despite the low number of rotatable bonds, have a rather low lipophilicity (Log $P \sim 0$) and a large PSA (just at, or above the limit of 140 Å^2) rendering them less feasible for passive permeation and therefore less orally available. On the other hand, some compounds, e.g. panaxadiol, ganoderol A and B—due to their pronounced lipophilicity—might be quite insoluble in water and, therefore, orally available in very limited amounts, but at repeated exposure, could accumulate in the adipose tissues, where they could persist over longer periods of time.

In general, with the exception of hyperforin (which differs substantially from typical orally available molecules in molecular weight, flexibility, and lipophilicity), all studied compounds have a good chance of being absorbed after oral intake, e.g. as an extract in tonic or as a part of food. The Lipinski's rule-of-five is by no means exclusive; it solely defines descriptor ranges where there is an increased likelihood for a compound of being orally available. Therefore, a slight deviation in one or two of Lipinski's or Veber's descriptors from recommended values observed for a few of studied compounds does not imply that, after all, they could not be orally available.

4.2 VirtualToxLab Binding Profiles

Binding-mode hypotheses and toxic-potential values obtained by the automatic docking and scoring protocol as implemented in the VirtualToxLab are summarized in Table 2. The color intensity correlates with the predicted affinity: dark gray cells indicate hits, i.e. computationally identified complementarity of the compound with a particular binding pocket (having at least one feasible binding pose) and favorable thermodynamics of transfer from aqueous environment to the binding site. For a better understanding of the following paragraphs, selected compounds discussed in detail are depicted in Fig. 1.

Compounds with low molecular weight (e.g. anaferin, anahygrine, cuscohygrine, isopelletierine withasomnine, hispidin, and oosporein) would seem to be too small for effectively occupying the binding site of any of the screened targets. In the VirtualToxLab, these compounds do not display any significant binding affinity and, consequently, their computed toxic potential is low. No favorable binding mode could be computationally identified for the topologically complex and pronouncedly hydrophobic hyperforin. The rigid pharmacophore—the spatial arrangement of functional groups attached to complex polycyclic scaffolds—of all ginkgolides, bilobalide, Asiatic, and carnosic acid is not complementary to any binding site of the targets currently implemented in the VirtualToxLab—even though explicitly allowing for ligand flexibility and local induced-fit in our simulations. No favorable interaction with any of the 16 targets could neither be identified for (*R*)-ganodone, nor for (*S*)-ganodone. Thus, for all compounds mentioned above, no effect on the symptoms of aging could be

t2.1 **Table 2**

t2.2 Color-coded binding profiles and toxic potential values for studied compounds from the VirtualToxLab

Molecule	Toxic Potential	AR	AhR	CYP1A2	CYP2C9	CYP2D6	CYP3A4	ERα	ERβ	GR	ERG	LXR	MR	PPARγ	PR	TRα	TRβ
Anaferine	0.454																
Anahygrine	0.369																
Auscohygrine	0.318																
Isopelletierine	0.270																
Withaferin A	0.492																
Withanone	0.520																
14-β1Hydroxywithanone	0.392																
Withadienolide	0.468																
Withanolide A	0.612																
Withasomnine	0.309																
Ginkgolide A	0.359																
Ginkgolide B	0.354																
Ginkgolide C	0.419																
Ginkgolide J	0.309																
Ginkgolide P	0.432																
Bilobalide	0.282																
Miroestrol	0.556																
Deoxymiroestrol	0.560																
Isomiroestrol	0.500																
Panaxadiol	0.547																
Falcarinol	0.551																
Panaxicol	0.573																
Panaxatriol	0.532																
Protopanaxadiol	0.571																
Asiatic acid	0.259																
Carnosic acid	0.361																
Hyperforin	0.308																
Ganoderol A	0.543																
Ganoderol B	0.623																
(R)-Ganodone	0.252																
(S)-Ganodone	0.302																
Lucidone	0.390																
Ganoderenic acid A	0.479																
Oosporein	0.033																
Hispidin	0.296																

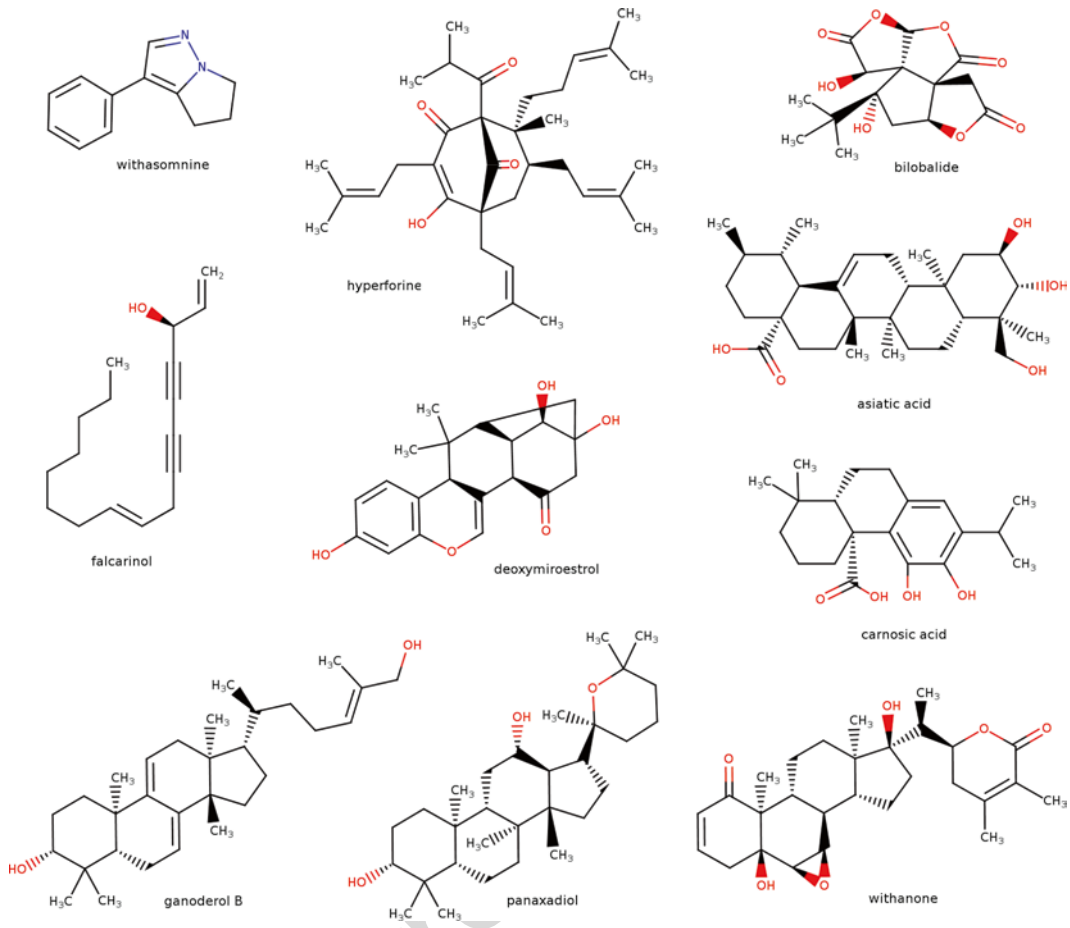


Fig. 1 Structural formulas of selected representative compounds

deducted based on the results from the VirtualToxLab. This, however, does not exclude other modes of action, i.e. effects triggered through binding to targets other than nuclear receptors, enzymes of the cytochrome P450 family, and the hERG potassium channel.

Several rings as well as H-bond donor and acceptor functionalities of the essentially rigid (according to Veber "completely rigid" as terminal methyl and hydroxyl groups are not counted as rotatable in that very concept) miroestrol derivatives closely resemble the pharmacophore of the naturally occurring female hormone 17β-estradiol. This results in an increased affinity toward nuclear receptors having steroidal structures as natural ligands, especially toward α and β estrogen receptors (Table 2). Upon binding to the estrogen receptor β (ERβ; Fig. 2), some of the polar atoms of miroestrol derivatives (carbonyl, ring oxygen atom, hydroxyl group) are not involved in any favorable interaction and offer possibilities for modification, while hydroxyl groups corresponding to ones at polar ends of the estradiol should be preserved, if binding

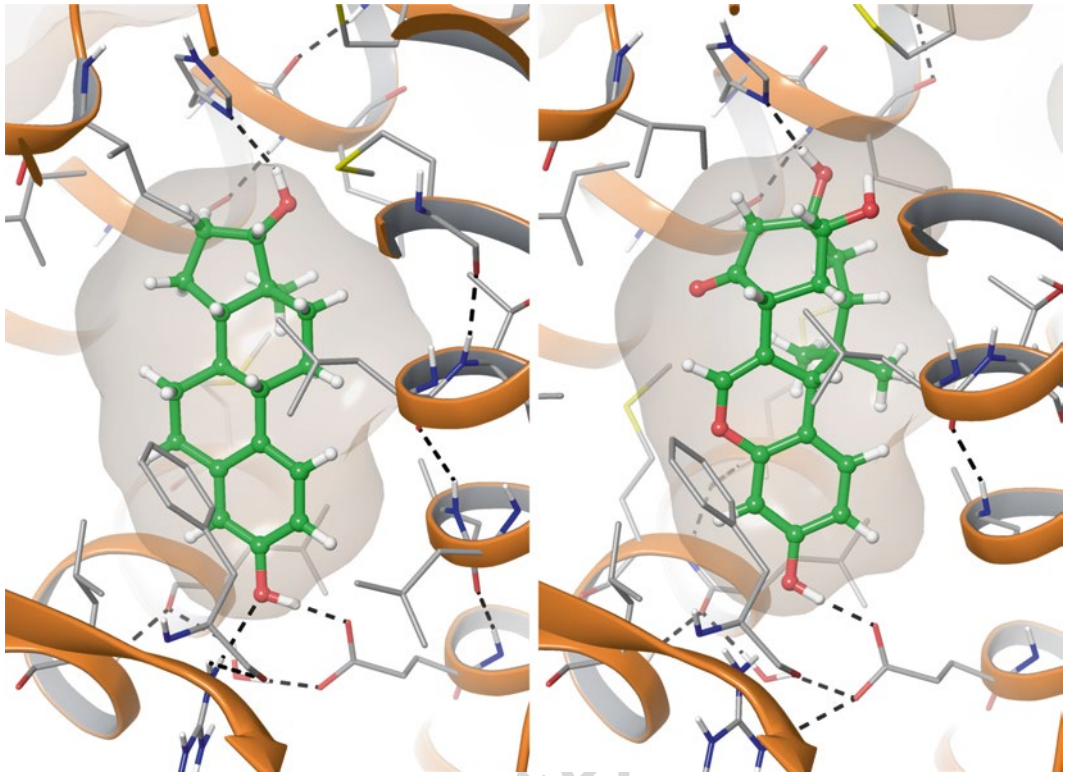


Fig. 2 17β-Estradiol (*left*, PDB entry 2J7X) and deoxymiroestrol (*right*, docked pose) bound to the estrogen receptor β

to ERβ is desired. A short molecular-dynamics simulation using the ligand–protein complex from the VirtualToxLab as the starting structure confirmed that these hydroxyl groups form stable H-bonds to the receptor (Fig. 3a). The hydroxyl group attached to the aromatic ring (corresponding to position 3 in ring A of estradiol) forms a direct H-bond with Glu305 (present during 99 % of the entire simulation time) and a water-mediated H-bond with either Arg346 (55 %) or Leu339 (15 %). The hydroxyl group mimicking the one at the 17β-position in the ring D of estradiol donates an H-bond to His475 (45 %) or Gly472 (36 %). As all three miroestrol derivatives are of comparable shape and size with estradiol, an agonistic effect is to be expected, which would seem to support the idea of administering a preparation from *Pueraria mirifica* containing miroestrols as estradiol mimicking molecules for relieving from symptoms associated with low estrogen levels in aging women. Obviously, instead of a rejuvenation, in men such compounds would cause an undesired feminization.

The steroidal scaffold of compounds from *Withania somnifera* (withanolides and similar), *Panax ginseng* (panaxadiol, panaxatriol, protopanaxadiol), and *Ganoderma lucidum* (ganoderol A and B, lucidone, ganoderenic acid A) suggests that such compounds may

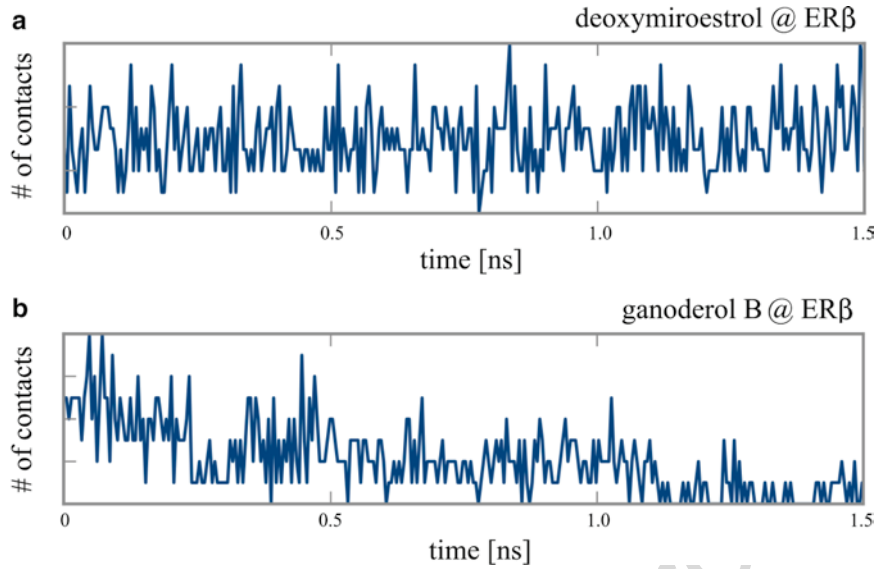


Fig. 3 Stability of protein–ligand interactions in MD simulations of (**a**) deoxymiroestrol and (**b**) ganoderol B at the estrogen receptor β (x-axis: simulation time, y-axis: number of protein-ligand contacts/interactions)

bind to nuclear receptors. However, most of them differ from typical natural steroidal agonists, because they feature a bulky and at least partially rigid substituent (6-membered lactone or pyran ring) at the position 17 of the cyclopentanoperhydrophenanthrene scaffold, which requires certain space for a proper accommodation and therefore could trigger induced-fit changes in the binding site leading to destabilization of the receptor structure—in this context, only partial agonistic or even antagonistic effects could be expected. In addition, the scaffold of these compounds is decorated with polar hydroxyl groups at positions different from those in natural ligands, which cannot form H-bonds with the same thermodynamic efficiency like those of latter do. Molecular-dynamics simulations of ligand–protein complexes using the highest-ranked binding pose from the VirtualToxLab as input structures showed that such hydrogen bonds have either a transient character (frequent interchange) or completely disappear early in the course of simulation (Fig. 3b), which greatly reduces their contribution toward the binding free energy (enthalpic terms). Such an unstable intermolecular interaction has been observed also for extremely flexible compounds like falcarinol and panaxicol. The computed data suggest that any potential beneficial effects of this subgroup of compounds in the context of rejuvenation might stem from weaker and not too specific binding, possibly at multiple nuclear receptors. The interactive analysis of the 4D ensemble of predicted binding modes used for scoring usually shows multiple poses with significant contributions toward the binding free energy, but with largely different orientation within the binding site accompanied by changes of side-chain conformations (local induced-fit; Fig. 4).

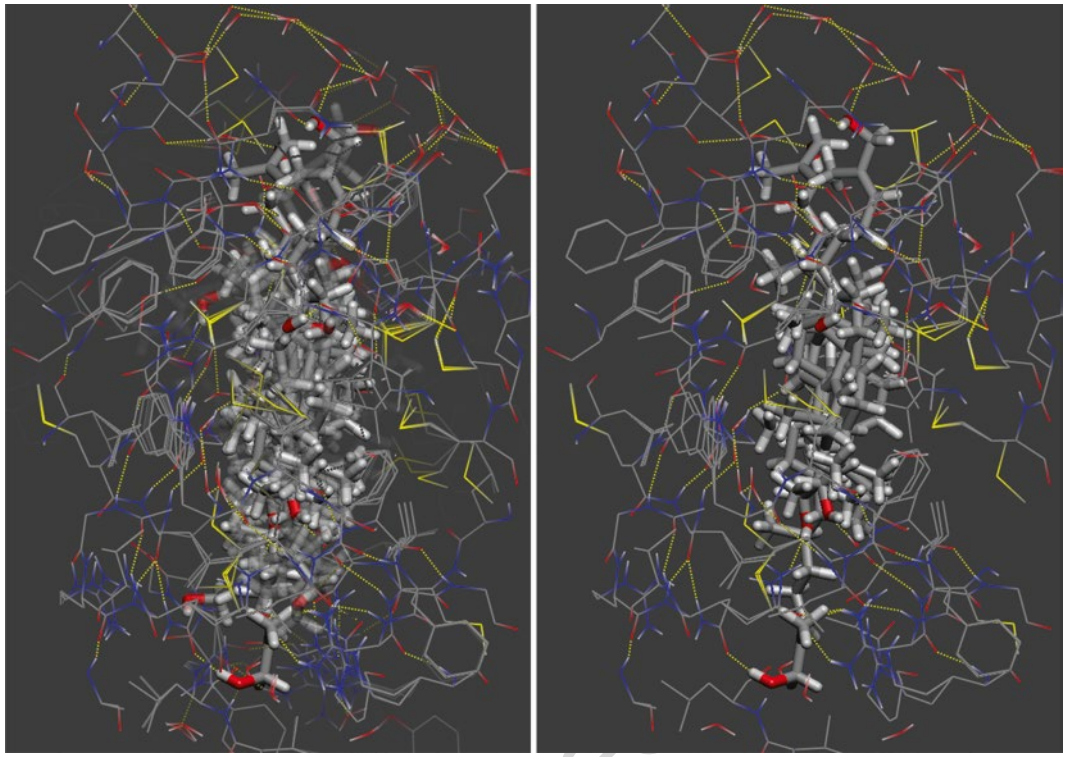


Fig. 4 Multiple binding modes (4D view with Boltzmann-scaled color intensities) observed for ganoderol B bound to the glucocorticoid receptor. *Left*: all 12 poses; *right*: top three poses contributing 58 %, 23 %, and 13 % to the total binding energy, respectively

Some compounds from the *Panax* species showed binding also to cytochromes (e.g. protopanaxadiol at CYP450 2D6), which might cause an enzyme inhibition and thus interfere with metabolic functions in liver cells.

At this place, we would like to point out that any outcome of an in silico screening in predictive toxicology, but especially the negative one, has to be interpreted with caution, as the applied methods and approximated model systems simply cannot provide a completely realistic answer to our scientific problem (e.g. due to a non-exhaustive conformational sampling, limited simulation time, and incomplete support for global conformational changes of target molecules, inaccuracies, or complete absence of force-field parameters).

5 Concluding Notes

In silico analyses of compounds, which are associated with rejuvenating effects based on traditional medicines, showed that a large majority of them fulfill the criteria for oral availability. This means

that after ingestion they would be able to reach the systemic circulation, while some of them could even cross the blood–brain barrier and exert their effects in the central nervous system.

Computed data—in the form of binding modes at the atomic level featuring favorable H-bonding as well as hydrophobic interaction patterns with associated binding free energies obtained by state-of-the-art methodologies—seem to provide some support for potential natural hormone-mimicking effects, particularly the group of miroestrol derivatives and to a smaller extent also for some steroid-like secondary metabolites occurring in the species *Withania*, *Panax*, and *Ganoderma*, but also uncover the risk associated with compound's inappropriate use, lack of selectivity, and possible interference with cytochromes.

The dynamic stability of interactions between ligand and target obtained by the automated docking was explored by means of MD simulations: while a few compounds exhibit stable and well-defined binding modes to some nuclear receptors further confirming their predicted binding potential, the others form only labile interactions suggesting that the scoring function might have overestimated their binding potential.

Positive findings regarding potential biological effects described in this study highlight the importance of a proper toxicological characterization of natural compounds occurring in preparations recommended by the traditional medicine, as their uncontrolled or excessive application or unintended use might affect human health negatively.

References

1. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 23:3–25
2. Veber DF, Johnson SR, Cheng H-Y, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623
3. Carrupt P-A, Testa B, Gaillard P (1997) Computational approaches to lipophilicity: methods and applications. Rev Comp Chem 11:241–315
4. Wang R, Fu Y, Lai L (1997) A new atom-additive method for calculating partition coefficients. J Chem Inf Comput Sci 37:615–621
5. Kellogg GE, Abraham DJ (2000) Hydrophobicity: is Log$P_{o/w}$ more than the sum of its parts? Eur J Med Chem 35:651–661
6. Tetko IV, Tanchuk VY, Villa AE (2001) Prediction of n-octanol/water partition coefficients from PHYSPROP database using artificial neural networks and E-state indices. J Chem Inf Comput Sci 41:1407–1421
7. Tetko IV, Tanchuk VY (2002) Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. J Chem Inf Comput Sci 42:1136–1145, http://www.vcclab.org/lab/alogps/
8. Kalgutkar AS (2011) Handling reactive metabolite positives in drug discovery: what has retrospective structure–toxicity analyses taught us? Chem Biol Interact 192:46–55
9. Spyrakis F, Kellogg GE, Amadasi A, Cozzini P (2007) Scoring functions for virtual screening. Front Drug Des Discov 3:317–379, Bentham Science Publishers Ltd., eISBN: 978-1-60805-201-1
10. Polgár T, Keserü GM (2007) Structure-based virtual screening. Front Drug Des Discov

3:477–502, Bentham Science Publishers Ltd., eISBN: 978-1-60805-201-1

11. Vedani A, Dobler M, Hu Z, Smieško M (2015) OpenVirtualToxLab—a platform for generating and exchanging in silico toxicity data. Toxicol Lett 232:519–532, available at http://www.virtualtoxlab.org/
12. Kumar S, Pandey K (2013) Chemistry and biological activities of flavonoids: an overview. Sci World J 2013:162750. doi:10.1155/2013/162750
13. Allen FH (2002) The Cambridge Structural Database: a quarter of a million crystal structures and rising. Acta Crystallogr B 58:380–388
14. QikProp, version 4.0 (2014) Schrödinger, LLC, New York
15. Desmond Molecular Dynamics System, version 3.8 (2014) D. E. Shaw Research, New York. Maestro-Desmond Interoperability Tools, version 3.8 (2014) Schrödinger, New York

Chapter 8

In Silico Model for Developmental Toxicity: How to Use QSAR Models and Interpret Their Results

Marco Marzo, Alessandra Roncaglioni, Sunil Kulkarni, Tara S. Barton-Maclaren, and Emilio Benfenati

Abstract

Modeling developmental toxicity has been a challenge for (Q)SAR model developers due to the complexity of the endpoint. Recently, some new in silico methods have been developed introducing the possibility to evaluate the integration of existing methods by taking advantage of various modeling perspectives. It is important that the model user is aware of the underlying basis of the different models in general, as well as the considerations and assumptions relative to the specific predictions that are obtained from these different models for the same chemical. The evaluation on the predictions needs to be done on a case-by-case basis, checking the analogs (possibly using structural, physicochemical, and toxicological information); for this purpose, the assessment of the applicability domain of the models provides further confidence in the model prediction. In this chapter, we present some examples illustrating an approach to combine human-based rules and statistical methods to support the prediction of developmental toxicity; we also discuss assumptions and uncertainties of the methodology.

Key words Developmental toxicity, OECD, QSAR, Predictive reliability, Similarity

1 Introduction

The assessment of information on developmental and reproductive toxicity (DART) (*see* **Note 1**) represents an important analysis for understanding the toxicological characteristics of chemicals and their effects during pregnancy, as well as on male and female fertility [1]. For instance, in the field of drug discovery it is important to discriminate drugs that are safe for mother and child in gestation or conversely, that are safe for mother but toxic for child in gestation. In addition, in Europe the characterization of DART is a mandatory requirement for all the different tonnage bands defined by the REACH regulation [2]. For regulatory agencies, the evaluation of a chemicals potential to induce DART is an important consideration when carrying out human health risk assessments.

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_7,

The various effects related to the DART potential of chemicals are apically defined by different in vivo tests that follow the OECD Test Guidelines (TG) [3] as outlined below and described in Table 1. These include:

OECD TG 414 [4]: Prenatal Development Toxicity Study.

OECD TG 416 [5]: Two-Generation Reproduction Toxicity.

OECD TG 421 [6]: Reproduction/Developmental Toxicity Screening Test.

OECD TG 422 [7]: Combined Repeated Dose Toxicity Study with the Reproduction/Development Toxicity Screening Test.

OECD TG 426 [8]: Developmental Neurotoxicity Study.

OECD TG 443 [9]: Extended One-Generation Reproductive Toxicity Study.

One of the in vivo protocols most commonly used to test developmental toxicity is OECD TG 414 (Prenatal Development Toxicity Study) [4]. This guideline test is conducted using female rats or rabbits. Route of exposure may vary with the chemical and may be modified to incorporate the relevant human route of predominant exposure, however the substance is usually administered orally. Exposure to the test substance starts at the beginning of implantation and finishes at either the end of organogenesis or the end of the period of gestation. At completion of the selected treatment period, dams are sacrificed and fetuses are weighed, sexed, and examined in detail for external, visceral, and skeletal alterations. OECD TG 421 [6] is a screening test guideline designed to provide initial toxicological information on reproductive and developmental effects such as gonadal function, mating behavior, conception, and development of the conceived and parturition. Female rats are administered the chemicals from 2 weeks before mating, through the pregnancy until 4 days after delivery; males are treated at least 2 weeks before mating, throughout the mating period and until approximately 2 weeks after mating. The same method is used within OECD TG 422 [7], but it is a repeated dose toxicity study. The main differences between the two guidelines are that TG 421 is a reduced one-generation reproduction study, and TG 422 is a combination of a 28-day toxicity study and a reduced one-generation reproduction study. OECD TG 416 [5] is the most comprehensive evaluation of the effects of chemicals on the male and female reproductive systems and on offspring development. It is a two-generation study in rats and consists of an exposure to chemicals for two generations until the third week of age of the second generation (F2). During this experiment a large number of endpoints are evaluated including: reproductive performance and fertility, growth and survival of offspring, achievement of developmental landmarks, potential endocrine-mediated effects, and developmental neuro- and immuno-toxicity. OECD TG 443

t1.1 **Table 1**
t1.2 **Overview of the different test guidelines for DART**

t1.3 t1.4 Test guideline	Design	Endpoints	Advantage (+)/limitation (–)
t1.5 t1.6 t1.7 t1.8 t1.9 t1.10 t1.11 t1.12 OECD TG 414: Prenatal Development Toxicity Study	At least from implantation to 1 or 2 days before expected birth 3 dose levels plus control $N=20$ pregnant females	An implantation, resorptions, fetal growth, morphological variations, and malformations.	+ Malformations are assessed in all fetuses. – The dosing period includes only the prenatal period. – The effects assessment includes only effects in fetus.
t1.13 t1.14 t1.15 t1.16 t1.17 t1.18 t1.19 t1.20 t1.21 t1.22 t1.23 t1.24 t1.25 t1.26 OECD TG 416: Two-Generation Reproduction Toxicity	Exposure before mating for at least one spermatogenic cycle until of second generation. Three dose levels plus control $N=20$ pregnant females	Fertility, estrus cyclicity and sperm quality, growth, developmental and viability, anogenital distance if triggered, sexual maturation, histopathology organs, brain, and target organs.	+ Exposure covers all sensitive periods. + Effect assessment in F1 and F2. + Includes assessment of semen quality and estrus cyclist. – Anogenital distance only assessed in F2 if triggered. – Malformations of reproductive organs only investigated in 1 per sex litter.
t1.27 t1.28 t1.29 t1.30 t1.31 t1.32 t1.33 t1.34 t1.35 OECD TG 421 and 422: Reproduction/ Developmental Toxicity Screening Test and Combined Repeated Dose Toxicity Study with the Reproduction/ Development Toxicity Screening Test	From 2 weeks prior to mating until at least day 4 postnatally. Three dose levels plus control. $N=8$–10 pregnant females	Fertility, pregnancy length and birth, fetal and pup growth, and survival until day 3.	+ Short-term test. – Limited exposure period. – Limited number of endpoints. – Limited sensitivity due to number of animals.
t1.36 t1.37 t1.38 t1.39 t1.40 t1.41 t1.42 t1.43 t1.44 OECD TG 426: Developmental Neurotoxicity Study	At least from implantation throughout lactation. Three dose levels plus control. $N=20$ recommended, less than 16 not appropriate	Birth and pregnancy length, growth, developmental and viability, physical and functional maturation. Behavioral changes, brain weight, and neuropathology.	+ Exposure covers most of the sensitive periods. – No exposure before mating and from weaning to sexual maturation. – Mating and nursing behavior is not assessed.

(continued)

Table 1
(Continued)

Test guideline	Design	Endpoints	Advantage (+)/limitation (–)
OECD TG 443: Extended One-Generation Reproductive Toxicity Study	Exposure from 2 weeks before mating to 6 weeks post-mating. If test continues on F2, the same treatment it will do at the F1 Three cohort and three dose levels plus control $N = 20$ pregnant female	Fertility, dystocia, gestation length, fetal survival, viability, post-implantation loss, litter size and weight, sex ratio, litter.	+ Systemic evaluation of repro and developmental toxicity. + Sensitive to endocrine influence.

[9] is a variation of OECD TG 416 [5]. TG 443 does not take into consideration the F2 generation, but does a complete analysis of the F1 generation to assess development of immune system function, developmental neurotoxicity, as well as reproductive function and additional endocrine-sensitive endpoints. Finally, OECD TG 426 [8] is focused on developmental neurotoxicity. This test guideline study consists of administering the test substance to the female, preferably rats, from mating to lactation and analyses are carried out on both dams and pups. Litters are evaluated for different neurotoxicity effects. The evaluation consists of observations to detect gross neurologic and behavioral abnormalities, including the assessment of physical development, behavioral ontogeny, motor activity, motor and sensory function, and learning and memory; included is also the evaluation of brain weights and neuropathology during postnatal development and adulthood. Dams are tested to assess effects in pregnant and lactating females and may also provide comparative information between dams and their offspring.

As reflected by the complexity of the various test guidelines designed to examine the potential for a chemical to induce reproductive and developmental effects, it is clear that there are numerous possible in vivo responses to chemicals related to the assessment of DART including fetal growth (fetal growth retardation, fetal weight decrease), fetal survival (fetal death, post-implantation loss, pre-implantation loss), structural dysmorphogenesis, visceral organ toxicity, neuropathology, and immunology. Given the need for an assessment of DART to account, with fidelity, for this broad spectrum of in vivo effects, an in silico approach to support the assessment of DART is consequently also a complex process.

Chemical inventories across various jurisdictions including Canada, EU, and US are populated with substances that have limited or no experimental toxicological data. As such, this poses a major challenge globally to regulatory agencies committed to addressing the potential to impact human health. The European legislation REACH requires specific assessment of DART [2] for substances present in the European market; the large numbers of these substances requiring evaluation supports the need for the development of new in silico models to address existing gaps in the empirical data. Currently, predictive models for DART are very few owing to the problems previously described which are essentially related to the high number of poorly understood complex biological processes producing DART effects that in silico methods currently cannot mimic. In addition, few experimental data are available and this is also an issue, since in silico models and the reliability of the predictions are based on validation with experimental biological data.

In this chapter, two publicly available in silico models present in the VEGA (Virtual models for Evaluating the properties of chemicals within a Global Architecture) platform [10] and two statistically based commercial in silico models are described with respect to their predictive application for developmental toxicity. Within the VEGA platform [10], Computer-Assisted Evaluation of industrial chemical Substances According to Regulations (CAESAR)—which addressed REACH—and Procter & Gamble (PG) models are considered; CAESAR is a statistical model made by descriptor, PG is a model made "by experts" that is an adaptation of the Proctor & Gamble (P&G) model. The two commercial models include the Leadscope Model Applier and Multicase CASE Ultra.

2 Materials and Methods

2.1 Applicability Domain

The guide to the VEGA software is described in Subheading 2.4 of Chapter 5. For the developmental toxicity models (the CAESAR and P&G implemented in VEGA), if the applicability domain index (ADI) value is lower than 0.7 the result of the QSAR is not supported by the evaluation done by VEGA on the similar chemicals and as a result the uncertainty of the prediction is higher. In this case, higher reliability of the result can only be achieved using a second QSAR model to confirm the assessment (*see* **Note 2**). Indeed, it is always recommended to have an independent support to the prediction done by a model. This is a general rule, and in the case of the models implemented within VEGA, the software provides an evaluation of the likeliness of the prediction based on the results for substances similar to the target compound. In addition, another independent support of the prediction can be the result of a second model, if in agreement with the result of the first model.

The ADI for developmental toxicity model also determines if the descriptors of the target compounds are inside the model descriptor range. This specific check on the descriptors does not apply to the PG model which is based on fragments. We also notice that, compared to mutagenicity models, the developmental toxicity models are based on smaller data sets, hundreds of compounds versus thousands of them. Thus in general, it is more difficult to find similar compounds, and the ADI values are lower for developmental toxicity models.

2.2 CAESAR-VEGA

The CAESAR model has been described in detail [11]. It is now implemented also within VEGA, offering a better assessment of the applicability domain; this platform is recommended for using the model.

2.2.1 Toxicity Data Source

The CAESAR data set is composed of 292 compounds that include 201 that are positive and 91 that are negative. The compounds were extracted from Arena et al. [12] and subsequently assessed by human experts on the basis of their experience within the CAESAR project (http://www.caesar-project.eu/). In practice, classification of these compounds was conducted using U.S. Food and Drug Administration (FDA) categories, as adopted within the original paper from Arena et al. [12], and then merging the categories into two: FDA categories A and B are considered as non-toxicant, whereas categories C, D, and X are considered toxicant. This data set was then split into two distinct data sets: one data set of 234 compounds which was used as the training set, and the other data set of 58 compounds which was used as the test set.

2.2.2 Description of the Model

The CAESAR model is based on descriptors. The model is a QSAR classification model based on a Random Forest method implemented using WEKA open-source libraries (*see* **Note 3**).

2.2.3 Model Statistics

Cooper statistics for the CAESAR model on its training set are [11]: accuracy = 1.00; specificity = 1.00; and sensitivity = 1.00. The meaning of these parameters has been described within Chapter 5. The performance of the model using the test set of 58 compounds was found to be: accuracy = 0.84; specificity = 0.59; sensitivity = 0.95. These figures are surely closer to the practical situation, compared to the much better results described on the training set.

2.2.4 Interpretation of the Output

CAESAR developmental toxicity model classifies chemicals as toxicant or non-toxicant. The evaluation of the output is done on the basis of the applicability domain index (ADI), provided by VEGA. The basis of this, and the different components of the ADI, shown to the user, have been described in Chapter 5, and will be discussed with examples below in this chapter.

2.3 PG-VEGA

2.3.1 Toxicity Data Source

The PG data set is composed of 716 compounds that include 665 compounds that are positive, 16 that are negative for DART, and 35 have insufficient or incomplete data. This data set was compiled by Wu et al. [13] and authors used values from different sources: they took 260 chemicals that were originally used in the development of a DART approach for Threshold of Toxicological Concern (TCC) for DART [14]; all these values meet the criteria of the TTC approach described by Kroes et al. [15] and have reliable NOAELs for DART endpoints. Other data were taken from the works of Maślankiewicz et al. [16] and Schardein [17]; in the case of pharmaceutical chemicals values were obtained from the ReproRisk® database. Data present in the P&G data set are primarily from studies that show one or more positive in vivo testing results in a mammalian species. When data on in vivo mammalian studies were not available, authors used a weight of evidence approach using data from in vitro assays or non-mammalian in vivo tests. For more details on the empirical data included in this data set consult appendix I of Wu et al. [13]. To develop the PG model in VEGA, only data for developmental toxicity was taken into account; this resulted in a training set of 641 compounds with the related information about each compound category for the developmental toxicity property.

2.3.2 Description of the Model

VEGA implements the P&G decision tree [13]. From the original data set of 716 chemicals, only 641 compounds were included in the training set that have experimental data for developmental toxicity; remaining compounds having data only for reproductive toxicity were excluded in order to have a separate model only for prediction of developmental toxicity. It is composed of 25 categories and six nodes. Each category is composed of one or more groups of chemicals; 5 categories are specific for receptor-binding (in total ten receptors are involved), and the other 20 are chemical structural-related categories, whereas the "nodes" discriminate query compounds on the basis of general chemical features. If a compound belongs to a category, it is classified as toxic for developmental toxicity; if it is not associated with any category then it is classified as non-toxic for developmental toxicity. The PG model has scaffolds for each category that describe groups of chemicals that compose each category. Using the scaffold and the possible substituents, it is possible to generate a list of virtual chemicals that can be toxic compounds taking into account the mode of action (as indicated by Wu et al. [13]) of positive compounds. The VEGA model implements a virtual library of 185,950 structures generated in this way. This model tries to find an exact match between the given compounds and any of the virtual compounds in the library. If a match is found, a prediction of "toxicant" is given, otherwise a "non-toxicant" prediction is provided.

2.3.3 Model Statistics

Cooper statistics (as described in Chapter 5) for the developmental toxicity on this training set are: accuracy = 0.85, specificity = 0.44, sensitivity = 0.89.

2.3.4 Interpretation of the Output

PG model classifies chemicals as toxicant or non-toxicant (*see* **Note 4**). The evaluation of the output is done on the basis of the applicability domain index (ADI), provided by VEGA. The basis of this, and the different components of the ADI, shown to the user, have been described in Chapter 5, and will be discussed with examples below in this chapter.

2.4 Multicase CASE Ultra

There are several CASE Ultra models designed for predicting developmental toxicity of chemicals to a variety of mammals including humans. For the purpose of this paper, the CASE Ultra model that predicts the developmental toxicity to mammals is being considered.

2.4.1 Toxicity Data Source

The training set of the model is composed of 275 chemicals. The empirical toxicity data on these chemicals were obtained from the Chemical Informatics Program (CIP) reprotox database [18, 19]. The majority of these data were created using five publicly available reprotox sources: (1) Reproductive Toxicology Center System (REPROTOX), (2) Shepard's Catalog of Teratogenic Agents, (3) Teratogen Information System (TERIS), (4) The Registry of Toxic Effects of Chemical Substances (RTECS), and (5) The Physicians' Desk Reference (PDR). In addition, a small portion of the reproductive toxicity data were obtained from International Agency for Research on Cancer (IARC) monographs and from original literature articles cited in RTECS.

2.4.2 Description of the Model

CASE Ultra is a fragment-based QSAR model that uses machine learning techniques to learn automatically from training data. It is influenced primarily by the Multiple Computer Automated Structure Evaluation (MCASE) methodology [20, 21]. On the basis of a hierarchical algorithm, MCASE uses Simplified Molecular-Input Line-Entry System (SMILES) codes to generate all possible 2–10 consecutive atom-molecular fragments (hydrogen atoms excluded) of preloaded training compounds with associated toxicity data. The program then statistically compares the fragments of active and inactive chemicals in the training set. Fragments that are primarily associated with active chemicals are identified as "positive alerts" (biophores). Conversely, the program also identifies molecular fragments primarily associated with inhibition of activity or "deactivating alerts" (biophobes). Further, the program identifies modulating factors based on physiochemical descriptors, calculated parameters such as highest occupied molecular orbital and lowest

unoccupied molecular orbital energies, octanol–water partition coefficient, and presence of certain molecular (sub)structures that correlate with enhanced or diminished activity. These parameters are combined to develop a QSAR that generates an estimate of the potential toxicity of the query compound.

2.4.3 Model Statistics

The ratio of positive to negative in the model training set is 166:59. Cooper statistics for this model are: specificity = 0.50, sensitivity = 0.85.

2.4.4 Interpretation of the Output

The CASE Ultra model reports prediction as percent probability of being positive. Generally, probabilities lying below 40 % are considered as non-toxic and those above 60 % as toxic. The zone lying between 40 % and 60 % is designated as the gray zone and probabilities in that area are considered inconclusive. The model also assesses query chemical for presence of contributing positive alerts as well as for presence of unknown fragments; that is fragments not present in any of the training set chemicals used in building the model. In addition to this, the model algorithm also looks for presence of positive modulators (i.e. fragments that support activity) and presence of negative modulators (i.e. fragments that reduce activity). The overall percent probability is computed based on this entire analysis.

2.5 Leadscope Model Applier

There are several models available within the Leadscope Model Applier suite that are designed for predicting a number of different developmental toxicity effects caused by exposure of mammals to chemicals. For the purpose of this paper, the model that predicts an effect related to "fetal survival," i.e. "fetal death mouse" is being considered.

2.5.1 Toxicity Data Source

The training set of the model contains 978 chemicals. It is composed of data obtained from the Chemical Informatics Program (CIP) reprotox database [18, 19].

2.5.2 Description of the Model

The developmental toxicity model pertaining to fetal survival (i.e. fetal death mouse) was built using the Leadscope Prediction Data Miner software. It considers molecular descriptors that include structural features and eight calculated properties [22]. The structural features include Leadscope default hierarchy features plus the predictive scaffolds generated with default settings, whereas the eight calculated properties are parent molecular weight, Log *P*, polar surface area, hydrogen bond acceptors, hydrogen bond donors, number of rotational bonds, and Lipinski score (rule violation). The developmental toxicity was modeled for study calls (e.g. fetal death in mouse), where the positive calls were trained as binary 1 and negative calls as binary 0.

2.5.3 Model Statistics

The ratio of positive to negative chemicals in the model training set is 406:572. Cooper statistics for this model on its training set are: accuracy = 0.68, specificity = 0.94, sensitivity = 0.41. The cross-validation statistics are: accuracy = 0.64; specificity = 0.90; sensitivity = 0.37.

2.5.4 Interpretation of the Output

The outcome of the Leadscope Model Applier QSAR prediction is given as the probability of being a developmental toxicant on a scale of 0 (non-toxic) to 1 (toxic). The prediction results are presented as the "prediction status" and the "positive prediction probability." The prediction status can be "positive," "negative," and "not-in-domain." The higher the probability is, the greater chance there is of the test chemical being toxic for the given endpoint. The model domain is defined for two factors: (1) containing structural model features in addition to property descriptors; (2) being within a similar structure group with at least 30 % similarity (this is set by the model developer). Additionally, the Model Applier allows analog browsing in these databases after a prediction has been made on a test set of compounds which is an added value because it provides an expert user the ability to also look at available empirical/mechanistic data to support the prediction.

3 Interpreting the Results

A comprehensive assessment of predictions is the most critical aspect related to the interpretation of results estimated by (Q)SAR models. VEGA facilitates the interpretability of (Q)SAR predictions by breaking down several critical aspects of the applicability domain as described in Chapter 5. Nevertheless, possible misinterpretations can still take place; the following examples will provide further insights into the application of (Q)SAR models as well as into the analysis and interpretation of (Q)SAR results (*see* **Note 5**).

3.1 Case Study 1: Dichlorobenzene (Fig. 1)

Systematic Name: 1,2-Dichlorobenzene.

CAS Registry Number: 95-50-1.

SMILES: c1ccc(c(c1)Cl)Cl.

Experimental value: Not available.

CAESAR results: Prediction is non-toxic, but the result may not be reliable.

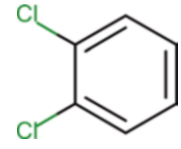


Fig. 1 The structure of dichlorobenzene

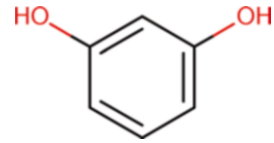


Fig. 2 The structure of the most similar chemical within the CAESAR model for the case study 1

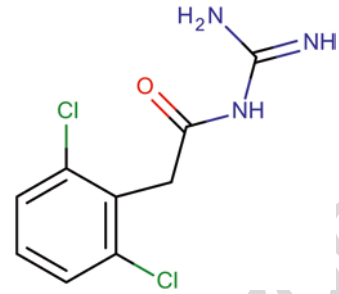


Fig. 3 The structure of the second most similar chemical within the CAESAR model for the case study 1

In the CAESAR model, the query molecule is not present in the training set. Even though the compounds identified as similar by the model have a relatively high similarity index of 0.729 (Fig. 2), it can be observed that the molecular structure is too different compared to the query molecule. The most similar compounds found in the training set have two substituents linked at the benzene, but substituents are in different positions with respect to the query compound and they are oxygen atoms instead of chlorine atoms. This difference may be quite important for the specific endpoint. The second most similar compound (Fig. 3) is a molecule that has two chlorine atoms, but also a more complex chain. From a chemical point of view, the second chemical is surely more complex than the first one. This can directly influence the toxicity profile of the chemical.

The accuracy index is low (0.527) because the model was not able to predict the property correctly for one of the two most similar compounds found in the training set.

Accordingly, the ADI is 0.714 and the prediction may not be conclusive.

Model PG results: prediction is toxic, and the result appears reliable.

In the library of the PG model there is no experimental value for the query compound. However, in the training set of the model there are some similar structures with benzene rings and chloride substituents in different positions (Figs. 4, 5, and 6). The three most similar compounds (Figs. 4, 5, and 6) are experimentally all

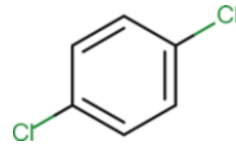


Fig. 4 The structure of the most similar chemical within the PG model for the case study 1

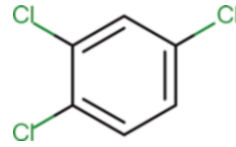


Fig. 5 The structure of the second most similar chemical within the PG model for the case study 1

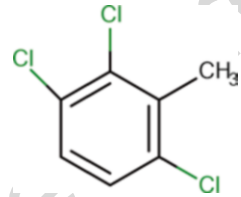


Fig. 6 The structure of the third most similar chemical within the PG model for the case study 1

developmental toxicant, and they have two or more chlorine atoms linked to the benzene ring. Therefore, we can deduce that the presence of chlorine substituents on the benzene ring makes the molecule toxic for developmental toxicity. The ADI is high (0.974) and indeed the prediction of the model for the first two similar molecules is in accordance within the experimental value. As such, the prediction is considered reliable.

Model CASE Ultra results: prediction is toxic, and there is high confidence in the result.

The compound was not used to build the model, and it lies within the applicability domain of the model. The model identified one positive alert and 86.4 % of the molecules (19 out of 22) in the training set that contained this alert and were found to be positive for this endpoint (i.e. developmental toxicity to mammals). Moreover, one positive modulator was also identified in the query chemical structure. No negative modulators of this alert or unknown fragment were found in the query chemical. Based on this analysis, a high confidence can be assigned to the computed probability of 0.72 indicating potential toxicity.

Model Applier results: prediction is non-toxic, but there is low–moderate confidence.

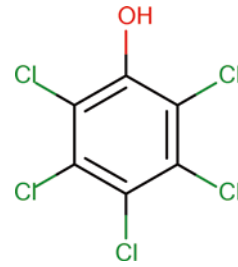


Fig. 7 The structure of the most similar chemical within the Model Applier program for the case study 1

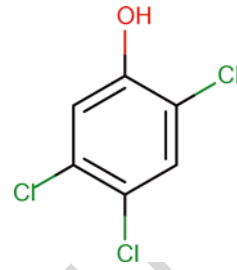


Fig. 8 The structure of the second most similar chemical within the Model Applier program for the case study 1

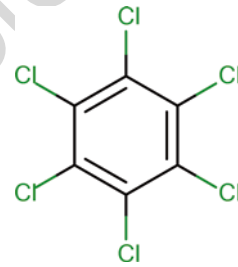


Fig. 9 The structure of the third most similar chemical within the Model Applier program for the case study 1

The compound was not used to build the model, however there are some similar structures in the training set, i.e. a benzene ring with one or more chlorine (and other) substituents in different positions (Figs. 7, 8, 9, and 10). The compounds shown in (Figs. 7 and 8) are reported non-toxic whereas compounds shown in (Figs. 9 and 10) are reported toxic for this endpoint. In this case, the model predicted correctly only the compounds (7) and (8). The analogs do not match structurally that well with the query structure and the query chemical shows only three matching structural features with the training set chemicals. Therefore, there is low–moderate confidence in the model computed prediction probability of 0.32 indicative of potential non-toxicity.

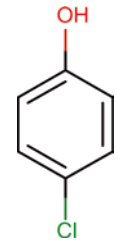


Fig. 10 The structure of the fourth most similar chemical within the Model Applier program for the case study 1

Final assessment: The tested models do not contain the experimental value for the target compound in their databases, thus only predictions can be used. The CAESAR model predicts the query compound as non-toxic but with low ADI. The reported analogs are also not very similar to dichlorobenzene. Similarly, Leadscope Model Applier predicts the query compound non-toxic but, due to the absence of good analogs and fewer matching features, a low–moderate confidence is assigned to the prediction. Conversely, the CASE Ultra model predicts the query chemical as toxic. In addition, the PG model predicts the molecule as toxic and there are some similar compounds in agreement with this assessment. The overall analysis based on the available information suggests that this compound has the potential to induce developmental toxicity with moderate certainty as predictions obtained from both the PG model and the CASE Ultra model have high confidence.

3.2 Case Study 2: 2-Methylresorcinol (Fig. 11)

Systematic Name: 2-Methyl-1,3-benzenediol.

CAS Registry Number: 23-22-3.

SMILES: Oc1cccc(O)c1C.

Experimental value: Not available.

CAESAR results: Prediction is non-toxic; the result appears reliable.

In the CAESAR model the query compound is not present in the training set, but there is a molecule quite similar to the query compound (similarity index (SI) 0.928) (Fig. 12). The model prediction for the most similar molecule is correct, but for the second most similar structure (Fig. 13) the prediction is wrong; accordingly, the model has a low accuracy index (0.518). The second structure (Fig. 13) is quite different compared to the query compound based on the presence of an amide group; this affects the overall ADI of the target compounds: 0.803. The prediction of query compounds has an ADI of 0.803 which is considered to be quite high.

Model PG results: prediction is non-toxic, but the result appears to be unreliable.

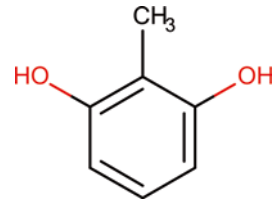


Fig. 11 The structure of 2-methylresorcinol

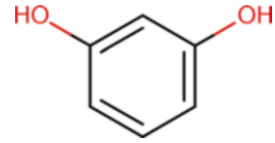


Fig. 12 The structure of the most similar chemical within the CAESAR model for the case study 2

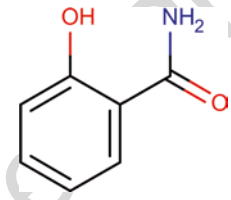


Fig. 13 The structure of the second most similar chemical within the CAESAR model for the case study 2

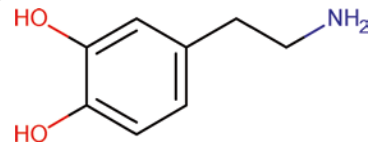


Fig. 14 The structure of the most similar chemical within the PG model for the case study 2

The library of the PG model does not have an experimental value for the query compounds and there are not many similar compounds in the training set; the most similar compound has a similarity index (SI) of 0.846, which is considered to be acceptable (Fig. 14). This similar compound has two phenolic groups as the target, however it also has a propilamine which can change the assessment of the compound. Moreover, five similar compounds out of six are experimentally reported as developmental toxicants (i.e. toxic), and only one similar compound is reported as non-toxic, but the model predicts it as toxic. As such, there is disagreement between the prediction for the query compound and this similar compound. Hence, ADI is very low: 0.639.

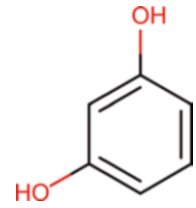


Fig. 15 The structure of the most similar chemical within the Model Applier program for the case study 2

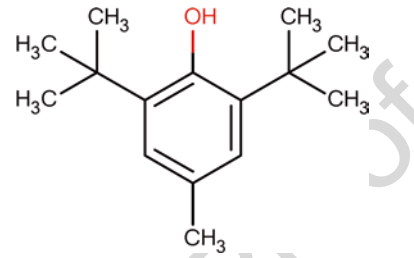


Fig. 16 The structure of the second most similar chemical within the Model Applier program for the case study 2

Model CASE Ultra results: prediction is inconclusive.

The compound was not used to build the model. The compound lies within the applicability domain of the model. No contributing positive alerts and no unknown fragments were found by the model. However, model computed probability of 0.49 fell inside the gray zone (0.40–0.60) around the model's current classification threshold (50.0 %), therefore, the results of activity prediction were considered inconclusive.

Model Applier results: prediction is non-toxic, and there is high confidence in the result.

The compound was not used to build the model, however there are some similar structures in the training set, i.e. a benzene ring with hydroxyl substituents in different positions as illustrated in Figs. 15, 16, 17, and 18. The compounds (Fig. 15) to (Fig. 17) are reported non-toxic for the endpoint, whereas compound (Fig. 18) was reported toxic. When compared with their empirical data, compounds (Fig. 15) to (Fig. 17) were found to be correctly predicted by the model whereas compound (Fig. 18) was found to be incorrectly predicted. The query chemical shows seven matching structural features with the training set chemicals. From the perspective of the presence of similar functional groups and matching features, there is enough justification for the prediction. Therefore, a high confidence is assigned to the model computed prediction probability of 0.24, which is indicative of potential non-toxicity.

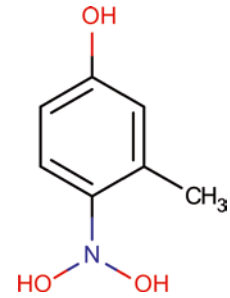


Fig. 17 The structure of the third most similar chemical within the Model Applier program for the case study 2

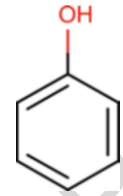


Fig. 18 The structure of the fourth most similar chemical within the Model Applier program for the case study 2

Final assessment: Models do not have an experimental value for methyl resorcinol. The ADI for CAESAR is quite high and the compound is predicted as non-toxic. Similarly, Leadscope Model Applier predicted the compound as non-toxic with high confidence. Model PG has difficulty to predict this compound based on (partially) related compounds. Similarly, CASE Ultra was not able to predict the activity. Therefore, in the final assessment, predictions by Leadscope and CAESAR are taken into account. Thus, CAESAR and Leadscope models illustrate how the information on structural analogs of query chemical can contribute to the overall assessment and in this particular example to make an overall conclusion of non-toxic.

3.3 Case Study 3: dl-Norepinephrine (Fig. 19)

Systematic Name: 4-(2-Amino-1-hydroxyethyl)-1,2-benzenediol.
CAS Registry Number: 138-65-8.
SMILES: Oc1ccc(cc1(O))C(O)CN.
Experimental value: Not available.
CAESAR results: Prediction is toxic, the result appears reliable.

The CAESAR model does not have an experimental value for the query compound. The most similar structure found in the training set (Fig. 20) is very similar to the query compound and is toxic. The only structural difference between the query compound and similar compound is at the terminal nitrogen where an extra methyl group is present. All the other compounds have a good SI

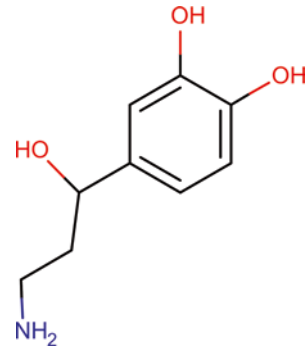


Fig. 19 The structure of dl-norepinephrine

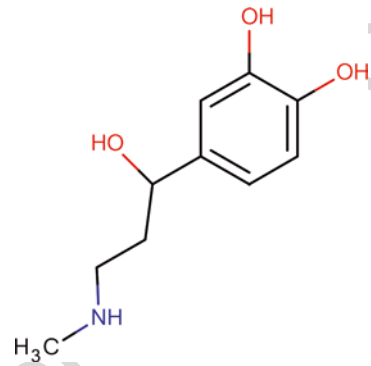


Fig. 20 The structure of the most similar chemical within the CAESAR model for the case study 3

and are almost all toxic except one that is non-toxic. In this case, the CAESAR model is able to predict all the similar compounds correctly, so the prediction has a very high ADI (0.97).

Model PG results: prediction is toxic, and the result appears reliable.

The library of the PG model does not have an experimental value for the query compound. Structures similar to the query compound were found, however unlike the CAESAR model, the PG library makes more errors in the prediction of the similar compounds. Notably, the model is able to correctly predict structures that are most similar to the query compound. Their similarity was greater than 0.94, as defined by the VEGA software [23] (*see* **Note 2** *for values of the similarity and thresholds applied*). For this reason, the prediction by PG is considered to be relatively less reliable than that obtained by the CAESAR model, yet acceptable.

Model CASE Ultra results: prediction is inconclusive.

The compound was not used to build the model. No contributing positive alerts were detected. However, one unknown fragment was identified in the chemical. Moreover, the prediction was found

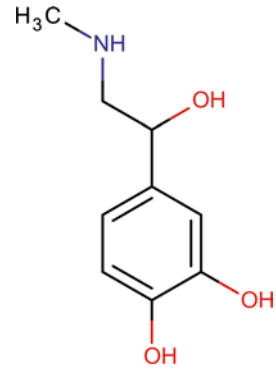


Fig. 21 The structure of the most similar chemical within the Model Applier program for the case study 3

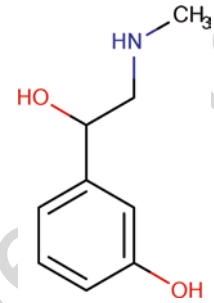


Fig. 22 The structure of the second most similar chemical within the Model Applier program for the case study 3

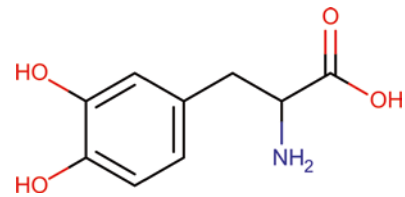


Fig. 23 The structure of the third most similar chemical within the Model Applier program for the case study 3

to fall inside the gray zone (0.40–0.60) and therefore the results of activity prediction were considered inconclusive.

Model Applier results: prediction is non-toxic, and there is low–moderate confidence in the result.

The compound was not used to build the model, however there are some similar structures in the training set (Figs. 21, 22, 23, and 24). Based on the empirical information, compounds (Fig. 21) and (Fig. 22) are non-toxic and compounds (Fig. 23) and (Fig. 24) are toxic. When using Model Applier, compounds

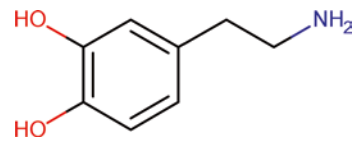


Fig. 24 The structure of the fourth most similar chemical within the Model Applier program for the case study 3

(Fig. 21) and (Fig. 22) are correctly predicted and compounds (Fig. 23) and (Fig. 24) are incorrectly predicted. Like the query structure, compounds (Fig. 23) and (Fig. 24) have a free amino group which in structures (Fig. 23) and (Fig. 24) is substituted with a *N*-methyl group. From the perspective of analogous structures and their predictions, there is insufficient justification for the prediction on the query chemical. Chemicals which are apparently quite similar have different toxicity, and the reason for the different property values is not clear. Therefore, there is low to moderate confidence in the model computed prediction probability of 0.17, which is indicative of potential lack of toxicity for the query chemical.

Final assessment: Model training sets do not have experimental value for norepinephrine. Both the CAESAR and PG models predict the target chemical as toxic with high reliability. The CAESAR model has more similar structures than the PG Model and predicts correctly almost all similar structures found in the training set. We clarify that the ADI algorithm within CAESAR uses the first two similar molecules (however, we always recommend the user to consider all six similar substances). Since these two models are in agreement, this compound can be considered as a toxic prediction with higher certainty. Conversely, the results obtained from CASE Ultra were found to be inconclusive. The non-toxic prediction obtained from Leadscope Model Applier has only low to moderate confidence for this endpoint, and further investigation of similar compounds suggests that they are toxic. This result obtained from Leadscope model introduces uncertainty in the overall assessment based on CAESAR and PG library. The inconclusive results obtained from CASE Ultra do not help in this case. Thus, the chemical may be assumed toxic, but with a moderate margin of uncertainty.

4 Conclusions

Developmental toxicity is one of the most challenging endpoints in the area of QSAR-based predictive toxicology. Some models on this complex endpoint are available. For molecules that fall within the model's applicability domain the predictions can be moderately reliable, when the criteria, defined by each model, are

met, such as high ADI value (as explicitly indicated in the summary cover page). Even though read across can be used for data gap filling for this type of endpoint, the QSAR models discussed above when applied with expert judgment may potentially support chemical screening as an initial starting point for exploring the potential for developmental toxicity of query compounds. In fact, the case studies presented in this chapter illustrate that in absence of experimental data, the potential for toxicity of a query compound may be projected by expertly weighing the predictions from multiple QSAR models and data from compound's analogs, provided that there is sufficient agreement on the results provided by the different models (*see* **Note 5**). If there are reasons for moderate to high uncertainty, due to the limited number of chemicals with experimental data, and of the limited performance of the models, no final conclusion can be achieved.

5 Notes

1. Please notice that DART includes a wide series of effects. The models illustrated in this chapter are based on a single defined endpoint (e.g. Leadscope Model Applier) or reflect the analysis of the many different effects that exist within a broad spectrum of possible DART endpoints (CaseUltra, CAESAR, P&G). Hence, it is important to interpret (especially) a negative/non-toxic with diligence/supplement with further empirical data from analogs/read across.
2. Threshold values for different properties of the applicability domain could change for different models. In particular in CAESAR model for mutagenicity, threshold values for similarity are 0.85 and 0.7, while CAESAR developmental toxicity model values are 0.8 and 0.7. If the similarity value is below 0.7, the two substances should not be considered similar. Accuracy and concordance thresholds for CAESAR mutagenicity model are 0.9 and 0.5, for developmental toxicity model are 0.8 and 0.5 and thresholds for ADI in mutagenic model are 0.9 and 0.7 and for developmental toxicity model are 0.8 and 0.7. These values are influenced by the number of compounds of the training set.
3. A very similar model is also implemented within T.E.S.T. [24]. Indeed, there are some similarities between T.E.S.T. and VEGA models. The few differences between them pertain to (1) difference in algorithm that is used to represent the chemical structures, and this may cause minor alterations in the results for some chemicals, (2) difference in the approach to refer to similar chemicals and calculation of the applicability domain. For instance, T.E.S.T. shows similar compounds and the relative statistics for the training and test sets separately,

providing information useful to evaluate the performance inside and outside the training set. Conversely, the approach of VEGA is to provide an overall evaluation on the reliability of the model for the specific chemical, showing the sub-indices as described in Chapter 5, using all chemicals of the training and test sets.

4. Even if PG model does not use the decisional tree present in Wu et al. [13], the reference at the category described in Wu et al. [13] output is present in VEGA.
5. There are some commercial or freely available software programs that can predict developmental toxicity. In addition to the models described here, other examples of freely available models are T.E.S.T. (Toxicity Estimation Software Tool) [24] and OECD QSAR Toolbox [25].

References

1. Hood R (2011) Developmental and reproductive toxicology: a practical approach, 3rd edn. CRC, Boca Raton, FL
2. Regulation (EC) No. 1907/2006 of the European Parliament and of the Council, of December 18, 2006 concerning the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission
3. OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
4. OECD (2001) Test No. 414: prenatal development toxicity study, OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
5. OECD (2001) Test No. 416: two-generation reproduction toxicity, section 4. OECD Publishing, Paris
6. OECD (1995) Test No. 421: reproduction/developmental toxicity screening test, OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
7. OECD (1996) Test No. 422: combined repeated dose toxicity study with the reproduction/developmental toxicity screening test, OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
8. OECD (2007) Test No. 426: developmental neurotoxicity study, OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
9. OECD (2012) Test No. 443: extended one-generation reproductive toxicity study, OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
10. Virtual models for evaluating the properties of chemicals within a global architecture. http://www.vega-qsar.eu/
11. Cassano A, Manganaro A, Martin T, Young D, Piclin N, Pintore M, Bigoni D, Benfenati E (2010) CAESAR models for developmental toxicity. Chem Central J 4(Suppl 1):S4
12. Arena VC, Sussman NB, Mazumdar S, Yu S, Macina OT (2004) The utility of structure-activity relationship (SAR) models for prediction and covariate selection in developmental toxicity: comparative analysis of logistic regression and decision tree models. SAR QSAR Environ Res 15:1–18
13. Wu S, Fisher J, Naciff JM, Laufersweiler MC, Lester C, Daston G, Blackburn K (2013) A framework for identifying chemicals with structural features associated with potential to act as developmental or reproductive toxicants. Chem Res Toxicol 26:1840–1861
14. Laufersweiler M, Gadagbui B, Baskerville-Abraham I, Maier A, Willis A, Scialli A, Carr G, Felter S, Blackburn K, Daston G (2012) Correlation of chemical structure with reproductive and developmental toxicity as it relates to the use of the threshold of toxicological concern. Regul Toxicol Pharmacol 62:160–182
15. Kroes R, Renwick A, Cheeseman M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos J, Wurtzen G (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem Toxicol 42:65–83

16. Maślankiewicz L, Hulzebos E, Vermeire T, Muller J, Piersma A (2005) Can chemical structure predict reproductive toxicity? RIVM report 601200005/2005
17. Schardein JL (2000) Chemically induced birth defects. Marcel Dekker Inc., New York
18. Matthews EJ, Kruhlak NL, Benz RD, Contrera JF (2007) A comprehensive model for reproductive and developmental toxicity hazard identification: I. Development of a weight of evidence QSAR database. Regul Toxicol Pharmacol 47:115–135
19. Matthews EJ, Kruhlak NL, Benz RD, Ivanov J, Klopman G, Contrera JF (2007) A comprehensive model for reproductive and developmental toxicity hazard identification: II. Construction of QSAR models to predict activities of untested chemicals. Regul Toxicol Pharmacol 47:136–155
20. Chakravarti SK, Saiakhov RD, Klopman G (2012) Optimizing predictive performance of CASE Ultra expert system models using the applicability domains of individual toxicity alerts. J Chem Inf Model 52:2609–2618
21. Klopman G (1992) MULTICASE 1. A hierarchical computer automated structure evaluation program. Quant Struct Act Relat 11:176–184
22. Yang C, Cross K, Myatt GJ, Blower PE, Rathman JF (2004) Building predictive models for protein tyrosine phosphatase 1B inhibitors based on discriminating structural features by reassembling medicinal chemistry building blocks. J Med Chem 47:5984–5994
23. Floris M, Manganaro A, Nicolotti O, Medda R, Mangiatordi GF, Benfenati E (2014) A generalizable definition of chemical similarity for read-across. J Cheminform 6:39
24. Quantitative Structure Activity Relationship, Toxicity Estimation Software Tool (TEST). http://www.epa.gov/nrmrl/std/qsar/qsar.html. Accessed 19 Jun 2015
25. The OECD QSAR Toolbox. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm

Chapter 9

In Silico Models for Repeated-Dose Toxicity (RDT): Prediction of the No Observed Adverse Effect Level (NOAEL) and Lowest Observed Adverse Effect Level (LOAEL) for Drugs

Fabiola Pizzo and Emilio Benfenati

Abstract

The preclinical stage in drug development requires the determination of repeated-dose toxicity (RDT) in animal models. The main outcome of RDT studies is the determination of the no observed adverse effect level (NOAEL) and the lowest observed adverse effect level (LOAEL). NOAEL is important since it serves to calculate the maximum recommended starting dose (MRSD) which is the safe starting dose for clinical studies in human beings. Since in vivo RDT studies are expensive and time-consuming, in silico approaches could offer a valuable alternative. However, NOAEL and LOAEL modeling suffer some limitations since they do not refer to a single end point but to several different effects and the doses used in experimental studies strongly influence the final results. Few attempts to model NOAEL and LOAEL have been reported. The available database and models for the prediction of NOAEL and LOAEL are reviewed here.

Key words Repeated-dose toxicity, NOAEL, LOAEL, Drug safety, In silico models, Chronic toxicity

1 Introduction

Repeated-dose toxicity (RDT) studies are designed to determine the effects of repeated oral, dermal, or inhalation exposure to a substance over a specific period of time [1]. Characterization of the toxicological profile of the test substance after repeated exposure is the primary goal of RDT study. RDT tests provide detailed information to identify the adverse effects, the potential target organs or systems (reproductive system, liver, kidney, central nervous system, endocrine system), and the persistence or reversibility of the effects [2].

Toxicity after repeated dosing must also be tested to contribute to the development of safe medicinal products that are to be given repeatedly [3].

Drug development is a long, complex, and expensive process. The typical procedure comprises three major steps: discovery,

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_9,

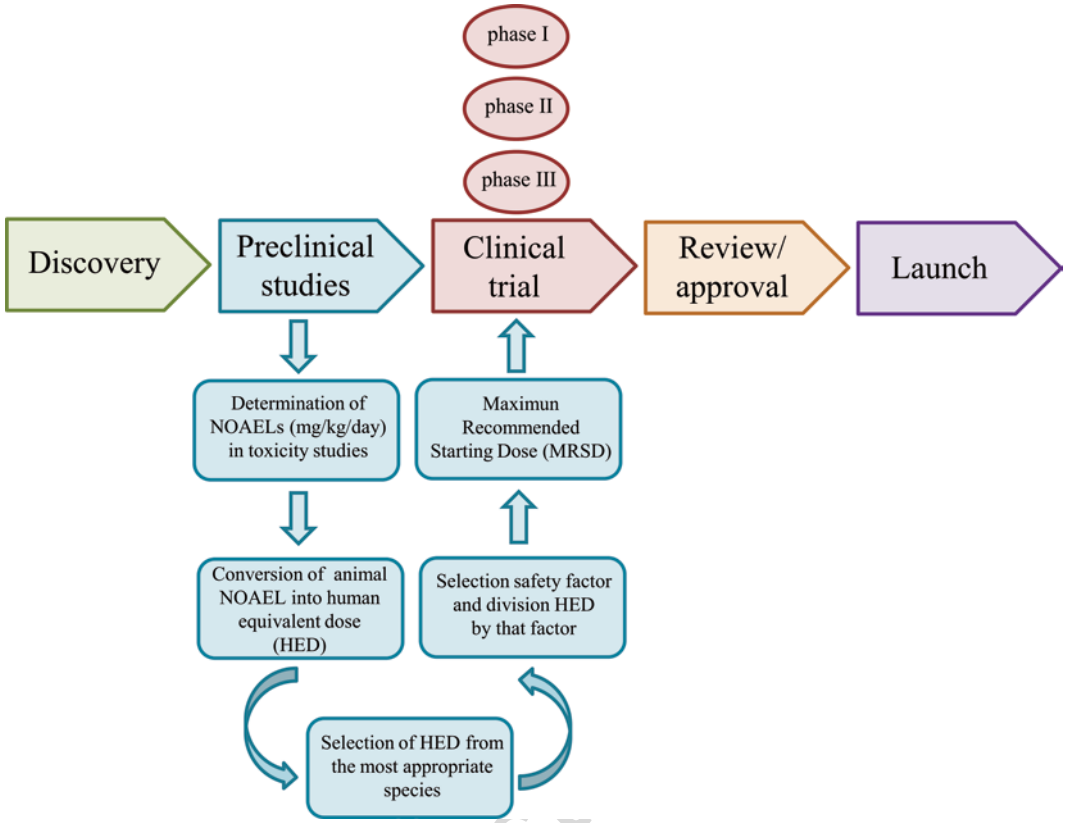


Fig. 1 Scheme of the typical drug development

preclinical development, and clinical trial [3, 4] (Fig. 1). Clinical trials involving daily chronic dosing require RDT studies on animal models (two species, one non-rodent) in the preclinical stage [3]. The no observed adverse effect level (NOAEL) and the lowest observed adverse effect level (LOAEL), the main outcomes of these studies, are of the utmost importance in the non-clinical risk assessment. Although the definitions of NOAEL and LOAEL are debated [5], generally, NOAEL is the highest dose without any biologically significant adverse effects, while LOAEL refers to the lowest exposure at which adverse effects are seen (Fig. 2). NOAEL, determined in non-clinical safety studies in the most appropriate animal species, gives important information for the first dose in humans [6]. NOAEL is essential to calculate the maximum recommended starting dose (MRSD), the dose used in the first human study (clinical trial) [7] (Fig. 1).

Besides pharmaceuticals [8, 9], other regulatory contexts require RDT testing to assess the potential risks of a substance: industrial chemicals [10], agrochemicals [11, 12], biocides [13], and cosmetics [1, 14].

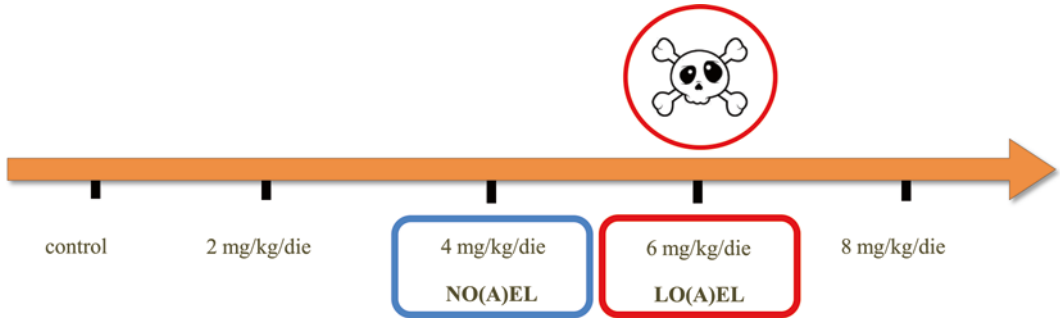


Fig. 2 Identification of the lowest observed adverse effect level (LOAEL) and no observed adverse effect level (NOAEL)

Considering the high cost of drug failure and withdrawal due to toxicity found in the development process, the potential toxicity of a drug needs to be determined as soon as possible [15]. The importance of the results of RDT studies for the evaluation of the safety of chemicals is undeniable, but the in vivo tests are time-consuming and very expensive [16]. The possibility of obtaining the same information using non-testing methods is tempting, though considering the peculiar nature of NOAEL and LOAEL, their computational modeling is a challenge. Few attempts have been made to model NOAEL and LOAEL. A review of the software, databases available, and published models is presented here.

2 LOAEL and NOAEL Databases

Databases containing NOAEL and LOAEL values are available, with a high percentage of overlap between the different sources (Table 1). Generally, for NOAEL and LOAEL, the measurement unit is expressed as mg/kg body weight/day. In order to build accurate computational models, the quality of the chemical structures and data is crucial [17]. In addition, for LOAEL and NOAEL, not only is the final number important but other supporting information is too, such as route and duration of exposure, species and strain used, space between doses, and organ level effects, in order to properly assess the quality and the potential use of these data for modeling.

The RepDose database, developed by Fraunhofer ITEM as part of a project funded by the European Chemical Industry Council (CEFIC), contains experimental NOAEL and LOAEL values for 655 chemicals related to oral (gavage, feeding, and drinking water) or inhalation studies in rodents exposed to the substance over at least 14 days. The chemicals in the database have a limited number of functional groups since complex and multi-functional chemical structures such as pharmaceuticals, inorganic

Table 1
RDT databases for modeling

Name	No. of substances	Developers	Exposure route	Exposure duration	Animal models	Available at	Copyright	Drugs
RepDose	655	Fraunhofer ITEM	Oral, inhalation	At least 14 days	Rodents	http://fraunhofer-repdose.de/	Yes	No
Munro	613	Munro et al. (1996)	Oral	Up to 910 days	Rodents, rabbit	QSAR OECD Toolbox	No	No
HESS	500	NITE	Oral	28–120 days	Rats	QSAR OECD Toolbox	No	No
IRIS	Over 500	US EPA	Many	Many	Rodents, non-rodents	http://www.epa.gov/iris/intro.htm	No	No
COSMOS	1660	COSMOS project	Many	Many	Rodents, non-rodents	http://cosmosdb.cosmostox.eu	No	Yes
ToxRef	310	US EPA	Many	Many	Rodents, non-rodents	QSAR OECD Toolbox	No	Yes

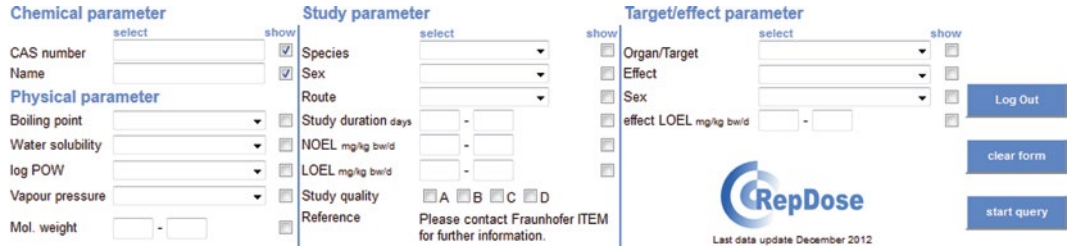


Fig. 3 Query form of the RepDose database

or metal compounds, and mixtures were eliminated [18]. A score (A, B, C, D) indicating the data quality is also provided. Details on the animals used (strain, sex, number per dose group) and the exposure (duration and route, postexposure observation period, and dose groups) are also provided. The database includes toxicological (effect data include all target organs with all associated effects and corresponding LOAEL) and physicochemical (molecular weight, solubility in water, physical state, boiling point, dissociation constant, octanol-water partition coefficient, and vapor pressure) data. The RepDose database is available at http://fraunhofer-repdose.de/, and access is free on registration by the user. A user-friendly query screen (Fig. 3) puts several questions regarding the influence of structural features and physicochemical data on LOAEL, target organs, and effects [18]. Although all the data in the database are displayed, their use is restricted.

Munro et al. [19] provide NOAEL and LOAEL values for 613 organic compounds related to sub-chronic, chronic, reproductive toxicity, and teratogenic studies in rodents and rabbits. For each compound the chemical name, CAS number, structural classification using the decision tree of Cramer et al. [20], species, sex, route of exposure, doses tested, study type, duration, end points, NOAEL, and LOAEL references are reported. The data come from four sources: US National Toxicology Program (NTP) technical reports (post-1984), the toxicological monographs prepared by the Joint FAO/WHO Expert Committee on Food Additives (JECFA), the Integrated Risk Information System (IRIS) database, and the Developmental and Reproductive Toxicology (DART) database. The compounds in the Munro database represent a variety of chemicals (e.g., pesticides, food additives, industrial chemicals). To demonstrate that a study is rigorous enough to detect toxic effects, a compound needs to have both NOAEL and LOAEL to be included in the database; however, in some cases, the LOAEL is not available because the substances are major food ingredients and had no toxicity at the highest dose tested in well-conducted studies [19]. The database is downloadable from the QSAR OECD Toolbox; otherwise, the publication provides a paper version of the database.

The Hazard Evaluation Support System (HESS) database comprises 500 chemicals for which RDT data were obtained from test reports of Japanese CSCL by the Ministry of Health, Labour and Welfare, the National Institute of Technology and Evaluation (NITE), and the Ministry of Economy, Trade and Industry (METI) and from reports produced by the US NTP [21]. All these tests were conducted in compliance with GLP principles. This database contains detailed RDT data related to sub-chronic and chronic (28–120 days) oral exposure in rats. The HESS database, freely downloadable from QSAR OECD Toolbox, provides information for the target compounds such as CAS number, chemical name, SMILES, exposure route and duration of the studies, animal used (strain, sex), toxicological data (organ, tissue, effects, largest and smallest doses used) and NOAEL/LOAEL values.

The Integrated Risk Information System (IRIS) is a publicly available repository, developed by the US Environmental Protection Agency (EPA) that contains information on over 500 chemicals. It provides descriptive and quantitative chronic health information on chemicals found in the environment in order to support risk decision-making [22]. Two main categories of effects are present in IRIS database: non-cancer (oral reference doses and inhalation reference concentrations: RfDs and RfCs) and cancer effects. NOAEL and LOAEL are reported with a detailed summary of the studies containing information on the species used, route and duration of exposure, concentrations tested, and target organs. The user can consult data on the EPA website (http://cfpub.epa.gov/ncea/iris/index.cfm?fuseaction=iris.showSubstanceList); substances are listed in alphabetical order.

The COSMOS database [23] contains 12,538 toxicological studies for 1660 compounds. Two datasets are available: US FDA PAFA and oRepeatToxDB. The first contains 12,198 studies across 27 end points including both repeated-dose (in this case the lowest effect level, LEL, is reported) and genetic toxicity data. ORepeatToxDB, assembled by the COSMOS consortium, contains 340 in vivo repeated-dose toxicity studies from different sources (EC REACH project, US NTP) for 228 chemicals. It reports observed toxicological effects together with the sites at which the effect occurred. Figure 4 reports the typical output of a COSMOS database query. The user needs to be registered for a free account. The COSMOS database was built in the context of the EC project SEURAT-1, partly funded by Cosmetics Europe.

The Toxicity Reference Database (ToxRefDB), developed by US EPA [24], comprises thousands of animal toxicity studies (reporting NOAEL and LOAEL) after testing hundreds of different chemicals. ToxRefDB is freely downloadable from the QSAR OECD Toolbox or can be consulted at the US EPA website (http://actor.epa.gov/toxrefdb/faces/Home.jsp).

Although none of these databases contains only NOAEL and LOAEL data for drugs, some of them cover pharmaceuticals.

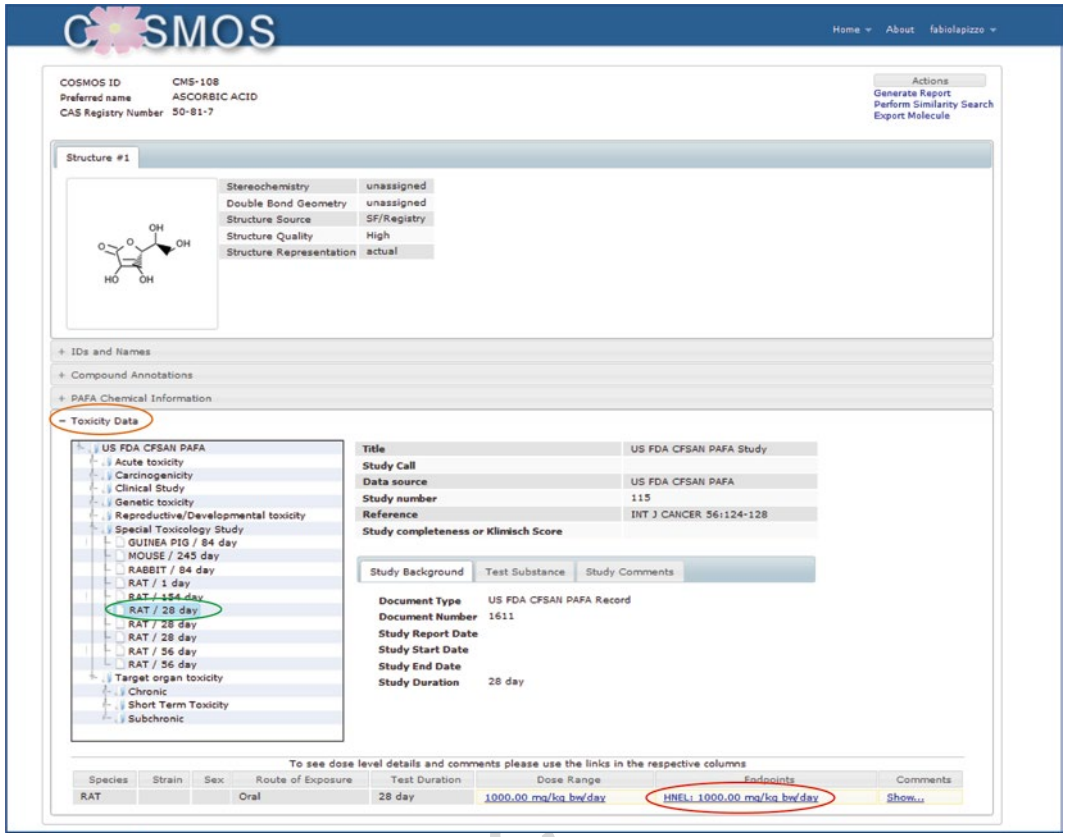


Fig. 4 Typical output of a query using the COSMOS database. In the toxicity data section (*orange*), the exposure duration and the animal used for the in vivo experiment (*green*) are indicated, and the RDT study is reported at the bottom of the screen (*red*) as highest no effect level (HNEL)

3 In Silico Models for the Prediction of LOAEL and NOAEL

A limited number of in silico models are available for the prediction of LOAEL and NOAEL. Published models and software are reviewed here.

3.1 Published Models

The models described here were not built primarily to predict NOAEL and LOAEL for pharmaceuticals; indeed, the compounds used for modeling came from different industrial and environmental contexts. The performances are close to acceptability and do offer a good starting point for the development of a reliable model that can be used in a multidisciplinary context. Table 2 provides a general overview of the literature-based models.

One of the most recent models for the prediction of RDT is described in Toropov et al. [25]. They modeled NOAEL for 113 organic compounds using the Monte Carlo method and three molecular descriptors. The dataset was split three times and the average performances in the training set (97 compounds) in terms

t2.1 **Table 2**
t2.2 **General overview of published models for the prediction of LOAEL and NOAEL**

t2.3 t2.4 Reference	End point	Modeling method	Descriptors	Data used	Training set size	Test set size	Drugs
t2.5 t2.6 Toropov et al. (2015)	NOAEL	Monte Carlo	3	28 and 90 days oral exposure in rats	97	16	No
t2.7 t2.8 Gadaleta et al. (2014)	LOAEL	*k*-NN	Fingerprints + 3 structural keys	90 days oral exposure in rats	254	179	No
t2.9 t2.10 Toropova et al. (2014)	NOAEL	Monte Carlo	8	28 days oral exposure in rats	~180	~21	No
t2.11 t2.12 Sakuratani et al. (2013)	LOAEL	Read-across	None	From 28 to 90 days oral exposure in rats	500	None	No
t2.13 t2.14 Mazzatorta et al. (2008)	LOAEL	GA-PLS for selecting descriptors and LOO-SMLR for model generating	19	Longer than 180 days oral exposure in rat	445	None	No
t2.15 t2.16 t2.17 Garcia-Domenech et al. (2006)	LOAEL	Furnival-Wilson algorithm for selecting descriptor; MLR for (LOO cross-validation) model generating	11	Chronic studies in rats	87	16	No
t2.18 t2.19 De Julián-Ortiz et al. (2005)	LOAEL	Stepwise procedure for selecting descriptor; LDA for model generating	15	Chronic studies	86	17	No
t2.20 t2.21 Matthews et al. (2004)	MRTD/NOEL	Structural alerts	None	Human, oral route, 3–12 months	1309	Leave more out	Yes

of R^2 and RMSE were, respectively, 0.52 and 0.61. In the test set (16 compounds), the performance in terms of R^2 and RMSE ranged from 0.62 to 0.73 and from 0.44 to 0.52, respectively.

Gadaleta et al. [26] using the *k* nearest neighbors (*k*-NN) algorithm, a computational technique based on the concept of similarity, built a model for the prediction of LOAEL. However, to improve the performance, the basic algorithm was refined by setting additional conditions, and a target chemical must fulfill all those rules to be considered reliably predicted. The training and test sets of the model comprised, respectively, 254 and 174 organic compounds, and R^2 for the two sets ranged from 0.632 to 0.769 and from 0.552 to 0.682, considering the different *k*. This model will be implemented in the VEGA (http://www.vega-qsar.eu/) platform and will be freely available.

Toropova et al. [27] modeled 218 NOAEL data (28 days of oral exposure in the rats) using the Monte Carlo method. *R2* for the training and test sets ranged from 0.679 to 0.718 and from 0.61 to 0.66, respectively, considering the different splits.

Sakuratani et al. [28] identified 33 chemical categories related to individual types of toxicity on the basis of mechanistic knowledge starting from a training set of 500 chemicals with RDT data related to oral exposure between 28 and 120 days in rats. Chemicals were assigned to a given category, and then the LOAEL was derived as the result of a data gap-filling approach by read-across on other chemicals in the category. This model does not provide figures for the LOAEL but can be used to identify the target organ most likely to be affected by the target chemical. The category library has been implemented and is available through the Hazard Evaluation Support System (HESS) integrated computational platform.

A further model for the prediction of LOAEL was developed by Mazzatorta et al. [29], applying an integrated approach of genetic algorithm (GA) and partial least squares (PLS). Selected descriptors (19 from DRAGON) were used to develop a LOAEL predictive model through a leave-one-out stepwise multiple linear regression (LOO-SMLR) starting from a set of 445 chronic toxicity data (180 days or more of oral exposure in rats) selected from several sources. The final dataset included pesticides, drugs, and natural products. This model performed as follows: R^2 0.570 and RMSE 0.700. No external validation was done, so the real predictive model's power is not known. However, the performances of LOO cross-validation were q^2 0.500 and RMSE 0.727.

De Julián-Ortiz et al. [30] used a dataset of chronic LOAEL data for 234 compounds compiled from different sources (US Environmental Protection Agency, EPA, and National Cancer Institute/National Toxicology Program, NTP) to build a multilinear regression model (MLR). They selected 15 topological descriptors by a Furnival-Wilson algorithm from among those in the

DESCRI program. MLR and the Furnival-Wilson algorithm were also applied to a smaller but more homogeneous dataset (86 compounds). The results on the first 234 compounds were quite poor (R^2 0.524 and RMSE 0.74). However, the performance on the second dataset (86 compounds) was significantly better (R^2 0.647 and RMSE 0.66). In both cases no external validation was done.

García-Domenech et al. [31] applied the same techniques (Furnival-Wilson for descriptor selection and MLR for model building) on the same 86 molecules used by De Julián-Ortiz et al. [30]. The model, based on six descriptors, was validated on 16 external chemicals. Performances in the training set were R^2 0.795 and RMSE 0.517; q^2 0.719 and RMSE 0.564 in LOO cross-validation and R^2 0.712 and RMSE 0.853 in external validation.

To the best of our knowledge, Matthews et al. [32], Toropova et al. [27], and Toropov et al. [25] are the only studies that report attempts at NOAEL modeling.

Matthews et al. [32] used Maximum Recommended Therapeutic Dose (MRTD) data for 1309 pharmaceutical substances for classification modeling. The MRTD (or Maximum Recommended Daily Dose, MRDD) was determined from clinical trials that employed an oral route of exposure and daily treatments, usually for 3–12 months. The MRTD is derived from human clinical trials and is an upper dose limit beyond which the efficacy of a drug does not increase and/or adverse effects start to outweigh the beneficial ones [33]. MRTD and NOEL for drugs are directly related in humans [32]. An analysis of the MRTD database indicated that most drugs do not show efficacy or adverse effects at a dose approximately ten times lower than the MRTD. Based on this observation, Matthews et al. [32] calculated NOEL as MRTD/10. Chemicals with low MRTD/NOEL were considered strongly toxic, whereas those with higher values were labeled as safe, and structural alerts were identified on this basis. The predictive ability of this model was evaluated through leave-more-out external validation (40 compounds were removed from the training data set of 120 selected test chemicals), and the results showed that the model gave good predictions of toxicity for the test chemicals; the positive predictivity and specificity were high, at, respectively, 92.5 % and 95.2 %, whereas the sensitivity was lower (74.0 %).

3.2 Software

Two software are available for the prediction of LOAEL, both commercial. The first is Toxicity Prediction by Komputer Assisted Technology (TOPKAT), developed by Accelrys®. The TOPKAT model aims to predict chronic oral LOAEL in rats (studies lasting 12 or more months were considered) and has been described in Mumtaz et al. [34]. Starting from a dataset of 234 heterogeneous chemicals, the model was built using a stepwise regression analysis with 44 descriptors selected from an initial pool of electronic,

topological, symmetry descriptors and molecular connectivity indices. The performance of the model was tested comparing the predicted with the experimental LOAEL. About 55 % of the compounds were predicted within a factor of 2 and more than 93 % within a factor of 5 [34].

Over the years the TOPKAT model for LOAEL prediction has been refined, including more data in the training set. Using the expanded training set (393 chemicals), models for five chemical classes were developed (acyclics, alicyclics, heteroaromatics, single benzenes, and multiple benzenes). Venkatapathy et al. [35] tested the predictive performance of the five sub-modules using a large dataset of 656 substances and the R^2 ranged between 0.78 (multiple benzenes) and 0.98 (alicycles). TOPKAT was further validated by Tilaoui et al. [36] using 340 chemicals not included in the TOPKAT training set. TOPKAT correctly predicted (with an error lower than 1 log unit) only 33 % of these chemicals [16].

Another software for LOAEL prediction has been developed by Molcode Ltd. using RDT data in the rat. Information about this model is available from the QSAR Model Reporting Format (QRMF) document. The model is proprietary, but the training and test sets are available. The model was developed using multilinear regression, and the descriptors were chosen through a stepwise selection. There were 76 compounds in the training set, and in order to validate the real ability of the model to predict LOAEL, an external dataset containing 18 compounds was used. In terms of R^2, the performance of the Molcode model gave, respectively, 0.79 and 0.725 for the training and test set; a definition of applicability domain was also provided.

These software are not built using only pharmaceutical compounds. However, they can be used for the prediction of LOAEL for drugs.

4 Uncertainty of LOAEL and NOAEL Data

The development of non-animal testing for RDT is difficult mainly because the complex underlying processes, which include effects on different organs and tissues and different time scales [2]. NOAEL and LOAEL have been criticized as conceptually inappropriate for providing quantitative estimates for toxicity, and it has been proposed to replace them with the benchmark dose [37].

Besides the fact that many organs and tissues are involved, other aspects make the NOAEL and LOAEL data uncertain. NOAEL and LOAEL are not derived or calculated from a dose-effect curve but can only be identified from the doses. This means that they both depend on the study design, particularly the spaces between doses. Consequently, different NOAEL and LOAEL

values may be obtained for the same substance using different study designs or different exposure doses. There is a further intrinsic uncertainty in LOAEL experimental data. The "true" LOAEL (the real dose of the substance that starts to generate an effect) may be anywhere between the NOAEL and the LOAEL.

This uncertainly is probably big, but how big cannot be measured. This is another problem of the NOAEL and LOAEL approach, as in risk assessment quantifying the uncertainties involved is crucial for establishing protective human exposure limits [38]. The variability of the responses between animals in the dose groups, the definition of the "adversity" of an effect, and the statistical methods supporting this definition are other aspects that raise the level of uncertainty of NOAEL and LOAEL [39].

5 Conclusion

The NOAEL and LOAEL of substances are required for human health hazard assessments under different regulatory contexts (pharmaceutical, biocides, REACH, cosmetics) [2]. In vivo RDT studies are very expensive and time-consuming and involve a large number of animals. In vivo RDT has been banned for the safety assessment of cosmetics [1], and REACH legislation [10] requires to use as few animals as possible to evaluate the toxicity of substances. Therefore, there is a pressing need to find a valid alternative.

However, considering the uncertainty of NOAEL and LOAEL values, the in silico models are extremely complex because all this uncertainty will be implicitly transferred into the data predicted by a model. Moreover, considering the QSAR approach, there is a no solid mechanistic basis to support the statistical association between a set of molecular descriptors and the systemic effects [2].

Despite the limitations of each single alternative approach, the combination and interpretation of data from different alternative techniques, such as QSARs, physiologically based pharmacokinetic modeling (PBPK), read-across, threshold of toxicological concern (TTC), and in vitro methods, may be useful to gain more reliable predictions of NOAEL and LOAEL.

Acknowledgments

This work was supported by the project HEALTH-F5-2010-267042 ToxBank (supporting integrated data analysis and servicing of alternative testing methods in toxicology) funded by the European Commission and Cosmetics Europe under the Seventh Framework Programme.

References

1. Scientific Committee on Consumer Safety (SCCS) (2012) The SCCS's notes of guidance for the testing of cosmetic ingredients and their safety evaluation. http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/sccs_s_006.pdf. Accessed 08 June 2015
2. Worth A, Barroso J, Bremer S et al (2014) Alternative methods for regulatory toxicology-a state-of the-art review. JRC Science and Policy reports. Report EUR 26797 EN
3. Steinmetz KL, Spack EG (2009) The basics of preclinical drug development for neurodegenerative disease indications. Neurology 9:S2
4. Marchetti S, Schellens JHM (2007) The impact of FDA and EMEA guidelines on drug development in relation to Phase 0 trials. Br J Cancer 97:577–581
5. Dorado MA, Engelhardt JA (2005) The no-observed adverse level in drug safety evaluations: use, issue, and definition(s). Regul Toxicol Pharmacol 42:265–274
6. Bakhtiar R (2008) Biomarkers in drug discovery and development. J Pharmacol Toxicol Methods 57:85–91
7. Food and Drug Administration (FDA), Department of Health and Human Services, Center for Drug Evaluation and Research (CDER) (2005) Guidance for Industry. Estimating the Maximum Safe Starting Dose in the initial clinical trials for therapeutics in adult healthy volunteers
8. European Medicines Agency (2010) Guideline on repeated dose toxicity. Committee for Human Medicinal Products. Reference number CPMP/SWP/1042/99 Rev
9. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2010) Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals M3(R2). http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002941.pdf. Accessed 06 June 2015
10. European Commission (2006) Regulation (EC) No 1907/2006 of the European Parliament and the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off J Eur Union L396:1–849
11. European Commission (2009) Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Off J Eur Union L309:1–47
12. European Commission (2013) Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market.
13. Union E (2012) Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products. Off J Eur Union L167:1–116
14. European Commission (2009) Regulation (EC) No 1223/2009 of the European Parliament and of the Council of 30 November 2009 on cosmetic products
15. Dearden JC (2003) *In silico* prediction of drug toxicity. J Comput Aid Mol Des 17:119–127
16. Tsakovska I, Lessigiarska I, Netzeva T, Worth AP (2007) A mini review of mammalian toxicity (Q)SAR models. QSAR Comb Sci 27:41–48
17. Przybylak KR, Madden JC, Cronin MTD et al (2012) Assessing toxicological data quality: basic principles, existing schemes and current limitations. SAR QSAR Environ Res 23:435–459
18. Bitsch A, Jacobi S, Melber C et al (2006) RepDose: a database on repeated dose toxicity studies of commercial chemicals—a multifunctional tool. Regul Toxicol Pharmacol 46:202–210
19. Munro IC, Ford RA, Kennepohl E et al (1996) Correlation of structural class with no-observed-effect levels: a proposal for establishing a threshold of concern. Food Chem Toxicol 34:829–867
20. Cramer GM, Ford RA, Hall RL (1978) Estimation of toxic hazard—a decision tree approach (and errata sheet). Food Cosmet Toxicol 16:255–276
21. Hayashi M, Sakuratani Y (2011) Development of an evaluation support system for estimating repeated dose toxicity of chemicals based on chemical structure. In: Wilson AGE (ed) New Horizons in predictive toxicology. Current status and application. RSC Publishing, Cambridge, UK

22. Persad AS, Cooper GS (2008) Use of epidemiologic data in Integrated Risk Information System (IRIS) assessments. Toxicol Appl Pharmacol 233:137–145
23. Anzali S, Berthold MR, Fioravanzo E et al (2012) Development of computational models for the risk assessment of cosmetic ingredients. IFSCC Mag 15:249–255
24. Martin MT, Judson RS, Reif DM et al (2009) Profiling chemicals based on chronic toxicity results from the U.S. EPA ToxRef Database. Environ Health Perspect 117:392–399
25. Toropov AA, Toropova AP, Pizzo F et al (2015) CORAL: model for no observed adverse effect level (NOAEL). Mol Divers 19:563–575. doi:10.1007/s11030-015-9587-1
26. Gadaleta D, Pizzo F, Lombardo A et al (2014) A *k*-NN algorithm for predicting OralSub-chronic toxicity in the rat. ALTEX 31:423–432
27. Toropova AP, Toropov A, Veselinović JB et al (2014) QSAR as a random event. Environ Sci Pollut Res Int 22:8264–8271. doi:10.1007/s11356-014-3977-2
28. Sakuratani Y, Zhang H, Nishikawa S et al (2013) Hazard evaluation support system (HESS) for predicting repeated dose toxicity using toxicological categories. SAR QSAR Environ Res 24:351–363
29. Mazzatorta P, Estevez MD, Coulet M et al (2008) Modeling oral rat chronic toxicity. J Chem Inf Model 48:1949–1954
30. Julián-Ortiz JV, García-Domenech R, Gálvez J et al (2005) Predictability and prediction of lowest observed adverse effect levels in a structurally heterogeneous set of chemicals. SAR QSAR Environ Res 16:263–272
31. García-Domenech R, de Julián-Ortiz JV, Besalú E (2006) True prediction of lowest observed adverse effect levels. Mol Divers 10:159–168
32. Matthews EJ, Kruhlak NL, Benz RD et al (2004) Assessment of the health effects of chemicals in humans: I. QSAR estimation of the Maximum Recommended Therapeutic Dose (MRTD) and No Effect Level (NOEL) of organic chemicals based on clinical trial data 1. Curr Drug Discov Technol 1:61–76
33. Contrera JF, Matthews EJ, Kruhlak NL et al (2004) Estimating the safe starting dose in phase I clinical trials and no observed effect level based on QSAR modeling of the human maximum recommended daily dose. Regul Toxicol Pharmacol 40:185–206
34. Mumtaz MM, Knau LA, Reisman DJ et al (1995) Assessment of effect levels of chemicals from quantitative structure-activity relationship (QSAR) models. I. Chronic lowest-observed-adverse-effect level (LOAEL). Toxicol Lett 79:131–143
35. Venkatapathy R, Moudgal CJ, Bruce RM (2004) Assessment of the oral rat chronic lowest observed adverse effect level model in TOPKAT, a QSAR software package for toxicity prediction. J Chem Inf Comput Sci 44:1623–1629
36. Tilaoui L, Schilter B, Tran LA, Mazzatorta P et al (2006) Integrated computational methods for prediction of the lowest observable adverse effect level of food-borne molecules. QSAR Comb Sci 26:102–108
37. Sand S, Victorin K, Filipsson AF (2008) The current state of knowledge on the use of the benchmark dose concept in risk assessment. J Appl Toxicol 28:405–421
38. Vermeire TG, Baars AJ, Bessems JGM et al (2007) Toxicity testing for human health risk assessment. In: van Leeuwen CJ, Vermeire TG (eds) Risk assessment of chemicals, an introduction, 2nd edn. Springer, The Netherlands
39. Paparella M, Daneshian M, Hornek-Gausterer R et al (2013) Food for thought…uncertainty of testing methods-what do we (want to) know? ALTEX 30:131–144

Chapter 10

In Silico Models for Acute Systemic Toxicity

Julien Burton, Andrew P. Worth, Ivanka Tsakovska, and Antonia Diukendjieva

Abstract

In this chapter, we give an overview of the regulatory requirements for acute systemic toxicity information in the European Union, and we review the availability of structure-based computational models that are available and potentially useful in the assessment of acute systemic toxicity. The most recently published literature models for acute systemic toxicity are also discussed, and perspectives for future developments in this field are offered.

Key words Acute systemic toxicity, Regulation, Organ-specific toxicity, In silico model

1 Introduction

Acute systemic toxicity comprises the general adverse effects that occur after a single or multiple exposure of an animal to a substance within 24 h and during an observation period of at least 14 days. The substance may be administered orally, by inhalation, or dermally.

Acute mammalian toxicity tests are often the first in vivo toxicity tests to be performed on a chemical. In recent years there have been considerable efforts to replace, reduce, or refine these animal tests by applying alternative approaches, including both in vitro and in silico models. An increasing number of models are available to predict acute mammalian toxicity. This is partly due to the fact that a reasonable number of datasets are openly available for modeling. However, the reliability of the in vivo data can be highly variable, and the metadata provided is often insufficient to determine the suitability of the data for modeling purposes. Another challenge is related to the multiple mechanisms leading to this complex effect, which is typically expressed as a single numerical value (LD_{50} for oral and dermal toxicity; LC_{50} for inhalational toxicity). In addition there are also differences between

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_10, © Springer Science+Business Media New York 2016

the routes of administration and species, and different data should be modeled separately [1].

Target organs, such as the liver, kidneys, heart, lungs, and brain, can be affected by exogenous chemicals to the extent that they cease to function. Thus, the use of QSAR models for organ/system specific toxicity would be extremely helpful when predicting acute systemic toxicity. A limited number of QSAR models for specific target organ and tissue effects are available.

The information obtained from acute systemic toxicity studies is used in the hazard assessment of chemicals occurring in food, industrial chemicals, biocides, pesticides, and cosmetics. In this chapter, we give an overview of the regulatory requirements for acute systemic toxicity information in the European Union, the software packages available for assessment of acute systemic toxicity and organ- and system-specific toxicity, as well as the databases available for obtaining such data. Since comprehensive reviews of literature QSAR studies are available elsewhere [2–5], we focus here on some of the more recently published literature models for acute systemic toxicity. Some of these software and literature models are documented in the JRC's QSAR Model Database (http://qsardb.jrc.ec.europa.eu/qmrf/).

2 Regulatory Context in the European Union

For the assessment of acute systemic toxicity, only in vivo tests are currently accepted by regulatory bodies (Table 1). However, in vivo acute systemic toxicity studies are prohibited for cosmetic substances and products [14].

The endpoint measured in the majority of the standard assays is animal morbidity or death. Evident signs of toxicity (i.e., clear signs of toxicity indicating that exposure to the next highest concentration would cause severe toxicity in most animals within the observation period) are only used in the oral fixed dose procedure (FDP), which causes less suffering and is, therefore, more humane.

Table 1
In vivo methods currently available for acute systemic toxicity

Exposure route	OECD	EU test method
Oral	TG 420: fixed dose procedure [6]	B.1 bis [7]
	TG 423: acute toxic class method [8]	
	TG 425: up and down procedure [9]	B.1 tris [7]
Dermal	TG 402 [10]	B.3 [7]
Inhalation	TG 403 [11]	B.2 [12]
	TG 436 (acute toxic class method) [13]	B.52 [12]

The assessment of acute systemic toxicity is one component in the safety evaluation of substances and represents a standard information requirement within several pieces of EU chemicals legislation, including the Regulation on Classification, Labelling and Packaging (CLP) of substances and mixtures [15], the Regulation concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) [16], the Biocidal Products Regulation [17], the Plant Protection Products Regulation [18], and the Cosmetic Products Regulation [14]. In preclinical drug development [19], however, these studies are no longer required to support first clinical trials in man. The information needed can be obtained from appropriately conducted dose-escalation studies or short-duration dose ranging studies that define a maximum tolerated dose in the general toxicity test species [20, 21]. Further information on the regulatory requirements in the EU is given in Prieto et al. [22].

3 Software for Predicting Acute Systemic Toxicity

Several software tools capable of predicting endpoints related to systemic toxicity are available, as reviewed previously [23]. An updated list is given in Table 2 and some updates on the programs are described below.

Among the commercial software programs covering a broad spectrum of systemic toxicological effects is ACD/Labs Percepta, which is developed and marketed by Advanced Chemistry development Inc. (http://www.acdlabs.com/). The platform has two modules related to systemic toxicity prediction—Acute Toxicity Prediction Module and Health Effects Prediction Module. The Acute Toxicity predictor has been built using experimental data for more than 100,000 compounds extracted from the Registry of Toxic Effects of Chemical Substances (RTECS) and former European Chemical Substances Information System (ESIS) databases. It provides three different software components related to acute mammalian toxicity:

- LD_{50}—Provides predictions of LD_{50} values for rats and mice according to various routes of administration. Prior to modeling, the original experimental data were converted to logarithmic form (pLD_{50}) in order to maintain linear relationship with used descriptors. The final prediction results returned to the user are converted back to LD_{50} values (mg/kg). The predictive model for pLD_{50} has been derived using GALAS (Global, Adjusted Locally According to Similarity) modeling methodology.
- Hazards—A knowledge-based expert system that identifies and visualizes hazardous structural fragments.

t2.1 **Table 2**

t2.2 **Software tools for systemic toxicity endpoints**

t2.3 Software (and developer)	Availability	Endpoint							
		Acute (oral) toxicity	Chronic (oral) toxicity	Hepatotoxicity	Nephrotoxicity (+ urinary tract toxicity)	Neurotoxicity	Cardiotoxicity	Immunotoxicity[a]	Cytotoxicity
t2.4 t2.5 Accelrys Discovery Studio, including TOPKAT (BIOVIA)	Commercial	●	●	●	●		●		
t2.6 t2.7 ACD/Percepta (Advanced Chemistry Development, Inc.)	Commercial	●		●	●		●		
t2.8 t2.9 ADMET Predictor (Simulations Plus Inc.)	Commercial	●		●			●	●	
t2.10 t2.11 t2.12 t2.13 admetSAR (Laboratory of Molecular Modeling and Design, East China University of Science and Technology)	Freely available	●							
t2.14 CASE Ultra (MultiCASE Inc.)	Commercial	●	●	●	●				
t2.15 Derek Nexus (Lhasa Ltd)	Commercial			●	●	●	●	●	
t2.16 t2.17 Lazar (In silico toxicology GmbH)	Freely available		●						
t2.18 Leadscope (Leadscope)	Commercial			●	●	●	●		
t2.19 t2.20 MetaDrug/ToxHunter™ (Thomson Reuters)	Commercial			●	●				
t2.21 t2.22 Molcode Toolbox (Molcode Ltd)	Commercial	●	●				●		●

t2.23 t2.24 t2.25	OpenVirtualToxLab (Biographics Laboratory 3R)	Freely available for academic organizations			●			●	
t2.26 t2.27 t2.28	Pallas System including ToxAlert and HazardExpert Pro (CompuDrug Inc.)	Commercial					●		●	●
t2.29	PASS (geneXplain GmbH)	Commercial	●		●				
t2.30 t2.31 t2.32 t2.33	Pred-hERG 2.0 (Laboratory for Molecular Modeling and Drug Design, Federal University of Goiás.)	Freely available						●	
t2.34 t2.35 t2.36	PROTOX (Charite University of Medicine Institute for Physiology)	Freely available	●						
t2.37	TerraQSAR (TerraBase)	Commercial	●						
t2.38	T.E.S.T. (US EPA)	Freely available	●						
t2.39 t2.40 t2.41 t2.42	TIMES (Laboratory of Mathematical Chemistry, University "Prof. Dr. Asen Zlatarov")	Commercial	●		●				
t2.43 t2.44 t2.45	Tox-Comp.net (Faculty of Pharmacy, Jagiellonian University Medical College)	Freely available						●	

t2.46 [a]Immunotoxicity other than skin sensitization

- Categories—Classifies compounds into one of five Globally Harmonised System (GHS) categories for acute oral toxicity.

The Health Effects module predicts the probability of a compound having a health effect on a particular organ or organ system (blood, cardiovascular system, gastrointestinal system, kidney, liver, and lungs). The models are based on data collected from chronic, sub-chronic, acute toxicity and carcinogenicity studies with adverse effects reported in particular organs or organ systems.

A common goal of toxicity prediction is to distinguish between toxicologically active and inactive compounds. Since multiple mechanisms are involved in systemic toxicity, this requires the availability of predictive tools that are able to cover a wide region of the activity space. This is the main feature of the expert systems that make assessments on the basis of structural alerts covering a spectrum of structural properties associated with the complex endpoint. One commonly used expert system, developed and marketed by Lhasa Ltd (http://www.lhasalimited.org/), is Derek Nexus which is a development of the former Derek for Windows (DfW). This contains knowledge rules derived from the known relationship between a given substructure and a toxicological effect of the molecule and applies these rules to predict potential toxicological effects of compounds. Derek Nexus generates a prediction by comparing the structural features of the target compound with a toxicophore encoded as structural pattern(s) in its knowledge base. The final predictions are derived from a reasoning scheme which takes into account the presence of a toxicophore in the query structure ('structural alert') and a limited number of calculated molecular properties, which, taken together, return an "uncertainty term" for the prediction itself. For some alerts, supporting examples are provided and the system states whether the query compound already exists as an example in the knowledge base. Literature references are also included to enable the user to assess the applicability of the structural alert to the predicted structure and to allow for an expert knowledge assessment. Derek Nexus covers multiple endpoints, including hepatotoxicity, nephrotoxicity, and cardiotoxicity.

CASE Ultra (http://www.multicase.com/) is further development of MCASE methodology and falls in the range of fragment based QSAR expert systems [24]. The CASE Ultra model mainly consists of a set of "positive alerts" (biophores), and "deactivating alerts" (biophobes), i.e., those fragments that are identified as statistically significant for increasing/decreasing the activity. The improvement of CASE Ultra over its predecessor is related to the identified alerts that are no longer limited to linear paths of limited size or limited branching pattern. In addition the training sets can be larger than 8000 molecules. The applicability domains

of individual toxicity alerts within the models quantitatively define the necessary structural environment of the toxicity alerts.

The statistically based program TOPKAT (http://accelrys.com/) uses multiple QSARs on small and homogenous sets of data. It is now a part of QSAR, ADMET and Predictive Toxicology module within Biovia Discovery Studio platform. The rat oral LD_{50} module in TOPKAT comprises 19 regression analyses developed using experimental values of approx. 4000 chemicals from RTECS, including pesticides and industrial chemicals. The rat oral LD_{50} module in MCASE (named A56) is based on and comprises data for 7920 chemicals from the FDA, WHO, and NTP datasets. Tunkel and coworkers [25] compared the performance of the TOPKAT and MCASE rat LD_{50} modules against an external test set of 73 organic compounds covering 32 chemical categories retrieved from submissions to the EPA High Production Volume (HPV) Challenge Program (http://www.epa.gov/chemrtk/). The predictive accuracy of each software tool was assessed by applying the EPA's New Chemical classification approach (http://www.epa.gov/oppt/newchems/index.htm), from the low-concern class (>2000 mg/kg) to the high-concern class (<15 mg/kg). While neither model was able to classify all 73 compounds, TOPKAT correctly classified 67 % of the chemicals, while MCASE classified 70 % correctly. However, it should be noted that the test set used was significantly skewed toward "low concern" chemicals, which both models predicted correctly with a high degree of accuracy (82 % and 100 % correct for TOPKAT and MCASE, respectively). Moreover, a high degree of false negatives was found for moderate and high concern HPV chemicals (TOPKAT, 72 %; MCASE, 100 %), suggesting that these programs are less reliable for the identification of more toxic compounds. The authors also compared the model outputs against the GHS five-tier scheme for classification of rat oral acute toxicants (<5, 5–50, 50–300, 300–2000, and 2000–5000 mg/kg), which is similar to the one adopted by EPA (<15, 15–50, 50–500, 500–2000, >2000 mg/kg). When compared against the GHS scheme, the ability of TOPKAT and MCASE to produce correct classifications was 73 % and 70 %, respectively, for the HPV test set chemicals, thereby changing slightly with respect to the EPA scheme, albeit enough to invert the rank order of these models.

VirtualToxLab is an in silico technology for estimating the toxic potential of chemicals [26] based on an automated protocol that simulates and quantifies the binding of small molecules towards a series of proteins, known or suspected to trigger adverse effects. The interface to the technology allows building and uploading molecular structures, viewing and downloading results and rationalizing any prediction at the atomic level by interactively analyzing the 3D binding mode of a compound with its target protein(s) in real-time. The VirtualToxLab has been used to

predict the toxic potential for over 2500 compounds and the free platform, OpenVirtualToxLab, is accessible (in client-server mode) over the Internet. It is free of charge for universities, governmental agencies, regulatory bodies, and nonprofit organizations.

The LeadScope software (http://www.leadscope.com) links chemical and biological data that allows exploration of large sets of chemical compounds, their properties, and biological activities. Chemical structures are organized in a taxonomy of familiar structural features each combined with common substituents—the common building blocks of medicinal chemistry [27]. LeadScope provides QSAR models for diverse physiological adverse effects including cardiological, hepatobiliary, and urinary endpoints.

Other software tools available for predicting acute toxicity (LD_{50}) to rat/mouse are also available, such as TerraQSAR (http://www.terrabase-inc.com/), ADMET Predictor (http://www.simulations-plus.com), Molcode Toolbox (http://molcode.com/). The TerraQSAR software is based on neural network methodology and includes models for predicting both oral and intravenous LD_{50} values in mice and rats. ADMET Predictor includes a number of in-built models for ADMET, and allows new predictive models to be built from the user's data. ADMET Predictor's Toxicity Module provides predictions of various toxicity endpoints including hepatotoxicity, carcinogenicity, acute rat toxicity, and cardiotoxicity. Molcode Toolbox has a range of modules for predicting toxicological endpoints, including intravenous acute LD_{50} values and in vitro cytotoxicity (IC_{50} values) (from the Registry of Cytotoxicity). The models are well documented and the underlying experimental data is made available with references and structure files (MDL molfiles).

4 Databases Containing Information on Acute Systemic Toxicity

Sources of rat LD_{50} values which may be suitable for the development of QSARs are listed in Table 3. Some recent updates are discussed in the section below.

In particular, Acutoxbase [29] was developed in the context of the EU FP6 project 'A-Cute-Tox' (http://www.acutetox.eu/), which aimed to optimize and "pre-validate" an in vitro testing strategy for predicting acute human toxicity ([30, 31]; Prieto and Kinsner-Ovaskainen 2015). While the database is not available, the in vitro and animal data are published in several publications [30–32].

Recently the COSMOS database has been developed as a part of the COSMOS project (http://www.cosmostox.eu/), one of seven projects forming the Seurat-1 research cluster (http://www.seurat-1.eu/). Version 1 of the COSMOS database (http://cosmosdb.cosmostox.eu/) contains 12,538 toxicity studies for

1660 compounds across 27 endpoints, including acute toxicity data for 1697 compounds tested on different animal species, as well as in vitro data.

The Hazardous Substances Data Bank (HSDB) is a part of NLM's Toxicology Data Network (TOXNET) [33]. It contains chemical substance information with one record for each specific chemical or substance, or for category of chemicals or substances. HSDB has approximately 5600 chemicals and substances, with information for toxicity and human exposure. All data comes from public scientific sources. HSDB's content is peer-reviewed by a group of experts.

The Registry of Toxic Effects of Chemical Substances (RTECS) database includes basic toxicity information for: prescription and non-prescription drugs, food additives, pesticides, fungicides, herbicides, solvents, diluents, chemical wastes, reaction products of chemical waste, and substances used in industrial and household situations. It covers six categories of toxicity data including acute toxicity data. In vitro toxicology data has been added as well. Accelrys now produces the RTECS files using existing data selection criteria and rules established by NIOSH (http://accelrys.com/products/databases).

In order to be useful for QSAR development, datasets should be first curated, i.e., the accuracy of the structures should be verified and the quality of biological data should be reviewed. It is useful to provide a reference to the source of the experimental data. In addition, inorganic and organometallic compounds, salts, and compound mixtures are often removed from the analysis. For the development of QSARs, LD_{50} values should be converted to log[1/(mol/kg)] (if originally expressed as mol/kg or mg/kg). Finally, approximate LD_{50} values should be converted to discrete values, and multiple LD_{50} values from different labs/experiments should be converted to a single value. The ChemIDplus and ZEBET databases have been recently employed as data sources for QSAR analyses [34, 35].

5 Prediction of Organ-Specific and System-Specific Toxicity

5.1 Ability to Predict In Vivo Toxicity

Some currently available software tools (e.g., TOPKAT and MCASE) are useful for predicting acute toxicity in categorical terms (e.g., in terms of GHS classifications). The performance of different software tools in predicting acute toxicity has been investigated [36, 37]. In these studies, ACD and T.E.S.T. have performed well.

In the scientific literature, local QSAR models have been generated for sets of congeneric compounds (organophosphates, aromatic amines, anilines, etc.) and are scattered over many original publications. Some of these studies have also explored the use of

in vitro data as additional descriptors in the derivation of so-called quantitative structure activity-activity relationships, QSAAR [38]. QSAAR modeling revealed good potential for acute toxicity prediction, particularly in cases when a significant correlation exists between in vivo data (LD_{50}) and in vitro cytotoxicity (IC_{50}), and the additional inclusion of physicochemical parameters serves to improve the correlation. In practical terms, QSAAR could be particularly useful if high-throughput screening methods are used to generate the in vitro data.

Despite their limited applicability when taken individually, local QSAR models might be usefully combined into an expert system for toxicity predictions. As a part of the efforts to develop global QSAR models for acute toxicity Raevsky and coworkers [39] proposed the so-called Arithmetic Mean Toxicity (AMT) modeling approach, which produces local models based on a k-nearest neighbors approach. Arithmetic mean toxicity values of one or more pairs of analogues (nearest neighbors) are considered as the toxicity of the chemical of interest. Recently a classification model based on 436 Munro database chemicals and developed using Dragon descriptors has been proposed as a tool for chemical screening [40]. Kleandrova et al. [3] have developed a multitasking (mtk) QSTR model based on ANN (artificial neural networks) for simultaneous prediction of acute toxicity by considering different routes of administration, different breeds of laboratory animals, and the reliability of the experimental conditions. The model is based on a diverse dataset comprising 1494 chemicals retrieved from CHEMBL (http://www.ebi.ac.uk/chembldb).

A consensus approach has been exploited in some studies where the models are built by using a combinatorial QSAR modeling approach, including multiple descriptors and employing several statistical modeling methods. It has been claimed that the predictive accuracy of consensus QSAR models is superior to the individual ones [34, 41]. In addition, several research studies [35, 42, 43] have demonstrated the ability to improve quantitative predictions for structurally diverse datasets when high throughput bioactivity data are used in combination with traditional molecular descriptors. This can also be regarded as an example of the QSAAR approach. These hybrid approaches and their underlying datasets are publicly available via the ChemBench web portal (https://chembench.mml.unc.edu/).

5.2 Ability to Predict Non-apical Toxicities

The feasibility of using in vitro cytotoxicity data for the prediction of in vivo acute toxicity has been investigated in a number of research programs [28, 44, 45]. Over 70 % correlation has been established between in vitro basal cytotoxicity and rodent LD_{50} values [46]. The applicability of 3T3 Neutral Red Uptake Cytotoxicity Assay for the identification of substances with an $LD_{50} > 2000$ mg/kg has been evaluated by the EURL ECVAM Scientific Advisory

Table 3
Databases containing acute toxicity information

Database	Availability	Information
Acutoxbase, linked to the EU FP6 project 'A-Cute-Tox'; http://www.acutetox.eu/	Database not available, but the data are included in several publications (see text)	The following data are available for 97 reference chemicals (i.e., 52 % drugs, 31 % industrial chemicals, 12 % pesticides, 5 % others): • In vitro: approx. 100 in vitro assays including general acute cytotoxicity, metabolism-mediated toxicity, biokinetics, and organ-specific toxicity. • In vivo: Over 2200 LD_{50} values in rodents (rat and mouse) and other animals (e.g., guinea pig, dog) with various administration routes (oral, intravenous, etc.) compiled from published literature. For 97 reference chemicals, nearly 2800 human acute poisoning cases from clinical/forensic reports are also available.
COSMOS Database; http://cosmosdb.cosmostox.eu/	Freely available through the Internet after registration	Includes US FDA PAFA acute toxicity data.
CEBS, developed by the US NIEHS; http://cebs.niehs.nih.gov/	Freely available through the Internet	Includes in vivo study data and acute dose of a small number of known hepatotoxicants to rat.
ChemIDplus, developed by the US NLM; http://chem.sis.nlm.nih.gov/chemidplus/	Freely available through the Internet, structure-searchable	Toxicity data is available for over 400,000 chemical records, of which over 300,000 include chemical structures that are retrieved from TOXNET® (TOXicology Data NETwork; http://toxnet.nlm.nih.gov). It includes HSDB (Hazardous Substances Data Bank, an older subset of the RTECS database). A search for rat and mouse oral LD_{50} values found 15,866 and 33,009 records, respectively.
Food Safety Acute Toxicity Database; https://www.leadscope.com/toxicity_databases/regulatory_databases/	Commercial	Contains acute oral toxicity (LD_{50}) data from US FDA CFSAN PAFA database for1070 food additives and 1633 tests. Test systems include mainly • Rats: 950 chemicals • Mice: 366 chemicals Other test systems include rabbits, guinea pigs, dogs, and monkey.

(continued)

Table 3
(continued)

Database	Availability	Information
RTECS, originally compiled and maintained (until 2001) by the US NIOSH and currently maintained by Accelrys Technologies. Structure-searchable through the Accelrys Toxicity Database; http://accelrys.com/products/databases/bioactivity/toxicity.html Also searchable via other databases including the Leadscope Toxicity Database; http://www.leadscope.com/databases/	Commercial	Rat acute oral toxicity (LD_{50}) and acute inhalation toxicity (LC_{50}) data are compiled from the open scientific literature for approx. 7000 compounds (organic, inorganic and mixtures), including approx. 4000 organic compounds.
HSDB—TOXNET database; http://toxnet.nlm.nih.gov	Freely available through the internet	Toxicology database that focuses on potentially hazardous chemicals. Contains nonhuman toxicity values for almost 3000 chemicals.
Registry of Cytotoxicity (RC) database	Freely available on request from BfR ZEBET (zebet@bfr.bund.de)	Based on the publication by Halle [28], this comprises rodent acute oral LD50 values and published IC50 values from diverse in vitro cytotoxicity assays on approximately 550 chemicals

CEBS chemical effects in biological systems, *HSDB* Hazardous Substances Data Bank, *RTECS* registry of toxic effects of chemical substances; *TOXNET NLM's* Toxicology Data Network, *US NLM* US National Library of Medicine, *US NIEHS* US National Institute of Environmental Health Sciences, *US NIOSH* US National Institute of Occupational Safety and Health, *BfR ZEBET* Centre for Documentation and Evaluation of Alternatives to Animal Experiments of the German Federal Institute for Risk Assessment

Committee (ESAC). It was recommended however that the results should always be used in combination with other information sources. For instance, the assay is recommended as a component of an Integrated Approach to Testing and Assessment (IATA) [47]. A reason for the absence of a clear relationship between basal cytotoxicity and in vivo acute toxicity could be that specific organ toxicity is the most sensitive parameter for acute toxicity. Common sensitive systems and organs include nervous, cardiovascular, immune system, kidneys and liver, lungs and blood. IATA proposed for acute systemic toxicity are a combination of complementary approaches (in vitro, ex vivo, in silico, in chemico) that address functional mechanistic endpoints tied to adverse outcomes of regulatory concern [48].

As summarized in Table 4, there is a limited number of literature models for predicting toxicities at tissue and organ levels. A list of software applications is provided in Table 2. They are based on expert system or regression/categorical QSAR models. In the case of ligand–protein interactions, molecular modeling approaches are mainly used. Among the commonly used software tools, Derek Nexus provides over 500 structural alerts for a range of organ and system-specific toxicities, and other miscellaneous endpoints. Models for predicting liver toxicity are further covered in Chapter 11 (Hewitt et al.).

Some of these models are based on the concept of reactivity-based toxicity. The covalent binding of reactive electrophiles to cellular targets (i.e., nucleophilic sites of macromolecules) has the potential to initiate a chain of biological effects (e.g., depletion of glutathione and protein thiols) resulting in specific organ and system toxicities.

Among the few comprehensive studies covering a range of organ toxicities and relying on a broad structural space in the training set are the models published by Matthews et al. [49]. These models were developed for urinary tract toxicities of drugs. For each organ, a number of toxicity endpoints were considered in the QSAR analysis. The investigation utilizes four software programs: MC4PC (versions 1.5 and 1.7); BioEpisteme (version 2.0); MDL-QSAR (version 2.2); Leadscope Predictive Data Miner (LPDM version 2.4). The four QSAR programs were demonstrated to be complementary and enhanced performance was obtained by combining predictions from two programs. The best QSAR models exhibited an overall average 92 % coverage, 87 % specificity, and 39 % sensitivity. These results support the view that a consensus prediction strategy provides a means of optimizing predictive ability.

In the work of Myshkin et al. [51], a detailed ontology of toxic pathologies for 19 organs was created from the literature in a consistent way to capture precise organ toxicity associations of drugs, industrial, environmental, and other compounds. Models for nephrotoxicity and for more specific endpoints related to these organ injuries were developed using a recursive partitioning algorithm. The models performed better at the prediction of distinct organ toxicity subcategories than general organ toxicity, reflecting the well-known tendency of QSAR models to have a better predictive performance for more specific endpoints.

In a more recent study, Lee et al. [50] present QSAR models for three common patterns of drug-induced kidney injury, i.e., tubular necrosis, interstitial nephritis, and tubulo-interstitial nephritis. Binary classification models of nephrotoxin versus non-nephrotoxin with eight fingerprint descriptors were developed based on heterogeneous pharmacological compounds data. Two types of data sets were used for construction of the training set, i.e., parent compounds of pharmaceuticals (251 nephrotoxins and

t4.1 **Table 4**

t4.2 **Overview of published organ and system-specific toxicity QSAR models**

t4.3 t4.4 Model, reference	Endpoint	Statistical method/software	Statistical parameters	Training set	Test set	Class(es) studied	Significant parameters	Note
t4.5 t4.6 t4.7 t4.8 t4.9 t4.10 Urinary tract toxicity [49]	Six types of urinary tract injury (acute renal disorders, nephropathies, bladder disorders, kidney function tests, blood in urine, urolithiases)	Software programs: MC4PC, BioEpisteme, MDL-QSAR Leadscope Predictive Data Miner	Best consensus models: sensitivity 56 %, specificity 78 %	≈1600	n/r	Multiple	n/r	Consensus models based on two programs increased sensitivity to 56 %
t4.11 t4.12 t4.13 t4.14 t4.15 t4.16 t4.17 t4.18 Nephrotoxicity [50]	Tubular necrosis (TN), interstitial nephritis (IN), and tubulointerstitial nephritis (TIN)	SVM	Best parent compounds-based TIN Model: CA = 0.80 and MCC = 0.32 Best Metabolite-based Models: CAs—0.84, 0.85, and 0.83; MCCs—0.69, 0.69, and 0.62 for TN, NI, and TIN models, respectively	487 parent compounds/624 metabolites	338 parent compounds/156 metabolites	Multiple	Topological fingerprints implemented in PaDEL-Descriptor	Metabolite sets consist of major urinary metabolites of pharmaceuticals in parent compound sets
t4.19 t4.20 t4.21 t4.22 Nephrotoxicity [51]	Nephrotoxicity, kidney necrosis, kidney relative weight gain, nephron injury	Recursive partitioning algorithm, as (ChemTree™ software)	Sensitivity and specificity above 90 %	172–847 depending on the model	42–154 depending on the model	Multiple	Two-dimensional structural descriptors	Models are available in the MetaDrug/ ToxHunterTM systems pharmacology suite.
t4.23 t4.24 t4.25 Nephrotoxicity, hematotoxicity [52]	Renal tubular necrosis, hemolytic anemia			16		Derivatives of 1,2-and 1,4-naphthoquinone	Structural alerts	SAR analysis, outlining important structural alerts
t4.26 t4.27 t4.28 t4.29 t4.30 t4.31 t4.32 t4.33 Nephrotoxicity [53]	cGST and MGST1 enzyme activity	LR	r^2 (specific activities for MGST1-catalyzed reactions) = 0.943	9	n/r	Haloalkenes	E_{LUMO}	The relation between nephrotoxicity of the haloalkanes and E_{LUMO}, reflect their propensity for conjugation reactions catalyzed by glutathione transferase enzymes
t4.34 t4.35 t4.36 t4.37 t4.38 t4.39 t4.40 t4.41 t4.42 Acute and delayed neurotoxicity [54]	Inhibitory activity and pairwise selectivity toward serine esterases including acetylcholinesterase and neuropathy target esterase	MLR (Hansch analysis), MFTA	Hansch analysis best models: $r = 0.699–0.993$ MFTA best models: $r^2 = 0.57–0.96$; $q^2 = 0.47–0.91$	Hansch analysis: 7/9 MFTA: 18–52	n/r	Organophosphorus compounds	Hansch analysis: hydrophobicity of substituent R in the general formula (RO)2P(O)X MFTA: effective charge, Q, on an atom; effective van der Waals radius, Re; group lipophilicity, Lg	

Acute and delayed neurotoxicity [55]	Inhibitory activity and pairwise selectivity toward serine esterases including acetylcholinesterase and neuropathy target esterase	FSMLR, BPNN, MFTA, CoMSIA	FSMLR models: Q^2_{DCV} range 0.180–0.778. BPNN models: Q^2_{DCV} range 0.601–0.800 MFTA models: r^2 range 0.62–0.96 CoMSIA models: r^2 range 0.8–0.93, q^2 range 0.62–0.80	30–58	n/r	*O*-Phosphorylated oximes	FSMLR: fragmental descriptors of up to eight non-hydrogen atoms, Log *P* MFTA: effective charge, *Q*, on an atom; effective van der Waals radius, Re; group lipophilicity, Lg	
Neurotoxicity [56]	Seven endpoints related to neurotoxicity: including effects on vesicular and membrane transporter-mediated uptake of dopamine, glutamate and gamma-aminobutyric acid	PCA; PLS	Two significant principal components (t1 and t2) explaining 51 % of the variation in the in vitro screening data (t1 = 37 % and t2 = 14 %) The PLS model (response formation of the chemically reactive oxygen containing species): $Q^2 = 0.63$	20	n/r	Non-dioxin-like polychlorinated biphenyls	67 chemical descriptors including partial atomic charges and parameters describing size and substitution pattern; Significant parameters (PLS model): molecular size, expressed as number of substituents and total surface area	PCA used to derive general toxicity profiles from the in vitro screening data
Neurotoxicity [57]	EC_{30}*rat*, EC_{30}*mouse*	TOPS-MODE approach	Model (rat): $r = 0.902$, $s = 0.252$, $F(6,38) = 27.5$, $r_{cv} = 0.902$, RMSECV = 0.273 Model (mice): $r = 0.901$, $s = 0.250$, $F(7,38) = 23.4$, $r_{cv} = 0.881$, RMSECV = 0.264	45/46	n/r	Non-congeneric series of solvents	Spectral moments, multiplication of moments, indicator of Hydrogen bond capacity of groups, experimental values of BP	Structural fragments, responsible for difference in neurotoxicity are analyzed
Neurotoxicity (in vitro [58])	Acetylcholinesterase inhibition, IC_{50}; in vitro	Pharmacophore modeling, using Catalist software	$r > 0.98$	8	n/r	Organophosphorous compounds	Pharmacophore models include at least one hydrogen bond acceptor site and 2–3 hydrophobic sites	
Lung toxicity [59]	IL-8 gene expression; in vitro	MLR Software: ADMEWORKS	Not reported statistics of the MLR model; 75 % accuracy of external validation (classification: upregulation class/downregulation class)	54	8	Multiple	WTPT3, MOLC4, V5CH, SYMM2, S3C, CRB_LEADL, and OPERA RULEI	The 54 chemicals are classified into two groups based on the gene expression of IL-8, namely upregulation class and downregulation class
Lung toxicity [60]	Cellular membrane damage of immortalized human lung epithelial cells (via lactate dehydrogenase release); in vitro	MLR, PCA, LDA classification	Models for TiO_2: $r^2 = 0.70–0.77$ Models for ZnO: $r^2 = 0.19–0.49$	24 measurements from five different TiO_2 features and 18 measurements from six different ZnO features	n/r	TiO_2 and ZnO nanoparticles	TiO_2: size in ultrapure water, concentration in ultrapure water, and zeta potential in ultrapure water. ZnO: size in ultrapure water, size in CCM, and concentration	The goodness of the classification measured by the resubstitution error

(continued)

Table 4
(continued)

Model, reference	Endpoint	Statistical method/software	Statistical parameters	Training set	Test set	Class(es) studied	Significant parameters	Note
Lung toxicity [61]	Activity of phospholipases A2 (PLA2), C (PLC), and D (PLD); in vitro	LR	r = 0.97, 0.98, 0.99, for PLA2, PLC, and PLD, respectively	4	n/r	Lipid ozonation products	E_{HOMO}, E_{LUMO}, and the net charge on the H49 atom	
Immunotoxicity [62]	Each endpoint corresponds to one out of 1418 assays, 36 molecular and cellular targets, 46 standard type measures related to immunotoxicity	TOPS-MODE approach	Accuracy = 91.76 %	6747	1156	Multiple	spectral moments	Multi-target QSAR (mt-QSAR)
Immunotoxicity [63]	Binding to the aryl hydrocarbon receptor (AhR)	CoMFA	Best model: r^2 = 0.858 q^2 = 0.684; r^2 (test set) = 0.851	59	19	Polychlorinated dibenzo-dioxins, polychlorinated dibenzo-furans, and polychlorinated biphenyls	Steric and electrostatic fields	
Immunotoxicity [64]	Log ED_{50}	MLR	r = 0.964 r_{cv} = 0.884	209	n/r	Polychlorinated diphenyl ethers	A parameter of electrostatic equilibrium on molecular surface	
Myelotoxicity [65]	pIC50 for human CFC-GEMM cells (colony-forming cell granulocyte, erythroid, megakaryocyte, macrophage) and GM-CFC (granulocytes monocytes colony-forming cell)	PCA, PLS Pentacle software: GRIND toxicophore-based descriptors calculated and PLS-DA performed	Best PLS model: r^2 = 0.79 q^2 = 0.72 r^2 (test set) = 0.67 RMSEP = 0.69 PLS-DA: Accuracy (test set prediction) = 86 %	14	21	Multiple	Volsurf descriptors, 2D structural and electrotopological E-states descriptors	Pentacle software used to calculate GRID toxicophore-based descriptors
Cardiotoxicity [66]	hERG channel blocking							Review
Cardiotoxicity [67]	hERG channel blocking							Review

AE adverse effect, *BP* boiling point, *BPNN* Backpropagation Neural Networks, *CA* classification accuracy, *cGST* cytosolic glutathione transferases, *Clog* calculated partition coefficient *P*, CoMFA comparative molecular field analysis, *CoMSIA* Comparative Molecular Similarity Index Analysis, *CRB LEADL* count of rotatable bonds, *ELUMO* energy of the lowest unoccupied molecular orbital, *FSMLR* Fast Stepwise Multiple Linear Regression, *hERG* human Ether-a-go-go Related-Gene, *IL-8* interleukin-8, *LDA* linear discriminant analysis, *LR* linear regression, *MCC* Matthews correlation coefficient, *MFTA MLR* multiple linear regression, Molecular Field Topology Analysis, *MGST1* microsomal glutathione transferase 1, *MOLC4* path-2 molecular connectivity, *n/r* not reported, *OPERA RULEI* the rule based on Lipinski's rule, *RMSECV* root mean square error of cross validation, *RMSEP* root-mean-squares error of prediction, *SVM* support vector machine, *Q2* cross-validation parameter, *Q2DCV* determination coefficient for double cross-validation, *r* correlation coefficient, *S3C* third order cluster MC Simple, *SYMM2* geometrical symmetry, *TOPS-MODE* topological sub-structural molecular design, *V5CH* fifth order chain MC Valence, *WTPT3* sum of atom indexes for all heteroatoms

387 non-nephrotoxins) and their major urinary metabolites (307 nephrotoxins and 233 non-nephrotoxins). Thus the study reflects the fact that the nephrotoxicity of a pharmacological compound is induced by the parent compound as well as its metabolites. The results of a tenfold cross-validation and external validation procedures showed a high accuracy of the models (better than 83 % for external validation sets).

For kidney toxicity, local QSARs have been developed for specific chemical classes, such as the haloalkenes. These high-volume chemicals used in industrial, synthetic, and pharmaceutical applications are common environmental pollutants. Many haloalkenes are known to be nephrotoxic in rodents after bioactivation via the cysteine conjugate beta-lyase pathway, which is triggered by formation of hepatic glutathione S-conjugates, a reaction catalyzed by cytosolic and microsomal glutathione transferases [68]. The study by Jolivette and Anders [53] relates the nephrotoxicity of nine haloalkenes to their lowest unoccupied molecular orbital energies, E_{LUMO}, reflecting their propensity for conjugation reactions catalyzed by glutathione transferase enzymes.

Very few QSAR studies of neurotoxicity have been published. An example is the work of Estrada et al. [57]. Their models are based on the TOPS-MODE approach, which provides a means of estimating the contributions to neurotoxicity in rats and mice of a series of structural fragments.

Organophosphorus (OP) compounds are well-known neurotoxic agents. These chemicals are potent inhibitors of serine esterases, the most critical of which is the widely distributed nervous system enzyme acetylcholinesterase (AChE). This well established mechanism of action underlies the usefulness of molecular modeling approaches like 3D QSAR and pharmacophore modeling to predict the inhibition potency of OPs. Several published models are based on these approaches [54, 55, 58, 63].

Among the commonly used software tools, Derek Nexus estimates neurotoxicity using a number of structural alerts: gamma-diketone or precursor, acrylamide or glycidamide, nitroimidazole, carbon disulfide or precursor, pyrethroid, 1-methyl-1,2,3,6-tetrahydropyridine, lead or lead compound, and organophosphorus ester.

Few studies have been published in relation to other organs/systems. Immunotoxicity can refer to immunosuppression in humans (caused, for example, by benzene and halogenated aromatic hydrocarbons), autoimmune disease (for example the pesticide dieldrin induces an autoimmune response against red blood cells, resulting in hemolytic anemia), and allergenicity (chemicals which stimulate the immune system can cause allergies or hypersensitivity reactions such as anaphylactic shock). Thus, immunotoxicity refers to a wide variety of biological effects, many of which involve complex biochemical networks. Tenorio-Borroto

et al. [62] have trained and validated a multi target-QSAR model for high-throughput screening of drug immunotoxicity using TOPS-MODE approach. Yuan et al. [63] have studied the key molecular features of polychlorinated dibenzodioxins, polychlorinated dibenzofurans, and polychlorinated biphenyls for determining binding affinity to the aryl hydrocarbon receptor (AhR)—an intracellular receptor which has been correlated to immunotoxicity, thymic atrophy, weight loss and acute lethality. CoMFA (Comparative Molecular Field Analysis) was applied to generate 3D QSAR models. In a study by Hui-Ying et al. [64], linear relationships between immunotoxicity values (log ED_{50}) and other biological activities of polychlorinated diphenyl ethers and their structural descriptors were established by multiple linear regression. It was shown that the structural descriptors derived from molecular electrostatic potentials together with the number of the substituting chlorine atoms on the two phenyl rings can be used to express the quantitative structure–property relationships of polychlorinated diphenyl ethers.

Evaluation of hematotoxicity is important step in early drug design. Particularly it is a common dose-limiting toxicity associated with anticancer drugs. The first attempt to build in silico models to predict the myelosuppressive activity of drugs from their chemical structure was made by Crivori et al. [65]. Two sets of potentially relevant descriptors for modeling myelotoxicity (i.e., 3D Volsurf and 2D structural and electrotopological E-states descriptors) were selected and PCA (Principal Component Analysis) was carried out on the entire set of data (38 drugs). The first two principal components discriminated the highest from the least myelotoxic compounds with a total accuracy of 95 %. In addition, a highly predictive PLS (Partial Least Squares) model was developed by correlating a selected subset of in vitro hematotoxicity data with Volsurf descriptors. After variable selection, the PLS analysis resulted in a one-latent-variable model with r^2 of 0.79 and q^2 of 0.72.

In contrast to other organ-specific effects, the in silico modeling of cardiotoxicity has been a rather productive field. This is because drug cardiotoxicity is one of the main reasons for drug related fatalities and subsequent drug withdrawals. In recent years, the hERG channel has been extensively investigated in the field of cardiotoxicity prediction as it has been found to play a major role in both cardiac electrophysiology and drug safety. Because hERG assays and QT animal studies are expensive and time consuming, numerous in silico models have been developed for use in early drug discovery. The earliest attempts to identify whether a molecule is a hERG blocker include a set of simple rules based on structural and functional features, but these rules are not always reliable predictors for identifying hERG blockers. In order to give more accurate predictions of hERG blockage, a wide range of QSAR models have been developed based on a variety of statistical

techniques and machine learning methods, including multiple linear regression, partial least square (PLS), k-nearest neighbor algorithm (kNN), linear discriminant analysis (LDA), artificial neural networks (ANN), support vector machine (SVM), self-organizing mapping (SOM), recursive partitioning (RP), random forest, genetic algorithm, and naive Bayesian classification (NBC). Most of these QSAR models are classifiers and only a few regression models have been reported.

Pharmacophore modeling has also been employed to develop ligand-based prediction models of hERG channel blockers. Since the crystal structure of the hERG channel is not available, all structure-based studies on its blockage are performed on homology models and are more qualitative and descriptive rather than predictive. For example they have been used for molecular docking, molecular dynamics simulations and free energy calculations to explore the hERG-blocker interactions.

Reviews by [66] and Villoutreix and Taboureau [67] summarize the advances and challenges in computational studies of hERG channel blockage. It is expected that the development of in silico models for hERG-related cardiotoxicity will stay active in the coming years in order to design drugs without undesirable side effects.

6 Conclusions

The modeling of acute systemic toxicity has largely focused on QSARs for predicting LD_{50} values and for categorizing chemicals according to ranges of LD_{50} values. For these purposes, which are potentially useful in the regulatory assessment of chemicals, the in silico models seem to perform as well as in vitro cytotoxicity methods. The developments in this field can be attributed to the availability of extensive LD_{50} datasets and a wide range of machine learning techniques. Many of these datasets, and software tools derived from the datasets, are in the public domain.

The emergence of mechanism-based toxicology (e.g., adverse outcome pathways) is a tremendous opportunity to improve current models with better biological knowledge. Indeed, the time of global (and scientifically dubious) QSARs predicting LD_{50} based on chemical properties for the whole chemical space is probably coming to an end. Future models should target specific toxicity mechanisms on the basis of current biological knowledge. Historically, this was actually done implicitly by focusing model building on very limited chemical classes (supposedly acting via the same mechanism). According to this approach, global LD_{50} models would be the sum of a multitude of accurate predictors dedicated to describe well-defined mechanisms of action. In this context, the use of biological (in vitro) descriptors in combination

with traditional molecular descriptors provides a promising means of building local QSAARs based on mechanistically based chemical classes.

In general, the modeling of organ-specific and system-specific effects represents an underdeveloped field, ripe for future research but far from regulatory applications, which typically rely on the assessment of lethality. A notable exception concerns the modeling of receptors and ion channels implicated in specific organ pathologies, such as the hERG channel in relation to cardiotoxicity. The development of models for upstream (molecular and cellular) effects represents a more scientifically meaningful exercise which also promises to unify the traditional regulatory distinction between the acute and repeat dose toxicity.

A future research initiative could include, for example, reexamination of the datasets for hepatobiliary and urinary tract toxicities of drugs with a view to developing more accessible models and assessing their applicability to chemicals other than pharmaceuticals. In addition, the concept of reactivity-based toxicity, now established as a plausible mechanism for hepatocyte toxicity, could be further exploited using data from hepatocyte cultures and cell lines. In some areas, such as immunotoxicity, short-term progress seems unlikely. The complexity of such effects probably means that systems biology approaches will be more appropriate.

In general, the development of models for organ-specific and system-specific effects will depend on a new generation of databases, such as the COSMOS database, which contain high quality data that are structured and annotated according to meaningful chemical and biological ontologies.

Acknowledgement

This work was partially funded by the European Community's Seventh Framework Program (FP7/2007–2013) COSMOS Project under grant agreement no 266835 and from Cosmetics Europe. The authors are grateful to Dr. Pilar Prieto (JRC) for critically reviewing this work.

References

1. Combes R, Grindon C, Cronin MTD, Roberts D, Garrod J (2008) Integrated decision-tree testing strategies for acute systemic toxicity and toxicokinetics with respect to the requirements of the EU REACH legislation. ATLA 36(Suppl 1):91–109
2. Devillers J, Devillers H (2009) Prediction of acute mammalian toxicity from QSARs and interspecies correlations. SAR QSAR Environ Res 20:467–500
3. Kleandrova VV, Luan F, Speck-Planche A, Cordeiro MN (2015) In silico assessment of the acute toxicity of chemicals: recent advances and new model for multitasking prediction of toxic effect. Mini Rev Med Chem 15(8):677–686
4. Lapenna S, Fuart-Gatnik M, Worth A (2010) Review of QSAR models and software tools for predicting acute and chronic systemic toxicity, JRC technical report EUR 24639 EN. Publications Office of the European Union,

Luxembourg, http://publications.jrc.ec.europa.eu/repository/

5. Tsakovska I, Lessigiarska I, Netzeva T, Worth Andrew P (2008) A mini review of mammalian toxicity (Q)SAR models. QSAR Comb Sci 27(1):41–48
6. OECD (2001) Guideline for testing of chemicals, 420, acute oral toxicity—fixed dose method. OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
7. European Union (2008a) Regulation (EC) No 440/2008 of 30 May 2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Off J Eur Union L142
8. OECD (2001) Guideline for testing of chemicals, 423, acute oral toxicity—acute toxic class method. OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
9. OECD (2001) Guideline for testing of chemicals, 425, acute oral toxicity—up-and-down procedure. OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
10. OECD (1987) Guideline for testing of chemicals, 420, acute dermal toxicity. OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
11. OECD (2009) Guideline for testing of chemicals, 403, acute inhalation toxicity. OECD guidelines for the testing of chemicals, section 4. OECD Publishing, Paris
12. European Union (2014) Commission Regulation (EU) No 260/2014 of 24 January 2014 amending, for the purpose of its adaptation to technical progress, Regulation (EC) No 440/2008 laying down test methods pursuant to Regulation (EC) No 1907/2006 of the European Parliament and of the Council on the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Off J Eur Union L81:1–253
13. OECD (2009) Guideline for testing of chemicals, 436, acute inhalation toxicity—acute toxic class method. OECD guidelines for the testing of chemicals, Section 4. OECD Publishing, Paris
14. European Union (2009) Regulation (EC) No 1223/2009 of the European Parliament and the Council of 30 November 2009 on cosmetic products. Off J Eur Union L342:59–209
15. European Union (2008b) Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labelling and packaging of substances and mixtures, amending and repealing Directives 67/548/EEC and 1999/45/EC, and amending Regulation (EC) No 1907/2006. Off J Eur Union L353
16. European Union (2006) Regulation (EC) No 1907/2006 of the European Parliament and the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. Off J Eur Union L396:1–849
17. European Union (2012) Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products. Off J Eur Union L167:1–116
18. European Union (2009) Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC. Off J Eur Union L309:1–47
19. ICH (2009) International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals M3(R2). Recommended for adoption at step 4 of the ICH process on 11 June 2009
20. Chapman K, Creton S, Kupferschmidt H, Bond GR, Wilks MF, Robinson S (2010) The value of acute toxicity studies to support the clinical management of overdose and poisoning: a cross-discipline consensus. Regul Toxicol Pharmacol 58:354–359
21. Robinson S, Delongeas JL, Donald E, Dreher D, Festag M, Kervyn S, Lampo A, Nahas K, Nogues V, Ockert D, Quinn K, Old S, Pickersgill N, Somers K, Stark C, Stei P, Waterson L, Chapman K (2008) A European pharmaceutical company initiative challenging the regulatory requirement for acute toxicity studies in pharmaceutical drug development. Regul Toxicol Pharmacol 50:345–352
22. Prieto P, Burton J, Graepel R, Price A, Whelan M, Worth A (2014) EURL ECVAM strategy to replace, reduce and refine the use of animals in the assessment of acute mammalian systemic toxicity, JRC report EUR 26797 EN. Publications Office of the European

Union, Luxembourg, http://publications.jrc.ec.europa.eu/repository/

23. Fuart-Gatnik M, Worth AP (2010) Review of software tools for toxicity prediction. JRC report EUR 24489 EN. Publications Office of the European Union. http://publications.jrc.ec.europa.eu/repository/
24. Chakravarti SK, Saiakhov RD, Klopman G (2012) Optimizing predictive performance of CASE Ultra expert system models using the applicability domains of individual toxicity alerts. J Chem Inform Model 52(10): 2609–2618
25. Tunkel J, Mayo K, Austin C, Hickerson A, Howard P (2005) Practical considerations on the use of predictive models for regulatory purposes. Environ Sci Technol 39(7):2188–2199
26. Vedani A, Dobler M, Smieško M (2012) VirtualToxLab—a platform for estimating the toxic potential of drugs, chemicals and natural products. Toxicol Appl Pharmacol 261(2): 142–153
27. Roberts G, Myatt GJ, Johnson WP, Cross KP, Blower PE Jr (2000) LeadScope: software for exploring large sets of screening data. J Chem Inf Comput Sci 40(6):1302–1314
28. Halle W (2003) The registry of cytotoxicity: toxicity testing in cell cultures to predict acute toxicity (LD50) and to reduce testing in animals. Altern Lab Anim 31:89–198
29. Kinsner-Ovaskainen A, Rzepka R, Rudowski R, Coecke S, Cole T, Prieto P (2009) Acutoxbase, an innovative database for in vitro acute toxicity studies. Toxicol In Vitro 23(3):476–485
30. Kinsner-Ovaskainen A, Prieto P, Stanzel S, Kopp-Schneider A (2013) Selection of test methods to be included in a testing strategy to predict acute oral toxicity: an approach based on statistical analysis of data collected in phase 1 of the ACuteTox project. Toxicol In Vitro 27(4):1377–1394
31. Prieto P, Kinsner-Ovaskainen A, Stanzel S, Albella B, Artursson P, Campillo N, Cecchelli R, Cerrato L, Díaz L, Di Consiglio E, Guerra A, Gombau L, Herrera G, Honegger P, Landry C, O'Connor JE, Páez JA, Quintas G, Svensson R, Turco L, Zurich MG, Zurbano MJ, Kopp-Schneider A (2013) The value of selected in vitro and in silico methods to predict acute oral toxicity in a regulatory context: results from the European Project ACuteTox. Toxicol In Vitro 27(4):357–376
32. Hoffmann S, Kinsner-Ovaskainen A, Prieto P, Mangelsdorf I, Bieler C, Cole T (2010) Acute oral toxicity: variability, reliability, relevance and interspecies comparison of rodent LD50 data from literature surveyed for the ACuteTox project. Regul Toxicol Pharmacol 58:395–407
33. Fonger GC, Hakkinen P, Jordan S, Publicker S (2014) The National Library of Medicine's (NLM) Hazardous Substances Data Bank (HSDB): background, recent enhancements and future plans. Toxicology 325:209–216
34. Zhu H, Martin TM, Ye L, Sedykh A, Young DM, Tropsha A (2009) Quantitative structure–activity relationship modeling of rat acute toxicity by oral exposure. Chem Res Toxicol 22(12):1913–1921
35. Zhu H, Ye L, Richard A, Golbraikh A, Wright FA, Rusyn I, Tropsha A (2009) A novel two-step hierarchical quantitative structure-activity relationship modeling work flow for predicting acute toxicity of chemicals in rodents. Environ Health Perspect 117(8):1257–1264
36. Diaza RG, Manganelli S, Esposito A, Roncaglioni A, Manganaro A, Benfenati E (2015) Comparison of in silico tools for evaluating rat oral acute toxicity. SAR QSAR Environ Res 26(1):1–27
37. Norlén H, Berggren E, Whelan M, Worth A (2012) An investigation into the use of computational and in vitro methods for acute systemic toxicity prediction, JRC report EUR 25473 EN. Publications Office of the European Union, Luxembourg, http://publications.jrc.ec.europa.eu/repository/
38. Lessigiarska I, Worth AP, Netzeva TI, Dearden JC, Cronin MTD (2006) Quantitative structure-activity-activity and quantitative structure-activity investigations of human and rodent toxicity. Chemosphere 65(10):1878–1887
39. Raevsky OA, Grigor'ev VJ, Modina AE, Worth AP (2010) Prediction of acute toxicity to mice by the arithmetic mean toxicity (AMT) modelling approach. SAR QSAR Environ Res 21(1):265–275
40. Chavan S, Nicholls IA, Karlsson BC, Rosengren AM, Ballabio D, Consonni V, Todeschini R (2014) Towards global QSAR model building for acute toxicity: Munro database case study. Int J Mol Sci 15(10):18162–18174
41. Lu J, Peng J, Wang J, Shen Q, Bi Y, Gong L, Zheng M, Luo X, Zhu W, Jiang H, Chen K (2014) Estimation of acute oral toxicity in rat using local lazy learning. J Cheminform 6:26
42. Low Y, Sedykh A, Fourches D, Golbraikh A, Whelan M, Rusyn I, Tropsha A (2013) Integrative chemical-biological read-across approach for chemical hazard classification. Chem Res Toxicol 26(8):1199–1208
43. Sedykh A, Zhu H, Tang H, Zhang L, Richard A, Rusyn I, Tropsha A (2011) Use of in vitro

HTS-derived concentration-response data as biological descriptors improves the accuracy of QSAR models of in vivo toxicity. Environ Health Perspect 119(3):364–370

44. Ekwall B, Barile FA, Castano A et al (1998) MEIC evaluation of acute systemic toxicity. Part V. I. The prediction of human toxicity by rodent LD50 values and results from 61 in vitro methods. Alternatives to laboratory animals. ATLA 26(Suppl 2):617–658
45. Spielmann H, Genshow E, Liebsch M, Halle W (1999) Determination of the starting dose for acute oral toxicity (LD50) testing in the up and down procedure (UDP) from cytotoxicity data. Altern Labor Anim ATLA 27(6):957–966
46. Clemedson C, Ekwall B (1999) Overview of the final MEIC results: I. The in vitro–in vitro evaluation. Toxicol In Vitro 13(4-5):657–663
47. EURL ECVAM (2013) EURL ECVAM recommendation on the 3T3 neutral red uptake cytotoxicity assay for acute oral toxicity testing (2013) JRC technical report EUR 25946 EN. Publications Office of the European Union, Luxembourg, https://eurl-ecvam.jrc.ec.europa.eu/eurl-ecvam-recommendations
48. Rovida C, Alépée N, Api AM, Basketter DA, Bois FY et al (2015) Integrated Testing Strategies (ITS) for safety assessment. ALTEX Altern Anim Exp 32(1):25–40
49. Matthews EJ, Ursem CJ, Kruhlak NL, Benz RD, Sabaté DA, Yang C, Klopman G, Contrera JF (2009) Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: part B. Use of (Q)SAR systems for early detection of drug-induced hepatobiliary and urinary tract toxicities. Regul Toxicol Pharmacol 54(1):23–42
50. Lee S, Kang Y, Park H, Dong M, Shin J, No K (2013) Human nephrotoxicity prediction models for three types of kidney injury based on data sets of pharmacological compounds and their metabolites. Chem Res Toxicol 26(11):1652–1659
51. Myshkin E, Brennan R, Khasanova T, Sitnik T, Serebriyskaya T, Litvinova E, Guryanov A, Nikolsky Y, Nikolskaya T, Bureeva S (2012) Prediction of organ toxicity endpoints by QSAR modeling based on precise chemical histopathology annotations. Chem Biol Drug Des 80:406–416
52. Munday R, Smith BL, Munday CM (2007) Structure-activity relationships in the haemolytic activity and nephrotoxicity of derivatives of 1,2- and 1,4-naphthoquinone. J Appl Toxicol 27(3):262–269
53. Jolivette LJ, Anders MW (2002) Structure-activity relationship for the biotransformation of haloalkenes by rat liver microsomal glutathione transferase 1. Chem Res Toxicol 15(8): 1036–1041
54. Makhaeva GF, Radchenko EV, Palyulin VA, Rudakova EV, Aksinenko AY, Sokolov VB, Zefirov NS, Richardson RJ (2013) Organophosphorus compound esterase profiles as predictors of therapeutic and toxic effects. Chem Biol Interact 203(1):231–237
55. Makhaeva GF, Radchenko EV, Baskin II, Palyulin VA, Richardson RJ, Zefirov NS (2012) Combined QSAR studies of inhibitor properties of O-phosphorylated oximes toward serine esterases involved in neurotoxicity, drug metabolism and Alzheimer's disease. SAR QSAR Environ Res 23(7–8):627–647
56. Stenberg M, Hamers T, Machala M, Fonnum F, Stenius U, Lauy AA, van Duursen MB, Westerink RH, Fernandes EC, Andersson PL (2011) Multivariate toxicity profiles and QSAR modeling of non-dioxin-like PCBs—an investigation of in vitro screening data from ultra-pure congeners. Chemosphere 85(9): 1423–1429
57. Estrada E, Molina E, Uriarte E (2001) Quantitative structure-toxicity relationships using TOPS-MODE. 2. Neurotoxicity of a non-congeneric series of solvents. SAR QSAR Environ Res 12(5):445–459
58. Yazal JE, Rao SN, Mehl A, Slikker W Jr (2001) Prediction of organophosphorus acetylcholinesterase inhibition using three dimensional quantitative structure-activity relationship (3D-QSAR) methods. Toxicol Sci 63(2): 223–232
59. Hosoya J, Tamura K, Muraki N, Okumura H, Ito T, Maeno M (2011) A novel approach for a toxicity prediction model of environmental pollutants by using a quantitative structure-activity relationship method based on toxicogenomics. ISRN Toxicol 2011:515724
60. Sayes C, Ivanov I (2010) Comparative study of predictive computational models for nanoparticle-induced cytotoxicity. Risk Anal 30(11):1723–1734
61. Kafoury RM, Huang MJ (2005) Application of quantitative structure activity relationship (QSAR) models to predict ozone toxicity in the lung. Environ Toxicol 20(4):441–448
62. Tenorio-Borroto E, Peñuelas-Rivas CG, Vásquez-Chagoyán JC, Castañedo N, Prado-Prado FJ, García-Mera X, González-Díaz H (2014) Model for high-throughput screening

of drug immunotoxicity—study of the antimicrobial G1 over peritoneal macrophages using flow cytometry. Eur J Med Chem 72:206–220

63. Yuan J, Pu Y, Yin L (2013) Docking-based three-dimensional quantitative structure-activity relationship (3D-QSAR) predicts binding affinities to aryl hydrocarbon receptor for polychlorinated dibenzodioxins, dibenzofurans, and biphenyls. Environ Toxicol Chem SETAC 32(7):1453–1458
64. Hui-Ying X, Jian-Wei Z, Gui-Xiang H, Wei W (2010) QSPR/QSAR models for prediction of the physico-chemical properties and biological activity of polychlorinated diphenyl ethers (PCDEs). Chemosphere 80(6):665–670
65. Crivori P, Pennella G, Magistrelli M, Grossi P, Giusti AM (2011) Predicting myelosuppression of drugs from in silico models. J Chem Inf Model 51(2):434–445
66. Wang S, Li Y, Xu L, Li D, Hou T (2013) Recent developments in computational prediction of HERG blockage. Curr Top Med Chem 13(11):1317–1326
67. Villoutreix BO, Taboureau O (2015) Computational investigations of hERG channel blockers: new insights and current predictive models. Adv Drug Deliv Rev 15:00022–00028
68. Anders MW, Dekant W (1998) Glutathione-dependent bioactivation of haloalkenes. Annu Rev Pharmacol Toxicol 38(1):501–537
69. Prieto and Kinsner-Ovaskainen (2015). Short commentary to Human in vivo database now on ACuteTox home page [Toxicol. In Vitro 27 (2013) 2350-2351]. Toxicol In Vitro. 2015 Mar;29(2):415

Chapter 11

In Silico Models for Hepatotoxicity

Mark Hewitt and Katarzyna Przybylak

Abstract

In this chapter we review the challenges of predicting human hepatotoxicity. Principally, this is our partial understanding of a very complex biochemical system and our ability to emulate that in a predictive capacity. We give an overview of the published modeling approaches in this area to date and discuss their design, strengths, and weaknesses. It is interesting to note the shift during the period of this review in the direction of evidenced-based approaches including structural alerts and pharmacophore models. Proposals on how best to utilize the data emerging from modeling studies are also discussed.

Key words Liver, Hepatotoxicity, In silico or computational prediction, QSAR, Expert system

1 Introduction

Toxicity of new medicinal compounds to the liver is perhaps the most significant hurdle to overcome during drug development. Often termed "drug-induced liver injury (DILI)," these adverse effects can range in nature from subtle elevations in serum enzymes, to acute and chronic hepatocellular injuries (steatosis, necrosis, cirrhosis), cholestatic injuries, and neoplasia [1]. Unfortunately, DILI accounts for a significant proportion of drugs (>25 %) being terminated during development or withdrawn from the market [2].

Given the protective/metabolic function of the liver, it is perhaps not surprising that hepatotoxicity is frequently encountered. Given the livers high blood flow and first-pass metabolism it is a certainty that a proportion of the diverse pharmaceutical products in use today are hepatotoxic (via metabolic conversion). Unwanted interaction between the liver and pharmaceuticals is a major hurdle which can often result in the loss of drug efficacy and/or hepatotoxicity. Despite preclinical and clinical safety assessments, liver toxicity remains a main cause of drug development failures and subsequent market withdrawal due to the poor predictivity of idiosyncratic toxicity in animal models [3, 4].

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_11, © Springer Science+Business Media New York 2016

The need to predict whether a new drug is likely to lead to hepatotoxicity is clear. Information relating to the likelihood of liver toxicity is critical in order to increase patient safety, reduce the frequency of drug withdrawals/failures and to further increase our understanding of liver toxicity.

Interestingly, despite a clear need to predict these effects, computational studies in this area have only started to emerge in the last decade [1, 5]. Such methods are well-suited to the rapid screening of large numbers of compounds, offering significant time and cost savings over traditional animal-based screening approaches. Furthermore, computational screening has been successfully established for other endpoints, including skin sensitization and mutagenicity [6, 7]. When coupled with supporting in vitro data they provide a powerful tool capable of predicting toxicity and, in certain cases, determining the mechanism of that toxicity. However, as stated, computational models for DILI have only recently started to surface and those that have been published are often limited in their scope and predictive capability.

The reason for this is simple; predicting toxicity to the liver is far from simple!

The task of predicting DILI is difficult because (a) the liver is an intricate and complex organ with numerous biological and metabolic pathways that can lead to downstream toxicities and (b) many of these toxicological pathways are poorly understood or remain unknown. Furthermore, as already introduced, DILI can take many forms and range in severity. With the absence of a single "catch all" biomarker that can be used as a metric of hepatotoxicity, actually measuring these affects in patients is very challenging.

Furthermore, toxicity to the liver can occur in a dose dependent manner (termed intrinsic toxicity) or in a non-dose dependent manner (termed idiosyncratic toxicity) [8]. Typically, intrinsic liver toxicity accounts for approximately 80 % of cases, where the observed toxicity can be related to a particular mechanism of action (pharmacological, toxicological, or chemical) triggered by the drug or its metabolite(s). Idiosyncratic toxicity is very difficult to predict and is thankfully a relatively rare occurrence. The susceptibility of particular patients to idiosyncratic DILI has been the focus of much research [9], but the prediction of idiosyncratic effects remains a herculean task.

2 Prediction of Hepatotoxicity

It is crucial to develop predictive screening systems and mechanistic models capable of detecting hepatotoxicity as early as possible in the drug development process. However, accurate prediction of organ toxicity is very challenging due to the complexity of the underlying mechanisms, which are very often not known.

Moreover, the lack of specific and selective biomarkers that can be used to detect hepatotoxicity leads to a shortage of reliable in vivo and in vitro data from which to derive predictive models. Most likely as a result of these limitations, the first in silico models were described in the literature only at the beginning of the last decade [10, 11]. The bulk of available computational models for liver toxicity have been published more recently [1].

Published models for the prediction of hepatotoxicity can be classified as belonging to one of two approaches [12]:

(A) The development of statistically based structure–activity relationship (SARs) of varying complexity. This modeling approach utilizes existing DILI data to derive a model able to predict a quantitative estimation of hepatotoxicity.

(B) The development of qualitative "models" based on expert knowledge, directly related to chemical structure and molecular features. Most often, these qualitative approaches result in the development of structural alerts or three-dimensional pharmacophore models.

These models can be further subdivided based upon (1) the endpoint being modeled (general hepatotoxicity or a specific aspect (e.g., steatosis)), (2) the type of variable(s) (descriptors) used to develop the model, or (3) the type of data being modeled (in vivo or in vitro). Figure 1 depicts how the 21 published models that are the subject of this chapter can be divided using these four differentiating criteria.

Statistical models are generally built from a training dataset of chemical structures and their associated toxicity data, expressed either in quantitative or qualitative terms, using an appropriate algorithm. Therefore, they are often referred to as "(quantitative) structure–activity relationships" ((Q)SARs). In contrast, expert systems apply expert knowledge to a predictive environment and are usually not statistically based. The knowledge is based on the observed toxicity of known compounds, together with an understanding of toxicological mechanisms, metabolism and chemical reactivity [13].

The development of statistical models is usually faster (if suitable data are available) than that of expert systems, since expert systems require extensive study and integration with existing literature sources and are usually evidence-based (examples and supporting documentation is supplied along with a prediction). Therefore, statistical models are the most common. Approximately 75 % of the existing predictive models for liver toxicity have been developed using an array of different statistical methodologies, including discriminant analysis [14], Bayesian models [15, 16], Artificial Neutral Networks (ANN) [14], k-Nearest Neighbor (kNN) [17, 18], Random Forest (RF) [18, 19], and specialist QSAR software [20].

In terms of endpoint, most in silico models are focused towards the prediction of general hepatotoxicity (positive/negative irrespective of the mechanism/toxicity outcome) [5, 10, 14, 15, 18, 19, 21–26]. However, it is important to stress that this trend seems to be changing in recent years as the number of approaches considering more specific endpoints is increasing. Examples of these specific endpoints include elevations of liver serum enzymes [17], cholestasis and jaundice [20], hepatosteatosis [27, 28], and hepatic histopathologic effects including hypertrophy, injury, and proliferative lesions [29].

It is interesting to see that the majority of in silico approaches have utilized variables representing only chemical structure [10, 11, 14–17, 20, 21, 23–25]. It is perhaps not surprising given that QSAR models traditionally relate chemical structure to observed activity, but it seems here that the complex nature of the liver may warrant the use of biological descriptors to describe the biological process/systems at work. Only three models, discussed later, employed both chemical and biological descriptors and are referred to as hybrid models [18, 19, 29].

Finally, considering the nature of endpoint data used for modeling, most models have been developed using in vivo data. This can be broken down further into human data [10, 14–17, 19, 20, 24–26] and animal data [18, 29] which may be further subdivided into data from different species [23]. Only two models have been built using in vitro data [11, 21] and a further two models utilizing both in vitro and in vivo data [23, 30]. The 21 in silico models considered in this chapter can be subdivided by their differentiating characteristics as described by Fig. 1. The models will be discussed in the context of these categories and the strengths and weaknesses of different modeling methods will be highlighted. Potential future developments in the area are also speculated.

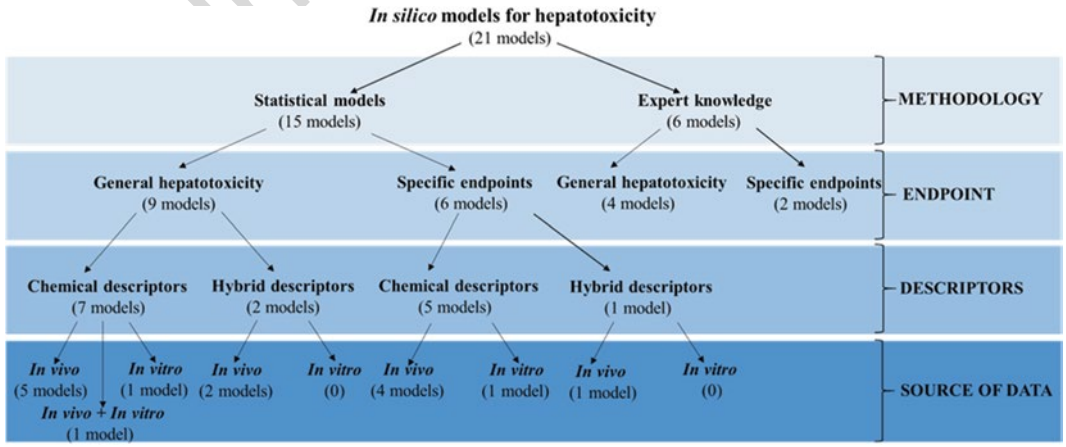


Fig. 1 Summary of published in silico models for predicting liver toxicity between 2000 and 2015

2.1 Statistically Derived Quantitative Models

A large proportion of the published hepatotoxicity models are statistical in their nature. The predictive element of these models is the statistical correlation of toxicity with one or more dependent variables. The approach used to identify and model this correlation varies considerably both in terms of methodology and complexity. Usually, statistically derived models are developed using sophisticated modeling software and tools.

The premise of any (Q)SAR model is the relationship between chemical structure (described using a number of descriptors) and biological activity (e.g., liver toxicity). This enables predictions of such activity to be made for new substances based on their chemical structure. The algorithms used to construct these models comprise of simple linear regression, complex multi-variant data modeling, data mining, and classification approaches [31]. Every statistical model has to be internally and externally validated to show its true predictive power and reliability [32, 33]. The predictive performance is usually evaluated by sensitivity (correctly predicted positive chemicals), specificity (correctly predicted negative chemicals) and accuracy (correctly predicted positive and negative chemicals). High sensitivity and specificity of a model guarantees correct classification of toxicologically active and inactive compounds. Therefore, it is the most important feature when aiming to detect potential hepatotoxic drugs in early drug development, since the consequences of misclassifying a toxic (positive) drug are severe (i.e., the possibility of a toxic drug reaching clinical trials) [14]. Of course, in drug development poor specificity can be a significant problem since many negative compounds may be dropped from further development unnecessarily.

Table 1, at the end of this section, describes the 15 diverse hepatotoxicity models discussed in this chapter and gives details of the methodologies employed, the endpoint modeled, the type of descriptors utilized, and the source of hepatotoxicity data.

2.1.1 Statistical Models for In Vivo General Hepatotoxicity Endpoint Using Chemical Descriptors

As already stated, most of the available in silico models have been developed based on in vivo data and are used to predict a general hepatotoxicity endpoint. These models consider intrinsic hepatotoxicity, idiosyncratic hepatotoxicity or a combination of these. The majority have been developed based only on the chemical features of the training set.

One of the first published in silico models was developed by Cheng and Dixon (ID 1 in Table 1) and is predicting intrinsic liver toxicity in humans [10]. Data for 382 drug and drug-like compounds (of various therapeutic classes) were collected from the literature. Amongst them, there were 149 chemicals which caused dose-dependent hepatocellular, cholestatic, neoplastic and other liver injuries. The authors employed a modeling method known as recursive partitioning (RP) [34, 35] with an ensemble approach [36], wherein the overall model is actually an average of numerous

t1.1 **Table 1**

t1.2 **Statistically based in silico models to predict hepatotoxicity available in the literature**

t1.3 t1.4 ID	Statistical method	Endpoint	Descriptors	Type and size of data	Validation	Predictive performance	Ref
t1.5 t1.6 t1.7 t1.8 t1.9 1	Ensemble recursive partitioning	Dose-dependent hepatotoxicity	Chemical: 25 1D and 2D descriptors filtered by Monte Carlo	In vivo data for 382 chemicals	IV: LOO, leave 10 % out EV: 54 chemicals	IV: 78 % (LOO), 76 % (leave 10 % out) SEN; 90 % (LOO), 75 % (leave 10 % out) SPE; 85 % (LOO), 76 % (leave 10 % out) ACC EV: 81 % ACC	[10]
t1.10 t1.11 t1.12 t1.13 t1.14 2	Molecular interaction fields and SIMCA	Cell proliferation; LDH activity; ATP levels; Caspases 3 and 7 levels	Chemical: steric and electrostatic IFO descriptors from SYBYL	In vitro (human hepatocyte cells) for 654 chemicals	CV for 27 NSAIDs	IV: 52 % ACC CV: 93 % SEN; 85 % SPE; 83 % ACC (for 6 NSAIDs)	[11]
t1.15 t1.16 t1.17 t1.18 3	QSAR (MLR using Sigma Stat)	LC_{50} (mM)	Chemical: log *P* EHOMO and μ calculated by MOPAC	In vitro for rat and human cells for 12 halobenzenes	Not reported	$r^2 = 0.966$ for rat cell, 0.993 for human cell, 0.846 for induced rat cell	[21]
t1.19 t1.20 t1.21 t1.22 t1.23 t1.24 t1.25 t1.26 t1.27 t1.28 t1.29 4	LDA, ANN, OneR	Idiosyncratic hepatotoxicity	Chemical: 3D RDF descriptors	In vivo human for 74 drugs	CV, EV for 13 drugs and three pairs of similar chemicals with opposite toxic potential	LDA model Train: 88 % SEN, 93 % SPE, 90 % ACC Test: 84 % SEN, 91 % SPE, 88 % ACC ANN model Train: 92 % SEN, 90 % SPE, 91 % ACC Test: 75 % SEN, 80 % SPE, 78 % ACC OneR model Train: 77 % SEN, 100 % SPE, 84 % ACC Test: 75 % SEN, 98 % SPE, 81 % ACC EV: for 3 pairs: 83 % ACC; for 13 drugs 69 % ACC	[14]

t1.30 t1.31 t1.32 t1.33 t1.34 t1.35 t1.36 t1.37	5	QSAR software: MC4PC, BioEpisteme, MDL-QSAR, Leadscope	(1) Hepatobiliary: liver disorders; jaundice and cholestasis; liver enzymes; gall bladder disorders; bile duct disorders, (2) urinary tract.	Chemical 2D descriptors	In vivo human for 1660 chemicals	IV: LMO, LOO EV for 18 toxic chemicals	IV: for LMO 39 % SEN, 86 % SPE For consensus model: 56 % SEN, 78 % SPE EV: 89 % ACC	[20, 45, 46]
t1.38 t1.39 t1.40 t1.41	6	Ligand-based Bayesian model	Idiosyncratic hepatotoxicity	Chemical 2D descriptors and FCFP of maximum diameter of 6	In vivo human for 295 chemicals	EV for 237 chemicals	EV: 56 % SEN, 67 % SPE, 60 % ACC	[15]
t1.42 t1.43 t1.44 t1.45 t1.46	7	kNN	Five liver serum enzymes: ALP, ALT, AST, LDH, GG	Chemical: MolConnZ-topological indices and Dragon descriptors	In vivo human for 490 chemicals	IV, Y-randomization, EV	EV: for composite liver enzymes: 74 % SEN, 94 % SPE, 84 % ACC	[17]
t1.47 t1.48 t1.49 t1.50 t1.51 t1.52 t1.53	8	SVM and clustering by chemical similarity	Hepatotoxicity	Chemical: 2D molecular fragments and Dragon descriptors	In vivo and in vitro human, rodent, and non-rodent for 951 chemical	Fivefold CV, EV for 246 chemicals and 18 chemicals toxic in non-rodents	Fivefold CV internal: 62–67 % ACC Fivefold CV external: 62–67 % ACC EV: for 246 chemicals 65–67 % ACC	[23]
t1.54 t1.55 t1.56 t1.57 t1.58 t1.59 t1.60 t1.61	9	kNN, SVM, RF, DWD	Hepatotoxicity	Chemical: Dragon and MOE Biological: toxicogenomics	In vivo rat for 127 chemicals	Fivefold CV	For chemical descriptors: CV: 45–56 % SEN, 60–77 % SPE, 55–61 % ACC For biological descriptors: CV: 57–67 % SEN, 77–84 % SPE, 69–76 % ACC For hybrid descriptors: CV: 76–77 % ACC	[18]

(continued)

Table 1
(continued)

ID	Statistical method	Endpoint	Descriptors	Type and size of data	Validation	Predictive performance	Ref
10	Ensemble of mixed learning algorithms	Hepatotoxicity	Chemical: PaDEL descriptors	In vivo human for 1087 chemicals	Fivefold CV, Y-randomization, EV for 187 chemicals divided into three sets	CV: 64 % SEN, 63 % SPE, 64 % ACC EV1: 68 % SEN, 71 % SPE, 70 % ACC EV2: 64 % SEN, 37 % SPE, 51 % ACC EV3: 62 % SEN, 62 % SPE, 62 % ACC	[24]
11	13 QSAR models developed using Bayesian methodology	13 hepatotoxic side effects	Chemical: functional class fingerprints (FCFP_6)	In vivo human for 888 chemicals	IV: LOO EV for three sets: LTKD, Pfizer, and O'Brien	LOO for 13 models: >71 % SEN, >94 % SPE, >93 % ACC EV LTKD: 66 % SEN, 67 % SPE, 66 % ACC EV Pfizer: 52 % SEN, 73 % SPE, 60 % ACC EV O'Brien: 56 % SEN, 93 % SPE, 70 % ACC	[16]
12	Machine learning methodology DT	Hepatotoxicity	Chemical: 82 Mold descriptors	In vivo human for 197 chemicals	Tenfold CV EV for three sets	CV: 58 % SEN, 78 % SPE, 70 % ACC EV1: 66 % SEN, 72 % SPE, 69 % ACC EV2: 58 % SEN, 67 % SPE, 62 % ACC EV3: 61 % SEN, 66 % SPE, 63 % ACC	[25]
13	RF	Hepatotoxicity	Chemical: CDK, Dragon, MOE Biological: HIATs	In vivo human for 292 chemicals	Fivefold CV	CV for chemical descriptors: 67 % SEN, 54 % SPE, 60 % ACC CV for biological descriptors: 67 % SEN, 87 % SPE, 77 % ACC CV for hybrid descriptors: 71 % SEN, 74 % SPE, 73 % ACC	[19]

14	Six machine learning analysis	Hepatic histopathologic effects: hypertrophy, injury and proliferative lesions	Chemical: 726 descriptor Biological: 124 bioactivity from ToxCast21	In vivo rat	Tenfold CV	CV: 84 % ACC for hypertrophy; 80 % ACC for injury; and 80 % ACC for proliferative lesions	[29]
15	PLS-DA	LXR binding potential involved in liver steatosis	Chemical: 6 PaDEL and 5 RDKit	LXRβ binding affinity for 356 LXR binders	Not reported		[52]

IV internal validation, *EV* external validation, *CV* cross-validation, *LOO* leave-one-out, *SEN* sensitivity, *SPE* specificity, *ACC* accuracy, *SIMCA* Soft Independent Modeling of Class Analogy, *LDH* lactate dehydrogenase, *ATP* adenosine triphosphate, *IFO* Idiotropic Field Orientation, *NSAIDs* nonsteroidal anti-inflammatory drugs, *QSAR* quantitative structure–activity relationship, *MLR* multiple linear regression, *EHOMO* energy of highest occupied molecular orbital, *LDA* linear discriminant analysis, *ANN* artificial neural networks, *RDF* radial distribution function, *LMO* leave many out, *FCFP* functional class fingerprint, *kNN* k-Nearest Neighbor, *ALP* alkaline phosphatase, *ALT* alanine aminotransferase, *AST* aspartate aminotransferase, *GG* glutamyl transpeptidase, *SVM* support vector machine, *RF* random forest, *DWD* distance weighted discrimination, *MOE* molecular operating environment, *DT* decision tree, *CDK* chemistry development kit, *HIAT* hepatocyte imaging assay technology

models developed from random subsets of the training set. The RP technique involves the use of a decision tree to split the training dataset into predominantly toxic or predominantly nontoxic molecules based on the independent variables. Twenty-five descriptors were selected from 1D molecular similarity scores and 2D structural information using a Monte Carlo linear regression algorithm. As a result, 151 different trees were generated with the RP approach. A compound was predicted using each of the 151 trees as being toxic or nontoxic and then the ensemble average was used to obtain the final prediction. Leave-one-out (LOO) and leave-10 %-out validation techniques yielded an overall concordance of 85 % and 76 %, respectively. The external validation of 54 compounds (23 toxic) gave a similar order of accuracy (81 %). This study showed the usefulness of the ensemble approach, using a diverse training dataset to build a model that can be applied to a broad range of chemical classes. Furthermore, a measure of predictive confidence is also supplied. However, a potential drawback of an ensemble approach is observed when the combination of models makes the method less transparent and more difficult (or impossible) to investigate the underlying mechanisms.

The next model (ID 4), developed by Cruz-Monteagudo, employed a number of different modeling methods to predict hepatotoxicity; linear discriminant analysis (LDA), artificial neural networks (ANN), and machine learning algorithms [14]. In this study, 33 compounds associated with idiosyncratic hepatotoxicity and 41 chemicals not associated with liver toxicity were collected from the literature. The models used 3D Radial Distribution Function (RDF) descriptors, which give information about interatomic distances in the entire molecule, ring types, planar and nonplanar systems, atom types, and bond distances. The best predictive performance was obtained with the LDA model, which correctly classified 86.4 % of compounds. Furthermore, based on the LDA model, a "desirability" analysis was performed in order to ascertain the characteristics, or descriptor values, that a drug candidate should have to ensure a lower idiosyncratic hepatotoxicity potential. For the external validation, two small datasets were used. The first set consisted of three pairs of chemically and pharmacologically related drugs having opposite observed toxicological profiles, including toxic troglitazone vs. nontoxic pioglitazone (insulin resistance drugs), toxic tolcapone vs. nontoxic entacapone (catechol-*O*-methyltransferase (COMT) inhibitors), and toxic clozapine vs. nontoxic olanzapine (psychotropic drugs). In this case, LDA and OneR predicted hepatotoxicity with the same accuracy of 83.3 %. The second external set was created from 13 published drugs, all hepatotoxic, and was used to validate the LDA model. Nine out of the 13 drugs were classified correctly and provide evidence that the computational approaches could be applied in early drug discovery to minimize the selection of chemicals with idiosyncratic hepatotoxicity.

Another model (ID 6) for idiosyncratic hepatotoxicity was developed by Ekins et al. [15]. They used a training set of 295 compounds (containing 158 DILI-inducers) and an external validation set of 237 molecules (114 DILI-inducers) to develop a liver toxicity prediction model using a Bayesian classification approach [37]. 2D molecular descriptors and extended connectivity functional class fingerprints of maximum diameter 6 (ECFC_6) were used to differentiate the active from inactive molecules and also to highlight chemical substructures known to be important for DILI, such as ketones, diols, and α-methyl styrene. In addition, the authors applied SMILES Arbitrary Target Specification (SMARTS) filters published by several pharmaceutical companies to all 532 molecules to evaluate whether such reactive substructures could be readily detected by any of these filters. The best predictivity was obtained for the Bayesian model which correctly classified 56.0 % of active chemicals and 66.7 % of inactive compounds. The external validation resulted in 59.9 % accuracy. Regarding the SMARTS filters, the Abbott filters resulted in more stringent classification, giving a reasonable sensitivity of 66.9 %, but a relatively low specificity of 40.3 %. A significant outcome of this study was the provision of the structural and DILI classification data that can be used as a foundation for developing future computational models, as well as filters, in the early stages of the drug development process. It is evident that approaches such as the one above are not yet capable of delivering acceptable levels of predictivity. However, their potential application of drug screening makes them of great interest.

Exploring the premise that no single learning algorithm is optimal for toxicity modeling problems, Liew et al. applied an ensemble of mixed learning algorithms and mixed features to develop a model to predict hepatic adverse effects (ID 10) [24]. The authors obtained the list of available drugs on the market from the US Food and Drug Administration (US FDA) Orange Book [38], which were then screened for adverse hepatic effects by checking the reports on adverse reaction in each drug's monograph. A final set of 1274 drugs was obtained which were split into a modeling set of 1087 and a validation set of 187 compounds. Using PaDEL descriptors [39] calculated for the training set, a total of 617 base classifiers were selected using three algorithms: support vector machine (SVM), k-nearest neighbor (kNN), and Naïve Bayes (NB). The remaining 187 compounds were divided into three different external validation sets. Two of them were aimed at verifying the model's ability to predict "severely" toxic compounds and structurally similar chemicals but of opposing toxicity status. The outcome of this was that 22 of 23 withdrawn drugs or those with black warnings were predicted correctly. However, for the structurally similar chemicals with opposite hepatotoxicity potential, only 30 % of nontoxic drugs were predicted correctly. The inability of the model to separate the non-hepatoxic

chemicals was probably due to the similarity of the true negative compounds to positive training compounds, coupled with the inherent difficulty to separate highly similar compounds by QSAR, which by definition expects that structurally related chemicals have similar activities. The third external set of 120 drugs gave the most reliable evaluation of model performance resulting in a sensitivity of 81.9 %, specificity of 64.6 % and overall accuracy of 75 %. The ensemble model was able to identify the positive compounds quite well, but it was less successful in classifying negative chemicals, especially when they were structurally similar. In general, this study again demonstrated the usefulness of an ensemble methodology when applied to large and diverse datasets similarly to the Cheng and Dixon study [10].

It is very important, especially in the case of such a complex endpoint as hepatotoxicity, to correctly annotate a drugs' potential to induce toxicity. The accuracy and utility of a predictive model depends largely on how to annotate the potential of a drug to cause hepatotoxicity in a reliable and consistent way. To address this issue, Chen et al. used the high quality US FDA-approved drug labeling DILI dataset to construct a QSAR model for hepatotoxicity (ID 12) [25]. Within this dataset most DILI-concern drugs are (1) withdrawn from the market; (2) labeled with a boxed warning; or (3) indicated in the warning and precautions section. The authors divided the 387 drugs into a training set of 197 drugs (containing 81 positives) and test dataset of 190 drugs (95 positives). They then used a Decision Tree (DT) algorithm and Mold molecular descriptors to develop a QSAR model to predict hepatotoxicity in humans. The model consisted of six decision trees using 82 descriptors. Its predictive performance was first assessed by ten-fold cross validation giving an overall accuracy of 69.7 %. Then external validation was undertaken applying the test set and two additional (independent) validation datasets: Green dataset consisting of 214 hepatotoxins and 114 drugs with no evidence of hepatotoxicity [22] and the Xu dataset consisting of 132 hepatotoxins and 109 negative compounds [40]. The accuracy obtained in each external validation was between 61.6 and 68.9 %. The external validation also showed that the drugs with consistent annotations among these three validation sets were better predicted (69.1 % accuracy) than drugs with inconsistent annotations (58.8 % correctly predicted). Finally, the applicability of the model was examined. To this aim, 2000 repetitions of cross-validation based on the training set were performed to identify therapeutic subgroups in which the QSAR model had higher or lower accuracy than the overall accuracy. As a result, 22 therapeutic subgroups with high-prediction confidence and 18 therapeutic categories with low prediction confidence were identified. Some drugs in the higher confidence subgroups, such as: analgesic, antibacterial agents and antihistamines, are well documented either to cause or

not to cause DILI. Focusing only on the therapeutic categories with high prediction confidence, the accuracy of model increased to 73.6 %. So, the therapeutic categories can be used to define the chemical structure space, where the model has better predictive power. This study demonstrates that using relatively large datasets with high quality annotations and focusing on the therapeutic subgroups where the model performs best is crucial in developing reliable predictive models, especially for very complex endpoint, such as liver toxicity.

2.1.2 Statistical Models for In Vitro General Hepatotoxicity Using Chemical Descriptors

Considering the scarcity of in vitro data, only one study employed such data to predict general hepatotoxicity (ID 3). It is not a typical in silico predictive model, as it focuses mostly on the validation of the in vitro method itself using isolated hepatocytes, which includes QSARs examining physicochemical properties of chemical congeners responsible for observed cytotoxic activity [21]. The authors investigated the molecular mechanism of hepatotoxicity for 12 halobenzenes in rat and human hepatocytes. A relatively good correlation ($r^2 = 0.90$) between LC_{50} measured in phenobarbital (PB)-induced rat hepatocytes and in vivo toxicity in PB-induced male Sprague-Dawley (SD) rats was found. Moreover, the QSAR was used to identify the metabolic activating pathway in halobenzene toxicity. It was found that toxicity in normal rat and human hepatocytes was strongly correlated with hydrophobicity ($\log P$), ease of oxidation (energy of Highest Occupied Molecular Orbital (EHOMO)) and the asymmetric charge distribution according to the arrangement of halogen substituents (dipole moment, μ). This suggests that the mechanism of toxicity is similar in both species and involves the interaction between halogens and cytochrome CYP450 for oxidation. In the case of PB-induced rat hepatocytes, halobenzene toxicity was correlated only with $\log P$ and dipole moment, but not EHOMO. This can indicate that ease of oxidation is no longer of significance in the underlying toxicity. This study is significant as it allows for better understanding of hepatotoxic mechanism(s) for that class of chemical. This knowledge is critical for the future prediction of hepatotoxicity.

2.1.3 Statistical Models for In Vivo and In Vitro General Hepatotoxicity Using Chemical Descriptors

Only a single example could be found where a combination of in vivo and in vitro data was used to develop a computational model for hepatotoxicity (ID 8) [23]. Given the success of ensemble modeling approaches previously applied, pooling together all supporting or descriptive data seems a logical step in order to try to explain and increase user confidence when predicting complex endpoints. Fourches et al. constructed a large and diverse dataset for liver toxicity using a novel approach of text mining from the published literature. The authors extracted 14,000 assertions linking compounds to different degrees, or types, of hepatotoxicity (from the cellular level to the whole organ) across different species:

including humans and rodents (mostly rat and mouse). A final dataset of 951 compounds was obtained following a data curation process. The data were classified into "class 1" consisting of 248 chemicals inducing liver effects in humans only and "class 2" consisting of 283 compounds inducing no liver toxicity in humans, but causing liver effects in rodents. The authors used hierarchical cluster analysis to identify groups of chemicals sharing similar molecular motifs corresponding to similar liver effect profiles in humans and rodents. As reported by Liew et al. [24] in their previous study, Fourches et al. again identified clusters of structurally similar molecules that possessed different liver effect profiles. This presents a significant challenge for modeling approaches fundamentally based on the premise that structurally similar compounds should act in a similar manner. It is possible that, descriptor-based approaches such as these are not sensitive enough to distinguish these compounds and opens the door to structural alert-based approaches which are discussed later in this chapter.

In addition, the authors also developed Support Vector Machine (SVM)-based models to predict whether a compound would be expected to produce adverse liver effects in humans. Predictive performance was assessed by internal and external five-fold cross-validation, giving accuracies ranging from 61.9 to 67.5 % and 55.7–72.6 % for internal and external validation, respectively. After removal of structural outliers using an implementation of the applicability domain, an accuracy of 67.8 % was obtained for an external validation dataset of 222 compounds.

Further examination of the external validation set highlighted 18 chemicals reported as liver toxicants in non-rodents only. This study confirmed low cross-species concordance of liver effects (40–45 %), which is in agreement with previous investigations [41, 42]. On the other hand, it showed the reasonably good predictivity of cheminformatics techniques using data generated by automated text mining with limited manual curation. The data mining technique seems to be feasible to search for the evidence of toxicity for compounds of interest that can be used to create in silico models.

2.1.4 Statistical Models for In Vivo Specific Hepatotoxicity Endpoints Using Chemical Descriptors

Hepatotoxicity is a complex beast, a result of multiple mechanisms, many of which are still poorly understood or are not yet known. Moreover, there are various types of liver injury which can occur, such as acute and chronic hepatocellular injuries (steatosis, necrosis, cirrhosis); cholestatic injuries; neoplasia; and elevated levels of liver serum enzymes (aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (ALP)) [8, 43]. That given, "global" modeling of general hepatotoxicity seems almost like trying to paint the Mona Lisa using only one brush with a single color. Much information would be lost. If you truly aim to be able to understand and predict hepatotoxicity with confidence, it seems logical that models should be developed for

specific endpoints of liver injury initiated by a single mechanism of action. Indeed, the focus in many areas of toxicity is shifting in the direction of trying to predict single molecular initiating events (MIEs) which then, once triggered, cause a cascade of effects leading to one or more toxicity outcomes. Such information is being termed an Adverse Outcome Pathway (AOP) (*see* also Chapter 14). Indeed, an AOP specifically for liver steatosis is one such development by the Organisation for Economic Cooperation and Development (OECD) [44]. A battery of such models used in combination would provide an incredibly powerful tool.

The US FDA conducted a three-part investigation to create a human health effects database and subsequently developed QSAR models to predict the hepatobiliary (liver enzyme disorders, cytotoxic injury, cholestasis and jaundice, bile duct disorders, gall bladder disorders) and urinary tract (acute renal disorders, nephropathies, bladder disorders, kidney function tests, blood in urine, urolithiases) adverse effects of drugs. Furthermore, they described specific properties of drugs that caused these adverse effects (ID 5) [20, 45, 46]. A dataset of about 1660 chemical structures was constructed from two pharmaceutical post-market surveillance databases maintained by the US FDA: a Spontaneous Reporting System (SRS) and an Adverse Event Reporting System SRS (AERS), and from the published literature. Five specific endpoints were considered: liver enzyme disorders, cytotoxic injury, cholestasis and jaundice, bile duct and gall bladder disorders. The authors employed four QSAR modeling programs to construct predictive models and model performance was optimized by adjusting the ratio of active to inactive drug molecules in the training sets. An average sensitivity of 39.3 % and specificity of 86.5 % was obtained in the internal leave many out (LMO) validation procedure of the four programs. To improve the low sensitivity, consensus models were constructed by a combination of two programs. This resulted in an average sensitivity and specificity of 56.2 % and 78.4 %, respectively. In the external validation of 18 new drugs, which were removed from market because of serious hepatotoxicity effects, 16 compounds were predicted correctly by at least one program, but only two drugs were assigned as hepatotoxic by all four programs. These studies demonstrated that QSAR technology is a useful (albeit data-hungry) tool providing decision support information in drug discovery. However, given its multifaceted nature, prediction of hepatotoxicity remains a significant challenge and the use of multiple models in combination could be a method of increasing performance and user confidence. Moreover, the US FDA study also provided molecular insights into the mechanisms responsible for some adverse effects, and this was investigated further in the third part of this study [46].

Rogers et al. employed the US FDA Human Liver Adverse Effects Database (HLAED) containing 490 chemicals with five

serum enzyme markers of liver toxicity: ALP, ALT, AST, lactate dehydrogenase (LDH), and γ-glutamyl transpeptidase (GGT) to build QSAR models using a kNN method (ID 7) [17]. Approximately 200 compounds covering a wide range of clinical data, structural similarity, and balanced (40/60) active/inactive ratios were selected for modeling and divided into multiple training/test and external validation sets. Since the kNN technique is based on interpolating activities of the nearest neighbors, it was necessary to introduce an applicability domain to avoid making predictions for compounds that differed substantially from the training set molecules [47]. Four hundred topological descriptors generated by MolConnZ (eduSoft LC, Ashland, VA) and 1664 Dragon descriptors (v.5.4, Talete SRL, Milano, Italy) were used to construct the models for the five endpoints as well as for the composite liver endpoint created from all five liver enzymes endpoints. Sensitivities >73 % and specificities >94 % were obtained in external validations. It was interesting to note that only three endpoints (ALT, AST, and the composite score) had a relatively broad coverage among the 490 drugs in the database. This is in agreement with the fact that ALT and AST are routine, widely used clinical chemistry biomarkers for liver toxicity. The examination of the applicability of these developed models, using three chemical databases: World Drug Index (WDI), Prestwick Chemical Library (PCL), and Biowisdom Liver Intelligence Module, showed low coverage. For example, 80 % of chemicals in the WDI database were outside the applicability domain of the models. The authors also verified the predictions for compounds from these three external datasets, by comparing model-based classification with reports in the publically available literature. For many compounds, the predictions could not be verified, because of the lack of reports of toxicity in the literature. This is a common problem encountered in many hepatotoxicity modeling studies. The lack of data is a limiting factor as is the questionable quality and relevance of what is available.

The model for the composite endpoint was also further validated using five pairs of structurally similar chemicals with opposing liver toxicity effects. The outcome of this external validation was equivocal. Two pairs were outside of the models applicability domain and only one pair was predicted correctly. Building on the similar experiences noted above, this may suggest that in some cases chemical mechanism(s) alone may not account for the toxic potential. It is possible in these cases that the differential toxicity may arise from metabolic transformations, complex disease pathways, or other risk factors dependent on genetic polymorphism and/or environmental conditions. This study clearly illustrates that the limitations of in silico methodologies result from their restricted applicability domains as well as a lack of understanding of the complexities of human risk factors and DILI pathways.

Liu et al. utilized the clinical and post-marketing data from the computer-readable side effect resource (SIDER) database [48] and identified 13 types of hepatotoxic side effects (HepSEs) based on MedDRA ontology, including bilirubinemia, cholecystitis, cholelithiasis, cirrhosis, elevated liver function tests, hepatic failure, hepatic necrosis, hepatitis, hepatomegaly, jaundice, liver disease, fatty liver, and liver function test abnormalities [16]. Firstly, these 13 side effects were used to discriminate drugs that do and do not cause DILI using the Liver Toxicity Knowledge Base Benchmark Dataset (LTKB-BD) [49] and the PfizerData [22]. For the LTKB-DB, classification accuracy was 91 %; for the PfizerData the accuracy was significantly lower (74 %). In the next step, using the SIDER database, QSAR models for every HepSEs were generated using a Bayesian methodology and these were then combined to form a DILI prediction system (DILIps) (ID 11). Finally, the authors implemented a "rule of three" (RO3) criterion (a chemical being positive in at least three HepSEs) into DILIps which increased classification accuracy. The predictive performance of DILIps was examined using three external databases: LTKB-DB, PfizerData and a dataset published by O'Brien et al. [50] and yielded prediction accuracies of 60–70 %.

Liu et al. also applied the RO3 criterion to drugs in DrugBank to investigate their DILI potential in terms of protein targets and therapeutic categories. Two therapeutic categories showing a higher risk for causing DILI were identified (anti-infective for systemic use and musculoskeletal system drugs). These findings are consistent with current knowledge that most of the anti-infective drugs are very often associated with liver injuries. One hundred thirty-four protein targets related to drugs inducing liver toxicity have been identified using pathway analysis and co-occurrence text mining with most of these targets being associated with multiple HepSEs. This study provides an interesting example of the translation of clinical observations into an in silico tool which can be used to screen and prioritize new drug candidates or chemicals and to avoid those that might cause hepatotoxicity.

In recent years, a number of new initiatives and international projects have been undertaken to develop in silico models to predict the harmful effects of chemicals to humans considering different endpoints such as liver injury. One such example is the COSMOS project [51] (belonging to the larger research initiative—SEURAT-1). The main aim of COSMOS is to develop publically available tools and workflows to predict the safety to humans following the use of cosmetic ingredients. Among them is the development of computational methods to evaluate the potential of chemicals to bind to liver X receptor (LXR), activation of which leads to liver steatosis (ID 15) [52]. Using different techniques such as molecular modeling to assess the LXR binding potential and applying PaDEL or RDKit descriptors, QSAR models based

on Partial Least Squares Discriminant Analysis (PLS-DA) were developed and implemented into the freely available KNIME Platform [52]. These models, used together with the molecular modeling methods and structural alerts as discussed within this chapter, are forming integrated in silico strategies for screening of potential steatosis inducers.

2.1.5 Statistical Models for In Vitro Specific Hepatotoxicity Endpoints Using Chemical Descriptors

Only one in silico model (ID 2) has been found that predicts in vitro specific hepatotoxicity endpoints measured by cell proliferation, lactate dehydrogenase (LDH) for membrane integrity, intracellular ATP levels for cell vitality, and levels of caspases 3 and 7 for cell apoptosis [11]. The authors applied molecular interaction fields (Idiotropic Field Orientation for Comparative Molecular Field Analysis (IFO-CoMFA)) as structural descriptors and Soft Independent Modeling of Class Analogy (SIMCA) to classify the hepatotoxicity of 654 drugs from the Sigma-RBI Library of Pharmaceutically Active Compounds (LOPAC) [11]. Each of the four assays showed good discrimination between the toxic and nontoxic chemicals. The greatest accuracy of 52 % was obtained for a hierarchical ranking model, which combined all four assays (again demonstrating that ensemble/consensus models show promise). A significant improvement in predictive performance (accuracy of 88 %) was achieved with a model constructed for a set of 27 nonsteroidal anti-inflammatory drugs (NSAIDs) using data from the LDH assay. The cross-validation confirmed the good performance of this model giving an accuracy of 71 % and 83 % for a training set of 21 NSAIDs and a test set of six NSAIDs, respectively. The poor predictivity of the global IFO-SIMCA approach for the large, diverse dataset of biologically active compounds and significant improvement for single pharmacological class chemicals' model showed that for endpoints based on specific cytotoxicity indicators only models for closely related class of chemicals may be useful. This possibly indicates that they are applicable only to a single mechanism of action within structurally related compounds. This is the main limitation of this approach, as it constricts the applicability of the model. However, local models such as this often demonstrate superior levels of predictivity, hence are useful in limited chemical space.

2.1.6 Statistical Models for In Vivo General Hepatotoxicity Using Hybrid Descriptors

Significant progress has been made in analytical and biomedical techniques in recent years which has resulted in the development of hundreds of new high-throughput screening (HTS) assays. The US Environment Protection Agencies (EPA's) Toxicity Forecaster (ToxCast) program uses these HTS assays to screen environmental chemicals for bioactivity [53, 54]. Within two phases of this program, 1057 chemicals were measured using more than 800 HTS assay endpoints including biochemical assays, cell-based assays, cell-free assays, and multiplexed transcription reporter assays. These data provide valuable information about the molecular

mechanism(s) of toxicity and help to identify the pathways related to adverse effects. Three studies using both chemical and biological descriptors have been identified. The main objective of these studies was to investigate if chemical descriptors and biological descriptors could be complementary in the prediction of hepatotoxicity.

One of the first studies applying chemical and biological descriptors to develop models for hepatotoxicity was conducted by Low et al. (ID 9) [18]. In contrast to many other in silico studies, the authors utilized only the animal data obtained from subchronic (28 days of treatment) assay in rats for 127 drugs studied in the Japanese Toxicogenomics Project [56]. The chemical was assigned as a liver toxicant if it exhibited histopathological characteristics of hepatotoxicity. Conversely, a compound was deemed non-hepatotoxic if it did not result in adverse histopathological features. When the observations were inconclusive, serum chemical indicators including ALT, AST, ALP, TBL, and gamma-glutamyl transpeptidase (GGT) were considered. The authors built conventional QSAR models using only chemical descriptors. They then applied toxicogenomic data to differentiate the hepatotoxins from non-hepatotoxins and finally hybrid hepatotoxicity classifiers were developed. For modeling purposes, statistical methodologies including: kNN, SVM, RF and Distance Weighted Discrimination (DWD) were applied using internal and a fivefold external cross-validation. The evaluation of predictivity showed that the accuracy of QSAR models based on chemical descriptors was generally poor (55–61 %). Conversely, models employing 85 selected toxicogenomics descriptors showed significantly improved predictive performance with accuracies as high as 76 %. The authors examined the spatial distribution of compounds in their chemical and toxicogenomics descriptor space which showed that 50 % of structurally similar pairs of compounds had opposing toxicities. On the other hand, amongst pairs of compounds with the most similar gene expression profilers, only 23 % exhibited opposing toxicity. It shows that pairs of compounds with similar gene expression profiles are more likely to have the same hepatotoxicity potential than pairs of chemically similar compounds. Of note here is that when hybrid models, combining both chemical and biological descriptors, were constructed they demonstrated similar accuracy (68–77 %) to those models based only on toxicogenomics data but the use of both chemical and biological descriptors provides additional insights into understanding DILI. The study confirmed that hepatotoxicity is a very complex endpoint and cannot be predicted effectively based only on the chemical characteristics of drugs. Such hybrid models look very promising as predictive and prioritization tools and allow for a better understanding of the mechanisms of hepatotoxicity.

A second study employing hybrid descriptors was conducted by Zhu et al. (ID 13) [19]. The authors constructed models based

on chemical descriptors and in vitro cell-imaging information taken from human hepatocyte imaging assay technology (HIAT) that measures the intensity of biochemical indicators, such as lipids, glutathione (GSH), reactive oxygen species (ROS) [40]. The models were built based on a dataset of 292 diverse chemicals (156 positive) using RF and fivefold cross validation methodologies. For each model the applicability domain was defined to control the distance between the predicted compound and its closest neighbor in the dataset. The main purpose of this research was comparing the prediction performance of models with a single type of descriptor (chemical or HIAT) with hybrid models. The hybrid models were constructed by combination of HIAT descriptors with chemical descriptors calculated using three programs (CDK-HIAT, Dragon-HIAT, and MOE-HIAT). These three hybrid models were combined into a consensus model. The models with chemical descriptors alone showed the poorest predictivity with accuracies between 57 % (for CDK descriptors) and 63 % (for MOE descriptors). Similar to the study conducted by Low et al. [18], this research confirmed that structural properties alone are incapable of capturing the complex mechanisms of liver toxicity. The highest accuracy (77 %) and specificity (87 %) were obtained from the HIAT model. However, the consensus hybrid model showed the greatest sensitivity (74 %). Since the HIAT model had the highest specificity and consensus model-best sensitivity, both models were applied together to distinguish liver toxicants from nontoxic chemicals. Ninety-eight of 158 DILI-inducers and 96 of 136 non-inducers were predicted correctly by both models. Careful investigation of the 39 false negative compounds revealed that at least three types of mechanisms are not captured by the models: (1) drugs that may cause liver toxicity only in high dosage, e.g., naltrexone; (2) metabolic activation, e.g., tianeptine; and (3) blockage of bile secretion, e.g., norethindrone. Ideally, QSAR models should be mechanistically interpretable to help understand the underlying mechanisms of toxicity. In this study, the distribution of molecular fragments among the toxic and nontoxic chemicals was investigated together with the analysis of biological descriptors. Forty-seven molecular fragments showed a significantly higher probability of being present in DILI-inducers than in non-inducers. Most of these fragments were associated with amine-derivatives, aromatic rings and alkyl chloride fragments. Furthermore, three of HIAT descriptors: the tetramethylrhodamine methyl ester (TMRM) intensity, ROS and a reduced intracellular GSH level were ranked as the most important indicators of DILI. These findings proved, for example, that the redox cycling of nitroaromatic drugs can generate reactive oxygen species (represented as ROS intensity HIAT descriptor) which are indicators of oxidative stress in hepatocytes. A further HIAT descriptor, TMRM, is an indicator of mitochondrial abnormality which can generate

superoxide and damage endogenous macromolecules. This study showed that chemical and biological descriptors can be complementary and enhances the prediction accuracy of hepatotoxicity and can aid in rational mechanistic interpretation.

2.1.7 Statistical Models for In Vivo Specific Hepatotoxicity Endpoints Using Hybrid Descriptors

A recent study conducted by Liu et al. utilized the in vitro bioactivity data from ToxCast together with chemical structure descriptors for 677 chemicals to predict in vivo hepatotoxicity (ID 14) [29]. Of the 677 compounds, 214 were classified as hepatotoxic based on rat liver histopathological observations in chronic studies and were categorized into three hepatotoxicity groups: (1) hypertrophy (161), (2) injury (101), and (3) proliferative lesions (99). The remaining 463 chemicals were classified as non-hepatotoxic. The authors built the models using six machine learning algorithms: LDA, NB, SVM, classification and regression trees (CART), kNN, and an ensemble of these classifiers (ENSMB). Three types of descriptors were used to build the models: 726 chemical descriptors from QikProp, OpenBabel, PaDEL, and PubChem; 125 ToxCast HTS bioactivity descriptors and hybrid descriptors (the combination of chemical and bioactivity descriptors). Because of the skewed ratio of positive to negative chemicals in every hepatotoxicity category, undersampled, balanced datasets have been prepared: 160 positive and negative chemicals for hypertrophy, 100 positive and negative chemicals for injury, and 90 positive and negative chemicals for proliferative lesions. For each of the three categories, classifiers of hepatotoxicity were built using imbalanced and balanced datasets for three types of descriptors: chemical, bioactivity, and hybrid. Predictive performance was evaluated using tenfold cross-validation and repeated 100 times. For each step in the cross-validation loop, the subset of best descriptors was filtered. The best predictive accuracy for hypertrophy (84 %), injury (80 %) and proliferative lesions (80 %) was obtained for hybrid descriptors. Using undersampled balanced datasets improved the sensitivity, but reduced the specificity of classifiers compared to the imbalanced datasets.

In general, classifiers with bioactivity descriptors have better specificity than models with chemical descriptors only, but have lower sensitivity. However, the best predictive statistics in terms of balanced accuracy, sensitivity and specificity were obtained for hybrid classifiers for both balanced and imbalanced datasets. This study showed that using both types of descriptors is more relevant for building predictive models, since they reflect the synergies between structural features, molecular mechanisms and cellular functions. The interpretation of these selected descriptors is important for the understanding of underlying mechanisms of hepatotoxicity and can help to establish the adverse outcome pathways (AOPs) as highlighted previously in this chapter. The analysis of the descriptors suggested that the classifiers may be related to AOPs

initiated by the pregnane X receptor (PXR), farnesoid X receptor (FXR), and vitamin D receptor (VDR). Overall, this study demonstrates the usefulness of HTS assays for characterizing the in vivo hepatotoxicity and the benefit of using both types of descriptors reflecting bioactivity and chemical structure.

2.1.8 Statistical Models Summary

The performance of statistical models generally suffers when predicting complex toxicity endpoints such as hepatotoxicity, a phenotype with multiple complex mechanisms and many that remain unknown. This literature review of the existing statistical models for predicting hepatotoxicity has confirmed that there is no easy solution to the problem of correctly identifying hepatotoxins. The shortage of reliable data, the lack of sensitive biomarkers and the multifaceted nature of hepatotoxicity itself, all serve to complicate an already complex problem. Since hepatotoxicity is so complex a phenomenon, it could not be predicted with high confidence based solely on the structural properties of the chemicals. It was found that the application of both chemical and biological information together and modeling specific endpoints of liver injury, initiated by a single mechanism of action rather than the effect as a whole, can significantly improve the identification of potential hepatotoxins. Moreover, multiple studies showed that the ensemble methodology that combines different models had improved the final performances when compared with the best performing individual model.

2.2 Qualitative (Expert Knowledge-Based) Models

In contrast to the quantitative models discussed up to this point, a number of qualitative approaches have also been explored. These are summarized later in this section by Table 2 following the discussion of these models.

2.2.1 Development of Structural Alerts

The development of structural alerts has been an area of considerable interest in recent years. Their transparency and ability to incorporate (or elucidate) mechanistic information offers an advantage over other, statistically derived, approaches.

2.2.1.1 Egan et al. (2004): Structural Alerts for Hepatotoxicity

Over a decade ago, Egan et al. provided an excellent review of in silico methods to predict various aspects of drug safety (ID 1 in Table 2) [5]. The authors own contribution to this review was the development of a structural alert-based approach for the prediction of liver toxicity. From a dataset of 244 drugs (54 of which were withdrawn from the market or abandoned during development owing to hepatotoxicity) a series of 74 computational alerts were developed. These alerts were based on an extensive review of the literature and were often accompanied with mechanistic reasoning for their observed hepatotoxicity. It is interesting to note that 56 of the 74 alerts were based on functional groups and were related to the formation of reactive (or otherwise toxic)

Table 2
Table summarizing expert knowledge-based models for liver toxicity

ID	Endpoint	Type and size of data	No. of structural alerts	Validation	Predictive performance	Ref
1	Hepatotoxicity	In vivo human data for 244 compounds	74 developed	No data	No data	[5]
2	Hepatotoxicity	In vivo data for 1266 compounds	38 developed	External validation using 626 chemicals	SEN (46 %), SPE (73 %), and ACC (56 %)	[22]
3	Hepatotoxicity	In vivo human data for 951 compounds	16 developed	N/A	N/A	[30]
4	Hepatosteatosis	PDB and ChEMBL	N/A	Validation using the 251 ChEMBL compounds and 951 Fourches et al. dataset	N/A	[28]
5	Hepatotoxicity	In vivo human data for 577 compounds	12 molecular fragments	Not reported	Not reported	[29]
6	Steatosis	Pharmacophore built on the three most active agonists	None—pharmacophore model	External validation using a test set of 21 agonists	N/A	[27]

metabolites. The remainder were based on whole molecule similarity and were more complex, often with limited or no mechanistic rationale. No attempt was made here to assess their predictive performance since the authors aim was to extract and investigate structural alerts for hepatotoxicity.

Unlike the statistical models in the previous section of this chapter, qualitative methods such as structural alerts are not statistically derived models. In fact, they should not be considered as "models" at all. They serve as a direct link showing that a particular molecular fragment/feature is associated with observed hepatotoxicity. No quantitative measure is provided. Interest in structural alerts is increasing. Since they are developed in an evidence-based manner and may contain mechanistic information, they are completely transparent and user confidence in their application is generally higher than that of statistical models.

This is not to say that structural alerts are simple to generate. Each structural alert must be carefully defined. Too general in nature and it will be flagged up in almost all compounds and will not

differentiate toxicity classes. Too specific (rigid) may restrict its application to a single compound and not extend to derivatives containing the actual fragment initiating the toxicity. All of this, coupled with the need to research and define mechanistic rationale makes structural alert definition a complex and time-consuming task.

Irrespective of their origins, the beauty of structural alerts is that they can be coded into computational systems which allow for rapid screening of compound libraries. Egan et al. packaged the knowledge extracted from the literature, linked this to defined structural alerts and developed a system capable of making mechanistically supported predictions of likely hepatotoxicity in humans.

2.2.1.2 Greene et al. (2010): The Interest in Structural Alerts Grows

Green et al. further develop the concept of generating structural alerts for hepatotoxicity (ID 2) [22]. The authors highlight the presence of Derek for Windows (DfW), a commercial prediction system developed by Lhasa Ltd. [56]. In recent years this has been rebranded as Derek Nexus as already introduced in Chapter 10. This knowledge-based expert system emulates human reasoning and utilizes the approach described by Egan et al. [5] to make predictions based on structural alerts and associated mechanistic knowledge. Version 8 of this software contained structural alerts for several endpoints, many of which were well established (e.g., carcinogenicity). However, at the time this study was performed, only two structural alerts for hepatotoxicity were present in DfW's knowledgebase.

Green et al. highlighted this shortfall and published a study aimed at developing a number of additional structural alerts. Importantly, this study investigated whether it is possible to use publically available data to develop structural alerts for hepatotoxic potential. This study goes into some detail of how a dataset of known hepatotoxins was divided into various chemical/therapeutic classes. This article also starts to introduce the concept of using structural similarity to generate structural alerts from clusters of structurally related compounds.

Thirty-eight new structural alerts were identified in this study based on human and/or animal data. Each was incorporated into a customized version of DfW (*see* Fig. 2) together with supporting examples and mechanistic information gathered from the literature. Importantly, these alerts were externally validated using a large Pfizer-developed dataset of 626 compounds (*see* Fig. 3 for examples of compounds containing identified alerts). The predictive performance of these alerts in the customized DfW knowledge base are summarized in Table 2.

The importance of developing structural alerts and embedding these into a tool such as DfW is clear. SARs in the form of structural alerts for complex endpoints can be elucidated from the open literature. The additional support of case studies and mechanistic rationale extracted from the literature is where a structural alert approach

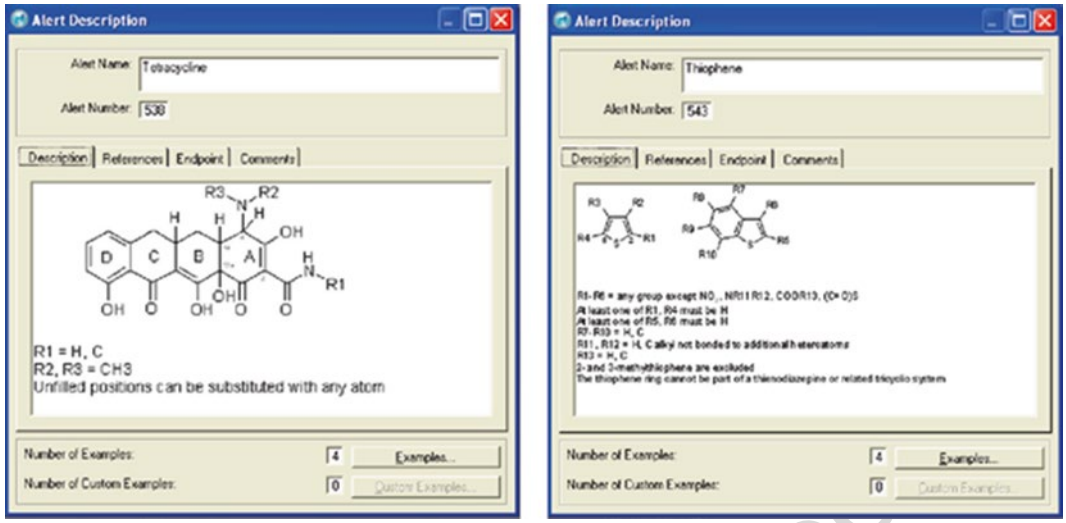


Fig. 2 Example alert describing SARs developed for tetracyclines and thiophenes. Reprinted with permission from Green et al. Chem. Res. Toxicol. 23, 1215–1222. Copyright 2015 American Chemical Society

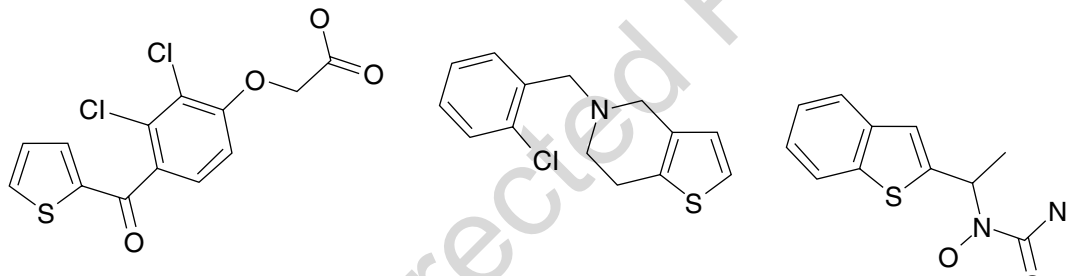


Fig. 3 Drugs containing a thiophene ring and associated with hepatotoxicity. Reprinted with permission from Green et al. Chem. Res. Toxicol. 23, 1215–1222. Copyright 2015 American Chemical Society

differs from traditional quantitative modeling. As a screening tool, a prediction along with transparent supporting evidence is very powerful. Of course, at the same time, the approach of developing structural alerts in this manner drives research into mechanisms of liver toxicity and injury which is of equal importance.

2.2.1.3 Hewitt et al. (2013): A Scheme for Generating Structural Alerts for Human Hepatotoxicity

Driven by the continued need to predict hepatotoxicity and the growing utilization of structural alerts, our contribution to this area has been in the development of a general scheme for structural alert development (ID 3) [30]. Focusing purely on publically accessible data, our aim was to develop an approach (using freely available tools) capable of yielding mechanistically supported structural alerts as previously described [5, 22]. Given the scarcity of high quality hepatotoxicity data, the broad spectrum of possible endpoints to consider and the complex nature of the mechanisms involved,

defining such alerts is a considerable challenge. Furthermore, our focus was set solely on predicting human hepatotoxicity utilizing compiled clinical data for 951 structural diverse compounds. Given that hepatotoxicity is often not evident until identified during post-marketing surveillance, it seems logical to conclude that current histopathological liver findings in rats do not model the idiosyncratic effects seen in humans [41, 42]. Conversely, Lhasa Ltd. (the developers of Derek Nexus) recently presented a poster showing that the alerts available in Derek Nexus which are developed using human data cannot predict the liver findings in rats [57].

In our study, structural similarity scores were used to highlight chemical categories of structurally related (and hepatotoxic) compounds (using the freely available Toxmatch software [58]). Eighty-two such categories were identified and each was manually inspected for validity. Following this validation step, 16 unique structural categories were identified and researched in detail to propose a mechanistic rationale. The common structural fragment of each category was extracted and taken to be the structural alert for that class. Each alert was further validated by using that structural alert to repopulate the original category. Examination of the resulting hits proved useful in highlighting alerts that were too general or restricted in terms of their definition.

An example of an alert generated from a small chemical category (Table 3) is shown in Fig. 4. This category contains a number of phenothiazine derivatives commonly used as antipsychotics. The common structural fragment was extracted and formed the structural alert as shown in Fig. 4. Searching the literature for a mechanistic rationale to explain the observed hepatotoxicity for this chemical class quickly revealed multiple implications in mitochondrial toxicity (*see* Hewitt et al. for more details). As was often the case, categories contained one or more members which were recorded as non-hepatotoxins. Here, perphenazine was classified as such in the Fourches et al. dataset. However, further literature

Table 3
Showing the category members formed using structural alert 6 (depicted)

Compound	Hepatotoxicity
Chlorpromazine	Positive
Perazine	Positive
Perphenazine	Negative
Prochlorperazine	Positive
Thioridazine	Positive
Triflupromazine	Positive

(*See* also Fig. 4)

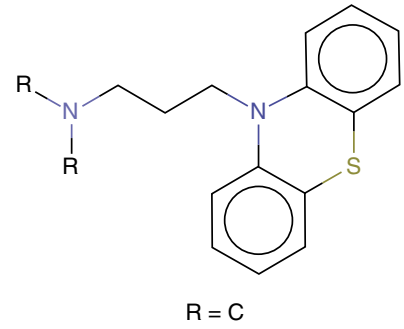


Fig. 4 Showing the category members formed using structural alert 6 (depicted) (*see* also Table 3)

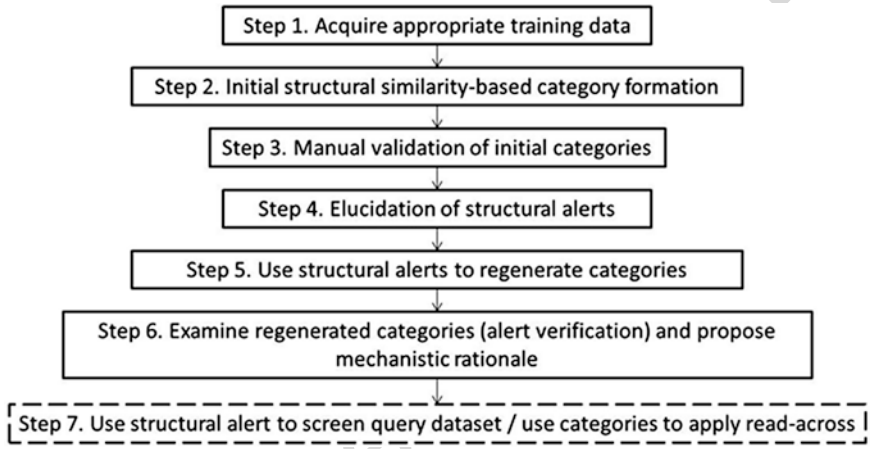


Fig. 5 Strategy for the development of structural alerts for the prediction of hepatotoxicity (taken with permission from Hewitt et al. [30])

searching suggested this to be an incorrect classification since perphenazine has been associated with liver effects in humans.

As such, this is not solely a process of extracting knowledge from a given dataset, but acts to highlight instances where the literature can be used synergistically to support and extend our current knowledge.

The aim of the article by Hewitt et al. was not to create a comprehensive suite of hepatotoxicity alerts, but to develop and publish a generic scheme for their development using freely available tools. Given the limitations of publically assessable data and our incomplete understanding of hepatotoxicity, developing a system sufficiently capable of predicting hepatotoxicity in humans is a herculean task. A dynamic scheme such as that proposed by Hewitt et al., updated regularly with new data leading to new alerts and renewed mechanistic understanding, is likely to be the most productive approach.

The general 7-step strategy proposed in this work is summarized in Fig. 5. As with all modeling approaches, the first step is to acquire

an appropriate dataset suitable for modeling (defined chemical structures, clear toxicity annotations, etc.). The second step is to form groupings of structurally related compounds (often termed chemical categories). A manual validation step is then required in order to remove any duplicate categories or those exhibiting too wide a range of chemical diversity. Step 4 is when each category is inspected and a common structural feature is identified. This feature becomes the structural alert. In order to assess the selectivity of the alerts generated, step 5 involves using these alerts to screen the original dataset. Step 6 then examines the resulting category members (which may contain compounds with the alert but not previously assigned to the category) This stage quickly highlights alerts that are too general in nature since the repopulated category tends to contain multiple new compounds (many of which often demonstrate no toxicity). If developed well, this category adds a supportive element to the alert demonstrating a category of example toxic compounds. The second stage of step 6 adds mechanistic support to the structural alert. Each alert (and its category members) is investigated in detail to define or propose a mechanistic basis for the toxicity observed. This stage is time consuming with no guarantee of success, but in most cases mechanistic rationale could be identified and this gives a much greater weighting (and user confidence) in their use. The final step proposed in the Hewitt et al. article (step 7) highlights that, at this stage, the structural alerts are read to be used to screen query datasets. Furthermore, it is stressed that the chemical categories themselves should not be forgotten and have a potential role in read-across; a process whereby measures of structural similarity can be used to match a query chemical to those in a library. These reference compounds (or category members) can then be used to estimate the properties/toxicity of the query compound based on their similarity.

As with the study by Greene et al., the power of structural alerts is their ability to be built into a platform capable of screening large numbers of compounds for the presence of each alert. The 16 alerts developed in his study were combined into a predictive tool and were made available on the predictive modeling platform developed within the eTOX Project [59]. Here, the structural alerts were coded as SMARTS and were incorporated into the KNIME platform [52]. This automated the screening procedure and allowed for an input file to be uploaded and rapidly screened.

2.2.1.4 Steinmetz et al. (2015): Focusing the Search

Working as part of the COSMOS Project, Steinmetz et al. (ID 4) [28] employed a slightly different approach to the problem. Instead of elucidating structural alerts and then investigating their mechanism(s) of action, they began with a known mechanism of interest (interaction with the retinoic acid receptor (RAR) which has been linked with liver steatosis) (It is interesting to note that the retinoid class was previously highlighted as a structural alert in

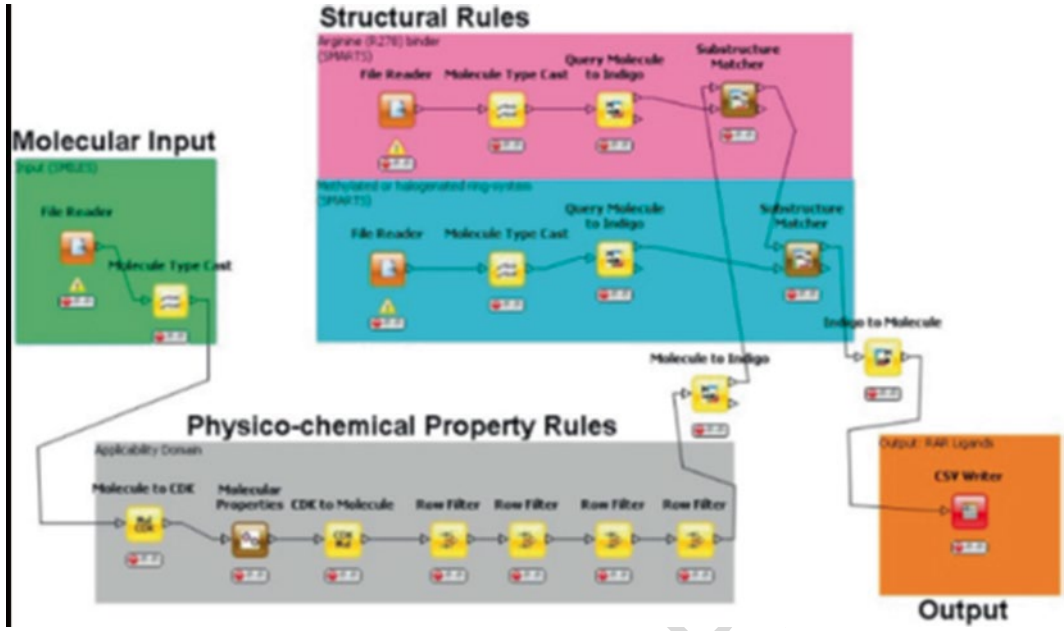


Fig. 6 KNIME workflow developed by Steinmetz et al. to predict RAR ligands (taken with permission from Steinmetz et al. [28])

Hewitt et al. [30].) Subsequent analysis then solely focuses on known RAR ligands to identify structural alerts for this mechanism of action. This is synonymous with the local versus global modeling approaches previously discussed with regards to the statistically derived models (multiple versus single mechanisms of action).

In contrast to previous works, Steinmetz at al. combined a small number of structural alerts together with a set of physicochemical property filters to highlight potential RAR ligands. These filters were based on the physicochemical characteristics of the known RAR ligands considered in the study.

Again, predictions were made via the development of a KNIME workflow containing the alerts as well as automated physicochemical property calculations and filters (*see* Fig. 6). The KNIME workflow then acts as a very powerful screening tool able to identify potential RAR ligands.

2.2.1.5 Liu et al. (2015): Boosting the Validity of Structural Alerts

The most recent example of structural alerts for human liver toxicity at the time of writing this chapter was an article by Liu et al. (ID 5) [29]. Their focus was on the validity of structural alerts. As stated in the article, a limitation of employing libraries of structural alerts is that they will effectively reduce the chemical space available for new drug discovery. Liu et al. highlight that more than half of the oral drugs currently on the market match to one or more structural alerts published for hepatotoxicity, suggesting that these alerts are either too general in their design or they are failing to take into

account other factors, such as metabolism. They go on to discuss the development of robust, statistically validated, structural alerts.

In the publication of Hewitt et al., structural alerts were often developed using categories containing both hepatotoxic and non-hepatotoxic compounds. The conflicting "non-hepatotoxic" compounds could often be rebuttled following detailed literature searches suggesting these classifications to be false. Furthermore, with the dataset considered in the Hewitt et al. study, the absence of clinical reports for hepatotoxicity lead to a non-hepatotoxic classification.

Liu et al. proposed to ensure the relationship of alert and toxicity using a statistical approach (utilizing p values) to highlight the robustness of this relationship in a quantitative manner. Alerts based on categories containing nontoxic compounds will therefore show reduced statistics and less robustness than those based solely on toxic compounds. However, as mentioned previously, it is important to ensure the validity of the nontoxic classification before proceeding in this manner.

2.2.2 Development of Pharmacophore Models

As introduced earlier in this chapter, the development of pharmacophore models is another qualitative approach to the prediction of hepatotoxicity. It is important to stress from the outset that pharmacophore models, depending upon how they are utilized, can provide quantitative information. Pharmacophore models can be seen to extend the theory of structural alerts and transform the two-dimensional representation of a structural alert into a three-dimensional scaffold, overlaid with information of important physicochemical features. (This is not to be confused with chemotypes which are effectively two-dimensional structural alerts with encoded physicochemical data).

2.2.2.1 Tsakovska et al. (2014)

Tsakovska et al., partners in the COSMOS Project, recently published a pharmacophore study focussing on a particular mechanism of action thought to be a key factor in the elucidation of liver steatosis (ID 6) [27]. As in the Steinmetz et al. study, efforts are focused onto a single mechanism of action, in this case concentrating on the activation of the peroxisome proliferator-activated receptor gamma (PPARγ).

A pharmacophore model was developed following analysis of the interactions between PPARγ and the three most active full agonists (rosiglitazone and two compounds termed compound 544 and 570). The pharmacophore was evaluated using a dataset of full agonists and the pharmacophore features were evaluated.

The structure of one of the full PPARγ agonist (rosiglitazone) is shown in Fig. 7.

The three most active agonists are aligned on top of one another to define the characteristics of the PPARγ pharmacophore (Fig. 8). In this study, four polar atoms and functional groups

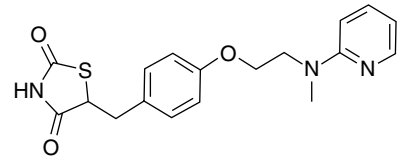


Fig. 7 Structure of rosiglitazone

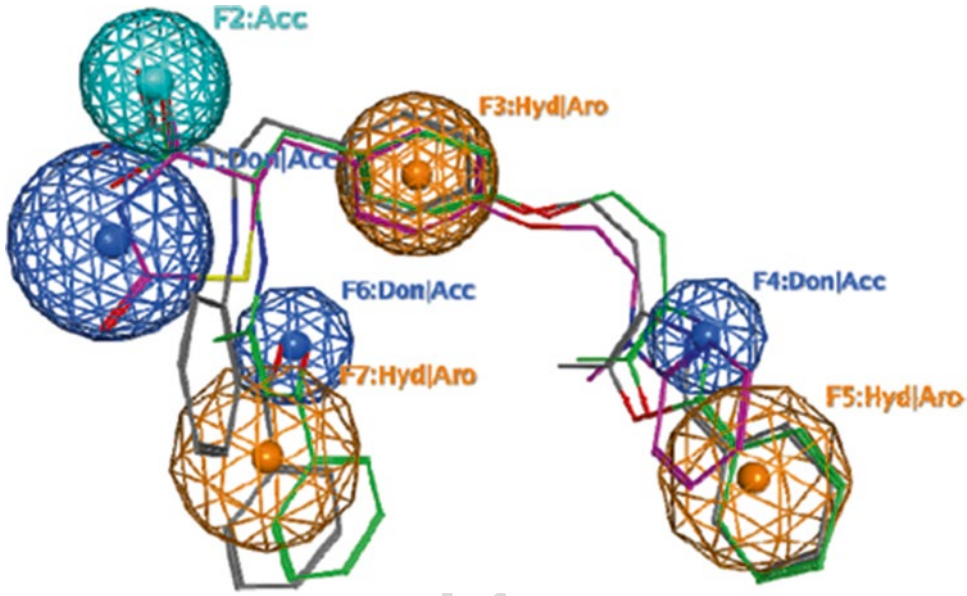


Fig. 8 Pharmacophore model of PPARγ full agonists (rosiglitazone, carbon atoms in *magenta*; compound 544, carbon atoms in *green*; compound 570, carbon atoms in *grey*) (taken from Tsakovska et al. [27])

capable of performing hydrogen bonding and ionic interactions (F1, F2, F4 and F6) and three hydrophobic and aromatic features (F3, F5 and F7) were determined to be important pharmacophore features of the most active agonists. The role of each feature and its interactions within the binding region of PPARγ are then considered.

This scaffold can be used to screen libraries of compounds for likely PPARγ binders. In its most simplistic form, the presence/absence of each pharmacophore feature can be used to predict activity. More complex application included assessment of the three-dimensional positioning of these features and the interactions these have with the PPARγ complexes.

Pharmacophore models extend beyond structural alerts in their ability to tease out information relating to the binding interactions between receptor and ligand. As such, if a particular interaction is known to be a prerequisite for activity, it can be explored and extended to find other groups/molecules which possess this ability. They therefore have a significant role in the drug development process given their possible applications in rational drug design.

3 Fitting Together the Different Pieces of the Puzzle and Future Directions

The mechanisms by which a compound can elicit toxicity to the liver are complex and diverse in nature. Attempting to then predict the hepatotoxicity of a new compound using a single approach is a very difficult task. It has already been seen that, on multiple occasions, authors have combined not only model predictions, but also model types in search of better and more reliable hepatotoxicity prediction [10, 24].

An emerging theme from all of these studies is that individual models have differing abilities to predict hepatotoxicity within a defined region of chemical space. As such, it is unlikely that a single model will ever be able predict such a complex endpoint as hepatotoxicity. Further integration of available datasets, mechanistic insights and available models for DILI is likely the only way to increase both prediction accuracy and application across chemical space. A system combining quantitative statistically derived models, structural alerts and pharmacophore models each bringing strengths (and weaknesses) is an exciting prospect and something that should be further explored. It is foreseeable that mechanistically based structural alerts could be used to screen large databases and populate a define category relating to a single mechanism of action. Local QSAR models could then be developed on this subset of data based on relevant descriptors. Moreover, it has been shown that most predictive methods discussed are based solely on descriptors of chemical structure and properties. Consideration and inclusion of biological information, such as toxicogenomics, can further help detect potential liver toxicants. Such biological descriptors may also provide further insights in the mechanisms at play in liver toxicity.

One of the major limitations currently is the lack of high quality hepatotoxicity data. To improve the prediction of potential hepatotoxins more effort should be focused towards developing specific and sensitive biomarkers for DILI. If this were possible, it would lead to more reliable hepatotoxicity data which then can be used for developing models to predict DILI. Similarly, a more detailed understanding of the mechanisms of liver injury would be of tremendous benefit and may invert the current approach of modeling with the subsequent addition of mechanistic reasoning. If we could better understand a causal mechanism of DILI (again relating to AOPs), perhaps we could design a model/alert based purely on the mechanism (e.g., what are the characteristics a chemical must possess in order to trigger mitochondrial toxicity?). These characteristics can then be used for screening and possibly further structural alert generation.

The generation of predictive systems for liver toxicity is rapidly gaining pace. With emerging modeling methods, technologies and

advances in all areas of science, it is likely that we are standing on the precipice of a modeling explosion. Careful consideration must now be made in how best to manage this emerging knowledge to best effect. In recent years, many regulatory agencies, institutions, EU Projects and working groups have established programs to help understand and detect DILI. These include the Virtual Liver Project (v-Liver™) established by US EPA [60], the Drug-Induced Liver Injury Network (DILIN) set up by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) in the USA [61], the Virtual Liver Network project initiated by the German Federal Ministry for Education and Research [62], and multiple EU Projects such as Mechanism based Integrated systems for the prediction of Drug Induced Liver Injury (MIP-DILI) [63]. Whilst duplication of effort is inevitable to some degree, what must be ensured is that both data and knowledge generated by these initiatives is shared. Just as combining models to form an ensemble seems to be beneficial for predictive performance, it is likely that a combined international ensemble effort is the only way we can successfully begin to tackle the prediction of liver toxicity in humans.

4 Conclusions

Hepatotoxicity has been a problem for many years. Unfortunately, the same can also be said for predictive models aimed at predicting these effects. It is only in the past decade that models/systems for predicting hepatotoxicity have started to emerge. It is fair to say that the modeling community are currently limited by the amount and quality/reliability of the data available to them. Coupled with an endpoint as complex as hepatotoxicity, the scale of the challenge is obvious. That said, it can be seen from the models discussed in this chapter that progress is being made, our knowledge of the processes behind liver toxicity is growing and our ability to tackle this problem is increasing. Given the diversity of the modeling approaches seen in these studies and the general transition towards ensemble/consensus approaches in this area, it is likely that the next decade will be equally as productive.

Acknowledgement

The funding provided by the eTOX project, grant agreement number 115002, under the Innovative Medicines Initiative Joint Undertaking (IMI-JU), the European Community's Seventh Framework Program (FP7/2007-2013) COSMOS project under grant agreement number 266835, and from Cosmetics Europe is gratefully acknowledged.

References

1. Przybylak KR, Cronin MTD (2012) In silico models for drug-induced liver injury—current status. Expert Opin Drug Metab Toxicol 8:201–217
2. Schuster D, Laggner C, Langer T (2005) Why drugs fail—a study on side effects in new chemical entities. Curr Pharm Des 11:3545–3559
3. Holt MP, Ju C (2006) Mechanisms of drug-induced liver injury. AAPS J 8:E48–E54
4. Kaplowitz N (2005) Idiosyncratic hepatotoxicity. Nat Rev Drug Discov 4:489–499
5. Egan WJ, Zlokarnik G, Grootenhuis PDJ (2004) In silico prediction of drug safety: despite progress there is abundant room for improvement. Drug Discov Today 1:381–387
6. Patlewicz G, Dimitrov SD, Low LK et al (2007) TIMES-SS-a promising tool for the assessment of skin sensitization hazard. A characterization with respect to the OECD validation principles for (Q)SARs and an external evaluation for predictivity. Regul Toxicol Pharmacol 48:225–239
7. Benigni R, Bossa C (2008) Structure alerts for carcinogenicity, and the salmonella assay system: a novel insight through the chemical relational databases technology. Mutat Res 659:248–261
8. Zimmerman HJ (1999) Hepatotoxicity: the adverse effects of drugs and other chemicals on the liver. Lippincott Williams & Wilkins, Philadelphia, PA
9. Li AP (2002) A review of the common properties of drugs with idiosyncratic hepatotoxicity and the "multiple determinant hypothesis" for the manifestation of idiosyncratic drug toxicity. Chem Biol Interact 142:7–23
10. Cheng A, Dixon SL (2003) In silico models for the prediction of dose-dependent human hepatotoxicity. J Comput Aided Mol Des 17:811–823
11. Clark RD, Wolohan PR, Hodgkin EE et al (2004) Modelling in vitro hepatotoxicity using molecular interaction fields and SIMCA. J Mol Graph Model 22:487–497
12. Marchant CA (2006) Virtual ADMET assessment. In: Testa B, Turski L (eds) Target selection and maturation. IOS Press, Amsterdam, p 237
13. Marchant CA, Fisk L, Note RR et al (2009) An expert system approach to the assessment of hepatotoxic potential. Chem Biodivers 6:2107–2114
14. Cruz-Monteagudo M, Cordeiro MN, Borges F (2008) Computational chemistry approach for the early detection of drug-induced idiosyncratic liver toxicity. J Comput Chem 29:533–549
15. Ekins S, Williams AJ, Xu JJ (2010) A predictive ligand-based Bayesian model for human drug-induced liver injury. Drug Metab Dispos 38:2302–2308
16. Liu Z, Shi Q, Ding D, Kelly R et al (2011) Translating clinical findings into knowledge in drug safety evaluation—drug induced liver injury prediction system (DILIps). PLoS Comput Biol 7(12):e1002310. doi:10.1371/journal.pcbi.1002310
17. Rodgers AD, Zhu H, Fourches D et al (2010) Modeling liver-related adverse effects of drugs using k-nearest neighbor quantitative structure-activity relationship method. Chem Res Toxicol 23:724–732
18. Low Y, Uehara T, Minowa Y et al (2011) Predicting drug-induced hepatotoxicity using qsar and toxicogenomics approaches. Chem Res Toxicol 24:1251–1262
19. Zhu XW, Sedykh A, Liu SS (2014) Hybrid in silico models for drug-induced liver injury using chemical descriptors and in vitro cell-imaging information. J Appl Toxicol 34:281–288
20. Matthews EJ, Ursem CJ, Kruhlak NL et al (2009) Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: part B. Use of (Q)SAR systems for early detection of drug induced hepatobiliary and urinary tract toxicities. Regul Toxicol Pharmacol 54:23–42
21. Chan K, Jensen NS, Silber PM, O'Brien PJ (2007) Structure–activity relationships for halobenzene induced cytotoxicity in rat and human hepatoctyes. Chem Biol Interact 165:165–174
22. Greene N, Fisk L, Naven RT et al (2010) Developing structure-activity relationships for the prediction of hepatotoxicity. Chem Res Toxicol 23:1215–1222
23. Fourches D, Barnes JC, Day NC et al (2010) Cheminformatics analysis of assertions mined from literature that describe drug-induced liver injury in different species. Chem Res Toxicol 23:171–183
24. Liew CY, Lim YC, Yap CW (2011) Mixed learning algorithms and features ensemble in hepatotoxicity prediction. J Comput Aided Mol Des 25:855–871
25. Chen M, Hong H, Fang H et al (2013) Quantitative structure-activity relationship models for predicting drug-induced liver injury based on FDA-approved drug labeling annotation and using a large collection of drugs. Toxicol Sci 136:242–249

26. Liu J, Mansouri K, Judson RS et al (2015) Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. Chem Res Toxicol 28:738–751
27. Tsakovska I, Al Sharif M, Alov P et al (2014) Molecular modelling study of the PPARγ receptor in relation to the mode of action/adverse outcome pathway framework for liver steatosis. Int J Mol Sci 15(5):7651–7666
28. Steinmetz FP, Mellor CL, Meinl T et al (2015) Screening chemicals for receptor-mediated toxicological and pharmacological endpoints: using public data to build screening tools within a KNIME workflow. Mol Inform 34:171–178
29. Liu R, Yu X, Wallqvist A (2015) Data-driven identification of structural alerts for mitigating the risk of drug-induced human liver injuries. J Cheminform 7:4
30. Hewitt M, Enoch SJ, Madden JC et al (2013) Hepatotoxicity: a scheme for generating chemical categories for read-across, structural alerts and insights into mechanism(s) of action. Crit Rev Toxicol 43(7):537–555
31. Tralau T, Oelgeschläger M, Gürtler R et al (2015) Regulatory toxicology in the twenty-first century: challenges, perspectives and possible solutions. Arch Toxicol 89:823–850
32. Golbraikh A, Tropsha A (2002) Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. J Comput Aided Mol Des 16:357–369
33. Williams A, Tkachenko V, Lipinski C et al (2010) Free online resources enabling crowdsourced drug discovery. Drug Discov World 10:33–39
34. Breiman L, Friedman JH, Olshen RA et al (1984) Classification and regression trees. Wadsworth International Group, Belmont, CA
35. Hawkins DM, Kass GV (1982) Automatic interaction detection. In: Hawkins DH (ed) Topics in applied multivariate analysis. Cambridge University Press, Cambridge, UK, pp 269–302
36. Dixon SL, Villar HO (1999) Investigation of classification methods for the prediction of activity in diverse chemical libraries. J Comput Aided Mol Design 13:533–545
37. Xia X, Maliski EG, Gallant P, Rogers D (2004) Classification of kinase inhibitors using a Bayesian model. J Med Chem 47:4463–4470
38. Orange book: approved drug products with therapeutic equivalence evaluations. http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm
39. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32(7):1466–1474
40. Xu JJ, Henstock PV, Dunn MC et al (2008) Cellular imaging predictions of clinical drug-induced liver injury. Toxicol Sci 105:97–105
41. Olson H, Betton G, Stritar J et al (1998) The predictivity of the toxicity of pharmaceuticals in humans from animal data. An interim assessment. Toxicol Lett 10:535–538
42. Olson H, Betton G, Robinson D et al (2000) Concordance of the toxicity of pharmaceuticals in humans and in animals. Regul Toxicol Pharmacol 32:56–67
43. Farrell GC (1994) Drug-induced liver disease. Churchill Livingstone, New York
44. https://aopkb.org/index.html
45. Ursem CJ, Kruhlak NL, Contrera JF et al (2009) Identification of structure activity relationships for adverse effects of pharmaceuticals in humans. Part A: use of FDA post-market reports to create a database of hepatobiliary and urinary tract toxicities. Regul Toxicol Pharmacol 54:1–22
46. Matthews EJ, Kruhlak NL, Benz RD et al (2009) Identification of structure-activity relationships for adverse effects of pharmaceuticals in humans: Part C. Use of QSAR and an expert system for the estimation of the mechanism of action of drug-induced hepatobiliary and urinary tract toxicities. Regul Toxicol Pharmacol 54:43–65
47. Tropsha A, Golbraikh A (2007) Predictive QSAR modelling workflow, model applicability domains, and virtual screening. Curr Pharm Des 13:3494–3504
48. Kuhn M, Campillos M, Letunic I et al (2010) A side effect resource to capture phenotypic effects of drugs. Mol Syst Biol 6:6
49. Chen M, Vijay V, Shi Q, Liu Z, Fang H et al (2011) FDA-approved drug labelling for the study of drug-induced liver injury. Drug Discov Today 16:697–703
50. O'Brien PJ, Irwin W, Diaz D et al (2006) High concordance of drug-induced human hepatotoxicity with in vitro cytotoxicity measured in a novel cell-based model using high content screening. Arch Toxicol 80:580–604
51. http://cosmostox.eu
52. http://knimewebportal.cosmostox.eu/webportal/#/Public/Nuclear%20Receptor%20Binding/LXR%20Binding%20Potential
53. Berthold MR, Cebron N, Dill F et al (2008) KNIME: the Konstanz Information miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L, Decker R (eds) Studies in classification, data analysis, and knowledge organization. Springer, Berlin

54. Kavlock RJ, Chandler K, Houck KA et al (2012) Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 25:1287–1302
55. Judson RS, Houck KA, Kavlock RJ et al (2010) In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118: 485–492
56. Uehara T, Ono A, Maruyama T et al (2010) The Japanese toxicogenomics project: application of toxicogenomics. Mol Nutr Food Res 54:218–227
57. Lhasa Ltd (2015) Analysis of human and in vivo data for hepatotoxicity modelling. http://www.lhasalimited.org/Public/Library/2015/Analysis%20of%20human%20and%20in%20vivo%20data%20for%20hepatotoxicity%20modelling.pdf
58. Ideaconsult Ltd (2012) Toxmatch structural similarity tool (version 1.07). http://ihcp.jrc.ec.europa.eu/our_labs/computational_toxicology/qsar_tools/toxmatch
59. http://www.etoxproject.eu/
60. http://www.epa.gov/ncct/virtual_liver
61. Fontana RJ, Watkin PB, Bonkovsky HL et al (2009) DILIN Study Group, Drug-Induced Liver Injury Network (DILIN) prospective study: rationale, design and conduct. Drug Saf 32:55–68
62. http://www.virtual-liver.de
63. http://www.mip-dili.eu/

Chapter 12

In Silico Models for Ecotoxicity of Pharmaceuticals

Kunal Roy and Supratik Kar

Abstract

Pharmaceuticals and their active metabolites are one of the significantly emerging environmental toxicants. The major routes of entry of pharmaceuticals into the environment are industries, hospitals, or direct disposal of unwanted or expired drugs made by the patient. The most important and distinct features of pharmaceuticals are that they are deliberately designed to have an explicit mode of action and designed to exert an effect on humans and other living systems. This distinctive feature makes pharmaceuticals and their metabolites different from other chemicals, and this necessitates the evaluation of the direct effects of pharmaceuticals in various environmental compartments as well as to living systems. In this background, the alarming situation of ecotoxicity of diverse pharmaceuticals have forced government and nongovernment regulatory authorities to recommend the application of in silico methods to provide quick information about the risk assessment and fate properties of pharmaceuticals as well as their ecological and indirect human health effects. This chapter aims to offer information regarding occurrence of pharmaceuticals in the environment, their persistence, environmental fate, and toxicity as well as application of in silico methods to provide information about the basic risk management and fate prediction of pharmaceuticals in the environment. Brief ideas about toxicity endpoints, available ecotoxicity databases, and expert systems employed for rapid toxicity predictions of ecotoxicity of pharmaceuticals are also discussed.

Key words Database, Ecotoxicity, Endpoints, Expert system, In silico, Pharmaceuticals, QSAR

1 Introduction

A significant amount of pharmaceuticals and their metabolites have been found in the various environmental compartments causing damage to the environment and hazard to the living systems. Due to an increase in application of human and veterinary medicines manyfold, pharmaceuticals and their metabolite residues have been found in rivers, sewage effluents, streams and in surface, ground, and potable water, creating a big concern for the ecologists [1]. The primary routes of entrance of pharmaceuticals into the environment are domestic, hospital, and industrial wastes [2]. Pharmaceuticals are excreted in urine or feces as a mixture of unchanged chemicals and metabolites and enter into the environment through septic tank and sewage systems [1]. On the other

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_12, © Springer Science+Business Media New York 2016

hand, ecotoxicity data of pharmaceuticals are available in the literature for less than 1 % of the drugs, and only a small number of pharmaceuticals and their residues have been subjected to risk assessment employing ecotoxicological tests.

Pharmaceuticals are intentionally designed to have a specific mode of action and exert an effect on specific organs, tissues, cells, or biomolecules in humans, mammals, or other vertebrates, and many of them are persistent in the body [3]. As a consequence, when pharmaceuticals and their unaltered metabolites enter into the environment by different means, they can affect humans as well as other living species. There are many drugs whose specific effects or modes of action are not well known, and they often produce effects through several modes of action. These distinguished features make pharmaceuticals dissimilar from others and this is the sole reason to assess the potential acute and chronic effects of pharmaceuticals in diverse environmental compartments. It is quite apparent that the toxic effects of pharmaceuticals on diverse organisms in aquatic as well as nonaquatic environment are due to their long persistent and bio-accumulative nature [4]. In view of the serious issue of pharmaceutical toxicity to the environment, it is vital to categorize the proper source, occurrence, effects, and fate of each individual pharmaceutical product as well as to perform the risk assessment and risk management of ecotoxicological effects of the pharmaceutical chemicals and their metabolites [1, 2].

Antibiotics are one of the majorly used pharmaceuticals in human and veterinary medicines. The world consumption of antibiotics has risen radically in the last decade, also increasing the elimination of their metabolites in their original form. Most antibiotics are poorly metabolized after ingestion, probably resulting in a fraction of antibiotics from 25 to 75 % leaving the bodies in an unaltered form after consumption [5]. Additionally, a high percentage of the antibiotics added to the animal feed are excreted in urine or manure. In some cases, as much as 90 % of the antibiotic administered orally may pass through the animal unchanged and excreted in urine and manure. Thereafter, these antibiotics can enter surface and groundwater and be strongly adsorbed in soils and are not readily degradable [6]. Vidaver [7] estimates that 53,000 ha of fruit and vegetable plants are sprayed annually with antibiotics. For example, streptomycin and oxytetracycline are registered by the US Environment Protection Agency (USEPA) for use in plant agriculture. Utilization of transgenic plants to produce inexpensive antibiotics may also be a cause of environmental hazards due to the existence of crop residues, roots, and root exudates in the soil which can act as a continuous source of residual antibiotics to soil fauna and flora [8].

While pharmaceuticals and their metabolite residues are detected in rivers, streams, sewage influents and effluents, surface,

ground, and potable waters [9], it may be noted that the drinking water treatment methods reduce residues, but they are incapable of removing the contaminant pharmaceuticals absolutely. According to a nationwide study of "emerging pollutants" in waters, the US Geological Survey (USGS) tested for pharmaceuticals in 139 rivers in 30 states of the USA, detecting diverse therapeutic classes of biologically active compounds [10]. The cardiovascular drug propranolol has been reported downstream from the sewage treatment plant [11]. The antiepileptic drugs carbamazepine and clofibrate are two most persistent pharmaceuticals which have been detected in the environment [2]. Major detected drugs in rivers were beta blockers (e.g., metoprolol up to 1.54 μg/l) and beta-sympathomimetics, estrogens (e.g., 17β-estradiol up to 0.013 μg/l) [12], analgesic and anti-inflammatory drugs (e.g., Diclofenac up to 1.2 μg/l) [13], and also antibiotics (e.g., erythromycin up to 1.7 μg/l) [12], as well as lipid-lowering agents (e.g., clofibrinic acid up to 0.2 μg/l) [14] and antiepileptic drugs (e.g., carbamazepine up to 2.1 μg/l) [13]. Presence of clofibric acid, propylphenazone, and diclofenac has been reported in the drinking water of Berlin in the concentration range of several hundreds of nanograms per liter [15]. Paracetamol, diclofenac, and carbamazepine were monitored in drinking water in Southern France [16], and clofibric acid and diazepams were detected in treated drinking water in Milan, Italy [17]. Psychoactive and illicit drugs amphetamine, cocaine and its metabolite benzoylecgonine, morphine, 6-acetylmorphine, 11-nor-9-carboxy-delta-9-tetrahydrocannabinol, methadone and its main metabolite 2-ethylidene-1,5-dimethyl-3,3-diphenylpyrrolidine have been detected in surface and waste waters [18]. Schultz and Furlong found highest concentrations of antidepressant drugs venlafaxine, citalopram, and bupropion 1000 ± 400 ng/l, 90 ± 20 ng/l, and 60 ± 40 ng/l, respectively, in samples collected downstream from a water reclamation plant [19]. The maximum determined concentration of fluoxetine was 0.099 ng/l in wastewater treatment plant (WWTP) effluents in Canada [20].

Nonprescription drugs like caffeine, cotinine, and acetaminophenone are found in samples of potable water collected near Atlanta, Georgia [21]. Tauber detected carbamazepine and gemfibrozil in drinking waters in ten cities in Canada that were examined for a 44-drug subset consisting pharmaceuticals including sulfonamides, quinolones, tetracyclines, and macrolide antibiotics [22]. Oraine and Pettigrove identified and quantified ibuprofen (0.93 μg/l) and ibuprofen methyl ester (4.95 μg/l) in finished water in alarming quantity [23]. Median concentrations of 0.02 μg/l and 0.12 μg/l were reported for ciprofloxacin and norfloxacin, respectively, for samples from 139 surface streams across the USA. Ciprofloxacin in the range 0.7–124.5 μg/l was found in wastewater of a Swiss hospital [24]. Hellweger et al. [25] claimed

that environmental concentrations of tetracycline in surface waters are usually less than 0.11 mg/l, although higher values of up to 6.8 mg/l have been observed. Estrogens, a sex hormone, have been detected in plasticizers and preservatives, while 17α-ethinylestradiol (EE2) used as a component of contraceptive pills has been identified in ground and tap water samples [26].

The presence of human and veterinary pharmaceuticals and their residues into the environment has impelled the introduction of different risk assessment guidelines in the European Union by the European Medicines Evaluation Agency (EMEA) and in the USA by the Food and Drug Administration (FDA). According to the European Commission guideline [27], a medicinal product for human use must be accompanied by environmental risk assessment data. The EMEA has released a guideline for the assessment of potential environmental risks in 2006 [28]. According to the US FDA guidelines for the risk assessments of human drugs, applicants have to provide an environmental assessment report when the expected concentration of the active pharmaceuticals in the aquatic environment is ≥1 μg/l [29]. Additionally, the FDA Center for Drug Evaluation and Research (CDER) issued a guidance document "*Guidance for Industry for the Submission of an Environmental Assessment in Human Drug Application and Supplements*" in 1995 [30]. In case of veterinary medicines, environmental risk assessments have been required in the USA since about 1980 and Europe since 1997 [31].

The need for a practical approach in gathering data on the environmental toxic effects of pharmaceuticals has been identified by the European Union Commission's scientific committee on toxicity, ecotoxicity, and environment (CSTEE). The four classes of special environmental feature-specific concerns, which are stereotypically not evaluated in traditional ecotoxicity testing under EU directive 1488/94 [28] are antibiotics [resistance issue], antineoplastics [mutagenicity], sex hormones [endocrine disruption], and cardiovascular high potential hazard. Therefore, it is acknowledged that a prioritization technique needs to be developed for environmental risk assessment of pharmaceuticals, and this should follow the general scheme for chemicals according to the REACH guidelines [27], where the implication of in silico methods specifically the quantitative structure–activity relationship (QSAR) method is stressed.

In this perspective, to make the information regarding ecotoxicity of diverse pharmaceuticals available, different government and nongovernment regulatory authorities are recommending the application of fast and economical in silico methods for prediction of the elementary physicochemical and fate properties of pharmaceuticals as well as their ecological and direct human health effects before they reach into market for usage. Computer-aided toxicity models allow for the effects of pharmaceuticals (physicochemical

properties, toxicological activity, distribution, fate, etc.) to be easily predicted. These predictions may be obtained from the knowledge of chemical structure alone, provided that the structure can be described in two or three dimensions. Employing these methods, ecotoxicity information on pharmaceuticals may be obtained without toxicity testing, and/or even before synthesis of the compound. Therefore, use of QSAR as one of the non-experimental methods is significant in order to lessen time, animal usage and cost involvement in design, development, and discovery process of drugs and/or pharmaceuticals.

There is a significant lack of knowledge about the environmental fate of a huge number of pharmaceuticals and their metabolites. On the contrary, only a limited number of in silico models have been developed so far to predict the risk of pharmaceuticals to the environment. This chapter aims to provide information regarding occurrence of pharmaceuticals and their residues in the environment, their persistence, environmental fate, and toxicity as well as application of in silico methods to predict risk and fate properties of pharmaceuticals to the environment. Concise ideas about ecotoxicity endpoints, available ecotoxicity databases and expert systems employed for rapid ecotoxicity predictions of pharmaceuticals are discussed in this chapter.

2 Ecotoxicity of Pharmaceuticals: A General Overview

2.1 Source and Entry Routes

Identification of proper sources and routes of entry of pharmaceuticals into diverse environmental compartments is the first step to get a proper view of the ecotoxicity problem due to pharmaceuticals. The most obvious and common pathways for environmental contamination of pharmaceuticals are discussed below.

(a) *Urine and feces*: Major and most common entry routes for pharmaceuticals into the environment are via urine and feces of the patients. Not only active ingredients, but also the metabolites are excreted through the urine and feces as many drugs are metabolized into hydrophilic compounds for excretion. The risk of these metabolites is completely different from the parent drugs in majority of cases which make the risk assessment study more critical one.

(b) *Direct exposure of diagnostic compounds*: Contrast media like diatrizoate, iohexol, iomeprol, and iopromide are used as diagnostic tools for capturing detailed X-ray images of soft tissues. Iodinated X-ray contrast media are highly hydrophilic substances which are extensively applied and eliminated without proper treatment; as a result they persist for a long time in the environment [32].

(c) *Household disposal*: Either out-of-date or unwanted medicines are discarded through the sink/toilet or via waste collection, before being taken to landfill sites where they appear as terrestrial ecosystem contaminants. Less than 20 % users had ever been given instructions about medication dumping by a health care provider. In a study, causes for possessing unused medication were found to be due to an alteration of medication by the doctor (48.9 %), or self-discontinuation (25.8 %) [33]. The most common method of disposal was to throw unused medicines in the trash (76.5 %) or flush them down the drain (11.2 %) [33].

(d) *Manufacturers*: According to the regulation of the Good Manufacturing Practices (GMP), the active pharmaceutical emissions during manufacturing have been thought to be insignificant. But recently it has been found that in Asian countries concentrations up to several milligrams per liter can be found in effluents for single compounds [34].

(e) *Hospital influent and effluent*: Point sources such as hospital effluents are likely to be another significant source. There are up to 16 pharmaceuticals including antiepileptics and anti-inflammatories which were found in the hospital waste water according to a study [35]. Several studies suggested the existence of the pharmaceuticals in the effluent and influent of the sewage treatment plants and it was proved that the elimination of the pharmaceuticals is partial [35].

(f) *Animal husbandry and veterinary medicine*: Veterinary medicines and their metabolites are also excreted through urine and feces. Apart from the potential for direct soil contamination, there is also the risk of run-off with heavy rain, thus potentially contaminating both the surrounding surface and groundwater. Other sources include direct application in aqua farming, manure run-off, run-off from the application of sewage sludge and manure on farmland as fertilizers, or, finally, via landfill leaching [36].

(g) *Aquaculture*: Sewage Treatment Plant (STP) sludge is habitually employed as fertilizer on agricultural land which is a rich source of non-suspected drugs [37]. According to the Food and Agriculture Organization (FAO), antibiotics have been utilized in aquaculture primarily for therapeutic purposes and as prophylactic agents. Antibiotics authorized for use in aquaculture are florfenicol, oxytetracycline, sarafloxacin, premix, erythromycin sulfonamides potentiated with ormethoprim, or trimethoprim [38].

(h) *Plant agriculture*: Antibiotics are comprehensively employed to control bacterial diseases of plants. Streptomycin with oxytetracycline to a minor extent is very commonly used antibiotic in

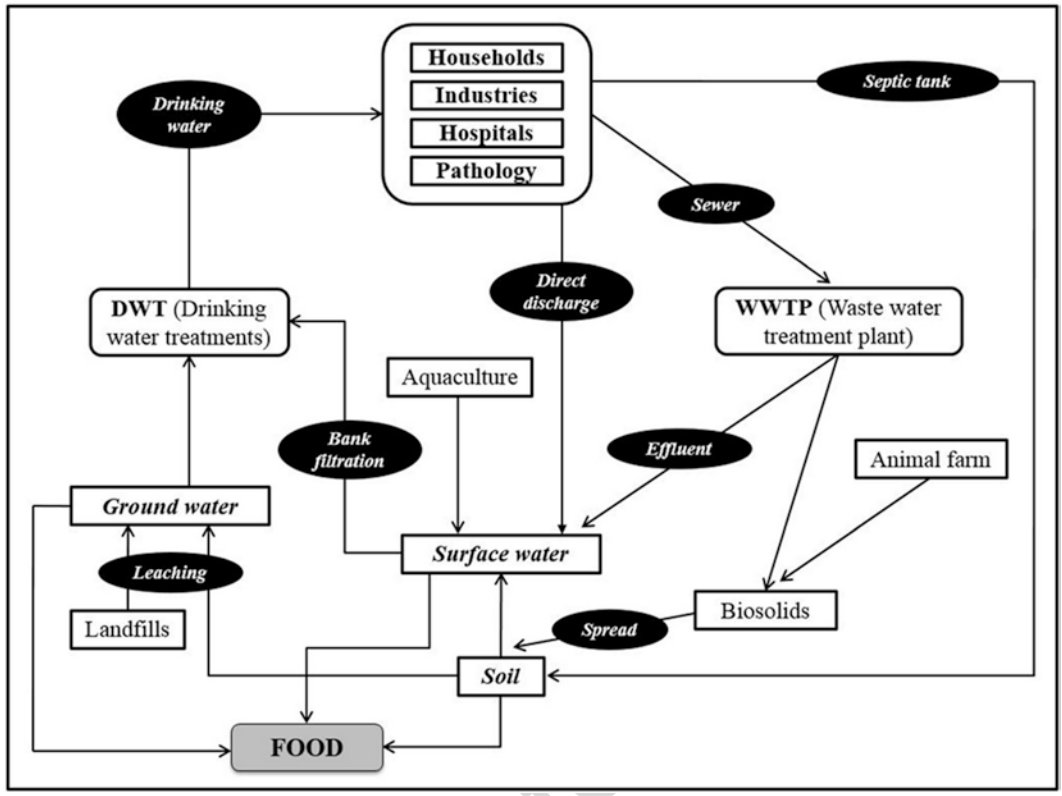


Fig. 1 Common sources, routes and fate of pharmaceuticals

plant agriculture in controlling bacterial diseases of tree fruits. Primary uses are on apple, pear, and related fruit trees for the control of fire blight caused by *Erwinia amylovora*. According to a report, antibiotics applied to plants account for less than 0.5 % of total antibiotic use in the USA [39]. In Fig. 1, we have represented different sources, routes, fate of pharmaceuticals.

2.2 Occurrence

Pharmaceuticals are among the most common personal care products in day to day life. Medicines are regularly used in human and veterinary health care, farming, and aquaculture in the modern era. Country specific consumption for groups of drugs in defined daily doses (DDDs) can be found for Europe on the European Surveillance of Antimicrobial Consumption (ESAC) homepage [40]. In the last decade, a large number of studies covering occurrence of pharmaceuticals in water bodies, sewage treatment plants, manure, soil, and air dust have been published. The most concerning issue is that under the environmental conditions, these molecules can be neutral, cationic, anionic, or zwitterionic which make the risk assessment study of pharmaceuticals more difficult. In Table 1 we have presented the reported concentrations of

t1.1 **Table 1**

t1.2 **Various therapeutic classes of pharmaceuticals and their reported concentration in different samples of different countries as well as toxicological endpoints and**

t1.3 **probable ecotoxicity data**

t1.4 t1.5 Class of drugs	Name of drugs	Country	Sample	Concentrations reported (ng/l)	Toxicological endpoint	Ecotoxicity data (mg/l)	Ref.
t1.6 NSAIDs	Acetylsalicylic acid	Romania	River water	<30–37.2 (±4.6)	*D. subspicatus* EC_{50} (growth inhibition)	106.7	[41–45]
t1.8		Japan	STP influent	470–19,400	*D. magna* EC_{50} (48 h) (immobilization)	88.1	[45]
t1.10	Salicylic acid	Canada	STP influent River Water	554.3–2178.2, 130.4–371.5	*V. fischeri* EC_{50} (30 min)	90	[46]
t1.12	Diclofenac	Spain	STP influent	200–3600	*D. magna* EC_{50} (48 h) (immobilization)	68	[41]
t1.14		Switzerland	STP influent	1300–2900	*D. subspicatus* EC_{50} (growth inhibition)	72	[41]
t1.16		Canada	STP influent	32–448	*L. minor*, EC_{50} (7 days) (growth inhibition)	7.5	[46]
t1.18		Greece	STP influent	12–560	*O. mykiss* LOEC (28 days) (cytological alterations)	0.001	[47]
t1.21		Germany	Groundwater	590	*D. subspicatus* EC_{50} (growth inhibition)	71.9	[45]
t1.23		USA	Drinking water	<0.25	*P. subcapitata* NOEC (96 h) (growth inhibition)	10	[44]
t1.26		UK	STP influent	901–1036	*D. magna* EC_{50} (48 h) (immobilization)	22.43	[44]
t1.28	Fenoprofen	Japan	STP influent	9.68–80.6	–	–	[45]
t1.29	Ibuprofen	Spain	STP influent	34,000–168,000	*D. magna* EC_{50} (48 h) (immobilization)	108	[41]
t1.31		Switzerland	STP influent	1750–4500	*D. subspicatus* EC_{50} (growth inhibition)	315	[41]

t1.33		Canada	STP influent	2235.2–6718.3	*L. minor*	22	[41]
t1.34					EC_{50} (7 days) (growth inhibition)		
t1.35		Sweden	STP influent	3590	NOEC (14 days) (survival)	20	[47]
t1.36		Italy	River water	78.50	*Gammarus pulex*	0.01	[48]
t1.37					LOEC (behavior)		
t1.38		USA	Groundwater	3110	*L. minor*	4.01	[49]
t1.39					EC_{50} (7 days) (growth inhibition)		
t1.40		UK	STP influent	7741–33,764	*O. latipes*	>100	[50]
t1.41					LC_{50} (96 h) (mortality)		
t1.42		South Korea	STP influent	10–137	*Hydra attenuata*	22.36	[41]
t1.43					LC_{50} (96 h) (morphology)		
t1.44		Germany	River water	8.7–32	NOEC (21 days) (survival)	5.36	[50]
t1.45	Indomethacin	Spain	STP effluent	160–390	*T. platyurus*	16.14	[3]
t1.46					LC_{50} (24 h) (mortality)		
t1.47		South Korea	River water	<1–33.5	*O. latipes*	81.92	[3]
t1.48					LC_{50} (96 h) (mortality)		
t1.49	Ketoprofen	Canada	STP effluent	8–351	–	–	[46]
t1.50		Sweden	STP effluent	940	–	–	[47]
t1.51		Spain	STP effluent	131	–	–	[51]
t1.52		Germany	River water	<26	–	–	[51]
t1.53		USA	STP effluent	23 (±6.8 %)	–	–	[51]
t1.54	Mefenamic acid	UK	STP influent	136–363	*T. platyurus*	3.95	[52]
t1.55					LC_{50} (24 h) (mortality)		
t1.56		Japan	STP influent	4.45–396	*O. latipes*	8.04	[52]
t1.57					LC_{50} (96 h) (mortality)		
t1.58		China	River water	ND to 22.4	–	–	[52]
t1.59				(±3.1)			
t1.60		Spain	STP effluent	40–60	–	–	[52]
t1.61	Naproxen	Canada	STP effluent	271.4–7962.3	*D. magna*	174	[46]
t1.62					EC_{50} (48 h) (immobilization)		
t1.63		Sweden	STP influent	3650	*L. minor*	24.2	[47]
t1.64					EC_{50} (7 days) (growth inhibition)		
t1.65		USA	Drinking water	<0.5	*T. platyurus*	84.09	[53]
t1.66					LC_{50} (24 h)		

(continued)

Table 1
(continued)

Class of drugs	Name of drugs	Country	Sample	Concentration reported (ng/l)	Toxicological endpoint	Ecotoxicity data (mg/l)	Ref.
		Spain	STP influent	109–455	*P. subcapitata* EC_{50} (72 h) (growth inhibition)	31.82	[53]
		USA	River water	31 (±5.5 %)	*B. calyciflorus* EC_{50} (48 h) (growth inhibition)	0.56	[53]
		South Korea	STP effluent	20–483	*D. magna* EC_{50} (48 h) (immobilization)	166.3	[45]
	Paracetamol	Spain	STP influent	29,000–246,000	*V. fischeri* EC_{50} (15 min)	567.5	[54]
		USA	Groundwater	380	*D. magna* EC_{50} (48 h) (immobilization)	30.1	[54]
		UK	Surface water	<50	*D. rerio* (*zebrafish*) LC_{50} (48 h)	378	[3]
		UK	STP influent	5529–69,570	*O. latipes* LC_{50} (48 h)	>160	[54]
		Taiwan	Hospital effluent	62,250	*D. magna* EC_{50} (96 h) (immobilization)	26.6	[54]
		South Korea	STP effluent	1.8–19	*S. subspicatus* EC_{50} (72 h)	134	[54]
Blood lipid lowering agents	Bezafibrate	Italy	River water	0.79–2.75	EC_{50} (96 h) (morphology)	25.85	[55]
		Brazil	River water	<25	*Hydra attenuata* LC_{50} (96 h) (morphology)	70.71	[55]
		Spain	STP effluent	40–130	LOEC (96 h) (morphology)	1	[55]
	Clofibric acid	Brazil	Drinking water	<10–30	*D. subspicatus* EC_{50} (growth inhibition)	115	[41]
		Italy	River water	0.41–5.77	*L. minor* EC_{50} (7 days) (growth inhibition)	12.5	[48]

		UK	STP influent	<20–651	*S. subspicatus* EC_{50} (72 h)	89	[48]
		Spain	STP influent	25–58	*D. magna* EC_{50} (immobilization)	106	[51]
		Greece	STP influent	ND	*D. magna* EC_{50} (48 h) (immobilization)	72	[51]
		Germany	Groundwater	2–40	*D. magna* EC_{50} (48 h) (immobilization)	>200	[3]
	Gemfibrozil	Canada	STP effluent	80.1–478.2	*Hydra attenuata* LC_{50} (96 h) (morphology)	22.36	[55]
		Sweden	STP influent	710	LOEC (96 h) (morphology)	1	[55]
		USA	Drinking water	0.43	NOEC (96 h) (morphology)	0.1	[55]
		China	River water	ND to 22.4 (±3.1)	*V. fischeri* EC_{50} (24 h) (bioluminescence)	64.6	[56]
		South Korea	STP effluent	3.9–17	*V. fischeri* EC_{50} (48 h) (bioluminescence)	45.1	[56]
		Spain	STP effluent	470–3550	*D. magna* EC_{50} (48 h) (immobilization)	42.6	[56]
	Atorvastatin	USA	Drinking water	<0.25	*L. gibba* LOEC (7 days) (growth parameters)	0.3	[57]
		Canada	STP influent	76 (±3)	–	–	[57]
	Lovastatin	Canada	STP influent	49 (±2)	–	–	[57]
	Pravastatin	Canada	STP influent	117 (±6)	–	–	[57]
	Simvastatin	Canada	STP influent	4 (±0)	–	22.8	[57]
Antibiotics	Ciprofloxacin	USA	Surface water	20	–	–	[42]
		USA	WWTP effluent	100–160	–	–	[68]
		USA	STP influent	ND to 1000	–	–	[58]
		Germany	Surface water	60	–	–	[68]
		Germany	WWTP effluent	600	–	–	[68]

(continued)

Table 1
(continued)

Class of drugs	Name of drugs	Country	Sample	Concentration reported (ng/l)	Toxicological endpoint	Ecotoxicity data (mg/l)	Ref.
		Switzerland	Surface water	5–18	–	–	[68]
		Switzerland	WWTP effluent	55–405	–	–	[68]
		France	WWTP effluent	60	–	–	[68]
		Italy	River water	ND to 26.15	–	–	[48]
		Sweden	WWTP effluent	30	–	–	[68]
		Sweden	STP influent	90–300	–	–	[58]
	Enrofloxacin	Portugal	STP influent	121.8–447.1	*V. fischeri* EC_{50} (15 min) (luminescence)	326.89	[58]
		Japan	WWTP influent	7–85	–	–	[68]
		USA	STP influent	250	*D. magna* EC_{50} (48 h) (immobilization)	131.7	[58]
	Levofloxacin	South Korea	River water	ND to 87.4 (±13)	*D. magna* EC_{50} (21 days) (reproduction)	0.34	[52]
		Japan	WWTP influent	255–587	–	–	[68]
	Norfloxacin	USA	Surface water	120	*S. obliquus* IC_{50} (48 h) (growth inhibition)	38.49	[42]
		Portugal	STP influent	191.2–455.0	*S. capricornutum* EC_{50} (growth inhibition)	16.6	[59]
		Sweden	STP influent	72–174	NOEC (growth inhibition)	4.01	[59]
		China	Surface seawater	<13	NOEC (growth inhibition)	4.02	[59]
		China	WWTP influent	460	–	–	[68]
		China	WWTP effluent	85–320	–	–	[68]

		Japan	WWTP influent	155–486	–	–	[68]
	Ofloxacin	USA	STP influent	ND to 1000	*T. platyurus* LC50 (24 h) (mortality)	33.98	[60]
		USA	WWTP effluent	110–1000	–	–	[68]
		Portugal	STP influent	ND	*D. magna* EC_{50} (24 h) (immobilization)	31.75	[60]
		Sweden	STP influent	<6–287	*C. dubia* EC_{50} (24 h) (immobilization)	17.41	[60]
		China	Harbor seawater	5.2–10	*C. dubia* EC_{50} (7 days) (growth inhibition)	3.13	[60]
	Naldixic acid	Japan	WWTP influent	7–40	–	–	[68]
		Taiwan	STP influent	26–372	–	–	[3]
	Ampicillin	Taiwan	Hospital effluent	21	*V. fischeri* EC_{50} (15 min) (luminescence)	2627	[58]
	Penicillin G	China	STP influent	153,000 ± 4000	*M. aeruginosa* EC_{50} (growth rate)	0.006	[43]
	Cephalexin	China	Surface seawater	<13–182	–	–	[60]
	Lincomycin	USA	Surface water	60	*B. calyciflorus* LC_{50} (24 h) (mortality)	24.94	[60]
		USA	Groundwater	320	*T. platyurus* LC_{50} (24 h) (mortality)	30.00	[60]
		Italy	River water	3.13–248.90	*B. calyciflorus* EC_{50} (48 h) (growth inhibition)	0.68	[60]
	Clarithromycin	Italy	River water	0.49–20.30	*B. calyciflorus* LC_{50} (24 h) (mortality)	35.46	[60]
		Taiwan	STP influent	59–1433	*B. calyciflorus* EC_{50} (48 h) (growth inhibition)	12.21	[60]
		South Korea	River water	ND to 443 (±14)	*D. magna* EC_{50} (24 h) (immobilization)	25.72	[52]

(continued)

Table 1
(continued)

Class of drugs	Name of drugs	Country	Sample	Concentration reported (ng/l)	Toxicological endpoint	Ecotoxicity data (mg/l)	Ref.
	Erythromycin	Italy	Po River water	1.4–15.9	*L. minor* EC_{50} (7 days) (growth inhibition)	5.62	[3]
		South Korea	STP effluent	8.9–294	*T. platyurus* LC_{50} (24 h) (mortality)	>100	[3]
	Sulfachloropyridazine	Korea	STP influent	<30–476	*V. fischeri* EC_{50} (15 min)	26.4	[3]
	Sulfadiazine	Italy	River water	236	*M. aeruginosa* EC_{50} (72 h) (growth inhibition)	0.135	[3]
	Sulfadimethoxine	USA	Surface water	60	*V. fischeri* EC_{50} (15 min)	>500	[54]
		USA	Groundwater	46–68	*D. magna* EC_{50} (48 h) (immobilization)	248	[54]
		Taiwan	Hospital effluent	ND	*D. magna* EC_{50} (96 h) (immobilization)	204.5	[54]
		Luxembourg	STP influent	0.3–6	*O. latipes* LC_{50} (48 h)	>100	[54]
		Italy	River water	28	*S. capricornutum* EC_{50} (growth inhibition)	2.30	[59]
	Sulfamethazine	USA	Groundwater	76–215	*V. fischeri*, EC_{50} (15 min)	344.7	[54]
		USA	STP influent	160	*O. latipes*, LC_{50} (96 h)	>100	[54]
		Luxembourg	STP influent	0.3-2	*O. latipes* LC_{50} (48 h)	>100	[54]
	Sulfamethoxazole	USA	Surface water	150	*V. fischeri* EC_{50} (15 min)	78.1	[54]
		USA	Groundwater	1110	*D. magna* EC_{50} (48 h) (immobilization)	189.2	[54]
		USA	Drinking water	0.32	*D. magna* EC_{50} (96 h) (immobilization)	177.3	[54]
		Taiwan	STP influent	179–1760	*O. latipes* LC_{50} (96 h)	562.5	[54]

		South Korea	STP effluent	3.8–407	LOEC (96 h) (morphology)	10	[55]
		Sweden	STP influent	<80–674	EC_{50} (48 h) (growth inhibition)	9.63	[60]
		Italy	Drinking water	13–80	*T. platyurus*	35.36	[60]
					LC_{50} (24 h) (mortality)		
	Sulfathiazole	Luxembourg	STP influent	0.3–2	*V. fischeri*	>1000	[3]
					EC_{50} (15 min)		
		South Korea	STP influent	<30–531	*D. magna*	35	[3]
					LOEC (21 days) (reproduction)		
	Chlortetracycline	USA	Surface water	420	*M. aeruginosa*	0.05	[43]
					EC_{50} (growth rate)		
		Taiwan	Hospital effluent	ND	*S. capricornutum*	3.1	[43]
					EC_{50} (growth rate)		
	Oxytetracycline	USA	Surface water	340	*Hydra attenuata*	>100	[55]
					LC_{50} (96 h) (morphology)		
		Italy	River water	ND to 19.2	*Hydra attenuata*	40.13	[55]
					EC_{50} (96 h) (morphology)		
	Tetracycline	USA	Surface water	110	*L. minor*	1.06	[60]
					EC_{50} (7 days) (growth inhibition)		
		Taiwan	STP influent	46–234	*S. capricornutum*	2.2	[43]
					EC_{50} (growth rate)		
		Luxembourg	STP influent	0.3–85	*D. magna*	340	[58]
					NOEC (48 h) (immobilization)		
		China	Surface seawater	<13–122	–	–	[58]
	Metronidazole	Taiwan	STP influent	1–294	*D. magna*	1000	[3]
					LOEC (48 h) (immobilization)		
	Trimethoprim	USA	Drinking water	<0.25	*D. magna*	167.4	[54]
					EC_{50} (48 h) (immobilization)		
		Serbia	River water	25	*D. magna*	120.7	[54]
					EC_{50} (96 h) (immobilization)		
		South Korea	STP effluent	10–188	*O. latipes*	>100	[54]
					LC_{50} (96 h)		
		China	Surface seawater	<13–21.8	*D. magna*	149	[58]
					EC_{50} (48 h) (immobilization)		

(continued)

Table 1
(continued)

Class of drugs	Name of drugs	Country	Sample	Concentration reported (ng/l)	Toxicological endpoint	Ecotoxicity data (mg/l)	Ref.
Sex hormones	17α-Estradiol	USA	Surface water	30	–	–	[42]
		France	Groundwater	0.8–3.5	–	–	[56]
	17β-Estradiol	USA	Surface water	9	*O. latipes* NOEC (21 days)	<0.0293	[42]
		Japan	STP influent	13.3–25.8	–	–	[62]
		China	Rivers water	ND to 7.5 (±0.4)	–	–	[62]
		South Korea	STP effluent	<1.0	–	–	[62]
		Germany	STP influent	11.8 (±5.1)	–	–	[62]
		France	Groundwater	0.3–1.3	–	–	[62]
	Estriol	USA	Surface water	19	–	–	[3]
		Italy	STP influent	23–48	–	–	[3]
		South Korea	STP effluent	8.9–25	–	–	[3]
	Estrone	USA	Surface water	27	–	–	[3]
		USA	Drinking water	<0.20	–	–	[3]
		Japan	STP influent	28.7–197	–	–	[3]
	17α-Ethinylestradiol	USA	Surface water	73	*P. promelas* LOEC (21 days) (plasma VTG induction)	0.000001	[63]
		USA	Drinking water	<1.0	*P. promelas* LOEC (21 days) (ultrastructure testes)	0.000001	[63]
		Germany	STP influent	8.8 (±8.0)	–	–	[63]
		Italy	STP influent	ND	–	–	[63]
		France	Groundwater	0.5–3.0	–	–	[61]

t1.276	Antiepileptic	Carbamazepine	Spain	STP influent	120–310	*D. magna* EC_{50} (48 h) (immobilization)	>100	[41]
t1.278			Finland	STP influent	290–400	*D. subspicatus* EC_{50} (growth inhibition)	74	[41]
t1.280			Romania	River water	<30–75.1 (±6.1)	*L. minor* EC_{50} (7 days) (growth inhibition)	25.5	[41]
t1.282			Sweden	STP influent	1680	*Gammarus pulex* LOEC (behavior)	0.00001	[47]
t1.284			Germany	Groundwater	900	*T. platyurus* LC_{50} (24 h) (mortality)	>100	[52]
t1.286			USA	Drinking water	6.8	*O. latipes* LC_{50} (96 h)	45.87	[52]
t1.288			Japan	STP influent	14.9–270	*O. latipes* LC_{50} (48 h)	35.4	[54]
t1.290			South Korea	Drinking water	<1.0	*V. fischeri* EC_{50} (30 min)	>81	[44]
t1.292		Carbamazepine-10,11-epoxide	Spain	STP influent	300–500	–	–	[3, 41]
t1.293			France	STP influent	ND-27	-	-	[3, 41]
t1.294	β-Blockers	Atenolol	Finland	STP influent	510–800	*T. platyurus* LC_{50} (24 h) (mortality)	>100	[52]
t1.296			Sweden	STP influent	30	*O. latipes* LC_{50} (96 h)	>100	[52]
t1.298			Italy	River water	3.44–39.43	*D. subspicatus* EC_{50} (growth inhibition)	620	[48]
t1.300			USA	Drinking water	0.47	*D. magna* EC_{50} (48 h) (immobilization)	313	[64]
t1.302			Spain	Hospital effluent	100–122,000	*P. promelas* NOEC (28 days) (growth)	3.2	[64]
t1.304			South Korea	River water	ND to 690 (±26)	*P. promelas* LOEC (28 days) (growth)	10	[64]

(continued)

Table 1
(continued)

Class of drugs	Name of drugs	Country	Sample	Concentration reported (ng/l)	Toxicological endpoint	Ecotoxicity data (mg/l)	Ref.
	Metoprolol	Finland	STP influent	980–1350	*D. magna* EC_{50} (48 h) (immobilization)	>100	[41]
		Sweden	STP influent	160	*D. subspicatus* EC_{50} (growth inhibition)	7.3	[47]
		Taiwan	STP influent	14–597	*L. minor* EC_{50} (7 days) (growth inhibition)	>320	[41]
	Sotalol	Finland	STP influent	640–830	–	–	[3]
		Germany	Ground water	560	–	–	[3]
	Propranolol	Sweden	STP influent	50	*D. magna* EC_{50} (48 h) (immobilization)	7.5	[47]
		Taiwan	Hospital effluent	54	*D. subspicatus* EC_{50} (growth inhibition)	5.8	[41]
		UK	STP influent	60–119	*L. minor*, EC_{50} (7 d) (growth inhibition)	114	[41]
		Spain	Hospital effluent	200-6500	*T. platyurus* LC_{50} (24 h) (mortality)	10.31	[52]
		South Korea	River water	ND to 40.1 (±3)	*O. latipes* LC_{50} (96 h)	11.40	[52]
Antidepressants	Fluoxetine	USA	Surface water	12	*H. azteca* LOEC (28 days) (growth)	0.1	[65]
		USA	Groundwater	56	*H. azteca* NOEC (28 days) (growth)	0.033	[65]
		USA	Drinking water	0.64	*D. magna* NOEC (21 days) (newborns' lenght)	0.0089	[65]
		South Korea	STP effluent	1.7	*D. magna* LOEC (21 days) (newborns' lenght)	0.031	[65]
		Norway	STP influent	0.4–2.4	*P. antipodarum* NOEC (reproduction)	0.013	[65]

		Canada	STP influent	3.1 (±0.3)–3.5	*Gammarus pulex* LOEC (behavior)	0.0001	[65]
	Norfluoxetine	USA	Drinking water	0.77	–	–	[66]
		Norway	STP influent	0.7 (±13.1)–9.3	–	–	[66]
		Canada	STP influent	1.8 (±0.3)–4.2	–	–	[66]
	Fluvoxamine	Norway	STP influent	0.4–3.9	*P. subscapitata* IC_{50} (96 h) (growth inhibition)	4.003	[3]
	Paroxetine	Norway	STP influent	0.6–12.3	*D. magna* EC_{50} (48 h) (immobilization)	2.5	[3]
	Sertraline	Norway	STP influent	1.8–2.5	*V. fischeri* EC_{50} (30 min) (inhibition)	10.72	[66]
		Canada	STP influent	6.0 (±0.4)	*V. fischeri* LOEC (30 min) (inhibition)	4.5	[66]
Antineoplastic	Cyclophos-phamide	Romania	River water	<30–64.8 (±8.0)	*P. subcapitata* EC_{50} (72 h) (growth inhibition)	>100	[67]
		Italy	STP influent	<1.9–9.0	*P. subcapitata* NOEC (72 h) (growth inhibition)	>100	[67]
		Switzerland	STP influent	2.0–6	*D. magna* LOEC (21 days) (reproduction)	100	[67]
	Methotrexate	Italy	STP influent	<0.83–12.6	*V. fischeri* EC_{50} (30 min)	1220	[3]
	Tamoxifen	UK	STP influent	143–215	*B. calyciflorus* LC_{50} (24 h) (mortality)	0.97	[67]

(continued)

Table 1
(continued)

Class of drugs	Name of drugs	Country	Sample	Concentration reported (ng/l)	Toxicological endpoint	Ecotoxicity data (mg/l)	Ref.
X-ray contrast media	Diatrizoate	Germany	STP effluent	250	–	–	[3]
	Iohexol	Australia	STP influent	2800–4760	–	–	[3]
	Iopamidol	Germany	Ground water	300	–	–	[3]
		Australia	STP influent	400–620	–	–	[3]
	Iopromide	South Korea	STP effluent	1170–4030	*D. magna* EC_{50} (48 h) (immobilization)	>1000	[3]
		Germany	STP effluent	4400	*V. fischeri* EC_{50} (30 min)	>10,000	[3]
		Australia	Ground water	168	*S. subspicatus* EC_{50} (72 h) (growth inhibition)	>10,000	[3]
		USA	STP influent	ND to 17	*P. putida* EC_{10} (16 h) (growth inhibition)	>10,000	[3]
		Spain	STP influent	6600	*D. rerio* NOEC (28 days)	>100	[3]
	Iomeprol	Australia	STP influent	<730	–	–	[3]

LOEC lowest-observed-effect concentration, *NOEC* no-observed-effect-concentration, *STP* sewage treatment plant, *WWTP* waste water treatment plants

diverse pharmaceuticals from various therapeutic classes in different samples of different countries and probable ecotoxicity data to particular toxicological endpoints [3, 41–68].

2.2.1 Waterbodies

The presence of pharmaceuticals in the various waterbodies in the environment has been quite extensively studied by different research groups. Quinolones (predominantly ciprofloxacin) and other pharmaceuticals have been detected in the effluent of hospitals up to a low μg/l range. Another study reveals that β-lactams (including penicillins, cephalosporins, carbapenems, monobactams, β-lactamase inhibitors) were detected in the lower μg/l range in hospital effluent and in the influent of a municipal STP [69]. NSAIDs have the higher concentrations recorded in surface water, ranging between 0.4 ng/l and 15 μg/l, diclofenac, paracetamol, and ibuprofen being the most quantitatively found [70]. Drugs like caffeine with a maximum concentration of 6 μg/l and sulfamethoxazole with 1.9 μg/l in the USA, carbamazepine up to 1.3 μg/l in Germany and in Canada, gemfibrozil up to 790 ng/l, ranitidine up to 580 ng/l, atenolol with 241 ng/l in Italy, and metformin up to 150 ng/l are detected in surface water [71]. In the effluent of WWTP and STP, the concentrations of estrogenic compounds usually are below 50 ng/l, but there are unexpected high concentrations of estriol and 17α-estradiol (about 590 ng/l and 180 ng/l respectively) found in the USA [72].

2.2.2 Manure and Soil

Antibiotics have been detected in soil in concentrations in the mg/kg range [73]. Generally, the concentrations of pharmaceuticals detected in the soils are quite low when compared with that of pharmaceuticals in water resource. According to the literature, the six most common pharmaceuticals found in soil are the antibacterials (trimethoprim, sulfadiazine, and triclosan), analgesics (ibuprofen and diclofenac) and antiepileptic (Carbamazepine). Extensive studies have detected tetracyclines and sulfonamides in liquid manure at concentrations of up to 20 and 40 mg/l, respectively. Antibiotics like virginiamycin, sarafloxacin, tetracycline, oxytetracycline, chlortetracycline, and cyclosporine A have quite slow biodegradability in soil. Tylosin disappeared soon after the application of manure. Hamscher et al. [74] detected tetracycline and chlortetracycline in 10 out of 12 soil samples. The highest average concentration of 86.2 μg/kg (0–10 cm), 198.7 μg/kg (10–20 cm), 171.7 μg/kg (20–30 cm) tetracycline, and 4.6–7.3 μg/kg (in all three sublayers) chlortetracycline were found. Carbamazepine is the most frequent compound detected in soil among five studies [75].

2.2.3 Air Dust

Several comprehensive reports have been published on environmental concentrations of antibiotics in dust originating from a pig-fattening house [76]. In a large-scale pig production, veterinary antibiotics are hugely used. This production system is represented as a considerable source of dust.

2.3 Effects

2.3.1 Antibiotics

Pharmaceuticals may have potential adverse effects on aquatic and terrestrial organisms by directly reaching into the environment. Organisms like bacteria, fungi, and microalgae are primarily affected as antibiotics are designed to inhibit the microorganisms. Antibiotics have the potential to affect the microorganisms in sewage systems and waste water treatment plant too. The inhibition of wastewater bacteria may seriously affect organic matter degradation and nitrification process which is a vital step in wastewater purification and elimination of toxic ammonia [77]. Lincomycin showed significant inhibition of the nitrification activity [78]. Ciprofloxacin was found to be active against *Vibrio fischeri* at a concentration of 5 mg/l [79]. Thomulka and McGee [80] have performed two bioassays to evaluate the toxicity of antibiotics like novobiocin, chloramphenicol, tetracycline, ampicillin, and streptomycin to *Vibrio harveyi,* and approximately no toxic effects were identified after short incubation times where the employed endpoint was luminescence. Common receptors have been identified in plants for a number of antibiotics affecting transcription and translation (tetracyclines, macrolides, lincosamides, aminoglycosides, and pleuromutilins), metabolic pathways such as folate biosynthesis (sulfonamides), chloroplast replication (fluoroquinolones), and fatty acid biosynthesis (triclosan) [81].

Antimicrobials can affect the degradation of organic matter in large extent as well as have effects upon sediment's microbial community [82]. Strong inhibitory effects on several bacteria and diminution in the length of the hyphae of lively molds in forest soil have been observed when antibiotics are added in concentrations of 10 mg/kg soil. A transitory effect on sulfate reduction was detected when antibiotics were mixed to sediment [83]. Allergic risks may arise from the high exposure of antibiotics dust particle in the air. Tylosin and sulfamethazine, which occurred in 80 % and 65 % of the samples respectively, are drugs with known allergic potential. Therefore, the high incidence of the asthma disease occurred among children living on farms. A survey on dust in pig fattening buildings in Europe exposed an average concentration of inhalable airborne dust of 2.2 mg/m^3 [84]. Chloramphenicol is extensively employed in farming resulting in severe hazardous effects including myelosuppression to farmers; that is why it was totally banned for food-producing animals within the EU and the USA in 1994 [85].

Another important aspect is the emergence of resistance due to enormous application of antibiotics in human medicine, veterinary medicine, and animal husbandry. Resistance is one of the most concerning issue in medical field due to its accumulating and accelerating nature. On the contrary, the techniques combating resistance are diminishing in power and number. Antibiotics in sub-inhibitory concentrations can have an influence on cell

functions and modify the genetic expression of virulence factors or the transfer of antibiotic resistance. The most prominent medical examples are vancomycin-resistant enterococci (VRE), methicillin-resistant *Staphylococcus aureus* (MRSA), and multiresistant pseudomonads [86].

2.3.2 Analgesics and Nonsteroidal Anti-inflammatory Drugs (NSAIDs)

Cleuvers [45] evaluated that acute toxicities of NSAIDs were relatively low, with half-maximal effective concentration (EC_{50}) values obtained using *Daphnia* in the range from 68 to 166 mg/l and from 72 to 626 mg/l in the algal test. With EC_{50} values of 23.6 mg/l (ibuprofen), 23.8 mg/l (diclofenac), and 38.2 mg/l (naproxen), chronic ecotoxicity was somewhat higher, but still the values are far above the concentrations detected in surface water. A prominent confirmation of diclofenac residues in dead cattle has been observed in Pakistan [87]. Only in Germany, in 2002, 93.5 million prescriptions for NSAIDs were made with a transaction volume of about 1562 million Euros [88]. Due to higher usage and pharmacokinetic and pharmacodynamic properties, analgesics and anti-inflammatory drugs can reach considerable (up to >1 μg/l) concentrations in the environment. Few NSAIDs are detected in very low doses even in drinking water. Reports suggested the presence a concerning amount of diclofenac and ibuprofen in Swiss lakes and rivers, as well as in water bodies from the UK, Spain, Brazil, Greece, and the USA [15].

Diclofenac seems to be the compound having the highest acute toxicity with the effective concentrations below 100 mg/l within the class of NSAIDs. Short-term acute toxicity was analyzed in algae and invertebrates, phytoplankton was found to react more sensitively [lowest EC_{50} (96 h) = 14.5 mg/l] than zooplankton [lowest EC_{50} (96 h) = 22.43 mg/l] [89]. Diclofenac is commonly found in wastewater at median concentration of 0.81 μg/l whereas the maximal concentration in wastewater and surface water is up to 2 μg/l [90]. Acetylsalicylic acid affected reproduction in *D. magna* and *D. longispina* at concentrations of 1.8 mg/l [90]. Water flea *Daphnia magna* population growth rate was considerably reduced for concentrations ranging from 0 to 80 mg/l due to chronic toxicity of ibuprofen. Acute toxicity tests showed that naproxen had LC_{50} and EC_{50} values within the 1–100 mg/l range for the water flea *Ceriodaphnia dubia*, the rotifer *Brachionus calyciflorus*, and the fairy shrimp *Thamnocephalus platyurus*. But the most sensitive reported species was *D. magna* for which EC_{50} values were 30.1 or 50 mg/l. Another most commonly prescribed NSAID is paracetamol which is present in concentration below to 20 ng/l to 4.3 μg/l in STP effluents; in surface waters, the values can reach 78.17 μg/l, which are values higher than the predicted no-effect concentration (PNEC) of 9.2 μg/l [3]. Hence, paracetamol might represent a threat for nontarget organisms.

2.3.3 Blood Lipid-Lowering Agents

Statins have the capability to subdue synthesis of the juvenile hormone in insects and may also produce detrimental effect to protozoan parasites, inhibiting growth and development. Reports suggested that a proliferation of peroxisomes in rodent livers is caused by fibrates. Embryonic development of nontarget organisms that share these receptors can be stopped by simply inhibiting cellular differentiation. Fibrates present in the micromolar concentration range are sufficient to cause it in zebrafish (*Danio rerio*) and amphibians [3]. Quinn et al. [91] classified gemfibrozil as toxic (EC_{50} between 1 and 10 mg/l) and bezafibrate as harmful for nontarget organisms (EC_{50} between 10 and 100 mg/l).

Clofibrate is classified as harmful to aquatic organisms as it showed LC_{50} values in the range of 7.7–39.7 mg/l. The fish *Gambusia holbrooki* [LC_{50} (96 h) = 7.7 mg/l] seems to be the most sensitive organism to acute clofibrate concentrations [92]. Clofibrate has an immunosuppressive action in mammalian hosts, suppressing the production of IgM but not IgE antibodies, allowing an amplified number of encysted larvae of the nematodes *T. spiralis* and *Trichinella nelsoni* to occur and a decrease in the rate of exclusion of adult worms from the intestines, although the effects differed between parasite species and host strain [93]. Fibrates have been assessed by conventional toxicity tests and the following no-observed-effect-concentration (NOEC) were found for clofibric acid in *C. dubia* [NOEC (7 days) = 640 μg/l], the rotifer *B. calyciflorus* [NOEC (2 days) = 246 μg/l], and in early life stages of zebrafish [NOEC (10 days) = 70 mg/l] [94]. Clofibrate was observed to produce no effect on in vitro growth of *T. bruceii* but did reduce the incidence of *P. berghei* and the invasiveness and development of *Acanthomoeba culbertsoni* in exposed mammalian hosts [95]. Lovastatin hinders the egg production of the trematode *S. mansoni* and subsequently there is a decline in pathogenic granulomas typically associated with the eggs in the mammalian liver [96].

2.3.4 Beta-blockers

Beta-blockers act by competitive inhibition of beta-adrenergic receptors which is critical for normal functioning in the sympathetic branch of the vertebrate autonomic nervous system. Among beta-blockers, propranolol shows the highest acute toxicity and highest log K_{ow} which proves the fact that it is a strong membrane stabilizer than other examined beta-blockers [97]. Undefined antagonists such as propranolol may be active in fish as they contain β_2-receptors in heart and liver as well as in reproductive tissues [98]. There is a prominent evidence that propranolol not only has chronic cardiovascular toxicity, but also has toxic effect on reproduction system. The NOEC and lowest-observed-effect-concentration (LOEC) of propranolol affecting reproduction in *C. dubia* were 125 and 250 μg/l, and reproduction was affected after 27 days of exposure in *H. azteca* at 100 μg/l [97].

Beta-blockers may also affect parasite functional biology. Aqueous exposure of propranolol may negatively affect swimming behavior, survival, and phototaxis of free living aquatic stages of trematodes. Propranolol may also considerably decrease the number of *Dirofilaria immitis* nematode larvae capable of finishing third-stage molt, and in vitro prevent the growth of the malaria parasite *Plasmodium falciparum* [99]. Fathead minnows exposed to atenolol throughout embryo-larval growth showed NOEC and LOEC values for growth rate of 3.2 mg/l and 10 mg/l, respectively [3]. At 48-h exposure to propranolol, LC_{50} values of 29.8, 1.6, and 0.8 mg/l were obtained for *H. azteca*, *D. magna*, and *C. dubia*, respectively, while acute exposure to nadolol did not affect the survival of the invertebrates [3]. Encystment of the protozoan *Entamoeba invadens* was inhibited in the presence of metoprolol [100].

2.3.5 Antineoplastic Drugs

Antineoplastic drugs are designed to kill the proliferating cells in cancer. As a consequence, a parallel effect can be expected on normally growing eukaryotic organisms. It is expected that antineoplastic drugs possess mutagenic, genotoxic, teratogenic, carcinogenic, and fetotoxic properties, and 14–53 % of the administered drugs can be excreted in unchanged form through urine [101]. Methotrexate revealed teratogenicity for fish embryos with an EC_{50} of 85 mg/l after 48 h of exposure and acute effects in the ciliate *Tetrahymena pyriformis* with an EC_{50} for 48 h of 45 mg/l [102]. Due to immunosuppressant property, methotrexate and cyclophosphamide are reported to cause a proliferation in disease incidence and intensity in host–parasite systems [103]. Acute toxicity of methotrexate is reported on highly proliferative species like the ciliate *Tetrahymena pyriformis* [EC_{50} (48 h) = 45 mg/l] [104]. On the contrary, cyclophosphamide appears to have a little effect on them. Methotrexate has been shown to have no or little effect on certain protozoans including *Toxoplasma gondii*, *Babesia bovis*, and *Leishmania tropica*, perhaps as they have different mechanisms of drug metabolism [105]. Development and growth of helminths in both mammalian and bird hosts were detrimentally effected by methotrexate and cyclophosphamide. Abnormal teratogenicity was noticed in fish embryos at higher concentrations [EC_{50} (48 h) = 85 mg/l]. *Biomphalaria glabrata*, a freshwater snail is largely affected with the long-term exposure to methotrexate [106]. Doxorubicin, tamoxifen, and methotrexate have all been reported as effective parasiticide agents against many protozoan species [107].

2.3.6 Neuroactive Compounds (Antiepileptics, Antidepressants)

A very limited number of studies on the effects of neurological agents on host–parasite dynamics have been studied, despite phenothiazine has been used as a parasiticide for long time [108]. The serotonin re-uptake inhibitor (SSRI) fluoxetine is deceptively the

most acute toxic human pharmaceutical with toxicity ranging from EC_{50} (48 h, alga) = 0.024 mg/l to LC_{50} (48 h) = 2 mg/l so far [2]. Sertraline exhibits highly toxic properties to rainbow trout (LC_{50} of 0.38 mg/l) at a 96-h exposure [109]. SSRIs were also tested on algae by evaluating the growth inhibition induced. Chronic toxicity tests proved that the organisms were sensitive with NOEC values below 1 mg/l [110]. *C. vulgaris* was shown to be the least sensitive species for all SSRIs tested. Fluvoxamine provided escalation to the highest EC_{50} values for all algae species tested (3563–10,208 μg/l).

Under the category of benzodiazepines, diazepam and nitrazepam were identified to increase the number of microfilariae of *Setavia cervi* liberated from the lungs into the peripheral blood circulation in rats [111]. Caffeine was found to stimulate the growth of *Plasmodium gallinaceum* and *P. falciparum*, while the antipsychotic haloperidol and the mood stabilizer valproic acid effectively inhibited the in vitro growth of *T. gondii* [112]. Diazepam and carbamazepine (antiepileptics) are classified as potentially detrimental to aquatic organisms as most of the acute toxicity data are below 100 mg/l. Conventional toxicity studies showed chronic toxicity of carbamazepine in *C. dubia* [NOEC (7 days) = 25 μg/l], in the rotifer *B. calyciflorus* [NOEC (2 days) = 377 μg/l], and in early life stages of zebrafish [NOEC (10 days) = 25 mg/l] [94]. Carbamazepine is carcinogenic to rats but does not have mutagenic properties in mammals [113]. It is also lethal to zebrafish at the 43 μg/l level and produces sublethal changes in *Daphnia* sp. at 92 μg/l [113]. Growth of *D. magna* was inhibited for concentrations of carbamazepine above 12.7 mg/l, showing acute toxicity at 17.2 mg/l [113].

2.3.7 Sex Hormones

Sex hormones are one of the extremely important biologically active compounds emerged as most serious aquatic environmental toxicants due to extensive use of human contraceptives. Exposure of mammalian hosts infected with the blood trematode *S. mansoni* to contraceptive pills resulted in a noteworthy modification in a range of liver cell's ultrastructure and function. Ethinylestradiol (EE2) is a synthetic estrogen found in oral contraceptive pills with noticeable estrogenic effects in fish. The life-cycle exposure of fathead minnows to EE2 concentrations below 1 ng/l produced a noteworthy decline in fertilization success, an increased egg production and decreased expression of secondary male sex characteristics. Life-long exposure of zebrafish to 5 ng/l to EE2 has led to reproductive failure due to the nonexistence of secondary male sex characteristics [63]. Exposure to 17β-estradiol caused an increased susceptibility to the protozoan *T. gondii* in mice, while increased pathology occurred in mammals infected with *Leishmania mexicana amazonensis* and exposed to either estradiol or testosterone [114]. Estradiol increased the susceptibility of cyprinids to

hemoflagellates by the suppression of lymphocyte proliferation [115]. At relatively high concentrations, hydrocortisone can cause an increase in the intensity of ectoparasitic infections in fish.

2.3.8 Antiparasitic Compounds

A study was performed on farms in the UK and the report suggested that concentrations of antiparasitic compounds of 0.112 mg/kg (doramectin) to 1.85 (ivermectin) mg/kg in dung were found. On the contrary, in same place, concentrations of these drugs in soil were considerably lower up to 0.046 mg/kg [116]. In a study performed in Slovenia, it was found that high concentrations of abamectin and doramectin were found in feces (0.2–0.8 mg/kg and 0.4–1.2 mg/kg, respectively) during the first 20 days after treatment, reaching concentrations of about 0.2 mg/kg after 70 and 50 days, respectively [117]. Grønvold et al. [118] found that ivermectin and fenbendazole affect the survival of the nematode *Pristionchus maupasi* at concentrations higher than 3 mg dung/kg (w/w) and 10–20 mg dung/kg, respectively. Svendsen et al. [119] showed that ivermectin and the fenbendazole did not affect earthworms. However, the disappearance of dung was affected by the avermectin but not by the fenbendazole. Avermectin B_{1A} with LC_{50} value of 17.1 mg/kg in soil was found with the compost worm *Eisenia fetida* [120]. Eprinomectin did not affect survival or biomass of the earthworm species *Lumbricus terrestris* in laboratory tests at concentrations up to 0.43 mg/kg dung (w/w) or 3.3 mg/kg dung [121].

2.3.9 Antivirals

Tamiflu [oseltamivir ethylester-phosphate (OP)] and Relenzas (zanamivir) belong to a novel class of antiviral drugs under the neuraminidase inhibitors category. National storing of neuraminidase inhibitors in the USA began with the emergence of the 2009 influenza pandemic (H1N1) [122]. Tamiflu tablet largely dominated Relenza (disk inhaler) due to its relative ease of administration. Tamiflu is a prodrug, which is converted to the active drug oseltamivir carboxylate (OC) in the liver. About 80 % of an oral dose of Tamiflu is excreted as OC in the urine and the remaining portions are excreted as OP in the feces. Therefore, both the parent chemical and its bioactive metabolite eventually are projected to reach a mean of 2–12 mg/l in WWTPs during a moderate and severe pandemic [122]. Current evidences suggested that rivers receiving WWTP effluent would also be exposed to OC throughout a pandemic. The OC concentrations between 293 and 480 ng/l have been recorded in rivers receiving WWTP effluent during the 2009 pandemic [123].

2.3.10 Pharmaceuticals Mixtures

Pharmaceuticals are identified as multicomponent mixtures rather than isolated pure substance in diverse environmental compartments. Majority of pharmaceuticals will either be transformed by physical and chemical means and/or subsequently biotransformed

by some organisms. Multicomponent mixtures are the foremost concerning issue for the ecotoxicity. The following characteristics also make their joint toxic effects a major issue for hazard and risk assessment:

1. The toxicity of a mixture has always a synergistic effect than the effects produced by a single component.
2. A mixture can have a substantial ecotoxicity, even if all components exist only in low concentrations that do not aggravate noteworthy toxic effects if acting separately on the exposed systems.

A combination of fluoxetine and clofibric acid is lethal for more than 50 % of a water-flea (*Daphnia*) population after an exposure of 6 days, although the individual drugs did not show any significant effect when present separately at same concentrations [124]. A substantial swing in sex ratio was perceived after an exposure to a three-component mixture of erythromycin, triclosan, and trimethoprim. Again, individual components did not elicit significant individual effects. These studies are very important to show that mixture effects have to be taken into consideration to identify the effects of pharmaceuticals.

2.4 The Environmental Risk Assessment

Exposure assessment is the procedure of determining or assessing the intensity, frequency, and extent of environment and human exposure to an existing pharmaceutical product, or of estimating theoretical exposure that might rise from the discharge of new pharmaceuticals into the environment. The concept of "exposomics," which integrates a top-down and bottom-up approach to identification of relevant exposure biomarkers, will be an important component of future exposure science [125]. The major aims of environmental risk assessment (ERA) should be risk mitigation and risk management. In order to alleviate or accept risks, a risk assessment has to be performed both for products and for activities followed by generation of report based on the characteristics of the product, its possible environmental exposure, fate and effects, and risk extenuation strategies. The inference of the report should be based on sound scientific reasoning supported by adequate studies. If other applicable data are accessible, they should also be submitted.

The outline of the registration process and the ERA consist of European Commission and Council directives and regulations on registration, European policy, case law, and global (trade) agreements. The decision-making process and the risk models should elevate the expenses to society in terms of ecotoxicity and financial loss. Also the assessment method itself should obstruct neither product development nor timely action to eradicate hazards.

2.4.1 Risk Assessment Approaches

The most commonly employed approaches for risk assessment are hazard identification, dose-response assessment, exposure assessment, and risk characterization of pharmaceuticals and its metabolites in various environment compartment [126].

Hazard Identification

The first step for risk assessment is hazard identification which supports the intensity of risk for a particular product. Although in vitro test studies provide useful data on the toxicity of environmental hazards, the majority of scientists rely heavily on the outcome of animal toxicity tests for hazard identification. As a consequence, a greater stress should be provided on the implication of in vitro assays in human cells and QSAR analysis, as well as the use of computational techniques in systems biology [127].

Dose-Response Assessment

Identification of the threshold dose of the toxic effect of any product is very much essential for scientific risk assessment. Dose-response information over a wide range of test concentrations should be assessed employing Quantitative high throughput screening (q-HTS). There should be availability of sensitive assays capable of detecting toxicity at very low doses or below environmental levels experienced by human populations. Statistical approaches can be used to estimate yardstick concentrations for adaptive and adversarial responses and to assess critical concentrations [128]. As discussed in subheading "Hazard Identification", the extrapolation techniques will be required to interpret in vitro test results to in vivo utilizing an appropriate internal tissue dose metric [129].

Dose and Species Extrapolation

The major problems of risk assessment are low-dose and interspecies extrapolation. In silico models and expert systems have supported such extrapolations, including linear and threshold models for low-dose extrapolation and body weight or surface area alterations for interspecies extrapolation. New extrapolation complications are dose extrapolation of molecular and cellular pathway responses, and extrapolation from the short-term in vitro to longer term in vivo exposure. In vitro to in vivo extrapolation and physiologically based pharmacokinetic (PBPK) models are amenable to sensitivity, variability, and uncertainty analysis employing conventional tools [130]. Computational biology systems will back the application of tools for determining variability and uncertainty from the pharmacologically based pharmacokinetics (PBPK) information as the pathway components imitate more targeted molecular elements and their interactions [131].

Exposure Assessment

In present scenario, human exposure assessment is made principally on the measured levels of environmental agents in the human environment [132]. In few cases, internal dose measures may also be calculated using biomonitoring [133] or pharmacokinetic modeling [134]. For superior exposure assessment, the focus should be

more on direct measures of critical toxicity pathway agitations in humans by employing innovative biomonitoring techniques coupled with advanced new high throughput approaches [135].

2.4.2 Environmental Risk Assessment Modeling of Pharmaceuticals

The risk assessment model consists of the risk assessment process, including their harmonization and communication with the risk management process. The risk model interprets the safety issues in quantities like probabilities, concentrations, dosages, and risk quotients of each pharmaceutical product. The simplest approaches to estimating concentrations of a pharmaceutical in diverse compartments are provided in the guidance for environmental assessments for regulatory drug approvals by the US FDA [30] or the EU EMA [28]. In Fig. 2, the risk assessment is harmonized with risk management process.

Before designing or modeling a toxicological study, it is very beneficial to assess exposure of any pharmaceutical by the following way [136]:

- The exposure is measured in form of the environmental concentration (occurrence) to which the biological system is exposed, the duration and frequency being not on the concentrations to which each individual is actually exposed. The actual exposure is subjected to many other factors such as, the fate, sorption effects, metabolism and transformation processes.
- The life stage and behavioral patterns should also be taken into account for any organism or living system.

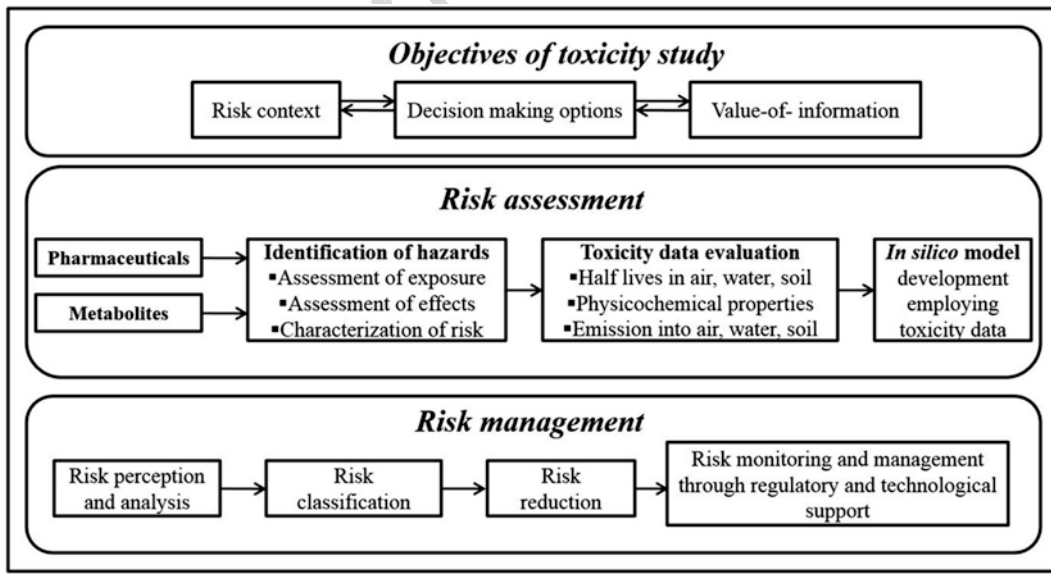


Fig. 2 Possible steps for risk assessment and risk management

- The bioavailability and toxicokinetics of the drug are studied.
- The pathways and target sites in the biological system are explored.
- The mode of action which depicts steps and processes to molecular and functional effects is determined.
- As pharmacokinetics and pharmacodynamics can influence the dose of pharmaceuticals, one has to consider these aspects to assess the dose which ultimately reaches the environment taking into account possible absorption, distribution, and elimination mechanisms.
- The hazard due to the inherent toxicity of the pharmaceutical according to its chemical properties is also studied.

2.5 Risk Management

Risk management is "the process of identifying, evaluating, selecting, and implementing actions to reduce risk to human health and to ecosystems. The goal of risk management is scientifically sound, cost-effective, integrated actions that reduce or prevent risks while taking into account social, cultural, ethical, political, and legal considerations" [137]. For eco-friendly risk management, one may select a combination of apposite tactics to balance risks, costs and benefits, taking into account social values and economic considerations.

2.5.1 Implementation of Precautionary Measures

The application of pharmaceuticals and their after use toxic effects cannot be stopped but the probable risk of pharmaceutical products related to environmental can be controlled by implementing proper precaution and safety measures. The EMEA 2006 guideline demonstrates following steps as safety measures for risk management:

1. Calculation of product risks initially
2. Proper product labeling and summary product characteristics (SPC)
3. Package leaflet (PL) for each pharmaceutical for patient use to inform the probable toxic effects
4. Appropriate and safe storage of pharmaceutical product
5. Safe and proper scientific disposal of pharmaceuticals

2.5.2 Lessening the Input of Pharmaceuticals into the Environment

To diminish the occurrence of pharmaceuticals into the different compartments of the environment, one has to follow the principle of sustainability where the entire life cycle of a pharmaceutical has to be taken into consideration to categorize the opportunities for risk management. For diminishing the input of pharmaceuticals into the environment, following steps can employed effectively [138].

Training and Awareness

The most important step to reduce the occurrence of pharmaceuticals in the environment is proper training and awareness of users who are the major source points. A proper usage and disposal of pharmaceutical is the responsibility of the shareholders and people using the compounds, including patients, doctors and nurses, and pharmacists. Industrial sectors should have the major role to treat the failed active pharmaceutical under quality control category properly before it reach to the environment. Additionally, each pharmaceutical product should consist of materials safety data sheet (MSDS) intended to provide workers and emergency personnel with procedures for handling or working with that substance in a safe manner, information such as physical data, toxicity, health effects, first aid, reactivity, storage, disposal, protective equipment, and spill-handling procedures. Appropriate and effective risk management strategies need basic knowledge of entry routes of pharmaceuticals. Therefore, one has to identify the bulk of drug flows connected with the diverse sources of pharmaceuticals such as households, industries, hospitals and pharmacy.

Improved Sewage Treatment

The most technical and extensively considered approach for risk management is improvement of sewage treatment. Analyzing Table 1, one can easily identify the presence of threatening amount of pharmaceutical wastes after sewage and waste water treatment also. The purpose of advanced and improved sewage and waste water treatment is to further reduce the ecotoxicity, hormonal effects and pathogenic effects of the effluent. In recent years, advanced effluent treatment has been studied extensively. The advanced treatment of sewage influents and effluents as well as waste water treatment can be done employing photochemical oxidation processes, filtration, and application of powdered charcoal and constructed wetlands [139].

Green and Sustainable Pharmacy

The third approach is evolving from the knowledge of green and sustainable pharmacy which states that substitution of the compound with a more environmentally benign compound [138]. Though this approach is less practiced, in terms of sustainability, it appears to be the most encouraging one in the long run. The prime principle of green chemistry is easy and fast degradability of pharmaceuticals after their application. Understanding of full life cycle of drugs will lead to a different understanding of the functionality necessary for a pharmaceutical.

Additionally, other crucial issues like (a) development of improved drug delivery systems so that lower doses are required; (b) upgradation of packaging and package sizes to prolong shelf life and lessen the amount of the product that expires and rejection of unused products; and (c) changes in prescription and animal farming practices are substantial options for minimizing or

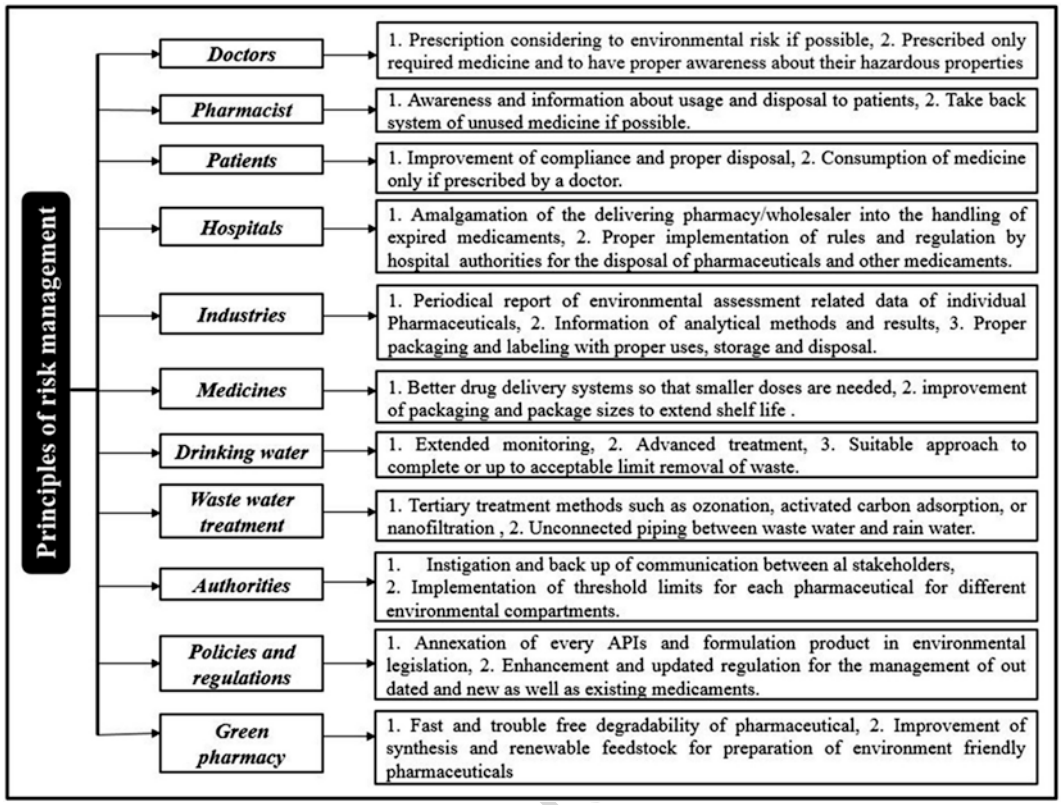


Fig. 3 Probable actions to be taken for reduction of the occurrence of pharmaceuticals in the environment by different stakeholders

eliminating emissions to the environment. Potential processes and measures to decrease the environmental toxicity by carious stakeholders are addressed in Fig. 3 for a better understanding.

3 Regulatory Agencies for the Risk Assessment and Management of Ecotoxicity Pharmaceuticals

Immense exposure of pharmaceuticals and their metabolites to the environment is a matter of concern and a burning global issue at recent times. The risk effects are not only related with the environment, it is also directly related to human health to a large extent. As a consequence, release of these pharmaceutical products, their risk assessment as well as risk management are controlled and regulated at local, national and international levels by different governments and regulatory agencies worldwide. As experimental data of environmental fate and toxicity of pharmaceuticals are absent or some time not sufficient, there is a strong urge to predict physical and chemical properties, environmental fate, ecological effects and health effects of pharmaceuticals and their metabolites. Several

government organizations have been applying the approaches of structure–activity relationship (SAR) and QSAR to develop the predictions for untested existing as well as newly introduced pharmaceuticals. To establish proper identification of environmental hazards, their risk assessment and fate modeling, SAR and QSAR approaches along with other predictive in silico tools are employed by Australian, Canadian, Danish, European, German, Japanese, Dutch, and US Government organizations [28, 30, 140–144].

QSAR models can be generated for prediction of the following ecotoxicity related properties or effects:

1. Physicochemical properties
2. Toxic potential and potency
3. Environmental distribution and fate in different compartments (air, water and soil) of environment
4. Biokinetic processes (absorption, distribution, metabolism, and excretion) of pharmaceuticals and their metabolites

Areas where QSARs can be applied by governmental regulatory agencies are as follows:

1. Prioritization of existing pharmaceuticals for toxicity testing to environment.
2. Classification and labeling of new pharmaceuticals according to their safe use.
3. Risk assessment of new and existing pharmaceuticals.
4. Guiding experimental design of regulatory tests or testing strategies.
5. Providing mechanistic information
6. Filling up the large data gaps.
7. Building a proper database of each pharmaceutical to different species regarding environmental toxicity.
8. Development of expert systems for each therapeutic classes for different compartments of the environment.
9. Construction of efficient interspecies models to extrapolate data from one species to another species when data of a particular species is absent.

Global regulatory authorities and agencies [28, 30, 140–144] for the risk identification, risk assessment and finally risk management of ecotoxicity pharmaceuticals are listed in Table 2.

The most common endpoints associated with various test methods proposed under Organization for Economic Co-operation and Development (OECD) are the following ones:

- *Physical-chemical properties*: Most commonly evaluated properties are melting point, boiling point, vapor pressure, octanol–water partition coefficient, organic carbon–water partition coefficient, and water solubility.

Table 2
Global regulatory bodies and agencies for the hazard and risk assessment of ecotoxicity pharmaceuticals

Regulatory agencies	Note
EMEA (European agency for the evaluation of medicinal products)	The EMEA has prepared a draft guideline for the environmental risk assessment of human medicinal products. It demonstrates the scope and legal basis for risk assessment of pharmaceuticals and outlines the general considerations and the recommended step-wise procedure for their risk assessment. As the risks cannot be excluded completely, this guideline outlines precautionary and safety measures to be considered. The guideline is based on the risk assessment paradigms for industrial chemicals and biocides, but it also considers the specific features of pharmaceuticals, e.g., the use of available pharmacological information. Previously environmental risk assessments were mainly based on acute ecotoxicity data, but in recent times EMEA draft has proposed to include pharmacokinetic and pharmacodynamic data for environmental risk assessment. Such an approach is presently also taken within the European Union project ERA Pharm. The phased approach in the environmental risk assessment by EMEA is divided into three different Phases: 1. Phase I: *Pre-screening* and *estimation of exposure* based only on the drug substance, irrespective of its route of administration, pharmaceutical form, metabolism and excretion. 2. Phase II Tier A: *Screening* and *initial prediction of risk*. In this phase, all relevant data should be taken into account, e.g., data on physical-chemical properties, primary and secondary pharmacodynamics, toxicology, metabolism, excretion, degradability, and persistence of the drug substance and/or relevant metabolites. 3. Phase II Tier B: *Extended* and *substance and compartment-specific risk assessment*. At the end of Tier B, information from the refined data set is available comprising information on route(s) of excretion; and qualitative and quantitative information on excreted compounds, and possibly additional long-term toxicity data
EU-CSTEE (European Union Commission's scientific committee on toxicity, ecotoxicity and environment)	The CSTEE identified the need for a proactive approach in obtaining data on the environmental effects of pharmaceuticals. Thus, it is recognized that a prioritization procedure needs to be developed for environmental risk assessment of pharmaceuticals, and that this should follow the general scheme for chemicals described in the White Paper for the EU chemicals policy, i.e., REACH guideline, where the uses of QSTRs are stressed. QSTR is the first step in gaining more general knowledge on the risk assessment issue as an alternative to non-animal method. In contrast to the amount of analytical data, information about the ecotoxicological effects of drug residues is scrubby. To create a broader basis for the evaluation of the ecotoxicological relevance of pharmaceutical compounds, proper documentation of their effects and the reason should be identified.

(continued)

Table 2
(continued)

Regulatory agencies	Note
US-FDA (US Food and Drug Administration) and *CDER* (Center for Drug Evaluation and Research)	An assessment of risk to the environment is required for manufacture, use and distribution of human drugs under the National Environment Policy Act of 1969 and an environmental assessment procedure was developed by the FDA as a part of the registration procedure for new human pharmaceutical drugs. Along with it, in 1995, the FDA-CDER issued a new guidance document for the Submission of an Environmental Assessment in Human drugs. In the same year, the US FDA initiated a retrospective review on ecotoxicity data submitted in environmental assessments over the preceding decade. In this respect, in 1997 the FDA implemented a Note for Guidance paper in which all drugs entering the aquatic compartment at levels below 1 μg/l, Predicted Environmental Concentration ($PEC_{EFFLUENT}$) were exempted from a detailed risk assessment.
MHLW (The Ministry of Health, Labor and Welfare of Japan)	The MHLW constructed a research group to build up a concept on the regulation of pharmaceuticals for environmental safety in 2007. The regulation system is similar to that of general chemicals in Japan and the Guideline by EMEA. The main function of this group is to establish a risk-benefit analysis committee for the pharmaceuticals which have a high risk for environmental organisms and to human health. The risk assessment is judged by the PEC/PNEC (Predicted Environmental Concentration/Predicted No Effect Concentration) ratio or ΣPECi/PNECi. In addition, the Organization for Pharmaceutical Safety and Research (OPSR) conducted compliance reviews on application data. This was followed by the integration of the aforementioned Evaluation Center, OPSR, and part of the Medical Devices Center to form a new independent administrative organization, the Pharmaceutical and Medical Devices Agency (PMDA). The MHLW and PMDA handle a wide range of activities from clinical studies to approval reviews, reviews throughout post-marketing stage, and pharmaceutical safety measures.
NICNAS	The National Industrial Chemicals Notification and Assessment Scheme (NICNAS) is a statutory scheme administered by the Australian Government Department of Health. NICNAS was established in July 1990 under the *Industrial Chemicals (Notification and Assessment) Act 1989* (the Act). A range of state, territory and Commonwealth government agencies share regulatory responsibility for chemical safety in Australia, with each chemical being regulated according to its use, whether as a pharmaceuticals, veterinary medicine, pesticide, food additive or industrial chemical. The major responsibility of NICNAS are: • Assessing new industrial chemicals for human health and/or environmental effects • Maintaining the Australian Inventory of Chemical Substances (AICS) • Circulation of information on the human health and environmental impacts of chemicals and recommending on their safe use • Registering new industrial chemicals

UBA (Federal Environment Agency)	The German Medicines Act provides that the Federal Environment Agency (UBA) is responsible for the environmental risk assessment. The UBA started assessing the environmental impact of veterinary and human pharmaceuticals in an authorization routine in 1998 and 2003, respectively. The UBA already assessed around 180 veterinary and around 240 human pharmaceutical formulations. Filtering concepts established between UBA and the authorization agency responsible for veterinary medicines focused the ERA on antibiotics, parasiticidal substances and analgesics. Cytostatic medicines, hormones and contrast agents dominated the human medicine dossiers assessed by UBA.
SECIS (Swedish Environmental Classification and Information System for pharmaceuticals)	The SECIS is an authorized regulatory body which was initiated in 2005 by the Swedish Association for the Pharmaceutical Industry. The rationale of the classification system is to offer the public and health care sectors with environmental information about all active pharmaceutical ingredients (API) on the Swedish market up to till date. In 2004, the Swedish Medical Product Agency (MPA) concluded in a report to the Swedish government that there is a lack of environmental toxicity data for the majority of the APIs available on the Swedish market. To improve risk management decision making, sufficient knowledge about environmental exposures and effects in nontarget species for all relevant pharmaceutical substances is needed. Within SECIS, the pharmaceutical companies provide environmental data and classify their products according to predefined criteria and a guidance document. The guidance document is developed for the purposes of SECIS, but it is based on the European Medicines Agency (EMA) guideline for environmental risk assessment of pharmaceuticals and the European Commission Technical Guidance Document (TGD).
AEA (Australian Environment Agency)	The AEA applies the latest methodologies to environmental risk assessment. It advises clients on the environmental hazards and potential risks associated with the production, use, and disposal of chemicals. AEA has undertaken extensive reports for the Department of Sustainability, Environment, Water, Population and Communities (DSEWPaC), particularly with respect to their environmental assessments performed on new and existing agricultural and veterinary chemicals for the Australian Pesticides and Veterinary Medicine Authority (APVMA), and industrial chemicals for the National Industrial Chemicals Notification and Assessment Scheme (NICNAS).
VICH (International Cooperation on Harmonization of Technical Requirements for Registration of Veterinary Medicinal Products)	VICH is a trilateral (EU–Japan–USA) program aimed at harmonizing technical requirements for veterinary product registration was officially launched in April 1996. The initiative to begin the harmonization process came about in 1983 when the first International Technical Consultation on Veterinary Drug Registration (ITCVDR) was held. Veterinary medicinal products (VMPs) are regulated for environment safety as described in Environmental Impact Assessment for VMPs; Phase I in 2000 and Phase II in 2004.

- *Ecological effects on endpoints*: Acute fish-toxicity, long-term toxicity, acute Daphnia toxicity, algal toxicity, terrestrial toxicity, marine organism toxicity, microorganism toxicity in sewage treatment plant.
- *Environmental fate*: Biodegradation, hydrolysis in water, atmospheric oxidation, and bioaccumulation;
- *Human health effects*: Acute oral, acute dermal, acute inhalation, eye irritation, skin irritation, skin sensitization, repeated dose toxicity, genotoxicity, reproductive toxicity, developmental toxicity, systemic toxicity, mutagenicity, carcinogenicity, etc.

OECD's database on risk assessment models:

In silico models that are employed by the OECD countries to predict health or environmental hazards, exposure potential, and probable effects were organized into a searchable database. This database is intended as an information resource only. The models are listed by countries and by the property or effect included. The models can be useful as a screening tool, when there is a lacking of chemical-specific data, for establishing priorities for chemical assessment and for identifying issues of potential concern [140]. Areas of assessment and category of information for predicting human health and environment according to OECD's guidelines are represented in Figs. 4 and 5, respectively.

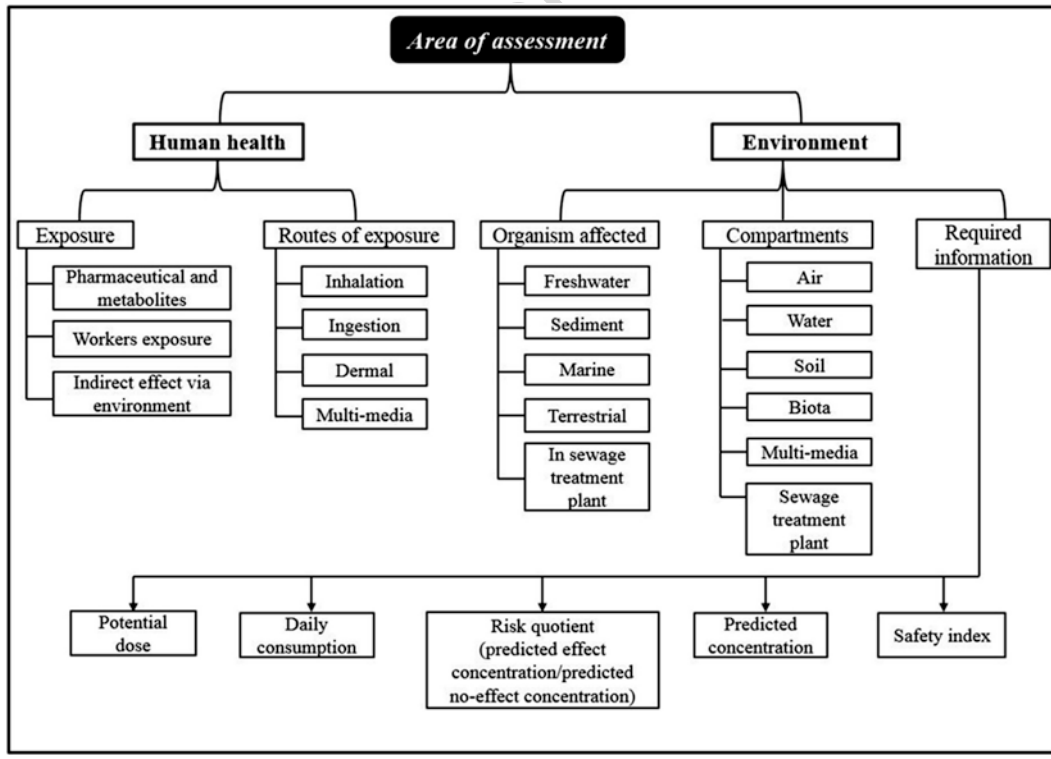


Fig. 4 Areas of assessment and risk models for predicting human health and environment according to the OECD database

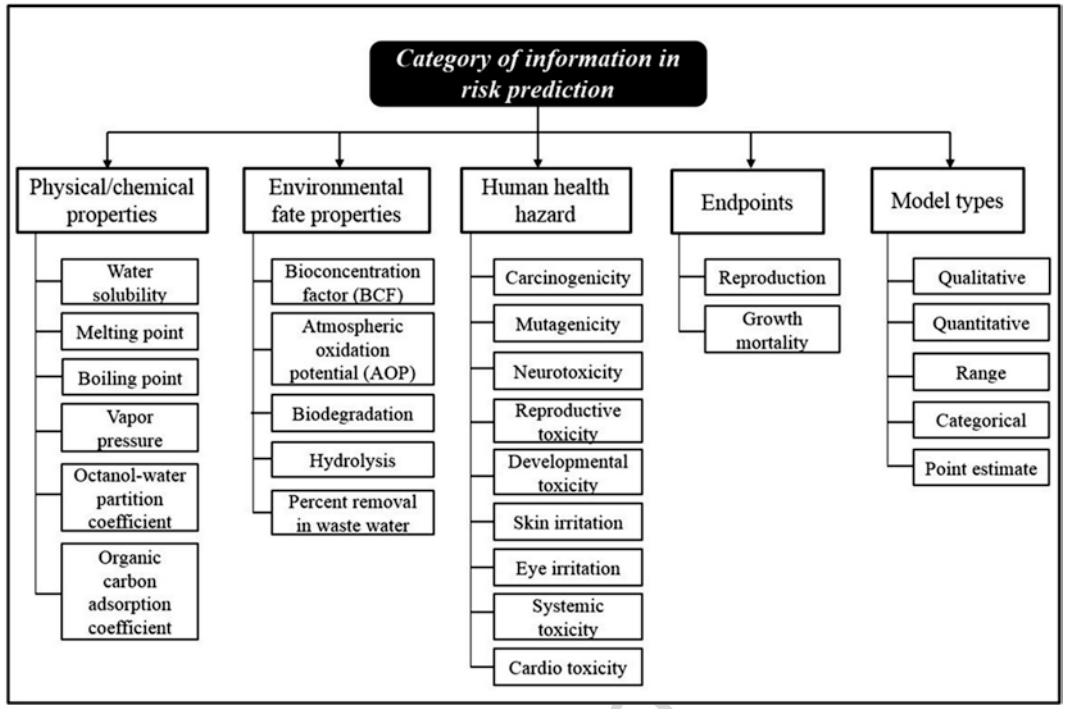


Fig. 5 Category of information included in predicting health and environmental effects according to the OECD guidelines

4 In Silico Modeling of Ecotoxicity Using SAR and QSAR Approaches

The toxic potential of large quantities of industrial chemicals including pharmaceuticals, cosmetics, pesticides and other synthetic or semisynthetic chemicals is often required to be assessed by using standard animal models, comprising the basic test protocol for risk assessments for their approval as a registered product to launch into the market. With increasing concern about the environmental pollution and human health, the manufacture, storage, distribution, and release of these hazardous substances after their application to the environment are controlled and regulated at various levels by different governments and regulatory agencies worldwide. Applications of analogues, SAR and QSAR of different pharmaceuticals are also providing useful information in a regulatory decision making context in the absence of experimental data [140]. Most commonly employed predictive in silico tools are depicted in Fig. 6.

Among the available in silico predictive models for ecotoxicity, majority of the models are constructed employing QSAR techniques. Therefore, in this book chapter, a special importance is given to the discussion of QSAR models. The QSAR approach

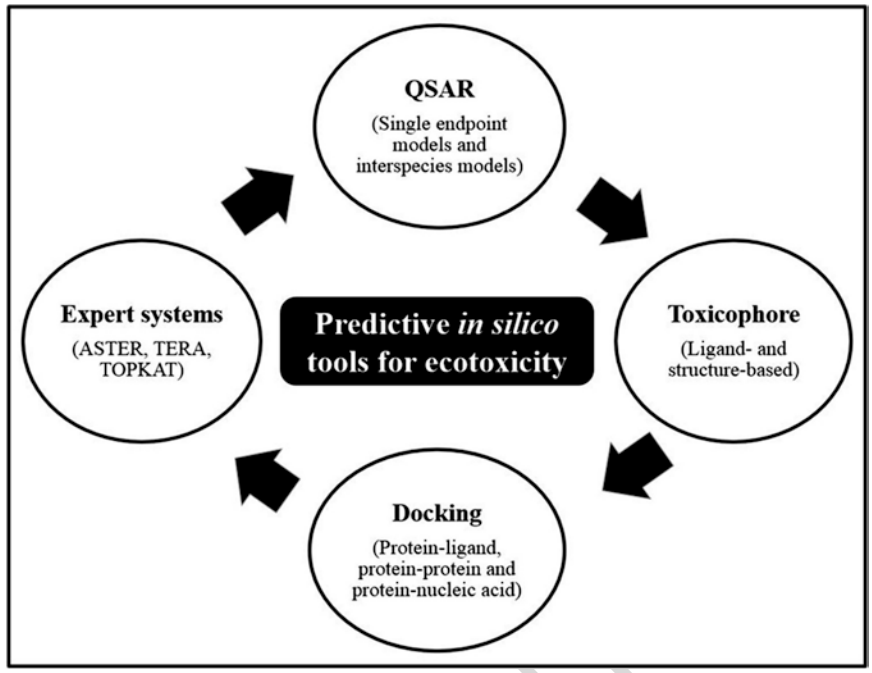


Fig. 6 Predictive in silico tools for the prediction of ecotoxicity of pharmaceuticals

attempts to correlate structural/molecular properties with biological activities/toxicities, for a set of compounds by means of statistical methods. As a result, a simple mathematical relationship is established:

$$\text{Biological activity or toxicity} = f\left(\text{chemical structure or property}\right).$$

Applications of QSAR can be extended to any molecular design purpose, prediction of different kinds of biological activities and toxicities, lead compound optimization, classification, diagnosis, and elucidation of mechanisms of drug action, toxicity prediction of environmental toxicants (pollutant pharmaceuticals, chemicals, gas, etc.), and prediction of drug-induced toxicity [145]. The major objective of structure–activity/toxicity relationship modeling is to investigate and identify the decisive factors for the measured activity/toxicity for a particular system, in order to have an insight of the mechanism and behavior of the studied system. For such a purpose, the employed strategy is to generate a mathematical model that connects experimental measures with a set of chemical descriptors determined from the molecular structure for a set of compounds. The derived model should have as good predictive capabilities as possible to predict the studied biological/toxicological or physicochemical behavior for new compounds. The factors governing the events in a biological system are represented by a

multitude of physicochemical descriptors, which can include parameters to account for hydrophobicity, electronic properties, steric effects, and topology, among others [145].

With the constant progress of QSAR techniques, many methods, algorithms, and techniques have been discovered and applied in QSAR studies. The development of a QSAR model follows five major steps:

1. Selection of a dataset with series of known response data
2. Calculation of descriptors
3. Splitting of the dataset into training and test sets for model development and its subsequent validation
4. Construction of models using different chemometric tools, and
5. Validation of the developed model based on internal and external validation statistics

Additionally, the development of 3D-QSAR models includes two more steps for their successful execution: conformation analysis of the molecules and their alignment status with respect to the most active compound. The most important feature for an acceptable and reliable QSAR model is predictive capability for new set of compounds. The predictive quality of the developed model is determined based on different validation statistics. Thus, validation of QSAR models plays the most crucial role in defining the applicability of the QSAR model for the prediction of untested compounds. Initially, verification of the correlation between chemical features of the molecules and the biological activity/toxicity was of prime interest during the development of a QSAR model. Later, the focus gradually shifted toward the predictive power of the model than simply unveiling the quantitative relationships [146].

To validate a QSAR model, one has to follow OECD principles for acceptable predictions in order to make the model as a reliable screening tool for future toxicity prediction of untested pharmaceuticals. A meeting of QSAR experts held in Setúbal, Portugal in March 2002 reported guidelines for the validation of QSAR models for regulatory purposes. The OECD principles were agreed by OECD member countries, QSAR and regulatory communities at the 37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology in November 2004. These principles are listed here: Principle 1: a defined endpoint; Principle 2: an unambiguous algorithm; Principle 3: a defined domain of applicability; Principle 4: appropriate measures of goodness-of fit, robustness, and predictivity; Principle 5: a mechanistic interpretation, if possible [147]. Different quality metrics for QSAR models can be categorized into two classes: one determining the fitting ability of the model while the other analyzing the predictive potential of the developed model [146].

4.1 Why In Silico Models Should Be Developed for Ecotoxicity Predictions of Pharmaceuticals?

4.1.1 The 3R Concept

The 3R concept represents three words "Reduction," "Replacement," and "Refinement". The concept brought about an imperative notification about animal experimentation in the scientific communities. The word 'Reduction' refers to the diminution in number of animals used to get results of a defined precision. Next, 'Replacement' corresponds to the use of nonliving resources to replace conscious living higher animals, and 'Refinement' means decline in the severity or cruelty of inhuman methodologies applied to the experimental animals [148]. As a consequence, to establish the 3R concept, in silico techniques are one of the front runners.

4.1.2 Ban of Animal Experimentation

There are different social as well as governmental organizations that consider reduction or complete ban of animal experimentations [149]. Here, we have listed a few of them:

1. The European Centre for the Validation of Alternative Methods (ECVAM) was established in the year 1991 that agrees the principle of 3Rs.
2. The European Convention for the Protection of Vertebrate Animals used for Experimental and Other Scientific Procedures.
3. Council Directive 86/609/EEC on the Approximation of Laws, Regulations and Administrative Provisions of the Member States Regarding the Protection of Animals Used for Experimental and Other Scientific Purposes.
4. Johns Hopkins Center for Alternatives to Animal Testing (CAAT), a US based organization focussing on the reduction of animal experimentations.
5. The testing ban on the finished cosmetic products applies since 11 September 2004; the testing ban on ingredients or combination of ingredients applies since 11 March 2009. The marketing ban applies since 11 March 2009 for all human health effects with the exception of repeated-dose toxicity, reproductive toxicity, and toxicokinetics. For these specific health effects, the marketing ban applies since 11 March 2013, irrespective of the availability of alternative non-animal tests [150].
6. India and Israel have also banned animal testing for cosmetic products, while the USA has no such ban in place [151].
7. China is the only major market where testing all cosmetics on animals is required by law, and foreign companies distributing their products to China must also have them tested on animals. [152] China has announced that its animal testing requirement will be waived for shampoo, perfume, and other so-called "non-special use cosmetics" manufactured by Chinese companies after June 2014. "Special use cosmetics," including hair regrowth, hair removal, dye and permanent wave products, antiperspirant, and sunscreen, will continue to warrant mandatory animal testing.

4.1.3 Regulatory Decision Making

SARs and QSARs are employed to predict aquatic toxicity, physical or chemical properties, and environmental fate parameters as well as to predict specific health effects of organic chemicals by Australian, Canadian, Danish, European, German, Japanese, Dutch, and US Government organizations.

4.1.4 Filling Data Gaps

Acceptable toxicity data of pharmaceuticals to environment and human health is considerably less than 5 % [153]. Computer-aided prediction has the competency to assist in the prioritization of pharmaceuticals for testing, and for predicting specific toxicities to allow for classification. As the number of reliable models for toxicity predictions is increasing, they can be employed as one of the major sources for filling the missing data of pharmaceutical toxicity to ecosystem.

4.1.5 Development of Understanding of Biology and Chemistry

In the modeling of acute toxicological endpoints, much has been gained regarding mechanisms of action. For many modeling approaches, it may be assumed that compounds fitting the same QSAR are acting by the same mechanism of action. This has allowed workers to define the chemical domain of certain mechanisms. There are countless examples where knowledge of biology and chemistry has been advanced by modeling in the field of toxicological and fate effects [154].

4.1.6 Cost and Time Reduction

Toxicity study is very costly in terms of the animals employed for testing and time taken. Even a simple ecotoxicological assay may cost several thousand dollars, and a 2-year carcinogenicity assay may cost several million dollars. Cost is a clear issue to fill the data gaps for the many new compounds that have not been tested. On the other hand, prediction of various toxicity endpoints for pharmaceuticals at an early stage of design can save a large amount of expenses for such compounds which may be found toxic at a later stage of drug development program [155].

4.1.7 Identification of New Toxicological Problems

The development of computational techniques not only allows for the prediction of the potential risk of pharmaceuticals but also allows for rational direction to be given to the testing programs.

5 Review of Literature on In Silico Ecotoxicity Modeling of Pharmaceuticals

Kar and Roy [156] have constructed robust quantitative interspecies toxicity correlation models for *Daphnia magna* and fish evaluating the ecotoxicity of structurally diverse 77 pharmaceuticals. They have demonstrated that the keto group and the  (aasC) fragment are principally responsible for higher toxicity of pharmaceuticals to *D. magna*. On the other hand, for fish toxicity, along with the keto group, structural fragments like X=C=X, R–C(=X)–X, and

R–C≡X are largely responsible for the toxicity. The interspecies models were also used to predict fish toxicities of 59 pharmaceuticals (for which Daphnia toxicities were present) and Daphnia toxicities of 30 pharmaceuticals (for which fish toxicities were present). They established that the interspecies correlation study would permit an improved and inclusive risk assessment of pharmaceuticals for which toxicity data was missing for a particular endpoint.

Das et al. [157] attempted to develop interspecies correlation models taking rodent toxicity as dependent variable so that any drug without reported rodent toxicity can be predicted using fish, daphnia, or algae toxicity data which can be further extrapolated to human toxicity. Interspecies extrapolation QSAR models were developed employing multiple validation strategies. Analyzing the models, the authors concluded that heteroatom atom count and charge distribution were significant determinants of the rodent toxicity, and that the atom level log *P* contributions of various structural fragments and various extended topochemical atom (ETA) indices reflecting electronic information and branching pattern of molecules were important determinants for the rodent toxicity. In addition, from interspecies aquatic toxicity modeling, it was established that apart from the algae toxicity, atom level log *P* contributions of different fragments, charge distribution, shape, and ETA parameters were important in describing the daphnia and fish toxicities in the interspecies correlation models with algae toxicity. The toxicity of chemicals to rodents bears minimum interspecies correlation with other mentioned nonvertebrate and vertebrate toxicity endpoints.

The acute toxicity was predicted (>92 %) using a generic quantitative structure–toxicity relationship (QSTR) model developed by Sanderson and Thomsen [158] suggesting a narcotic mechanism of action (MOA) of 275 pharmaceuticals. An analysis of model prediction error suggests that 68 % of the pharmaceuticals have a nonspecific MOA. Authors have compared the measured effect data to the predicted effect concentrations using ECOSAR regarding the predictability of ecotoxicity of pharmaceuticals and accurate hazard categorization relative to Global Harmonized System (GHS). Molecules were predicted using the model resulting in 71 % algae, 74 % daphnia, 83 % fish datasets that could be compared.

Escher et al. [159] constructed QSAR models with the total toxic potential of mixtures of the β-blockers and related human metabolites for the phytotoxicity endpoint. They have assumed two scenarios for this study. In the first scenario, the metabolites lose their explicit activity and act as baseline toxicants. In the second scenario, the metabolites reveal the identical specific mode of action like their parent drug. β-Blockers are secondary amines and are, therefore, fully protonated at environmental pH. The authors accounted for their positive charge in the QSAR analysis and have

experimentally determined the liposome–water partition ratios at pH 7 to make QSAR analysis more robust.

Berninger and Brooks [160] considered the mammalian Acute to Therapeutic Ratio (ATR) to predict pharmaceuticals which may result in comparatively high Acute to Chronic Ration (ACR) in fish models. The authors identified a statistically significant relationship between mammalian ATRs and fish ACRs ($p < 0.001$, $r^2 = 0.846$). In this model, they only included chronic responses of fish to pharmaceuticals which appear to have been elicited through a therapeutic MOA for calculating ACRs and for statistical analysis of the relationship with mammalian ATRs. Utilizing this approach, mammalian ATR values can be used for predicting pharmaceuticals with higher fish ACRs if the chronic response used in ACR calculation is reasonably linked to the therapeutic MOA of a pharmaceutical.

Sanderson et al. [161] employed the US EPA generic aquatic (Q)SAR model ECOSAR to screen more than 2800 pharmaceuticals and provided a baseline to fill the screening data regarding parent pharmaceuticals environmental toxicity. The model can be used to predict both acute and chronic aquatic toxicity.

Sanderson and Thomsen [162] overestimated the toxicity for 70 % of the 59 pharmaceuticals by ECOSAR v3.20 which contains both measured and modeled data. For the remaining 30 % pharmaceuticals, more than 94 % of the predictions underestimated toxicity by less than a factor of 10. This is an indication that a narcosis based model is conservative relative to experimental values around 70 % of the time, thus implying that for at least 70 % of the Active Pharmaceutical Ingredients (APIs), the acute mode of action (MOA) can be elucidated by baseline toxicity. The authors have observed determination coefficients (r^2) ranging from 0.73 to 0.76 between all the modeled Log EC_{50} and Log K_{ow}. The slopes of the Log EC_{50}–Log K_{ow} regressions based on measured data from the USA National Oceanic and Atmospheric Administration (NOAA) database for both fish and daphnia equal −0.86 which suggest a narcotic MOA.

Lienert et al. [163] assessed the ecotoxicological risk potential of 42 pharmaceuticals from 22 therapeutic classes, including metabolites formed in humans. They considered each parent drug and its metabolites as a mixture of equally acting compounds, and in case when effect data were missing, they estimated these with QSAR models. They have collected data on the identity and excretion pathways of human metabolites and, where available, experimental ecotoxicity data (EC/LC_{50}) from pharmaceutical compilations and from diverse literature sources. They have compiled physicochemical data like structure, molecular weight, octanol–water partition coefficient K_{ow}, acidity constant pK_a mainly from the Physical Properties Database (http://www.syrres.com/esc/physprop.htm). Moreover, they have generated a risk quotient

($RQ_{mixture}$) using simple predictions of drug concentrations in wastewater which can be useful for risk assessment of pharmaceuticals.

Christen et al. [164] developed VirtualTox Lab [165] to predict the effects of pharmaceuticals in the aquatic system. The study leads to the inference that the mode of action perception is most appropriate for the identification of highly active compounds (HC). As suggested by the authors, modification can be done by balancing this concept by the QSAR model (VirtualTox Lab), whereas the fish plasma model seemed to be less apposite due to the requirement of environmental concentration above 10 ng/l for the identification of a risk. The practice of the VirtualTox Lab will support the mode of action concept and may be beneficial to recognize surplus targets of the pharmaceutical to assess the ecotoxicity.

Escher et al. [166] predicted baseline toxicity of the 100 molecules using established QSARs for algae, daphnia, and fish. The QSARs were selected from the Technical Guidance Document of the EU. The logarithm of D_{lipw} (liposome water distribution coefficient) was employed in the model development for baseline toxicity to calculate the toxicity of the compound towards the stated species.

The environmental risk assessment of 26 pharmaceuticals and personal care products have been performed by De García et al. [167] based on the ecotoxicity values generated by bioluminescence and respirometry assays. Then the compounds were classified following the Globally Harmonized System of Classification and Labelling of Chemicals by predictions using the US EPA ecological structure–activity relationship (ECOSAR™). The real risk of impact of these pharmaceuticals in wastewater treatment plants (WWTPs) and in the aquatic environment was predicted according to the criteria of the European Medicines Agency. According to their studies, in at least two ecotoxicity tests, 65.4 % of the PPCPs showed prominent toxicity to aquatic organisms. There study showed some type of risk for the aquatic environments and/or for the activated sludge of WWTPs for pharmaceuticals like acetaminophen, ciprofloxacin, clarithromycin, clofibrate, ibuprofen, omeprazole, triclosan, parabens, and 1,4-benzoquinone.

Here we have discussed available in silico models on ecotoxicity of pharmaceuticals. Due to the limited availability of the in silico models on ecotoxicity of pharmaceuticals, there is a need to develop more in silico models in order to reduce time and cost involvement as well as reduction of animal usage in getting relevant data and for better and fast risk assessment of pharmaceuticals. It is not possible to experimentally study toxic effects of each pharmaceutical in different species. Most active pharmaceutical ingredients have available rodent toxicity information. As a result, if this data could be extrapolated or modeled to different other species,

this would be a noteworthy resource for prioritization of pharmaceuticals with regards to diverse environment hazards. However, very limited papers have been published on interspecies models to predict environmental toxicity for pharmaceuticals, and there are relatively few statistical models available to bridge the chronic toxicity data information gap [168, 169].

6 Endpoints

Toxicity of a molecule should be assayed on specific toxicity endpoints for the generation of data which are employed commonly to develop in silico models. This why a clear concept is required about the endpoints or test batteries as they are employed for the experimental toxicity studies and for understanding the mode of toxicity with respect to that particular endpoint [110]. We list the most commonly employed endpoints for this purpose in Table 3.

7 Databases

A good quality of ecotoxicological data of pharmaceuticals and their metabolites is required for the development of accurate and reproducible in silico models. A significant number of chemical/drug/agrochemical/pesticide toxicity databases towards environment are publicly accessible, and such numbers are growing. But one cannot deny that the existing databases are very few compared to drug discovery compound libraries. Recent initiatives requiring superior use of in silico technologies have called for transparency and expansion of toxicity database information that is available to the public at no cost. Table 4 represents publicly available toxicity databases describing environmental as well as human health effects of pharmaceuticals useful in risk assessment, risk management, safety evaluation, and hazard characterization.

8 Expert Systems

Expert systems allow for the direct entry of a structure into software followed by the calculation or prediction without the requirement to compute descriptors and re-perform the modeling process. This makes expert system a more convenient option for toxicity prediction over traditional QSARs. Expert systems have been frequently employed by regulatory agencies, academia and industries worldwide for more efficient and fast prediction. The foremost criterion of toxicity prediction is to differentiate between toxicologically active and inactive molecules. Multiple mechanisms can lead to the identical toxic effect and this intricacy requires the

t3.1 **Table 3**

t3.2 **Most commonly employed test batteries (endpoints) for the modeling of ecotoxicity**

t3.3 Endpoints	Species	Portrayal
t3.4 Algae	*Chlorella vulgaris*	Chlorella is a unicellular green alga comprising a major component of the phytoplankton.
t3.5	*Chlorella pyrenoidosa*	
t3.6 t3.7 t3.8	*Pseudokirchneriella subcapitata* (*Selenastrum capricornutum*)	One of the prime producers of the aquatic ecosystem and ideal test organisms for toxicological studies. Ecotoxicity is measured by growth rate inhibition of green alga *P. subcapitata*.
t3.9 t3.10 t3.11	*Scenedesmus obliquus*	*Scenedesmus obliquus* (Chlorophyta, Chlorococcales) is a common cosmopolitan green alga, often occurring as almost a pure culture in fresh water plankton. It can grow in industrial wastewaters of different origins showing good adaptation ability and it is a very versatile microalga as a test battery.
t3.12 t3.13 t3.14 t3.15	*Scenedesmus vacuolatus*	A green alga of the Chlorophyceae. It is colonial and non-motile. The species has been used in the prediction of photoinduced toxicity of polycyclic aromatic hydrocarbons by in silico modelers. Also, Predictive modeling studies has been performed for the ecotoxicity of ionic liquids (ILs) towards the *Scenedesmus vacuolatus*.
t3.16 t3.17 Bacterium	*Escherichia coli*	A gram-negative, facultatively anaerobic, rod-shaped bacterium of the genus *Escherichia*. *E. coli* is frequently used as a model organism in microbiology and modeling studies.
t3.18 t3.19 t3.20	*Vibrio fischeri*	A gram-negative rod-shaped bacterium found globally in marine environments which has bioluminescent properties and is found predominantly in symbiosis with various marine animals. Predominantly employed in the research of microbial bioluminescence, quorum sensing, and bacterial–animal symbiosis.
t3.21 t3.22 t3.23	*Bacillus flexus*, *Pseudomonas fluorescens* and *Vibrio natriegens*	They are good model systems for studying marine biofouling.

Crustaceans	*Daphnia magna* *Daphnia pulex* *Daphnia ambigua* *Daphnia melanica*	*Daphnia* are members of the order Cladocera, and are one of the small aquatic crustaceans commonly called water fleas. These invertebrate species in aquatic food webs have been used as a representative test species for ecotoxicological evaluation of industrial chemicals. Generally, immobilization test is done by *Daphnia*.
	Thamnocephalus platyurus	Small crustaceans employed as a biological species for ecological toxicity testing of chemicals.
Duckweed	*Lemna minor*	*L. minor*, one form of aquatic vascular plant, is most commonly used among duckweeds for in silico models. Growth inhibition tests of duckweeds are used to identify the chemical toxicity.
	Lemna gibba	*Lemna gibba* is used in testing the phytotoxicity of pesticides and other environmental chemicals to higher plants.
Enzyme	Acetylcholinesterase	It plays the most important role in autonomic nervous system function. It catalyzes the hydrolysis of acetylcholinesters with a relative specificity for acetylcholine. Commonly, (a) enzyme inhibition data of the acetylcholinesterase from electric eel (*Electrophorus electricus*), (b) the AMP deaminase and (c) the antioxidant enzyme system of mouse liver are important for toxicity prediction and in silico model development.
Fish	Channel Catfish Ovary (CCO)	CCO is the cell line of choice for the propagation and diagnosis of Channel Catfish Virus (CCV). It is the standard for diagnosing Channel Catfish Virus Disease (CCVD) in farm reared Channel Catfish. Prediction of ILs has been performed by using this endpoint in recent time.
	Zebrafish (*Danio rerio*)	Zebrafish plays an important role in ecotoxicology as a prominent model vertebrate. It is standardized under the OECD and is employed to test chemicals, pharmaceuticals as well as industrial effluents.
	Fathead minnow (*Pimephales promelas*)	*Pimephales promelas* is the EPA recommended vertebrate species for freshwater chronic toxicity tests. In these tests, larvae are exposed for 7 days to different concentrations of effluent or to receiving water. Test results are based on the survival and weight of the larvae. As the fathead minnow is fairly tolerant of harsh conditions, it can be found in bodies of water that may be uninhabitable to other fish, such as waste drainage sites. It has also been studied to investigate the effects of these waste materials on the aquatic life. Effects induced by progestins are largely studied employing Fathead minnow.

(continued)

Table 3
(continued)

Endpoints	Species	Portrayal
Mammalian cells	Human keratinocyte cell line (HaCaT)	HaCaT cells are a spontaneously immortalized, human keratinocyte line that has been widely used for studies of skin biology and differentiation. In recent times, this cell line is modeled for the cytotoxicity prediction of metal oxide by different group of researchers.
	CaCo-2	A continuous cell of heterogeneous human epithelial colorectal adenocarcinoma cells. Permeability coefficients across the cellular membranes of Caco-2 cells are generally employed for modeling.
	HeLa	A prototypical cells of the human epithelium used in scientific research. It is the oldest and most commonly used human cell line which was derived from cervical cancer cells. Largely employed for anticancer activity.
	Prostate cancer cell line (PC3)	A human prostate cancer cell line used in modeling of prostate cancer inhibitors.
	Human malignant melanoma (Fem-X)	Not so popular, but recently used by group of authors to derive QSAR models.
	HT-29	HT-29 is a human colorectal adenocarcinoma cell line with epithelial morphology. These cells are sensitive to the chemotherapeutic drugs used in standard treatment for colorectal cancer.
	Rat cell line—IPC-81	Promyelotic leukemia rat cell line IPC-81 is frequently used in cytotoxicity assays of ILs.
Protozoa	*Tetrahymena thermophila* *Tetrahymena pyriformis*	Free-living unicellular ciliated protozoa and one of the largely employed endpoint for the assessment of the environmental toxicity.
Tadpoles	*Bufo vulgaris formosus* *Rana japonica* (Japanese brown frog)	Tadpoles, a common and sensitive species, the larva of the frogs. They are typical amphibious animals bridging the gap between aquatic and terrestrial animals. They are recurrently used for toxicity testing purposes and risk assessments and have been recommended by the EU-TGD.
Yeast	*Saccharomyces cerevisiae*	One form of budding yeast and one of the most intensively studied eukaryotic model organisms in molecular and cell biology. Small in size, accessible, reproduction time quick, and potentially economic. Considered as important species for ecotoxicity prediction.

t4.1 **Table 4**

t4.2 **List of available databases comprising information of the ecotoxicity due to pharmaceuticals[a]**

t4.3 Database name	Web accessibility	Information
t4.4 ACToR	http://actor.epa.gov/actor/faces/ACToRHome.jsp	Publicly available data on industrial chemicals, pesticides and drinking water contaminants by US EPA National Center for Computational Toxicology. The database consists of chemical structure, physicochemical values, and provides in vitro and in vivo toxicology data
t4.7 BDSM	http://systemsanalysis.louisville.edu/	Developmental toxicity database and open-source software for discovery of developmental toxicants by University of Louisville
t4.9 Cal/EPA	http://www.oehha.ca.gov/risk/ChemicalDB/index.asp	State of California EPA Toxicity Criteria Database of chronic reference exposure levels and cancer potency information
t4.11 CCRIS	http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?CCRIS	Chemical carcinogenesis research information system (CCRIS) created by US National Library of Medicine (NLM) contains carcinogenicity and mutagenicity test results for over 8000 chemicals
t4.14 CPDB	http://potency.berkeley.edu/	Carcinogenic potency database (CPDB) is a widely used resource of the results of 6540 chronic, long-term animal cancer tests on 1547 chemicals developed by University of California, Berkeley. The CPDB provides easy access to the bioassay literature and analyses of experiments that have been published over the past 50 years in the general literature through 2001 and by the National Cancer Institute/National Toxicology Program through 2004
t4.20 Danish (Q)SAR Database	http://ecbQSTR.jrc.ec.europa.eu/	Repository of estimates from over 70 QSTR models and health effects for 166,072 chemicals created by Danish EPA
t4.22 DEMETRA	http://www.demetra-tox.net/	DEMETRA is a EC-funded project in which a huge number of QSTR models have been developed for the prediction of different ecotoxicological endpoints namely rainbow trout LC_{50} (96 h), water flea LC_{50} (48 h), honey bee LD_{50} (48 h)
t4.25 DevTox	http://www.devtox.org	Developmental toxicity study data and control database for various strains of common laboratory animals

(continued)

Table 4
(continued)

Database name	Web accessibility	Information
DSSTox	http://www.epa.gov/ncct/dsstox/index.html	Distributed Structure-Searchable Toxicity database (DSSTox) developed by National Center for Computational Toxicology, US EPA. It is a network of downloadable, structure-searchable, standardized chemical structure files associated with toxicity data
ECOTOX	http://cfpub.epa.gov/ecotox/	Another USEPA database of single chemical toxicity information for aquatic and terrestrial life
ESIS	http://ecb.jrc.ec.europa.eu/esis/	European Chemical Substances Information system (ESIS) provides information on chemicals related to risk and safety
EXTOXNET	http://extoxnet.orst.edu/ghindex.html	EXtension TOXicology NETwork is a cooperative effort of University of California-Davis, Oregon State University, Michigan State University, Cornell University, and the University of Idaho. Databases related to pesticide toxicology. Pesticide Information Profiles (PIPs) and Toxicology Information Briefs (TIBs) can be obtained
GAC	http://www.niehs.nih.gov/research/resources/databases/gac/index.cfm	Genetic Alterations in Cancer (GAC) is a database that quantifies specific mutations found in cancers induced by environmental chemicals. Its created by US NIH/NIEHS
GAP	http://www.ils-inc.com/services/information-sciences	US EPA and IARC Genetic Activity Profile (GAP) database; it provides quantitative genotoxicity results of ≈500 chemicals to support hazard classification of human carcinogens
Gene-Tox	http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?GENETOX	Created by US NLM, consists of genetic toxicology test data for over 3000 chemicals
HERA	http://www.heraproject.com/RiskAssessment.cfm	Human and Environmental Risk Assessment (HERA) is a voluntary industry program in Brussels, Belgium. The database consists of toxicity assessment on ingredients and household cleaning products; toxicity and risk data on ingredients supplied and formulated by European manufacturers
Household Products Database	http://householdproducts.nlm.nih.gov/	Database with MSDSs of household products with health effects ratings and produce chemical information

IARC Monograph	http://monographs.iarc.fr/	International Agency for Research on Cancer Monograph (IARC) is developed by World Health Organization (WHO) which identifies environmental factors that can increase the risk of human cancer. These include chemicals, complex mixtures, occupational exposures, physical agents, biological agents, and lifestyle factors
IRIS	http://cfpub.epa.gov/ncea/iris/index.cfm	Integrated Risk Information System (IRIS) is under National Center for Environmental Assessment (NCEA), US EPA. The database is a compilation of electronic reports on 540 environmental chemical substances and their potential to cause human health effects
ISSCAN	http://www.iss.it/ampp/dati/cont.php?id=233&lang=1&tipo=7	ISSCAN is a database on chemical carcinogens (long-term carcinogenicity bioassay on rodents) created by Istituto Superiore di Sanità, Italy
ITER	http://www.tera.org/iter/	International Toxicity Estimates for Risk (ITER) is developed by TERA (Toxicity Excellence for Risk Assessment). The database consists of human health risk values and cancer classifications for over 600 environmental chemicals
IUCLID	http://iuclid.echa.europa.eu/	International Uniform ChemicaL Information Database (IUCLID) is a software application to capture, store, maintain and exchange data on intrinsic and hazard properties of chemical substances
JECDB	http://dra4.nihs.go.jp/mhlw_data/jsp/SearchPageENG.jsp	JECDB is a Chemical Toxicity Database by Japanese Ministry of Health, Labor and Welfare which contains toxicity test reports of environmental chemicals
JRC QSTR Database	http://ecb.jrc.ec.europa.eu/QSTR/background/	European Commission, Joint Research Centre's database of REACH relevant QSTRs
KATE	http://kate.nies.go.jp	KAshinhou Tool for Ecotoxicity (KATE) is created by Japanese National Institute for Environmental Studies (NIES), Ministry of the Environment (MoE), Government of Japan. It uses a structural domain named *C*-judgement and performs categorization of chemicals as potential hazards
LAZAR	http://lazar.in-silico.de/	Structure–Activity Relationships database provides QSTR predictions for liver toxicity, mutagenicity, and carcinogenicity

(continued)

Table 4
(continued)

Database name	Web accessibility	Information
MRL	http://www.atsdr.cdc.gov/mrls/index.html	Minimal Risk Levels for hazardous substances (MRL) is developed by US DHHS/ATSDR. The MRL is an estimate of the daily human exposure to a hazardous substance that is likely to be without appreciable risk of adverse non-cancer health effects over a specified duration of exposure
N-class database, KemI	http://apps.kemi.se/nclass/	Based on substance environmental hazard data (available), the N-Class database calculates aquatic classifications of preparations both according to Directive 1999/45/EEC and to the Globally Harmonized System (GHS)
NTP	http://ntp.niehs.nih.gov/	National Toxicology Program initiated by US NIH/NIEHS
OECD HPV database	http://cs3-hq.oecd.org/scripts/hpv/	It stores at least a minimum set of data (including acute aquatic toxicity) necessary to determine a potential hazard. Compounds are searchable by chemical name, CAS, sponsoring country and stage
RAIS	http://rais.ornl.gov/	Risk Assessment Information System (RAIS); it deals with chemical-specific toxicity values Sponsored by the US Department of Energy (DOE), Office of Environmental Management, Oak Ridge Operations (ORO) Office through a contract between Bechtel Jacobs Company LLC and the University of Tennessee
Riskline, KemI	http://apps.kemi.se/riskline/	It contains information on both environment and health. The database is produced by the Swedish Chemicals Inspectorate, Sweden
RITA	http://www.item.fraunhofer.de/reni/public/rita/index.php	Registry of Industrial Toxicology Animal-data (RITA) is generated by Fraunhofer Institute of Toxicology and Experimental Medicine (ITEM) Hannover for comparing and interpreting rodent carcinogenicity studies and tumor data. As per May 2011, RITA includes total toxicity case-studies of 60,896 on rat and 25,335 on mice
SCOGS	http://www.fda.gov/Food/FoodIngredientsPackaging/GenerallyRecognizedasSafeGRAS/GRASSubstancesSCOGSDatabase/default.htm	Database resource on toxicology and safety reports made by the selected Committee on GRAS Substances by US FDA/CFSAN

STITCH	http://stitch.embl.de/	Search Tool for Interactions of Chemicals (STITCH) is developed at the Novo Nordisk Foundation Center for Protein Research; University of Copenhagen, European Molecular Biology Laboratory, the Swiss Institute of Bioinformatics, the Biotechnology Center (BIOTEC) of the Technische Universität Dresden and the University of Zurich. It is a knowledge database to explore known and predicted interactions between proteins and small-molecule chemicals for understanding of molecular and cellular functions. Contains interactions for over 74,000 small molecules and over 2.5 million proteins in 630 organisms
TEXTRATOX	http://www.vet.utk.edu/TETRATOX/index.php	TEXTRATOX is a product of the University of Tennessee Institute of Agriculture. It is a collection of aquatic toxic potency data for more than 2400 industrial organic compound
TOXNET	http://toxnet.nlm.nih.gov/	Databases on toxicology, hazardous chemicals, environmental health, and toxic releases by US NLM
ToxRefDB	http://www.epa.gov/ncct/toxrefdb/	Database of standard toxicity test results for pesticides and other environmental chemicals including acute, subchronic, chronic, reproductive, and developmental toxicity in support of the ToxCast program by US EPA
Toxtree	http://ecb.jrc.ec.europa.eu/QSTR/QSTR-tools/index.php?c=TOXTREE	Open-source application that places chemicals into categories and predicts various kinds of toxic effects by applying decision tree approaches created by European Commission, Joint Research Centre
TSCATS	http://www.ntis.gov/products/ots.aspx	Toxic Substances Control Act Test Submissions (TSCATS) is an online database of chemical testing results and adverse effects of chemicals on health and ecological systems constructed by US Department of Commerce National Technical Information Service Alexandria, Virginia. The collection currently exceeds 25,000 titles of studies that are submitted to the US Environmental Protection Agency by US Industry under several section of the Toxic Substance Control Act
USGS	http://137.227.231.90/data/acute/acute.html	US Geological Survey (USGS), Columbia Environmental Research Center developed the database for the aquatic acute toxicity tests
WikiPharma	www.wikipharma.org	WikiPharma provides an easily accessible, comprehensive and up-to-date overview of effects caused by pharmaceuticals on nontarget organisms developed within the Swedish research program MistraPharma (www.mistrapharma.se). The database currently contains basic information for 831 active pharmaceutical ingredients (APIs) representing 35 different drug classes. Effect data have been identified and included for 116 of these substances and ecotoxicity test data have been extracted from 156 different sources

[a]Note that many database contains pharmaceuticals, organic pollutants, agrochemicals and pesticides under the categories of chemicals

accessibility of predictive tools that are able to discriminate manifold regions in the activity space. This necessitates the development of so-called expert systems, which try to cover broader structural and activity regions in comparison to the local models. Table 5 summarizes different freely available and commercial expert systems to predict endpoints related toxicity predictions.

9 Green and Ecological Pharmacy

The role of green chemistry and principles are very important for risk management of pharmaceuticals. The principles of green chemistry state that the functionality of a chemical should not only comprise the properties of a chemical essential for its application, but also quick and trouble free degradability after its usage. Improvement of synthesis and renewable feedstock are very important issues for preparation of environment friendly pharmaceuticals. Employing these principles and the awareness of green chemistry to pharmaceuticals are necessary [138]. In this perspective, a system called "benign by design" can be considered which means easy degradability after application is considered even before a pharmaceutical's synthesis. This approach is not completely new. For instance, it is a general practice during the development of pharmaceuticals that adverse side effects are to be taken into consideration. This can also result in economic rewards in the long run and will fit into green pharmacy [170]. But one has to note that a pharmaceutical may also lose its specific therapeutic action due to the structural modification while introducing green chemistry. However, this approach can be employed at least for the optimized and new synthesis routes [170]. Again, it is true that finding good lead compounds is a major task even without considering the environment toxicity issue. However, there is no requirement to find a new lead compound at first. The modification of known lead structures can be the best option to do. Responding to the green and justifiable pharmacy challenge may also result in new marketing opportunities with help of appropriate and scientific research within industry and academia.

10 Overview and Conclusion

A huge number of reports and publications have been published in the last decade about the ecotoxicity due to human and veterinary pharmaceuticals, but it is still too meager to permit us to execute a systematic and precise risk assessment and appropriate risk management. There is still a huge need of filling the missing data gaps in our knowledge. Due to biologically active and persistence nature of pharmaceuticals, they are one of the most serious threats to

Table 5
A complete list of exert systems to predict various endpoints of ecotoxicity

Expert system	Website	Explanatory note
ASTER (ASsessment Tools for the Evaluation of Risk)	http://www.epa.gov/med/Prods_Pubs/aster.htm	ASTER is an integration of ECOTOX database and a structure–activity based expert system. ASTER is freely available to provide high quality data for discrete chemicals, when available in the associated databases (i.e., ECOTOX and EcoChem), and QSTR models estimate when data are lacking.
CAESAR (Computer Assisted Evaluation of industrial chemical Substances According to Regulations)	http://www.caesar-project.eu/	CAESAR is an EC funded project which is specifically dedicated to develop QSTR models for the REACH legislation. Five endpoints considered under CAESAR are bioconcentration factor, skin sensitization, carcinogenicity, Mutagenicity, developmental toxicity.
DEREK (Deductive estimation of risk from existing knowledge)	http://www.lhasalimited.org/indexphp?cat=2&sub_cat=64	DEREK, a Knowledge-based expert system, developed in collaboration with industrial partners, which makes it predictions based on structural alerts, reasoning rules and examples contained within its knowledge base. Currently 21 structural alerts for teratogenicity, or teratogenic endpoints are considered under this expert system.
DfW	–	Knowledge-based expert system created with knowledge of structure–toxicity relationships. It consists of 361 alerts covering a wide range of toxicological endpoints. The skin sensitization knowledge base in DfW was initially developed in collaboration with Unilever in 1993 using its historical database of GPMT data for 294 chemicals. The DfW version 9.0.0 contains 64 alerts for skin sensitization.
ECOSAR (ECOlogical Structure–Activity Relationships)	http://www.epa.gov/oppt/exposure/docs/episuitedl.htm	ECOSAR is freely available from the US EPA which utilizes a number of class-specific log K_{ow}-based QSTRs to predict the toxicity (both short-term and long-term) of chemicals. Hazard assessment of environmentally occurring pharmaceuticals to fish, daphnids, and green algae can be performed employing ECOSAR.

(continued)

Table 5
(continued)

Expert system	Website	Explanatory note
HazardExpert Pro	http://www.compudrug.com/	Teratogenicty, reproductive toxicity predicted based on structural fragments.
MCASE/MC4PC	http://www.multicase.com/products/products.htm	MCASE is a knowledge-based commercial system which develops QSTR models and evaluates the structural features for non-congeneric molecules and identifies the substructures that may be responsible for the response. MCASE contains models for several fish species (blue gill, FHM, rainbow trout, red killifish). There are more than 180 modules covering various areas of toxicology and pharmacology endpoints including skin sensitization currently marketed by MultiCASE Inc. Available databases are as follows: Retinoids (76 compounds); developmental toxicity (66 antifungal triazole alcohols; composite dataset of 275 compounds); developmental toxicants (mouse 101, rat 134, rabbit 66, humans 119, hamster 192 compounds); FDA/TERIS developmental toxicity (humans 323 compounds); developmental toxicants in FDA teratogenicity (rabbit 812, rat 1286, mouse 794, miscellaneous mammal 1409 compounds)
OASIS & TIMES	http://www.oasis-lmc.org/software.php	OASIS is commercial software which uses the response-surface approach for modeling acute toxicity for two types of toxico-chemical domains: reversible acting chemicals (noncovalent ones) and irreversible (covalent ones) bioreactive chemicals. Interspecies correlations for acute toxicity to 17 aquatic species, such as fish, snail, tadpole, hydrozoan, crustacean, insect larvae, and bacteria have been developed. The TIssue MEtabolism Simulator (TIMES) platform is used to predict the individual and interspecies models for acute aquatic toxicity.
OECD (Q)SAR Application Toolbox	http://www.oecd.org/document/23/0,3343,en_2649_34379_33957015_1_1_1_1,00.html	It allows the user to develop categories and perform read-across, QSTR and trend analyses. A platform that will allow chemical information management, similarity searches and toxicological profiling.
ONCOLOGIC	http://www.epa.gov/oppt/sf/pubs/oncologic.htm	OncoLogic is a desktop computer program that evaluates the probability that a chemical may induce cancer. OncoLogic predicts cancer-causing potential by applying the rules of structure activity relationship (SAR) analysis, mimicking the decision logic of human experts, and incorporating knowledge of how chemicals cause cancer.

OSIRIS property explorer	http://www.organic-chemistry.org/prog/peo/tox.html	OSIRIS is an on-line system which predicts reproductive effects on the basis of structural fragments which developed from the analysis of 3570 compounds with reproductive effects listed in RTECS.
PASS	http://ibmc.p450.ru/PASS//	PASS assesses similarity of molecules to those with known activity. It predicts over 30 endpoints relevant to reproductive toxicity. The employed endpoints are abortion inducer, alkylator, carcinogenic, DNA intercalator, DNA repair enzyme inhibitor, DNA synthesis inhibitor, DNA topoisomerase ATP hydrolyzing inhibitor, DNA topoisomerase inhibitor, DOPA decarboxylase inhibitor, embryotoxic, estradiol 17β-dehydrogenase stimulant, estrogen agonist, estrogen antagonist, ER modulator, estrone sulfatase inhibitor, estrone sulfotransferase stimulant, fertility enhancer, gynecological disorders treatment, menopausal disorders treatment, mutagenic, retinoic acid α-receptor agonist, retinoic acid β-receptor agonist, retinoic acid receptor agonist, retinoic acid receptor antagonist, retinoid X receptor agonist, retinoid acid receptor agonist, spermicide, teratogen, testosterone agonist, toxic, uterine relaxant, uterine stimulant.
SARET (Structure–Activity Relationships for Environmental Toxicology)	http://www.ibmh.msk.ru	SARETbase and SARETmodel are used as computer programs for computation of descriptors. SARETbase includes the information on more than 190 characteristics for 8500 substances: chemical structure, physicochemical properties (density, boiling and melting points, partition coefficients of octanol–water, etc.), adverse effect doses and concentrations for acute and chronic exposure. The SARETmodel is prepared for statistical analysis of data and calculation of unknown parameters of substances on the basis of (Q)SARs. The application of SARET provides the information essential to assess the hazard of chemicals and to approximate their unknown characteristics.
TERA (Tools for environmental risk assessment)	http://www.ogm-dss.isprambiente.it/index.xhtml	TÉRA is based on a fuzzy inference engine using the personal experience and knowledge of ERA experts. TERA includes the information on approximately 200 characteristics for more than 13,000 chemical substances. All information collected in SARET and TERA is verified and specified on the basis of both Russian and foreign literature data including official documents, open publications. In addition, TERA contains information for 194 mixtures, 182 polymers, 346 dyes, 1080 non-organic compounds, 1407 remedies, 1260 agrochemicals. More than 1000 compounds contained in TERA are not presented in the Registry of Toxic Effects of Chemical Substances (RTECS) TERA contains information useful for human, environmental and ecological risk assessment and management. TERA is a tool for assessment of multi-domain risk, assessment of carcinogenic potency risk, Prediction of lead concentrations in blood of fetus, children, adults (system LRISK), health risk connected with lead exposure, prediction of emission of chemical substances and there distribution in different media, parameters used for setting priority of chemical substances in risk assessment and risk assessment using epidemiological data

(continued)

Table 5
(continued)

Expert system	Website	Explanatory note
TerraQSTR-FHM	http://www.terrabase-inc.com	It is a commercial software and a stand-alone neural network-based program to compute the acute toxicity of organic chemicals to the FHM using a proprietary neural network algorithm
TIMES-SS (Times MEtabolism Simulator platform)	Marketed by LMC University "As Zlatarov," Bourgas, Bulgaria	TIMES-SS is a hybrid expert system which can encode structure toxicity and structure metabolism relationships through a number of transformations simulating skin metabolism and interaction of the generated reactive metabolites with skin proteins. The skin metabolism simulator mimics metabolism using 2D structural information. The covalent reactions with proteins are described by 47 alerting groups.
TOPKAT (TOxicity Prediction by C(K)omputer Assisted Technology)	http://accelrys.com/products/discovery-studio/admet.html	TOPKAT is a statistical commercial expert system. Under TOPKAT, QSTR models developed from a huge number of heterogeneous databases of toxicological information using substructural fragments and (electro)-topological indices. Developmental toxicity potential are taken from FDA/TERIS. The program uses a range (Q)SAR models for assessing acute toxicity to FHM and Daphnia. The TOPKAT *LD50* (acute oral toxicity) modeling approach has been used by the Danish EPA in their project to develop QSTR models for evaluation of dangerous properties of around 47,000 organic substances on the European Inventory of Existing Commercial chemical Substances (EINECS) list.
Toxmatch	http://ecb.jrc.ec.europa.eu/QSTR/QSTR-tools/index.php?c=TOXMATCH	An open-source computer program of Joint Research Centre (EC) that encodes several chemical similarity indices in order to facilitate the grouping of chemicals, thereby supporting the development of chemicals categories and the application of read-across between analogues.

human health and environment stability. Additionally, their specific modes of action and specific effects on living systems make pharmaceuticals distinctly different from other chemicals. This sole feature is sufficient reason to assess the potential effects of pharmaceuticals in diverse environmental compartments. The problem is more horrifying as the occurrence level of pharmaceuticals in different environmental compartments is largely varied. The variations in drug occurrences from country to country and also within the different regions of a country make the assessment of pharmaceuticals a troublesome job for the environmental scientist. The interactions between pharmaceuticals and natural stressors of aquatic and terrestrial communities remain to be unexplained. Along with that, the proper risk assessment of mixtures of pharmaceutical products is another area where more introspection is required in present times.

In this book chapter, the hazardous effects of the most common therapeutic classes of pharmaceutical to the living ecosystems and environment are discussed. Furthermore, specific information on the sources, fate, and effects of pharmaceuticals in the environment and their possible negative impact on different ecosystems are explored. There is a lack of sufficient information and scientific data on effects of long-term exposure to nontarget organisms. It is also important to assess the presence of pharmaceuticals and their metabolites and transformation products in several environmental compartments. One can find only a few reports on the quantitative effects of pharmaceuticals, but the effects of metabolites are not sufficiently explored by the scientific community. One has to accept that the identification of risk assessment and management are not sufficient if they are not properly implemented in right way. In these perspectives, the major role should be played by government authorities and agencies by implementing various guidelines and rules for the reduction of toxicity of pharmaceuticals to the environment.

Scarcity of adequate ecotoxicity data related to the diverse classes of pharmaceuticals and their metabolites has stalled appropriate computational modeling and development of expert systems. As a consequence, there are only a very limited number of models developed so far for the risk assessment of pharmaceuticals and their metabolites as well as for the pharmaceutical mixtures. Hence, a sufficient number of models should be developed to address the risk assessment and risk management in an efficient way by minimizing the requirement of time, animal testing and cost. This will also help in gathering the ecotoxicity data as soon as a new pharmaceutical product comes to the market. In this perspective, expert systems are more reliable and results may be easily available in no time. There is a need of more expert systems for prediction of toxicity of pharmaceuticals from diverse classes of therapeutic actions and their metabolites against different endpoints. It is true

that in silico techniques cannot substitute "wet" experiments but both of them can be utilized together for a better risk management of pharmaceuticals in near future.

References

1. Kümmerer K (2001) Pharmaceuticals in the environment: sources, fate, effects and risks. Springer-Verlag, Berlin
2. Halling-Sørensen B, Nors Nielsen S, Lanzky PF et al (1998) Occurrence, fate and effects of pharmaceutical substances in the environment—a review. Chemosphere 36(2):357–393
3. Santosa LHMLM, Araújoa AN, Fachinia A et al (2010) Ecotoxicological aspects related to the presence of pharmaceuticals in the aquatic environment. J Hazard Mater 175:45–95
4. Li WC (2014) Occurrence, sources, and fate of pharmaceuticals in aquatic environment and soil. Environ Pollut 187:193–201
5. Rivas J, Encinas A, Beltran F, Grahan N (2011) Application of advanced oxidation processes to doxycycline and norfloxacin removal from water. J Environ Sci Health A Tox Hazard Subst Environ Eng 46:944–951
6. Kumar K, Gupta SC, Chander Y et al (2005) Antibiotic use in agriculture and its impact on the terrestrial environment. Adv Agron 87:1–54
7. Vidaver AK (2002) Uses of antimicrobials in plant agriculture. Clin Infect Dis 34(Suppl 3):S107–S110
8. Rooklidge SJ (2004) Environmental antimicrobial contamination from terraccumulation and diffuse pollution pathways. Sci Total Environ 325:1–13
9. Ternes TA, Bonerz M, Schmidt T (2001) Determination of neutral pharmaceuticals in wastewater and rivers by liquid chromatography electrospray tandem mass spectrometry. J Chromatogr A 938:175–185
10. Kolpin DW, Furlong ET, Meyer MT et al (2002) Pharmaceuticals, hormones, and other organic wastewater contaminants in U.S. streams, 1999–2000: a national reconnaissance. Environ Sci Technol 36:1202–1211
11. Daughton CG, Ternes TA (1999) Pharmaceuticals and personal care products in the environment: agents of subtle change? Environ Health Perspect 107:907–937
12. Adler P, Steger-Hartmann T, Kalbfus W (2001) Distribution of natural and synthetic estrogenic steroid hormones in water samples from southern and middle Germany. Acta Hydrochim Hydrobiol 29:227–241
13. Ternes T (1998) Occurrence of drugs in German sewage treatment plants and rivers. Water Res 32:3245–3260
14. Ahrer W, Scherwenk E, Buchberger W (2001) Occurrence and fate of fluoroquinolone, macrolide, and sulphonamide antibiotics during wastewater treatment and in ambient waters in Switzerland. In: Daughton CG, Jones-Lepp T (eds) Pharmaceuticals and personal care products in the environment: scientific and regulatory issues, Symposium series 791. American Chemical Society, Washington, DC, pp 56–69
15. Heberer T (2002) Tracking persistent pharmaceutical residues from municipal sewage to drinking water. J Hydrol 266:175–189
16. Rabiet M, Togola A, Brissaud F et al (2006) Consequence of wastewater disposal on the contamination of the water resource by pharmaceuticals in a Mediterranean basin. Environ Sci Technol 40:5282–5288
17. Zuccato E, Calamari D, Natangelo M et al (2000) Presence of therapeutic drugs in the environment. Lancet 355:1789–1790
18. Castiglioni S, Zuccato E, Chiabrando C et al (2008) Mass spectrometric analysis of illicit drugs in wastewater and surface water. Mass Spectrom Rev 27:378–394
19. Schultz MM, Furlong ET (2008) Trace analysis of antidepressant pharmaceuticals and their select degradates in aquatic matrixes by LC/ESI/MS/MS. Anal Chem 80:1756–1762
20. Metcalfe CD, Miao XS, Koenig BG et al (2003) Distribution of acidic and neutral drugs in surface waters near sewage treatment plants in the lower Great Lakes, Canada. Environ Toxicol Chem 22:2881–2889
21. Frick EA, Henderson AK, Moll DM, et al (2001) Presented at the Georgia Water Resources Conference, Athens, GA
22. Tauber R (2003) Quantitative analysis of pharmaceuticals in drinking water from ten Canadian cities. Enviro-Test Laboratories, Winnipeg, Manitoba
23. Loraine GA, Pettigrove ME (2006) Seasonal variations in concentrations of pharmaceuticals and personal care products in drinking water and reclaimed wastewater in Southern California. Environ Sci Technol 40:687–695

24. Fink L, Dror I, Berkowitz B (2012) Enrofloxacin oxidative degradation facilities by metal oxide nanoparticles. Chemosphere 86:144–149
25. Hellweger FL, Ruan X, Cherchia E et al (2011) Applicability of standard antibiotic toxicity testes to the ambient aquatic environment. Ann Environ Sci 5:61–66
26. Seiler JP (2002) Pharmacodynamic activity of drugs and ecotoxicology can the two be connected? Toxicol Lett 131:105–115
27. European Commission, Directive 2006/121/EC of the European Parliament and of the Council of 18 December 2006 amending Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances in order to adapt it to Regulation (EC) No. 1907/2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) and establishing a European Chemicals Agency. Off J Eur Union, L 396/850 of 30.12.2006, Office for Official Publications of the European Communities (OPOCE), Luxembourg
28. EMEA (2006) Guideline on the environmental impact assessment of medicinal products for human use. European Agency for the Evaluation of Medicinal Products, London, UK (Report no. CPMP/SWP/4447/00)
29. FDA-CDER (1998) Guidance for industry-environmental assessment of human drugs and biologics applications, revision 1. FDA Center for Drug Evaluation and Research, Rockville
30. FDA: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER) (1998) CMC 6—revision 1
31. Boxall ABA, Kolpin D, Halling Sørensen B et al (2003) Are veterinary medicines causing environmental risks. Environ Sci Technol 36:286–294
32. Ternes TA, Hirsch R (2000) Occurrence and behavior of X-ray contrast media in sewage facilities and the aquatic environment. Environ Sci Technol 34:2741–2748
33. Persson M, Sabelström E, Gunnarsson B (2009) Handling of unused prescription drugs-knowledge, behaviour and attitude among Swedish people. Environ Int 35:771–774
34. Li D, Yang M, Hu J et al (2008) Determination and fate of oxytetracycline and related compounds in oxytetracycline production wastewater and the receiving river. Environ Toxicol Chem 27:80–86
35. José Gómez M, Petrovic M, Fernández-Alba AR et al (2006) Determination of pharmaceuticals of various therapeutic classes by solid-phase extraction and liquid chromatography-tandem mass spectrometry analysis in hospital effluent wastewaters. J Chromatogr 1114:224–233
36. Kreuzig R, Höltge S, Brunotte J et al (2005) Test plat studies on runoff of sulfonamides from manured soil after sprinkler irrigation. Environ Toxicol Chem 24:777–781
37. Xia K, Bhandari A, Das K et al (2005) Occurrence and fate of pharmaceuticals and personal care products (PPCPs) in biosolids. J Environ Qual 34:91–104
38. Serrano PH (2005) Responsible use of antibiotics in aquaculture. Fisheries technical paper 469. Food and Agriculture Organization of the United Nations (FAO), Rome
39. http://www.apsnet.org/online/feature/Antibiotics/. Accessed 4 June 2015
40. http://www.esac.ua.ac.be/main.aspx?c=*ESAC2&n=1066l. Accessed 4 June 2015
41. Cleuvers M (2003) Aquatic ecotoxicity of pharmaceuticals including the assessment of combination effects. Toxicol Lett 142:185–194
42. Harnagea-Theophilus E, Gadd SL, Knight-Trent HA et al (1999) Acetaminophen-induced proliferation of breast cancer cells involves estrogen receptors. Toxicol Appl Pharmacol 155:273–279
43. Halling-Sørensen B (2000) Algal toxicity of antibacterial agents used in intensive farming. Chemosphere 40:731–739
44. Ferrari B, Paxeus N, Lo Giudice R et al (2003) Ecotoxicological impact of pharmaceuticals found in treated wastewaters: study of carbamazepine, clofibric acid, and diclofenac. Ecotoxicol Environ Saf 55(3):359–370
45. Cleuvers M (2004) Mixture toxicity of the anti-inflammatory drugs diclofenac, ibuprofen, naproxen and acetylsalicylic acid. Ecotoxicol Environ Saf 59:309–315
46. Verenitch SS, Lowe CJ, Mazumder A (2006) Determination of acidic drugs and caffeine in municipal wastewaters and receiving waters by gas chromatography-ion trap tandem mass spectrometry. J Chromatogr A 1116:193–203
47. Bendz D, Paxéus NA, Ginn TR et al (2005) Occurrence and fate of pharmaceutically active compounds in the environment, a case study: Höje River in Sweden. J Hazard Mater 122:195–204
48. Calamari D, Zuccato E, Castiglioni S et al (2003) Strategic survey of therapeutic drugs in the Rivers Po and Lambro in Northern Italy. Environ Sci Technol 37:1241–1248

49. Barnes KK, Kolpin DW, Furlong ET et al (2008) A national reconnaissance of pharmaceuticals and other organic wastewater contaminants in the United States: I) Groundwater. Sci Total Environ 402:192–200
50. Pounds N, Maclean S, Webley M et al (2008) Acute and chronic effects of ibuprofen in the mollusc Planorbis carinatus (Gastropoda: Planorbidae). Ecotoxicol Environ Saf 70: 47–52
51. Hernando MD, Heath E, Petrovic M et al (2006) Trace-level determination of pharmaceuticals residues by LC–MS/MS in natural and treated waters. A pilot-survey study. Anal Bioanal Chem 385:985–991
52. Kim JW, Ishibashi H, Yamauchi R et al (2009) Acute toxicity of pharmaceutical and personal care products on freshwater crustacean (Thamnocephalus platyurus) and fish (Oryzias latipes). J Toxicol Sci 34:227–232
53. Isidori M, Lavorgna M, Nardelli A (2005) Ecotoxicity of naproxen and its phototransformation products. Sci Total Environ 48: 93–101
54. Kim Y, Choi K, Jung J et al (2007) Aquatic toxicity of acetaminophen, carbamazepine, cimetidine, diltiazem and six major sulfonamides, and their potential ecological risks in Korea. Environ Int 33:275–370
55. Quinn B, Gagné F, Blaise C (2008) An investigation into the acute and chronic toxicity of eleven pharmaceuticals (and their solvents) found in wastewater effluent on the cnidarian, Hydra attenuate. Sci Total Environ 389: 306–314
56. Zurita JL, Repetto G, Jos A et al (2007) Toxicological effects of the lipid regulator gemfibrozil in four aquatic systems. Aquat Toxicol 81:106–115
57. Miao XS, Metcalfe CD (2003) Determination of cholesterol-lowering statin drugs in aqueous samples using liquid chromatography-electrospray ionization tandem mass spectrometry. J Chromatogr A 998:133–141
58. Park S, Choi K (2008) Hazard assessment of commonly used agricultural antibiotics on aquatic ecosystems. Ecotoxicology 17: 526–538
59. Heckamann LH, Callaghan A, Hooper HL et al (2007) Chronic toxicity of ibuprofen to Daphnia magna: effects on life history traits and population dynamics. Toxicol Lett 172:137–145
60. Isidori M, Lavorgna M, Nardelli A et al (2005) Toxic and genotoxic evaluation of six antibiotics on non-target organisms. Sci Total Environ 346:87–98
61. Vulliet E, Wiest L, Baudot R et al (2008) Multi-residue analysis of steroids at sub-ng/L levels in surface and ground-waters using liquid chromatography coupled to tandem mass spectrometry. J Chromatogr A 1210:84–91
62. Zuehlke S, Duennbier U, Heberer T (2005) Determination of estrogenic steroids in surface water and wastewater by liquid chromatography-electrospray tandem mass spectrometry. J Sep Sci 28:52–58
63. Pawlowski S, van Aerle R, Tyler CR et al (2004) Effects of 17α-ethinylestradiol in a fathead minnow (Pimephales promelas) gonadal recrudescence assay. Ecotoxicol Environ Saf 57:330–345
64. Winter MJ, Lillicrap AD, Caunter JE et al (2008) Defining the chronic impacts of atenolol on embryo-larval development and reproduction in the fathead minnow (Pimephales promelas). Aquat Toxicol 86:361–369
65. Péry ARR, Gust M, Vollat B et al (2008) Fluoxetine effects assessment on the life cycle of aquatic invertebrates. Chemosphere 73:300–304
66. Vasskog T, Anderssen T, Pedersen-Bjergaard S et al (2008) Occurrence of selective serotonin reuptake inhibitors in sewage and receiving waters at Spitsbergen and in Norway. J Chromatogr A 1185:194–205
67. Grung M, Källqvist T, Sakshaug S et al (2008) Environmental assessment of Norwegian priority pharmaceuticals based on the EMEA guideline. Ecotoxicol Environ Saf 71:328–340
68. Frade VMF, Dias M, Teixeira ACSC et al (2014) Environmental contamination by fluoroquinolones. Braz J Pharm Sci 50:41–54
69. Ye Z, Weinberg HS, Meyer MT (2007) Trace analysis of trimethoprim and sulfonamide, macrolide, quinolone, and tetracycline antibiotics in chlorinated drinking water using liquid chromatography electrospray tandem mass spectrometry. Anal Chem 79:1135–1144
70. Moder M, Braun P, Lange F et al (2007) Determination of endocrine disrupting compounds and acidic drugs in water by coupling of derivatization, gas chromatography and negative-chemical ionization mass spectrometry. Clean Soil Air Water 35:444–451
71. Mompelat S, Le Bot B, Thomas O (2009) Occurrence and fate of pharmaceutical products and by-products, from resource to drinking water. Environ Int 35:803–814
72. Martín J, Camacho-Muñoz D, Santos JL et al (2012) Occurrence of pharmaceutical compounds in wastewater and sludge from wastewater treatment plants: removal and ecotoxicological impact of wastewater dis-

charges and sludge disposal. J Hazard Mater 239–240:40–47

73. Martínez-Carballo E, González-Barreiro C, Scharf S et al (2007) Environmental monitoring study of selected veterinary antibiotics in animal manure and soils in Austria. Environ Pollut 148:570–579
74. Hamscher G, Sczesny S, Höper H et al (2002) Determination of persistent tetracycline residues in soil fertilized with liquid manure by high-performance liquid chromatography with electrospray ionization tandem mass spectrometry. Anal Chem 74:1509–1518
75. Ho YB, Zakaria MP, Latif PA et al (2012) Simultaneous determination of veterinary antibiotics and hormone in broiler manure, soil and manure compost by liquid chromatography-tandem mass spectrometry. J Chromatogr A 1262:160–168
76. Hamscher G, Pawelzick HT, Sczesny S et al (2003) Antibiotics in dust originating from a pig fattening farm: a new source of health hazard for farmers? Environ Health Perspect 111:1590–1594
77. Kümmerer K, Alexy R, Hüttig J (2004) Standardized tests fail to assess the effects of antibiotics on environmental bacteria. Water Res 38:2111–2116
78. Carucci A, Cappai G, Piredda M (2006) Biodegradability and toxicity of pharmaceuticals in biological wastewater treatment plants. J Environ Sci Health A Tox Hazard Subst Environ Eng 41:1831–1842
79. Hernando MD, DeVettori S, Martínez-Bueno MJ et al (2007) Toxicity evaluation with *Vibrio fisheri* test of organic chemicals used in aquaculture. Chemosphere 68:724–730
80. Thomulka KW, McGee DJ (1993) Detection of biohazardous materials in water by measuring bioluminescence reduction with the marine organism *Vibrio harveyi*. J Environ Sci Health A28:2153–2166
81. Brain RA, Johnson DJ, Richards SM (2004) Microcosm evaluation of the effects of an eight pharmaceutical mixture to the aquatic macrophytes Lemna gibba and Myriophyllum sibiricum. Aquat Toxicol 70:23–40
82. Sengelov G, Agerso Y, Halling-Sørensen B et al (2001) Bacterial antibiotic resistance levels in farmland as a result of treatment with pig slurry. Environ Int 28:587–595
83. Hansen AJ, Jensen J, Krogh PH (2000) Effects of the antibiotics oxytetracycline and tylosin on soil fauna. Chemosphere 40:751–757
84. Takai H, Pedersen S, Johnsen JO et al (1998) Concentrations and emissions of airborne dust in livestock buildings in northern Europe. J Agric Eng Res 70:59–77
85. Holt D, Harvey D, Hurley R (1993) Chloramphenicol toxicity. Adverse Drug React Toxicol Rev 12:83–95
86. Harrison PF, Lederberg J (1998) Antimicrobial resistance, issues and options. National Academic Press, Washington, DC
87. Oaks JL, Gilbert M, Virani MZ et al (2004) Diclofenac residues as the cause of population decline of vultures in Pakistan. Nature 427:630–633
88. Schwabe U, Paffrath D (2003) Arzneiverordnungs report. Springer Verlag, Berlin, Germany
89. Ferrari B, Mons R, Vollat B et al (2004) Environmental risk assessment of six human pharmaceuticals: are the current environmental risk assessment procedures sufficient for the protection of the aquatic environment? Environ Toxicol Chem 23:1344–1354
90. Schwaiger J, Ferling H, Mallow U et al (2004) Toxic effects of the non-steroidal anti-inflammatory drug diclofenac: part I: histopathological alterations and bioaccumulation in rainbow trout. Aquat Toxicol 68:141–150
91. Isidori M, Nardelli A, Pascarella L et al (2007) Toxic and genotoxic impact of fibrates and their photoproducts on non-target organism. Environ Int 33:635–641
92. Nunes B, Carvalho F, Guilhermino L (2004) Acute and chronic effects of clofibrate and clofibric acid on the enzymes acetylcholinesterase, lactate dehydrogenase and catalase of the mosquitofish, *Gambusia holbrooki*. Chemosphere 57:1581–1589
93. Santamarina MT, Leiro J, Vilanova R et al (1987) The immunosuppressive effect of clofibrate and cyclosporine on IgM and IgE antibodies induced by *Trichinella spiralis*. Rev Iber Parasitol 47:289–293
94. Triebskorn R, Casper H, Scheil V et al (2007) Ultrastructural effects of pharmaceuticals (carbamazepine, clofibric acid, metoprolol, diclofenac) in rainbow trout (Oncorhynchus mykiss) and common carp (Cyprinus carpio). Anal Bioanal Chem 387:1405–1416
95. Misra SK, Sharma AK, Mehdi H et al (1986) Effect of cholesterol and alpha p-chlorophenoxyisobutyyrate on virulence in *Acanthamoeba culbertsoni* strain A-1 and C-7. Int J Parasitol 16:191–196
96. Vandewaa EA, Mills G, Chen GZ et al (1989) Physiological role of HMG-CoA reductase in regulating egg production by *Schistosoma mansoni*. Am J Physiol Regul Integr Comp Physiol 257:618–625

97. Huggett DB, Brooks BW, Peterson B et al (2002) Toxicity of select beta adrenergic receptor-blocking pharmaceuticals (β-blockers) on aquatic organisms. Arch Environ Contam Toxicol 43:229–235

98. Haider S, Baqri SSR (2000) β-Adrenoceptor antagonists reinitiate meiotic maturation in Clarias batrachus oocytes. Comp Biochem Physiol A Mol Integr Physiol 126:517–525

99. Ohnishi ST, Sadanaga KK, Katsuoka M et al (1989) Effects of membrane acting-drugs on *Plasmodium* species and sickle cell erythrocytes. Mol Cell Biochem 91:159–165

100. Coppi A, Merali S, Eichinger D (2002) The enteric parasite *Entamoeba* uses an autocrine catecholamine system during differentiation into the infectious cysts stage. J Biol Chem 277:8083–8090

101. Sanderson H, Brain RA, Johnson DJ et al (2004) Toxicity classification and evaluation of four pharmaceuticals classes: antibiotics, antineoplastics, cardiovascular, and sex hormones. Toxicology 203:27–40

102. Henschel K-P, Wenzel A, Diedrich M et al (1997) Environmental hazard assessment of pharmaceuticals. Regul Toxicol Pharmacol 25:220–225

103. Boroskova Z, Dvoroznakova E, Tomasovicova O et al (2001) The effect of cyclophosphamide on the cellular and humoral immune responses on experimental *Toxocara canis* infestation. Helminthologia 38:193–199

104. Grujić S, Vasiljević T, Laušević M (2009) Determination of multiple pharmaceutical classes in surface and ground waters by liquid chromatography-ion trap-tandem mass spectrometry. J Chromatogr A 1216:4989–5000

105. Ellenberger TE, Wright JE, Rosowsky A et al (1989) Wild-type and drug-resistant *Leishmania major* hydrolyze methotrexate to N-10-methyl-4-deoxy-4-aminopteroate without accumulation of methotrexate polyglutamates. J Biol Chem 264:15960–15966

106. Warren KS (1967) Studies on the treatment of molluscan *Schistosomiasis mansoni* with antibiotics, non-antibiotic metabolic inhibitors, molluscicides and antischistosomal agents. Trans R Soc Trop Med Hyg 61:368–372

107. Miguel DC, Yokoyama-Yasunaka JKU, Andreoli WK et al (2007) Tamoxifen is effective against *Leishmania* and induces a rapid alkalinization of parasitophorous vacuoles harbouring *Leishmania* (*Leishmania*) *amazonensis* amastigotes. J Antimicrob Chemother 60:526–534

108. Harwood PD (1953) The anthelmintic properties of phenothiazine. Exp Parasitol 2: 428–455

109. Minagh E, Hernan R, O'Rourke K et al (2009) Aquatic ecotoxicity of the selective serotonin reuptake inhibitor sertraline hydrochloride in a battery of freshwater test species. Ecotoxicol Environ Saf 72:434–440

110. Crane M, Watts C, Boucard T (2006) Chronic aquatic environmental risks from exposure to human pharmaceuticals. Sci Total Environ 367:23–41

111. Baqui A, Ansari JA (1987) Investigations on the response to various stimuli of the microfilariae of *Setaria cervi* in white rats. Helminthologia 24:33–42

112. Jones-Brando L, Torrey EF, Yolkem R (2003) Drugs used in the treatment of schizophrenia and bipolar disorder inhibit the replication of *Toxoplasma gondii*. Schizophr Res 62:237–244

113. Thacker PD (2005) Pharmaceutical data elude researchers. Environ Sci Technol 39:193A–194A

114. Arcay L (1985) Influence of sex hormones on the experimental infection produced by a strain of *Leishmania mexicana amazonensis* from Venezuela. Rev Latinoam Microbiol 27:195–207

115. Wang R, Belosevic M (1994) Estradiol increases susceptibility of goldfish to *Trypanosoma danilewskyi*. Dev Comp Immunol 18:377–387

116. Boxall AB, Fogg LA, Baird DJ et al (2005) Targeted monitoring study for veterinary medicines in the environment. Environment Agency, Bristol, UK

117. Erzen NK, Kolar L, Cerkvenik Flajs V et al (2005) Degradation of abamectin and doramectin on sheep grazed pasture. Ecotoxicology 14:627–635

118. Grønvold J, Svendsen TS, Kraglund HO et al (2004) Effect of the antiparasitic drugs fenbendazole and ivermectin on the soil nematode *Pristionchus maupasi*. Vet Parasitol 24:91–99

119. Svendsen TS, Hansen PE, Sommer C et al (2005) Life history characteristics of *Lumbricus terrestris* and effects of the veterinary antiparasitics compounds Ivermectin and fenbendazole. Soil Biol Biochem 37:927–936

120. Sun Y, Diao X, Zhang Q et al (2005) Bioaccumulation and elimination of avermectin B1a in the earthworm (*Eisenia fetida*). Chemosphere 60:699–704

121. Halley BA, Winter R, Yoon S et al (2005) The environmental safety of eprinomectin to earthworms. Vet Parasitol 128:109–114

122. Singer AC, Johnson AC, Anderson PD (2008) Reassessing the risks of Tamiflu use during a pandemic to the Lower Colorado

River. Environ Health Perspect 116: A285–A286

123. Soderstrom H, Jarhult JD, Fick J et al (2010) Levels of antivirals and antibiotics in the River Thames, UK, during the pandemic 2009. SETAC Europe annual meeting, Seville

124. Flaherty CM, Dodson SI (2005) Effects of pharmaceuticals on *Daphnia* survival, growth, and reproduction. Chemosphere 61:200–207

125. Rappaport SM (2011) Implications of the exposome for exposure science. J Expo Sci Environ Epidemiol 21:5–9

126. NRC (National Research Council) (1983) Risk assessment in the Federal Government: managing the process. National Academies Press, Washington, DC, Accessed 4 June 2015

127. NRC (National Research Council) (2007) Toxicity testing in the 21st century: a vision and a strategy. National Academies Press, Washington, DC, https://download.nap.edu/login.php?record_id=11970&page=%2Fdownload.php%3Frecord_id%3D11970. Accessed 4 June 2015

128. Wignall JA, Shapiro AJ, Wright FA et al (2014) Standardizing benchmark dose calculations to improve science-based decisions in human health assessments. Environ Health Perspect 122:499–505

129. Yoon M, Campbell J, Andersen ME et al (2012) *In vitro* to *in vivo* extrapolation of cell-based toxicity assay results. Crit Rev Toxicol 42:633–652

130. Wetmore BA, Wambaugh JF, Ferguson SS et al (2012) Integration of dosimetry, exposure, and high-throughput screening data in chemical toxicity assessment. Toxicol Sci 125:157–174

131. Zhang Q, Bhattacharya S, Andersen ME et al (2010) Computational systems biology and dose-response modeling in relation to new directions in toxicity testing. J Toxicol Environ Health B Crit Rev 13:253–276

132. US EPA (U.S. Environmental Protection Agency) (2010) Exposure assessment tools and models: what is an exposure assessment? http://www.epa.gov/oppt/exposure/pubs/exposurep.htm. Accessed 4 June 2015

133. Hays SM, Aylward LL (2008) Using biomonitoring equivalents to interpret human biomonitoring data in a public health risk context. J Appl Toxicol 29:275–288

134. Barton HA, Chiu WA, Woodrow SR et al (2007) Characterizing uncertainty and variability in physiologically based pharmacokinetic models: state of the science and needs for research and implementation. Toxicol Sci 99:395–402

135. Jones DP, Park Y, Ziegler TR (2012) Nutritional metabolomics: progress in addressing complexity in diet and health. Annu Rev Nutr 32:183–202

136. van Gestel CA, Jonker M, Kammenga JE (2010) Mixture toxicity: linking approaches from ecological and human toxicology. Taylor & Francis, USA

137. Presidential/Congressional Commission on Risk Assessment Risk Management (1997) Risk assessment and risk management in regulatory decision-making. Final report, vol. 2. PCRARM, Washington, DC. http://www.riskworld.com. Accessed 4 June 2015

138. Kümmerer K (2007) Sustainable from the very beginning: rational design of molecules by life cycle engineering as an important approach for green pharmacy and green chemistry. Green Chem 9:899–907

139. Nowotny N, Epp B, Von Sonntag C (2007) Quantification and modeling of the elimination behavior of ecologically problematic wastewater micropollutants by adsorption on powdered and granulated activated carbon. Environ Sci Technol 41:2050–2055

140. Fjodorova N, Novich M, Vrachko N et al (2008) Directions in QSAR modeling for regulatory uses in OECD member countries, EU and in Russia. J Environ Sci Health C 26:201–236

141. Knacker T, Duis K, Ternes T et al (2005) The EU-project ERAPharm—Incentives for the further development of guidance documents? Environ Sci Pollut Res 12:62–65

142. Cleuvers M (2002) Aquatic ecotoxicology of selected pharmaceutical agents—algal and acute Daphnia tests. UWSE-Z Umweltchem Ökotox 14:85–89

143. VICH (2004) Environmental impact assessment (EIAs) for veterinary medicinal products (VMPs)–Phase II, Draft Guidance, August, 2003, International cooperation on harmonization of the technical requirements for registration of veterinary medicinal products. http://vich.eudra.org/pdf/10_2003/g138_st4.pdf

144. http://www.nicnas.gov.au/. Accessed 4 June 2015

145. Selassie CD (2003) History of quantitative structure-activity relationships. In: Abraham DJ (ed) Burger's medicinal chemistry and drug discovery. Wiley, New York, pp 1–48

146. Roy K, Kar S (2015) How to judge predictive quality of classification and regression based QSAR models? In: Haq ZU, Madura J (eds) Frontiers of computational chemistry. Bentham, pp 71–120. http://ebooks.ben-

thamscience.com/book/9781608059782/chapter/128894/

147. Roy K, Kar S (2015) Importance of applicability domain of QSAR models. In: Roy K (ed) quantitative structure-activity relationships in drug design, predictive toxicology, and risk assessment, IGI Global, pp 180–211. http://dx.doi.org/10.4018/978-1-4666-8136-1.ch005
148. http://ec.europa.eu/research/biosociety/pdf/anim_al_see_final_report.pdf. Accessed 4 June 2015
149. Kar S, Roy K (2012) Risk assessment for ecotoxicity of pharmaceuticals—an emerging issue. Expert Opin Drug Saf 11:235–274
150. Kanter J (2013) E.U. Bans cosmetics with animal-tested ingredients. http://www.nytimes.com, March 11. Accessed 4 June 2015
151. Graef A (2013) Victory! India bans animal testing for cosmetics. http://care2.com, July 2. Accessed 4 June 2015
152. Bloomberg News (2013) China animal testing complicates L'Oreal's expansion. www.Bloomberg.com, August 22. Accessed 4 June 2015
153. Cronin MTD, Livingstone DJ (2004) Predicting chemical toxicity and fate. CRC, Washington, DC
154. Schultz TW, Cronin MTD, Walker JD (2003) Quantitative structure-activity relationships (QSARs) in toxicology: a historical perspective. J Mol Struct (Theochem) 622:1–22
155. Kar S, Roy K (2010) Predictive toxicology using QSAR: a perspective. J Indian Chem Soc 87:1455–1515
156. Kar S, Roy K (2010) First report on interspecies quantitative correlation of ecotoxicity of pharmaceuticals. Chemosphere 81: 738–747
157. Das RN, Sanderson H, Mwambo AE et al (2013) Preliminary studies on model development for rodent toxicity and its interspecies correlation with aquatic toxicities of pharmaceuticals. Bull Environ Contam Toxicol 90: 375–381
158. Sanderson H, Thomsen M (2009) Comparative analysis of pharmaceuticals versus industrial chemicals acute aquatic toxicity classification according to the United Nations classification system for chemicals. Assessment of the (Q)SAR predictability of pharmaceuticals acute aquatic toxicity and their predominant acute toxic mode-of-action. Toxicol Lett 187:84–93
159. Escher BI, Bramaz N, Richter M et al (2006) Comparative ecotoxicological hazard assessment of beta-blockers and their human metabolites using a mode-of-action-based test battery and a QSTR approach. Environ Sci Technol 40:7402–7408
160. Berninger JP, Brooks BW (2010) Leveraging mammalian pharmaceutical toxicology and pharmacology data to predict chronic fish responses to pharmaceuticals. Toxicol Lett 193:69–78
161. Sanderson H, Johnson D, Reitsma T et al (2004) Ranking and prioritization of environmental risks of pharmaceuticals in surface waters. Regul Toxicol Pharmacol 39:158–183
162. Sanderson H, Thomsen M (2007) Ecotoxicological quantitative structure–activity relationships for pharmaceuticals. Bull Environ Contam Toxicol 79:331–335
163. Lienert J, Güdel K, Escher BI (2007) Screening method for ecotoxicological hazard assessment of 42 pharmaceuticals considering human metabolism and excretory routes. Environ Sci Technol 41:4471–4478
164. Christen V, Hickmann S, Rechenberg B et al (2010) Highly active human pharmaceuticals in aquatic systems: a concept for their identification based on their mode of action. Aquat Toxicol 96:167–181
165. http://www.biograf.ch. accessed 4 June 2015
166. Escher BI, Baumgartner R, Koller M et al (2011) Environmental toxicology and risk assessment of pharmaceuticals from hospital wastewater. Water Res 45:75–92
167. de García MSAO, Pinto GP, García-Encina PA et al (2014) Ecotoxicity and environmental risk assessment of pharmaceuticals and personal care products in aquatic environments and wastewater treatment plants. Ecotoxicology 23:1517–1533
168. Sanderson H (2012) Challenges and directions for regulatory use of QSARs for predicting active pharmaceutical ingredients environmental toxicity. Curr Drug Saf 7:309–312
169. Kar S, Das RN, Roy K, Leszczynski J (2015) Can Toxicity for Different Species be Correlated? The Concept and Emerging Applications of Interspecies Quantitative StructureToxicity Relationship (i-QSTR) Modeling. Int J Quant Struct-Prop Relat 1(2): 23–51, doi: 10.4018/IJQSPR.2016070102
170. Kümmerer K (2009) The presence of pharmaceuticals in the environment due to human use-present knowledge and future challenges. J Environ Manage 90:2354–2366

Chapter 13

Use of Read-Across Tools

Serena Manganelli and Emilio Benfenati

Abstract

Read-across has become popular since the introduction of regulations, such as the European REACH regulation. This chapter provides instructions on how to use ToxRead, new freely available software for read-across analysis, and on how to interpret its output predictions for mutagenicity assessments.

This tool offers two seminal sources: a set of rules/structural alerts, which may explain the toxicity, and a similarity tool, associated with a large database of chemicals with their properties.

Key words Read-across, ToxRead, SAR, Structural alerts, Rules, Mutagenicity, REACH

1 Introduction

1.1 Use of Structural Alerts for Mutagenicity Assessment

Mutagenicity is one of the most important endpoints to evaluate toxicity towards humans; indeed, it is part of the CMR (Carcinogenic, Mutagenic, Reprotox) regulatory assessment. As discussed in Chapter 5, the most common assay to assess experimentally mutagenicity is the Ames test [1]. The Ames test makes use of genetically engineered *Salmonella typhimurium* and *E. coli* bacterial strains and it has an estimated inter-laboratory reproducibility of 85–90 % [2]. In Chapter 5, the involvement of mutagenicity assessment in different fields, such as drug discovery, is discussed along with the importance of characterization of this endpoint to fulfill European regulation requirements.

Human experts usually estimate toxicity by means of the identification of structural fragments known to be responsible for the toxic property under investigation. The knowledge of the biochemical mechanism of action of chemicals helps the expert in the determination of these fragments' activity. Once these moieties are found to be the reason for an effect, such chemical moieties can be codified into rules called structural alerts (SA), toxicophores, etc. In the last decades, many lists of fragments have been discovered and codified. Some of these lists of fragments have been extracted

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_13, © Springer Science+Business Media New York 2016

manually, such as rules extracted by Ashby [3, 4], Benigni–Bossa [5], and DEREK [6], while in other cases computer programs have been used, as in the case of Kazius [7], SARpy [8], MultiCASE [9, 10], and Ahlberg [11].

However, when the SA is lacking we cannot conclude that the chemical is not mutagenic: it is possible that the chemical contains a SA that is not yet identified; this may lead to false negatives. Also for this reason, the guidance on the use of models for mutagenicity for impurities within pharmaceuticals asks users to apply two independent approaches: one based on SA and one based on statistical criteria, such as quantitative structure–activity relationships (QSAR) [12].

It is also possible that there are chemicals, which contain the SA, but are non-toxic. Computer systems programs may also extract fragments related to the lack of effects. Differently from the detection of toxic fragments, which can be mechanism-based, the meaning of non-toxic fragments may be purely statistical, indicating that substances with those particular fragments are non-toxic.

One of the most recently developed software providing a useful application of probably the largest collection of SA for the mutagenicity assessment is ToxRead [13]. ToxRead is a program for read-across which offers guidance to the user for identifying similar chemicals that share the same fragments with the target compound.

1.2 Read-Across

Read-across is a method for data-gap filling where information from one or more chemicals is used to predict the same endpoint for a target chemical, which is similar in some key aspects related to that endpoint.

Two main problems can be encountered when filling data gaps with read-across. The first one is the difficulty in assessing the absence of toxicity, which seems to require a greater burden of proof for justification. The second one is how to deal with uncertainty and to what extent results are to be considered reliable. Different elements contribute to reliability: the quality and number of the experimental data used to perform read-across; the chemical similarity measures used; knowledge about how chemicals interact with biological systems; and supplementary data from other properties or in vitro assays. This information is not always available, but each element may contribute in a weight of evidence (WoE) approach [14].

A read-across approach may fulfill REACH information requirements, in order to avoid unnecessary testing, only if it meets the following criteria set out in Annex XI:

1. Results are adequate for the purpose of classification and labeling and/or risk assessment.

2. Results have adequate and reliable coverage of the key parameters addressed in the corresponding test methods.
3. An exposure duration comparable to or longer than the corresponding test method is covered, if this parameter is relevant.
4. Adequate and reliable documentation of the applied method is provided.

Read-across approaches may also be used to define further testing needs in integrated testing strategies to allow efficient targeting of testing. These approaches can also support a conclusion for a REACH endpoint using a WoE method.

The strategies to perform read-across based prediction are essentially four:

1. One-to-one (one analogue used to make an estimation for a single chemical).
2. One-to-many (one analogue used to make estimations for two or more chemicals).
3. Many-to-one (two or more analogues used to make an estimation for a single chemical).
4. Many-to-many (two or more analogues used to make estimations for two or more chemicals).

In general, strategies based on the assessment of a number of analogues may be more efficient and accurate than one-to-one approaches [15].

The crucial step of read-across is the identification of similar compounds. This is performed by means of the following approaches:

- "Analogue approach," which is based on a very limited number of chemicals (e.g., target substance + source substance).
- "Category approach," which is based on a more extensive range of analogues (e.g., three or more members) and there may be an apparent trend in property.

A strategy for grouping the chemicals in terms of similarity can be based on chemical structure, or on other common properties such as common precursor and/or breakdown products, or a constant pattern in the changing potency of the properties across the group (in the case of a quite consistent number of compounds). These criteria can be adopted one by one or can be integrated to strengthen the grouping hypothesis.

A straightforward way to find analogues of the target compound is to check its presence in existing categories or to apply expert knowledge to link this compound to an existing category. Different web-sources contain information on existing categories, such as US EPA (http://cfpub.epa.gov/hpv-s/), OECD (www.oecd.org/env/existingchemicals/data), Canada (http://www.

chemicalsubstanceschimiques.gc.ca/plan/index-eng.php), eChemportal (http://www.echemportal.org), and OECD QSAR Toolbox (www.qsartoolbox.org).

If this condition does not occur, similar compounds search can make use of a similarity assessment approach (pair-wise similarity or similarity to a group). This procedure is helpful even if the chemical is associated with an existing category, since it may lead to the identification of new information and more analogues. In one type of grouping (descriptor-based grouping), the structural similarities of the analogues can be explored by means of statistical approaches such as principal component analysis (PCA) or pattern recognition approaches (e.g., Kohonen neural maps). A wide array of descriptors is generated (constitutional, topological, and geometrical descriptors, molecular connectivity indices, physicochemical properties) for all the analogues; then, a suitable plot (e.g., PCA plot) allows visualizing similarities, trends and possible outliers. A second type of grouping (endpoint-based grouping) makes use of different experimental data and/or QSAR predictions generated for all the analogues and endpoints of interest. This information can predict trends as well as breakpoints in trends, and therefore possible subcategories. There are several available tools to identify analogues, such as ToxRead [13], OECD QSAR Toolbox [16], AMBIT [17], ToxMatch [18], Leadscope [19], AIM [20], and ChemIDplus [21].

The collection of experimental data for relevant analogues in a data matrix is the preliminary step for the subsequent read-across approach. Ecotoxicological information on the analogues can be obtained from the available in-house databases, and from querying external databases.

Finally, endpoint information for the target compound can be obtained using the corresponding information for relevant analogues [22, 23].

The expert attempts to identify the most similar cases with respect to the chemical structure, presence of functional groups, applicability of specific alerts, reasons for considering the parent compounds or its metabolites, and other approaches. This process is time-consuming and not easy to replicate.

To improve this, some automatic systems have been developed to assist the expert in performing read-across based analysis.

The QSAR Toolbox is a standalone software application, developed by the Organization for Economic Co-operation and Development (OECD) with the aim of filling gaps in (eco)toxicity data needed for chemicals' hazard assessment. The Toolbox integrates information and tools from different sources into a workflow [23].

The features of the Toolbox to perform read-across are the following:

1. "Profiling" based on the identification of relevant structural features and potential mechanisms or mode of action of a target substance.

2. "Grouping" based on the identification of other chemicals sharing the structural characteristics and/or mechanism/mode of action recognized for the target.
3. Data gap(s) filling, which makes use of existing experimental data.

Toolbox contains databases with results from experimental studies, plus regulatory inventories, and accumulated knowledge for structural alerts that can indicate the presence of hazardous and other properties. These alerts, named "profilers," encode SAR type information. Some examples are profilers for "DNA Binding," "Protein Binding," "Aquatic toxicity MOAs," etc. Aside from tools for read-across-based-estimations, the Toolbox also contains tools to perform trend analysis, and (Q)SAR models to predict missing experimental values [23].

AMBIT is a software for chemoinformatic data management, resulting from a Long-range Research Initiative of the European Chemical Industry Council (CEFIC LRI), and developed in collaboration with Procter & Gamble. The AMBIT system consists of a database and functional modules allowing a variety of searches and mining of data stored in the database. The AMBIT database stores more than 450,000 chemical structures and their identifiers such as CAS and EINECS numbers and InChI. It also contains attributes such as molecular descriptors, experimental data together with test descriptions, and literature references. The quality assured data is organized in searchable templates, offering features on chemicals information (structure, data, text), including REACH applicable PBT/vPvB and analogues assessment. AMBIT Discovery performs chemical grouping usable for read-across, and evaluates the applicability domain of a QSAR offering a variety of methods, including the use of different approaches for similarity assessments [17].

Under the EC funded projects CALEIDOS [24] and PROSIL [25], ToxRead, a new standalone application for read-across analysis, has been developed. ToxRead contains databases of compounds with their experimental activities, currently for two endpoints: mutagenicity and bioconcentration factor (BCF). From its databases, ToxRead arranges similar molecules sharing structural alerts and rules with the target compound, thus providing the expert an interactive tool for studying the target compound and performing a solid read-across analysis.

2 Materials

2.1 Optional Software for Structure Search and Normalization

A number of existing databases can be helpful to obtain quality assurance chemical structures, expressed as "Simplified Molecular Input Line Entry Specification" (SMILES) starting from a chemical identifier, such as CAS number (*see* **Note 1**). Additionally or

alternatively, any software application can be useful for chemical structure drawing and conversion into SMILES (*see* **Note 2**) and for SMILES normalization (*see* **Note 3**).

2.2 ToxRead: The Software

ToxRead aims to be an easy way to obtain and integrate the available knowledge, a systematic tool to indicate the uncertainty of the result, and a reproducible program to categorize the substances. ToxRead provides evidence on the evaluation of the relevance of the different structural alerts for the specific chemical of interest, indicating at the same time the most similar compounds, which contain these structural alerts [14]. The developed tool is based on an application supported by libraries of fragments, which visualize the substances and the structural alerts.

2.3 Database

Currently, the experimental values for mutagenicity are referred to 6,065 compounds extracted from the ANTARES project [26], which refers to the data from Hansen et al. [27], checked and pruned. The dataset contains (1) chemical structure, (2) CAS number, (3) common name as identifiers of each compound, and (4) experimental mutagenic activity. The structures are represented as SMILES strings, and the corresponding Ames test value (mutagenic or not mutagenic) is derived from several well-known sources such as Chemical Carcinogenesis Research Information (CCRIS) [28], Helma et al. [29], Kazius et al. [7], Feng et al. [30], VITIC [31], and the GeneTox databases [32].

2.4 The Implemented Rules

Currently, the program includes the following libraries of rules on mutagenicity:

1. Benigni–Bossa rules implemented within the Toxtree software [33].
2. SARpy rules [8].
3. 281 alerts manually extracted by human experts at Istituto di Ricerche Farmacologiche Mario Negri (IRFMN).
4. Rules automatically extracted by the Center for Advanced Studies, Research and Development in Sardinia (CRS4) within the LIFE PROSIL project [25].

The first two sets of rules are also present in the VEGA software [34]. The SARpy and Toxtree algorithms generating rules have already been described in Chapter 5. SARpy, CRS4 and IRFMN rules can be associated with both mutagenicity and non-mutagenicity. These are conceptually similar to the exclusion rules present in the Benigni–Bossa rulebase, but the exclusion rules within Toxtree are always associated with a positive toxic rule, while the rules for "non-toxicity" listed by SARpy, CRS4, and IRFMN can be more general and apply to all chemicals.

The rules for mutagenic and non-mutagenic activity are expressed as "SMiles ARbitrary Target Specification" (SMARTS)

strings [35]. This notation is an extension of the widely used SMILES notation (described in the introduction Chapter 1), adding the possibility of describing generic molecular patterns that can match with several compounds. Overall, 759 rules are present within the ToxRead program for mutagenicity.

3 Methods

3.1 ToxRead: The Workflow

ToxRead has been designed to be user-friendly. The simple workflow of ToxRead is described below. The user should insert the target molecule, encoded as a SMILES string in the blank space at the top of the user interface (Fig. 1). The user can choose the maximum number of similar compounds, which is three by default. These chemicals are identified using the algorithm implemented in VEGA and the similarity value is calculated as the weighted combination of a fingerprint, three structural keys based on molecular descriptors, and a series of other descriptors (constitutional, heteroatoms and specific functional groups considering the number of some features or functional groups and not only their presence/absence). The description of the similarity algorithm has been presented by Floris et al. [36].

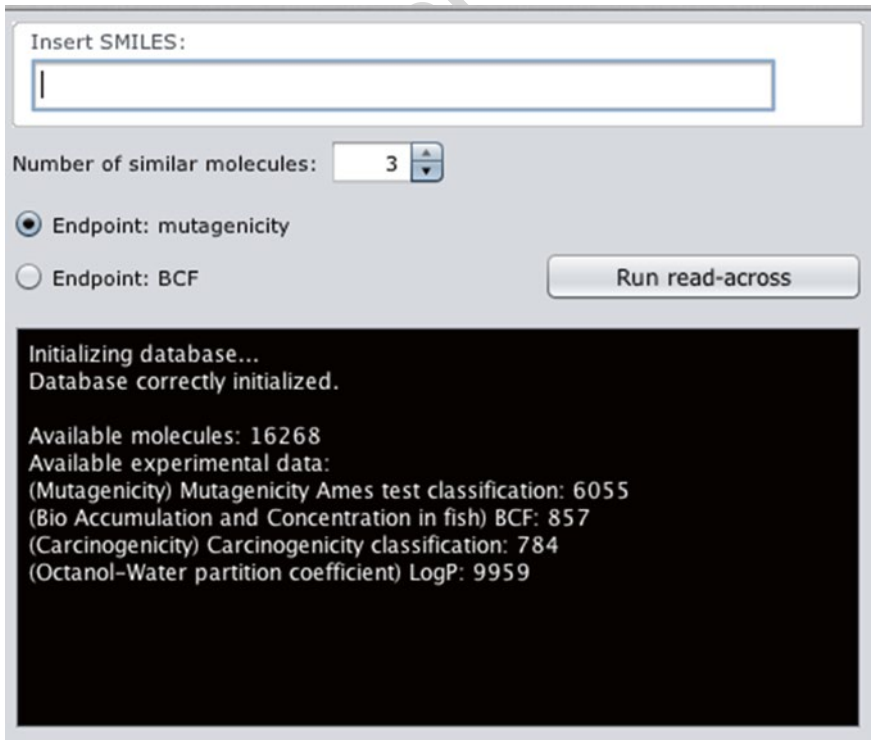


Fig. 1 Graphical user interface (GUI) of the ToxRead software

After the choice of the endpoint of interest (in this case mutagenicity), the user can run the software for the read-across analysis by clicking the "Run read-across" button.

Once the calculation has been performed, ToxRead displays the interactive chart (Fig. 2) showing the structural alerts for the specific chemical of interest, and indicating at the same time the most similar compounds which contain these structural alerts. The chemical in the first example is the same as reported in Chapter 5. The overall evaluation supports the prediction results obtained by VEGA. The second example is more challenging and it will advise the reader about more complex cases. It will take advantage of an integrated approach based on QSAR predictions and read-across. The purpose of this section is to provide an insight into the critical assessment of read-across predictions, to highlight relevant aspects that should be taken into account when analyzing ToxRead outputs, and to make clear how these aspects can be merged with the information from QSAR predictions used by means of a synergistic approach.

The next paragraph will also provide an explanation on how to interpreter SMARTS encoding rules through some practical examples.

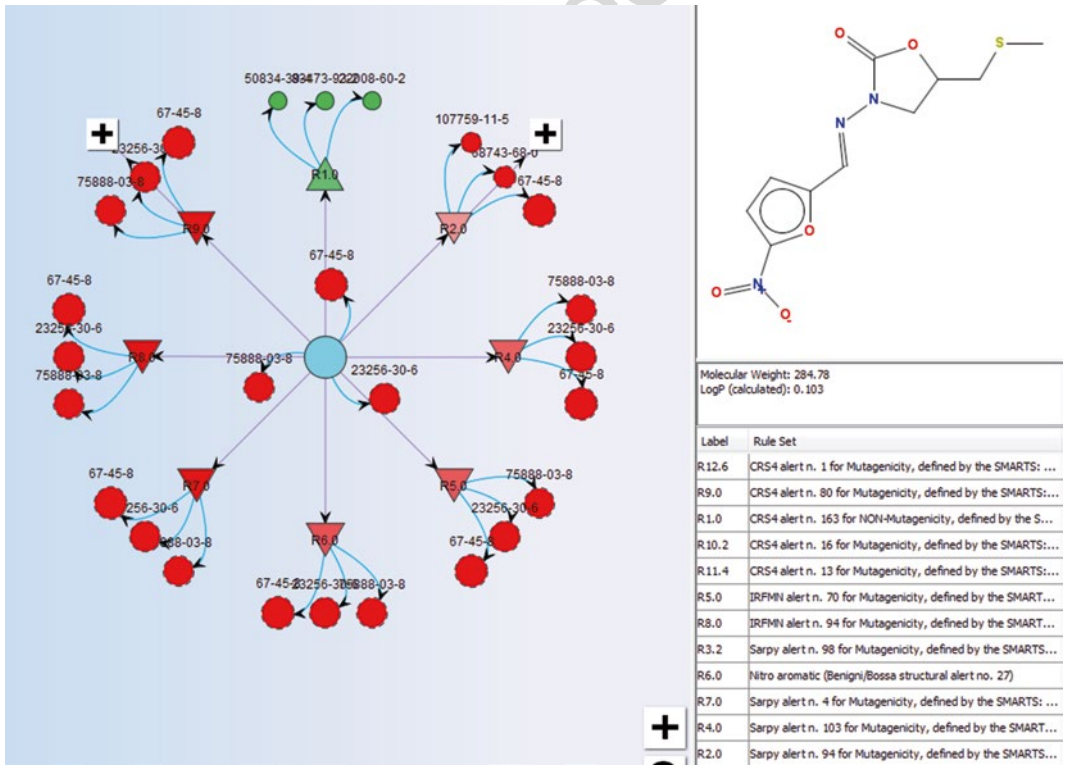


Fig. 2 ToxRead screen showing the similar compounds (represented by *circles*) and the rules (represented by *triangles*) found in the analysis of nifuratel

3.2 Example 1: Nifuratel

Systematic Name: 2-Oxazolidinone, 5-((methylthio)methyl)-3-(((5-nitro-2-furanyl) methylene)amino)-.

CAS Registry Number: 4936-47-4.

SMILES: O=C2OC(CN2(N=Cc1oc(cc1)[N+](=O)[O−])CSC.

Experimental activity: Mutagenic in Ames test [37].

The overall evaluation should keep into account the occurrence of rules/structural alerts in common (or not) between the target compound, and the similar chemicals, and this is provided by ToxRead. The target chemical is drawn at the center of the visualization panel; it is represented by a blue circle (*see* the example given in Fig. 2), with outgoing links to N similar chemicals (in this case three). The size of the circle of any similar compound is proportional to the similarity index in order to make the user aware of the relevance of each chemical. The color of the circle indicates whether the chemical is mutagenic (red) or not (green). This color-coding refers to the experimental value in the internal database. If one chemical is present more than once, the circle line is dashed. Moreover, all the available experimental values, such as BCF, Log *P* values, carcinogenicity, etc., appear allowing the user accomplishing evaluations that are more robust. Clicking on a chemical, the user can see its structure, CAS number, the similarity and experimental values associated with it.

The user should evaluate the similarity of the related chemicals, look at the structures and evaluate the similarity index. More relevance should be given to the most similar compounds, and particular attention is necessary if the similarity is below 0.75. The three most similar chemicals to the target nifuratel are shown in Fig. 3, according to their similarity indices, which have quite high values, respectively, 0.922, 0.899, and 0.882. The target chemical is also linked to several structural alerts (*see* Fig. 4), represented by triangles; those pointing upward are non-mutagenic and those

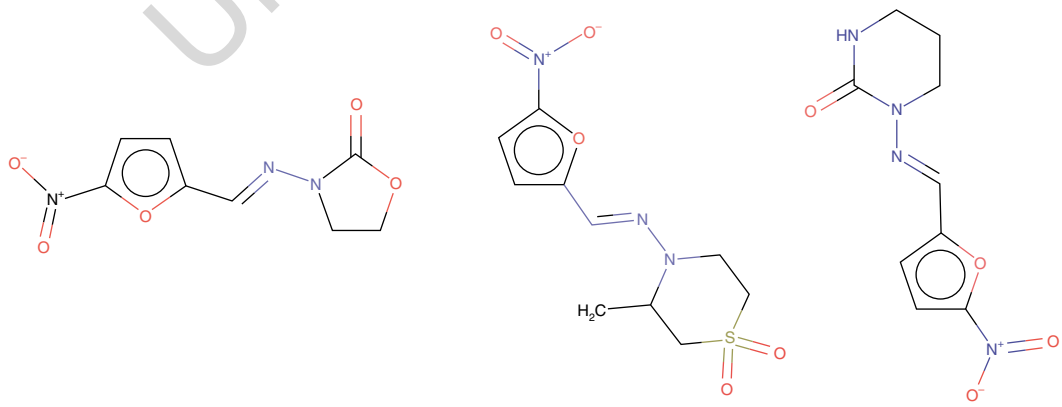


Fig. 3 The three most similar compounds of nifuratel found by ToxRead. From *left* to *right* the similarity index values are 0.922, 0.899, and 0.882

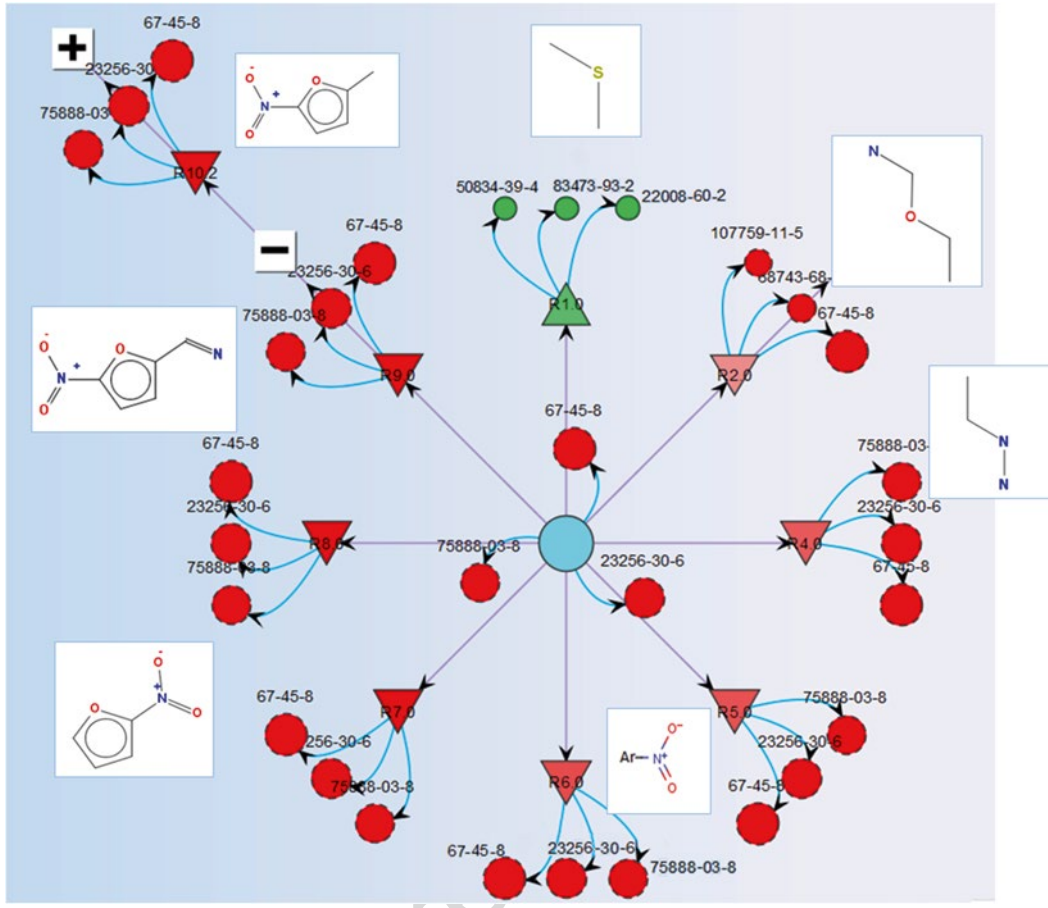


Fig. 4 ToxRead screen showing the chemical fragments encoded by the rules in the analysis of nifuratel

pointing downward are mutagenic. In addition, mutagenic alerts are red while non-mutagenic are green. Another immediate visual detail about the "validity" of a certain alert is that the saturation of the color is proportional to the percentage of toxic or non-toxic chemicals. The size of the triangles is proportional to the number of chemicals containing that SA in the training set. The rules are presented clockwise, starting at the top from the most accurate rule related to non-toxicity, proceeding with less accurate toxicity rules, and finally with more accurate toxicity rules. By clicking on a structural alert, the user can visualize its chemical structure, its explanation, the encoding SMARTS, its accuracy, and the *p*-value relative to the toxicity. By clicking on a specific button of a structural alert, it is also possible to visualize up to 100 similar chemicals presenting that structural alert. In this example, one rule of non-toxicity appears, indicated by a green triangle and seven of toxicity indicated by red triangles. This gives a first indication that there are reasons of possible concern.

The non-toxicity rule is a quite generic alert, generated by CRS4, and is expressed by the SMARTS "CSC," indicating an aliphatic thioether. Proceeding clockwise two SARpy alerts, with 60 % and 80 % mutagenic activity appear. These are defined respectively by the SMARTS "C(OCC)N" indicating an aliphatic amine linked to an alkoxy chain with at least two carbon atoms, and "NNCC" encoding a *N*-alkyl hydrazine group with a chain of at least two aliphatic carbons. The mutagenic activity (%) increases in the subsequent two alerts by IRFMN and Benigni–Bossa both referring to the generic nitroaromatic ring; the Benigni–Bossa alert does not include chemicals with *ortho-* distribution and with a sulphonic group on the nitroaromatic ring. This leads to a slight difference in the accuracies of these fragments, which are respectively 85 % and 87 %. The most accurate fragments with 100 % mutagenic activity are shown on the left of the graph. These fragments have a better coverage of the target compound, so we believe these fragments better explain the behavior of the compound of interest. Two of these are the SARpy and IRFMN alerts both encoding the 2-nitrofuran ring, while the last one indicates 2-nitrofuran with a methanimine group in position 5. This rule is marked with a "+" symbol. Clicking on this symbol, the sequence of hierarchically related rules appears. The rules appear in sequence from the most specific to the other, more generic ones, which may be fired for the target compound. This rule is connected to a series of more generic ones, such as the alert defined by the 5-alkyl-2-nitrofuran moiety. It is common to have conflicting results: similar compounds, which are both toxic and non-toxic, for instance, or the presence of both toxicity and non-toxicity structural alerts. In this case, the alert of non-toxicity is very generic and the similar compounds linked to it have a medium-low similarity index (<0.75). All these chemicals lack the mutagenic alerts reported for the most similar alerts. In particular, they do not have the nitroaromatic ring, and especially the nitrofuran, in their structures, which is responsible for the mutagenic activity of this molecule. This does not mean that the thioether fragment does not affect the mutagenic activity. However, in this case the alerts for toxicity seem to be prevalent and appear to be crucial in the activity exploitation. Thus, the overall conclusion from read-across is towards mutagenicity according to the Ames test. This evaluation supports results from QSAR predictions provided in Chapter 5 for this target compound.

3.3 Example 2: Spironolactone

Systematic Name: 17-Hydroxy-7alpha-mercapto-3-oxo-17alpha-pregn-4-ene-21-carboxylic acid, gamma-lactone acetate.

CAS Registry Number: 52-01-7.

SMILES: O=C1OC3(CC1)(CCC2C5C(CCC23(C))C4(C(=CC(=O)CC4)CC5SC(=O)C)(C)).

Experimental activity: unknown.

The target compound is shown in Fig. 5. For this compound QSAR predictions from the VEGA software will require the support of ToxRead for a better understanding of structural alerts provided by SARpy and Toxtree models and to get more information from the set of rules implemented in ToxRead which are not present in VEGA.

3.3.1 VEGA Results

The output of QSAR predictions from VEGA is equivocal because the models predictions are in disagreement and show very low values of ADI.

- **CAESAR results**: *Prediction is non-mutagenic but the result may not be reliable.*

Although similarity, concordance, and accuracy indices are high (respectively, 0.893, 1, and 1), ADI is equal to 0.567, and therefore, spironolactone could be out of the Applicability Domain of the model. This lack of reliability is caused by a low (0.6) value of the ACF index. The presence in the molecule of the thioacetyl group (*see* Fig. 6)—a fragment never found in the model's training set—is mainly responsible for this low ACF index.

- **SARpy results**: *Prediction is non-mutagenic but the result may not be reliable.*

The model identifies 13 inactive fragments. Some of these fragments are the same as those identified for Dexamethasone in Chapter 5, Fig. 10.

The values of similarity, concordance, accuracy, and ACF indices are the same as those observed when using CAESAR, producing the same ADI.

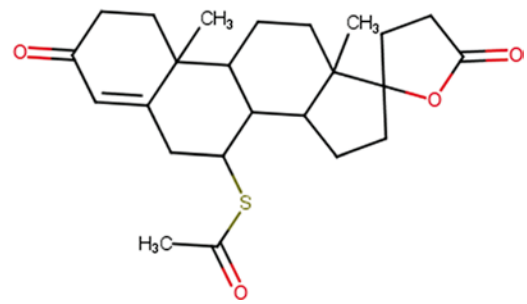


Fig. 5 Chemical structure of the target compound spironolactone

Fig. 6 The thioacetyl fragment, identified by CAESAR in spironolactone, which is not present in the training set molecules

- **TT-VEGA results**: *Prediction is mutagenic but the result may not be reliable.*

The model identifies the presence of the Benigni–Bossa structural alert SA10, indicating the generic α, β unsaturated carbonyl (*see* Chapter 5) as cause of mutagenicity of the target compound.

The predictions yielded by CAESAR and TT-VEGA are in disagreement since CAESAR does not contain the SA10 fragment in its subset of rules.

The unreliability of the TT-VEGA prediction is highlighted by the poor value of its ADI (0) that is determined by low values of the concordance, accuracy, and ACF indices (0, 0.496, and 0.6, respectively).

Indeed, even if the prediction yielded by TT-VEGA is characterized by a similarity index which is greater (0.8) than the corresponding index of CAESAR and SARpy, the experimental and the predicted values are in disagreement for all the similar compounds in the output.

3.3.2 ToxRead Results

This example can benefits from the support of ToxRead, which provides an insight into the analysis of the structural alerts provided by SARpy and Toxtree.

Figure 7 shows the graph of the second example.

The three most similar compounds are non-toxic and their similarity indices are greater than 0.75 (*see* Fig. 8). The target chemical is also linked to several structural alerts. The first rule is expressed by a non-toxicity alert generated by IRFMN, not provided by VEGA, indicating "1,2,6,7,8,9,10,11,12,13,14,15,16,17 -dodecahydrocyclopenta[a]phenanthren-3-one", which is present in the non-mutagenic similar compounds as well. This rule has a non-mutagenic activity of 100 % and a *p*-value of 0.00183. Proceeding clockwise, nine quite generic SARpy alerts of non-toxicity appear (three of these are shown in Fig. 7) which are hierarchically related to other rules for non-toxicity (*see* Fig. 9). The nine fragments, identified by VEGA as well, have non-mutagenic activity ranging from 69 % to 97 % and *p*-values $<10^{-6}$.

One of these rules, indicating the alkylthio- group, is linked to a mutagenic compound containing a bromine in position 3 of the cholestane and an ethylenedisulfonyl group (*see* Fig. 10). These two chemical moieties are not present in the target spironolactone. Two more alerts by CRS4 with low accuracy values come next the SARpy fragments. The last alert is the "α, β unsaturated carbonyl" Benigni–Bossa rule for mutagenicity (already provided by VEGA) with a very low prevalence of mutagenic activity of 49 % and *p*-value of 0.015. Thus, the overall evaluation highlights more reasons for non-mutagenicity than for mutagenicity through an integrated approach based on QSAR predictions, the fired structural alerts, similarity assessment, and chemical reasoning.

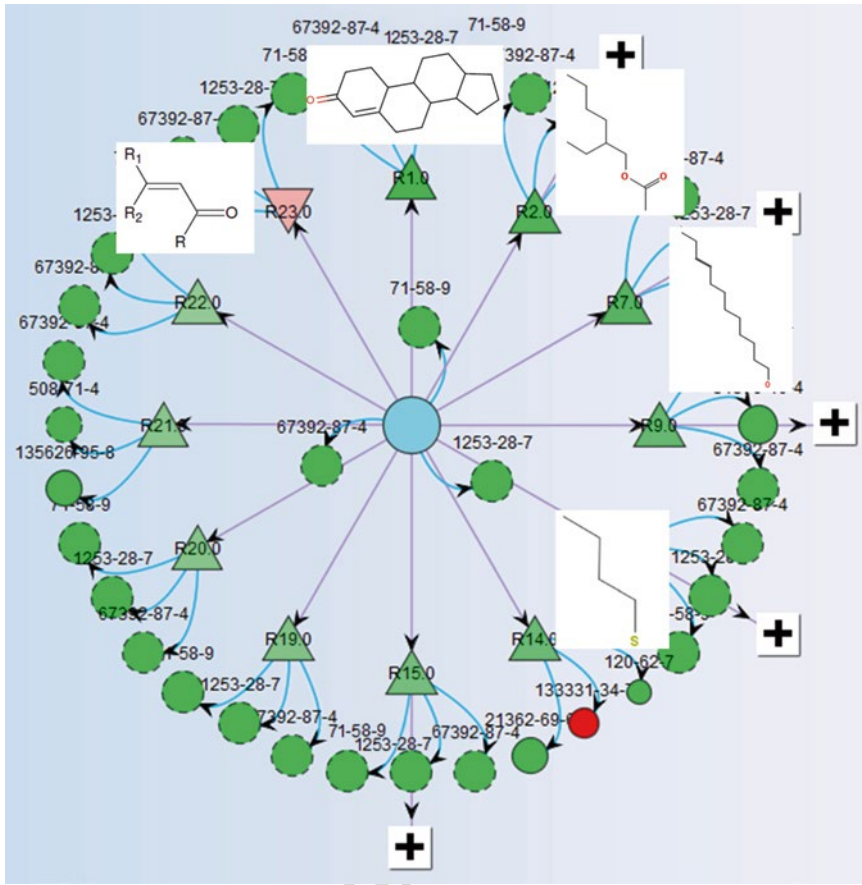


Fig. 7 ToxRead screen showing some chemical fragments encoded by the rules in the analysis of the spironolactone

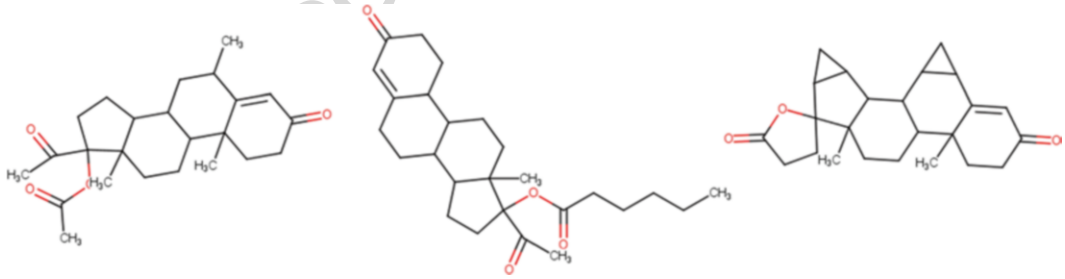


Fig. 8 The three non-mutagenic most similar compounds to spironolactone found by ToxRead. From *left* to *right* the similarity index values are 0.9, 0.897, and 0.882

In this example, ToxRead has enabled to perform a more detailed analysis of rules influencing VEGA predictions, and to identify a new key rule for the non-mutagenicity assessment. This highlights the need to avoid the use of QSAR and read-across as mutually exclusive methods and to combine them to obtain greater evidence for toxicity/non-toxicity.

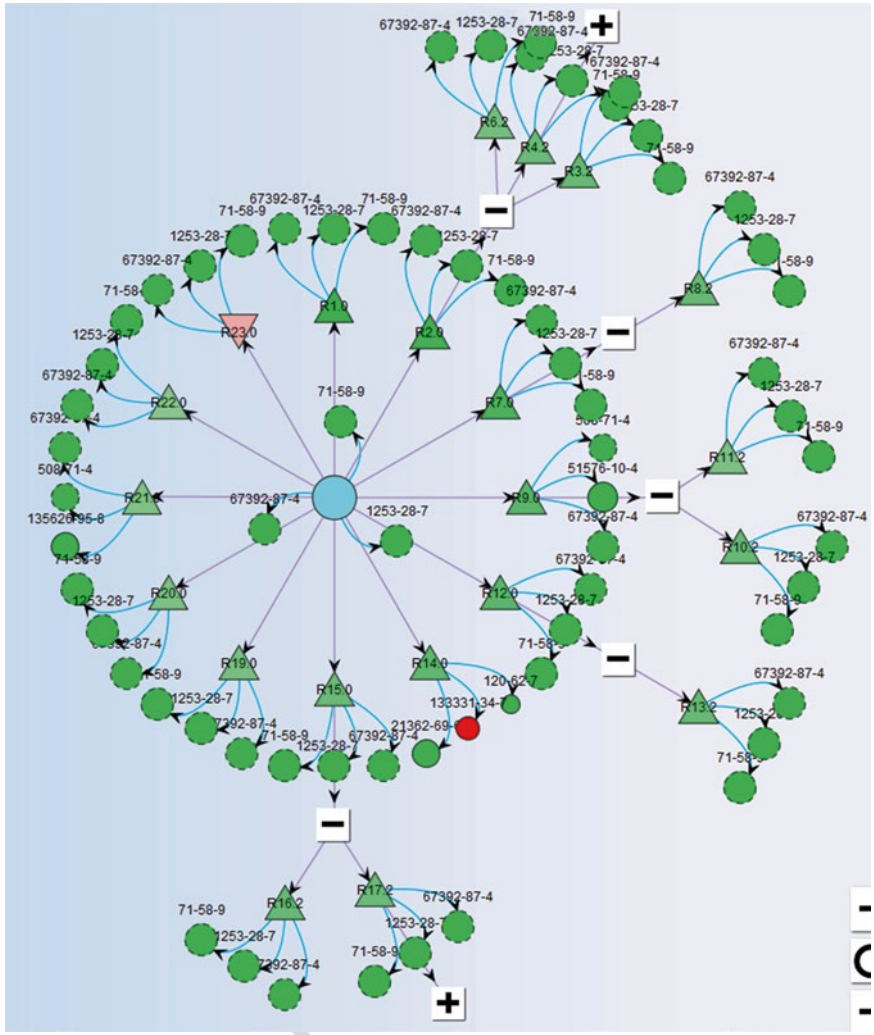


Fig. 9 ToxRead screen showing the similar compounds (*circles*) and the hierarchical rules (*triangles*) found in the analysis of spironolactone

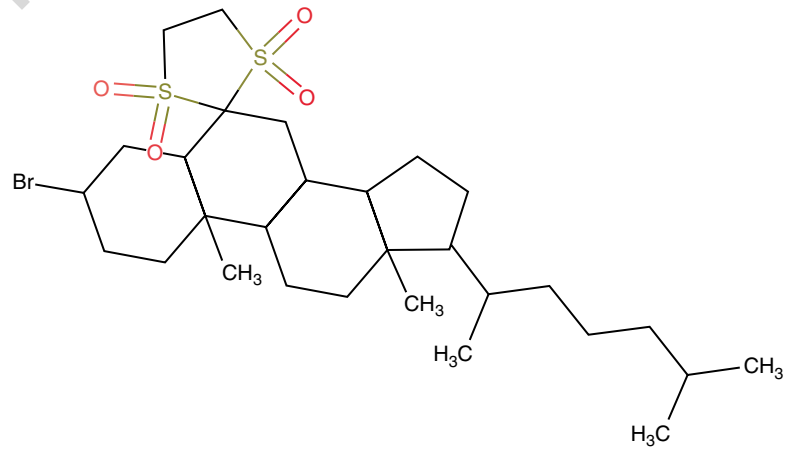


Fig. 10 The only mutagenic similar compound to spironolactone found by ToxRead

4 Conclusions

Read-across requires experts in toxicology, chemistry, biology, environmental sciences, and other fields. Experts may use different sets of rules and they may over-rely on past experience and miss new evidence. That is why expert reasoning may be irreproducible. In this chapter, we proposed two examples of read-across analysis, using ToxRead, which organizes the different elements for reasoning in a reproducible hierarchical structure. The most representative rules, sharing a larger sub-structure with the target compound, are indicated first, but the user can visualize the complete family of more general rules. Some rules within the same family may have an opposite label, because they are exceptions to toxic rules. These rules are related to a toxic effect, or lack of effect, and some act on the effect. This also introduces a more complex approach, than with existing software. In this way, the software assists the user in the read-across evaluation, pointing out the reasons for toxicity, lack of toxicity, and effects on toxicity. This means that ToxRead aims to improve the current issues related to the irreproducibility of read-across.

5 Notes

1. ChemIDplus [21], ChemSpider [38], OECD QSAR Toolbox [16] for searching structures expressed as "Simplified Molecular Input Line Entry specification" (SMILES).
2. Additionally or alternatively, any software application for chemical structures drawing and conversion into SMILES. Several programs can perform this task: VEGA [34], ACD/ChemSketch [39], MarvinSketch [40], OECD QSAR Toolbox [16].
3. istMolBase for SMILES normalization [41].

The list of databases/software is not exhaustive. Moreover, these applications are subject to different software licenses and terms, and conditions of use.

Acknowledgements

We acknowledge the LIFE + CALEIDOS and PROSIL projects.

References

1. Ames BN (1984) The detection of environmental mutagens and potential. Cancer 53:2030–2040
2. Piegorsch WW, Zeiger E (1991) Measuring intra-assay agreement for the Ames Salmonella assay. In: Hotorn L (ed) Statistical methods in toxicology, Lecture notes in medical informatics, vol 43. Springer-Verlag, Heidelberg, pp 35–41
3. Ashby J (1985) Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. Environ Mutagen 7:919–921

4. Ashby J, Tennant RW (1988) Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the US NCI/NTP. Mutat Res 204:17–115
5. Benigni R, Bossa C (2006) Structural alerts of mutagens and carcinogens. Curr Comput Aided Drug Des 2:169–176
6. Marchant CA, Briggs KA, Long A (2008) In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vitic. Toxicol Mech Methods 18:177–187
7. Kazius J, McGuire R, Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem 48:312–320
8. Ferrari T, Cattaneo D, Gini G et al (2013) Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction. SAR QSAR Environ Res 24:365–383
9. Klopman G (1992) MULTICASE 1. A hierarchical computer automated structure evaluation program. Quant Struct Act Rel 11:76–184
10. Klopman G, Rosenkranz HS (1994) Prediction of carcinogenicity/mutagenicity using MULTICASE. Mutation Res 305:33–46
11. Ahlberg E, Carlsson L, Boyer S (2014) Computational derivation of structural alerts from large toxicology data sets. J Chem Inf Model 54:2945–2952
12. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) (2008) Guidance for industry, genotoxic and carcinogenic impurities in drug substances and products: recommended approaches. http://www.fda.gov/ohrms/dockets/98fr/fda-2008-d-0629-gdl.pdf. Accessed 7 July 2015
13. ToxRead. http://www.toxgate.eu. Accessed 7 July 2015
14. Gini G, Franchi AM, Manganaro A et al (2014) ToxRead: a tool to assist in read-across and its use to assess mutagenicity of chemicals. SAR QSAR Environ Res 25:999–1011
15. European Chemicals Agency (2012) Practical Guide 6: how to report read-across and categories. http://echa.europa.eu/documents/10162/13655/pg_report_readacross_en.pdf/. Accessed 7 July 2015
16. OECD (2014) The OECD QSAR toolbox, version 3.3.2. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm. Accessed 7 July 2015
17. AMBIT, Version 2.4.11. http://ambit.sourceforge.net/. Accessed 7 July 2015
18. Toxmatch—OpenTox. http://www.opentox.org/dev/documentation/components/toxmatch. Accessed 7 July 2015
19. Leadscope. http://www.leadscope.com/. Accessed 23 July 2015
20. Analog Identification Methodology (AIM). http://www.epa.gov/opptintr/sf/tools/aim.htm
21. ChemIDplus. http://chem.sis.nlm.nih.gov/chemidplus/. Accessed 7 July 2015
22. European Chemicals Agency (2008) Guidance on information requirements and chemical safety assessment, Chapter R.6: QSARs and grouping of chemicals. http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf/. Accessed 23 July 2015
23. OECD (2014) Guidance on grouping of chemicals, 2nd ed. Environment Directorate Joint Meeting of the chemicals committee and the working party on chemicals, pesticides and biotechnology, Organization for Economic Co-operation and Development, ENV/JM/MONO(2014)4. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono%282014%294&doclanguage=en. Accessed 7 July 2015
24. The CALEIDOS project: chemical Assessment according to Legislation Enhancing the In silico Documentation and Safe use. Funded by the LIFE+ EC Programme 2013–2015. http://www.caleidos-life.eu/. Accessed 23 July 2015
25. The PROSIL project: promoting the use of in silico methods in industry. Funded by the LIFE+ EC Programme 2013–2016. http://www.life-prosil.eu/. Accessed 23 July 2015
26. The ANTARES project: alternative non-testing methods assessed for REACH substances. Funded by the LIFE+ EC Programme 2010–2012. http://www.antares-life.eu/. Accessed 7 July 2015
27. Hansen K, Mika S, Schroeter T et al (2009) Benchmark data set for in silico prediction of Ames mutagenicity. J Chem Inf Model 49:2077–2081
28. Chemical Carcinogenesis Research Information System (2009) NCRI informatics Initiative Homepage. http://www.cancerinformatics.org.uk/matrix/CCRIS.htm. Accessed 7 July 2015
29. Helma C, Cramer T, Kramer S, Raedt LD (2004) Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure-activity

relationships of noncongeneric compounds. J Chem Inf Comput Sci 44:1402–1411

30. Feng J, Lurati L, Ouyang H et al (2003) Predictive toxicology: benchmarking molecular descriptors and statistical methods. J Chem Inf Comput Sci 43:1463–1470
31. Judson PN, Cooke PA, Doerrer NG et al (2005) Towards the creation of an international toxicology information centre. Toxicology 213:117–128
32. Matthews EJ, Kruhlak NL, Cimino MC et al (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regul Toxicol Pharmacol 44:83–96
33. Benigni R, Bossa C, Jeliazkova N et al (2008) Benigni/Bossa rulebase for mutagenicity and carcinogenicity—a module of Toxtree. JRC Scientific and Technical Reports. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_23241_EN.pdf. Accessed 19 Jun 2015
34. Virtual models for evaluating the properties of chemicals within a global architecture. http://www.vega-qsar.eu/download.html. Accessed 7 July 2015
35. SMARTS—a language for describing molecular patterns. http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html. Accessed 23 July 2015
36. Floris M, Manganaro A, Nicolotti O et al (2014) A generalizable definition of chemical similarity for read-across. J Cheminform 6:39
37. Byeon WH, Hyun HH, Lee SY (1976) Mutagenicity of nitro furan nitroimidazol and nitrothiazole derivatives on salmonella microsome system. Kor J Microbiol 14:151–158
38. ChemSpider. http://www.chemspider.com/. Accessed 7 July 2015
39. ACD/ChemSketch for academic and personal use. http://www.acdlabs.com/resources/freeware/chemsketch/. Accessed 19 Jun 2015
40. Marvin, intuitive applications and API for chemical sketching, visualization and data exploration. http://www.chemaxon.com/products/marvin/. Accessed 19 Jun 2015
41. Kode from data to knowledge. http://www.kode-solutions.net. Accessed 7 July 2015

Part III

The Scientific and Society Challenges

Chapter 14

Adverse Outcome Pathways as Tools to Assess Drug-Induced Toxicity

[AU1] **Mathieu Vinken**

Abstract

Adverse outcome pathways (AOPs) are novel tools in toxicology and human risk assessment with broad potential. AOPs are designed to provide a clear-cut mechanistic representation of toxicological effects that span over different layers of biological organization. AOPs share a common structure consisting of a molecular initiating event, a series of key events connected by key event relationships, and an adverse outcome. Development and evaluation of AOPs ideally complies with OECD guidelines. AOP frameworks have yet been proposed for major types of drug-induced injury, especially in the liver, including steatosis, fibrosis, and cholestasis. These newly postulated AOPs can serve a number of purposes pertinent to safety assessment of drugs, in particular the establishment of quantitative structure-activity relationships, the development of novel in vitro toxicity screening tests, and the elaboration of prioritization strategies.

Key words AOP, Drug safety, Steatosis, Fibrosis, Cholestasis

1 Introduction

Predictive toxicology, based upon mechanistic information, has become a critical aspect of human risk assessment in the last decade. A major step in this direction came with the introduction of the mode-of-action concept, which relates to a series of key events (KEs) along a biological pathway from the initial chemical interaction to the adverse outcome (AO) [1]. The mode-of-action concept was originally used by the US Environmental Protection Agency (EPA) in the cancer field [2] but seemed equally exploitable for non-cancer points [3–6]. Another milestone was the well-known report published by the US National Academy of Science in 2007, outlining a vision on toxicology in the twenty-first century and placing toxicity pathways on the foreground [7]. These toxicity pathways denote cellular pathways that, when disturbed, can lead to adverse health effects [1]. Toxicity pathways align with adverse outcome pathways (AOPs), which have their roots in the area of ecotoxicology. An AOP refers to a conceptual construct

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_14, © Springer Science+Business Media New York 2016

that portrays existing knowledge concerning the linkage between a direct molecular initiating event (MIE) and an AO at a biological level of organization relevant to risk assessment (Fig. 1) [1, 8]. In comparison with the mode of action, the scope of an AOP is broader, as it starts with the exposure and can go up to the population level. Thus far, AOPs have been designed for a number of different human-relevant toxicological endpoints. In response to the increasing use of AOPs, the Organization for Economic Cooperation and Development (OECD) together with the US EPA, the US Army Engineer Research and Development Center, and the European Joint Research Center has initiated a project to facilitate the use of AOPs in assessing the safety of chemicals, called the AOP Knowledge Base (AOP-KB). The AOP-KB consists of four modules, namely, the AOP Xplorer, Effectopedia, the Intermediate Effects Database, and the AOP Wiki. The AOP Xplorer is a computational tool that enables automated graphical representation of AOPs and networks among them. Effectopedia is a modeling platform designed for collaborative development and utilization of AOPs. The Intermediate Effects Database hosts chemical-related data derived from non-apical endpoint methods and informs how individual compounds trigger MIEs and KEs. The AOP Wiki is a module of the AOP-KB that provides an open-source interface for rapid, widely accessible, and collaborative

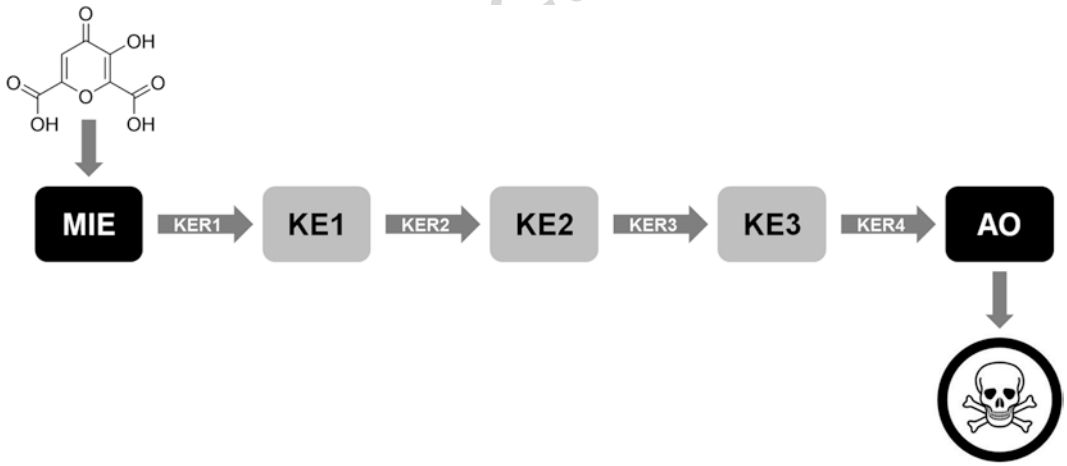


Fig. 1 *Generic structure of an AOP.* Each AOP consists of two anchors, namely, the molecular initiating event (MIE), which refers to the interaction of a chemical with a biological system at the molecular level, and the adverse outcome (AO), which is the actual apical toxicological endpoint. The entire response matrix between the MIE and AO is filled with key events (KEs), which represent changes in the biological state that are both measurable and essential to the progression of a defined biological perturbation leading to a specific AO. Subsequent KEs are connected by key event relationships (KERs), defining a link between both KEs and that facilitate inference or extrapolation of the state of the downstream KE from the known, measured, or predicted state of the upstream KE (adapted from [10, 11])

sharing of established AOPs and building new AOPs [9]. The AOP Wiki was launched in late 2014 and yet contains about fifty AOPs for several human-relevant toxicological endpoints, including drug-induced hepatotoxicity. These AOPs on liver toxicity will be scrutinized in this chapter while discussing AOP development, assessment, and applications in drug safety evaluation.

2 AOP Development and Assessment

2.1 Identification of the MIE, KEs, and AO

The MIE is considered as the first anchor of an AOP and refers to the interaction of a chemical with a biological system at the molecular level, such as ligand-receptor interactions or binding to proteins and nucleic acids. It hereby is of utmost importance to define the site of action of the MIE, as this directly dictates the nature of the AO. The latter is envisaged as the second AOP anchor and describes the actual apical toxicological endpoint. The AO may be located at different levels of biological organization, ranging from the cellular to the population level, and can relate to either a chronic or a systemic toxicological outcome or an acute or local adverse effect. A KE is defined as a change in biological state that is both measurable and essential to the progression of a defined biological perturbation leading to a specific AO. KEs do not provide a comprehensive molecular description of every aspect of the biological process involved per se. Rather, a limited number of KEs should be selected. These are normally those for which there is the most information to support assessment of weight of evidence in a regulatory context. The identification of the MIE, KEs, and AO may be the result of an in-depth survey of relevant scientific literature or may be retrieved from experimental studies. Basically, any type of information can be fed into an AOP, including structural data, "omics-based" data, *in chemico* data, in vitro data, and in vivo data [1, 9–12].

2.2 Description of the KERs

A KER is a scientifically based relationship that connects two KEs, defining a link between both KEs, and that facilitates inference or extrapolation of the state of the downstream KE from the known, measured, or predicted state of the upstream KE. Description of the KERs is a critical step in AOP development, which sets the stage for assessment of the AOP. KERs may either refer specifically to a direct linkage between a pair of KEs that are adjacent in an AOP or may indicate indirect linkages between a pair of KEs for which the relationship is thought to run through another KE or a gap in current understanding. At present, the vast majority of KERs in the AOP Wiki are rather of qualitative nature. However, from the risk assessment point of view, establishing quantitative KERs might be more desirable. These quantitative KERs may be defined in terms of correlations, dose-response relationships,

dose-dependent transitions, or points of departure. They may take the form of simple mathematical equations or sophisticated biologically based computational models that consider other modulating factors, such as compensatory responses or interactions with other biological variables [9–11].

2.3 AOP Assessment

Assessment of AOPs and evaluation of their suitability for application for regulatory purposes relies on (1) the confidence and precision with which the KEs can be measured; (2) the level of confidence in KERs based on biological plausibility, empirical support for the KER, and consistency of supporting data and among different biological contexts; and (3) weight of evidence for the hypothesized pathway. Therefore, overall assessment of AOPs is best supported by providing thorough descriptions of the KEs and KERs as well as robust consideration of weight of evidence for the essentiality of KEs and KERs [9–11]. Basically, AOP assessment relies on two sets of questions, which should be answered in an in-depth and scientifically sound way by AOP developers. The first set of questions focuses on weight-of-evidence assessment based on the Bradford-Hill criteria (Table 1), defining the minimal requirements for establishing a causal link between the different information blocks of the AOP [1, 13]. The second set of key questions has been proposed by the OECD and rather envisages a confidence assessment (Table 2) [1, 12].

Table 1
Bradford-Hill criteria for AOP weight-of-evidence assessment [1, 9, 13]

– Concordance of dose-response relationships
– Temporal concordance among the KEs and AO
– Strength, consistency, and specificity of association of the AO and the MIE
– Biological plausibility, coherence, and consistency of the experimental evidence
– Alternative mechanisms that logically present themselves and the extent to which they may distract from the postulated AOP
– Uncertainties, inconsistencies, and data gaps

Table 2
Key questions for testing AOP confidence [1, 9]

– How well characterized is the AOP?
– How well are the MIE and KEs causally linked to the AO?
– What are the limitations in the evidence in support of the AOP?
– Is the AOP specific to certain tissues, life stages, or age classes?
– Are the MIE and KEs expected to be conserved across *taxa*?

3 Liver Toxicity AOPs

3.1 Liver Steatosis

Steatosis is a prototypical type of drug-induced liver injury that refers to the process of abnormal retention of lipids, mainly triglycerides, within hepatocytes. It reflects the impairment of normal synthesis and elimination of triglycerides and is triggered by a plethora of drugs, such as valproic acid [14]. Steatosis can develop further into nonalcoholic steatohepatitis, which is characterized by hepatocellular injury and inflammation [15, 16]. Liver steatosis may occur in a microvesicular or in a macrovesicular pattern. In microvesicular steatosis, numerous small lipid droplets are present in the hepatocyte cytoplasm, which do not displace the cell nucleus. By contrast, large droplets that move the hepatocyte nucleus to the periphery are observed in macrovesicular steatosis [14, 17–19]. Since interaction of drugs with nuclear receptors is a frequent mechanism observed in liver steatosis, it has been considered as the main MIE in an established liver steatosis AOP (Fig. 2). In particular, activation of the liver X receptor induces an array of effects, such as enhanced transcription of genes encoding mediators of cholesterol and lipid metabolism. This leads to the increased influx

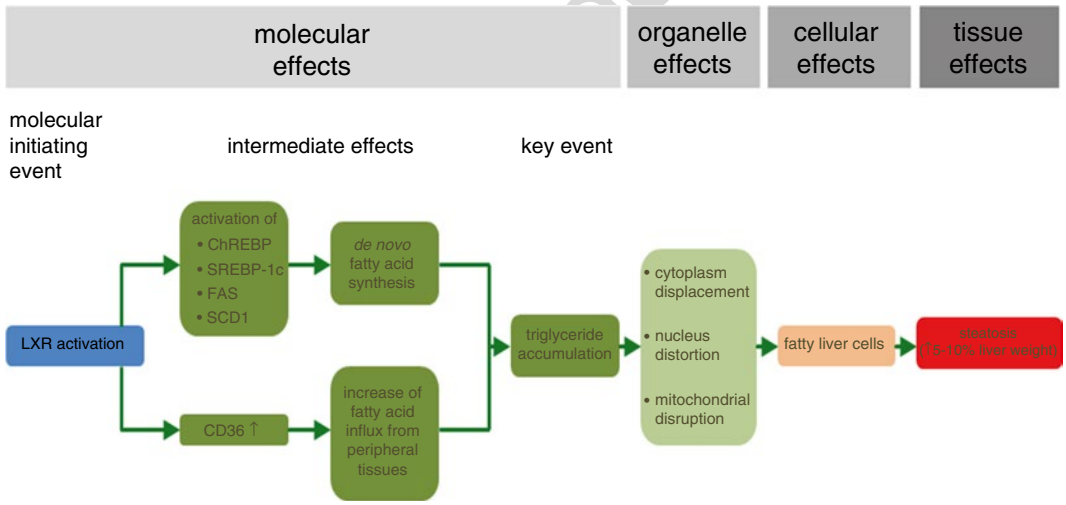


Fig. 2 *AOP for drug-induced liver steatosis*. Activation of the liver X receptor (LXR), which is the MIE (*blue*), induces a number of transcriptional changes, including activation of the expression of carbohydrate response element binding protein (ChREBP), sterol response element binding protein 1c (SREBP-1c), fatty acid synthase (FAS), and stearoyl-coenzyme A desaturase 1 (SCD1). As a result, de novo synthesis of fatty acids is enhanced in the liver. At the same time, fatty acid translocase (CD36) production is upregulated, which mediates increased hepatic influx of fatty acids from peripheral tissues. All together, these intermediate steps drive accumulation of triglycerides, which is considered a key event (*dark green*). At the organelle level, this evokes cytoplasm displacement, distortion of the nucleus, and mitochondrial disruption. This ultimately burgeons into the appearance of fatty liver cells (*orange*) and further into the clinical diagnosis of liver steatosis (*red*) (adapted from [20])

of fatty acids from peripheral tissues into the liver and equally drives de novo synthesis of fatty acids. Consequently, triglycerides tend to accumulate in hepatocytes, which is considered as a KE in this AOP. At the organelle level, hepatocellular lipid accumulation may provoke cytoplasm displacement, nucleus distortion, mitochondrial toxicity, and endoplasmic reticulum stress. All together, these effects underlie the acquisition of the typical fatty liver cell phenotype, which in turn causes a clinically relevant increase in liver weight [20]. This AOP has been generated according to OECD guidelines, including critical consideration of the Bradford-Hill criteria for weight-of-evidence assessment and the OECD key questions for evaluating AOP confidence [9, 20].

3.2 Liver Fibrosis

Liver fibrosis is a reversible wound-healing response to either acute or chronic cellular injury that reflects a balance between liver repair and scar formation. It can be activated by a number of drugs, such as methotrexate. A central event in liver fibrosis is the activation of hepatic stellate cells, which occurs in two phases, namely, the initiation stage and the perpetuation stage [14, 21–23]. In the initiation phase, quiescent hepatic stellate cells become responsive to growth factors. This may be triggered by a variety of signals, including reactive oxygen species and apoptotic bodies originating from dying hepatocytes. In the perpetuation phase, the primed hepatic stellate cells undergo several changes related to proliferation, contractility, fibrogenesis, chemotaxis, extracellular matrix degradation, and retinoid loss, whereby they adopt a myofibroblast-like phenotype. Hepatic stellate cell activation may be counteracted in a resolution phase through apoptosis, senescence, or reversion to the quiescent phenotype [21, 22]. Protein alkylation is considered as the MIE in an established AOP on liver fibrosis (Fig. 3), whereas the obvious AO at the organ level is liver fibrosis. Different steps at the cellular and tissue level have been defined, including hepatocyte injury and cell death, activation of Kupffer cells, expression of transforming growth factor beta 1, activation of hepatic stellate cells, oxidative stress and chronic inflammation, collagen accumulation, and changes in hepatic extracellular matrix composition. The postulated AOP has been assessed by evaluation of the strength of evidence that supports the MIE, the KEs, and the AO [9, 20].

3.3 Cholestasis

Cholestasis is another manifestation of drug-induced liver injury for which an AOP has been introduced (Fig. 4). Cholestasis can be caused by drugs such as bosentan. The MIE in this AOP is the direct *cis*-inhibition of the bile salt export pump. As a result of this, toxic bile acids accumulate into hepatocytes or bile canaliculi. These bile salts trigger a direct deteriorative response and an adaptive response [14]. At the cellular level, the deteriorative response is accompanied by the formation of the mitochondrial permeability pore, which leads to mitochondrial impairment, inflammation, the

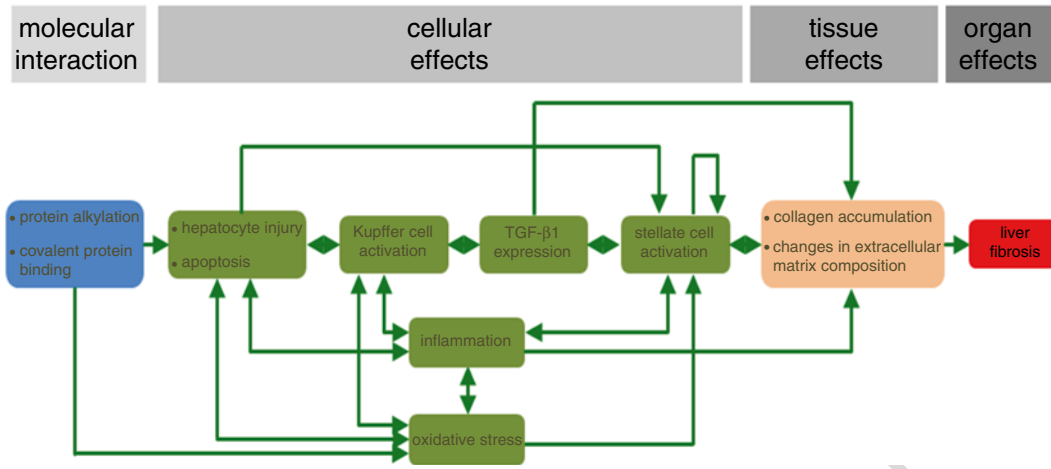


Fig. 3 *AOP for drug-induced liver fibrosis*. The MIE (*blue*) is considered protein alkylation and covalent protein binding in the liver. This serves as a trigger to provoke hepatocyte injury, including apoptosis, which in turn activates Kupffer cells. As a result, transforming growth factor beta 1 (TGF-β1) expression is induced, which is a key factor for stellate cell activation. The latter goes hand in hand with the occurrence of inflammation and oxidative stress. The different events at the cellular level (*green*) are interconnected in several ways. The overall end result is accumulation of collagen and changes in the extracellular matrix composition in the liver (*orange*), which becomes clinically manifested as the AO, namely, liver fibrosis (*red*) (adapted from [20])

production of reactive oxygen species and ultimately to the onset of cell death by both apoptotic and necrotic mechanisms [24, 25]. Because of the latter, cytosolic enzymes start to leak from hepatocytes and cholangiocytes and become measurable in the serum [26, 27]. A hallmark of cholestasis at the cellular level includes the induction of an adaptive response, which is aimed at counteracting bile accumulation and thus cholestatic liver injury. Accordingly, a complex machinery of transcriptionally coordinated mechanisms involving nuclear receptors is activated by bile acids, which collectively decrease the uptake and increase the export of bile acids and bilirubin into and from hepatocytes, respectively. Simultaneously, detoxification of bile acids is enhanced, while their synthesis becomes downregulated [28–30]. The increased effort of cholestatic hepatocytes to remove bilirubin causes bilirubinuria and hyperbilirubinemia. As a result, a yellowish pigmentation of the skin and the conjunctival membranes over the sclera becomes visible, known as jaundice. Furthermore, the elevated presence of bile acids in the serum is thought to account for the typical skin itching in cholestasis patients [26, 27, 30]. The development of this AOP was performed according to OECD guidance, including consideration of the Bradford-Hill criteria for weight-of-evidence assessment and the OECD key questions for evaluating AOP confidence. Proposed KEs are the accumulation of bile, the induction of oxidative stress and inflammation, and the activation of nuclear

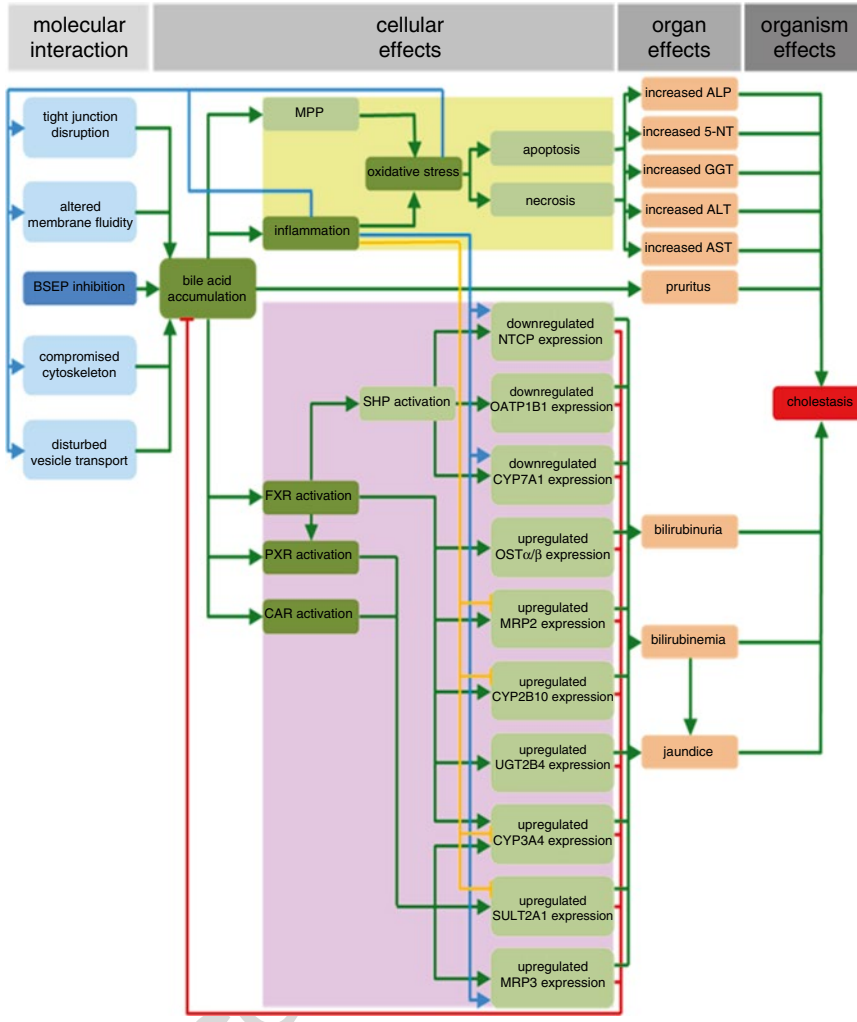


Fig. 4 *AOP for drug-induced cholestasis*. The response matrix between the MIE (*dark blue*) and AO (*red*), the inhibition of the bile salt export pump (BSEP) and cholestasis, respectively, spans over the cellular and organ levels. Identified KEs (*dark green*) include the accumulation of bile, the induction of oxidative stress and inflammation, and the activation of the nuclear receptors pregnane X receptor (PXR), farnesoid X receptor (FXR), and constitutive androstane receptor (CAR). Together with a number of intermediate steps, these KEs drive both a deteriorative cellular response (*yellow*), which underlies directly caused cholestatic injury, and an adaptive cellular response (*purple*), which is aimed at counteracting the primary cholestatic insults. Direct inducing and inhibiting effects are indicated with green and red arrows, respectively. Secondary inducing and inhibiting effects of oxidative stress and/or inflammation are indicated with blue and orange arrows, respectively [31] (5′-NT, 5′-nucleotidase; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CYP2B10/3A4/7A1, cytochrome P450 2B10/3A4/7A1; GGT, gamma-glutamyl transpeptidase; MPP, mitochondrial permeability pore; MRP2/3, multidrug resistance-associated protein 2/3; NTCP, sodium/taurocholate cotransporter; OATP1B1, organic anion transporter 1B1; OSTα/β organic solute transporter α/β; SHP, small heterodimeric partner; SULT2A1, dehydroepiandrosterone sulfotransferase; UGT2B4, uridine 5′-diphosphate-glucuronosyltransferase 2B4)

receptors. Furthermore, the AOP distinguishes direct adverse and indirect adaptive effects and takes a number of alternative MIE mechanisms into account [9, 31].

4 AOP Applications

4.1 Establishment of (Quantitative) Structure-Activity Relationships

As the MIE in each AOP involves a rather specific interaction of chemicals with biological systems, it can be used as the basis for generating structure-activity relationships, whether or not quantifiable. In turn, such information can be used for chemical grouping and read-across approaches, thus facilitating predictive and mechanism-based toxicology [1]. Using quantitative structure-activity relationship (QSAR) approaches, it has been demonstrated that chemicals with an ester bound to a carbon atom of a heterocyclic group or carbocyclic systems with a least one aromatic ring positively contribute to bile salt export pump inhibition, being the MIE in the AOP on drug-induced cholestasis, while the presence of hydroxyl groups bound to aliphatic carbon atoms has a negative contribution [32, 33]. In silico modeling further showed the role of hydroxyl groups in the interaction of chemicals with the bile salt export pump [34]. Two-dimensional and three-dimensional QSAR studies have also been performed on ligands of the liver X receptor, which constitutes the MIE in the AOP on drug-induced steatosis. By doing so, a number of chemical features, such as the presence of phenyl rings, chloro groups, and methyl moieties, have been identified as determinants of liver X receptor binding and activation [35].

4.2 Elaboration of Prioritization Strategies

Prioritization of chemicals denotes the process in which less complex, cheaper, and faster assays are used to determine which chemicals are subjected first to more complex, expensive, and slower testing [36]. AOPs have great potential with respect to prioritization strategies. Indeed, they can increase confidence in the integration of information, such as obtained from in vitro assays, for prioritizing chemicals for further assessment. The use of AOPs for the hepatotoxic endpoints described in this chapter in the context of the prioritization has not yet been described in current scientific literature. However, there are some examples for other adverse effects, including developmental toxicity. At present, the most promising alternative vertebrate models for screening of chemicals for developmental toxicity are fish embryos, in particular zebrafish. Using paraoxon, an acetylcholinesterase inhibitor, as a reference chemical, an AOP providing quantitative linkages across levels of biological organization during zebrafish embryogenesis has been proposed. Based on a series of experiments, it was found that normal acetylcholinesterase activity is not required for secondary motor neuron development and that acetylcholinesterase

inhibition, the MIE, may not be associated with an increased frequency of spontaneous tail contractions following paraoxon exposure. This AOP may support chemical screening and prioritization strategies with respect to developmental toxicity testing [37].

4.3 Development of In Vitro Tests

An essential step during AOP development is the labeling of KEs. In turn, this may serve as the basis for the characterization of biomarkers and simultaneously for the establishment of ex vivo, but especially in vitro, toxicity screening assays applicable for regulatory testing purposes. Furthermore, such new non-animal tests might be implemented into integrated testing strategies, thereby contributing to the refinement, reduction, and replacement of conventional in vivo testing. Reversely, by linking proposals for the development of in vitro test methods to KEs in an AOP, the relationship to hazard endpoints relevant for regulatory purposes can be established [1, 12].

5 Conclusions and Perspectives

Although conceptually not entirely new, AOPs have found their way to the human risk assessment arena in recent years, including the safety evaluation of drugs. The potential use of AOPs in this field is indeed considerably larger than the mode-of-action concept, as, at least ideally, it considers an exposure aspect and because it is not restricted to the tissue and individual level. However, despite the introduction of OECD guidance on AOP development and evaluation [1, 9], this area is still in its infancy and will greatly benefit from fine-tuning in the upcoming years (Table 3). A major criticism on AOPs nowadays is their simplicity and thus their poor reflection of complex toxicological processes. AOPs are presented as stand-alone linear events, yet the reality is likely to be much less straightforward, since parallel cascades and crossing of pathways may be involved. It is important that the overall toxicological scenario does not become lost when using AOPs. Furthermore, AOPs are to be considered as open and flexible structures that should be continuously refined by feeding in old and new data. Such iterative refinement exercises should ideally include the elaboration and quantification of the KERs as well as the specification of toxicokinetic conditions governing the activation of an AOP. Thus, classical kinetic determinants, like absorption, distribution, metabolism, and excretion, as well as more specific events, such as hormonal influences and adaptive responses, must be considered in AOP development. Another hurdle to overcome in the near future relates to the weight of evidence of data that are proposed to substantiate an AOP. Basically, anyone can propose an AOP, but not all AOPs are sufficiently supported by data. In order to develop confidence in the accuracy and utility of AOPs, there needs to be a

t3.1 **Table 3**
t3.2 **Major challenges for future AOP development**

t3.3 – Complying with the overall complexity of toxicological processes
t3.4 – Quantification and inclusion of dose-response relationships
t3.5 – Implementation of exposure data
t3.6 – Implementation of toxicokinetic data
t3.7 – Establishment of a transparent and objective evaluation system

transparent evaluation process that includes all stakeholders. In addition to hazard identification and the establishment of dose-response relationships, the risk assessment paradigm also includes implementation of exposure data. Thus far, this has gained little attention in the context of AOP development, thereby defining another challenge lying ahead. Several efforts are currently ongoing around the globe to tackle these issues, including at the OECD level [1, 9], the US Hamner Institutes of Health [38], the US Center for Alternatives to Animal Testing [39], and the European research program called Safety Evaluation Ultimately Replacing Animal Testing [40, 41]. Such projects are anticipated to yield robust and reliable AOP tools that can be used for a variety of purposes pertinent to toxicology and risk assessment, including the safety evaluation of new drug candidates.

Acknowledgments

This work was financially supported by the grants of the University Hospital of the Vrije Universiteit Brussel, Belgium (Willy Gepts Fonds UZ-VUB); the University of São Paulo, Brazil (USP); the São Paulo Research Foundation, Brazil (FAPESP); the Fund for Scientific Research, Flanders (FWO-Vlaanderen); the European Research Council (project CONNECT); the European Union (FP7); and Cosmetics Europe (projects DETECTIVE and HeMiBio).

References

1. OECD (2013) Proposal for a template and guidance on developing and assessing the completeness of adverse outcome pathways. Series on Testing and Assessment 184: 1–45
2. US EPA (2005) Guidelines for carcinogen risk assessment. Washington, DC
3. Bogdanffy MS, Daston G, Faustman EM et al (2001) Harmonization of cancer and noncancer risk assessment: proceedings of a consensus-building workshop. Toxicol Sci 61:18–31
4. Julien E, Boobis AR, Olin SS (2009) The key events dose-response framework: a cross-disciplinary mode-of-action based approach to

[AU2]

examining dose-response and thresholds. Crit Rev Food Sci Nutr 49:682–689
5. Meek ME, Bucher JR, Cohen SM et al (2003) A framework for human relevance analysis of information on carcinogenic modes of action. Crit Rev Toxicol 33:591–653
6. Seed J, Carney EW, Corley RA et al (2005) Overview: using mode of action and life stage information to evaluate the human relevance of animal toxicity data. Crit Rev Toxicol 35:664–672
7. NRC (2007) Toxicity testing in the 21st century: a vision and a strategy. The National Academies Press, Washington, DC
8. Ankley GT, Bennett RS, Erickson RJ et al (2010) Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. Environ Toxicol Chem 29:730–741
9. https://aopkb.org/ (consulted February 2015)
10. Villeneuve DL, Crump D, Garcia-Reyero N et al (2014) Adverse outcome pathway (AOP) development I: strategies and principles. Toxicol Sci 142:312–320
11. Villeneuve DL, Crump D, Garcia-Reyero N et al (2014) Adverse outcome pathway development II: best practices. Toxicol Sci 142:321–330
12. Vinken M (2013) The adverse outcome pathway concept: a pragmatic tool in toxicology. Toxicology 312:158–165
13. Hill AB (1965) The environment and disease: association or causation? Proc R Soc Med 58:295–300
14. Vinken M, Maes M, Vanhaecke T, Rogiers V (2013) Drug-induced liver injury: mechanisms, types and biomarkers. Curr Med Chem 20:3011–3021
15. Begriche K, Massart J, Robin MA et al (2011) Drug-induced toxicity on mitochondria and lipid metabolism: mechanistic diversity and deleterious consequences for the liver. J Hepatol 54:773–794
16. Cohen JC, Horton JD, Hobbs HH (2011) Human fatty liver disease: old questions and new insights. Science 332:1519–1523
17. Amacher DE (2011) The mechanistic basis for the induction of hepatic steatosis by xenobiotics. Expert Opin Drug Metab Toxicol 7:949–965
18. Ramachandran R, Kakar S (2009) Histological patterns in drug-induced liver disease. J Clin Pathol 62:481–492
19. Zimmerman HJ (2000) Drug-induced liver disease. Clin Liver Dis 4:73–96
20. Landesmann B., Goumenou M., Munn S., Whelan M. (2012): Description of prototype modes-of-action related to repeated dose toxicity. JRC Scientific and Policy Report, 75689
21. Friedman SL (2008) Mechanisms of hepatic fibrogenesis. Gastroenterology 134:1655–1669
22. Friedman SL (2010) Evolving challenges in hepatic fibrosis. Nat Rev Gastroenterol Hepatol 7:425–436
23. Lee UE, Friedman SL (2011) Mechanisms of hepatic fibrogenesis. Best Pract Res Clin Gastroenterol 25:195–206
24. Schoemaker MH, Conde de la Rosa L, Buist-Homan M et al (2004) Tauroursodeoxycholic acid protects rat hepatocytes from bile acid-induced apoptosis via activation of survival pathways. Hepatology 39:1563–1573
25. Woolbright BL, Jaeschke H (2012) Novel insight into mechanisms of cholestatic liver injury. World J Gastroenterol 18:4985–4993
26. Hofmann AF (2009) Bile acids and the enterohepatic circulation. In: Arias IM, Alter HJ, Boyer JL, Cohen DE, Fausto N, Shafritz DA, Wolkoff AW (eds) The liver: biology and pathobiology. Wiley-Blackwell, Oxford, pp 289–304
27. Padda MS, Sanchez M, Akhtar AJ, Boyer JL (2011) Drug-induced cholestasis. Hepatology 53:1377–1387
28. Zollner G, Trauner M (2006) Molecular mechanisms of cholestasis. Wien Med Wochenschr 156:380–385
29. Zollner G, Trauner M (2008) Mechanisms of cholestasis. Clin Liver Dis 12:1–26
30. Wagner M, Zollner G, Trauner M (2009) New molecular insights into the mechanisms of cholestasis. J Hepatol 51:565–580
31. Vinken M, Landesmann B, Goumenou M et al (2013) Development of an adverse outcome pathway from drug-mediated bile salt export pump inhibition to cholestatic liver injury. Toxicol Sci 136:97–106
32. Hirano H, Kurata A, Onishi Y et al (2006) High-speed screening and QSAR analysis of human ATP-binding cassette transporter ABCB11 (bile salt export pump) to predict drug-induced intrahepatic cholestasis. Mol Pharm 2:252–265
33. Saito H, Osumi M, Hirano H et al (2009) Technical pitfalls and improvements for high-speed screening and QSAR analysis to predict inhibitors of the human bile salt export pump (ABCB11/BSEP). AAPS J 11:581–589
34. Warner DJ, Chen H, Cantin LD et al (2012) Mitigating the inhibition of human bile salt export pump by drugs: opportunities provided

by physicochemical property modulation, *in silico* modeling, and structural modification. Drug Metab Dispos 40:2332–2341

35. Honorio KM, Salum LB, Garratt RC et al (2008) Two- and three-dimensional quantitative structure–activity relationships studies on a series of liver X receptor ligands. Open Med Chem J 2:87–96
36. Judson R, Kavlock R, Martin M et al (2013) Perspectives on validation of high-throughput assays supporting 21st century toxicity testing. ALTEX 30:51–56
37. Yozzo KL, McGee SP, Volz DC (2013) Adverse outcome pathways during zebrafish embryogenesis: a case study with paraoxon. Aquat Toxicol 126:346–354
38. Andersen ME, Clewell R, Bhattacharya S (2012) Developing in vitro tools sufficient by themselves for 21st century risk assessment. In: Gocht T, Schwarz M (eds) Towards the replacement of in vivo repeated dose systemic toxicity testing, vol 2. Imprimerie Mouzet, France, pp 347–360
39. http://caat.jhsph.edu/ (consulted February 2015)
40. http://www.seurat-1.eu/ (consulted February 2015)
41. Vinken M, Pauwels M, Ates G et al (2012) Screening of repeated dose toxicity data present in SCC(NF)P/SCCS safety evaluations of cosmetic ingredients. Arch Toxicol 86:405–412

Author Queries

Chapter No.: 14 0002663028

Queries	Details Required	Author's Response
AU1	Please check whether the e-mail address and affiliation are appropriate.	
AU2	Please provide complete bibliographic details for Ref. [2].	

Chapter 15

A Systems Biology Approach for Identifying Hepatotoxicant Groups Based on Similarity in Mechanisms of Action and Chemical Structure

[AU1] **Dennie G.A.J. Hebels, Axel Rasche, Ralf Herwig, Gerard J.P. van Westen, Danyel G.J. Jennen, and Jos C.S. Kleinjans**

Abstract

When evaluating compound similarity, addressing multiple sources of information to reach conclusions about common pharmaceutical and/or toxicological mechanisms of action is a crucial strategy. In this chapter, we describe a systems biology approach that incorporates analyses of hepatotoxicant data for 33 compounds from three different sources: a chemical structure similarity analysis based on the 3D Tanimoto coefficient, a chemical structure-based protein target prediction analysis, and a cross-study/cross-platform meta-analysis of in vitro and in vivo human and rat transcriptomics data derived from public resources (i.e., the diXa data warehouse). Hierarchical clustering of the outcome scores of the separate analyses did not result in a satisfactory grouping of compounds considering their known toxic mechanism as described in literature. However, a combined analysis of multiple data types may hypothetically compensate for missing or unreliable information in any of the single data types. We therefore performed an integrated clustering analysis of all three data sets using the R-based tool iClusterPlus. This indeed improved the grouping results. The compound clusters that were formed by means of iClusterPlus represent groups that show similar gene expression while simultaneously integrating a similarity in structure and protein targets, which corresponds much better with the known mechanism of action of these toxicants. Using an integrative systems biology approach may thus overcome the limitations of the separate analyses when grouping liver toxicants sharing a similar mechanism of toxicity.

Key words Systems biology, 3D Tanimoto, Protein targets, Meta-analysis, iClusterPlus, Hepatotoxicity, Chemical structure, Mechanism of action, Similarity, diXa

1 Introduction

Systems biology is an interdisciplinary field of study that focuses on complex interactions within biological systems. It uses a holistic approach that aims at integrating data from multiple sources to study the interactions between the components of biological systems and gain a wider understanding of how these interactions give rise to the function and behavior of that system, e.g., a pathway, a cell, etc.

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_15, © Springer Science+Business Media New York 2016

In other words, instead of taking apart a system and studying each of its individual components, systems biology focuses on integrating all these parts to reach a new level of understanding under the assumption that the whole is more than the sum of its parts.

Omics technologies are particularly useful for this purpose since they cover a large part of the changes in a certain part of the system, such as the transcriptome, the proteome, or the metabolome, thereby aiding the systems biology approach. However, despite the vast amount of information obtained from omics techniques, single omics analysis still does not always provide sufficient information to understand the behaviors of, for example, a cellular system. Therefore, a combination of multiple omics analyses and/or other data sources, the multi-omics (or multi-data source) approach, is needed to acquire a more precise picture of a system [1–5]. Combining multiple data types also has the advantage of being able to compensate for missing or unreliable information in any of the single data types and decreases the likelihood of false-positive findings.

In the field of hepatotoxicity, systems biology approaches are also receiving much attention [6–11]. Given the liver's vital role as a detoxification organ, it is not surprising that hepatotoxicity is the most prominent adverse reaction against drugs. As a result many newly developed candidate drugs fail in preclinical or clinical trials which is associated with a huge financial drain considering that the costs to develop a fully approved drug are around $800 million [12]. Failure to pick up hepatotoxicity in early stages is also contributable to the idiosyncratic nature of many adverse reactions, i.e., unusual individual reactions with very low frequency likely associated with differences in genetic make-up between individuals [13]. New screening methods, able to detect (idiosyncratic) drug-induced liver injury in the early stages of the research process, represent an important step toward efficient new drug development. Despite their poor predictive accuracy, animal models are still considered the gold standard toxicological approach for evaluating chemical toxicity and contribute substantially to the high costs involved in drug development [14]. In vitro systems are therefore increasingly studied with the ultimate goal of replacing animal models. Because of the time-saving nature and practicality of such systems, they are especially well suited to study drug metabolism, measure enzyme kinetics, evaluate toxicity mechanisms, and examine dose–response relationships using systems biology approaches [15]. The systems biology "map" of a hepatotoxic compound of interest may serve as a profile of its (idiosyncratic) toxicological mechanism. Studying large compilations of such compound profiles can thus assist in finding groups of compounds with similar (toxicological) mechanisms of action by comparing profiles and thereby assist in the early identification and elimination of compounds with a potential hepatotoxic effect.

In this chapter we will demonstrate a systems biology approach focused on compiling compound profiles from multiple data sources in order to group toxic compounds based on similarity. There are many data types available which can be used to obtain such similarity measures. Here, transcriptomics and proteomics data are of particular interest. While such omics data are excellent sources to explore the biological signaling cascades involved in hepatotoxic responses, including sources that focus more on the chemical similarities of the compounds may contribute significantly to the grouping of compounds with comparable hepatotoxic mechanisms. Given the crucial role of chemical structures with respect to xenobiotic metabolism in the liver, quantifying the chemical similarity of molecules is a very active field of research. In our multi-data source systems biology approach, we will therefore focus on a combination of these two approaches. A test data set will be used to illustrate an integrative analysis approach of a transcriptomics analysis and two chemical structure-based analyses. These three analysis approaches will first be explained in more detail separately. They involve a chemical structure similarity analysis based on the 3D Tanimoto coefficient, a chemical structure-based protein target prediction analysis, and a comprehensive transcriptomics meta-analysis. A hierarchical clustering-based grouping of the analysis results will be used to discuss the limitations of the individual methods by comparing the outcome with the known mode of action as described in literature. The multi-omics tool iClusterPlus will subsequently be presented as a means of overcoming these limitations and integrating multiple sources of information to improve grouping of similarly acting hepatotoxic compounds.

2 Data Set

To demonstrate the application of multisource data analysis on hepatotoxicity data, we queried the Data Infrastructure for Chemical Safety Assessment (diXa) data warehouse [16]. diXa is a recently created robust and sustainable infrastructure designed for storing toxicogenomics data. The warehouse is designed to store any type of omics data for every disease of interest and currently mostly contains transcriptomics data on hepatotoxicants and nephrotoxicants. The warehouse is connected to a portal with links to chemical information and molecular and phenotype data. diXa is publicly available through a user-friendly web interface, and new data can be readily deposited into diXa (http://wwwdev.ebi.ac.uk/fg/dixa/index.html, Fig. 1).

A selection of studies stored within diXa was downloaded to present as a use case in this chapter. The selection was based on an initial exploration of the data sets where we set out to include data

covering a wide range of experimental conditions (several doses and exposure times, in vitro and in vivo studies) and multiple species (rat and human). To improve data comparability, only studies using the same microarray platform (Affymetrix) were considered. Gene annotations were adjusted to their corresponding orthologues between species where needed. Using these criteria, nine studies were selected covering a total of 33 compounds as shown in Table 1.

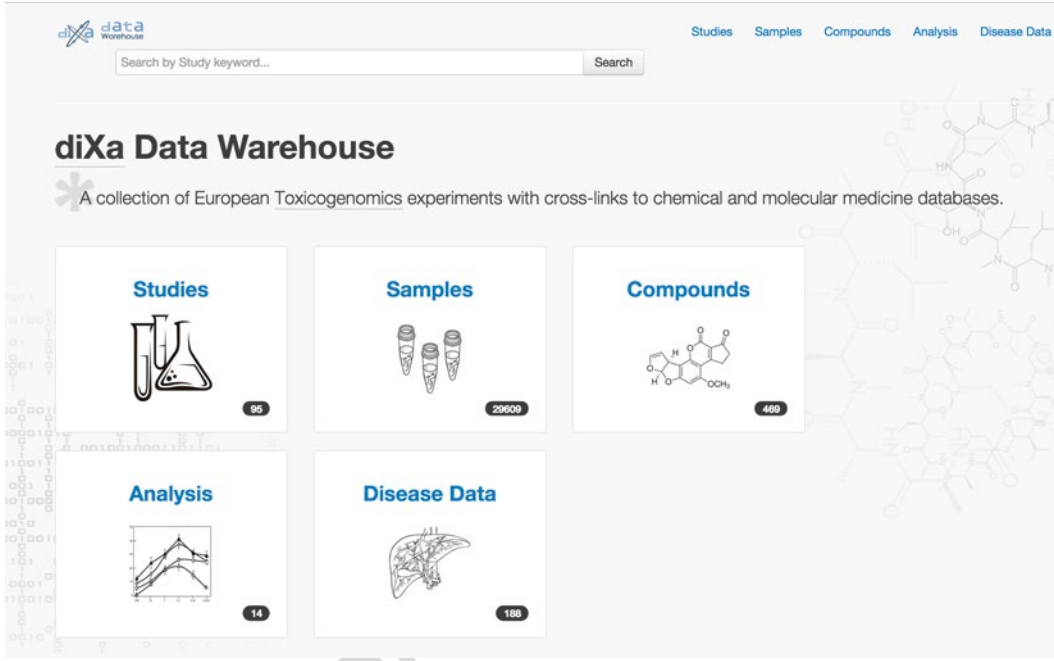


Fig. 1 The diXa data warehouse web portal provides immediate access to a wide range of transcriptomics studies

Table 1
Overview of studies included in analysis and the full list of hepatotoxic compounds

Project	Species	In vitro/in vivo	Cell/tissue type
carcinoGENOMICS	Homo sapiens	In vitro	HepaRG
	Homo sapiens	In vitro	HepG2
	Rattus norvegicus	In vitro	Primary rat hepatocytes
DrugMatrix	*Rattus norvegicus*	In vitro	Primary rat hepatocytes
	Rattus norvegicus	In vivo	Liver tissue

(continued)

Table 1
(continued)

	Project	Species	In vitro/in vivo	Cell/tissue type
t1.11	Predictomics	Homo sapiens	In vitro	HepG2
t1.12 t1.13	TG-GATEs	Homo sapiens	In vitro	Primary human hepatocytes
t1.14 t1.15		*Rattus norvegicus*	In vitro	Primary rat hepatocytes
t1.16		*Rattus norvegicus*	In vivo	Liver tissue
t1.17	Hepatotoxic compounds			
t1.18	1-Naphthyl isothiocyanate	Cyclophosphamide	Gemfibrozil	Phenobarbital
t1.19	Acetaminophen	Danazol	Ketoconazole	Pirinixic acid
t1.20	Aflatoxin B1	Diclofenac	Lomustine	Simvastatin
t1.21	Allyl alcohol	Doxorubicin	Methapyrilene	Sulindac
t1.22	Amiodarone	Ethanol	Nifedipine	Tamoxifen
t1.23	Azathioprine	Ethinyl estradiol	Nimesulide	Tetracycline
t1.24	Carbon tetrachloride	Fenofibrate	*N*-nitrosodimethylamine	Tolbutamide
t1.25	Clofibrate	Fluphenazine	Pemoline	Valproic acid
t1.26	Clomipramine			

t1.27 t1.28 Elaborate descriptions of all studies can be found in the diXa data warehouse (http://wwwdev.ebi.ac.uk/fg/dixa/index.html)

3 Tanimoto Similarity Score

Structural similarities between compounds may reflect similar mechanisms of action. Quantifying the similarity of two molecules is therefore a key concept in cheminformatics and pharmaceutical research. Although a close similarity between compounds can never guarantee an overlap in the mechanism of action, there is a strong correlation between the presence of certain structural subunits in a molecule and the eventual biological effect, which is a relationship that is often explored during the development of new pharmaceutical compounds. The Tanimoto coefficient [17] is a frequently used measure of chemical similarity and will be applied here to focus purely on the overlap in chemical properties of the compounds in the test data set.

3.1 Tanimoto Coefficient Procedure

Calculation of Tanimoto coefficient similarity scores can be performed in PubChem, which is an open repository for small molecules and their experimental biological activities [17]. Generating

Tanimoto scores is a very straightforward procedure and requires a list of compounds (compound name, PubChem compound database identifier (CID)) which can be uploaded. Subsequently structural similarity data will be calculated between each pair of compounds (https://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?p=clustering, Fig. 2). This resulting structure similarity matrix is then clustered using the single-linkage clustering algorithm.

The structural similarity in PubChem is either based on the Tanimoto score calculated from the 2D structure fingerprint or the 3D shape/feature similarity [17, 18]. The 2D structure fingerprint is based on an ordered list of binary substructures (i.e., fragments of a chemical structures) for chemical structures, in which each substructure is counted as either present or not present in the compound under investigation (e.g., an atomic element count, a type of ring system, atom pairing, atom neighbors, etc.). These fingerprints are used by PubChem for similarity neighboring and similarity searching [17].

A defining characteristic of 3D similarity methods, compared to 2D methods, is that they are applied at a conformer level instead of a compound level, thereby making it possible to consider the various distinct molecular conformations a compound can adopt in 3D space which may have biological relevance [19]. PubChem3D makes a distinction between two 3D similarity measures, i.e.,

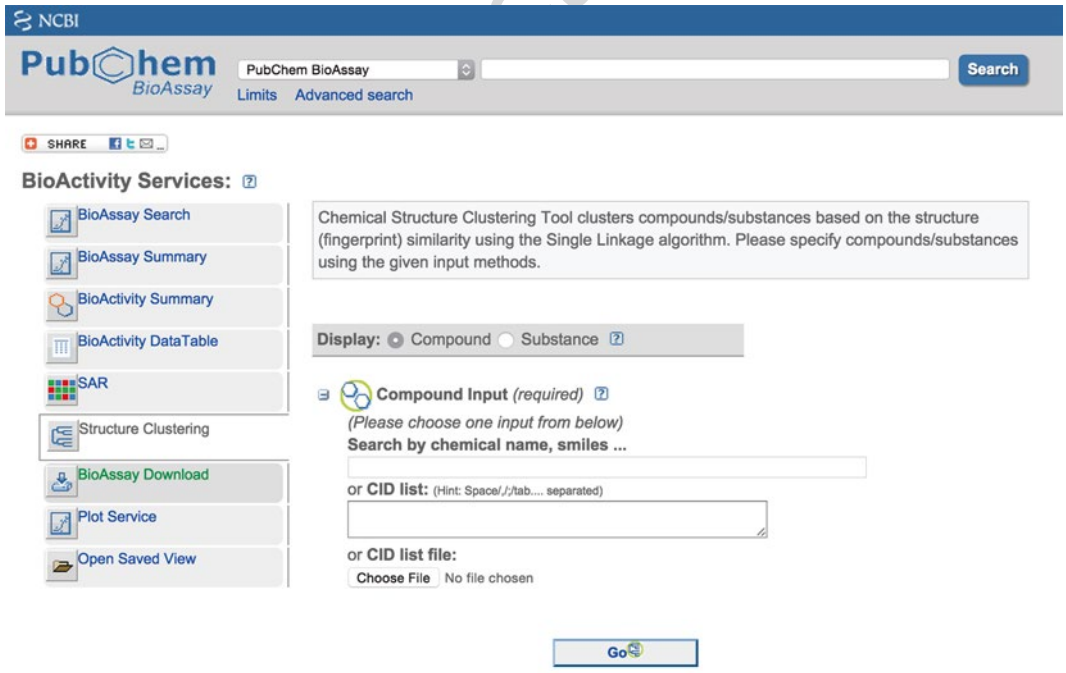


Fig. 2 The PubChem chemical structure clustering tool which generates a clustering dendrogram based on calculated Tanimoto scores (2D and 3D) for any list of compounds

shape-Tanimoto (ST) and color-Tanimoto (CT). The ST score is a measure of shape similarity, while the CT score quantifies the similarity of 3D orientation of functional groups or features by checking the overlap of fictitious "color" atoms which represent the six functional group types: hydrogen-bond donors, hydrogen-bond acceptors, anion, cation, hydrophobes, and rings. The ST and CT similarity metrics attempt to cover key aspects important for locating chemical structures that may have similar biological activity. In other words, the ST helps to identify compounds that can adopt a particular 3D shape (e.g., of a neurotransmitter bound in a particular conformational orientation in a postsynaptic membrane protein pocket), while the CT helps to identify compounds with similar 3D orientation of molecular features (e.g., necessary for making a hydrogen or ionic bond interaction of a neurotransmitter with its receptor). The assumption is then that compounds with highly similar 3D shape and feature orientations may also display similarities in their biological activity [19].

Given the importance of biological activity with respect to hepatotoxicity, in this chapter we will focus on the 3D Tanimoto scores. CID identifiers of the 33 compounds in our test data set were retrieved from PubChem, and 3D Tanimoto scores were calculated using the default options of a combined shape (ST) and feature (CT) similarity score (optimized for shape), which was followed by a clustering analysis (see paragraph 6).

4 Protein Target Analysis

Biological relevance and investigation of mode of action require an understanding of the proteins to which the compounds bind. Based on the chemical structure of the compounds, we can predict their interaction partners (protein targets) in an organism. This is done by comparing the structure of the compound to large curated literature-based databases of known compound–protein interactions such as DrugBank, ChEMBL, the Human Metabolome Database, and the Therapeutic Target Database [20–23]. In this chapter, we use the data in the ChEMBL database release 17 containing approximately 12 million data points [24, 25].

4.1 Protein Target Procedure

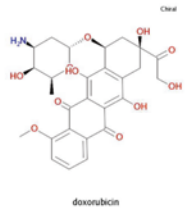
A multi-category naive Bayes statistical model trained on ChEMBL database release 17 was used for target prediction [25]. The compound structural features were encoded using extended-connectivity fingerprints with a diameter of six covalent bonds (ECFP6) as implemented in Pipeline Pilot (version 8.5, Accelrys Software Inc.) [26]. Target classes were limited to single protein targets with at least 30 active compounds (to ensure a robust model). Active was defined as having an activity better than 10 μM where the activity type was restricted to Ki, Kd, IC50, AC50, or EC50. In total

690,853 data points were used to construct the model. A multi-category model was then built for each of these proteins; herein relevant molecular features correlated to activity were identified by comparing the structure of actives per protein versus all of the other compounds (assumed inactive). Subsequently, each compound was scored with all 1282 models, and a ranked list of up to the top ten predicted protein targets for each compound was generated.

However, due to large differences in available data points per target (e.g., adenosine A2A receptor versus solute carrier organic anion transporter 1B1) and differences in average compound size per target (e.g., metabotropic glutamate receptors versus thrombin), the raw Bayesian score can differ significantly per protein target (per model class). To make the scores comparable, they were standardized in the form of *z*-scores [27]. The score per compound–protein pair was obtained for predictions by subtracting the mean score for the protein considered from the raw score and dividing this over the standard deviation for that protein (e.g., [1]). To obtain these values, after model training, the model was used to score all compounds in ChEMBL release 17 (1.3 million compounds). From this, a mean score per target and standard deviation per target were derived. Similarly, the mean score and standard deviation of compounds known to be active on a protein were calculated. After model predictions, targets with a standard score ≥2 were considered as a significant protein target for the

Compound name is: doxorubicin

AOJJSUZBOXZQNB-TZSSRYMLSA-N



Octanol / Water (ALogP): -4.4e-002

All predicted targets score >= 2

Z-Score	Z-Score Actives	Target Name	In Training
12.74	2.59	Breast cancer type 1 susceptibility protein:P38398:Homo sapiens	yes
8.03	-0.65	Multidrug resistance-associated protein 1:P33527:Homo sapiens	no
6.48	-0.47	Signal transducer and activator of transcription 6:P42226:Homo sapiens	yes
6.24	0.96	Peripheral myelin protein 22:Q01453:Homo sapiens	no
5.18	1.08	Hypoxia-inducible factor 1 alpha:Q16665:Homo sapiens	yes
5.12	0.55	Geminin:O75496:Homo sapiens	no
4.65	-0.02	Tyrosine-protein kinase FYN:P06241:Homo sapiens	yes
4.40	3.11	ATP-dependent DNA helicase Q1:P46063:Homo sapiens	yes
4.22	0.98	Thrombopoietin:P40225:Homo sapiens	no
4.20	1.10	Nuclear factor NF-kappa-B p105 subunit:P19838:Homo sapiens	no
4.17	2.64	DNA-(apurinic or apyrimidinic site) lyase:P27695:Homo sapiens	yes
4.17	-0.62	P-glycoprotein 1:P08183:Homo sapiens	yes
4.08	-1.17	Histone deacetylase 6:Q9UBN7:Homo sapiens	no
4.06	1.13	Bloom syndrome protein:P54132:Homo sapiens	no
3.55	0.02	Heat shock protein HSP 90-alpha:P07900:Homo sapiens	no

Fig. 3 Example output from the protein target analysis for the compound doxorubicin, showing compound structure and InChI key and the top 15 protein target *z*-scores and *z*-score actives

compound in question and reflect the enrichment of the score over randomness (i.e., all compounds in ChEMBL release 17) for the specific target of that compound in terms of standard deviations. Likewise *z*-score actives are calculated which show the difference a compound scores on this target compared to the average score of known actives for that protein. An example output for the compound doxorubicin is shown in Fig. 3. For further analyses as presented in paragraph 6, all calculated *z*-scores (significant and nonsignificant) were taken into account.

5 Gene Expression Meta-Analysis

An inherent problem of heterogeneous data sets is the experiment-specific variation which cannot be controlled for in a post hoc analysis. These variations stem from a variety of sources such as the use of different cell culture assays, differences in compound concentration and exposure time, and the use of different species (Table 1). To compensate for such variations, cross-study/cross-platform gene expression meta-analysis is a valid strategy to extract consistent information from a set of individual studies across a wide range of experimental conditions, including in vitro and in vivo data. In fact, combining data from in vitro and in vivo studies on liver carcinogens with gene expression data from human liver cancers was shown to improve carcinogenicity prediction [28]. Meta-analysis has been frequently applied in diseases with complex phenotypes such as cancer [29], Down syndrome [30], and diabetes mellitus type 2 [31]. A meta-analysis approach on hepatotoxicity-associated transcriptomics data can therefore be very valuable given the vast amount of heterogeneous data sets available in literature.

5.1 Meta-analysis Procedure

All experiments in the data set (*see* Table 1) have a case–control design comparing two groups of replicate samples. These groups are denoted as treatment and control, respectively, and constitute a test case. For a test case, the generated chips are normalized with each other using the R/Bioconductor framework.

The normalization accounts for three major influence factors in the hybridization data: background expression, probe binding affinity, and measurement variation. GC-RMA corrects for such effects [32]. In the background correction, it takes into account the GC content of the probe sequences, i.e., the number of G or C nucleotides in the sequence. A higher GC content is associated with a higher binding affinity of the probes due to three instead of two covalent bindings for single nucleotides. GC-RMA contains a position-specific model correcting the binding affinity between probes. Between chips unwanted effects are introduced by RNA extraction, pipetting, temperature fluctuations, hybridization

efficiency, and more. To reduce these effects, the quantile normalization is implemented in GC-RMA. Finally probe intensities are summarized into probe set expressions. GC-RMA uses median polish which proposes a linear model of a baseline hybridization with two factors, a probe effect and an array effect [33]. The model is fitted robustly with a median decomposition.

An advantage of the Affymetrix array design is the possibility to calculate a presence tag, i.e., the probability that the corresponding gene is effectively expressed and active in the sample under study. Non-expressed genes confuse the results with low intensities leading to high, unmotivated fold changes. The presence tag, or detection *p*-value, is based on a comparison of raw perfect-match values and corresponding mismatch values. Using a robust Wilcoxon test yields a *p*-value for each probe set which indicates whether or not the perfect-match probe signals are different from the mismatch probe signals and thus allows judging the expression of the corresponding gene.

Necessary for any meta-analysis is the consolidation of the different identifier types, different species, or different arrays [34]. The Ensembl database provides a stable reference for microarray studies (http://www.ensembl.org; version 74) and enables orthologue gene searches to allow for the combination of human and rat data. Since comparability of chip studies is hindered by the total number of probes and preprocessing issues between manufacturers, the analysis in this chapter constrains on Affymetrix arrays for case–control studies. Expression results from the arrays are mapped to Ensembl by the custom chip definition file (CDF) annotations [35].

The computation of gene expressions and presence tags is followed by a gene-wise evaluation of treatment versus control expressions. Expressions are assessed by two criteria: presence and alteration. For the approach of a meta-analysis, as presented in this chapter, the two criteria are condensed into a single score for every gene. The test case score *St* of a gene is computed as follows:

$$s_t = \begin{cases} \left|\log_2(r)\right|(1-10p) & , p \le 0.1 \\ 0 & , \text{else} \end{cases}$$

Here, r is the fold change and p is the average detection *p*-value. Thus, the fold change is corrected with its effective expression activity. The gene expression alteration in every study test case t is quantified with this score.

For every gene g, the scores from compound-specific studies are summarized constituting a gene–compound score *Sgc*:

$$s_{g-c} = \frac{T_g}{T_{g-c}} \sum_{g-c} s_t$$

So we sum up the gene scores over all test cases related to compound *c*. The sum is weighted by the quotient of *Tg* the number of test cases with gene scores divided by *Tgc* the number of test cases with scores for gene *g* and compound *c*. This weight compensates for genes which are not represented on every Affymetrix array, which is, for example, relevant for nonhomologous genes between human and rat. The results are discussed in the next paragraph together with the results of the other two analysis approaches.

6 Results of Individual Data Analysis Approaches

The Tanimoto 3D similarity scores are automatically processed in a clustering analysis, the results of which are shown in Fig. 4a. For comparison purposes the protein target *z*-scores and meta-analysis gene scores were also hierarchically clustered; this is shown in Fig. 4b, c (both Ward's clustering, using the "minimum increase in the sum of squares for error" method). This also allows for a more straightforward comparison of the individual analysis results with the integrative analysis covered in the next paragraph.

If we compare the two analysis methods based on chemical structure, i.e., the Tanimoto similarity scores and protein target *z*-scores, there is a number of subclusters that appear to correspond with certain protein target clusters. However, it is also apparent that the protein target scores tend to cluster into more distinct groups of compounds, whereas the Tanimoto dendrogram

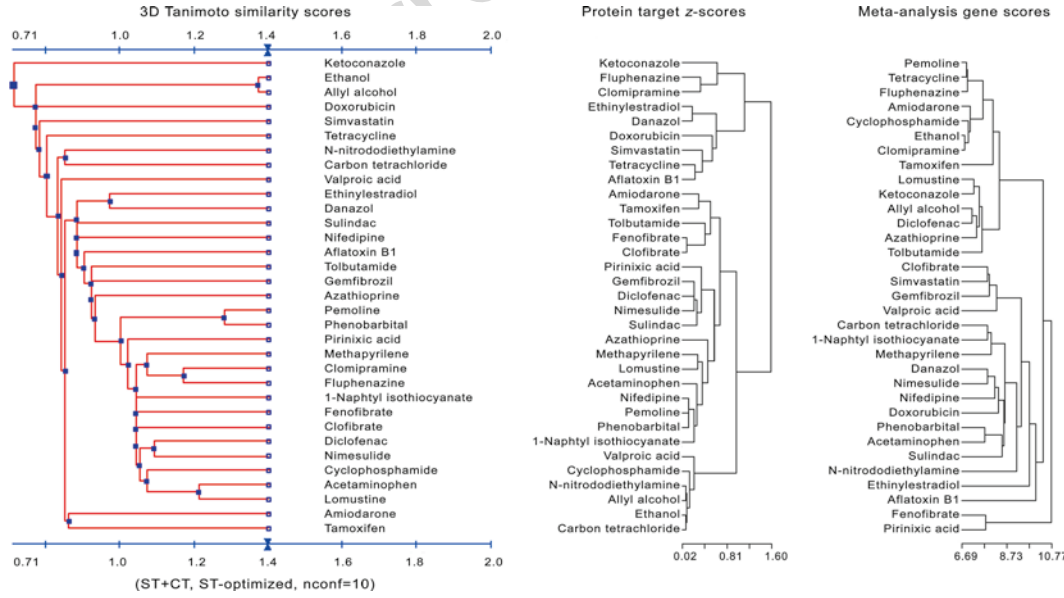


Fig. 4 Clustering dendrograms of the Tanimoto similarity scores (**a**), the protein target *z*-scores (**b**), and the meta-analysis gene scores (**c**)

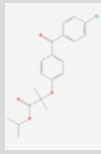
does not form any separate groups with the exception of the duo clusters ethanol/allyl alcohol and amiodarone/tamoxifen. These two duo clusters can also be readily recognized in the protein target dendrogram. Other small clusters which can also be distinguished in the Tanimoto dendrogram include ethinyl estradiol/danazol and pemoline/phenobarbital.

Strong disagreements between analyses become apparent when the meta-analysis is taken into account. Its dendrogram's clusters are quite inconsistent with the Tanimoto and protein score dendrograms, and no immediate overlap can be seen (Fig. 4). So the question arising is which one of these analyses is right? There is of course no straightforward answer to this. If we consider some of the grouped compounds in these dendrograms and compare them with what can be found in literature regarding known mechanism of action, we see that all three analyses cluster compounds as might be expected. We will use the following examples to illustrate this:

(a) Fenofibrate and pirinixic acid.

The meta-analysis suggests that fenofibrate and pirinixic acid induce a similar gene expression response, which indeed makes sense given the fact that they are both peroxisome proliferator-activated receptor alpha (PPARA) agonists [36]. The Tanimoto score analysis does not consider these compounds to be structurally related. Of course structural dissimilarity does not exclude the possibility of having a similar biological effect and vice versa. Small identical substructures in two molecules can already be enough to exert a similar effect even when the overall composition is very different. Conversely, a good example of compounds

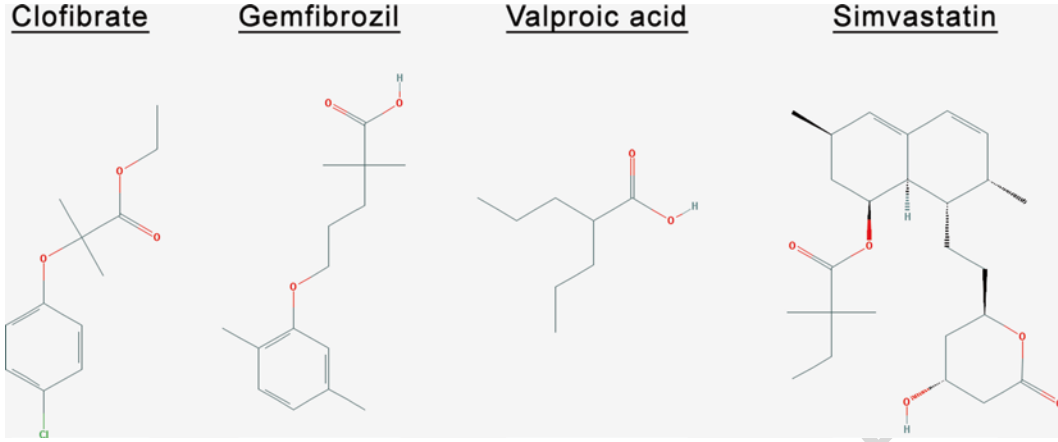
Table 2
List of significant protein targets (*z*-score >2) for fenofibrate and pirinixic acid compounds

Fenofibrate							Pirinixic acid		
Structure	Protein (HGNC)	*z*-Score	Protein (HGNC)	*z*-Score	Protein (HGNC)	*z*-Score	Protein (HGNC)	*z*-Score	Structure
	LSS	3.23	GLP1R	2.42	TRPA1	2.28	PTGES	4.82	
	PPARA	2.94	IGFBP3	2.39	CACNA1H	2.24	ALOX5	3.00	
	FFAR2	2.93	PPARD	2.39	P2RY1	2.22	PLA2G7	2.32	
	SCN2A	2.80	AKR1C2	2.39	CTSG	2.21	AKR1C2	2.24	
	SCN10A	2.79	CYP26A1	2.36	ELOVL6	2.20	CSNK2A2	2.16	
	PPARG	2.79	VCAM1	2.34	SRD5A2	2.13	PPARG	2.16	
	GIPR	2.60	ICAM1	2.29	SELE	2.06	CXCR2	2.10	
	ELANE	2.59	UTS2	2.28	MMP14	2.04	CTSA	2.00	

with high structural similarity but entirely different effects are the enantiomers of thalidomide; S-thalidomide is a severe teratogen, while R-thalidomide is a sedative with no teratogenic action. This difference in structure between fenofibrate and pirinixic acid also partially explains why the protein target analysis does not group these compounds together since this analysis is also based on the chemical (2D) structure using the ECFP6 fingerprints. However, if we take a closer look at the calculated *z*-scores of this analysis, there are also some inconsistencies with literature. Despite the fact that both compounds are PPARA agonists, PPARA is only a significant protein target for fenofibrate, not pirinixic acid. Another interesting observation is the significance of PPARD and PPARG for fenofibrate when this compound is usually not considered an agonist for these two PPARs [37]. The two top-scoring protein targets for pirinixic acid, prostaglandin E2 synthase-1 (PGES-1) and 5-lipoxygenase (ALOX5), also show an inconsistency with literature (Table 2). PGES and ALOX5 are only protein targets for pirinixic acid after substantial modification of the structure to an aminothiazole-featured pirinixic acid [38]. It thus appears that protein targets do not always reflect literature accurately, which may be related to drawbacks of the manual curation on which the algorithm is dependent.

(b) Clofibrate, gemfibrozil, valproic acid, and simvastatin.

The compounds clofibrate, gemfibrozil, valproic acid, and simvastatin form an obvious cluster in the meta-analysis but are completely scattered across the Tanimoto and protein target dendrograms. Clofibrate and gemfibrozil are PPARA agonists, while simvastatin, a statin compound, increases expression of PPARA and as such can have a similar effect [39]. Indeed there appears to be a cross-talk of statin signaling pathways and (agonist-induced) PPARA activity, and combination therapies of fibrates and statins are being used to treat dyslipidemia [40–42]. Valproic acid has a different mechanism of action and is used as an anticonvulsant and mood-stabilizing drug which has been attributed to the blockade of voltage-dependent sodium channels and increased brain levels of gamma-aminobutyric acid (GABA) [43]. However, it has also been found to be an activator of PPARD, but not PPARA or PPARG, although it is not a direct PPARD ligand [44, 45]. Valproic acid can therefore interact in the PPAR signaling cascades, which explains its similarity in gene expression with the other three compounds. Despite this similarity in gene expression and evidence in literature for overlap in mechanism of action, the Tanimoto and protein target analyses do not consider these compounds to be similar in their effect. However, a visual inspection of the molecular structures of these compounds does reveal a structural similarity, especially between clofibrate and gemfibrozil (Fig. 5). Moreover, the carboxylic (pentanoic) acid moiety in these two compounds is also present



[AU3]

Fig. 5 Molecular structures of clofibrate, gemfibrozil, valproic acid, and simvastatin

in valproic acid. This moiety is essential for fibrates to function as PPAR agonists [46]. Since both the 3D Tanimoto analysis and the 2D ECFP6-based protein target analysis take the entire structure into account, essential substructures that convey the similarity in working mechanism could be masked by a dissimilarity in the remainder of the molecule. Smaller molecules with structural similarities can therefore be expected to cluster together more readily as can be seen in the next example.

(c) Allyl alcohol, ethanol, carbon tetrachloride, and *N*-nitrosodimethylamine.

The compounds allyl alcohol, ethanol, carbon tetrachloride, and *N*-nitrosodimethylamine form clusters in the Tanimoto and protein target dendrograms but are completely scattered across the meta-analysis. Indeed their structures are very similar as shown in Fig. 6 which also contributes to the overlap in protein targets. While structural similarity does not guarantee similar gene expression responses, literature review does suggest that these compounds should share some common mode of action. For example, all four compounds are metabolized by the cytochrome P450 metabolizing enzyme CYP2E1 and/or alcohol dehydrogenase (ADH) causing oxidative stress which (partially) explains their hepatotoxic effects [47–52]. An explanation for the scattered clustering in the meta-analysis could lay in the fact that some essential information in the gene expression meta-analysis may get lost since we found that some compounds did cluster similarly to the protein target *z*-scores and Tanimoto scores when a distinction was made based on, for example, dose and exposure time (results not shown). Of course this is inherent to the approach of the meta-analysis, but could lead to problems with group identification if transcriptomic responses differ greatly between experimental conditions.

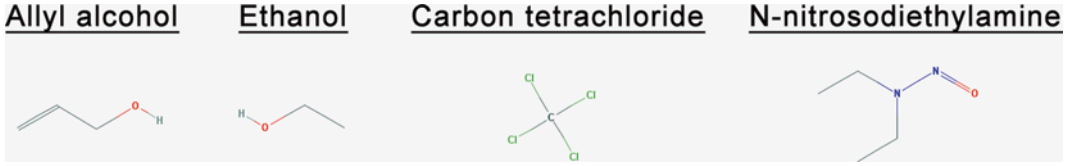


Fig. 6 Molecular structures of allyl alcohol, ethanol, carbon tetrachloride, and *N*-nitrosodimethylamine

7 Combined Analysis Using iClusterPlus

The individual analyses presented above reveal a number of shortcomings, which include (a) disagreements with the described mechanisms of action of compounds with respect to identified protein targets, (b) important similarities in compound substructures which are missed, and (c) a loss of important information when performing a cross-study/cross-platform meta-analysis. These limitations may be overcome by running an integrative clustering that takes into account all data in one single analysis and can resolve the considerable heterogeneity present in individual data sets. iClusterPlus is an R-based tool specifically designed for such a multi-data source integration using a joint latent variable model [53]. It is designed to perform pattern discovery that can integrate diverse data types such as binary values (e.g., somatic mutation data), categorical values (e.g., copy number gain, normal, loss), and continuous values (e.g., gene expression, protein levels) (Fig. 7).

Given multiple data types (e.g., gene expression, Tanimoto scores, protein target data, etc.) measured in the same set of samples and specified sparsity parameter values, iClusterPlus uses generalized linear regression to fit a regularized latent variable model-based clustering that generates an integrated cluster assignment based on joint inference across data types. The common set of latent variables represents distinct driving factors, which, geometrically speaking, form a set of principal coordinates that span a lower dimensional integrated subspace and collectively capture major biological variations, enabling rigorous analysis of the integrated genomic data [53]. The iClusterPlus package is available for download from the open-source software framework Bioconductor (http://www.bioconductor.org/).

7.1 iClusterPlus Results

Compounds with similar toxicity and/or mode of action were grouped using iClusterPlus by integrating meta-analysis gene scores, structural similarities, and protein target predictions. In order to guarantee that each data type has the same weight in the analyses, scaled Euclidian distances were used for meta-analysis gene scores and target predictions in the range of 0–1 (0 = most similar; 1 = most dissimilar), and for structural similarities the 3D Tanimoto scores were used in the range 0–2 (0 = most dissimilar; 2 = most similar).

The iClusterPlus analysis was performed using default settings except for the number of CPUs used for parallel computing (30 CPUs) and the lasso parameter λ which was rescaled to be between 0 and 0.1. These settings were used to determine the optimal number of clusters by calculating the percentage of total variation explained by the model for 2–21 clusters. The percentage explained variation typically increases as more clusters are introduced. The optimal number of clusters is where the curve of percentage explained variation levels off. Figure 8 shows the curve for the analysis with the three data types combined, where 16 clusters are indicated as the optimum number of clusters.

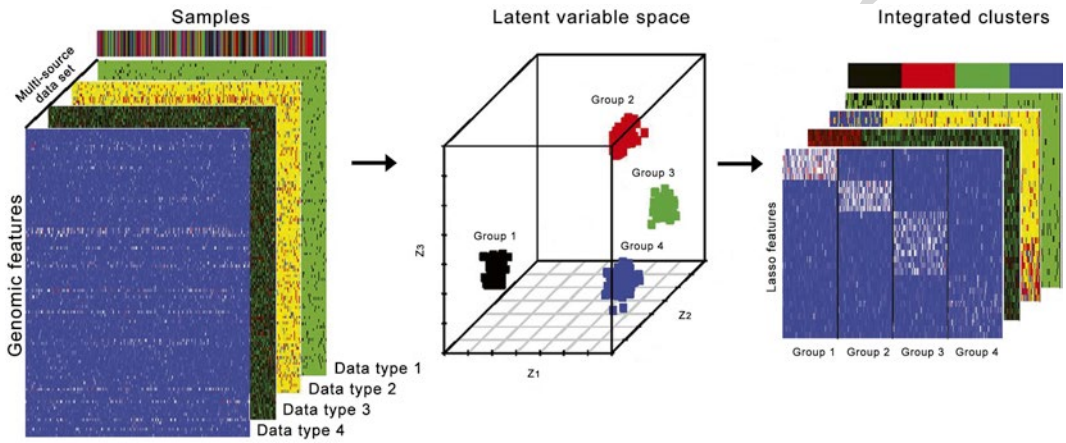


Fig. 7 The basic principle of iClusterPlus analysis. Adapted with PNAS permission from Ref. [53]

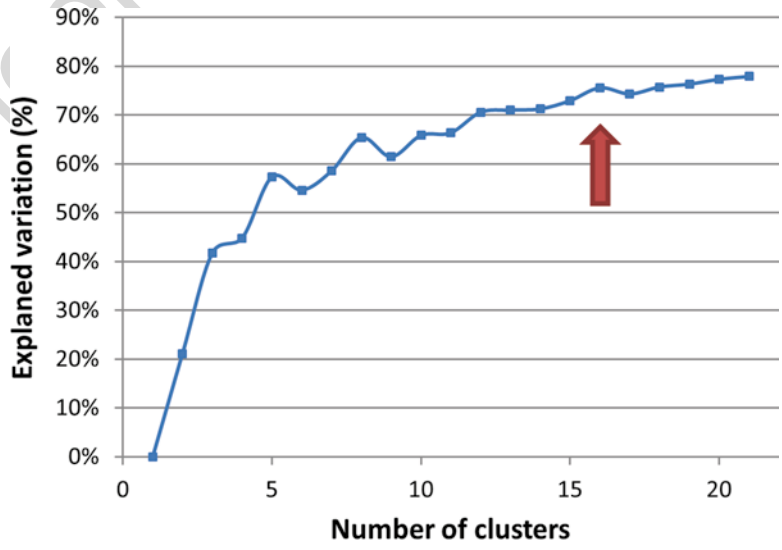


Fig. 8 Percentage explained variation curve for the analysis with the three data types combined. The arrow indicates the optimal number of clusters

There indeed appears to be a better grouping of compounds when all three approaches are combined (Fig. 9). For example, fenofibrate and pirinixic acid now cluster together (cluster #9) where they previously did only in the meta-analysis (Fig. 4). Protein targets in this case did not fully reflect the literature (which provides sufficient evidence for a similar mechanism of action), and the structures, while having some similarities, were found to be considered as different when taking into account the whole structure in the Tanimoto score analysis.

Clofibrate, gemfibrozil, simvastatin, and valproic acid previously grouped together in the meta-analysis which was supported by literature to a certain degree (all involved in peroxisome signaling), but structurally they are more dissimilar, and their protein targets are different because they work through different mechanisms (i.e., clofibrate and gemfibrozil are PPARA agonists, while simvastatin increases PPARD expression and valproic acid affects PPARD signaling). This is now much better reflected by the clustering in Fig. 9 where clofibrate and gemfibrozil cluster together

Compound	Cluster #
Ethinylestradiol	1
Aflatoxin B1	2
Amiodarone Tetracycline	3
Danazol	4
Doxorubicin Fluphenazine Ketoconazole	5
Tamoxifen	6
Simvastatin	7
Diclofenac Lomustine Tolbutamide	8
Fenofibrate Pirinixic acid	9
Cyclophosphamide Pemoline	10
N-nitrosodiethylamine	11
Clofibrate Gemfibrozil	12 12
1-Naphthyl isothiocyanate Acetaminophen Methapyrilene Nifedipine Nimesulide Phenobarbital Sulindac Valproic acid	13
Clomipramine	14
Azathioprine	15
Allyl alcohol Carbon tetrachloride Ethanol	16

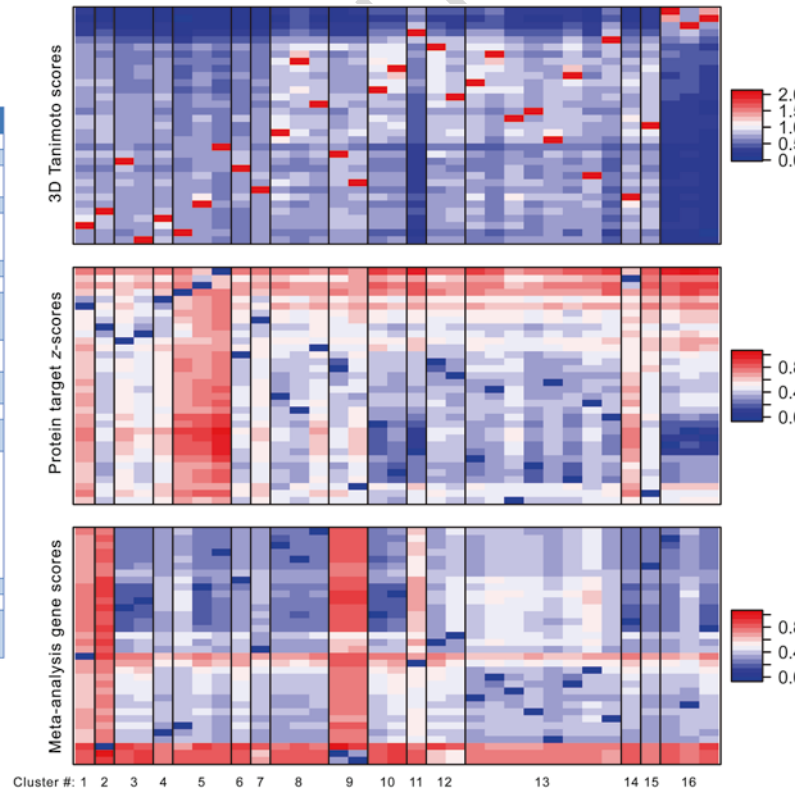


Fig. 9 iClusterPlus results, showing the grouping of the 33 compounds in the data set based on an integrated multisource analysis of protein target *z*-scores (Euclidian distances), meta-analysis gene scores (Euclidian distances), and 3D Tanimoto scores. The order of the compounds in the table corresponds with the column order in the clustering heatmap

(cluster #12), gemfibrozil forms a separate group (cluster #7), and valproic acid is clustered together with a set of other compounds (cluster #13). These compounds include the COX-2-selective, nonsteroidal anti-inflammatory drug nimesulide which is known to affect both GABA neurotransmission and PPARD signaling just like valproic acid [43, 54, 55] and phenobarbital, which is also an anticonvulsant that interacts with the GABAergic response [56].

According to literature, allyl alcohol, carbon tetrachloride, ethanol, and *N*-nitrosodimethylamine all have a somewhat similar metabolic mechanism and toxicity (CYP2E1/ADH metabolism, oxidative stress). Indeed these compounds had similar protein targets and a similarity in structure (Fig. 4, ethanol and allyl alcohol form a group and carbon tetrachloride and *N*-nitrosodimethylamine), but this was not reflected by the meta-analysis data. However, when separate doses and time points were investigated, this grouping was better (results not shown). The iClusterPlus analysis now also shows a much better grouping of these compounds with only NDEA forming a separate group (#11).

It thus appears that an integrated analysis of data from multiple sources potentially leads to an improved clustering of related hepatotoxic compounds.

8 Conclusion

In this chapter, we have presented an approach that focuses on integrating hepatotoxic compound-induced gene expression and (protein target-directed) chemical structural patterns in order to evaluate whether they can complement each other. The presented examples show that grouping compounds based solely on cross-study/cross-platform gene expression, 3D chemical structure, or protein targets can result in wrongly clustered compounds which have different toxicity or mode of action. To overcome these limitations, iClusterPlus is shown to be a promising tool for integrating data from several distinct sources and improving the clustering of related compounds which share a common mechanism of action. It should be pointed out though that evaluation of the identified groups is needed by (literature-based) expert judgment. Still, a systems biology approach where multiple data sources are used, especially when these data types focus on different aspects of compound (hepato)toxicity and/or chemistry, appears to be a promising way of handling big data sets and promoting the development of new pharmaceutical compounds. The flexibility of iClusterPlus with regard to data set types (e.g., binary, categorical, and continuous values) allows for many data sets to be included in the analysis if considered toxicologically relevant. Inclusion of other data sources, such as proteomics or fragment-based fingerprint methods, is likely to further improve the grouping of similar compounds.

References

1. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D (2015) Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet 16:85–97
2. Holzinger ER, Ritchie MD (2012) Integrating heterogeneous high-throughput data for meta-dimensional pharmacogenomics and disease-related studies. Pharmacogenomics 13:213–222
3. Reif DM, White BC, Moore JH (2004) Integrated analysis of genetic, genomic and proteomic data. Expert Rev Proteomics 1:67–75
4. Hamid JS, Hu P, Roslin NM, Ling V, Greenwood CM, Beyene J (2009) Data integration in genetics and genomics: methods and challenges. Hum Genom Proteomics. DOI: 10.4061/2009/869093
5. Hawkins RD, Hon GC, Ren B (2010) Next-generation genomics: an integrative approach. Nat Rev Genet 11:476–486
6. Shon J, Abernethy DR (2014) Application of systems pharmacology to explore mechanisms of hepatotoxicity. Clin Pharmacol Ther 96:536–537
7. Howell BA, Siler SQ, Watkins PB (2014) Use of a systems model of drug-induced liver injury (DILIsym((R))) to elucidate the mechanistic differences between acetaminophen and its less-toxic isomer, AMAP, in mice. Toxicol Lett 226:163–172
8. Bhattacharya S, Shoda LK, Zhang Q, Woods CG, Howell BA, Siler SQ, Woodhead JL, Yang Y, McMullen P, Watkins PB, Andersen ME (2012) Modeling drug- and chemical-induced hepatotoxicity with systems biology approaches. Front Physiol 3:462
9. Chen M, Vijay V, Shi Q, Liu Z, Fang H, Tong W (2011) FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov Today 16:697–703
10. Cui Y, Paules RS (2010) Use of transcriptomics in understanding mechanisms of drug-induced toxicity. Pharmacogenomics 11:573–585
11. Giuliano KA, Gough AH, Taylor DL, Vernetti LA, Johnston PA (2010) Early safety assessment using cellular systems biology yields insights into mechanisms of action. J Biomol Screen 15:783–797
12. DiMasi JA, Hansen RW, Grabowski HG (2003) The price of innovation: new estimates of drug development costs. J Health Econ 22:151–185
13. Senior JR (2008) What is idiosyncratic hepatotoxicity? What is it not? Hepatology 47: 1813–1815
14. Holmes AM, Creton S, Chapman K (2010) Working in partnership to advance the 3Rs in toxicity testing. Toxicology 267:14–19
15. Soldatow VY, Lecluyse EL, Griffith LG, Rusyn I (2013) In vitro models for liver toxicity testing. Toxicol Res (Camb) 2:23–39
16. Hendrickx DM, Aerts HJ, Caiment F, Clark D, Ebbels TM, Evelo CT, Gmuender H, Hebels DG, Herwig R, Hescheler J, Jennen DG, Jetten MJ, Kanterakis S, Keun HC, Matser V, Overington JP, Pilicheva E, Sarkans U, Segura-Lepe MP, Sotiriadou I, Wittenberger T, Wittwehr C, Zanzi A, Kleinjans JC (2015) diXa: a data infrastructure for chemical safety assessment. Bioinformatics 31:1505–1507
17. Bolton EE, Chen J, Kim S, Han L, He S, Shi W, Simonyan V, Sun Y, Thiessen PA, Wang J, Yu B, Zhang J, Bryant SH (2011) PubChem3D: a new resource for scientists. J Cheminform 3:32
18. Kim S, Bolton EE, Bryant SH (2012) Effects of multiple conformers per compound upon 3-D similarity search and bioassay data analysis. J Cheminform 4:28
19. Kim S, Bolton EE, Bryant SH (2011) PubChem3D: biologically relevant 3-D similarity. J Cheminformatics 3:26
20. Jenkins JL, Bender A, Davies JW (2006) In silico target fishing: predicting biological targets from chemical structure. Drug Discov Today Tech 3:413–421
21. Nidhi, Glick M, Davies JW, Jenkins JL (2006) Prediction of biological targets for compounds using multiple-category Bayesian models trained on chemogenomics databases. J Chem Inf Model 46:1124–1133
22. Southan C, Sitzmann M, Muresan S (2013) Comparing the chemical structure and protein content of ChEMBL, DrugBank, human metabolome database and the therapeutic target database. Mol Informat 32:881–897
23. Mugumbate G, Abrahams KA, Cox JA, Papadatos G, van Westen G, Lelievre J, Calus ST, Loman NJ, Ballell L, Barros D, Overington JP, Besra GS (2015) Mycobacterial dihydrofolate reductase inhibitors identified using chemogenomic methods and in vitro validation. PLoS One 10:e0121492
24. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP

(2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:D1100–D1107

25. ChEMBL Team (2013) ChEMBL release 17. DOI: 10.6019/CHEMBL.database.17
26. Rogers D, Hahn M (2010) Extended-connectivity fingerprints. J Chem Inf Model 50:742–754
27. Kreyszig E (1979) Applied mathematics. Wiley Press, New York
28. Caiment F, Tsamou M, Jennen D, Kleinjans J (2014) Assessing compound carcinogenicity in vitro using connectivity mapping. Carcinogenesis 35:201–207
29. Rhodes DR, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A 101:9309–9314
30. Vilardell M, Rasche A, Thormann A, Maschke-Dutz E, Perez-Jurado LA, Lehrach H, Herwig R (2011) Meta-analysis of heterogeneous down syndrome data reveals consistent genome-wide dosage effects related to neurological processes. BMC Genomics 12:229
31. Rasche A, Al-Hasani H, Herwig R (2008) Meta-analysis approach identifies candidate genes and associated molecular networks for type-2 diabetes mellitus. BMC Genomics 9:310
32. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193
33. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003) Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31:e15
34. Rasche A, Yildirimman R, Herwig R (2009) Integrative analysis of microarray data: a path for systems toxicology, General, applied and systems toxicology. Wiley, Hoboken, NJ
35. Dai M, Wang P, Boyd AD, Kostov G, Athey B, Jones EG, Bunney WE, Myers RM, Speed TP, Akil H, Watson SJ, Meng F (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 33:e175
36. Guo L, Fang H, Collins J, Fan XH, Dial S, Wong A, Mehta K, Blann E, Shi L, Tong W, Dragan YP (2006) Differential gene expression in mouse primary hepatocytes exposed to the peroxisome proliferator-activated receptor alpha agonists. BMC Bioinformatics 7(Suppl 2):S18
37. Ogata M, Tsujita M, Hossain MA, Akita N, Gonzalez FJ, Staels B, Suzuki S, Fukutomi T, Kimura G, Yokoyama S (2009) On the mechanism for PPAR agonists to enhance ABCA1 gene expression. Atherosclerosis 205:413–419
38. Hanke T, Dehm F, Liening S, Popella SD, Maczewsky J, Pillong M, Kunze J, Weinigel C, Barz D, Kaiser A, Wurglics M, Lammerhofer M, Schneider G, Sautebin L, Schubert-Zsilavecz M, Werz O (2013) Aminothiazole-featured pirinixic acid derivatives as dual 5-lipoxygenase and microsomal prostaglandin E2 synthase-1 inhibitors with improved potency and efficiency in vivo. J Med Chem 56:9031–9044
39. Seo M, Inoue I, Ikeda M, Nakano T, Takahashi S, Katayama S, Komoda T (2008) Statins activate human PPARalpha promoter and increase PPARalpha mRNA expression and activation in HepG2 cells. PPAR Res 2008:316306
40. Paumelle R, Blanquart C, Briand O, Barbier O, Duhem C, Woerly G, Percevault F, Fruchart JC, Dombrowicz D, Glineur C, Staels B (2006) Acute antiinflammatory properties of statins involve peroxisome proliferator-activated receptor-alpha via inhibition of the protein kinase C signaling pathway. Circ Res 98:361–369
41. Wierzbicki AS, Mikhailidis DP, Wray R, Schacter M, Cramb R, Simpson WG, Byrne CB (2003) Statin-fibrate combination: therapy for hyperlipidemia: a review. Curr Med Res Opin 19:155–168
42. Barnett J, Viljoen A, Wierzbicki AS (2013) The need for combination drug therapies in patients with complex dyslipidemia. Curr Cardiol Rep 15:391
43. Chateauvieux S, Morceau F, Dicato M, Diederich M (2010) Molecular and therapeutic potential and toxicity of valproic acid. J Biomed Biotechnol [AU4]
44. Lampen A, Carlberg C, Nau H (2001) Peroxisome proliferator-activated receptor delta is a specific sensor for teratogenic valproic acid derivatives. Eur J Pharmacol 431:25–33
45. Ren H, Aleksunes LM, Wood C, Vallanat B, George MH, Klaassen CD, Corton JC (2010) Characterization of peroxisome proliferator-activated receptor alpha – independent effects of PPARalpha activators in the rodent liver: di-(2-ethylhexyl) phthalate also activates the constitutive-activated receptor. Toxicol Sci 113:45–59

46. Kliewer SA, Xu HE, Lambert MH, Willson TM (2001) Peroxisome proliferator-activated receptors: from genes to physiology. Recent Prog Horm Res 56:239–263

47. Jia R, Cao LP, Du JL, Wang JH, Liu YJ, Jeney G, Xu P, Yin GJ (2014) Effects of carbon tetrachloride on oxidative stress, inflammatory response and hepatocyte apoptosis in common carp (Cyprinus carpio). Aquat Toxicol 152:11–19

48. Jimenez-Lopez JM, Cederbaum AI (2005) CYP2E1-dependent oxidative stress and toxicity: role in ethanol-induced liver injury. Expert Opin Drug Metab Toxicol 1:671–685

49. Jaeschke H, Gores GJ, Cederbaum AI, Hinson JA, Pessayre D, Lemasters JJ (2002) Mechanisms of hepatotoxicity. Toxicol Sci 65:166–176

50. Yang JW, Shin JS, Lee JJ, Chang HI, Kim CW (2001) In vitro model using mouse hepatocytes for study of alcohol stress. Biosci Biotechnol Biochem 65:1528–1533

51. Kwolek-Mirek M, Zadrag-Tecza R, Bednarska S, Bartosz G (2014) Acrolein-induced oxidative stress and cell death exhibiting features of apoptosis in the yeast saccharomyces cerevisiae deficient in SOD1. Cell Biochem Biophys

52. Kujawska M, Ignatowicz E, Murias M, Ewertowska M, Mikolajczyk K, Jodynis-Liebert J (2009) Protective effect of red beetroot against carbon tetrachloride- and N-nitrosodiethylamine-induced oxidative stress in rats. J Agric Food Chem 57:2570–2575

53. Mo Q, Wang S, Seshan VE, Olshen AB, Schultz N, Sander C, Powers RS, Ladanyi M, Shen R (2013) Pattern discovery and cancer gene identification in integrated cancer genomic data. Proc Natl Acad Sci U S A 110:4245–4250

54. Zhu B, Bai R, Kennett MJ, Kang BH, Gonzalez FJ, Peters JM (2010) Chemoprevention of chemically induced skin tumorigenesis by ligand activation of peroxisome proliferator-activated receptor-beta/delta and inhibition of cyclooxygenase 2. Mol Cancer Ther 9:3267–3277

55. Dhir A, Naidu PS, Kulkarni SK (2007) Neuroprotective effect of nimesulide, a preferential COX-2 inhibitor, against pentylenetetrazol (PTZ)-induced chemical kindling and associated biochemical parameters in mice. Seizure 16:691–697

56. Ghio L, Cervetti A, Respino M, Belvederi Murri M, Amore M (2014) Management and treatment of gamma butyrolactone withdrawal syndrome: a case report and review. J Psychiatr Pract 20:294–300

Author Queries

Chapter No.: 15 0002663029

Queries	Details Required	Author's Response
AU1	Please check whether the affiliations are appropriate as typeset.	
AU2	Please check whether the edits made to Tables 1 and 2 are appropriate.	
AU3	Note that Figs. 5 and 6 text are too small, and we could not able to increase the font size. Kindly provide us better version images.	
AU4	Please provide complete bibliographic details for Refs. [43] and [51].	

Chapter 16

In Silico Study of In Vitro GPCR Assays by QSAR Modeling

Kamel Mansouri and Richard S. Judson

Abstract

The US EPA's ToxCast program is screening thousands of chemicals of environmental interest in hundreds of in vitro high-throughput screening (HTS) assays. One goal is to prioritize chemicals for more detailed analyses based on activity in assays that target molecular initiating events (MIEs) of adverse outcome pathways (AOPs). However, the chemical space of interest for environmental exposure is much wider than ToxCast's chemical library. In silico methods such as quantitative structure-activity relationships (QSARs) are proven and cost-effective approaches to predict biological activity for untested chemicals. However, empirical data is needed to build and validate QSARs. ToxCast has developed datasets for about 2000 chemicals ideal for training and testing QSAR models. The overall goal of the present work was to develop QSAR models to fill the data gaps in larger environmental chemical lists. The specific aim of the current work was to build QSAR models for 18 G-protein-coupled receptor (GPCR) assays, part of the aminergic family. Two QSAR modeling strategies were adopted: classification models were developed to separate chemicals into active/non-active classes, and then regression models were built to predict the potency values of the bioassays for the active chemicals. Multiple software programs were used to calculate constitutional, topological, and substructural molecular descriptors from two-dimensional (2D) chemical structures. Model-fitting methods included PLSDA (partial least square discriminant analysis), SVMs (support vector machines), kNNs (k-nearest neighbors), and PLSs (partial least squares). Genetic algorithms (GAs) were applied as a variable selection technique to select the most predictive molecular descriptors for each assay. N-fold cross-validation (CV) coupled with multi-criteria decision-making fitting criteria was used to evaluate the models. Finally, the models were applied to make predictions within the established chemical space limits. The most accurate model was for the bovine nonselective dopamine receptor (bDR_NS) GPCR assay, for which the classification balanced accuracy reached 0.96 in fitting and 0.95 in fivefold CV, with only two latent variables. These results demonstrate the accuracy of QSAR models to predict the biological activity of chemicals specifically for each one of the studied assays.

Key words QSAR, GPCR, ToxCast, Toxicity, Machine learning

1 Introduction

Thousands of manufactured chemicals find their way to the environment, leading to the potential for exposure to humans and wildlife species. For most of these environmental chemicals, very limited toxicity information is available [1–6]. Time, costs, animal welfare issues related to traditional toxicology studies, and lack of regulatory authority are the main causes of this data gap.

Emilio. Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology 1425 vol. 1425,
DOI 10.1007/978-1-4939-3609-0_16, © Springer Science+Business Media New York 2016

For the US Environmental Protection Agency (EPA) and other regulatory agencies, there is a pressing need to develop new methods capable of quickly evaluating large numbers of environmental chemicals for potential toxicity at reasonable costs [7]. This need is being partially addressed by the use of high-throughput screening (HTS) approaches that have been developed by the pharmaceutical industry as drug discovery screening tools [8, 9]. Over the past decade, HTS has gained popularity as an adjunct to traditional toxicology testing methods. Since 2007, the EPA has been evaluating this approach through its ToxCast program [7, 10, 11]. ToxCast is being implemented in a phased approach and based on the fundamental hypothesis that toxicity is driven by interactions between chemicals and biomolecular targets such as receptors, ion channels, and kinases and on the capacity of in vitro data to reliably predict in vivo toxicity [11]. The ToxCast program uses in vitro biochemical assays to build large collections of toxicity data on environmental chemicals with potential human exposure, including pesticides, cosmetics, pharmaceuticals, and industrial chemicals [7]. The relevance of these classes to the environmental toxicity community as well as the high number of tested chemicals differentiates this program from any previous such efforts. Through its two first phases, 1063 chemicals were tested in a set of ~200 assays. The technologies used in these assays include cell-free systems, cell lines and primary cells, complex culture systems, and small model organisms [11–17].

From these data, several models have been developed that predict in vivo effects from the in vitro HTS data [18–22]. However, there is a long-acknowledged need to screen tens of thousands of chemicals for their potential toxicity in a fast and cost-effective way [23]. Therefore, the goals of the multiyear, multimillion dollar ToxCast program are not only identifying in vitro assays that can reliably indicate alterations in biological processes of relevance to in vivo toxicity but also to develop computational models based on multiple assays along with chemical properties to achieve higher predictive accuracy than single assays or molecular descriptors alone [24] and to combine in vitro bioassay-based predictive toxicity signatures with in silico models to allow prioritization of very large numbers of environmental chemicals for more detailed testing.

The use of in silico approaches for virtual screening and data gap filling is growing within the scientific community [3, 25]. Quantitative structure-activity relationships (QSARs) are recognized alternatives to empirical testing because of their ability to predict relevant toxicological and environmental endpoints in a rapid and cost-effective way [26, 27]. The conceptual basis of these modeling techniques is the congenericity principle, which is the hypothesis that similar structures are expected to exhibit similar biological behavior [28]. This leads to the possibility to predict biological activity of new chemicals based on existing experimental data for

structurally similar chemicals. Several guidance documents to build QSAR models have been published in the literature [3, 29, 30].

To demonstrate the utility of the ToxCast HTS data for constructing QSAR models, we focused on assays that measure chemical binding to G-protein-coupled receptors (GPCRs). The GPCR assays are a subset of cell-free HTS assays run in ToxCast, summarized by Sipes et al. These authors performed a cluster analysis of the cell-free assays and showed that the cluster of assays with the highest number of actives was a family of 18 aminergic GPCRs [17], which makes them interesting to model. To model these 18 assays, two QSAR strategies were applied. First we used all tested chemicals to build binary classification models; then regression models were applied on the active chemicals to estimate their activity concentration (defined as AC50 or concentration at which 50 % of maximal activity was seen). 2D chemical structures were curated and prepared for modeling, after which several classes of molecular descriptors were calculated, including constitutional, topological, and substructural descriptors. To pick the appropriate and most information-rich descriptors, genetic algorithms (GAs) were applied as a variable selection technique. Several different methods such as PLSDA, SVM, kNN, and PLS were used to fit the models for the 18 GPCR assays. All models were validated in fivefold cross-validation, and the applied methods were compared to select the best performing models to be used for the prioritization of untested chemicals. These chemicals represent a large fraction of those to which humans may be exposed through their inclusion in manufactured products.

2 Materials

2.1 ToxCast GPCR Assays

The cell-free assays in ToxCast, described by Sipes et al. [17], consist of 331 assay endpoints largely testing the potential of chemicals to bind receptors or to inhibit enzymatic activity. These assays included 77 GPCRs, 32 of which are in the aminergic class [17]. A total of ~1000 chemicals were tested in these assays and included in an unsupervised hierarchical cluster analysis. The aminergic GPCR category was associated with the highest number of active chemicals, and especially a cluster of 18 aminergic assays, listed in Table 1, that was considered for this study. For more details about the assays and chemicals, as well as access to the data used, see Sipes et al. 2013 [17] and the ToxCast dashboard [31].

2.2 Training Set

The chemicals tested in these assays include marketed and failed pharmaceuticals, air pollutants, antimicrobials, pesticides, and food additives. These chemicals were selected based on several criteria defined by EPA and other federal agencies (e.g., National Toxicology Program/National Institutes for Environmental

Table 1
Listing of 18 cell-free assays from the aminergic GPCR family

Assay	Gene symbol	Target name
hM1 to hM5 (five assays)	CHRM	Cholinergic receptor, muscarinic 1–5
gMPeripheral_NonSelective	M1	Muscarinic receptor peripheral
hAdrb2	ADRB2	Adrenergic receptor, beta-2, surface
bDR_NonSelective	DRD1	Dopamine receptor D1
h5HT2A	HTR2A	5-Hydroxytryptamine (serotonin) receptor 2A
rAdra1A,B	Adra1a,b	Adrenergic receptor, alpha-1A-B
rAdra1_NonSelective	Adra1a	Adrenergic receptor, alpha-1A
hH1	HRH1	Histamine receptor H1
gH2	Hrh2	Guinea pig histamine receptor H2
rAdra2_NonSelective	Adra2a	Adrenergic receptor, alpha-2A
hAdra2A	ADRA2A	Adrenergic receptor, alpha-2A
rmAdra2B	Adra2b	Adrenergic receptor, alpha-2B

Health Sciences, National Center for Advancing Translational Sciences/National Institutes of Health, US Food and Drug Administration) as well as international organizations such as the Organisation for Economic Co-operation and Development.

For this study, only the 1005 compounds tested for all 18 endpoints were considered in the training set so that all models could be calibrated on the same set of chemicals. After selecting the best QSAR model to be applied, missing values corresponding to the non-tested ToxCast chemicals were filled with predictions.

Assay-assay unsupervised hierarchical clustering of the 18 assays on the training set data (log-AC50 values) was performed using Euclidean distance as the similarity metric and Ward's linkage method for assembling clusters. The clustering dendrogram applied to a heatmap of the bioactivity of training chemicals shows two large clusters of the most similar GPCR assays, where one of them represents the cholinergic receptor group of five first assays from the left in Fig. 1: hM1 to hM5.

2.3 Standardization and Curation of Chemical Structures

When collected from different public sources, chemical structures usually contain duplicates and inconsistencies in the molecular representations which could lead to inaccuracies in modeling and the predictions of QSARs. Thus, a cleaning and standardization procedure is needed to prepare a set of unique QSAR-ready structures. A curation workflow was designed to process all chemical structures using the free and open-source data-mining

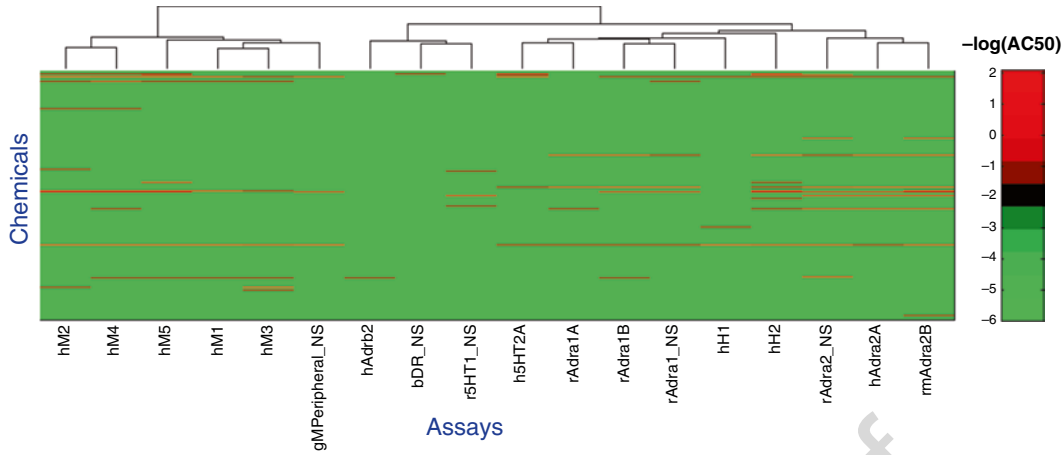


Fig. 1 Clustering of the 18 aminergic GPCR assays on the training set data

environment KNIME [32]. The workflow performed the series of steps described below [33]:

1. The original files containing structures in different formats were parsed and checked for valence imbalances relative to a set of rules and for the integrity of the required structural information to render the molecules. Invalid entries were corrected, if possible, or compared to structures retrieved from online databases for consistency using web services [34, 35] and removed if ambiguous.
2. A check was applied to remove inorganic compounds.
3. The structures were desalted, and inorganic counterions were removed.
4. A series of transformations was applied on the structures to standardize tautomers to unique representations (e.g., nitro zwitterionic form and azide mesomers, keto-enol tautomers, enamine-imine tautomers, ynol-ketene, and other conversions) [36–38].
5. Charged structures were neutralized, when possible, and then stereochemistry information was removed.
6. Explicit hydrogen atoms were added, and structures were aromatized according to Hückel's rules implemented in KNIME [32].
7. The duplicates were removed using standard InChI codes, because these are unequivocal identifiers.
8. A final filter was applied to remove chemicals containing metals, which often cause problems in molecular descriptor calculations.

After the structure standardization procedure and duplicate removal of the ToxCast data, we obtained a final set of 1005 QSAR-ready structures as a training set.

2.4 Descriptor Calculation and Variable Selection

The previously curated molecular structures were used to calculate molecular descriptors using the free and open-source software PaDEL and the commercial proprietary toolkit MOE [39, 40]. In PaDEL only 2D descriptors were selected. The use of 3D molecular structures could add valuable chemical information about the molecules. Thus, MOE descriptors were calculated after an energy minimization and geometry optimization of the 3D structures. However, there is a risk that the use of 3D descriptors can affect the predictability of the models on new molecules because the difference between conformers can lead to different 3D descriptor values, especially with very flexible molecules.

A total number of 1022 molecular descriptors were calculated including constitutional, topological, functional group counts; fragmental, atom-type E-state indices; and calculated physicochemical properties. In order to reduce collinearity among descriptors, a correlation filter with a threshold of 0.96 was applied. For each pair of descriptors with a correlation coefficient higher than the threshold, the one showing the largest pair correlation with all the other descriptors was excluded. Then, descriptors with constant, near constant, or at least one missing value were removed. The remaining reduced set consisted of 470 descriptors used for the subsequent modeling analysis.

Genetic algorithms (GAs) were then applied to find the optimal subset of molecular descriptors [41]. GAs start from an initial random population of chromosomes, which are binary vectors representing the presence or absence of molecular descriptors. An evolutionary process is simulated to optimize a defined fitness function, and new chromosomes are obtained by coupling the chromosomes of the initial population with genetic operations (crossover and mutation). This process was repeated 100 times for each one of the 100 runs with 0.01 probability of mutation and 0.5 probability of crossover on 30 chromosomes. The goodness of fit function to optimize the models was calculated in cross-validation. The final set of descriptors was picked based on the frequency of selection during the 100 GA runs.

3 Methods

3.1 Categorical Models

Three classification methods were applied in order to estimate the best relationship between chemical information, encoded in molecular descriptors, and the modeled activity of chemicals: k-nearest neighbors (kNNs) [42], partial least square discriminant analysis (PLSDA) [43, 44], and support vector machines (SVMs) [45]. The application of these methods, based on different mathematical strategies, aimed to better explore the chemical space and balance potential biases inherent in each single modeling algorithm. All calculations were carried out in MATLAB 8.2 (glnxa64) [46].

The kNN classification rule is conceptually quite simple: each predicted chemical is classified according to the classes of the k-closest chemicals, which means it is classified according to the majority of its k-nearest neighbors in the selected descriptor space [42]. In this work, the Euclidean metric was used to measure distances between molecules. The *k* value giving the lowest classification error in cross-validation was selected as the optimal one.

PLSDA is a classification technique that inherits the properties of partial least square (PLS) regression with the discrimination power of a classification technique [43, 44]. It finds fundamental relations between the matrix of descriptors and the class vector by calculating latent variables (LVs), which are orthogonal linear combinations of the original variables. PLSDA models optimize in cross-validation to find a compromise between the classification performance and the number of selected LVs.

SVM is a method that defines a decision boundary that optimally separates two classes by maximizing the distance between them [45, 47]. The decision boundary can be described as a hyperplane that is expressed in terms of a linear combination of functions parameterized by support vectors, which consist in a subset of training molecules. SVM algorithms search for the support vectors that give the best separating hyperplane using a kernel function. During optimization, SVM searches the decision boundary with maximal margin among all possible hyperplanes, where the margin can be intended as the distance between the hyperplane and the closest point for both classes. This procedure was carried out by means of a kernel based on a radial basis function; the learning level is governed by a cost (c) parameter. SVMs were calibrated using the library LIBSVM3.1 implemented in C [48].

3.2 Continuous Models

PLS is a powerful statistical method applied in chemometrics and other fields of scientific research [44]. A major advantage of this method is its ability to overcome the problem of singularity in a transformed matrix when the number of columns (variables) is larger than the number of rows (samples). PLS also compensates for the collinearity of the variables. This latter problem is solved by decomposing the descriptor matrix into orthogonal scores and loadings. Then, the modeled biological activity is correlated to the first columns of the scores instead of the original variables. In this way, PLS includes information from both the variables and the observed response in the calculation of the scores and loadings and aims to explain the maximum variance in the original variables as well as in the observed biological activity of the training samples. There are several implementations of PLS algorithms in the literature that give similar results, especially in the case of a single vector response. These may differ slightly when dealing with multivariate responses [49, 50].

3.3 Evaluation and Validation Criteria

Due to the very low number of active chemicals, the initial set was not divided into a training and a test set for external validation. However, not all of the chemicals within the list were used to select molecular descriptors and to build the models. During model optimization and descriptor selection, a cross-validation procedure with five groups was performed. Thus, this procedure is similar to constantly dividing the initial set into training and test sets, containing 80 and 20 % of the total number of chemicals, respectively. The selection was performed maintaining the class proportions, that is, the number of active test chemicals was proportional to the number of active training chemicals.

The classification models were evaluated on the basis of sensitivity (Sn) and specificity (Sp), which are the ability to correctly predict active and inactive chemicals, respectively. In particular, Sn and Sp were calculated using the number of true negatives, true positives, false negatives, and false positives. In addition, the balanced accuracy (BA) was calculated as the average of Sn and Sp. These indices were used in order to better estimate classification performance in the presence of a dataset with an unequal number of samples in each class. In this study BA, specificity, and sensitivity are expressed as ratios and not as percentages. The quality of regression models was evaluated using two groups of statistical indices:

- The goodness of fit parameters measuring the fitting ability. These indices are used to measure the degree to which the model is able to explain the variance contained in the training set. The coefficient of determination R^2 is one of the most used parameters. It is the square multiple correlation coefficient given by

$$R^2 = \frac{\sum_{i=1}^{n} \left(\check{y}_i - y_i \right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2}$$

- where $\hat{y}i$ is the estimated response and $\bar{y}$ is the average observed response over the n training compounds.

- The second parameter used is the root-mean-square error (*RMSE*) calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \check{y}_i \right)^2}{n}}$$

- The goodness of prediction parameters measures the true predictive ability of a model; these are related to the reliability of prediction in the validation step. These parameters are used

in the validation step, the most important being the predictive squared correlation coefficient Q^2. Different ways of calculating this parameter are available in the literature [30, 51]. In this work, the following formula was considered:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_{\mathrm{EXT}}} \left(y_i - \check{y}_i\right)^2 / n_{\mathrm{EXT}}}{\sum_{n_{\mathrm{TR}}}^{i=1} \left(y_i - \bar{y}\right)^2 / n_{\mathrm{TR}}}$$

- where n_{EXT} is the number of test compounds and n_{TR} is the number of training compounds.
- The second parameter commonly used is the root-mean-square error in prediction (*RMSEP*) calculated as follows:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{n}^{i=1} \left(y_i - \check{y}_i\right)^2}{n_{\mathrm{EXT}}}}$$

4 Results

4.1 Structure-Activity Relationship Analysis

A preliminary analysis of the dataset was performed using the self-organizing map (SOM) technique implemented in the Kohonen and CPANN Toolbox [52, 53]. Since all 18 GPCR assays are biologically similar and part of the same cluster [17], all GPRC active chemicals were considered in a supervised learning analysis to generate a categorical SOM based on the complete set of 470 descriptors. The resulting map (Fig. 2) demonstrates that there is a clear structure-activity relationship between the chemical information encoded in the descriptors and the observed biological activity in the in vitro assay results. The map's calculated BA in classifying actives and inactives was high in both calibration and fivefold cross-validation, with values of 0.82 and 0.67, respectively. Consequently, this demonstrated that activity in these assays has a strong structural component and that building more detailed QSAR models should lead to higher prediction scores.

4.2 Comparison of Modeling Approaches

In order to select an appropriate modeling method to be applied on the training set to build the models, three QSAR approaches were compared. GA coupled with PLSDA, kNN, and SVM was used to select the optimal subset from the list of 470 descriptors. These methods were applied first on assay hH1, which is associated with an average number of actives (37 actives out of a total of 1005 chemicals); second, on the combination of the five assays from the cholinergic receptor group shown as similar by the cluster analysis (Fig. 1) (a chemical active in any of the five assays is

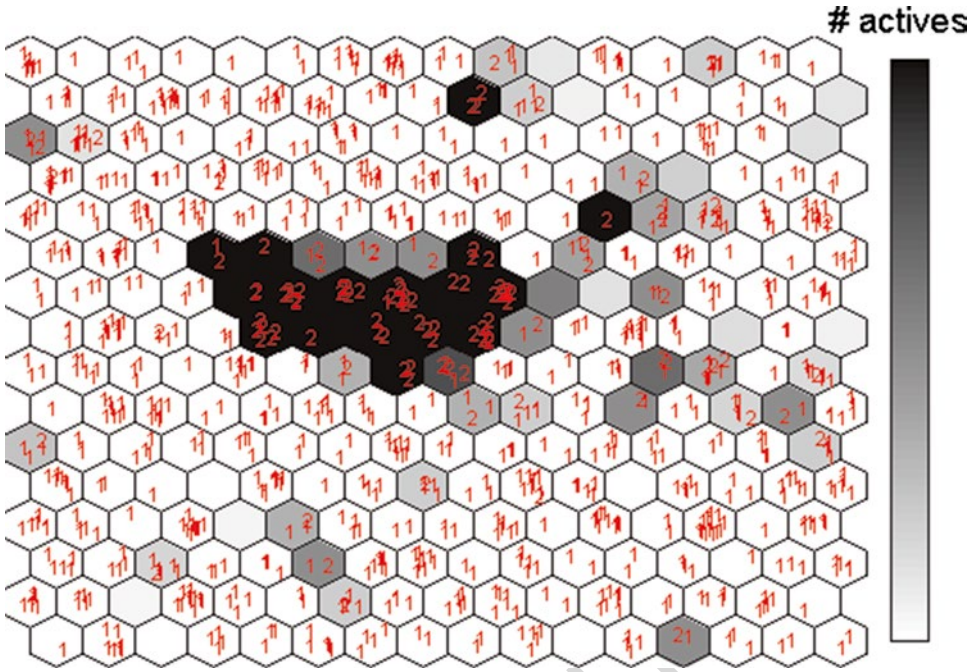


Fig. 2 Supervised SOM of all training set chemicals on the set of 470 descriptors. All active chemicals in the 18 assays were considered. The nodes are in gray scale indicating the number of active chemicals present in each node. The darker the node the more active chemicals it contains. In the nodes, "1" and "2" symbolize inactive and chemicals, respectively. The most similar chemicals situated in the same location on the map have overlapping symbols

t2.1 **Table 2**
t2.2 **Statistics of the PLSDA test models. LVs: number of latent variables for PLS, BA: balanced accuracy**

t2.3 Endpoint (number of actives)	Descriptors	LVs	BA fitting	BA five-fold CV
t2.4 hH1 (37)	26	5	0.92	0.91
t2.5 hM1-5 (76)	15	3	0.84	0.85
t2.6 All (115)	20	5	0.84	0.82

considered to be active); and third, on the combination of all 18 assays for a maximum number of active assays (a chemical active in any of the 18 assays is considered to be active). The best models of each approach were selected by maximizing the BA and minimizing the number of descriptors. The results are summarized in Tables 2, 3, and 4.

All three methods showed high performance on the three modeled datasets. However, PLSDA showed the highest BA in cross-validation. Both PLSDA and kNN are highly stable as they are associated with similar BAs in fitting and cross-validation. SVM, on the other hand, seems to overfit the models even with a

t3.1 **Table 3**
t3.2 **Statistics of the kNN test models. k: number of nearest neighbors, BA:**
t3.3 **balanced accuracy**

t3.4 Endpoint (number of actives)	Descriptors	*k*	BA Fitting	BA fivefold CV
t3.5 hH1 (37)	27	1	0.81	0.81
t3.6 hM1-5 (76)	19	4	0.77	0.79
t3.7 All (115)	22	1	0.77	0.78

t4.1 **Table 4**
t4.2 **Statistics of the SVM test models. c: cost of SVM fitting, BA: balanced**
t4.3 **accuracy**

t4.4 Endpoint (number of actives)	Descriptors	c	BA fitting	BA fivefold CV
t4.5 hH1 (37)	30	5	0.88	0.68
t4.6 hM1-5 (76)	12	10	0.95	0.77
t4.7 All (115)	16	10	0.94	0.76

relatively low number of descriptors. This could be due to a high number of super vectors employed. PLSDA and kNN demonstrate higher performances on hH1 compared to hM1-5 and the full list of 18 assays. This decrease in BA is probably due to the increased heterogeneity of the data after combining the assays. Thus, the models are expected to perform better on the single assays.

4.3 Selected Models

PLSDA was selected as the modeling approach in order to build individual QSAR models for the 18 GPCR assays. GA was used to pick the minimum set of the most information-rich descriptors for each of the assays. In addition, the modeling procedure aimed at building models minimizing the number of LVs and keeping a balance between Sn and Sp in both fitting and cross-validation to avoid overfitting and to maximize the predictive ability of the models.

Table 5 summarizes the 18 PLSDA models for the individual 18 GPCR assays. All models showed high performance with a high stability. The highest BA reached 0.96 in fitting and 0.95 in cross-validation for the bovine nonselective dopamine receptor bDR_NS. Even the model with the lowest BA shows good performance and stability (BA = 0.87 human muscarinic hM2). However, note that Sn is slightly higher than Sp which means that false positives can be expected in the model predictions. This behavior is, nevertheless, not contradictory to the general aim of this in silico study which is prioritizing new chemicals for testing.

In order to have a precise estimate of the potency of the active chemicals in terms of log-AC50 values (concentrations at which

t5.1 **Table 5**

t5.2 **Selected categorical and continuous models for the 18 GPCR assays. LVs: latents vairables for PLS. BA: balanced accuracy, Sn: sensitivity, SP:**

t5.3 **specificity, RMSE: root-mean-square error**

		Categorical models: PLSDA								Continuous models: PLS					
				Fitting			Five-fold CV					Fitting		five-fold CV	
t5.4 t5.5 Endpoint	Actives (/1005)	Descriptors	LVs	BA	Sn	Sp	BA	Sn	Sp	Descriptors	LVs	R^2	RMSE	Q^2	RMSE
t5.6 hH1	37	35	3	0.89	0.89	0.89	0.90	0.92	0.89	16	3	0.83	3.35	0.76	3.96
t5.7 gH2	54	21	3	0.93	0.96	0.89	0.93	0.96	0.89	18	3	0.65	4.76	0.52	5.56
t5.8 hM1	37	15	2	0.92	0.95	0.89	0.92	0.95	0.89	17	2	0.89	2.41	0.83	2.83
t5.9 hM2	50	21	2	0.87	0.82	0.92	0.87	0.82	0.92	27	3	0.78	3.10	0.63	4.03
t5.10 hM3	38	13	2	0.89	0.95	0.84	0.89	0.95	0.83	25	3	0.92	1.76	0.78	2.41
t5.11 hM4	50	24	2	0.89	0.88	0.89	0.90	0.90	0.89	18	2	0.86	2.46	0.81	2.75
t5.12 hM5	56	16	3	0.87	0.88	0.87	0.89	0.91	0.87	15	2	0.76	3.11	0.74	3.24
t5.13 gMPeripheral_NS	39	20	3	0.93	0.95	0.90	0.93	0.95	0.90	11	2	0.77	3.09	0.62	3.94
t5.14 rAdra1A	35	23	2	0.93	0.94	0.91	0.93	0.94	0.91	18	3	0.78	3.01	0.67	3.70
t5.15 rAdra1B	38	30	3	0.91	0.92	0.90	0.89	0.89	0.89	29	2	0.85	3.18	0.71	3.72
t5.16 rAdra1_NS	34	26	2	0.91	0.91	0.90	0.91	0.91	0.90	12	1	0.81	3.85	0.71	4.75
t5.17 rAdra2_NS	41	30	2	0.89	0.90	0.87	0.89	0.90	0.88	34	2	0.80	3.41	0.63	4.21
t5.18 hAdra2A	45	30	2	0.87	0.91	0.84	0.88	0.91	0.84	9	2	0.72	4.70	0.48	6.43
t5.19 rmAdra2B	51	25	2	0.87	0.84	0.90	0.88	0.86	0.89	17	3	0.76	2.66	0.63	3.27
t5.20 hAdrb2	26	31	2	0.90	0.92	0.87	0.90	0.92	0.87	23	3	0.87	2.83	0.77	3.73
t5.21 r5HT1_NS	17	21	2	0.90	0.94	0.86	0.90	0.94	0.86	8	3	0.98	1.66	0.96	2.57
t5.22 h5HT2A	25	24	2	0.95	1.00	0.90	0.95	1.00	0.90	14	2	0.88	2.52	0.74	3.31
t5.23 bDR_NS	16	28	2	0.96	1.00	0.91	0.95	1.00	0.91	18	3	0.98	0.66	0.92	1.48

activity is seen), regression models were built based on only the active compounds of each assay. For this purpose, GA was coupled with PLS to calibrate a continuous model corresponding to each of the 18 GPCR assays. Descriptors were selected from the same initial set used for categorical models. The resulting models are summarized in the same Table 5. Most models show high performance with relatively low numbers of descriptors and LVs. However, R^2 in fitting and Q^2 in cross-validation are not balanced for all models. This can be explained by the large difference in potency between the weak and strong actives that could present different chemical features that are not always easy to capture by machine learning algorithms when there are low number of chemicals used in the training process.

4.4 Molecular Descriptors

The machine learning and variable selection algorithms were applied independently to build the 18 assays. However, certain descriptors were selected in more than a single model. This redundancy between the models can be explained by the similarity between the assays since they belong to the same aminergic GPCR family. The selection of these descriptors highlights the importance of the encoded information to this class of GPCR assays. Table 6 lists the descriptors that were included in the categorical PLSDA models more than five times. Most of these descriptors are describing the electronic profile of the chemicals such as the electronegativity (e.g., AM1_LUMO and PEOE _VSA14) as well as certain structural features and functional groups (e.g., nBase and MDEC-22).

Table 6
The most selected descriptors in categorical models

Descriptor	Description	Software	Number of models
nBase	Number of basic groups defined by a list of SMARTS	PaDEL	12
GCUT_SLOGP_0	Eigenvalues of a graph adjacency matrix (GCUT) descriptor weighted by Crippen logP (SLOGP_0)	MOE	8
BCUT_SLOGP_0	Eigenvalues of the burden matrix (BCUT) weighted by Crippen logP	MOE	8
GCUT_PEOE_0	GCUT descriptor weighted by partial equalization of orbital electronegativity (PEOE) charges	MOE	7
AM1_LUMO	The energy (eV) of the lowest unoccupied molecular orbital calculated using the AM1 Hamiltonian	MOE	7
maxssCH2	Maximum atom-type E-state: -CH2-	PaDEL	7
MDEC-22	Molecular distance edge between all secondary carbons	PaDEL	7
PEOE _VSA14	PEOE descriptor based on van der Waals surface area (VSA)	MOE	6

Table 7 lists the descriptors that were included in the continuous PLS models more than five times. Most of these descriptors were also included in the categorical models. These descriptors encoded similar chemical information as in the categorical models: electronic profile described in terms of electronegativity (AM1_LUMO and PEOE _VSA14) and polarizability (SMR_VSA3) as well as substructural features (e.g., nBase and MDEC-22). In both types of models, nBase was selected the highest number of times which is an indication of the importance of the presence of these functional groups that can be captured by the following SMARTS notations: "[$([NH2]-[CX4])]," "[$([NH](-[CX4])-[CX4])]," "[$(N(-[CX4])(-[CX4])-[CX4])]," "[$([*;+;!$(*~[*;-])])]," "[$(N=C-N)]," and "[$(N-C=N)]." In Table 7, the frequency of selection of the overlapping descriptors between categorical and continuous models was higher than in Table 6. Also the total number of descriptors selected more than five times was higher in the case of continuous models. This could be explained by the promiscuity of the active chemicals among the studied group of in vitro assays.

Table 7
The most selected descriptors in continuous models

Descriptor	Description	Software	Number of models
nBase	Number of basic groups defined by a list of SMARTS	PaDEL	15
GCUT_PEOE_0	GCUT descriptor weighted by partial equalization of orbital electronegativity (PEOE) charges	MOE	13
BCUT_SLOGP_0	BCUT descriptor weighted by Crippen SlogP	MOE	10
PEOE_VSA14	PEOE descriptor based on van der Waals surface area (VSA)	MOE	9
GCUT_SLOGP_0	GCUT descriptor weighted by Crippen SlogP	MOE	8
AM1_LUMO	The energy (eV) of the lowest unoccupied molecular orbital calculated using the AM1 Hamiltonian	MOE	8
minaasC	Minimum atom-type E-state: :C:-	PaDEL	7
maxssCH2	Maximum atom-type E-state: -CH2-	PaDEL	7
MDEC-22	Molecular distance edge between all secondary carbons	PaDEL	7
SMR_VSA3	Atomic contribution to VSA using molar refractivity to capture polarizability	MOE	6
ATSc5	Centered Broto-Moreau autocorrelation – lag 5/ weighted by charges	PaDEL	6
maxaasC	Maximum atom-type E-state: :C:-	PaDEL	6
hmin	Minimum H E-state	PaDEL	6
XLogP	XlogP	PaDEL	6

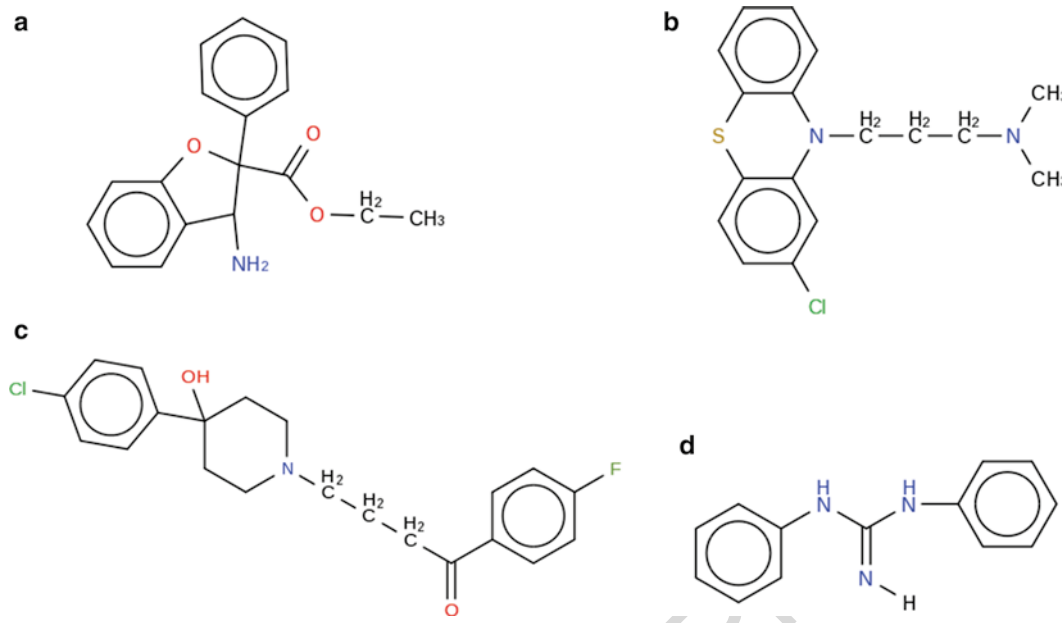


Fig. 3 Example structures from the training set. (**a**) Ethyl (2R,3S)-3-amino-2-phenyl-2,3-dihydro-1-benzofuran-2-carboxylate. Active in 18/18 assays, mean AC50 = 4.25 μM. (**b**) Chlorpromazine hydrochloride. Active in 17/18 assays, mean AC50 = 1.77 μM. (**c**) Haloperidol. Active in 11/18 assays, mean AC50 = 5.58 μM. (**d**) 1,3-Diphenylguanidine. Active in 5/18 assays, mean AC50 = 12.34 μM

4.5 Chemical Promiscuity

The ToxCast dataset was demonstrated to be containing several promiscuous chemicals from different chemical categories [17]. Examples of training set chemicals of different degrees of promiscuity (18/18, 17/18, 11/18, 5/18) are rendered in Fig. 3. This figure shows the number of assays in which the chemicals are actives, the average AC50 values, and the chemical structures and names. With some exceptions, the general trend we noticed is that the promiscuity increases with the potency of the chemicals; chlorpromazine hydrochloride, active in 17/18 assays, is more potent than 1,3-diphenylguanidine, which is active in only 5/18 assays. The common feature among these chemicals seems to be the two aromatic rings. The two most potent and promiscuous chemicals have in common a long chain in addition to the two phenyl groups.

The categorical models were applied to an additional 778 ToxCast chemicals that were not tested in the 18 studied assays, but which are of environmental/toxicological interest. Predicted actives using the QSAR models could then be candidates for in vitro testing. Additionally, AC50 values of actives were estimated using the continuous models. Figure 4 shows some chemicals of different degrees of promiscuity (18/18, 15/18, 10/18, 4/18) in the predicted ToxCast chemicals. Similarly to what was noticed in the tested ToxCast chemicals used as training set, the predictions also show a correlation between potency and promiscuity of the chemicals;

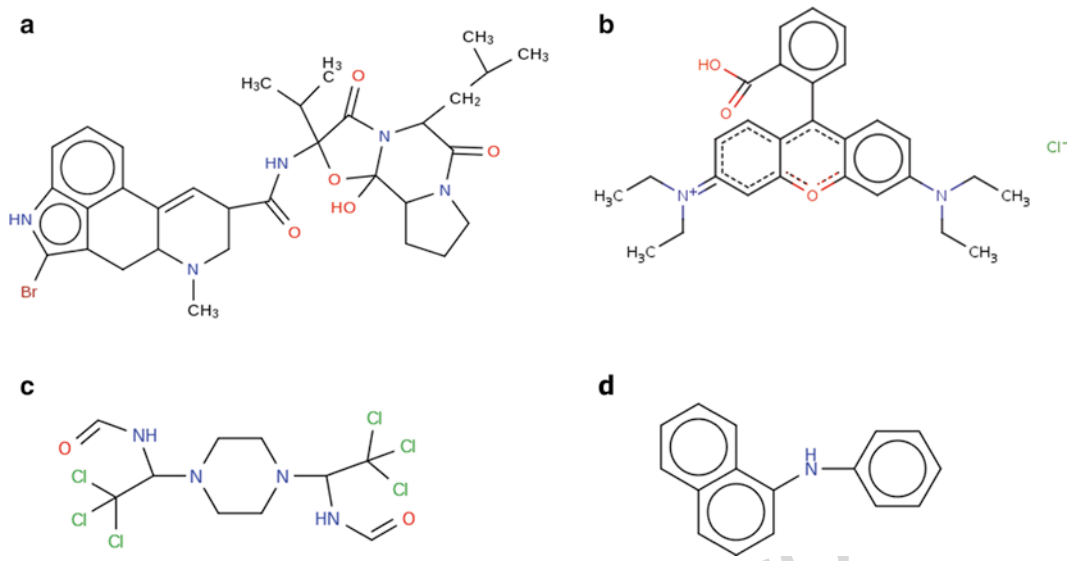


Fig. 4 Example structures of predicted chemicals. (**a**) Bromocriptine mesylate. Active in 18/18, mean AC50 = 7.89 μM. (**b**) Rhodamine B. Active in 15/18, mean AC50 = 10.42 μM. (**c**) Triforine. Active in 10/18, mean AC50 = 11.19 μM. (**d**) *N*-Phenyl-1-naphthylamine. Active in 4/18, mean AC50 = 16.56 μM

bromocriptine mesylate active in 18/18 assays is more potent than *N*-phenyl-1-naphthylamine active in 4/18 assays. Except for triforine, these promiscuous chemicals also seem to have in common the presence of aromatic rings. The list of the most promiscuous predicted chemicals is given in Table 8. Some chemicals from the 778 list were subsequently tested and confirmed to be actives as shown in Table 8 for hydramethylnon, emamectin benzoate, and chlorhexidine diacetate [31].

5 Conclusions

The ToxCast program was initiated by the EPA to develop the methods that forecast toxicity of chemicals based on their bioactivity profiles and hence to set priorities for further testing of environmental contaminants [11]. The number of chemicals being simultaneously analyzed using a large number of HTS technologies represents ToxCast's major departure from traditional toxicology testing. Hence, the central foundational element of the ToxCast effort is its chemical library. The current work demonstrates one approach to use the ToxCast assay data to help evaluate a large number of chemicals by building specific QSAR models for each of the studied 18 GPCR assays based on approximately 1000 ToxCast chemicals.

This QSAR modeling study focused on GPCRs, part of the largest class of human proteins that regulate vital biological and

Table 8
The most promiscuous predicted chemicals

CASRN	Chemical name	Number of active assays	Subsequently tested in the 18 assays
22260-51-1	Bromocriptine mesylate. Dopamine receptor agonist drug (Parlodel)	18	Non-tested
67485-29-4	Hydramethylnon	18	Tested in 1/18, active in 1/18 (hM1)
155569-91-8	Emamectin benzoate	17	Tested in 11/18, active in 11/18 assays
2353-45-9	FD&C Green no. 3	17	Non-tested
2390-60-5	Basic blue 7	17	Non-tested
27090-63-7	*N,N,N′,N′*-tetrabutyl-1,6-hexanediamine	16	Non-tested
3734-33-6	Denatonium benzoate	16	Non-tested
56-95-1	Chlorhexidine diacetate	16	Tested in 1/18, active in 1/18 (hM1)
67564-91-4	Fenpropimorph	16	Non-tested
95-38-5	1H-Imidazole-1-ethanol, 2-(8-heptadecenyl)-4,5-dihydro-	16	Non-tested
10081-67-1	4-(2-Phenylpropan-2-yl)-*N*-[4-(2-phenylpropan-2-yl)phenyl]aniline	15	Non-tested
25155-18-4	Methylbenzethonium chloride	15	Non-tested
5137-55-3	*N*-methyl-*N,N*-dioctyloctan-1-aminium chloride	15	Non-tested
63449-41-2	C8-18-Alkydimethylbenzyl ammonium chlorides	15	Non-tested
68959-20-6	*N,N*-didecyl-*N*-methyl-3-(trimethoxysilyl) propanaminium chloride	15	Non-tested
8001-54-5	Benzalkonium chloride	15	Non-tested
81-88-9	Rhodamine B	15	Non-tested
41372-20-7	Apomorphine hydrochloride hydrate	14	Non-tested
51229-78-8	3,5,7-Triaza-1-azoniatricyclo[3.3.1.13,7]decane, 1-(3-chloro-2-propenyl)-, chloride	14	Non-tested

physiological functions [54]. GPCRs are also regarded as major targets for drug discovery [55]. We built categorical and continuous models of 18 aminergic GPCR assays associated with the highest number of actives among the GPCR assays. A comparison of different QSAR methods found PLSDA to best predict in the categorical modeling procedure, followed by a similar method, PLS, for the

continuous models. GAs were coupled with these methods and used for feature selection to pick the most appropriate molecular descriptors for each model. High-accuracy classification and regression models were built using the curated data and then validated in fivefold cross-validation. In order to minimize the risk of overfitting that could affect the predictability of the models, a minimum number of descriptors and balance between fitting and validation performance were maintained as much as possible. For categorical models, the balance between Sn and Sp was taken into consideration.

The active chemicals from the training set and the chemicals predicted to be active presented similar structural features and high promiscuity due to the high similarity among the 18 assays.

This modeling procedure will be first extended on the remaining GPCR assays and then all ToxCast assays with a sufficient number of actives to carry a QSAR study and build accurate models. The built models will be used to prioritize a list of ~32 k unique chemicals called the "human exposure universe," which covers a wide range of man-made chemicals identified by the EPA as having significant potential for high exposure for humans.

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the US Environmental Protection Agency.

References

1. Egeghy PP, Judson R, Gangwal S, Mosher S, Smith D, Vail J, Cohen Hubal EA (2012) The exposure data landscape for manufactured chemicals. Sci Total Environ 414:159–166. doi:10.1016/j.scitotenv.2011.10.046
2. Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, Dellarco V, Henry T, Holderman T, Sayre P, Tan S, Carpenter T, Smith E (2009) The toxicity data landscape for environmental chemicals. Environ Health Perspect 117:685–695. doi:10.1289/ehp.0800168
3. Worth AP, Bassan A, Gallegos A, Netzeva TI, Patlewicz G, Pavan M, Tsakovska I, Vracko M (2005) The characterisation of (quantitative) structure-activity relationships: preliminary guidance. EUROPEAN COMMISSION JOINT RESEARCH CENTRE, Institute for Health and Consumer Protection Toxicology and Chemical Substances Unit Ispra (VA) Italy
4. Whitacre DM (2008) Reviews of environmental contamination and toxicology Volume 199. Springer Science & Business Media, LLC.
5. Boxall A (2009) Transformation products of synthetic chemicals in the environment. Springer, Dordrecht
6. Cronin MTD, Madden JC (2010) In silico toxicology: principles and applications. Royal Society of Chemistry, Cambridge
7. Dix DJ, Houck KA, Martin MT, Richard AM, Setzer RW, Kavlock RJ (2007) The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol Sci 95:5–12. doi:10.1093/toxsci/kfl103
8. Fliri AF, Loging WT, Thadeio PF, Volkmann RA (2005) Biological spectra analysis: linking biological activity profiles to molecular structure. Proc Natl Acad Sci U S A 102:261–266. doi:10.1073/pnas.0407790101
9. Janzen WP, Hodge CN (2006) A chemogenomic approach to discovering target-selective drugs. Chem Biol Drug Des 67:85–86. doi:10.1111/j.1747-0285.2005.00325.x
10. Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, Reif DM,

Rotroff DM, Shah I, Richard AM, Dix DJ (2010) In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project. Environ Health Perspect 118:485–492. doi:10.1289/ehp.0901392

11. Kavlock R, Chandler K, Houck K, Hunter S, Judson R, Kleinstreuer N, Knudsen T, Martin M, Padilla S, Reif D, Richard A, Rotroff D, Sipes N, Dix D (2012) Update on EPA's ToxCast program: providing high throughput decision support tools for chemical risk management. Chem Res Toxicol 25:1287–1302. doi:10.1021/tx3000939
12. Houck KA, Dix DJ, Judson RS, Kavlock RJ, Yang J, Berg EL (2009) Profiling bioactivity of the ToxCast chemical library using BioMAP primary human cell systems. J Biomol Screen 14:1054–1066. doi:10.1177/1087057109345525
13. Knight AW, Little S, Houck K, Dix D, Judson R, Richard A, McCarroll N, Akerman G, Yang C, Birrell L, Walmsley RM (2009) Evaluation of high-throughput genotoxicity assays used in profiling the US EPA ToxCast chemicals. Regul Toxicol Pharmacol RTP 55:188–199. doi:10.1016/j.yrtph.2009.07.004
14. Martin MT, Dix DJ, Judson RS, Kavlock RJ, Reif DM, Richard AM, Rotroff DM, Romanov S, Medvedev A, Poltoratskaya N, Gambarian M, Moeser M, Makarov SS, Houck KA (2010) Impact of environmental chemicals on key transcription regulators and correlation to toxicity end points within EPA's ToxCast program. Chem Res Toxicol 23:578–590. doi:10.1021/tx900325g
15. Rotroff DM, Beam AL, Dix DJ, Farmer A, Freeman KM, Houck KA, Judson RS, LeCluyse EL, Martin MT, Reif DM, Ferguson SS (2010) Xenobiotic-metabolizing enzyme and transporter gene expression in primary cultures of human hepatocytes modulated by ToxCast chemicals. J Toxicol Environ Health B Crit Rev 13:329–346. doi:10.1080/10937404.2010.483949
16. Rotroff DM, Dix DJ, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Reif DM, Richard AM, Sipes NS, Abassi YA, Jin C, Stampfl M, Judson RS (2013) Real-time growth kinetics measuring hormone mimicry for ToxCast chemicals in T-47D human Ductal carcinoma cells. Chem Res Toxicol 26:1097–1107. doi:10.1021/tx400117y
17. Sipes NS, Martin MT, Kothiya P, Reif DM, Judson RS, Richard AM, Houck KA, Dix DJ, Kavlock RJ, Knudsen TB (2013) Profiling 976 ToxCast chemicals across 331 enzymatic and receptor signaling assays. Chem Res Toxicol 26:878–895. doi:10.1021/tx400021f
18. Judson R, Houck K, Martin M, Knudsen T, Thomas RS, Sipes N, Shah I, Wambaugh J, Crofton K (2014) In vitro and modelling approaches to risk assessment from the U.S. Environmental Protection Agency ToxCast programme. Basic Clin Pharmacol Toxicol 115:69–76. doi:10.1111/bcpt.12239
19. Martin MT, Knudsen TB, Reif DM, Houck KA, Judson RS, Kavlock RJ, Dix DJ (2011) Predictive model of rat reproductive toxicity from ToxCast high throughput screening. Biol Reprod 85:327–339. doi:10.1095/biolreprod.111.090977
20. Benigni R, Bossa C, Giuliani A, Tcheremenskaia O (2010) Exploring in vitro/in vivo correlation: lessons learned from analyzing phase I results of the US EPA's ToxCast project. J Environ Sci Health Part C Environ Carcinog Ecotoxicol Rev 28:272–286. doi:10.1080/10590501.2010.525781
21. Benigni R (2013) Evaluation of the toxicity forecasting capability of EPA's ToxCast phase I data: can ToxCast in vitro assays predict carcinogenicity? J Environ Sci Health Part C Environ Carcinog Ecotoxicol Rev 31:201–212. doi:10.1080/10590501.2013.824188
22. Sipes NS, Martin MT, Reif DM, Kleinstreuer NC, Judson RS, Singh AV, Chandler KJ, Dix DJ, Kavlock RJ, Knudsen TB (2011) Predictive models of prenatal developmental toxicity from ToxCast high-throughput screening data. Toxicol Sci 124:109–127. doi:10.1093/toxsci/kfr220
23. Krewski D, Acosta D, Andersen M, Anderson H, Bailar JC, Boekelheide K, et al. (2010) Toxicity testing in the 21st century: a vision and a strategy. J Toxicol Environ Health. Part B, Critical reviews 13:51–138. doi:10.1080/10937404.2010.483176
24. Judson RS, Magpantay FM, Chickarmane V, Haskell C, Tania N, Taylor J, et al. (2015) Integrated Model of Chemical Perturbations of a Biological Pathway Using 18 In Vitro High-Throughput Screening Assays for the Estrogen Receptor. Toxicological Sciences 148:137–154; doi:10.1093/toxsci/kfv168
25. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: where have you been? where are you going to? J Med Chem 57:4977–5010. doi:10.1021/jm4004285
26. Muster W, Breidenbach A, Fischer H, Kirchner S, Müller L, Pähler A (2008) Computational

toxicology in drug development. Drug Discov Today 13:303–310. doi:10.1016/j.drudis.2007.12.007

27. Vedani A, Smiesko M (2009) In silico toxicology in drug discovery – concepts based on three-dimensional models. Altern Lab Anim ATLA 37:477–496
28. Todeschini R, Consonni V (2008) Handbook of molecular descriptors. Wiley-VCH Verlag GmbH, Weinheim, Germany
29. Dearden JC, Cronin MTD, Kaiser KLE (2009) How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). SAR QSAR Environ Res 20:241–266. doi:10.1080/10629360902949567
30. Consonni V, Ballabio D, Todeschini R (2009) Comments on the definition of the Q2 parameter for QSAR validation. J Chem Inf Model 49:1669–1678. doi:10.1021/ci900115y
31. US EPA-NCCT (2015) iCSS ToxCast Dashboard. http://actor.epa.gov/dashboard2/. Accessed 15 Jul 2015
32. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Sieb C, Thiel K, Wiswedel B (2007) KNIME: the Konstanz Information Miner. Stud Classif Data Anal Knowl Organ GfKL 2007
33. Mansouri K, Abdelaziz A, Rybacka A, Roncaglioni A, Tropsha A, Varnek A, et al. 2016. CERAPP: Collaborative Estrogen Receptor Activity Prediction Project. Environmental Health Perspectives; doi:10.1289/ehp.1510267
34. NIH (2015) The PubChem Project. http://pubchem.ncbi.nlm.nih.gov/. Accessed 26 Jan 2015
35. Royal Society of Chemistry (2015) ChemSpider API Services. http://www.chemspider.com/AboutServices.aspx. Accessed 28 Jan 2015
36. ChemAxon (2014) ChemAxon Standardizer–Cheminformatics platforms and desktop applications. http://www.chemaxon.com/products/standardizer/. Accessed 26 Nov 2014
37. Reusch W (2013) Examples of chemical reactions. http://www2.chemistry.msu.edu/faculty/reusch/virttxtjml/react2.htm. Accessed 25 Nov 2014
38. Sitzmann M, Filippov IV, Nicklaus MC (2008) Internet resources integrating many small-molecule databases. SAR QSAR Environ Res 19:1–9. doi:10.1080/10629360701843540
39. Yap CW (2011) PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. J Comput Chem 32:1466–1474
40. Molecular Operating Environment (MOE) (2015). Chemical Computing Group Inc, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada
41. Leardi R, Lupiáñez González A (1998) Genetic algorithms applied to feature selection in PLS regression: how and when to use them. Chemom Intell Lab Syst 41:195–207. doi:10.1016/S0169-7439(98)00051-3
42. Kowalski BR, Bender CF (1972) The K-nearest neighbor classification rule (pattern recognition) applied to nuclear magnetic resonance spectral interpretation. Anal Chem 44:1405–1411
43. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemom Intell Lab Syst 58:109–130. doi:10.1016/S0169-7439(01)00155-1
44. Ståhle L, Wold S (1987) Partial least squares analysis with cross-validation for the two-class problem: a Monte Carlo study. J Chemom 1:185–196. doi:10.1002/cem.1180010306
45. Cortes C, Vapnik V (1995) Support-vector networks. In: Mach. Learn. pp 273–297
46. MATLAB Version 8.2. (2013), The MathWorks, Inc., Natick, Massachusetts, United States
47. Boser BE, Guyon IM, Vapnik VN (1992) A training algorithm for optimal margin classifiers. In: Proc. Fifth Annu. Workshop Comput. Learn. Theory. ACM, New York, NY, USA, pp 144–152
48. Chang C-C, Lin C-J (2001) LIBSVM: a library for support vector machines. National Taiwan University, Department of Computer Science, Taipei 106, Taiwan
49. De Jong S (1993) SIMPLS: an alternative approach to partial least squares regression. Chemom Intell Lab Syst 18:251–263. doi:10.1016/0169-7439(93)85002-X
50. Mevik B-H, Wehrens R (2007) The pls package: principal component and partial least squares regression in R. J Stat Softw 18:1–24
51. Consonni V, Ballabio D, Todeschini R (2010) Evaluation of model predictive ability by external validation techniques. J Chemom 24:194–201. doi:10.1002/cem.1290
52. Ballabio D, Consonni V, Todeschini R. (2009) The Kohonen and CP-ANN toolbox: a collection of MATLAB modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks.Chemometrics and Intelligent Laboratory Systems, 98, 115–122

53. Ballabio D, Vasighi M. (2012) A MATLAB Toolbox for Self Organizing Maps and supervised neural network learning strategies. Chemometrics and Intelligent Laboratory Systems, 118, 24–32

54. Bourne H, Horuk R, Kuhnke J, Michel H (2007) GPCRs: from deorphanization to lead structure identification. Springer-Verlag Berlin Heidelberg. doi:10.1007/978-3-540-48982-5

55. Overington JP, Al-Lazikani B, Hopkins AL (2006) How many drug targets are there? Nat Rev Drug Discov 5:993–996. doi:10.1038/nrd2199

Chapter 18

QSAR Models at the US FDA/NCTR

Huixiao Hong, Minjun Chen, Hui Wen Ng, and Weida Tong

Abstract

Quantitative structure–activity relationship (QSAR) has been used in the scientific research community for many decades and applied to drug discovery and development in the industry. QSAR technologies are advancing fast and attracting possible applications in regulatory science. To facilitate the development of reliable QSAR models, the FDA had invested a lot of efforts in constructing chemical databases with a variety of efficacy and safety endpoint data, as well as in the development of computational algorithms. In this chapter, we briefly describe some of the often used databases developed at the FDA such as EDKB (Endocrine Disruptor Knowledge Base), EADB (Estrogenic Activity Database), LTKB (Liver Toxicity Knowledge Base), and CERES (Chemical Evaluation and Risk Estimation System) and the technologies adopted by the agency such as Mold^2 program for calculation of a large and diverse set of molecular descriptors and decision forest algorithm for QSAR model development. We also summarize some QSAR models that have been developed for safety evaluation of the FDA-regulated products.

Key words FDA, Databases, Liver toxicity, Endocrine disruptors

1 Introduction

1.1 Brief History of QSAR

Quantitative structure–activity relationship (QSAR) is a relationship that can be presented as a mathematical function for predicting biological activities of compounds based on their chemical structures. QSAR dates back to the nineteenth century, when a very simple equation (1) was proposed by Crum-Brown and Fraser for the curare-like paralyzing properties of a set of quaternized strychnines [1]:

$$\Phi = f(C) \quad (1)$$

In Eq. (1), f is a mathematical function that converts the relevant structural features characterizing the quaternizing group, C, to the biological activity, Φ. Richardson constructed a QSAR in a reciprocal function that can estimates the toxicity effect of ethers and alcohols based on their water solubility [2].

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_18, © Springer Science+Business Media New York 2016

In the twentieth century, the QSAR techniques were advanced to utilize multiple parameters and were applied to many fields by a lot of pioneers. The most notable contribution to the emerging QSAR field is the so-called Hansch equation [3]:

$$\mathrm{Log}1/C = a\pi + b\pi^2 + c\sigma + dE_s + \text{constant} \quad (2)$$

In Eq. (2), the constants π, σ, and E_S represent the hydrophobic, electronic, and steric substituents, respectively [4]. By the end of last century, the advancement in computer technology and the generation of a huge amount of scientific data [5, 6] progressed the field of QSAR to a new height. A lot of QSAR methods have been developed and applied by the scientific research community and in regulatory sciences [7–10], e.g., pharmacophore modeling [11–15], molecular docking [16–19], CoMFA [19], classification tree model [20], decision forest [21–27], and support vector machine [28], to name a few.

1.2 The Role of QSAR at the FDA

Most of the US FDA-regulated products contain chemicals such as drugs, food additives, and cosmetic ingredients. Both benefit and risk of a product are important to the agency to protect the public health of the Americans. When some specific efficacy and toxicological data are needed but not available from experiments, alternative estimations are used to inform if further evidence from experiments is required for regulatory decision-making. With the advancements in computational technology and QSAR methods, QSAR can be used to make a reasonably accurate prediction quickly and plays more and more roles in regulatory sciences.

The rationale for applications of QSAR models to assess the efficacy and safety of the chemicals found in FDA-regulated products is illustrated in Fig. 1. When the data on the efficacy and safety are lacking for the chemicals, experimental data contained in the application or obtained through literature mining are used by the agency to inform regulatory actions as indicated by the dash-line arrows in Fig. 1. Alternatively, the efficacy and safety data are estimated using QSAR models as depicted by the solid-line arrows. In order to construct QSAR models, endpoint data for other chemicals should be gathered as the training set. Tools and algorithms are then applied to the training set to construct QSAR models to predict the required endpoint data for the chemicals in the FDA-regulated products. To facilitate the development of reliable QSAR models, the FDA had invested a lot of efforts in constructing chemical databases with a variety of efficacy and safety endpoint data, as well as in the development of computational algorithms.

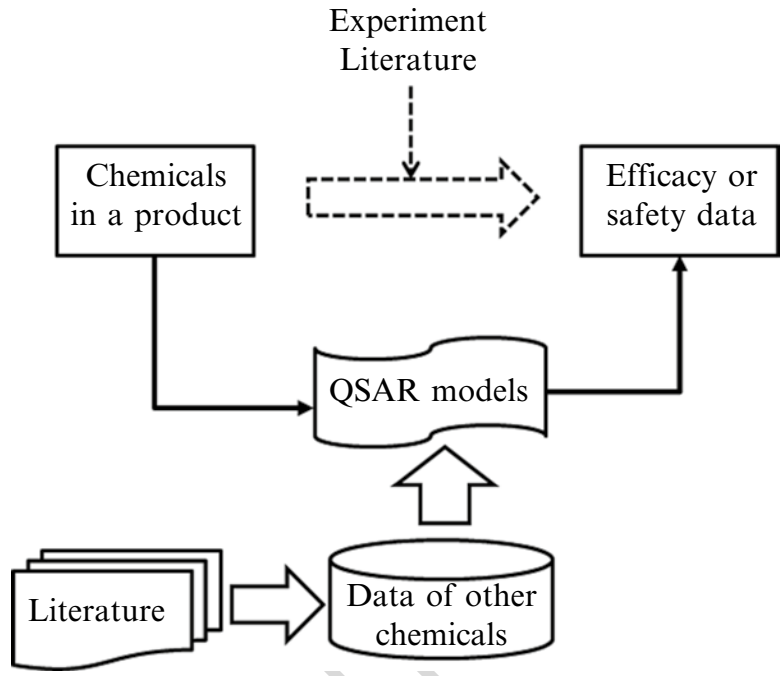


Fig. 1 Illustration of the role of QSAR in assessment of efficacy and safety for chemicals in the FDA-regulated products

2 Databases

Many chemical databases have been constructed and have been/could be used for developing QSAR models. In this chapter, we briefly discuss some of the often used databases.

2.1 EDKB

Endocrine disruptors (EDs) are exogenous compounds that act like hormones in the endocrine system of humans and other vertebrates. The endocrine activity of EDs has the potential to cause numerous adverse outcomes, e.g., disrupting the physiological function of endogenous hormones and altering homeostasis. The EDKB (Endocrine Disruptor Knowledge Base) (http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase/) is a database that was developed at the FDA's National Center for Toxicological Research (NCTR) to address these concerns [6]. It can be used to identify, prioritize, and inform the need for further thorough safety evaluation of chemicals with endocrine disruption potential in FDA-regulated products.

The EDKB database contains experimental data of different assays including estrogen receptor (ER) binding [29], androgen receptor (AR) binding [30], uterotrophic, cell proliferation, and reporter gene assays for more than 1800 chemicals. Detailed infor-

mation for each compound such as chemical structure, assay type, potency, etc. was organized in a manner that facilitates an efficient search strategy. A user-friendly interface has been implemented for the quick navigation of the database, efficient searching for information on chemicals with endocrine-related assay data, and graphical view of searched results. The search engine implemented in the EDKB enables searching by one or a combination of fields: chemical structure (including exact search and similarity search), chemical name, molecular formula, CAS registration number, experiment source, molecular weight, and so on. Cross-links to other publicly available and related databases are provided. Figure 2 shows a screenshot of the EDKB interface. Since its introduction to the scientific community, the EDKB has been a major data source for endocrine disruptor research, and many QSAR models have been developed based on the data contained in it.

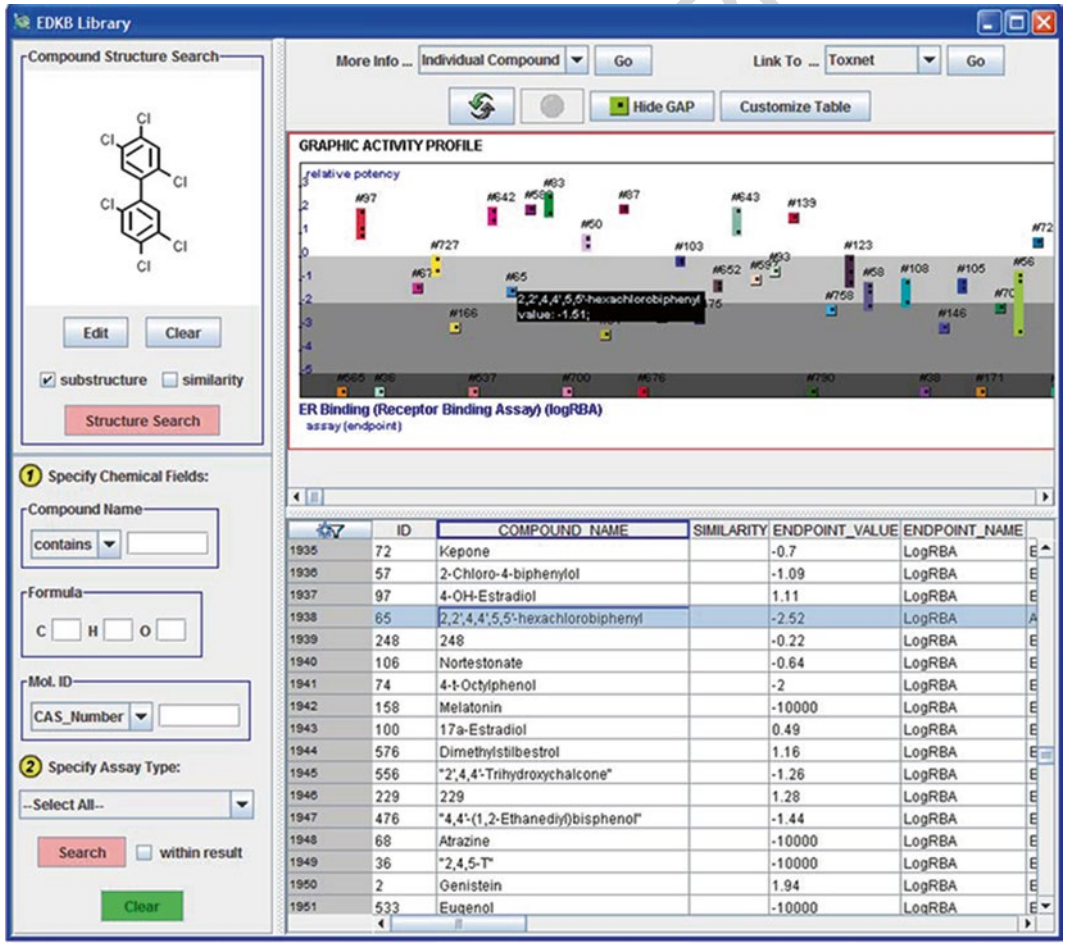


Fig. 2 Screenshot of EDKB

2.2 EADB

Chemicals of potential endocrine activity can interact with receptors in the body's endocrine system. ER is the major receptor that has been widely studied, and many endocrine disruptors can bind ER. Apart from being the target of endocrine disruptors, the ER is also a therapeutic target for treatment of various medical conditions. A huge amount of efforts in product safety evaluation in terms of estrogenic activity and drug development utilizing estrogenic chemicals have generated estrogenic activity data for a large number of chemicals, but these data exist in various sources and formats. This restricts the full utilization of these data, e.g., in benefit–risk ratio assessment in regulatory science and for discovering potent lead compounds in drug development. To facilitate the full utilization of the available data, FDA/NCTR has developed Estrogenic Activity Database (EADB) (http://www.fda.gov/ ScienceResearch/ BioinformaticsTools/EstrogenicActivityDatabaseEADB/default.htm). EADB is freely available to the scientific community.

There are more than 18,000 estrogenic activity data points from 1284 assays for more than 8000 chemicals in EADB. The data were curated from experiments of 11 different species. The assay types include ER binding, reporter gene, cell proliferation, and in vivo assays. The chemicals have a wide structural space, and the activity data cover a wide range and thus are suitable for QSAR development. A set of functions have been developed to help users to easily use the database in evaluation of compounds for their potential endocrine activity.

EADB has different user interfaces as shown in Fig. 3 to accommodate different purposes and users with different knowledge backgrounds. The biological data focused interface (Fig. 3a) is developed for examination of chemical structures with a specific estrogenic activity, while the chemical structure focused interface (Fig. 3b) is designed for exploration of compounds for their estrogenic activity data. The main part of the database is the table that is located in the right of the screen. It is designed for displaying the database content and the querying results. The left panel of the window displays query and chemical structure. Structure searching and data filtering functions are available in EADB. The "individual compound" button pops up the molecule interface that is used to display chemical-related data such as the chemical identifications, physical and chemical properties, and activity data (Fig. 3b).

Table 1 summarizes the database functions implemented in EADB that help users to fully utilize the data curated in EADB. The diverse and comprehensive estrogenic data are a rich source for the development of QSAR models for predicting estrogenic activity of chemicals. EADB is being widely accessed in the scientific community since we made the database publicly available in 2013.

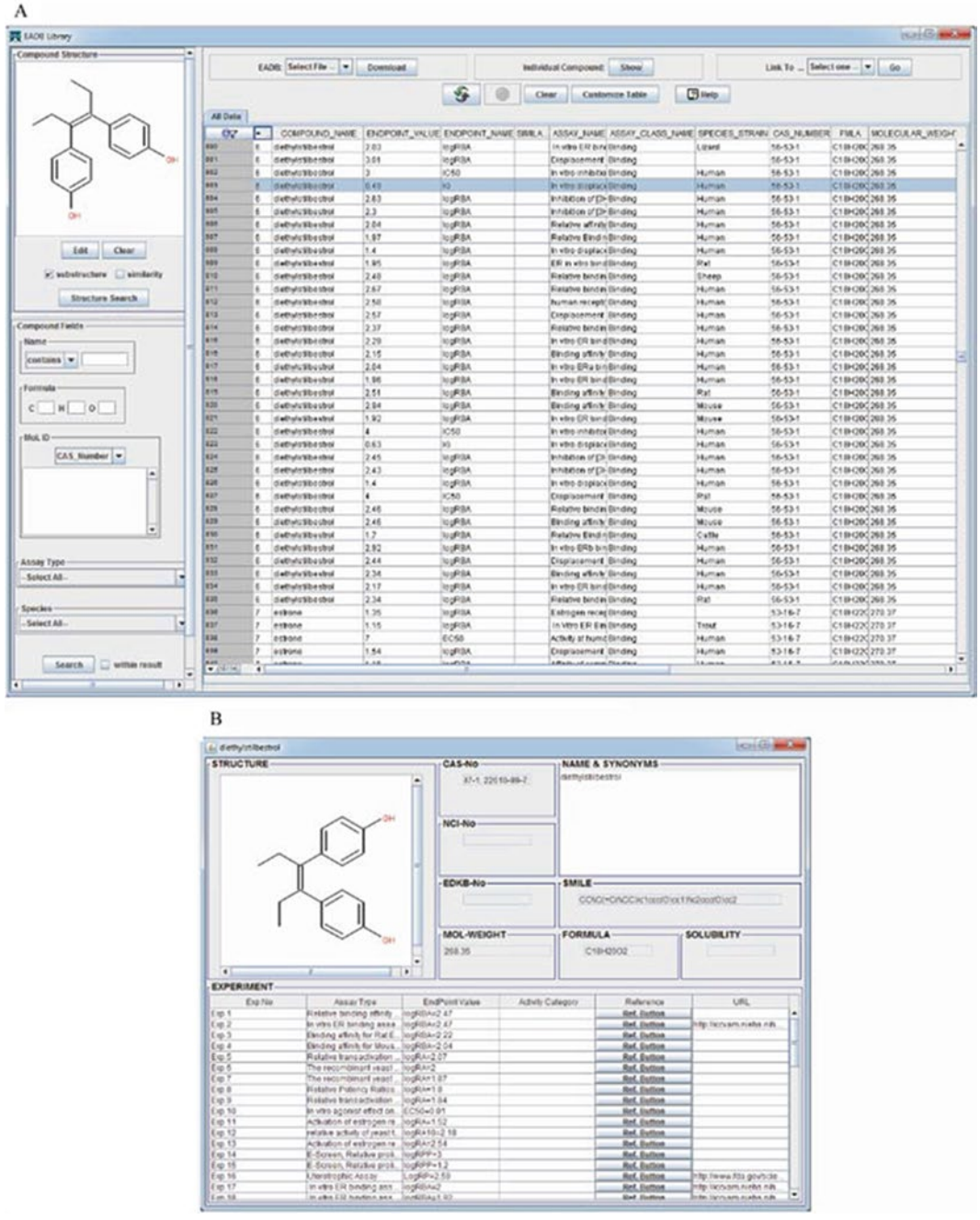


Fig. 3 EADB biological focused interface (**a**) and chemical structure focused interface (**b**). The query and filtering functions are implemented in the biological focused interface. The chemical structure focused interface can be opened by clicking the "Show" individual compound at the top of the biological focused interface

t1.1 **Table 1**
t1.2 **Database functions implemented in EADB**

t1.3 Function	Description
t1.4 Browsing	The database or search results can be browsed easily in different ways
t1.5 t1.6 t1.7 t1.8 Searching	Searching can be carried out on structure (substructure search, superstructure search, similarity search, full search, R-group search, and exclusion search) or data, including numerical data (various estrogenic activity data) and text data (assay descriptions and literature references), as well as logical combinations of multiple search operations
t1.9 t1.10 Updating	The database can be updated through adding new chemicals or estrogenic activity data, and editing the structures or data whenever errors are found
t1.11 Exporting	Structures and data can be exported in various formats

2.3 LTKB

Drug-induced liver injury (DILI) in humans is a significant risk for drug development, and creative approaches are needed to combat this risk [31]. Roughly over 1000 post-marketing drugs are considered to be potentially capable of inducing liver injury [32]. DILI is one of the main reasons for the issuing of "black box" warnings by the FDA and withdrawal drugs from the market [33]. Programs such as the National Institutes of Health's DILI Network and the "Virtual Liver Network" in Germany have been formed to enhance the understanding of the pathogenesis of DILI. LTKB (Liver Toxicity Knowledge Base), developed by the FDA/NCTR, aims to improve the understanding of DILI and facilitate the development of QSAR models for predicting DILI [34].

In the FDA's Advancing Regulatory Science Initiative, a variety of technologies such as in vitro assays, new animal tests, and in silico modeling are embraced to acquire a better understanding of the extrapolation from preclinical testing results to clinical setting. LTKB is designed as a knowledge base to incorporate all types of information needed for safety evaluation in terms of hepatotoxicity. The overall structure of LTKB is shown in Fig. 4. Diverse data covering multiple levels of biological complexity has been collected for most of the FDA-approved drugs in LTKB, such as toxicogenomics, mode of action, in vitro test results, histopathology, and adverse effects. Independently and in conjunction with other approaches, a predictive model for DILI was constructed at NCTR/FDA. The FDA drug labeling was used as primary source to annotate DILI risk to develop and evaluate these predictive models.

Since DILI is a heterogeneous disease and many factors are involved in DILI, incorrect predictions can be yielded for DILI based on a single type of data sources or using a single model. Consequently, combination of diverse data and integration of

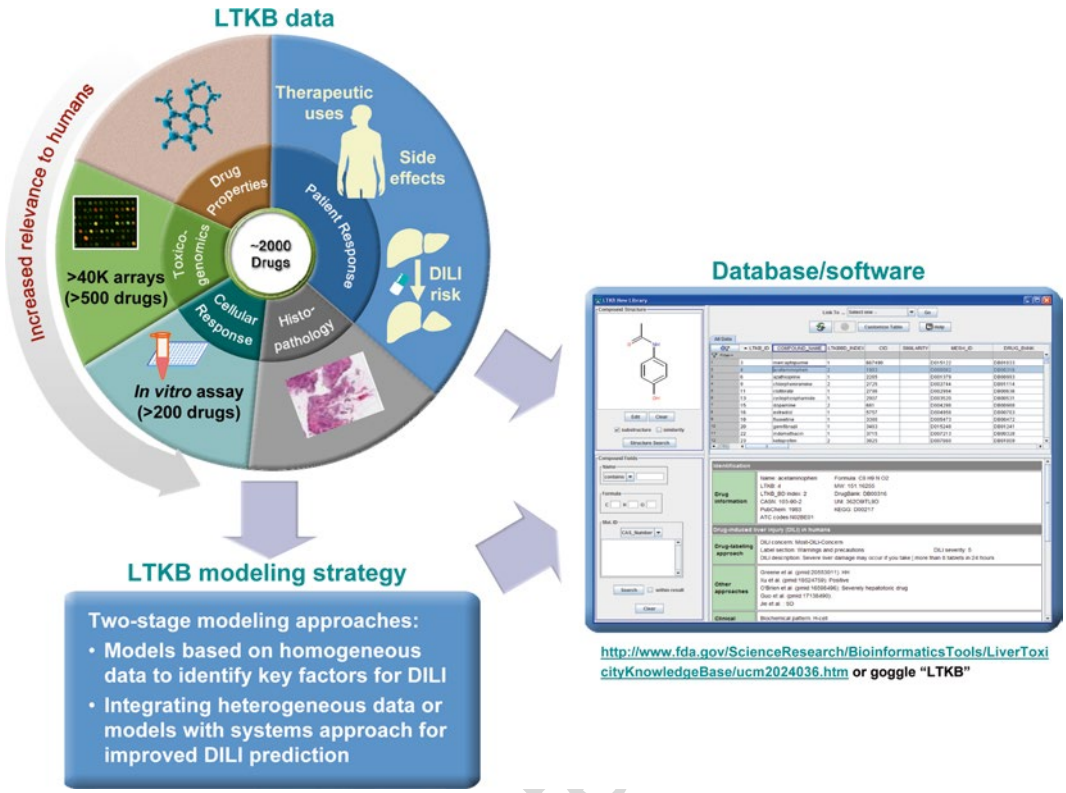


Fig. 4 Overview of the LTKB project. Three major components of LTKB were curation of drug-elicited data, development of predictive DILI models, and development of a software environment to publish both data and models

different models were used to advance model performance. Meanwhile, we established a "benchmark dataset" in which the drugs were well annotated with risks for DILI. Therefore, a standard set of defined drugs were offered to the research community to support assay development and QSAR model construction. Moreover, the LTKB team also utilized the datasets and tools developed by other institutes to for better understanding of DILI. Several other government-sponsored projects have adopted the LTKB benchmark drugs, including but not limited to ToxCast of the EPA and the Tox21 programs, a collaborative toxicological program among multiple US governmental agencies.

Three attributes can be used to identify a drug's potential for DILI: causality, incidence, and severity. The public data resource that can meet the three conditions is the FDA-approved drug labels. The unique characteristic of drug labels is the reflection of expert opinions based on clinical data and is continuously being enhanced with post-market surveillance data. The drugs are classified as most-DILI-concern, less-DILI-concern, and no-DILI-concern through mining the FDA-approved drug labels. The harmony

among published datasets is strong among most-DILI-concern and no-DILI-concern [35].

As illustrated in Fig. 4, LTKB comprehensively collected multiple types of data on the FDA-approved drugs, including physiochemical properties, in vitro data, toxicogenomic data, histopathological data, and adverse reactions. Multiple resources were utilized to collect the data. Particularly the two toxicogenomic databases—the Toxicogenomics Project in Japan (http://toxico.nibio.go.jp/open-tggates/search.html) and the DrugMatrix database of the National Institute of Environmental Health Sciences (https://ntp.niehs.nih.gov/drugmatrix/index.html)—have archived a total of over 40,000 microarray gene expression data from over 500 drug treatment [36] and are publicly available. Thus, they have been included in the LTKB.

Bioinformatics designs have been developed for the construction of DILI models using single analogous data that mirrors a single biological response. One bioinformatics strategy being used is the development of the QSAR model from data relating to roughly 500 drugs with chemical descriptors [27]. Meanwhile, we also utilized 164 oral drugs to identify a simple rule, namely, "rule of two"—a daily dose of ≥100 mg and a lipophilicity measured using log$P$≥3—to be associated with a significant risk of DILI in humans [37].

Integrating different types of data and predictive models is another critical effort of the LTKB. We found that individual models perform differently for drugs with different therapeutic uses [38], and methods such as consensus, hybrid, and/or hierarchical approaches tailored to therapeutic categories should be used to avoid the defects of "one size fits all" and increase prediction accuracy [39, 40].

2.4 SRS UNIIs

The SRS (Substance Registration System) (http://fdasis.nlm.nih.gov/srs) is a database system that is designed for the management of the substances contained in the FDA-regulated products. SRS contains some 7000 substances.

When registering, SRS generates unique ingredient identifiers (UNIIs) for the substances contained in the FDA-regulated products such as drugs, biologics, foods, and devices. A UNII is represented as an alphanumeric identifier that is nonproprietary, free, unique, unambiguous, and non-semantic. The UNIIs in SRS were generated based on the molecular structures and/or descriptive information of the substances registered. The UNIIs in SRS can be downloaded from http://fdasis.nlm.nih.gov/srs/jsp/srs/uniiListDownload.jsp.

The UNII codes are used by many FDA systems such as FDA's Structured Product Labeling, FDA Inactive Ingredient Query Application, and FDA Data Standards Council and

other governmental agencies such as NLM's Unified Medical Language System (UMLS), National Cancer Institute's Enterprise Vocabulary Service, and VA National Drug File Reference Terminology (NDF-RT).

SRS provides the most comprehensive product-related information and UNIIs. Therefore, it can also serve as a resource for development of QSAR models.

2.5 CERES

Chemical Evaluation and Risk Estimation System (CERES) is developed by the US FDA's Center for Food Safety and Nutrition for pre- and post-market review of food ingredients. It is a chemical evaluation and risk estimation system that is chemical centric. CERES contains internal and external chemical and toxicity data and knowledge on food additives using controlled vocabulary. It provides a variety of functions for data retrieval and structure and similarity search. Some QSAR models were developed and included in CERES.

3 Molecular Descriptors

In mathematics, a QSAR model, either qualitative or quantitative, is a mathematical function that describes the relationship between structures of chemicals and their biological activities. Alternatively, a QSAR model is a transformation that can be used to estimate the biological function of a chemical from its structure. It is very difficult, if not impossible, that a QSAR model predicts biological activity of a chemical by directly using its molecular structure (the red path in Fig. 5). In practices, a QSAR model is constructed to use chemical structures by indirectly describing the chemical structures in molecular descriptors rather the structures themselves. Molecular descriptors, the form of numerical descriptions that capture the structural characteristics, are easier to be encoded in a QSAR model. Therefore, a QSAR model (mathematical or statistical) can be developed to correlate the biological activity of interest with the molecular descriptors (or, in most cases, a subset of the descriptors) of the compounds (the blue path in Fig. 5).

QSAR is based on the assumption that the molecular structure of a chemical must contain features responsible for its physical, chemical, and biological properties. When described as numerical values, these features are known as molecular descriptors. The most frequently used molecular descriptors in the early stage of QSAR are empirical in nature such as substituent constants, partition coefficients, and various electronegativity-related parameters [41–47]. With the increased computational power and advances in sciences, other types of molecular descriptors such as quantum chemical, electronic, geometrical, constitutional, and topological descriptors are used in modern QSAR.

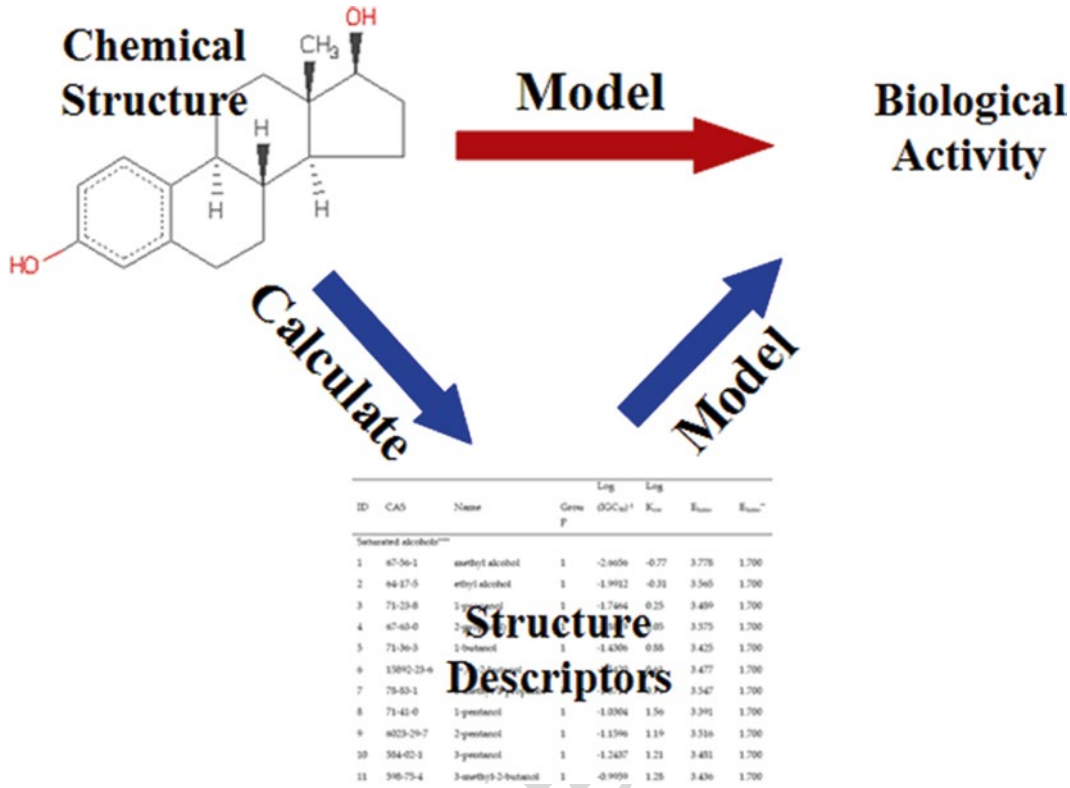


Fig. 5 Molecular descriptors bridge chemical structure to biological activity

Whether 3D molecular descriptors yield better predictive models than 2D descriptors is a long-time topic of debate in the scientific community. The argument for preference of 3D molecular descriptors emphasizes their capability of handling absolute stereochemistry and advantages of using force fields to model complexities of ligand binding to a receptor. However, some comparative studies [48–50] showed that 2D descriptors can perform as well as 3D descriptors in most applications, supporting the preference of 2D descriptors. Another advantage of 2D descriptors is the less computational cost because there is no need to deal with estimating the bioactive conformations required for 3D methods. The best set of molecular descriptors in the absolute sense may well be indeterminable.

The Mold^2 program was developed at the FDA/NCTR to enable the rapid calculation of a large and diverse set of molecular descriptors from both 1D and 2D chemical structure information [51]. Mold^2 is very fast and thus suitable not only for small datasets, as is normal in QSAR applications, but especially for the large databases typically in virtual screening chemicals. Calculation of the Mold^2 descriptors does not require 3D structures, and consequently, both the descriptors and models derived from them should be highly reproducible.

Table 2
Molecular descriptors in Mold²

Class	Subclass	Number of descriptors	Example of descriptors
1D	Counts for atoms	105	Number of O atoms
	Chemical physical property	2	Molecular weight
2D	Counts for atoms	80	Number of ring tertiary C
	Counts for bonds	9	Number of rotatable bonds
	Counts for functional groups	104	Number of carboxylic (aromatic)
	Chemical physical property	16	log*P*
	Structural features	13	Number of 5 member rings
	2D autocorrelation	96	Moran coefficient
	Balaban index	12	Normalized centric index
	Connectivity index	36	Randic connectivity index
	Detour index	24	Cyclicity index
	Distance (topological) index	73	Average atom eccentricity
	Eigen value-based descriptors	88	Folding degree index
	Information content	45	Mean information content
	Kier index	14	Kier flexibility
	Molecular walk counts	13	Total walk count
	Schultz index	4	Reciprocal Schultz index
	Topological charge index	21	Mean topological charge
	Wiener index	17	Normalized Wiener index
	Zagreb index	5	Quadratic index

Based on the 2D structure of chemical, the current version of Mold² calculates 777 molecular descriptors that are grouped as listed in Table 2 by their origin. Mold² primarily calculates constitutional and topological parameters as molecular descriptors. It is freely available to the public (http://www.fda.gov/ScienceResearch/BioinformaticsTools/Mold2/default.htm).

To demonstrate the applicability of Mold² molecular descriptors in QSAR, we compared the performance of QSAR models to correlate C_{max} (the maximum or "peak" concentration in serum of a drug observed after its administration) values to the structures of chemicals using Mold² and CODESSA (Comprehensive Descriptors for Structural and Statistical Analysis) which emphasizes descriptors obtained from quantum mechanical calculations. To construct the QSAR model using Mold² molecular descriptors, all of the 777 descriptors were first scaled to the range of 0–1. Then, a forward stepwise multiple linear regression method was used to construct an optimal regression model. In the construction of the model using CODESSA descriptors, all of the descriptors were scaled automatically from the software output. Thus, the descriptor values were directly used to build the QSAR model using the integrated best multiple linear regression (BMLR) algorithm [52]. For comparative purposes, a model involving the same number of parameters as the one generated using Mold² descriptors was selected.

Table 3
The best 5-parameter regression model obtained using CODESSA descriptors

X	ΔX	t	Descriptor
31.280	11.270	2.775	Intercept
0.714	0.143	4.994	Average information content (order 2)
7.935	2.127	3.731	FHACA fractional HACA (HACA/TMSA) (MOPAC PC)
0.069	0.020	3.419	Number of C atoms
35.900	11.790	3.044	Max *sigma–sigma* bond order
0.0008	0.0003	2.855	WNSA-2 weighted PNSA (PNSA2*TMSA/1000) (MOPAC PC)

$R^2 = 0.319$, $R^2_{cv} = 0.278$, $R^2_{test} = 0.229$, $F = 24.9$, Std. error of estimate = 1.10

Table 4
The best 5-parameter regression model obtained using Mold² descriptors

X	ΔX	t	Descriptor
0.168	0.571	0.294	Intercept
3.335	0.727	4.589	Mean of vertex distance information index
1.744	0.548	3.183	Information content order 3 index
3.416	1.097	3.114	Lowest eigenvalue from Burden matrix weighted by masses order-1
0.971	0.390	2.486	Maximal valence vertex electrotopological negative variation
1.008	0.678	1.487	Mean electrotopological states index

$R^2 = 0.302$, $R^2_{test} = 0.212$, $F = 23.1$, Std. error of estimate = 1.12

The two alternative regression models reported in Tables 3 and 4 are based on a dataset of 410 compounds with C_{max} values curated in the LTKB. The whole dataset was randomly split into the training (273 compounds) and test (137 compounds) subsets. The modeling results from using Mold², and CODESSA descriptors are plotted in Fig. 6a, b.

As shown in Fig. 6, both sets of molecular descriptors produced regression models of almost the same quality. The comparative study demonstrated that Mold² can be advantageously used for the purpose of QSAR as it can be ultrafast calculated.

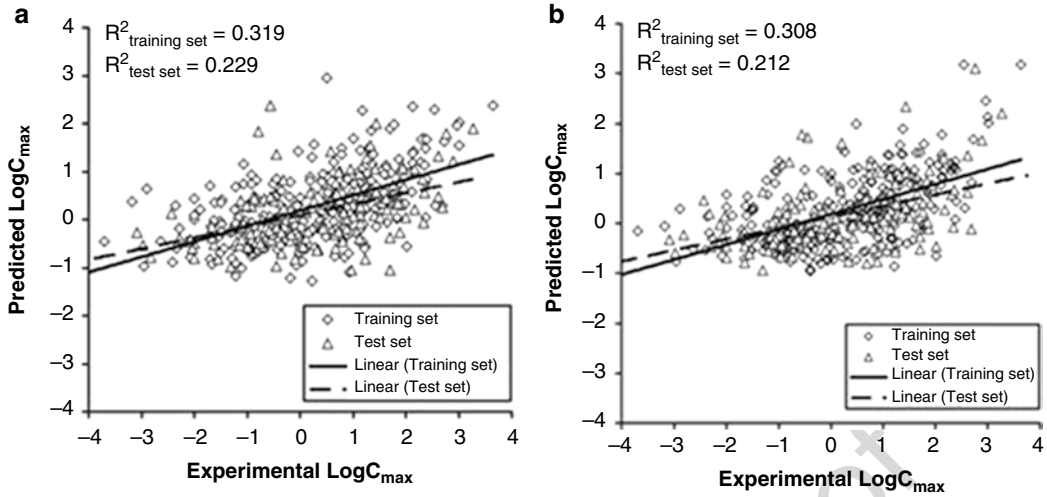


Fig. 6 Regression models based on CODESSA (**a**) Mold^2 (**b**) molecular descriptors. Experimental log (C_{max}) values were plotted against the predicted log (C_{max}) values. The training results were depicted by the diamond points, while the testing results were given by the triangle markers

4 QSAR Models Developed at the FDA/NCTR

4.1 QSAR Models for Predicting Estrogenic Activity

Chemicals that show potential endocrine activity could have adverse interactions with both humans and animals. They can directly or indirectly interact with many target proteins in the body's endocrine system. ER is the receptor that has been most studied for endocrine activity, and many endocrine active chemicals show estrogenic activity and could change the ER-mediated pathways. Several QSAR models have been developed for predicting estrogenic activity at the FDA/NCTR.

4.1.1 Tree-Based Model for Priority Setting

In 1996, the Food Quality Protection Act of 1996 (http://www.epa.gov/pesticides/regulating/laws/fqpa/) and the Safe Drinking Water Act Amendments of 1996 (http://water.epa.gov/lawsregs/guidance/sdwa/theme.cfm) were passed by the US Congress. These two acts requested the US Environmental Protection Agency (EPA) to screen and test for estrogenic, androgenic, and thyroid endpoints for a large number of chemicals in the environment. There were more than 87,000 chemicals in the environment for evaluation. The polymers or otherwise unlikely to bind to steroid receptors were filtered, leaving about 58,000 chemicals for evaluation. Experimental evaluation of such a large number of chemicals would require many years and extensive resources. Therefore, the US EPA adopted an approach requiring priority setting to rank the most potential chemicals for more resource-intensive and costly

experimental evaluations. Many biologic mechanisms such as receptor binding involve endocrine activity that has potential for endocrine disruption. Hence, rapid methods for characterizing ER binding activity are important priority setting of environmental chemicals. We developed a tree-based QSAR model for predicting ER binding potential of the 58,000 environment chemicals [20].

We used a training dataset having ER binding data from an established in-house rat ER binding assay (NCTR dataset) [20] for model development. Chemicals are classified as ER binders and non-binders by the tree-based model using a series of rules on the basis of descriptors. To evaluate a large number of initial 153 molecular descriptors and identify the ones most informative for the tree-based model, we selected the top ten descriptors using the genetic function approximation (GFA) approach. Several tree-based models were constructed based on the NCTR dataset using combined groups of three to six of the top ten descriptors. The model giving the best concordance was the final model. Five descriptors (phenolic ring index, log*P*, Jurs-PNSA-2, Jurs-RPCS, and shadow-XY fraction) were used in the final tree-based model as shown in Fig. 7. The presence or absence of the phenolic group in

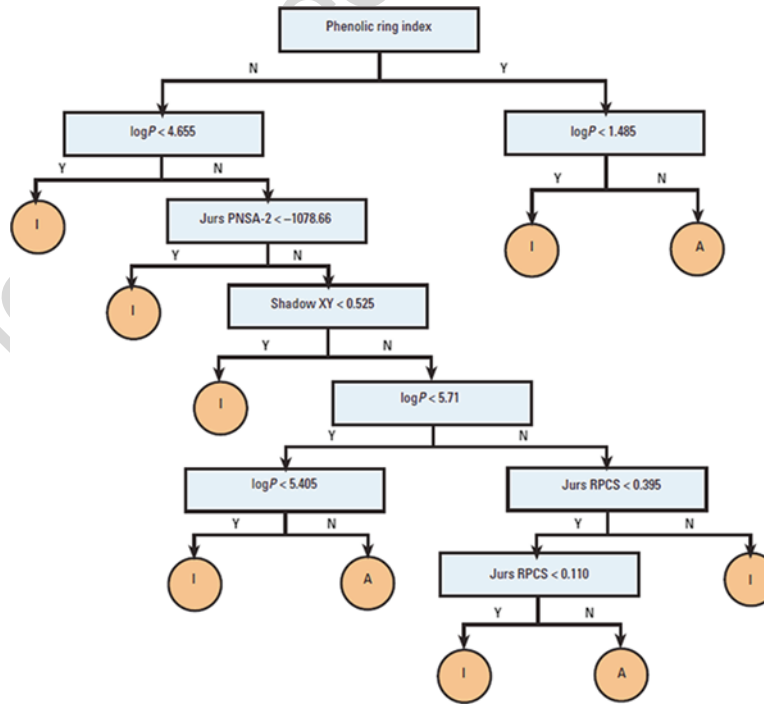


Fig. 7 Tree-based model. The model displays a series of yes/no (Y/N) rules to classify chemicals into active (A) and inactive (I) categories based on five descriptors: phenolic ring index, log*P*, Jurs PNSA-2, shadow-XY, and Jurs RPCS. The squares represent the rules; the circle represents the categorical results

a chemical was the phenolic ring index. The log*P* measures the hydrophobicity of a chemical [53]. Jurs-PNSA-2 and Jurs-RPCS combined molecular shape and electronic information and were used to characterize the positive-charged surface area of a molecule [54]. The breadth of a molecule was represented by a geometric descriptor, the shadow-XY fraction [55].

The training results of the tree-based model showed an accuracy of about 88 % for the NCTR dataset. More specifically, 123 of the 131 ER binders were correctly predicted to be active (sensitivity = 93.9 %), while 81 of the 101 non-ER binders were correctly predicted to be inactive (specificity = 80.2 %).

The dataset reported by Nishihara et al. [56] was then used as a test dataset to challenge the tree-based QSAR model constructed from the NCTR dataset. This dataset was generated using the yeast two-hybrid assay for 517 chemicals, most of which are pesticides and industrial chemicals. After removing the chemicals that lacked unique structures such as mixtures, the remaining 463 chemicals were used for the test. Sixty-two chemicals were defined as active using the criterion of activity >10 % of 10^{-7} M E_2 by Nishihara et al., while the majority of the chemicals were treated as inactive.

An accuracy of 82.5 % was yielded when applying the tree-based model to the Nishihara dataset. The sensitivity and specificity of the model were 87.1 % (54/62) and 81.8 % (328/401), respectively.

We applied the tree-based QSAR model into an integrated system that consists of rejection filters, structural alerts, and the tree-based model to prioritize the some 58,000 environmental chemicals. Of 58,230 chemicals in priority setting, the two rejection filters removed 16,689 chemicals as ER non-binders. The remaining 41,541 chemicals were predicted for their ER binding activity using the tree-based model and the structural alerts. The prediction yielded that 6903 chemicals were ER binders and 34638 chemicals ER non-binders. Our results suggested that less than 12 % (6903) of the original 58230 chemicals might need to be tested for their potential ER activity. Of the 6903 chemicals, only 104 chemicals had the most active ER activity as they were predicted to be active by more than three of the four models (the tree-based model and three structural alerts).

4.1.2 Docking Models for Predicting ER Agonists and Antagonists

Structurally diverse chemicals can bind the ER to change the conformation of the protein in a nonspecific way, altering normal estrogen signaling through genomic and non-genomic pathways [57, 58]. Depending on their binding to ER, xenoestrogens can be agonists, partial agonists, or antagonists, altering normal gene expression levels and functions modulated by endogenous hormones [59, 60].

Many in vivo and high-throughput in vitro assays have been developed and validated to screen for mimics that act either as

estrogens or antiestrogens. However, comprehensively testing hundreds of thousands of man-made chemicals would be too expensive [61]. The timeline would also be highly protracted, given that a few chemical classes have been tested in over a decade, barely the tip of the iceberg of the chemical universe. Finally, the validated experimental techniques are not comprehensive at the moment. As developmental endpoints, means to detect levels of no biological effect, mixture and metabolism effects, among other limitations, are not adequately represented. Therefore, a full assessment of endocrine activity potential across the universe of chemicals constitutes a daunting problem, and QSAR models are needed to reduce costs and streamline the process.

In silico methods have often been used to complement experimental studies in order to assist with data analysis as well as improve results. In this instance, rapid QSAR models can be used not only to help identify and prioritize which class of compounds to screen but also reduce the number of compounds to be tested. Docking is one of the popular QSAR techniques often used for ligand pose prediction, ligand binding affinity prediction, as well as identifying potential actives from a library of decoys in virtual screening [62].

QSAR models based on docking techniques demonstrated that docking has utility to differentiate potential ligands (binders) from decoys (non-binders). However, current docking QSAR models lack the ability to distinguish agonists from antagonists and are thus unable to obviate or reduce experimental assays for further understanding of the mechanisms of actions of xenoestrogen. Hence, we have developed a QSAR model based on docking that can differentiate ligands in accordance with likelihood of activating or inhibiting or blocking the activity of ER.

This QSAR model consists of two separate docking models (SDMs), one constructed using known agonists and the other was built from known antagonists [16]. Figure 8 shows the study design. Basically, two SDMs were constructed to form a competitive docking model (CDM) for differentiation of ER agonists from ER antagonists. The SDMs compete in determination of ER agonist or antagonist. The CDM used docking scores that estimates the non-covalent interactions between a chemical and the ER agonist conformation and ER antagonist conformation to select the preferred ER binding mode for the chemical. A chemical is predicted to be (in a winner-take-all strategy) the type, agonist or antagonist, corresponding to the most favorable docking score from the individual SDMs.

The rationale of this approach is the dynamic nature of competing ligand-protein complexes where agonists and antagonists impart different conformation changes not represented by a single rigid conformation found in prior docking models. The approach was tested using two sets of ER ligands (one extracted from PDB crystal structures and another from the DUD [63]). We used

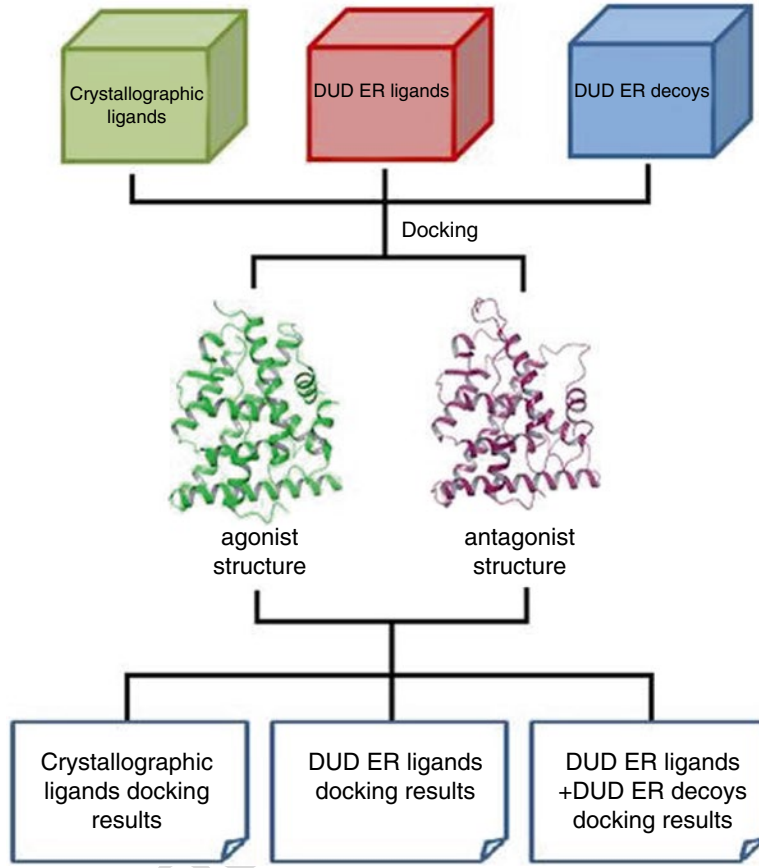


Fig. 8 Study design depicting the overall workflow. Three ligand sets are used for docking. While the first set of ligands is derived from the crystal structures available from the PDB, the second and third sets of ligands and decoys, respectively, are obtained from the DUD website. Results from the first and second sets of docking will be used to evaluate the ability of the CDM to differentiate agonists and antagonists, while the results from the second and third sets of dockings will be combined and used to calculate enrichment factors

enrichment factors (EFs) as the performance metric to assess the quality of our SDMs and CDM through virtual screening. Results obtained showed that the CDM could differentiate agonists from antagonists as depicted in Fig. 9.

4.1.3 Decision Forest Model for Predicting ER Binding

With EADB, a database that contains estrogenic activity data and structural information for more than 8000 chemicals collected by mining the literature and publicly available databases, it is expected that more accurate and reliable QSAR models for predicting estrogenic activity can be developed as the number of chemicals that can be used for training QSAR models are large, and consistency of

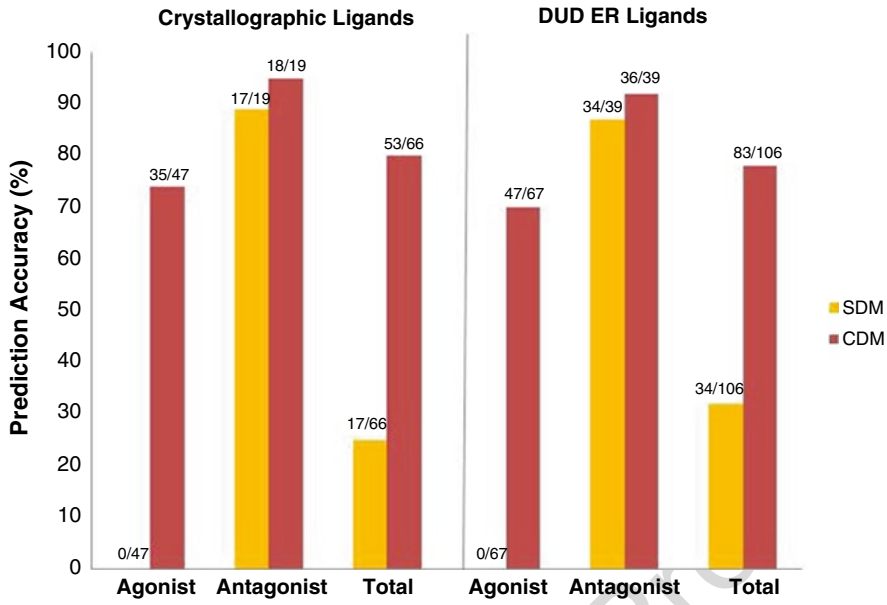


Fig. 9 The bar charts show the prediction accuracy of the SDMs (*yellow*) and CDM (*red*) for the crystallographic and DUD ER ligand sets. The bar heights denote the total number of ligands in each category. In all cases, CDM outperformed the SDMs, particularly in the case of agonist predictions

activity data for the same chemicals can be used to improve the data quality in the training dataset. To demonstrate the utility of EADB in the prediction of estrogenic activity, we developed a QSAR model for predicting ER binding activity.

First, we examined the data consistency for the 5497 chemicals that have ER binding data in EADB. We found 103 chemicals having discordant ER binding activity data (i.e., active in some assays but inactive in other assays), and, thus, we removed them from the QSAR model development. Of the rest 5394 chemicals having consistent ER activity data, 4719 are ER binders, while 675 are ER non-binders. The 777 Mold^2 molecular structures [51] for each of these 5394 chemicals were then calculated using the SDF files exported from EADB. Thereafter, we removed the molecular descriptors with constant value across all the 5394 chemicals. The values of the remaining 633 Mold^2 descriptors were then normalized from zero to one. Lastly, decision forest (DF) [21], the novel supervised machine learning algorithm, was used to build the QSAR model for prediction of ER binders using the normalized Mold^2 descriptors.

The fivefold cross validation was carried out to evaluate predictive power and robustness of the DF QSAR model as depicted in Fig. 10. For one iteration of the fivefold cross validation, the 5394 chemicals were randomly split into five equal portions. One portion of the chemicals was left for testing the DF QSAR model trained using the remaining four portions. This process was

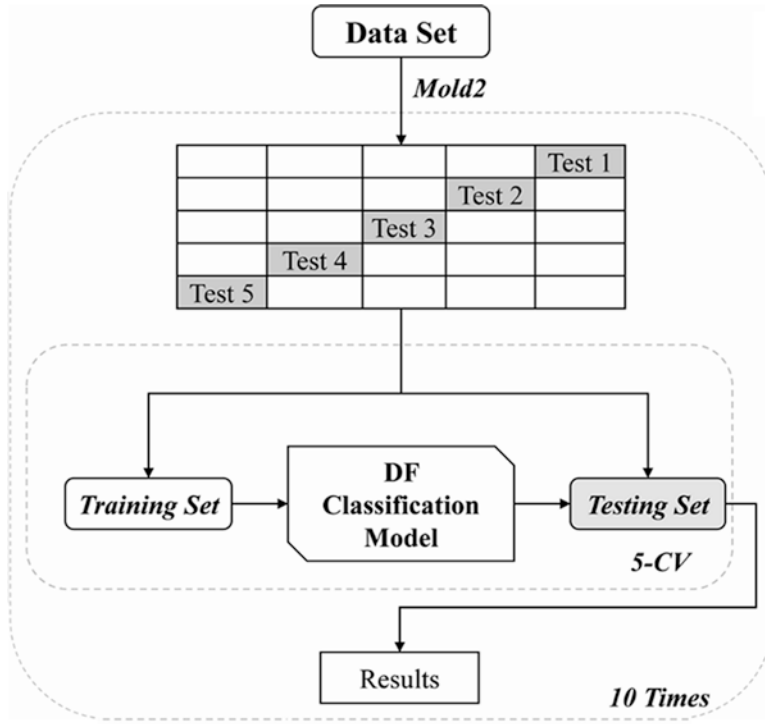


Fig. 10 The flowchart of fivefold cross validations. The dataset was first randomly split into five portions. Four portions were used to construct a DF model, and the remaining one portion was used to challenge the model. This procedure was iterated for five times by changing the challenge portions. The predictions of the five models were then used as measurement of the performance for the DF model. The random splits of the datasets into five portions were repeated ten times

repeated by changing the left portion of chemicals so that all the five portions were used as testing datasets. The prediction results yielded from the five models were then averaged to provide the estimate of model performance. To make the performance evaluation statistically robust, the fivefold cross validation was repeated ten times using different random divisions of the 5394 chemicals to ensure the results are not purely by chance.

The predictive performance of the DF QSAR model was summarized in Fig. 11. The mean accuracy, sensitivity, and specificity reached 93.84 % (standard deviation (SD) = 0.25 %), 98.03 % (SD = 0.21 %), and 64.53 % (SD = 2.51 %), respectively. The results demonstrated that EADB is a valuable resource and a convenient tool for developing high-quality QSAR models.

4.2 QSAR Models for Predicting Drug-Induced Liver Injury

The shrink in the number of recent drugs presented to the market causes detriment for the pharmaceutical industry. This is attributable to about 90 % of drug candidates approved for human testing failing in clinical trials [64]. Drug potency and

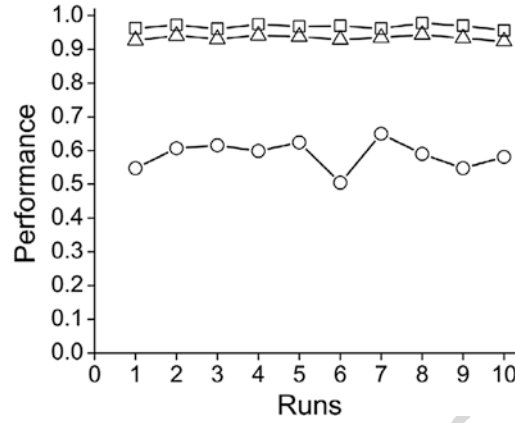


Fig. 11 Performance of the ten iterations of fivefold cross validation. Accuracy was plotted in *up triangles*, sensitivity in *squares*, and specificity in *circles*

toxicity are the main causes of drug failure, and DILI (drug-induced liver injury) is one of principal toxicity causes [31]. QSAR is a computational method that has shown to be useful for safety screening during the early stages of drug discovery [65]. The physicochemical nature of compounds like lipophilicity has been identified as an important risk factor for DILI when considered together with daily dose [37]. QSAR models have been published for the study of hepatotoxicity, and currently the performance of the majority of published models for DILI in humans has been less than satisfactory, with efficiencies of roughly 60 % or less [32], most notably when the models are tested by large external validation sets.

4.2.1 QSAR Model for DILI Prediction

To facilitate the development of better QSAR models for assessing DILI risk in humans, FDA-approved drug labeling data were used. Drug labeling is one of the few public data sources that can assess severity, causality, and incidence, a requirement for assessing a drug's potential for human hepatotoxicity [35]. The FDA-approved drug labeling separated drugs into three groups: most-DILI-concern, less-DILI-concern, and no-DILI-concern that were used to develop a QSAR model for predicting DILI. Three published datasets, namely, NCTR dataset [27], Xu et al. dataset [66], and Greene et al. dataset [67], were used as independent validation sets to measure the performance of the QSAR model developed from the training set. Mold^2 molecular descriptors [51] were used in the QSAR model. DF [21] was used to develop the QSAR model. Cross validation was used to measure the model performance. Permutation analysis was used to determine if a model's performance was different from random chance. Figure 12 gives an overall strategy of the QSAR model development and validation procedure.

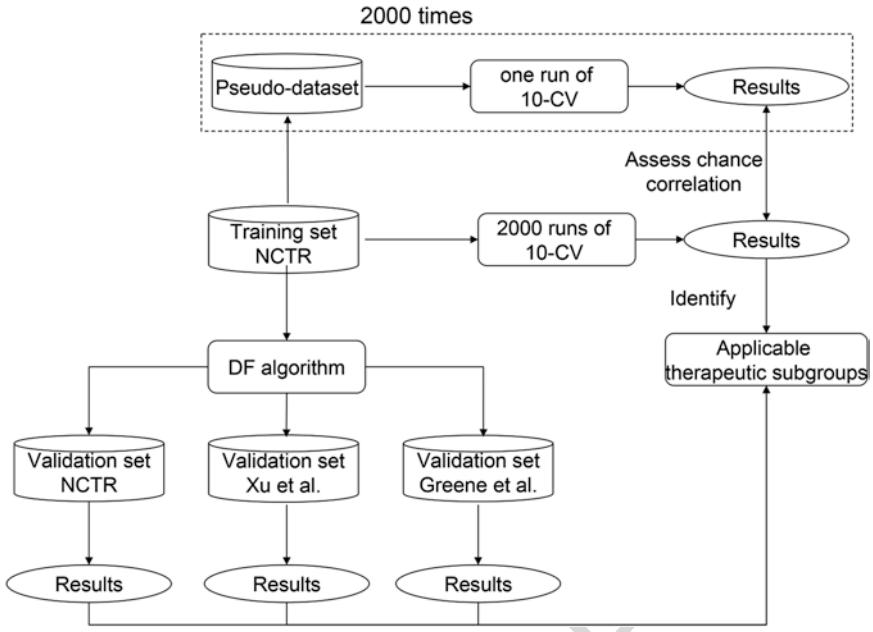


Fig. 12 Flowchart of the quantitative structure–activity relationship model development and validation procedure of the study

The QSAR model composed of six decision trees using 82 Mold^2 descriptors.

4.2.2 Cross Validation of the QSAR Model and Permutation Tests

The mean accuracy, sensitivity, and specificity listed in Table 5 are the results of 2000 repetitions of tenfold cross validation. The distributions of prediction accuracy for the 2000 cross validations and for 2000 permutations are plotted in Fig. 13. The low percentage of the permutation results (48.5 %) suggests that the medium accuracy value from the cross validations (69.7 %) was not by mere chance.

4.2.3 External Validation of the QSAR Model

The prediction performance of the QSAR model was evaluated using three validation datasets: the NCTR, Xu, and Greene validation sets. The performances of the QSAR model on the three external datasets are listed in Table 5.

We further compared the QSAR predictions between drugs with consistent and inconsistent DILI annotations across the three external validation sets. We observed that 70 % of the drugs with consistent annotations between the NCTR and Greene validation datasets were correctly predicted by the QSAR model, while only 58.8 % of the drugs with inconsistent annotations could be correctly predicted. The same trend was observed in the comparison between the NCTR and Xu validation datasets and between the Greene and Xu validation datasets.

t5.1 **Table 5**
t5.2 **Summary of internal cross validation and external validation results**

t5.3 t5.4		Cross validation (2000 runs)	External validation		
t5.5		NCTR training set[a]	NCTR validation set	Greene dataset	Xu dataset
t5.6	Accuracy	69.7 % ± 2.9 %	68.6 %	61.6 %	63.1 %
t5.7	Sensitivity	57.8 % ± 6.2 %	66.3 %	58.4 %	60.6 %
t5.8	Specificity	77.9 % ± 3.0 %	70.8 %	67.5 %	66.1 %
t5.9	PPV	64.6 % ± 4.3 %	69.2 %	77.2 %	68.4 %
t5.10	NPV	72.6 % ± 2.5 %	68.0 %	46.4 %	58.1 %
t5.11 t5.12	Drugs	197 (pos/ neg = 81/116)	191 (pos/ neg = 95/96)	328 (pos/ neg = 214/114)	241 (pos/ neg = 132/109)

t5.13 Cross validation results are averaged values of 2000 runs of tenfold cross validations. External validation results are
t5.14 prediction results on the three independent validation sets
t5.15 [a]Mean ± relative standard deviation

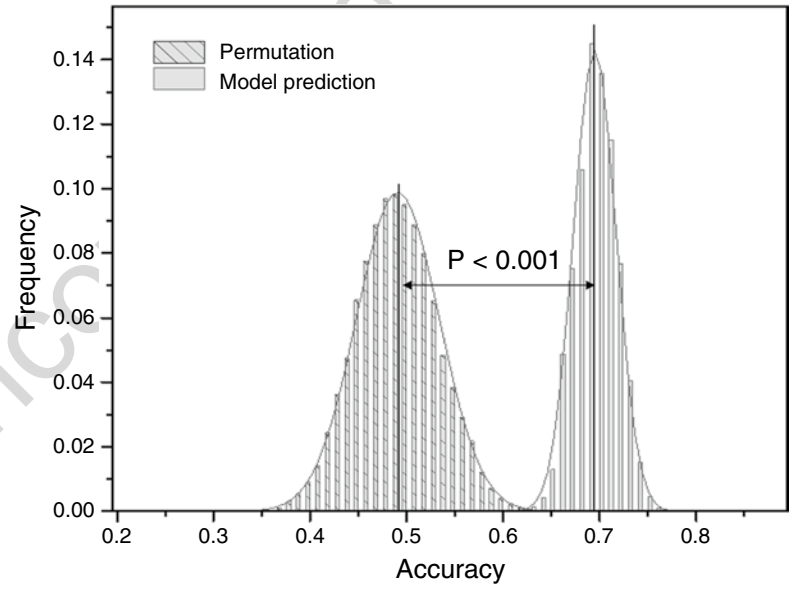


Fig. 13 Distribution of prediction accuracy of the 2000 runs of tenfold cross validation and the 2000 permutation analysis

4.2.4 Identification of High-Confidence Therapeutic Subgroups

The tenfold cross validations performed on the training set were used to explore the difference of the QSAR model's predictive performance for drugs in the therapeutic subgroups defined by the second level of the Anatomical Therapeutic Chemical (ATC). The result showed that the QSAR model had a higher prediction

Table 6
The high- and low-confidence therapeutic subgroups (second level of ATC classification) identified from cross validation of the QSAR model based on the training set

Confidence domain	Therapeutic subgroup (ATC code)
High confidence	Stomatological (A01), drugs for functional gastrointestinal disorders (A03), antidiarrheals, intestinal anti-inflammatory/anti-infective agents (A07), antihemorrhagics (B02), cardiac therapy (C01), antihypertensives (C02), vasoprotectives (C05), antipruritics, incl. antihistamines, anesthetics, etc (D04), antibiotics and chemotherapeutics for dermatological use (D06), corticosteroids for systemic use (H02), antibacterials for systemic use (J01), muscle relaxants (M03), analgesics (N02), nasal preparations (R01), throat preparations (R02), drugs for obstructive airway diseases (R03), cough and cold preparations (R05), antihistamines for systemic use (R06), ophthalmologicals (S01), otologicals (S02), ophthalmological and otological preparations (S03), contrast media (V08)
Low confidence	Vitamins (A11), anabolic agents for systemic use (A14), antithrombotic agents (B01), peripheral vasodilators (C04), calcium channel blockers (C08), agents acting on the renin-angiotensin system (C09), antifungals for dermatological use (D01), gynecological anti-infectives and antiseptics (G01), sex hormones and modulators of the genital system (G03), urologicals (G04), antivirals for systemic use (J05), antineoplastic agents (L01), endocrine therapy (L02), antiinflammatory and antirheumatic products (M01), anesthetics (N01), psycholeptics (N05), psychoanaleptics (N06), antiprotozoals (P01)

accuracy than the overall prediction for drugs in 22 therapeutic groups that were termed as high-confidence subgroups and had a lower prediction accuracy than the overall prediction for drugs in 18 therapeutic groups that were assigned as low-confidence subgroups (Table 6). The predictive accuracy values of the 40 therapeutic subgroups were plotted as a bar chart in Fig. 14, showing that the QSAR could be used for predicting DILI in humans, especially for the high-confidence therapeutic subgroups like analgesics, antibacterial agents, and antihistamines.

Consistent with the observation in the tenfold cross validations, the external validations also showed that the QSAR model performed better for drugs in the high-confidence therapeutic subgroups than drugs in the low-confidence therapeutic subgroups.

Our study demonstrated the possibility of constructing QSAR models for predicting DILI potential.

5 Future Perspectives

Generally speaking, all currently available descriptor software packages share a firm foundation in theory and practice that span for many decades. Likewise, there is a rich literature and proven track record of contributions of these software packages to chemistry, medicinal

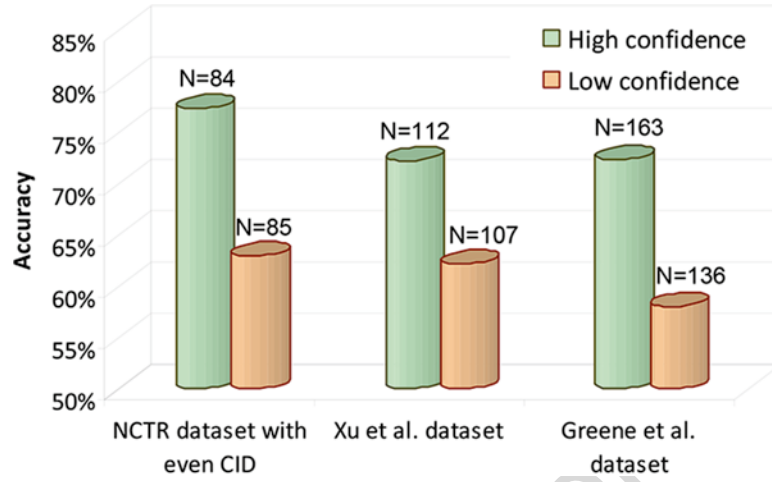


Fig. 14 Prognostic accuracy in the high- and low-confidence therapeutic subgroups derived from the three external validation sets

chemistry, and a myriad of other QSAR applications. Different software packages may emphasize on specific descriptor space domains corresponding to distinct aspects of chemical structure space or may offer various statistical or graphical functionalities. Improvement of current descriptor calculation software packages and development of new descriptors that can describe chemical structures more comprehensively to cover chemical structure characterization are anticipated in the near future. With more comprehensive molecular descriptors, it can be expected that better QSAR models will be obtained from high-quality data by the experienced and careful practitioner.

Multiple QSAR models have been and could be developed for a specific endpoint using different datasets, descriptors, and machine learning methods. In most cases, the models for predicting the same endpoint perform differently for some chemicals. The question as to how to utilize the models in applications is a challenge in the field. A common approach is to combine the results from multiple models using different consensus approaches. Many consensus methods have been used in QSAR modeling. For example, decision forest [21] uses consensus modeling by combining multiple well-learned models from different sets of descriptors of all samples using both majority voting and weighted voting. In a different way, random forest [68] ensembles unpruned classification trees constructed by using bootstrapping samples from the training samples and a subset of features randomly selected. Consensus methods had gained applications in regulatory science recently, as evidenced by the issuing of "Practical guide 2: How to report weight of evidence" (https://echa.europa.eu/documents/10162/13655/pg_report_weight_of_evidence_en.pdf) by

ECHA (European Chemicals Agency). It is expected that more consensus modeling methods will be developed and used for QSAR in the future.

There are diverse data available for QSAR. The amount of data for the same chemical grows quickly, making utilization of the data in QSAR very challenging. In addition to the challenges in capturing, curating, storing, visualizing, and sharing of the big and diverse data, fusion of the data will also be a challenging task for the development of more robust and accurate QSAR models. Fusion data from different assays, especially from different emerging technologies, will be a key step to fully utilize the knowledge in the QSAR research. As the big data solutions continue apace, we expect that more data fusion algorithms will be developed, and they will, in turn, improve QSAR in the near future.

Declaration

The views presented in this article do not necessarily reflect those of the US Food and Drug Administration.

References

1. Brown AC, Fraser TR (1868) On the connection between chemical constitution and physiological action; with special reference to the physiological action of the salts of the ammonium bases derived from strychnia, brucia, thebaia, codeia, morphia, and nicotia. J Anat Physiol 2(2):224–242
2. Richardson BW (1869) Lectures on experimental and practical medicine. Physiological research on alcohols. Med Times Gaz 2 (1869):703–706
3. Hansch C, Muir RM, Fujita T, Maloney PP, Geiger F, Streich M (1963) The correlation of biological activity of plant growth regulators and chloromycetin derivatives with Hammett constants and partition coefficients. J Am Chem Soc 85(18):2817–2824
4. Taft RW, Lewis IC (1959) Evaluation of resonance effects on reactivity by application of the linear inductive energy relationship. V. Concerning a σR scale of resonance effects 1,2. J Am Chem Soc 81(20):5343–5352
5. Shen J, Xu L, Fang H, Richard AM, Bray JD, Judson RS et al (2013) EADB: an estrogenic activity database for assessing potential endocrine activity. Toxicol Sci 135(2):277–291
6. Ding D, Xu L, Fang H, Hong H, Perkins R, Harris S et al (2010) The EDKB: an established knowledge base for endocrine disrupting chemicals. BMC Bioinformatics 11(Suppl 6):S5
7. Shi L, Tong W, Fang H, Xie Q, Hong H, Perkins R et al (2002) An integrated "4-phase" approach for setting endocrine disruption screening priorities –phase I and II predictions of estrogen receptor binding affinity. SAR QSAR Environ Res 13(1):69–88
8. Tong W, Fang H, Hong H, Xie Q, Perkins R, Anson J et al (2003) Regulatory application of SAR/QSAR for priority setting of endocrine disruptors: a perspective*. Pure Appl Chem 75(11):2375–2388
9. Tong W, Hong H, Xie Q, Shi L, Fang H, Perkins R (2005) Assessing QSAR limitations-A regulatory perspective. Curr Comput Aided Drug Des 1(2):195–205
10. Tong W, Perkins R, Fang H, Hong H, Xie Q, Branham W et al (2002) Development of quantitative structure-activity relationships (QSARs) and their use for priority setting in the testing strategy of endocrine disruptors. Regul Res Perspect 1(3):1–13
11. Hong H, Neamati N, Wang S, Nicklaus MC, Mazumder A, Zhao H et al (1997) Discovery of HIV-1 integrase inhibitors by pharmacophore searching. J Med Chem 40(6):930–936
12. Hong H, Neamati N, Winslow HE, Christensen JL, Orr A, Pommier Y et al (1998) Identification of HIV-1 integrase inhibitors based on a four-point pharmacophore. Antivir Chem Chemother 9(6):461–472

13. Neamati N, Hong H, Sunder S, Milne GW, Pommier Y (1997) Potent inhibitors of human immunodeficiency virus type 1 integrase: identification of a novel four-point pharmacophore and tetracyclines as novel inhibitors. Mol Pharmacol 52(6):1041–1055
14. Neamati N, Hong H, Mazumder A, Wang S, Sunder S, Nicklaus MC et al (1997) Depsides and depsidones as inhibitors of HIV-1 integrase: discovery of novel inhibitors through 3D database searching. J Med Chem 40(6): 942–951
15. Nicklaus MC, Neamati N, Hong H, Mazumder A, Sunder S, Chen J et al (1997) HIV-1 integrase pharmacophore: discovery of inhibitors through three-dimensional database searching. J Med Chem 40(6):920–929
16. Ng HW, Zhang W, Shu M, Luo H, Ge W, Perkins R et al (2014) Competitive molecular docking approach for predicting estrogen receptor subtype alpha agonists and antagonists. BMC Bioinformatics 15(Suppl 11):S4
17. Shen J, Zhang W, Fang H, Perkins R, Tong W, Hong H (2013) Homology modeling, molecular docking, and molecular dynamics simulations elucidated alpha-fetoprotein binding modes. BMC Bioinformatics 14(Suppl 14):S6
18. Tie Y, McPhail B, Hong H, Pearce BA, Schnackenberg LK, Ge W et al (2012) Modeling chemical interaction profiles: II. Molecular docking, spectral data-activity relationship, and structure-activity relationship models for potent and weak inhibitors of cytochrome P450 CYP3A4 isozyme. Molecules 17(3):3407
19. Hong H, Fang H, Xie Q, Perkins R, Sheehan DM, Tong W (2003) Comparative molecular field analysis (CoMFA) model using a large diverse set of natural, synthetic and environmental chemicals for binding to the androgen receptor. SAR QSAR Environ Res 14(5–6):373–388
20. Hong H, Tong W, Fang H, Shi L, Xie Q, Wu J et al (2002) Prediction of estrogen receptor binding for 58,000 chemicals using an integrated system of a tree-based model with structural alerts. Environ Health Perspect 110(1):29–36
21. Tong W, Hong H, Fang H, Xie Q, Perkins R (2003) Decision forest: combining the predictions of multiple independent decision tree models. J Chem Inf Comput Sci 43(2): 525–531
22. Tong W, Xie Q, Hong H, Shi L, Fang H, Perkins R et al (2004) Using decision forest to classify prostate cancer samples on the basis of SELDI-TOF MS data: assessing chance correlation and prediction confidence. Environ Health Perspect 112(16):1622–1627
23. Hong H, Tong W, Perkins R, Fang H, Xie Q, Shi L (2004) Multiclass decision forest-a novel pattern recognition method for multiclass classification in microarray data analysis. DNA Cell Biol 23(10):685–694
24. Xie Q, Ratnasinghe LD, Hong H, Perkins R, Tang ZZ, Hu N et al (2005) Decision forest analysis of 61 single nucleotide polymorphisms in a case-control study of esophageal cancer; a novel method. BMC Bioinformatics 6 (Suppl 2):S4
25. Hong H, Tong W, Xie Q, Fang H, Perkins R (2005) An in silico ensemble method for lead discovery: decision forest. SAR QSAR Environ Res 16(4):339–347
26. McPhail B, Tie Y, Hong H, Pearce BA, Schnackenberg LK, Ge W et al (2012) Modeling chemical interaction profiles: I. Spectral data-activity relationship and structure-activity relationship models for inhibitors and non-inhibitors of cytochrome P450 CYP3A4 and CYP2D6 isozymes. Molecules 17(3):3383–3406
27. Chen M, Hong H, Fang H, Kelly R, Zhou G, Borlak J et al (2013) Quantitative structure-activity relationship models for predicting drug-induced liver injury based on FDA-approved drug labeling annotation and using a large collection of drugs. Toxicol Sci 136(1): 242–249
28. Liu J, Mansouri K, Judson RS, Martin MT, Hong H, Chen M et al (2015) Predicting hepatotoxicity using toxcast in vitro bioactivity and chemical structure. Chem Res Toxicol 28(4):738–751
29. Blair RM, Fang H, Branham WS, Hass BS, Dial SL, Moland CL et al (2000) The estrogen receptor relative binding affinities of 188 natural and xenochemicals: structural diversity of ligands. Toxicol Sci 54(1):138–153
30. Fang H, Tong W, Branham WS, Moland CL, Dial SL, Hong H et al (2003) Study of 202 natural, synthetic, and environmental chemicals for binding to the androgen receptor. Chem Res Toxicol 16(10):1338–1358
31. Zhang M, Chen M, Tong W (2012) Is toxicogenomics a more reliable and sensitive biomarker than conventional indicators from rats to predict drug-induced liver injury in humans? Chem Res Toxicol 25(1):122–129
32. Chen M, Bisgin H, Tong L, Hong H, Fang H, Borlak J et al (2014) Toward predictive models for drug-induced liver injury in humans: are we there yet? Biomark Med 8(2):201–213
33. Chen M, Borlak J, Tong W (2014) Predicting idiosyncratic drug-induced liver injury – some recent advances. Expert Rev Gastroenterol Hepatol 8(7):721–723

34. Chen M, Zhang J, Wang Y, Liu Z, Kelly R, Zhou G et al (2013) The liver toxicity knowledge base: a systems approach to a complex end point. Clin Pharmacol Ther 93(5): 409–412
35. Chen M, Vijay V, Shi Q, Liu Z, Fang H, Tong W (2011) FDA-approved drug labeling for the study of drug-induced liver injury. Drug Discov Today 16(15–16):697–703
36. Chen M, Zhang M, Borlak J, Tong W (2012) A decade of toxicogenomic research and its contribution to toxicological science. Toxicol Sci 130(2):217–228
37. Chen M, Borlak J, Tong W (2013) High lipophilicity and high daily dose of oral medications are associated with significant risk for drug-induced liver injury. Hepatology 58(1):388–396
38. Yu K, Geng X, Chen M, Zhang J, Wang B, Ilic K et al (2014) High daily dose and being a substrate of cytochrome P450 enzymes are two important predictors of drug-induced liver injury. Drug Metab Dispos 42(4): 744–750
39. Chen M, Tung CW, Shi Q, Guo L, Shi L, Fang H et al (2014) A testing strategy to predict risk for drug-induced liver injury in humans using high-content screen assays and the "rule-of-two" model. Arch Toxicol 88(7):1439–1449
40. Chen M, Shi L, Kelly R, Perkins R, Fang H, Tong W (2011) Selecting a single model or combining multiple models for microarray-based classifier development? – a comparative analysis based on large and diverse datasets generated from the MAQC-II project. BMC Bioinformatics 12(Suppl 10):S3
41. Yang Z-Z, Wang C-S (2003) Atom-bond electronegativity equalization method and its applications based on density functional theory. J Theor Comput Chem 2(02):273–299
42. Arulmozhiraja S, Morita M (2004) Structure-activity relationships for the toxicity of polychlorinated dibenzofurans: approach through density functional theory-based descriptors. Chem Res Toxicol 17(3):348–356
43. Liu SS, Cui SH, Yin DQ, Shi YY, Wang LS (2003) QSAR studies on the COX-2 inhibition by 3, 4-diarylcycloxazolones based on MEDV descriptor. Chin J Chem 21(11): 1510–1516
44. Chiu T-L, So S-S (2004) Development of neural network QSPR models for Hansch substituent constants. 1. Method and validations. J Chem Inf Comput Sci 44(1):147–153
45. Chiu T-L, So S-S (2004) Development of neural network QSPR models for hansch substituent constants. 2. Applications in QSAR studies of HIV-1 reverse transcriptase and dihydrofolate reductase inhibitors. J Chem Inf Comput Sci 44(1):154–160
46. Zhihua L, Yuzhang W, Xuejun Q, Yuegang Z, Bing N, Ying W (2002) Use of a novel electro-topological descriptor for the prediction of biological activity of peptide analogues. Lett Pept Sci 9(6):273–281
47. Agrawal V, Mishra K, Khadikar P (2003) Multivariate analysis for modelling some antibacterial agents. Oxid Commun 26(1):14–21
48. McGregor MJ, Pallai PV (1997) Clustering of large databases of compounds: using the MDL "Keys" as structural descriptors. J Chem Inf Comput Sci 37(3):443–448
49. Brown RD, Martin YC (1997) The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding. J Chem Inf Comput Sci 37(1):1–9
50. Matter H, Pötter T (1999) Comparing 3D pharmacophore triplets and 2D fingerprints for selecting diverse compound subsets. J Chem Inf Comput Sci 39(6):1211–1225
51. Hong H, Xie Q, Ge W, Qian F, Fang H, Shi L et al (2008) Mold(2), molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. J Chem Inf Model 48(7): 1337–1344
52. Draper NR, Smith H, Pownell E (1966) Applied regression analysis. Wiley, New York
53. Leffler J, Grunwald E (1963) Rates and equilibrium constants of organic reaction. Wiley, New York
54. Stanton DT, Jurs PC (1990) Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. Anal Chem 62(21):2323–2329
55. Rohrbaugh RH, Jurs PC (1987) Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. Anal Chim Acta 199:99–109
56. Nishihara T, Takatori S, Kitagawa Y, Hori S (2000) Estrogenic activities of 517 chemicals by yeast two-hybrid assay. J Health Sci
57. Safe S, Kim K (2008) Non-classical genomic estrogen receptor (ER)/specificity protein and ER/activating protein-1 signaling pathways. J Mol Endocrinol 41(5):263–275
58. Lathe R, Kotelevtsev Y (2014) Steroid signaling: ligand-binding promiscuity, molecular symmetry, and the need for gating. Steroids 82c:14–22
59. Ng HW, Perkins R, Tong W, Hong H (2014) Versatility or promiscuity: the estrogen receptors, control of ligand selectivity and an update on subtype selective ligands. Int J Env Res Public Health 11(9):8709–8742

60. Schug TT, Janesick A, Blumberg B, Heindel JJ (2011) Endocrine disrupting chemicals and disease susceptibility. J Steroid Biochem Mol Biol 127(3–5):204–215
61. Falconer IR, Chapman HF, Moore MR, Ranmuthugala G (2006) Endocrine-disrupting compounds: a review of their challenge to sustainable and safe water supply and water reuse. Environ Toxicol 21(2):181–191
62. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 3(11): 935–949
63. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. J Med Chem 49(23):6789–6801
64. Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? Nat Rev Drug Discov 3(8):711–716
65. Merlot C (2010) Computational toxicology – a tool for early safety evaluation. Drug Discov Today 15(1):16–22
66. Xu JJ, Kalgutkar AS, Will Y, Dykens J, Tengstrand E, Hsieh F (2009) Predicting drug-induced hepatotoxicity in vitro, in silico and in vivo approach. In: Faller B, Urban L (eds) Hit and lead profiling. pp 345–85
67. Greene N, Fisk L, Naven RT, Note RR, Patel ML, Pelletier DJ (2010) Developing structure–activity relationships for the prediction of hepatotoxicity. Chem Res Toxicol 23(7):1215–1222
68. Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Chapter 19

A Round Trip from Medicinal Chemistry to Predictive Toxicology

Giuseppe Felice Mangiatordi, Angelo Carotti, Ettore Novellino, and Orazio Nicolotti

Abstract

Predictive toxicology is a new emerging multifaceted research field aimed at protecting human health and environment from risks posed by chemicals. Such issue is of extreme public relevance and requires a multidisciplinary approach where the experience in medicinal chemistry is of utmost importance. Herein, we will survey some basic recommendations to gather good data and then will review three recent case studies to show how strategies of ligand- and structure-based molecular design, widely applied in medicinal chemistry, can be adapted to meet the more restrictive scientific and regulatory goals of predictive toxicology. In particular, we will report:

- Docking-based classification models to predict the estrogenic potentials of chemicals.
- Predicting the bioconcentration factor using biokinetics descriptors.
- Modeling oral sub-chronic toxicity using a customized k-nearest neighbors (k-NN) approach.

Key words Docking-based classification models, Estrogenic potentials of chemicals, Bioconcentration factor, Biokinetics descriptors, Oral sub-chronic toxicity

1 Introduction

Predicting the effects of xenobiotics, not solely drugs, is far from being a winning bet. Their interplay with living organisms is in fact responsible for biological/toxicological actions which are often not easy to predict. On the other hand, predictions can be made on the basis of (a) in vivo experiments based on direct animal testing, (b) in vitro experiments making use of tissue culture cells, and (c) in silico simulations by employing computer models. It is widely acknowledged that in vivo and in vitro experiments are time demanding and expensive. Great efforts have been thus directed to develop in silico approaches. Such computational strategies allow a

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology 1425, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_19, © Springer Science+Business Media New York 2016

significant save in terms of money, time, and, above all, laboratory animals and provide reliable toxicological evidence in order to minimize or replace in vivo assays according to the "three Rs" principle (replacement, reduction, refinement) [1]. In our opinion, computational methods are thus complementary to experimentation and prospectively capable of replacing empirical testing. The tendency is thus that of moving from experiments to exploratory toxicology which can provide timely go/no-go decisions and represents a viable alternative for the prediction of biological/toxicological effects [2, 3].

In the present survey, we will review some ad hoc examples taken from our recent studies showing how adapting consolidated drug discovery strategies to the scientific and regulatory goals of exploratory toxicology. First of all, we will emphasize the importance of having high-quality data to ensure the derivation of trustable models. In this respect, some practical recommendations will be given. Then, we will discuss how applying molecular docking, perhaps the most popular structure-based method employed by medicinal chemists, to obtain classifiers for discerning estrogenic from non-estrogenic substances. In the second case studies, we will present how QSAR models can be derived and applied to predict the bioconcentration factor, a relevant ecotoxicological endpoint. In this respect, attention will be paid to the appropriate use of biokinetics descriptors and to the definition of the applicability domain to ensure both model transparency and adequacy. Finally, we will describe how customizing a k-NN algorithm to properly model oral sub-chronic toxicity. We will show how the implementation of user-adjustable rules can be very effective to increase the confidence in data prediction, which is the ultimate aim of computational toxicology.

2 Looking for High-Quality Data: Some Practical Recommendations

The advent of new regulations concerning the protection of human health and environment has strengthened the role of QSAR. Such methodology has today assumed the *status* of a mature discipline for both scientific and regulatory purposes. The pressing need of regulatory bodies and industries for the derivation of adequate QSAR models has led to issue some best practices, which are, at present, key elements for successful predictive in silico toxicological studies. Some seminal papers [4–6] have clearly demonstrated that the predictive potential of QSAR models is mostly dependent from the quality of chemical descriptors rather than from the sophistication of the employed optimization techniques. A high-quality data is therefore essential for obtaining trustable models. In this respect, several preliminary checks need to be taken into

account for steering away from even small structural mistakes whose occurrence can result in inaccurate molecular descriptors, which in the end are responsible for disappointing predictions. To circumvent this pitfall [7], great attention has been given to the data curation, a preprocessing treatment necessary to discard or amend chemical records, which are difficult to handle with conventional cheminformatics techniques. Normally, data curation is applied to filter out inorganic and organometallic compounds, counterions, salts, and mixtures. In addition, data curation is carried out to standardize the ring aromatization, to uniform specific chemotypes, to assign tautomeric forms, and to remove duplicates.

Since model reliability is strictly dependent on data quality (i.e., garbage in, garbage out), QSAR developers should also pay high attention in appropriately sizing the dataset and in fairly balancing structural classes or categories, which in real-life investigations are often unevenly represented. It would be advisable that the number of compounds in the dataset should not be too small since this could lead to the occurrence of chance correlation and overfitting; both these phenomena can deteriorate the real predictive power of models. Moreover, a small-sized dataset would be unsuitable for validation analyses. On the other hand, there is not an upper limit to define a maximum size. In this respect, a key role is played by the algorithm implemented for deriving QSAR as well as by the available resources (e.g., computer and time). For practical reasons, a too large dataset can be reduced by selecting a given subset of chemically diverse compounds or can be partitioned in clusters from which deriving multiple and independent models. However, some golden rules should be observed to split the initial dataset into a training set for model derivation and into a test and external set for model validation. In case of continuous response variables, at least 40 compounds should be considered: 20 compounds in the training set and 10 compounds in both test and external sets. Moreover, the response variables should cover a range at least five times larger than the experimental error and should be fairly distributed over such entire range. In case of classification or category response variables, at least 20 compounds per class are recommended: the training set should be made of no less than 10 compounds per class while test and external sets no less than 5 each.

Another reason of attrition in QSAR derivation is given by compounds, which are typical chemical singletons, being their structural features far away from those of all the other compounds within a dataset. In other words, they could behave as leverage (or structural) outliers. Other compounds could instead act as activity outliers as they rebut the basic QSAR assumption stating that similar compounds have similar properties. As reported in a number of

seminal works [8, 9], these compounds could originate the so-called cliffs of the descriptor space where a given response property (i.e., biological/toxicological response) changes dramatically for an even subtle structural variation. Actually, both these types of outliers can be real or sometimes due to accidental errors in reporting the chemical structure or in annotating the response variable. Normally, it is wise to remove them prior to model derivation as they will likely cause model instability and deeply affect predictions.

Moreover, high-quality molecular descriptors are essential to derive predictive and interpretable QSAR models [10]. Nowadays, it is quite easy to quickly calculate an overwhelming number of descriptors [11] related to two- or three-dimensional molecular aspects, although their mechanistic interpretation remains somewhat obscure to mid-level QSAR practitioners. Needless to say that medicinal chemists have long debated about chemical desirability, a concept inherent to the chemical meaning of QSAR model [12, 13]. We can guess that descriptors referring to the passage of xenobiotics across cellular membranes, for instance, may be desirable in a toxicological context. In this respect, we do believe that ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties would make the descriptors space more attractive for toxicological purposes and of adequate transparency for molecular and numerical modeling. ADMET properties are in fact important to study the fate and disposition of drugs and to monitor their behavior in the body at therapeutic doses (i.e., pharmacokinetic properties). Importantly, the studies of ADMET properties are not limited to drugs but can be extended to any chemical, including environmental pollutants, potentially affecting human health. In this respect, the term toxicokinetics and, even better, the more inclusive term biokinetics [14] are normally used to describe and, then, to predict unwanted toxic effects of xenobiotics on living system exposed to chemicals at any dosage regimen. The masterpiece by Waterbeemd [15] describes the progress made by medicinal chemistry in the attempt of refining ADMET properties in order to reduce the costly late-stage failures in drug development and thereby accelerating the drug discovery process. Such efforts have resulted in the wide introduction of ADMET-related descriptors implemented in in silico methods to predict the most relevant pharmacokinetic, metabolic, and toxicity endpoints.

3 Docking-Based Classification Models to Predict the Estrogenic Potentials of Chemicals

Predicting the endocrine disruptor potential of chemicals and, more specifically, their ability to interfere with the estrogen receptors (ERs) is a theme of utmost relevance [16]. Unlike previous

predictive models [17–19], we have recently described how the current availability of X-ray-solved target structures can be employed [20]. Importantly, accounting for physicochemical information on the biological target allows a larger applicability domain with respect to classical QSAR-like models.

We used a three-dimensional (3D) training dataset (hereafter referred to as EPA-ERDB) consisting of 1677 chemical structures shared by US EPA. For each chemical, the estrogenic/nonestrogenic action was derived from concentration-response data resulting from 18 high-throughput assays probing several sites of the mammalian ER pathway. Challengingly, the 1677 chemicals were unevenly distributed, being only 237 (14.13 %) chemicals designated as ER binders. To possibly cover a broader spectrum of possible biological actions of compounds comprised within the EPA-ERDB training dataset, eight ER crystal structures were retrieved from the Protein Data Bank (PDB) for docking simulations. All four possible ER classes were considered: (1) ERα bound to agonist, (2) ERα bound to antagonist, (3) ERβ bound to agonist, and (4) ERβ bound to antagonist. The 3D conformations of the 1677 chemicals in the training dataset were subjected to docking simulations performed by both GLIDE v.6.5 [21] and GOLD v.5.2 [22], two very popular software largely adopted in drug discovery projects. The ability of the selected docking protocols to discern binders from nonbinders was assessed using typical confusion matrix, which includes information about experimental and predicted matches and mismatches returned for each classification system. Next, docking performance was evaluated using the enrichment factor (EF), which represents the percentage of known binders found at a given percentage of the ranked database. In addition, we reported the EF at the early 1 % of the ranked dataset (i.e., EF1%). Predictive docking-based classification models are expected to return similar values for both EF1% and EFmax (a reference ideal value obtained by dividing the total number of chemicals by the total number of binders). All these data were derived from the obtained receiver operating characteristic (ROC) curves (*see* Fig. 1). The thresholds for defining the classes were set on the basis of the desired sensitivity (SE) values. The value of SE estimates the proportion of true positives that are correctly identified. In order to designate the estrogenic or nonestrogenic potential, two SE values equal to 0.25 and to 0.75 were set as thresholds to define, for each ER crystal, three probability binding classes as follows:

(a) SE $\leq$0.25, the class with high probability of binding (i.e., binder molecules).

(b) SE >0.75, the class with low probability of binding (i.e., nonbinder molecules).

(c) $0.25 < SE \leq 0.75$, the class with medium probability of binding (i.e., suspicious molecules).

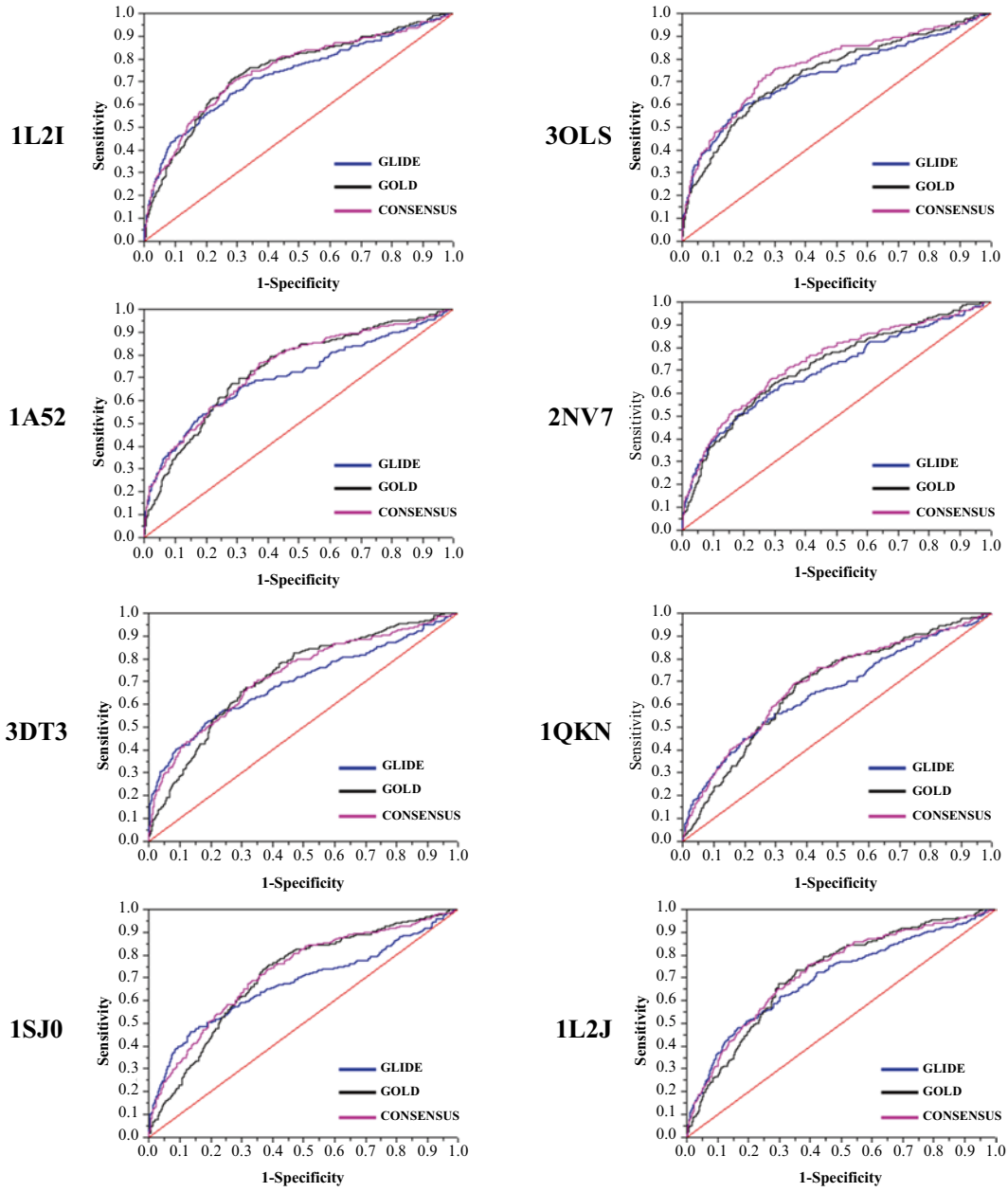


Fig. 1 ROC curves derived from ERα (PDB entries: 1L2I, 1A52, 3DT3, and 1SJ0) and ERβ structures (PDB entries: 3OLS, 2NV7, 1QKN, and 1L2J) are shown on the *left* and *right hand side*, respectively (taken from 20)

At a given threshold, the goodness of the classification was assessed using two parameters: (a) the positive predictive value (PPV) that is related to the probability that a chemical predicted as a binder (over-threshold) is actually a binder and (b) the negative predictive value (NPV) that is related to the probability that a

chemical predicted as a nonbinder (under-threshold) is actually a nonbinder. However, the pronounced asymmetry of data prompted us to compute the positive (+LR) and the negative likelihood ratio (−LR) for each of the SE-considered thresholds. Briefly, the greater the +LR is at a given threshold, the better the performance of the classification model. It is worthy to say that these likelihood ratios are independent from the data distribution within the training set.

We observed that, unequivocally, GLIDE detects a higher number of binders in the earliest fraction of the rank despite the lower AUC values. For all ER crystal structures, the ability to minimize FPs is higher with GLIDE with respect to GOLD, in agreement with the already discussed EF1% factors. Importantly, an opposite trend can be detected if the second threshold (SE = 0.75) is considered. GOLD returns PPV values higher than GLIDE. In other words, GLIDE ensures better performances in terms of ability to minimize FPs, whereas the interest is mostly oriented to the upper part of the ranking. Our results would suggest that the use of GLIDE or GOLD depends on the pursued goals. As shown, there is not a winning model, but rather a case-by-case evaluation should be made. Docking-based classification models have allowed to employ the wealth of physicochemical information contained in the native protein structures to screen large chemical collections and demonstrated to be helpful for immediately obtaining a preliminary idea of the estrogenic activity by simply comparing the docking score of a target chemical with those reported at the different SE-based thresholds.

4 Predicting the Bioconcentration Factor Using Biokinetics Descriptors

The bioconcentration factor (BCF) represents the ratio of the concentration of a substance in an aquatic organism with respect to that in water [23]. It is an endpoint of utmost relevance due to its costs and its (eco)toxicological impact. Its assessment should be done following the experimental test OECD 305, which requires for each substance more than hundreds of fishes, months for test execution, and tens of thousands of Euros [24]. The herein used data [25] comprises 851 chemicals, retrieved from the ANTARES dataset. The obtained dataset was split into three subsets: about 10 % (78 out of 851) of the compounds were randomly selected to form the blind set (BS), required for final validation. The remaining chemicals were split to ensure a uniform distribution of their experimental BCF values, applying the Venetian blinds method [26], to form training set (TS) and validation set (VS) containing 620 and 153 chemicals, respectively. These selection criteria were used to obtain two different and independent sets for model validation and to ensure the most realistic situation for the external compounds, so that statistics could explain the real capability of the model to predict new compounds, as it should be for regulatory purposes.

Many commercial and free software programs are available for the calculation of thousands of two-dimensional (2D) or three-dimensional (3D) descriptors. In the present work, we preferred to calculate a smaller number (i.e., 51) of ADMET (absorption, distribution, metabolism, excretion, and toxicity)-relevant descriptors that are closely related to pharmaceutical properties of organic molecules. To this end, we used QikProp 3.4 [27] included in Schrödinger 2011-1 suite [28]. Note that, as already mentioned, descriptors referring to the permeation of the membrane may be more desirable for a toxicological or pharmacological audience. A number of models were derived using the Monte Carlo approach (simulated annealing), multiple linear regression (MLR), and neural network algorithm (NN). Importantly, the obtained models could be flexibly adapted to play as classifiers using as thresholds those established in Annex XIII of REACH to classify chemicals. All substances that exceed the first threshold of log BCF = 3.3 are classified as bioaccumulative (B), while those having log BCF < 3.3 are classified as nonbioaccumulative (nB) according to the PBT (persistent, bioaccumulative, and toxic) definition; on the other hand, all substances that exceed the second threshold of log BCF = 3.7 are classified as very bioaccumulative (vB).

Among others, our attention was mostly engaged by a nine-descriptor model. Apart from robust statistics, particular attention was paid to the definition of the applicability domain (AD). Needless to say that predictions provided by models without a clearly defined AD are meaningless [29–31]. As previously described, its importance has also been remarked in REACH Annex XI, BPR Annex IV, and OECD principles for the derivation of acceptable QSARs. In our studies, we implemented a multi-step filter system to confidently designate chemicals within the AD only those having the matching criteria requested at any step. Such procedure ensures higher confidence and transparency irrespective of the accuracy of predictions [32].

The first independent filter accounted for the dataset structural diversity. Briefly, the occurrence of organic functional group (nested) was assessed using the QSAR Toolbox 3.0 software, released by OECD—2013. The second independent filter accounted for the chemical descriptors range. The minimum and maximum values of the nine descriptors in the model for TS chemicals were used a criterion of interval validity. In this respect, VS or BS chemicals whose descriptors violated even only one range were placed outside AD. The third independent filter was a geometrical trap based on the interpolation region space representing the smallest convex area whose borders describe the perimeter of a polygon containing TS compounds. In particular, the interpolation polygon was drawn using spatial coordinates of the first two principal components of the multivariate descriptor space of the nine-term model. The polygon area was reduced to include the top 98 % TS compounds (considering their closeness to TS centroid)

to avoid the inclusion of underrepresented areas likely increasing the prediction uncertainty. Finally, the leverage method was applied as fourth independent filter. Briefly, the leverage represents the compound distance from the model experimental space (that is the center of TS observations) and, thus, provides a measure of the degree of influence that a particular TS chemical structure has on the model or the degree of extrapolation for the prediction of VS and BS compounds. In this respect, VS and BS compounds having leverages exceeding the widely acknowledged threshold of $h^* = 3p'/n$ (where p' is the number of model variables plus one and n is the number of TS compounds) were placed outside model AD being poor reliable predictable [33].

The simultaneous application of multi-filter system has the effect of leaving outside AD: (a) a number of 20 (13 % of the initial) VS compounds with an indirect gain of r^2 from 0.635 to 0.765 and of RMSE from 0.794 to 0.616 and (b) a number of 7 (9 % of the initial) BS compounds with an indirect gain of r^2 from 0.623 to 0.659 and of RMSE from 0.841 to 0.817 (*see* Fig. 2).

The harmonic application of consolidated QSAR approaches employing pharmaceutically relevant descriptors and a multi-step filter system to designate chemicals inside/outside AD demonstrated to be very effective for modeling BCF data, an endpoint of utmost importance in both toxicological and regulatory terms.

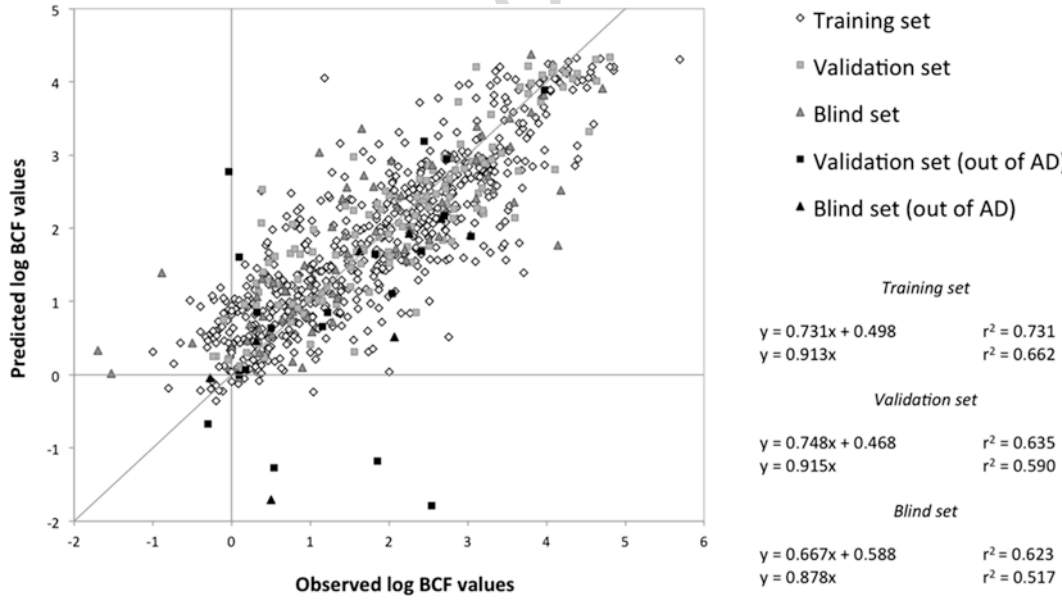


Fig. 2 Comparison of the experimental and predicted log BCF values obtained through the nine-descriptor BCF model. TS, VS, and BS chemicals are represented by *white diamonds*, *gray squares*, and *upside triangles*, respectively. VS and BS outside AD chemicals are represented by *black squares* and *upside triangles*, respectively. The continuous line represents the case of ideal correlation (taken from 25)

5 Modeling Oral Sub-chronic Toxicity Using a Customized k-Nearest Neighbors (k-NN) Approach

Repeated dose toxicity (RDT) is an important endpoint to toxicologically profile a given chemical after repeated administration. RDT studies are focused on the no observed (adverse) effect level (NO(A)EL) and on the lowest observed (adverse) effect level (LO(A)EL). The former is the higher experimental dose at which there is no appreciable response [34]; the latter indicates the lowest dosage at which adverse effects occur in comparison with a control group (e.g., onset of an adverse effect) [35]. The NO(A)EL and LO(A)EL are assessed by means of in vivo studies that can be based on various protocols accounting for different exposure period, animal model (rodent or non-rodent species) and exposure route (oral, inhalation or dermal) [36]. As a result, regulators explicitly require data relative to repeated dose toxicity.

We recently conducted a toxicological study [37] focused on RDT data for sub-chronic oral exposure (i.e., 90 days) in rats. Training set data was retrieved from different sources (i.e., Munro database, Hazard Evaluation Support System, EPA's Integrated Risk Information System). In particular, 254 chemicals were selected being the ones having unequivocal values of chronic toxicity studies (from 84 to 98 days) of oral exposure (gavage, diet, or drinking water). An external dataset comprising 179 chemicals was also used to challenge the predictive power of our models. External dataset data were taken and properly selected from the RepDose database.

A customized k-nearest neighbors (k-NN) approach for predicting sub-chronic oral toxicity in rats was used (*see* Fig. 3). The basic idea was that of predicting a given response on the basis of those observed in the most structurally similar chemicals. The straight application of the k-NN was however very disappointing. To overcome this limitation, the algorithm was ad hoc adapted by implementing several rules to better control the reliability of predicted chemicals. The gain in prediction and confidence was obtained for a given percentage of the dataset; the reasonable price to pay was that a number of compounds (those unmatching the new rules implemented in the k-NN) were left unpredicted as a precautionary measure. However, the use of restrictive conditions in modeling such a complex endpoint meets both the scientific and regulatory purposes established by international bodies for the protection of human health. In fact, providing few but highly reliable predictions represents a valuable prioritization strategy to generate trustable toxicological information on the substances and, at the same time, to support the use of alternative methods and thus to reduce the number of animals needed for in vivo testing.

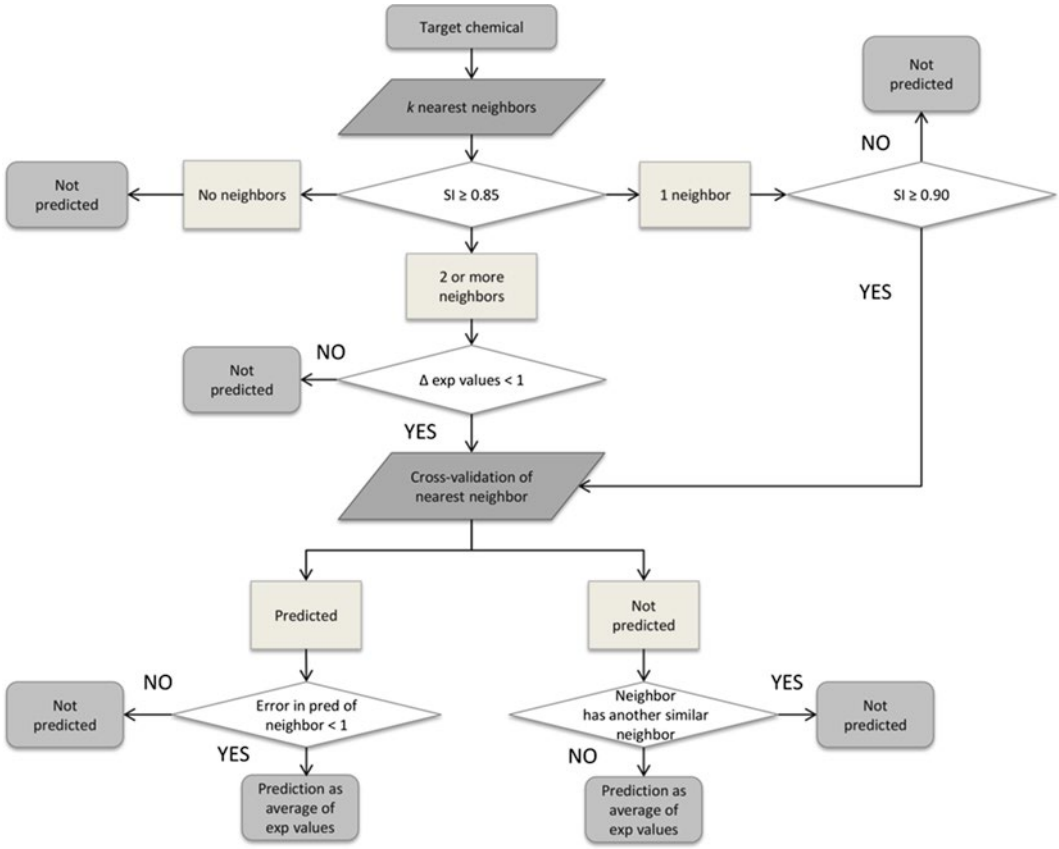


Fig. 3 Flowchart for the selection of the output predictions. SI or similarity index between the target chemical and its nearest neighbors; Δ exp values are the difference between experimental values of nearest neighbors; error in pred is the error in prediction returned in cross validation of a neighbor in the TS (taken from 37)

6 Conclusions and Perspectives

Exploratory toxicology employs in silico methods for their importance in scientific and regulatory context. Indeed, the need of protecting human health and environment has prompted public authorities, such as the US Environmental Protection Agency (US EPA) and the European Chemicals Agency (ECHA) to play a frontline role in the promotion of programs of predictive toxicology to assess the risk posed by chemicals. For instance, the European Commission (EC) has issued, in Annex XI of REACH and Annex IV of BPR, four conditions for using in silico in place of in vivo testing: (1) results have to be derived from a QSAR model whose scientific validity has been well established, (2) the substances are expected to fall within the applicability domain of the QSAR model, (3) results need to be adequate for the purpose of classification and labeling and/or risk assessment, and (4) adequate and reliable documentation of the applied method has to be

provided. Importantly, these recommendations for the implementation of the so-called non-testing methods are perfectly known to medicinal chemists, whose community is continuously discussing roles and goals. It is well known that medicinal chemists have in recent years already openly deplored the frequent temptation of discussing highly speculative computational predictions that are often the result of over-interpreted but not properly validated models. In this respect, a blacklist of simply decorative and colorful QSAR models has been matter of a strong skepticism, as recently pointed out by Cramer [38]. In this continuing debate, we do believe that modern medicinal chemists should be strongly committed to face the new challenge of exploratory toxicology, which implies more restrictive scientific and regulatory purposes (i.e., chemical prioritization, selecting compounds for further experimental testing, reducing the number of false negatives, harmful compounds predicted as safe). By discussing three case studies, we reported how successfully adapting consolidated structure- and ligand-based strategies, largely applied in drug discovery programs, to the goal of exploratory toxicology. Needless to repeat that a critical case-by-case assessment is necessary to prove the result reliability and to make trustable the adopted approach. Indeed, an informed interpretation of the results can make the difference. However, we are just at the beginning of a new fascinating journey requiring new scientific efforts and challenges.

References

1. Russell WMS, Burch RL (1959) The principles of humane experimental technique. Johns Hopkins Bloom Sch Public Health, Baltimore
2. Hornberg JJ, Laursen M, Brenden N et al (2014) Exploratory toxicology as an integrated part of drug discovery. Part I: why and how. Drug Discov Today 19:1131–1136
3. Nicolotti O, Benfenati E, Carotti A et al (2014) REACH and in silico methods: an attractive opportunity for medicinal chemists. Drug Discov Today 19:1757–1768
4. Young D, Martin T, Venkatapathy R et al (2008) Are the chemical structures in your QSAR correct? QSAR Comb Sci 27:1337–1345
5. Zhu H, Tropsha A, Fourches D et al (2008) Combinatorial QSAR modeling of chemical toxicants tested against Tetrahymena pyriformis. J Chem Inf Model 48:766–784
6. Tetko IV, Sushko I, Pandey AK et al (2008) Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection. J Chem Inf Model 48:1733–1746
7. Fourches D, Muratov E, Tropsha A (2010) Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. J Chem Inf Model 50:1189–1204
8. Maggiora GM (2006) On outliers and activity cliffs—why QSAR often disappoints. J Chem Inf Model 46:1535
9. Cruz-Monteagudo M, Medina-Franco JL, Pérez-Castillo Y et al (2014) Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? Drug Discov Today 19:1069–1080
10. Nicolotti O, Carotti A (2006) QSAR and QSPR studies of a highly structured physicochemical domain. J Chem Inf Model 46: 264–276
11. Todeschini R, Consonni V (eds) (2000) Handbook of molecular descriptors. Wiley, Weinheim
12. Nicolotti O, Gillet VJ, Fleming PJ et al (2002) Multiobjective optimization in quantitative structure-activity relationships: deriving accurate and interpretable QSARs. J Med Chem 45:5069–5080

13. Bickerton GR, Paolini GV, Besnard J et al (2012) Quantifying the chemical beauty of drugs. Nat Chem 4:90–98
14. Bouvier d'Yvoire M, Prieto P, Blaauboer BJ et al (2007) Physiologically-based Kinetic Modelling (PBK Modelling): meeting the 3Rs agenda. The report and recommendations of ECVAM Workshop 63. Altern Lab Anim ATLA 35:661–671
15. Van de Waterbeemd H, Gifford E (2003) ADMET in silico modelling: towards prediction paradise? Nat Rev Drug Discov 2:192–204
16. Diamanti-Kandarakis E, Bourguignon J-P, Giudice LC et al (2009) Endocrine-disrupting chemicals: an endocrine society scientific statement. Endocr Rev 30:293–342
17. Shi LM, Fang H, Tong W et al (2001) QSAR models using a large diverse set of estrogens. J Chem Inf Comput Sci 41:186–195
18. Devillers J, Marchand-Geneste N, Carpy A et al (2006) SAR and QSAR modeling of endocrine disruptors. SAR QSAR Environ Res 17:393–412
19. Jacobs MN (2004) In silico tools to aid risk assessment of endocrine disrupting chemicals. Toxicology 205:43–53
20. Trisciuzzi D, Alberga D, Mansouri K et al (2015) Docking-based classification models for exploratory toxicology studies on high-quality estrogenic experimental data. Future Med Chem 7:1921–1936. doi:10.4155/FMC.15.103
21. Small-molecule drug discovery suite 2014-4: glide, version 6.5 (2014) Schrödinger, LLC, New York
22. Jones G, Willett P, Glen RC et al (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267:727–748
23. Arnot JA, Gobas FA (2006) A review of bioconcentration factor (BCF) and bioaccumulation factor (BAF) assessments for organic chemicals in aquatic organisms. Environ Rev 14:257–297
24. OECD (2012) Test no. 305: bioaccumulation in fish: aqueous and dietary exposure, OECD guidelines for the testing of chemicals, section 3. OECD Publishing, 12.10.12, 72. (http://dx.doi.org/10.1787/9789264185296-en). / http://dx.doi.org/10.1787/9789264185296-enS. Consulted July 2015
25. Gissi A, Gadaleta D, Floris M et al (2014) An alternative QSAR-based approach for predicting the bioconcentration factor for regulatory purposes. ALTEX 31:23–36
26. Consonni V, Ballabio D, Todeschini R (2009) Comments on the Definition of the Q2 Parameter for QSAR Validation. J Chem Inf Model 49:1669–1678
27. Schrödinger Release 2011-1: QikProp, version 3.4 (2011) Schrödinger, LLC, New York
28. Schrödinger Release 2011-1 (2011) Schrödinger, LLC, New York
29. Jaworska J, Nikolova-Jeliazkova N, Aldenberg T (2005) QSAR applicability domain estimation by projection of the training set descriptor space: a review. Altern Lab Anim ATLA 33:445–459
30. Aptula AO, Roberts DW (2006) Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. Chem Res Toxicol 19:1097–1105
31. Roberts DW, Aptula AO, Patlewicz G (2006) Mechanistic applicability domains for nonanimal based prediction of toxicological endpoints. QSAR analysis of the Schiff base applicability domain for skin sensitization. Chem Res Toxicol 19:1228–1233
32. Schultz TW, Hewitt M, Netzeva TI et al (2007) Assessing applicability domains of toxicological QSARs: definition, confidence in predicted values, and the role of mechanisms of action. QSAR Comb Sci 26:238–254
33. Gramatica P (2010) Chemometric methods and theoretical molecular descriptors in predictive QSAR modeling of the environmental behavior of organic pollutants. In: Puzyn T, Leszczynski J, Cronin MT (eds) Recent advances in QSAR studies. Springer, Netherlands, pp 327–366
34. Sand S, Victorin K, Filipsson AF (2008) The current state of knowledge on the use of the benchmark dose concept in risk assessment. J Appl Toxicol 28:405–421
35. Sakuratani Y, Zhang HQ, Nishikawa S et al (2013) Hazard Evaluation Support System (HESS) for predicting repeated dose toxicity using toxicological categories. SAR QSAR Environ Res 24:351–363
36. SCCS—Scientific Committee on Consumer Safety (2012). The SCCS's notes of guidance for the testing of cosmetics substances and their safety evaluation 8th revision. (http://ec.europa.eu/health/scientific_committees/consumer_safety/docs/ sccs_s_006.pdf). Consulted April 2014
37. Gadaleta D, Pizzo F, Lombardo A et al (2014) A k-NN algorithm for predicting the oral subchronic toxicity in the rat. ALTEX 31:423–432
38. Cramer RD (2012) The inevitable QSAR renaissance. J Comput Aided Mol Des 26:35–38

Chapter 20

The Use of In Silico Models Within a Large Pharmaceutical Company

Alessandro Brigo and Wolfgang Muster

Abstract

The present contribution describes how in silico models are applied at different stages of the drug discovery process in the pharmaceutical industry. A thorough description of the most relevant computational methods and tools is given along with an in-depth evaluation of their performance in the context of potential genotoxic impurities assessment.

The challenges of predicting the outcome of highly complex studies are discussed followed by considerations on how novel ways to manage, store, share and analyze data may advance knowledge and facilitate modeling efforts.

Key words Drug discovery, Genotoxicity, TTC, Lead optimization

1 Introduction

Computational methods (in silico models) are widely used in the pharmaceutical industry for optimizing molecules during early drug development, not only for efficacy, but in parallel with regard to their toxicological as well as drug disposition properties. It is the fine balance of target potency, selectivity, favorable ADME (absorption, distribution, metabolism, excretion), and (pre)clinical safety properties that will ultimately lead to the selection and clinical development of a potential new drug [1, 2]. As a clinical candidate needs rigorous preclinical optimization in various aspects, multidimensional optimization (MDO) is a term often used to describe the intensive investigations during the first 3–4 years of drug discovery from the identification of the target to the selection of the best drug development compound. The current MDO process comprises the use of in silico, in vitro, as well as in vivo techniques. In general, in silico tools have the intrinsic advantages to be fast and not to need the physical presence of the test compounds and can therefore be applied very early in drug development. Theoretically, in silico models can be developed for all end points

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_20,

and organisms, but the availability of large enough, balanced, and high-quality datasets is the main drawback for reliable predictions. An excellent correlation with the in vitro/in vivo data, that is, high-sensitivity as well as high-specificity, easy-to-use, and easy-to-interpret in silico model, is a key requirement for its usefulness. In the past few years, computational toxicology prediction systems tremendously increased their predictive power for end points like genotoxicity, carcinogenicity, phototoxicity, phospholipidosis, GSH adduct formation, hERG inhibition, and CYP inductions, but still have not achieved the major breakthrough due to lack of sufficiently large datasets covering more complex toxicological end points (e.g., liver-, kidney-, cardiotoxicity). These are the critical toxicity end points, which needs to be addressed in the next years to weed out potential safety issues in the clinics. Recent initiatives and consortia (e.g., IMI/eTOX, ToxCast, and ToxBank) dealing with data sharing of preclinical in vivo toxicology studies and computational approaches have the potential of significantly improving these end point predictions and filling the data gaps [3–5].

This review will outline general considerations on the mainly applied expert systems—rule-based and statistical-based models—in toxicology and ADME for pharmaceuticals and their application in the early drug development process as well as their regulatory impact on the assessment of potential impurities arising in the manufacturing process. Recent improvements and future perspectives on the main challenge of predicting complex in vivo end points will be summarized and discussed.

2 In Silico Methods for the Prediction of Toxicity

As already described in Subheading 1 of this chapter, the thorough characterization of the safety profile of drug candidates is of great importance to ensure that no harm is posed to healthy volunteers and patients during and after clinical development throughout the entire compound lifecycle.

Drug toxicity can manifest itself in a number of ways and may interest one or more target organs or biological processes. In particular, carcinogenicity and liver, renal, cardiovascular, reproductive, and genetic toxicities are among the most significant safety issues that can prevent drug candidates to progress through clinical development or can cause the withdrawal of already marketed products. Overall, between 20 and 30 % of failures can be attributed to safety reasons [6–8].

Over that past few years, predictive computational approaches have found a significant role within drug discovery in helping scientists rank compounds classes and prioritize in vitro and in vivo experiments. A number of factors contributed to the increased importance of in silico methods in drug discovery: (1) wider avail-

ability of high-quality datasets (public domain, focused data sharing initiatives), (2) robust computational models that can provide reliable predictions[9], (3) pressure to reduce animal testing, (4) need to bring new drugs to the market faster and cheaper, (5) legislation on the assessment of potential genotoxic impurities, and (6) greater number of commercially available and open-source software tools.

The most widely used computational methods for the prediction of toxicity end points can be roughly divided into two main categories, rule-based and statistical-based systems, depending on what type of methods they use to make their classifications.

2.1 Rule-Based Systems

Computational tools included in this category store and manipulate knowledge to interpret information. They are often referred to as expert systems, which make use of a set of explicit rules (i.e., not implicitly embedded in a code) to make deductions and classifications. Such systems have the advantage that rules can be easily represented and developed by experts in the field of toxicology (or of any discipline the systems are applied to), rather than by information technology (IT) specialists. In addition, solid expert rules can be derived from limited amounts of data, as long as they are sufficiently representative of specific chemical and biological spaces.

Both commercial and open-source systems are available within the rule-based methodologies, and they include, among others, Derek Nexus [10–13], Toxtree [14], CASE Ultra Expert Rules [15], and Leadscope Expert Alerts System [16].

Derek Nexus is an expert, knowledge base system which contains structural alerts (SAs) and expert knowledge rules (derived from both public and proprietary data by scientists at Lhasa Ltd.) for a wide range of toxicological end points and applies these to make in silico predictions about the toxicity of chemical entities. The knowledge-based expert rules represent knowledge from literature, academic, industrial, and Lhasa Ltd. scientific experts and are regularly updated according to newly available experimental data and publications. In making predictions, the expert rules take into account not only the presence or absence of a structural alert but also the species and a few calculated physicochemical parameters (where applicable) in a process akin to the human-based logic of argumentation. Proprietary data donated, by Lhasa Ltd. members, has been used in the development of approximately 25 % of the bacterial in vitro (Ames test) mutagenicity alerts in Derek Nexus, and proprietary datasets are used to validate the performance of alerts for this, and other end points, to provide an indication of predictive performance within the chemical space of highest interest to users. In addition proprietary and customized alerts can be defined by users and implemented through the Derek Knowledge Editor.

The most recent version of Derek Nexus contains expert-derived functionality to provide negative predictions for bacterial in vitro mutagenicity in order to give more confidence on nonpositive predictions. If a query compound does not match a structural alert for mutagenicity, then it is compared to a Lhasa reference set of Ames test data, and a negative prediction is provided based on the features within the query compound [17]. In case of absence of alerts for end points other than mutagenicity, negative calls should be made with caution as alerts that are not part of the rulebase, hence unknown to the system, can still be relevant in the induction of certain toxicities.

Since Derek is an expert system, it has no training set in a strict sense as in QSAR-based systems, but there are example compounds for the alerts stored in its knowledge base.

Toxtree [14, 18] is a Java-based, freely available, open-source application for toxicity prediction. It was developed by IDEAconsult Ltd. (Sofia, Bulgaria) under the terms of a contract with the European Commission Joint Research Centre. The program is mainly based on structural alerts but also provides QSAR models for distinct chemical classes to refine the predictions. For mutagenicity, Toxtree implements the Benigni-Bossa rulebase [19] for carcinogenicity and mutagenicity. The alerts are only differentiated into genotoxic and a small number of non-genotoxic ones, without distinction between carcinogenicity and mutagenicity. Additionally, this module offers QSAR models for aromatic amines and α,β-unsaturated aldehydes, which should improve the predictivity for these specific chemical classes. However, the mutagenicity QSARs refer to *Salmonella typhimurium* TA100 only. With regard to structures that do not trigger any alert, the same considerations on negative predictions made for Derek Nexus apply.

CASE Ultra Expert Rules: As of version 1.5.2.0 of CASE Ultra, an *expert-rule system* is built using rules from expert knowledge or scientific literature for the prediction of bacterial mutagenicity [15]. A detailed description of the software is given in the section describing the statistical-based systems.

Leadscope Expert Alerts System: Leadscope Inc. produces several software modules applicable in the context of toxicological forecasting, particularly in the field of QSAR models. Recently, Leadscope developed a rule-based expert system for the prediction of mutagenicity, using an extensive high-quality genetic toxicity database containing the results of the bacterial mutagenesis assay along with chemical structures [20]. Firstly, the chemical structures were merged using a chemical registration system to assign a unique identifier to each chemical and merging entries on the basis of this identifier. Next, the graded end points for *Salmonella* and *E. coli* were combined from the different sources, resulting in a database of over 7,000 chemicals each with a positive/negative overall bacterial mutation call. The reference set also covers a

diverse collection of compounds since they have been derived from many different sources, including pharmaceuticals, pesticides, industrial chemicals, and food additives. Clustering led to 1,220 clusters with two or more examples and 1,049 singletons (clusters with one example). Once substructures are identified for alert definitions, the selected alerts are consolidated and organized hierarchically (i.e., parent/child). This helps in establishing a mechanistic explanation particularly where any child alert is lacking or has limited mechanistic information, as it may be inherited from the parent alert. When the expert alerts are used to make prediction, a score is calculated reflecting the precision of the alert [20]. In addition to the primary alert, it is also important to define any factors that would deactivate the alerts as a result of electronic or steric effects or by blocking an important metabolic step. In this context, the Leadscope software identified and quantitatively assessed deactivating factors using the 27,000 predefined structural features in Leadscope and generated new chemical scaffolds associated with negative bacterial mutagenicity. Any deactivating fragments identified were quantitatively evaluated using the reference set.

2.2 Statistical-Based Systems

Quantitative structure-activity relationship (QSAR) models are regression or classification models used in the chemical and biological sciences and other disciplines. Like other regression models, QSAR regression models relate a set of "predictor" variables (X) to the potency of the response variable (Y), while classification QSAR models correlate the predictor variables to a category value of the response variable.

The QSAR approach can be generally described as an application of data analysis methods and statistics to model development that could accurately predict biological activities or properties of compounds based on their structures. Any QSAR method can be generally defined as an application of mathematical and statistical methods to the problem of finding empirical relationships (QSAR models) in the form $\mathrm{Pi} = k'(\mathrm{D1}, \mathrm{D2}, \ldots, \mathrm{Dn})$, where Pi are biological activities (or other properties) of molecules; D1, D2, …, Dn are calculated (or, sometimes, experimentally measured) structural properties (or molecular descriptors) of compounds; k' is some empirically established mathematical transformation that should be applied to descriptors to calculate the property values for all molecules. The goal of QSAR modeling is to establish a trend in the descriptor values, which parallels the trend in biological activity [21].

Both commercial and open-source systems are available within the QSAR-based methodologies, and they include, among others, Sarah Nexus [22], CASE Ultra [15], Leadscope Model Applier [23], OECD Toolbox [24], Bioclipse [25], admetSAR, and Prous Institute Symmetry [26].

Sarah Nexus is a statistical system which utilizes a self-organizing hypothesis network (SOHN) model to generate predictions for mutagenicity [27]. This hierarchical model not only retrieves matching fragments, it also further refines these results by considering the structure's similarity to the query structure. The methodology retains those fragments that are perceived to be of greater value; fragments may be of various sizes and can even overlap, ensuring greater accuracy in predictions. Fragments are generated from the provided training set of molecules and not selected from lists of predetermined fragments. Both global (broad coverage, not adequately sensitive to local variations) and local (more accurate for fragments that fall inside their chemical space, narrower in scope) models are available in Sarah Nexus. If the query structure is not an exact match to a compound within the training set (for which a prediction of 100 % confidence is generated), the structure is fragmented and the software will select the most appropriate model for each fragment.

The structural explanation for the prediction provided by Sarah Nexus is conveyed by highlighting those fragment(s) that the model considers meaningful. Sarah Nexus provides a confidence score and a structural explanation for each prediction alongside direct access to supporting data to aid expert analysis [28].

CASE Ultra: CASE Ultra's algorithm is mainly influenced by the original MCASE methodology [29, 30], a traditional QSAR system, which can automatically generate a predictive model from a training set of non-congeneric compounds with associated biological or toxicity data. The training set ideally should contain examples of both active and inactive chemicals in a non-overly skewed ratio.

CASE Ultra can identify alerts that are not limited to linear paths of limited size or limited branching pattern, and the training sets could be larger than 8,000 molecules [31]. To build a model, CASE Ultra picks up one active chemical at a time from the training set and systematically generates a list of fragments for that chemical. Each fragment's relevance for activity is then determined using a two-objective criteria comprised of Shannon's entropy [32] as a fitness measure and the number of the active training set molecules containing this fragment (fragments that are optimal based on the two objectives, i.e., the ones that cannot be replaced by any other fragment without degrading one or both objectives, are selected and then sorted in descending order of the number of their active chemicals). A top few fragments (based on the aforementioned two-objective criteria, e.g., fragments that have low entropy as well as supported by higher number of active training chemicals) are selected. These fragments are considered as potential positive alerts. The fragment generation procedure produces simple linear chains of varying lengths and branched fragments as

well as complex substructures generated by combining simple fragments. When the algorithm has finished scanning all the active chemicals, a search is made in the accumulated list of the potential positive alerts to find the alert that covers the highest number of active chemicals, and it is added to the final list of positive alerts. This step is repeated until enough positive alerts were identified to cover all the active chemicals in the training set. Once a final set of positive alerts is identified, CASE Ultra attempts to build separate local QSARs for each positive alert in order to explain the variation in activity within the training set chemicals covered by that alert. In addition, deactivating alerts are found using a very similar process but by scanning inactive chemicals and finding fragments that occur mainly in inactive chemicals. This collection of positive and deactivating alerts constitutes a model for a particular end point and can be used for predicting activity in test chemicals. During prediction, a test chemical is scanned against the list of the model's positive and deactivating alerts, and if no positive alerts could be identified in it, the chemical is considered inactive. In general, if the test chemical contains one or more positive alerts, it is predicted as "active." However, this active prediction call can be changed if the local QSAR of the positive alert modifies the prediction. The presence of a deactivating alert alongside a positive alert renders the prediction call as "inactive." If more than one positive alert is present, then the one with the highest number of active chemicals is used, and in the case of more than one deactivating alert, the one with the highest number of inactive chemicals is used. If a test chemical contains a positive alert that has been seen in just one or two active training set chemicals, the prediction result is considered "inconclusive" because of the alert low statistical confidence. CASE Ultra recognizes unusual features/fragments in test chemicals that do not match training data (unknown structural fragments). The presence of more than three unknown structural fragments in the test chemical results in an "out of domain" call.

Leadscope Model Applier: The Leadscope software employs a fragment-based QSAR paradigm; however, the fragments are not paths of distinct lengths but are predefined in a hierarchically organized dictionary that is closely related to common organic/medicinal chemistry building blocks. For binary classification problems, such as the Ames test results, the algorithm identifies toxicity modulating fragments using a χ^2-test. Furthermore, the software is able to build superstructures from smaller fragments if they improve predictivity. Together with eight global molecular properties, the set of fragments is then used as a descriptor set in a partial least squares (PLS) logistic regression model of the activity class. Therefore, the predictions from this algorithm are continuous probabilities of class membership rather than binary outputs. The program also assesses the applicability domain by measuring the

distance to training set molecules. Typically, probabilities greater than 0.5 can be used to give an "active" prediction and probabilities smaller than 0.5 an "inactive" prediction, which is the standard procedure used by the Model Applier for pretrained models. The system can also annotate compounds as "out of domain" or with "missing descriptors" when a conclusive prediction cannot be made [23].

OECD Toolbox: The OECD Toolbox [24, 33] represents a free source of various models. The Toolbox is a software application intended to the use of governments, chemical industry, and other stakeholders in filling gaps in (eco)toxicity data needed for assessing the hazards of chemicals. The Toolbox incorporates information and tools from various sources into a logical workflow. Crucial to this workflow is grouping chemicals into chemical categories. The seminal features of the Toolbox are (1) identification of relevant structural characteristics and potential mechanism or mode of action of a target chemical, (2) identification of other chemicals that have the same structural characteristics and/or mechanism or mode of action, and (3) use of existing experimental data to fill the data gap(s). The Toolbox includes a number of models predicting several toxicological end points, such as skin sensitization, Ames mutagenicity, acute and repeat-dose toxicity, aquatic toxicity, and others [34].

Bioclipse [25]: It is an open-source cheminformatics toolkit with a wide array of toxicity models integrated, such as carcinogenicity, mutagenicity (Ames), hERG, aquatic tox (Daphnia), and a wide array of models from OpenTox [35]. The Ames mutagenicity model in Bioclipse is built using the dataset published by Kazius et al. in 2005 [36] containing 4337 chemical structures of which 2401 were classified as mutagen and 1936 non-mutagen. The datasets can be downloaded, and the software can be used to generate many molecule descriptors (using the CDK) [37, 38] and then QSAR models (through integration with the R statistical software). The software is considered not as user friendly as some commercial tools [39].

admetSAR: admetSAR [40] is a free website (http://lmmd.ecust.edu.cn:8000/) [41] that enables a single input SMILES structure to be used to rapidly predict scores against a wide range of ADME/Tox models (at the time of writing, 26 qualitative classification and 5 quantitative regression models). These datasets can also be downloaded as most are based on other publications. Each model has some statistics describing the model as well as a probability to provide more confidence in the result. The software is simple to use, and drawbacks appear to be the lack of batch processing operation, the "black box" nature of the models, and the lack of capability to build or update the models on the website [39].

Symmetry: Symmetry [26] is a platform that applies advanced machine learning techniques to a variety of structural features and physico-chemical properties of small molecules to provide quality predictions about biological effects. Available Symmetry algorithms include binary classification for active/inactive datasets, meta-classifiers to achieve consensus predictions for sets of binary models, and multi-label learning that yields ranking and probabilistic estimates of the possible outcomes. Symmetry offers a wide range of predictive models, including mechanism of action and phenotypic models, toxicity [42], and human adverse effects.

3 Assessment of Potential Genotoxic Impurities

3.1 ICH M7 Guideline

3.1.1 Background

The European Medicines Agency Committee for Medicinal Products for Human Use (CHMP) released in 2006 [43] a "Guideline on the Limits of Genotoxic Impurities," which describes an approach for assessing genotoxic impurities of unknown carcinogenic potential based on the TTC concept. In 2007 a question and answer document was published on the EMA website addressing several aspects of the practical implementation of the recommendations contained in the Guideline.

Genotoxicity is a broad term that typically describes a deleterious action on cellular genetic material. Chemicals may induce DNA damage by directly interacting with it (e.g., alkylating agents) or by acting on non-DNA targets (e.g., mitotic spindle poisons, inhibitors of topoisomerase, etc.). For DNA-reactive genotoxins, the mechanism by which they induce genetic damage is assumed to follow a linear no-threshold model; on the other hand, for molecules not interacting directly with DNA, the existence of a threshold concentration required to induce the damage is by and large accepted [44]. Impurities that belong to the second category of substances can be regulated according to the ICH Quality Guideline Q3C [45] which includes class 2 solvents. The thresholds or permissible daily exposures (PDE) are calculated from the no-observed-effect level (NOEL) obtained in the most relevant animal studies with the use of conservative conversion factors used to extrapolate the animal data to humans.

The CHMP Guideline suggests that the TTC concept should be applied to those genotoxic impurities that do not have sufficient evidence of a threshold-related mechanism of action. The reference values are taken from Kroes et al. [46], where a TTC of 0.15 μg/day is proposed for impurities presenting a structural alert for genotoxicity, corresponding to a 10^{-6} lifetime risk of cancer. In the case of pharmaceuticals, the Guideline suggests a 1 in 100,000 risk be applied, resulting in a TTC of 1.5 μg/day.

For drug substances, the identification thresholds above which impurities are required to be identified are within the range of 0.05

and 0.1 %. ICH Guidelines Q3A(R) [47] and Q3B(R) [48] state that even though the identification of impurities is not necessary at levels lower than or equal to the identification threshold, "analytical procedures should be developed for those potential impurities that are expected to be unusually potent, producing toxic or pharmacological effects at a level not more than the identification threshold." The Guideline recommends carrying out a thorough evaluation of the synthetic route along with chemical reactions and conditions, with the aim of identifying reagents, intermediates, starting materials, and readily predicted side products which may be of potential concern. Once all potential impurities are theoretically identified and listed, an initial assessment for genotoxicity is carried out by a scientific expert using computer tools such as QSAR and knowledge base expert systems. A thorough literature and internal archive (when applicable) search also needs to be completed, as a number of intermediates and reagents have often been tested in genotoxicity or carcinogenicity assays. The potential genotoxic impurities which may be present in an API are then classified into one of five classes described by Müller et al., in 2006 [49]; the purpose is to identify those impurities that pose a high risk and need to be limited to very low concentrations.

In 2006, a task force established under the umbrella of the US Pharmaceutical Research and Manufacturers of America (PhRMA) for the first time proposed the "staged TTC" concept to be applied to pharmaceuticals [49]. The task force was established as a response to various clinical holds imposed by the FDA on investigational drugs in clinical trial phases based on suspicions to contain genotoxic impurities at levels potentially associated with a risk for the volunteers or patients involved in these trials [50]. The staged approach allows levels of daily intake of mutagenic impurities higher than 1.5 μg as defined by the lifetime TTC, namely, 10 μg (for a 6–12-month duration), 20 μg (3–6 months), 40 μg (1–3 months), and 120 μg for not more than 1 month. The EMA adopted the staged TTC approach for limits of genotoxic impurities in clinical trials in the 2007 Q&A document (EMA 2010), but to be more conservative, it reduced the staged TTC limits proposed in the PhRMA paper by a factor of 2.

In 2008, the FDA issued a draft "guidance for Industry on Genotoxic and Carcinogenic Impurities in Drug Substances and Products: Recommended Approaches" (FDA 2008) which was largely similar to the EU guidance. However, this document has not been finalized because in 2009 the topic "genotoxic impurities" was adopted by ICH for development of a new internationally harmonized guideline. Since the topic was considered to include both safety and quality aspects, the projected Guideline was assigned to the M (multidisciplinary) series of the ICH process and designated as ICH M7 with the title "Assessment and Control

of DNA-Reactive (Mutagenic) Impurities in Pharmaceuticals to Limit Potential Carcinogenic Risk" [51].

In February 2013 a draft of the M7 Guideline was published in the three ICH regions for public consultation (step 3 of the ICH process). The document was adopted as a step 4 ICH Harmonised Tripartite Guideline in June 2014 (ICH 2014) and is currently on step 5, adopted by CHMP on 25 September 2014 and issued as EMA/CHMP/ICH/83812/2013 [51].

3.1.2 Key Aspects of the ICH M7 Guideline

The ICH M7 Guideline combines many of the principles set by the EU and the draft FDA Guidelines on genotoxic impurities. Some aspects, though, have been updated and clear recommendations can be identified. A thorough description of all key aspects of the ICH M7 Guideline, which are described elsewhere [50], is beyond the scope of the present contribution. It is nonetheless worthwhile mentioning few of the critical aspects that the ICH M7 Guideline does enforce:

1. Structure-based assessment of potentially mutagenic impurities has to be carried out using two in silico systems that complement each other: one should be a rule-based and one a statistics-based method (*see* Subheading 2 in this chapter).
2. The impurities classification system proposed by the ICH M7 Guideline has been derived from the scheme proposed by Müller et al. in 2006 [49], which identifies five classes of impurities as a function of data availability for the characterization of their mutagenicity and carcinogenicity potential.
3. ICH M7 replaced the term "genotoxic impurities" as applied by the EU Guideline on the Limits of Genotoxic Impurities with the term "DNA-reactive impurities" in order to specify that DNA-reactive compounds (i.e., that typically covalently bind to DNA-generating adducts, which, if unrepaired, can lead to point mutations and/or strand breakage) are those that fall within the scope of the Guideline. There is also the assumption that DNA-reactive (Ames-positive) compounds are likely carcinogens with no threshold mechanism.
4. For DNA-reactive (Ames-positive) compounds lacking rodent carcinogenicity data, a generic TTC value would be applied as an acceptable intake level that poses a negligible risk of carcinogenicity.
5. If rodent carcinogenicity data is available for a (potentially) mutagenic impurity, the application of the TTC concept is not warranted, and a compound-specific calculation of acceptable levels of impurity intake is recommended as is described in more detail in the Note 4 of the Guideline [51].
6. Compound-specific calculations for acceptable intakes can be applied case-by-case for impurities which are chemically similar

Table 1
Acceptable intakes for an individual impurity

Duration of treatment	≤1 month	>1–12 months	>1–10 years	>10 years
Daily intake (μg/day)	120	20	10	1.5

Table 2
Acceptable total daily intakes for multiple impurities

Duration of treatment	≤1 month	>1–12 months	>1–10 years	>10 years
Daily intake (μg/day)	120	60	30	5

to a known carcinogen compound class (class-specific acceptable intakes) provided that a rationale for chemical similarity and supporting data can be demonstrated (Note 5) [44, 51].

7. The acceptable intakes derived from compound-specific risk assessments can be adjusted for shorter duration of exposure. The TTC-based acceptable intake of 1.5 μg/day is considered to be protective for a lifetime of daily exposure. To address less-than-lifetime (LTL) exposures to mutagenic impurities in pharmaceuticals, a formula is applied in which the acceptable cumulative lifetime dose (1.5 μg/day × 25,550 days = 38.3 mg) is uniformly distributed over the total number of exposure days during LTL exposure. This allows higher daily intakes of mutagenic impurities than would be the case for lifetime exposure and still maintain comparable risk levels for daily and non-daily treatment regimens.

   Table 1 summarizes the levels for different duration.

8. As far as multiple impurities are concerned, when there are more than two mutagenic (i.e., Ames-positive) or alerting impurities, total mutagenic impurities should be limited as described in Table 2 for clinical development and marketed products.

3.2 Performance of Commercial Systems on Proprietary Compounds

In silico methods for the prediction of mutagenic activity have been available for many years, and they have been continuously improved in terms of technology and prediction results, also for greater availability of high-quality data.

The specific use of such in silico tools in the pharmaceutical industry, in the context of the evaluation of genotoxic impurities, has been recently summarized and reviewed by Sutter et al. [52]. The authors, representing a total of 14 pharmaceutical companies, compared the predictive value of the different methodologies analyzed in two surveys conveyed in the US and European

pharmaceutical industry: most pharmaceutical companies used a rule-based expert system as their primary methodology, yielding negative predictivity values of ⩾78 % in all participating companies. A further increase (>90 %) was often achieved by an additional expert review and/or a second statistics-based methodology. Also in the latter case, an expert review was encouraged, especially when conflicting results were obtained. The conclusion was that a rule-based expert system complemented by either expert knowledge or a second (Q)SAR model is appropriate. Overall, the procedures for structure-based assessment presented in the article by Sutter et al. [52] were already considered appropriate for regulatory submissions within the scope of ICH M7, which mandates the use two different methodologies: one expert-rule based and one statistical-based.

In order to comply with such Guideline specification, additional commercial in silico tools and novel models have been recently made available to the scientific community. Brigo *et al.* [53] evaluated three expert-rule systems (*Derek Nexus v.4.0.5* [13], *Toxtree v.2.6.6* [14], *Leadscope Expert Alerts v.3.2.4-1* [16]) and three statistical systems (*Sarah v.1.2.0* [22], *Leadscope Model Applier v.3.2.4-1* [23], *three models of CASE Ultra v.1.5.1.8* [15]—*GT1_7B, SALM2013, SALM2013PHARMA*) in an individual and combined fashion.

The evaluation was carried out using a large validation set of Ames mutagenicity data comprising over 10,000 compounds, 30 % of which are Roche proprietary data (Table 3). The Roche datasets include the vast majority of compounds (not only impurities) tested in the Ames Standard [54] and Microsuspension [55] protocols.

All programs have been applied as commercially available, without internal customization or follow-up expert knowledge.

Individual systems showed adequate performance statistics with public domain datasets (concordance, 74–95 %; sensitivity, 58–99 %; specificity, 51–96 %; *see* Fig. 1); however, there was a consistently significant drop in sensitivity with the Roche datasets,

Table 3
External validation sets

Dataset	Number of compounds	Positive	Negative
Roche Ames Standard	1,335	254	1,081
Roche Ames Microsuspension	1,785	190	1,595
LSDB	4,699	2,068	2,631
Hansen [56]	2,647	1,773	874
Total	10,466	4,285	6,181

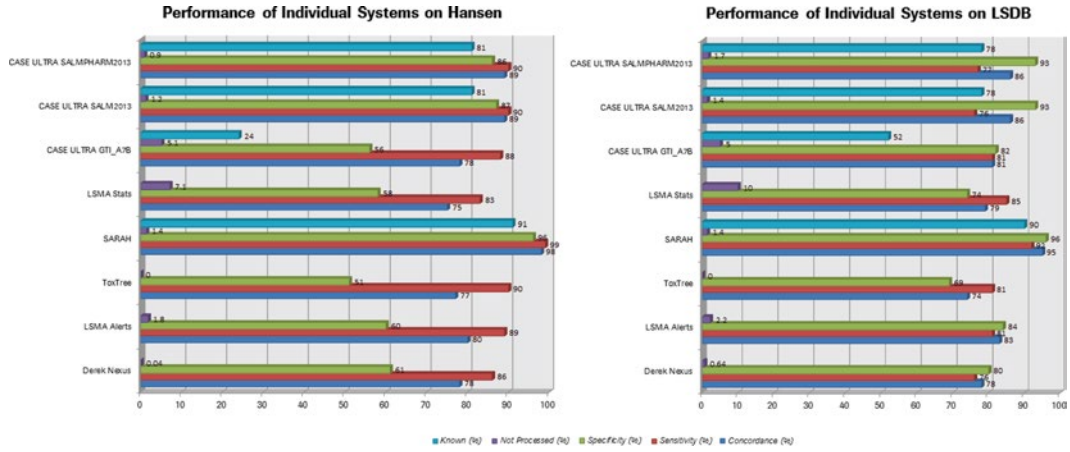


Fig. 1 Performance of individual systems on public datasets Hansen [56] and LSDB

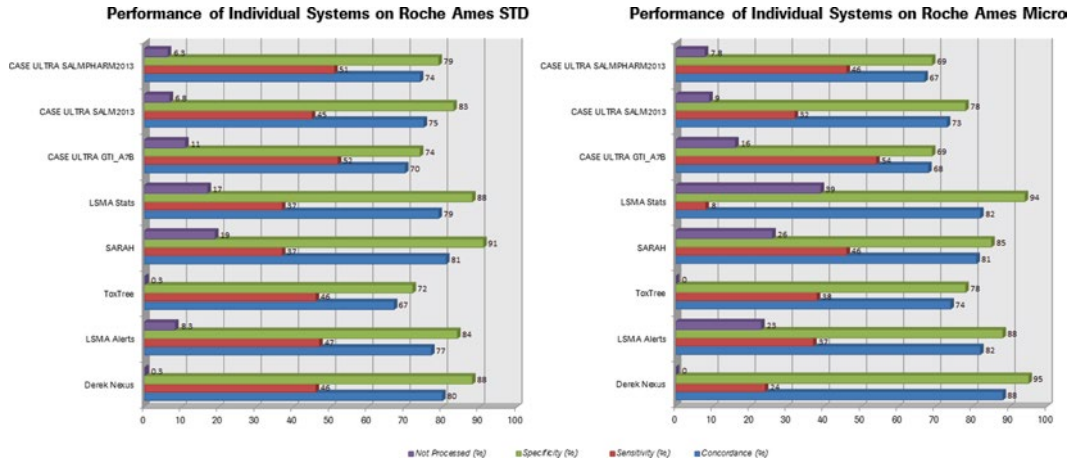


Fig. 2 Performance of individual systems on Roche Ames Standard and Ames Micro datasets

down, in one case, to single digit (concordance, 66–88 %; sensitivity, 8–54 %; specificity, 69–95 %; *see* Fig. 2). All systems showed good performance with "public validation sets," also due to the training set overlap, which went up to 91 % for Sarah (Fig. 1).

Expert-rule-based tools showed lower specificity with public domain datasets versus the statistical-based programs. Statistical tools showed a much higher number of compounds (up to 39 % in one case) outside of their applicability domains and, hence, not predicted (Fig. 2).

To evaluate the performance of the combined approach recommended by the ICH M7 Guideline, the Roche validation sets have been submitted to all possible combinations of one expert-rule-based and one statistical-based system (Figs. 3 and 4).

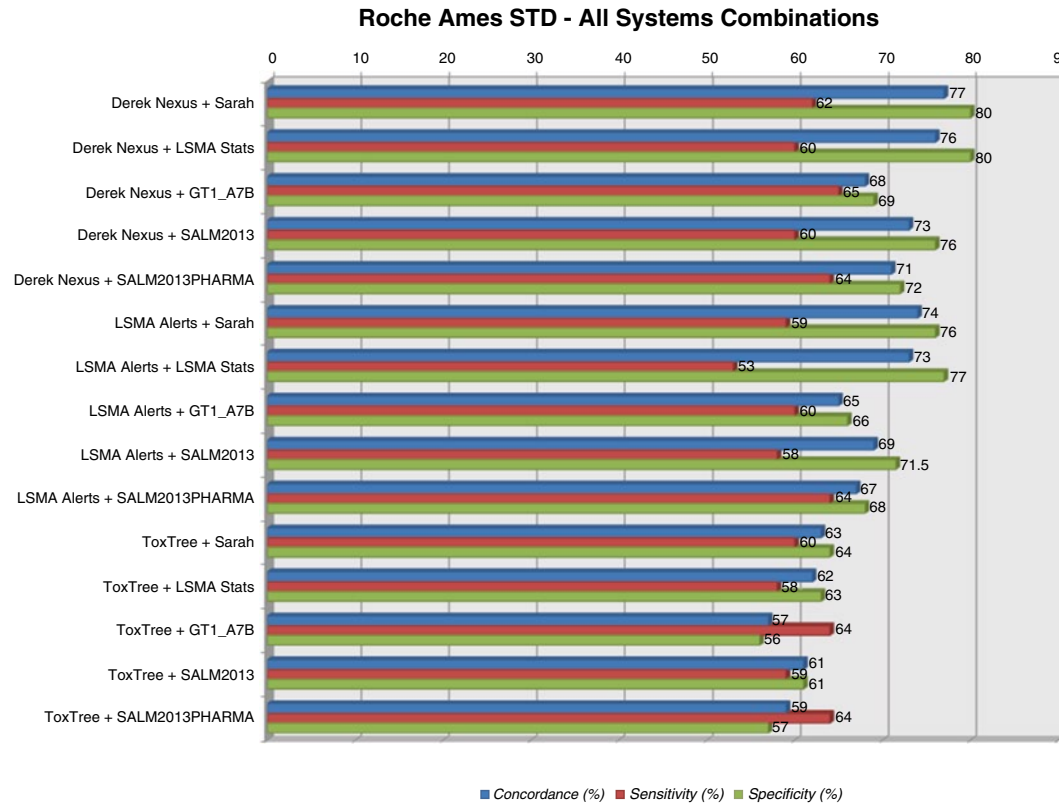


Fig. 3 Performance of combined systems on the Roche Ames Standard dataset

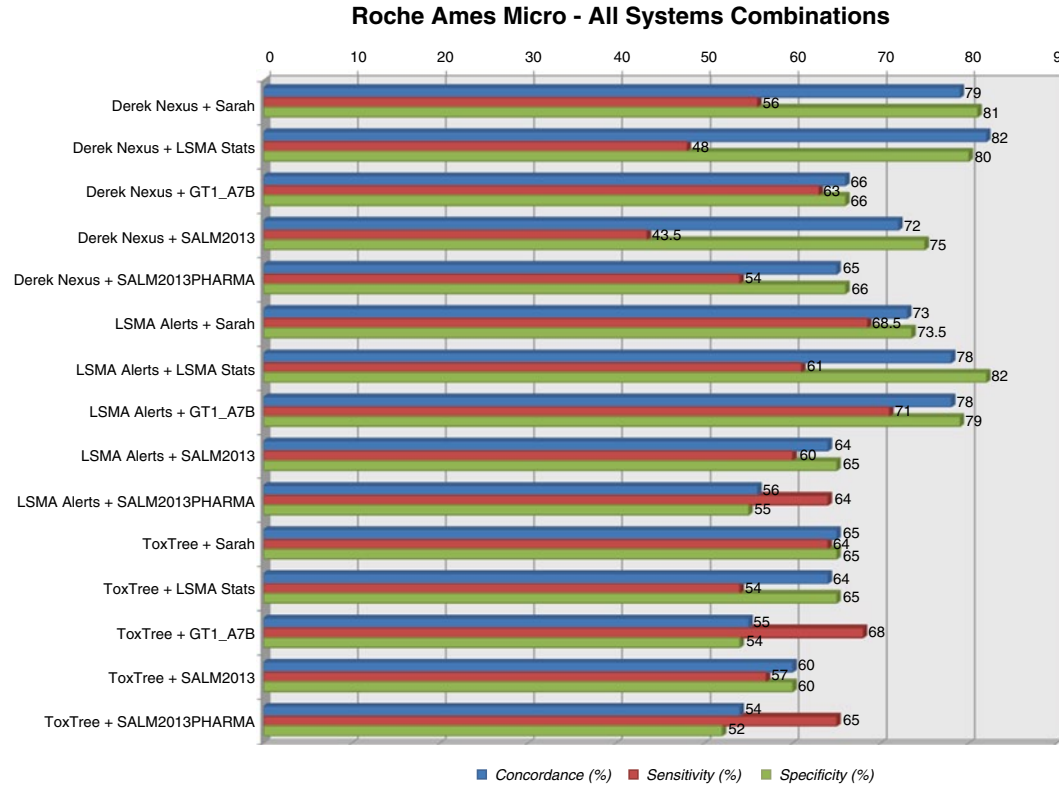


Fig. 4 Performance of combined systems on the Roche Ames Micro dataset

The combinations of all systems, compared to their individual performance with both Roche validation sets, improve the sensitivity to consistently above 50 %, up to 71 % for the combination "LSMA Alerts + SALM2013." As expected, specificity is generally lower than with individual systems, but its reduction is limited for the majority of combinations.

In order to assess the prediction tools with chemicals that fall within the potential genotoxic impurities chemical space, four subsets of both Roche validation sets have been generated with molecular weights (MW) ≤400, ≤350, ≤300, and ≤250. Such subsets cover the chemical space of the large majority of the potential genotoxic impurities tested in Roche over the past decade.

All programs have been tested against these subsets individually (Figs. 5 and 6) and in combination (Figs. 7 and 8)[53].

With individual systems, sensitivity shows a clear trend to increase proportionally to the decrease of MW. For example, in the Roche Ames Microsuspension set, sensitivity improves as follows: Derek from 27 to 64 %, Sarah from 51 to 76 %, Toxtree from 42 to 85 %, and CASE Ultra SALMPHARM2013 from 45 to 71 %. LSMA Alerts and LSMA Stats show an increase in sensitivity to 60 % up to MW ≤300, but there is a flection down to ~55 % for both programs for MW ≤250. In general, sensitivity increases significantly with low-MW subsets with almost all programs and models (Figs. 5 and 6). The only exception is CASE Ultra SALM2013 model, which keeps the same sensitivity values throughout all subsets (between 29 and 33 %) [53].

The evaluation of combined systems with low-MW Roche subsets shows a significant increase in sensitivity, up to over 90 % for sets with MW ≤300 and ≤250 with several combinations (Figs. 7 and 8). The increase in sensitivity is proportional to the decrease in MW; at the same time, there is a considerable decrease in specificity (<30 % in some cases). Such deltas are generally more pronounced in the Ames Micro dataset (Fig. 8) compared to the Roche Ames Standard dataset (Fig. 7). In the Ames Standard subsets, specificity and sensitivity values are consistently comprised between 70 and 80 % in nearly all Derek Nexus and LSMA Alerts combinations. In the latter combinations, values are a bit lower than 70 % at higher MW. Toxtree combinations show lower sensitivity and specificity values at higher molecular weights and greater gaps between sensitivity and specificity within the subsets MW ≤300 and MW ≤250 [53].

As far as the Roche Ames Micro set is concerned, the sensitivity is in the range of 90 % in the subset with MW ≤250 with several combinations, such as Derek Nexus + Sarah and Derek Nexus + GT1_A7B; LSMA Alerts + Sarah; LSMA Alerts + CASE Ultra models. Nearly all combinations with Toxtree gave sensitivity in the range of 90 %. Nearly all combinations of LSMA Alerts showed high sensitivity also in the subset with MW ≤300.

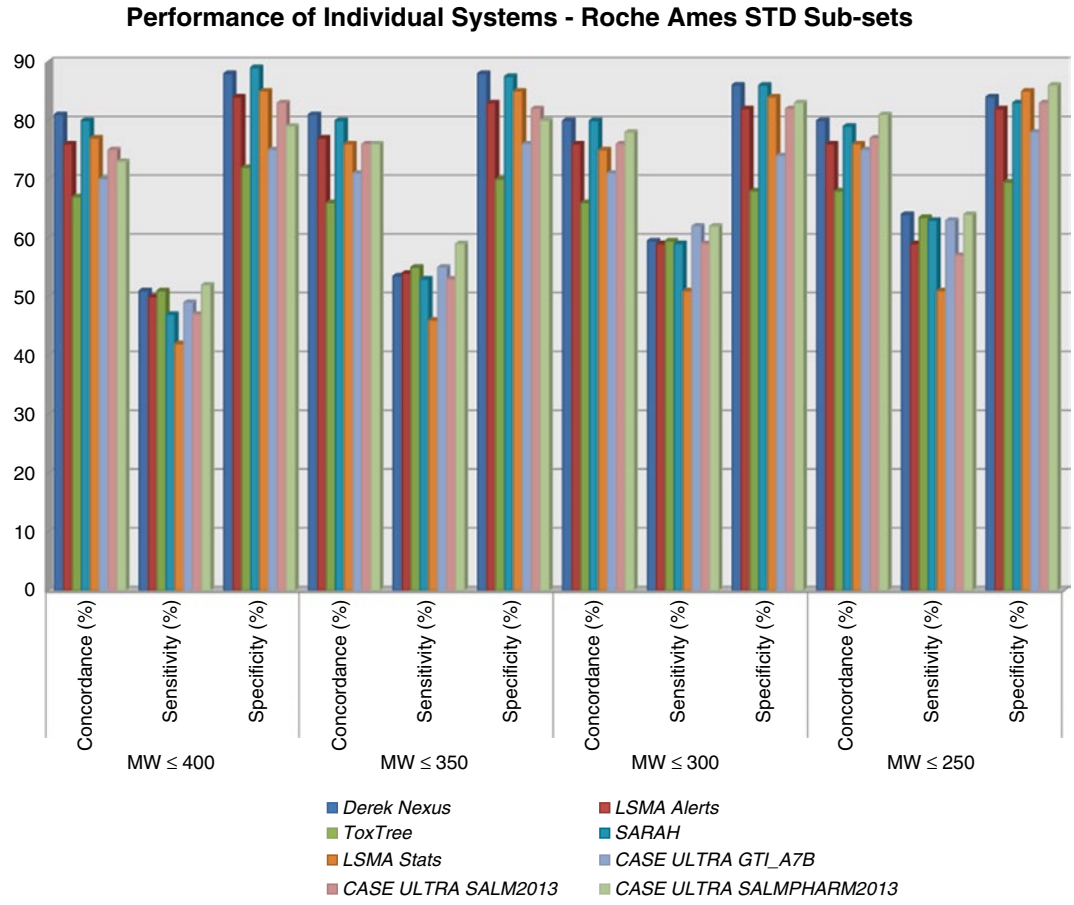


Fig. 5 Performance of individual systems on the Roche Ames Standard dataset filtered by MW

Looking at the plots in Fig. 8, it is evident that more balanced results are obtained with all Derek Nexus combinations: in other words, the sensitivity increases proportionally to the decrease of the MW at a moderate expense of specificity. Compared to this, LSMA Alerts combinations have overall lower specificity than Derek combinations. At the same time, Toxtree combinations, despite showing good sensitivity, have a greater corresponding decrease in specificity.

3.3 Improvement of In Silico Predictions with Proprietary Data

Validation exercises such as those described in Subheading 3.2 for mutagenicity or for other end points are typically very useful for the identification of specific gaps in the chemical space represented by the assessed models and tools.

In particular, when proprietary data are used as external validation sets, false predictions represent a valuable opportunity to improve the models and expand their overall applicability domain.

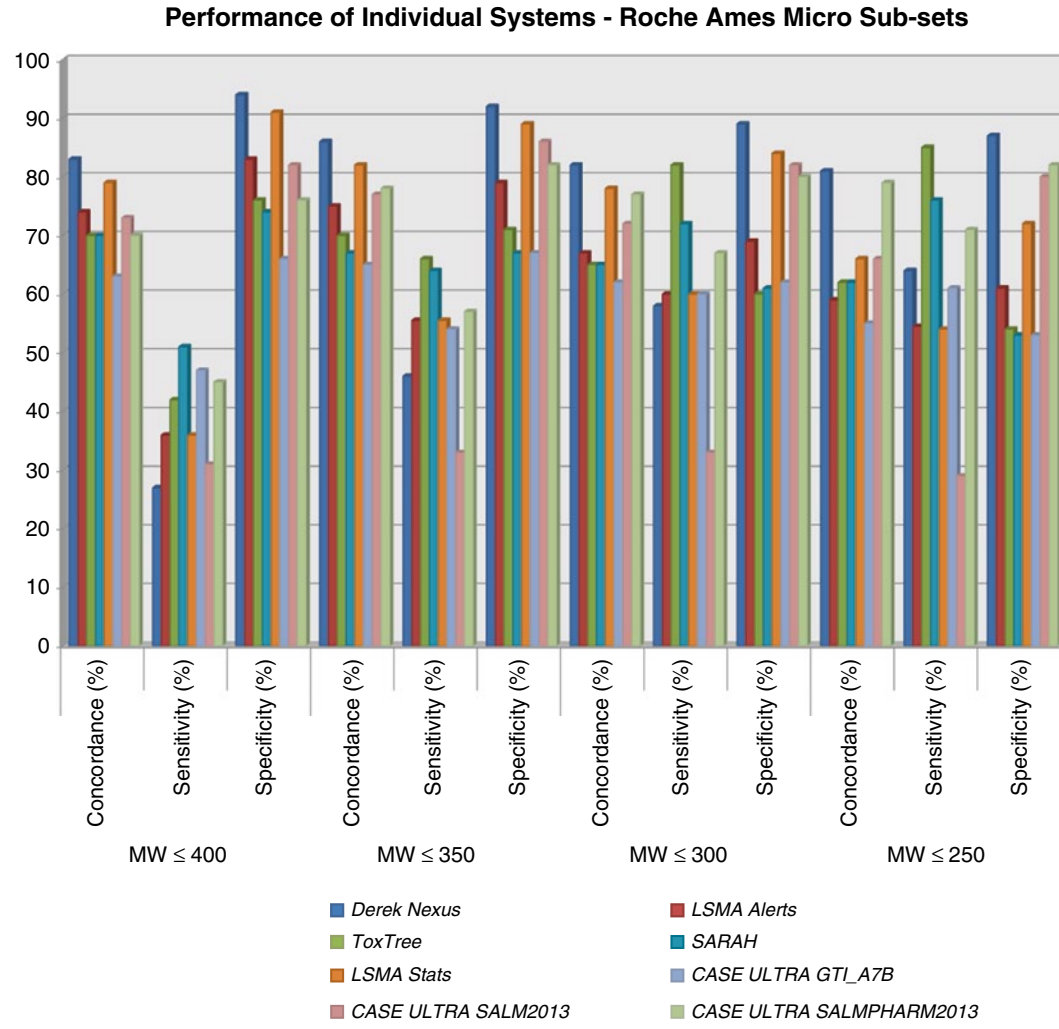


Fig. 6 Performance of individual systems on the Roche Ames Micro dataset filtered by MW

Roche recently undertook a similar exercise with Lhasa Ltd. in order to systematically include proprietary knowledge into the in silico prediction tools that are routinely used for early safety assessment. Data collected from Ames test, embryonic stem cell assay (teratogenicity), hERG inhibition in vitro screening, and micronucleus in vitro (chromosome damage) have been used to fill the gaps identified in the models adopted within the company.

These collaborative efforts, aimed at incorporating proprietary knowledge in prediction models, quickly translated into a significant increase in the prediction metrics (*see* Table 4), with sensitivity values that showed up to 60 % improvements.

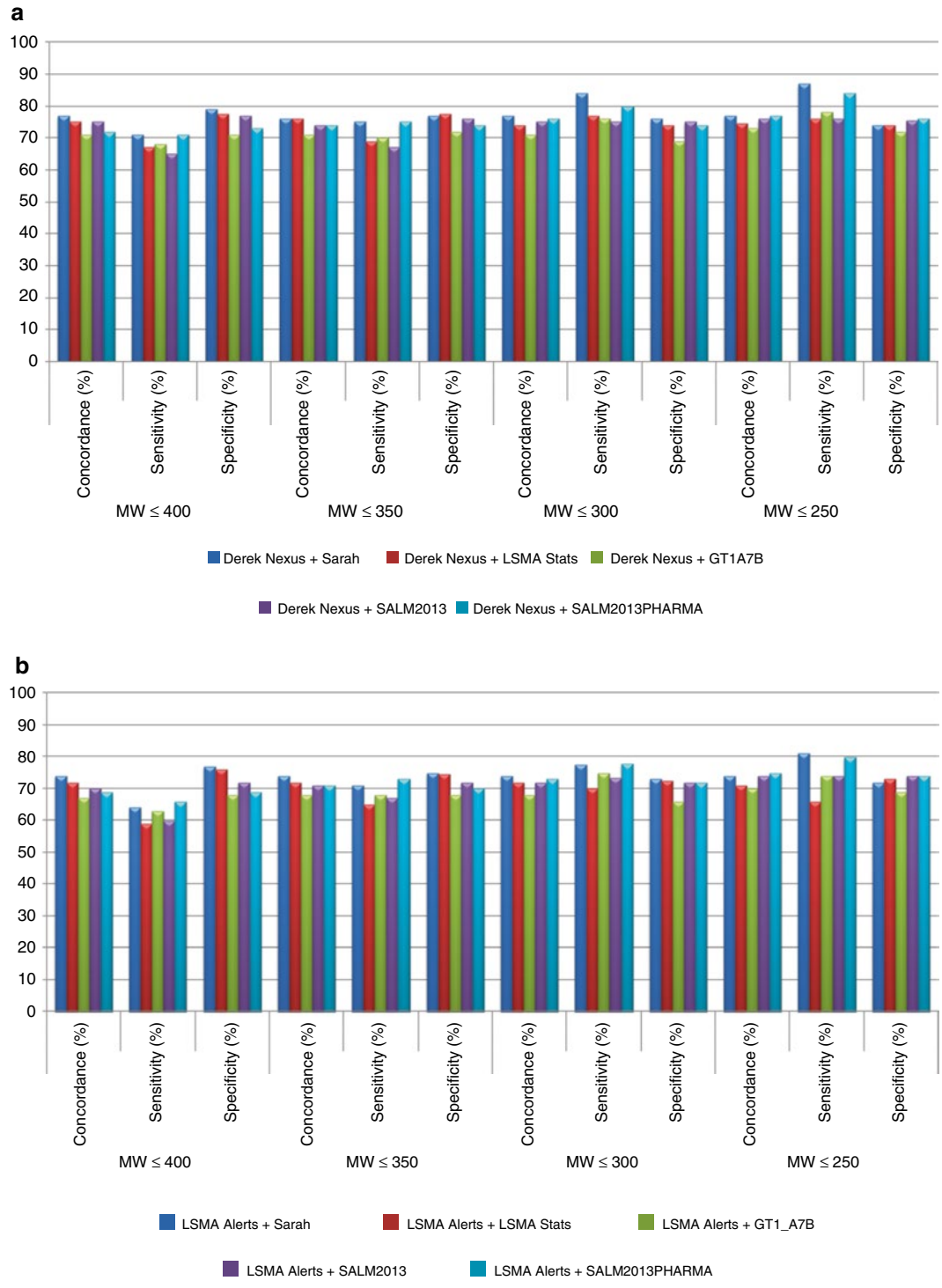


Fig. 7 Performance of combined systems on the *Roche Ames Standard* dataset filtered by MW

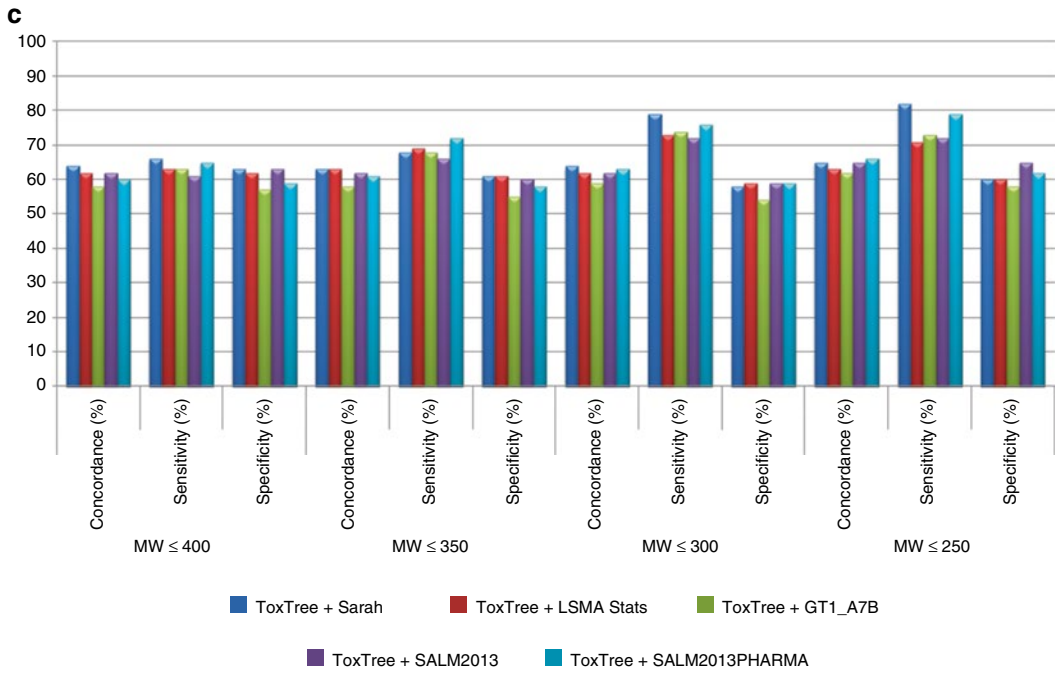


Fig. 7 (continued)

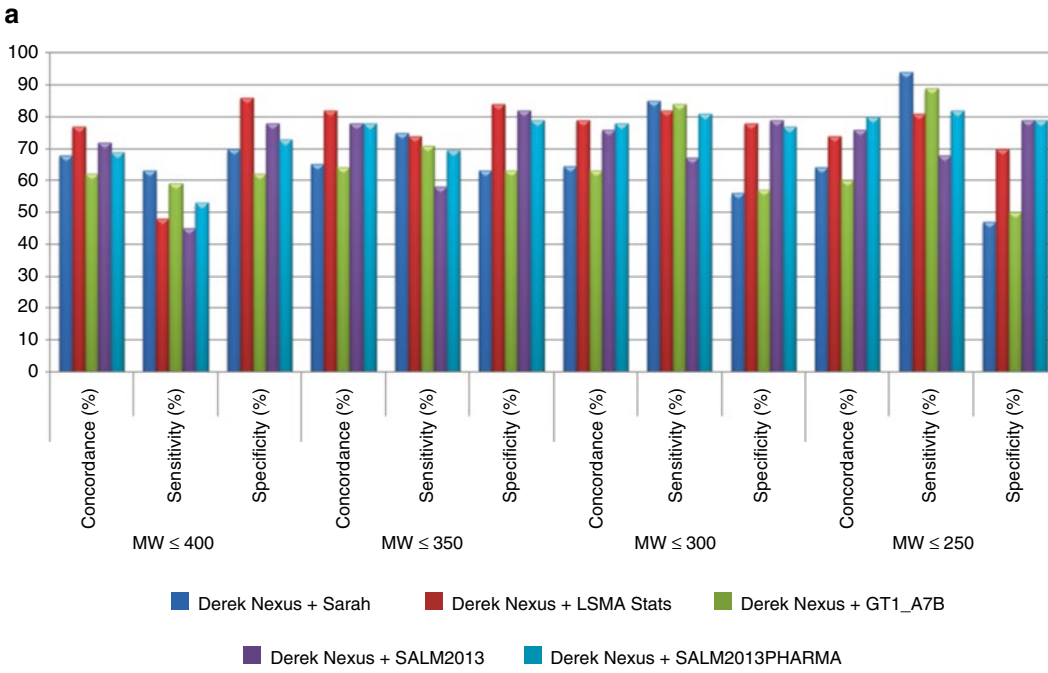
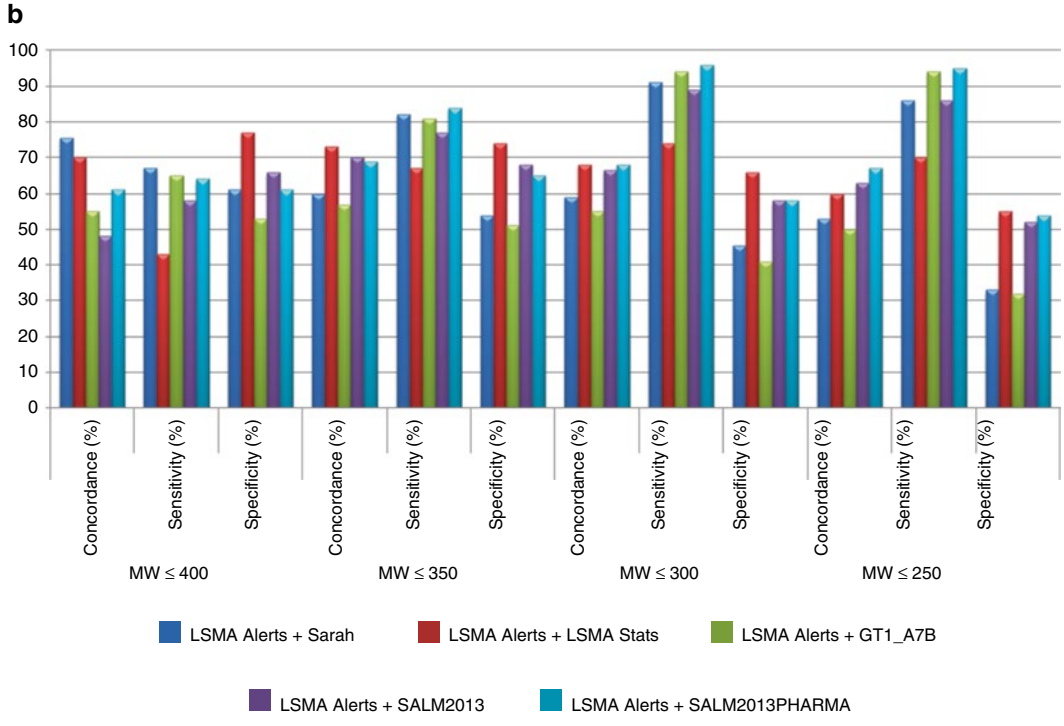


Fig. 8 Performance of combined systems on the *Roche Ames Micro* dataset filtered by MW



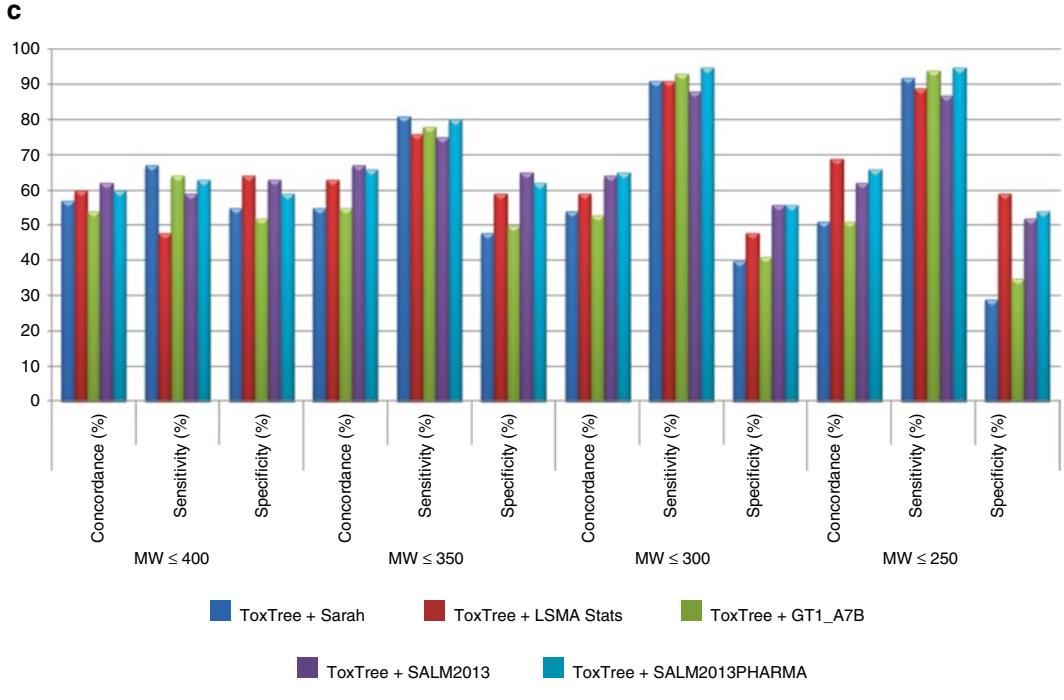


Fig. 8 (continued)

Table 4
Improvement in predictive performance of an in silico prediction tool including Roche proprietary data

End point		Sensitivity (%)	Specificity (%)	Positive predictivity (%)	Negative predictivity (%)	Balanced accuracy (%)
Mutagenicity	Previous	36	92	45	89	64
	Updated	69	89	55	94	79
Chromosome damage	Previous	5	97	34	76	51
	Updated	65	92	72	89	78
hERG inhibition	Previous	21	90	70	50	55
	Updated	63	67	69	61	65
Teratogenicity	Previous	3	96	17	79	50
	Updated	59	92	66	90	76

4 Role of In Silico Models in the Prediction of Toxicity in Drug Discovery

In silico approaches to predict potential toxicities and drug metabolism on the basis of the chemical structure are of particular interest to the pharmaceutical industry as having the potential to impact the early drug discovery process as well as in the candidate selection phase. Prediction models for the identification of metabolic soft spots and potentially toxic substructures can be easily applied to a large number of chemical structures and are therefore integrated already during HTS (high-throughput screening) or even earlier as an automatically attributed alert for all new chemical entities. At this early stage, only a basic in silico profiling can be done, as only the most well-validated end points can be reliably applied automatically and generated on the fly without an expert intervention. Later in the development, at the latest before the final candidate is selected, a more detailed in silico profiling also considering the whole profile of the compound is thoroughly conducted. According to the development scheme provided in Fig. 9, the further in silico tools and in vitro downstream activities are conducted.

4.1 Target Identification (TI), Target Assessment (TA), and Exploratory Work

The first step after the target has been identified as potential development opportunity is a target assessment (TA) conducted by nonclinical safety experts, using appropriate databases and public sources. A proper target/functionality assessment in healthy and diseased status contains pathway mapping, information from knockout and transgenic models, a target expression profile in relevant species, as well as a critical evaluation of potential off-target

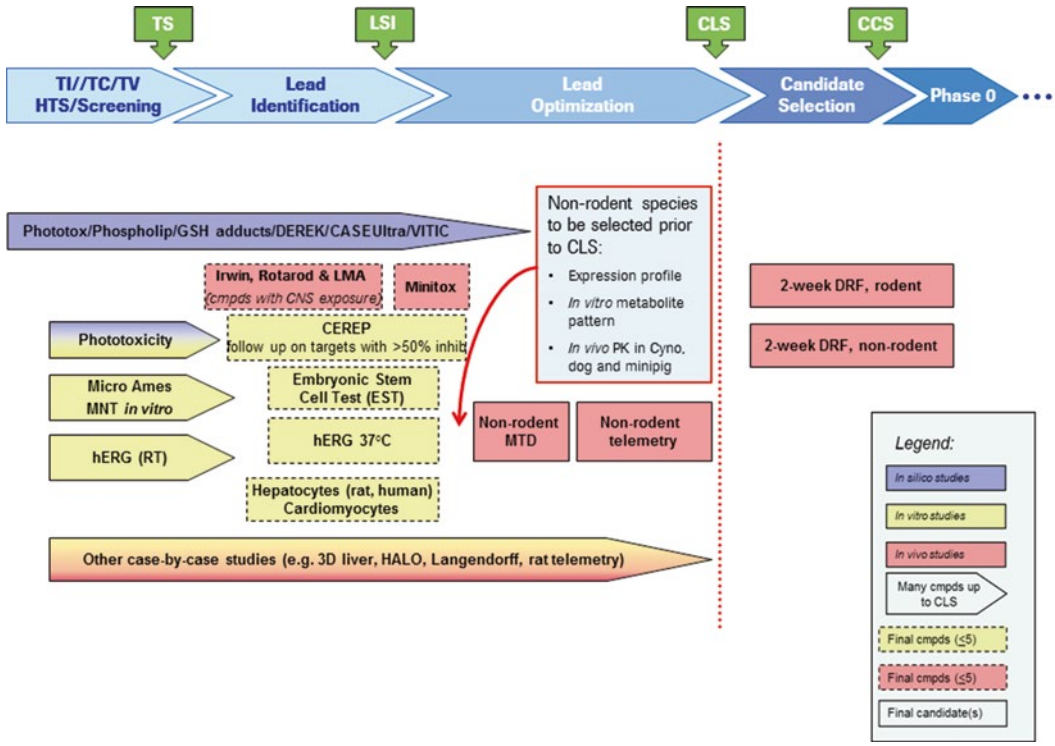


Fig. 9 Use of in silico tools and safety screening during the early drug development process

safety alerts (selectivity). Various software systems are available to assist the experts in these assessments (e.g., MetaCore [57], Symmetry [26]).

Modern in silico prediction software is able to calculate thousands of chemical structures on the fly and can be therefore applied very early in the drug development process. Immediately after the chemical structure is known, meaning chemical libraries are added to the companies' chemical database, a basic in silico prediction panel is applied using reliably validated toxicological end points like genotoxicity and carcinogenicity. As always, a large, homogeneous, and high-quality database is the prerequisite for reliable predictions. Therefore, in vitro screens which have been used within pharmaceutical companies for years containing data generated often in one single lab are the best sources for the development of highly predictive models. For example, an in silico model predicting the potential of drug-induced phospholipidosis (a reversible storage disorder characterized by accumulation of phospholipids within cells) has been developed. Based on more than 600 in vitro assay, an accuracy of 86 % led to a replacement of the in vitro by the in silico method. The model is calculating the free energy of amphiphilicity ($\Delta\Delta G_{AM}$) and log P value [58] of cationic amphiphilic drugs and can be applied in a high-throughput mode.

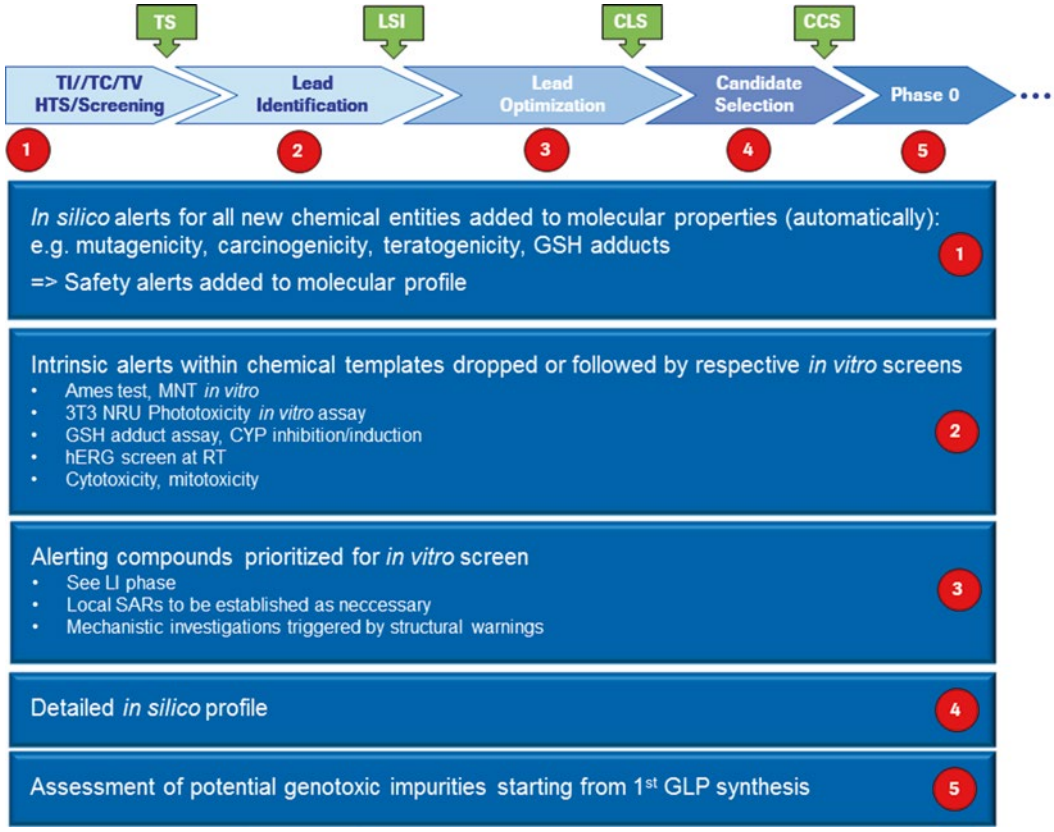


Fig. 10 Downstream activities following in silico alerts in the drug development process

Further end points, which can be used for on-the-fly predictions, are teratogenicity, GSH adduct formation, irritation, and skin sensitization.

At this early stage of development, the potential safety hazards identified by the application of an expert system in combination with a set of statistical models contribute to the overall compound profile, but are not used as a decision pathway (*see* Fig. 10).

4.2 Lead Identification (LI) Phase: Target Selected (TS) to Lead Series Identified (LSI)

The main goal during lead identification is to identify valid chemical templates for further optimizing the efficacy and selectivity on the target, ideally multiple discrete series. Besides computational chemistry tools to calculate physicochemical properties, virtual screening, structure-based design, QSAR analysis of both the desired target and off-target activities, and chemical structures are analyzed continuously in silico for possible structure-related safety concerns to identify major issues with the templates. Insights into the toxicological potential of a scaffold or series of structures early in the drug discovery process could help medicinal chemists to

prioritize particular scaffolds. Components of early avoidance of chemical structure safety liabilities include predictions for genotoxicity, carcinogenicity, hERG channel blockade, reactive metabolite formation, phospholipidosis, structural similarity to problematic molecules, CYP inductions, GSH adduct formation, and DMPK properties (cell penetration, microsomal stability, CYP3A inhibition). The in silico tools offer good guidance on what additional tests may be necessary or whether further characterization is warranted; however, they also have limitations [59].

Drug, metabolism and pharmacokinetics (DMPK) properties play a major role during lead identification. Numerous commercially available tools for the prediction of metabolites exist, such as METEOR [11, 17], MetabolExpert [60], and MetaSite [61]. Most software packages correctly predict metabolites that are detected experimentally. However, a relatively high incidence of false predictions of metabolites is common to most unspecified computerized systems. In the hand of drug metabolism experts, these software packages have a certain value for hypothesis generation and guiding to experimental approaches for the identification of drug metabolites. However, the generation of additional new local rules, intended to predict the activity of a single enzyme (and often only within a chemical series), can significantly improve the prediction accuracy.

Experimental follow-up of potential issues are conducted to build/refine safety plans moving forward. Even if in vitro assays clearly disprove identified in silico alerts, further spot-checking of the distinctive end point will be conducted to avoid creeping in of a structural liability. The in vitro results always overrule the in silico warnings provided that the corresponding assays could have been conducted under reliable conditions (e.g., solubility, stability). Chemical templates with identified and confirmed intrinsic metabolic and/or safety concerns will be eliminated (*see* Fig. 10).

4.3 Lead Optimization (LO) Phase: Lead Series Identified (LSI) to Clinical Candidate Selected (CLS)

The task of the LO phase is to take a lead and convert it into a candidate for preclinical evaluation. This phase is intensively accompanied by early safety in vitro screening in various areas: genotoxicity, hERG and other ion channels, cytotoxicity, hepatic toxicity, bone marrow toxicity, transporters, metabolite identification, metabolic stability, CYP induction/inhibition, reactive metabolites, off-target pharmacology/secondary pharmacology, and cross-species comparisons, where applicable. Further screens might be applied based on the target liabilities or already identified potential safety issues. If adverse in vitro activities appear, specified structure-activity relationships (SARs), so-called local SARs, will be established to support the discovery projects in optimizing the clinical candidates toward safety/DMPK in parallel to efficacy.

At this stage, first, fit for purpose in vivo studies are conducted to address early target or chemistry related safety concerns. The first general toxicology studies are maximal tolerated dose (MTD) and dose-range finding (DRF) studies generally performed in rodents and non-rodents. The value of performing exploratory drug safety studies before candidate nomination is to identify unwanted toxicities evident in a study of up to 14 days duration, as well as any potential toxicities anticipated based on a known cause for concern. In the absence of findings or the presence of findings that are judged manageable, these studies provide a greater comfort in the selection of a molecule for advancement into development with higher likelihood of success. Additional benefits of these studies are the identification of target organs to monitor in development and the selection of doses for the GLP toxicology studies. In addition, identification of the toxicity profile of a lead compound can be useful for the backup program where the goal is often an improved safety margin. In silico safety concerns might be included as part of pharmacokinetics/pharmacodynamics (PK/PD) characterization in vivo (disease) models to extract safety-relevant information and to build confidence in safety before expanding into larger regulatory animal studies.

During LO, every in silico alert is immediately followed up by the corresponding in vitro screen, in case, even in vivo studies might be frontloaded. To avoid late failures of optimized candidates, spot-checking of the potential development candidates without alerts is conducted if the resources and throughput of the assay allows. In case of screening alerts, creation of local SARs can result in significant acceleration of project by optimizing the chemical improvement rounds. Specific and tailor-made local models normally have a significantly higher accuracy, if continuously updated with new incoming screening results. Learnings and newly identified alerting substructures should be implemented in general rules and models (customized systems) to continuously improve the performance of the computational tools used for drug optimization (*see* Fig. 10).

4.4 Phase 0: After Clinical Candidate Selected (CLS) to Entry into Humans (EIH)

The main usage of in silico tools after the final candidate has been selected encompasses the assessment of potentially genotoxic impurities according to the ICH M7 Guideline as described in Chapter 3, as well as cross-reading and pathway analysis following an unexpected event in preclinical studies. Furthermore, a backup or fast-follower program will trigger dedicated in silico profiling and screening of the new molecules, based on experiences and identified issues of the frontrunner compound.

Apart from the use of in silico tools to assess genotoxic impurities, there are no computational assessments which are mandatory requirements from regulatory agencies, but in case in silico models have been applied during drug development and influenced the

testing strategy or triggered addition investigations, the information should be included in regulatory documents and adequately described.

5 Prediction of Complex End Points

5.1 Challenges in the Prediction of the Outcome of In Vivo Safety Studies

When the goal is the prediction of the outcome of certain assays, such as the Ames assay [54], in which the results can be roughly considered as binary, i.e., "yes" or "no" answer, in silico models have a higher chance to give a better performance if compared to more complex assays and studies.

The mechanism of action of a molecule leading to a specific readout plays a critical role in the predictive performance of in silico models as it is one of the biggest challenges of, for instance, QSAR models. "Do the descriptors have any physicochemical interpretation that is consistent with a known biological mechanism? [62]" is often a very difficult question to answer. In vitro chromosome damage (an assay used to establish the clastogenicity potential of test compounds) can also be considered binary (i.e., the test item is "clastogenic" or "not clastogenic"). However, the mechanisms of action that may lead to clastogenicity are manifold and may involve the interaction of the compound with a number of proteins or enzymes, the disruption of one or more biological pathways that ultimately lead to a clastogenicity outcome. This complexity is reflected in the performance of the in silico prediction tool described in Table 4 for the chromosome damage end point. Before the update of the model based on internal data and structures, the sensitivity was in the single digit, showing that the model was practically unable to identify any clastogenic compound within the validation set used. The update was successful in increasing the sensitivity value to 65 %; nonetheless, we need to bear in mind that due to the various mechanisms of action that can lead to clastogenicity, minor structural changes within a chemical class can have a large impact on the mechanism of action (e.g., the interaction with one or more proteins may be hampered, hence changing the final outcome of the assay).

Even greater challenges are offered to the prediction of the outcome of single-dose and repeat-dose toxicity in vivo studies. In the pharmaceutical industry, such studies are typically used to identify a maximum tolerated dose (MTD) and the NOAEL (non-observed-adverse-effect level) for a test compound, in addition to the identification of a general toxicity profile and significant target organs that may show toxicity upon exposure to the compound tested.

Since animal models are very complex and the number of readouts collected in such studies is extremely wide, the development of in silico models that can reliably predict such outcomes is

extremely challenging. For example, a typical repeat-dose study requires the use of a control group plus three dose groups: each animal is then carefully examined for clinical observations throughout the in-life part, including body weight and food consumption measurements as well as some behavioral evaluations; clinical pathology values are collected at different time points; urine analysis is performed; macroscopic and microscopic examinations are carried out on a number of selected organs; toxicokinetics values are then calculated using the test item concentrations measured in blood from the samples collected throughout the study, which could be of different durations, from 5 days till 39 weeks (up to 2 years for the rodent bioassays for the evaluation of carcinogenicity), and in different species.

The variations, permanent or transient, of the parameters and values briefly described above may depend on the pharmacological target, on the chemical structure and related physicochemical properties, and on background incidences due to adaptations or other factors, such as major differences in plasma exposures. Because of this variability, building an in silico model capable of predicting all these different "degrees of freedom" or "dimensions" is extremely challenging, in particular due to the fact that the identification of unequivocal mechanisms of action for whatever findings have been identified is not trivial. In addition, the development of robust SARs using the outcome of such studies is difficult because of the limited amount of publicly available data, and, even within large pharmaceutical companies, the number of chemically similar compounds tested in such long and expensive studies for each investigated pharmacological target is small (less than 5). This, of course, hampers the possibility to even develop local models since the number of similar compounds, designed for the same target, undergoing the same type of studies is rather limited.

Even if some sophisticated in silico models may become available for the prediction of the potential findings identified in, for example, repeat-dose studies, all the limitations described above and the difficulties to conduct a proper validation would make very difficult, within the pharmaceutical industry, to accept them for decision making on compounds prioritization or as guidance for chemical optimization.

5.2 Data Collection, Organization, Availability, and Interpretation for In Vivo Toxicity Studies

Within the industry, it has been recently recognized that the consolidation of the results of in vivo toxicity within appropriate tools making use of the right technology would allow the full exploitation of the knowledge that such data can provide.

Large pharma organizations can typically count on many years of drug discovery and research conducted across several sites on a significant number of therapeutic areas, pharmacological targets, and molecules. This translates into a large amount of complex datasets, stored in different repositories or Laboratory Information

Management Systems (LIMS) each designed specifically to accommodate the data type of interest (e.g., histopathology, clinical observation, clinical pathology, PK, etc.). The organization of such wealth of information to generate specific knowledge from the integration of all of these data types has been considered several times in the past by many pharmaceutical companies. However, due to limited resources or inadequate technology, the outcome of such initiatives has often been disappointing.

In more recent years, there has been a tremendous focus across industries, not only pharma, to extract knowledge and identify patterns or trends from large amounts of data, being either omics, market research, public preferences on digital movies rental [63], airplane estimated times of arrivals, or others. A lot of these initiatives often fall under the term "Big Data," generally underlying the intention of large organizations to look deeper into their databases and assess whether an improvement in the way such data are organized, stored, made accessible, and mined may provide any advantage for the business in terms of saving resources or increasing efficiency via surfacing hidden value.

Along this line, Roche has been working on a number of "Big Data" projects across several areas of research and IT. One of them had the goal to integrate all in vivo nonclinical safety data generated by the company over the past 30 years across three research sites, two in the USA and one in Switzerland. The goal was to ensure that all different data types that are part of in vivo studies (i.e., histopathology, clinical observations, PK, clinical pathology, etc.) were brought together electronically in such a way that they could all be searched and made available at the same time to the user community. The scope for such a platform, internally called SDI (i.e., Safety Data Integration), is to allow scientists to identify specific patterns of findings across species and their historical relevance and correlations between molecular structures and toxicological effects and, eventually, use the data to generate more reliable prediction algorithms. The application of a semantic data integration approach [64] for the harmonization of terms, formats, units, and taxonomy allowed the implementation of a nonclinical study warehouse including approximately 5,000 studies of different types which can be interrogated with very complex queries such as "Which compounds showed spleen hyperplasia and liver necrosis and lung leukocytosis and an AST increase >50 %?", returning an answer in a matter of seconds. The identification of studies and compound matching the query above, in the absence of properly designed data integration efforts, would have been extremely labor intensive and time consuming, if possible at all.

In addition, the SDI platform has been interfaced with other, already existing, internal databases, such as the chemical structures and the in vitro biology data repositories to further expand the data integration beyond toxicology allowing the users to assess the compound profile in almost its entirety.

5.3 Possible Model Generation

As far as model development is concerned, the advantage of the platform described in Subheading 5.2 is the high data granularity available, down to the single animal level.

One of the challenges in the development of predictive models for complex end points, such as hepatotoxicity, is that the modeler is forced to make a generic classification (hepatotoxic vs. non-hepatotoxic), often neglecting safety margins (vs. pharmacological activity), doses at which specific toxicity is seen, and ignoring the specific findings and whether it is transient or not. This is because, more often than not, such information is not easily available. All these factors make such classification relatively inaccurate: for example, paracetamol (or acetaminophen), an over-the-counter mild analgesic, commonly used to relieve headaches and reduce fever, is commonly classified as hepatotoxic (as its overdose can cause fatal liver damage [65]). However, at doses as high as up to 4 g per day in adults, paracetamol is regarded as totally safe and can comfortably be used (at lower doses, of course) even in infants. This example explains how critical and challenging a correct classification is: it is correct to classify paracetamol as hepatotoxic, since an overdose would likely cause a fatal liver failure? However, in drug development settings, what type of decision can be made on a compound predicted to be hepatotoxic by a model based on the information gathered, among others, from paracetamol? Should this molecule be discontinued and any further investigation stopped before knowing what safety margins might there be with regard to its intended therapeutic indication? Disregarding this molecule immediately after a positive prediction bears the risk of losing a potentially valuable compound. Continuing the investigations to further profile the molecule for future clinical development may be the best option to get to a more solid data-driven decision on its potential to become a drug. The bottom line is that, in this context, the prediction model will have a negligible impact on the decision.

In order to strengthen the reliability of in silico models for the prediction of complex end points, all information generated by in vivo single- and repeat-dose studies should be made available in a clear and searchable way at the highest possible level of details. This would allow experts to generate very specific models by making the correct compound classifications for very specific findings via a preliminary and careful data analysis. For example, it will be possible to have models for AST and ALT increases above 50 % vs. control groups or for the prediction of bilirubinemia, moving away from a nonspecific, for example, "hepatotoxicity" classification. This approach would, in principle, also make the identification of sound mechanisms of action for the specific observed toxicities a bit easier to address.

6 Future Perspectives

6.1 SEND Model and Data Exchange with FDA

On December 18, 2014, FDA issued the binding guidance titled "Providing Regulatory Submissions In Electronic Format—Standardized Study Data" [66] that requires Investigational New Drug (IND), New Drug Application (NDA), Abbreviated New Drug Application (ANDA), and Biologics License Application (BLA) submissions to be made in a standardized electronic format. The Clinical Data Interchange Standards Consortium (CDISC) Standard for Exchange of Nonclinical Data (SEND) is intended to guide the structure and format of standard nonclinical datasets for interchange between sponsors and contract research organizations (CROs) and for submission to the US FDA.

The current version of the SEND Implementation Guide (SENDIG v.3.0) is designed to support single-dose general toxicology, repeat-dose general toxicology, and carcinogenicity studies.

The guidance requires submission of nonclinical safety studies in SEND format for the study types currently supported. In the near future, the standard will be expanded to include additional study types, such as safety pharmacology (cardiovascular and respiratory) and developmental and reproductive toxicology, which will also be required.

The guidance further stipulates that published FDA-specific SEND validation rules will be enforced for all submitted datasets. The agency may refuse to file (for NDAs and BLAs) or refuse to receive (for ANDAs) an electronic submission that does not have study data in conformance to the required standards.

Under the guidance, supported studies (included in NDA, ANDA, and certain BLA submissions) starting after December 18, 2016, must be submitted in SEND.

For IND submissions supported studies starting after December 18, 2017 must be submitted in SEND.

Currently nonclinical safety data is provided as tabulated data within PDF study reports. Original electronic data, generated in-house, is normally stored on the originating LIMS systems until it is archived. In the case of CRO studies, original electronic data is typically not made available unless explicitly requested by the sponsor. The FDA now requires that, in addition to the PDF reports, the original electronic data also be submitted in SEND format.

While it is possible to build a SEND dataset manually, the process is labor intensive, error prone, and very difficult to validate. Given the fact that data comprising a study may come from multiple data sources, the challenge becomes unworkable.

An automated or semiautomated computerized system that can accurately and consistently transform original non-SEND data from multiple sources to the SEND standard and validate SEND

data following published rules is required. Oversight, tools, and processes for ensuring that source datasets are collected, curated, transformed to SEND, and made available for submission in an effective manner are also required.

Currently, FDA pharm/tox reviewers analyze the submitted study reports by manually extracting the tabulated data contained in the appendices of the PDF documents and loading them into any number of tools they see fit for visualizing and reviewing it. This first step is labor intensive and time consuming.

With the recently issued guidance for e-submissions, FDA reviewers have the opportunity to receive the study data directly in the appropriate format into one single platform called Nonclinical Information Management System (NIMS). FDA will use NIMS also to visualize the data, run their analyses, and draw their conclusions on the studies under review.

This approach will allow FDA reviewers to save time on data curation and formatting aspects and free resources for more in-depth scientific analyses, also leveraging the large amount of information and knowledge that NIMS will be capturing over the coming years.

Since clinical data is also electronically exchanged via standardized models (www.cdisc.org), it can be expected that one day, clinical and nonclinical data will be integrated under one single platform, which would represent a significant milestone in translational medicine arena.

6.2 Data Sharing Initiatives

Analysis of reasons for previous failures and exploitation of them should help in improving the efficiency of clinical development of new drugs and their safety profiles. So far, preclinical study reports have been rarely stored in a format that supports data mining or statistical analysis. Some pharmaceutical companies have realized these hidden treasures in their archives and started internal work to improve retrievability of their report data. It would clearly be of benefit to the whole industry to analyze these data across multiple companies in order to expand the chemical and biological space. However, extracting these data from the reports and building such a database requires considerable investment. Recent advances achieved in international initiatives, including IMI's eTOX project, have shown that sharing of preclinical data, both private and public, is achievable through the combination of legal (IP), IT, and honest broker concepts ([3, 67]; *see* Fig. 11).

The eTOX project aims to collect, extract, and organize preclinical safety data from pharmaceutical industry legacy study reports and publically available toxicology data into a searchable database to facilitate data mining and the development of innovative in silico models and software tools to predict potential safety liabilities of small molecules. The eTOX consortium consists of 13 pharmaceutical companies, 11 academic institutions, and 6 SMEs

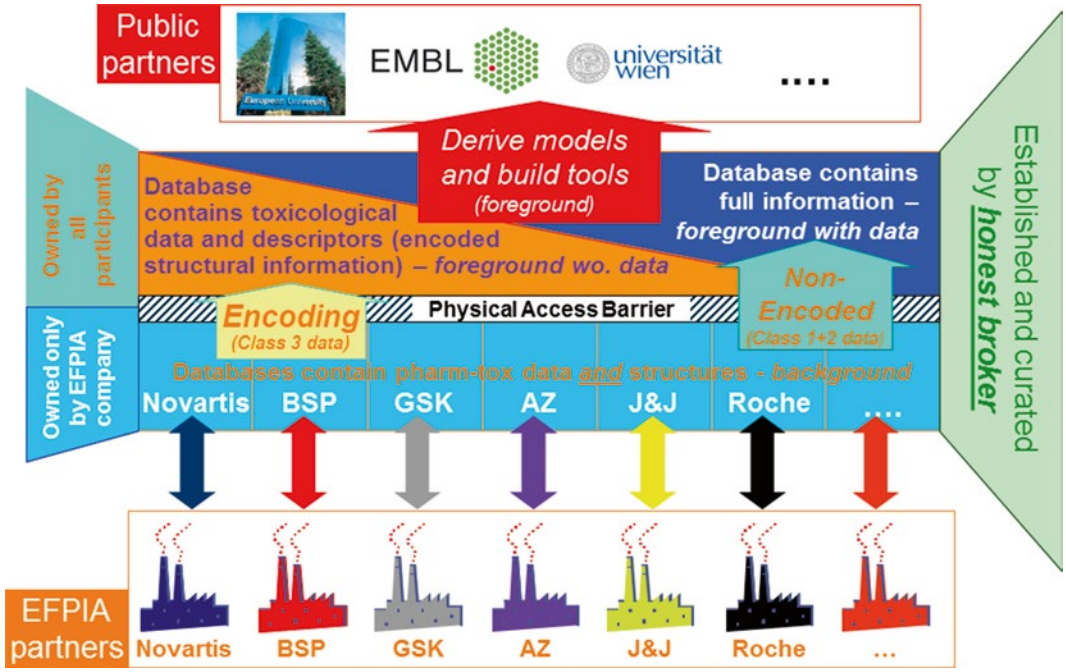


Fig. 11 The overall setup of the eTOX project

working together under the sponsorship of the Innovative Medicines Initiative (IMI) since 2010. The participating partners embrace expert knowledge in computational modeling, toxicology, pathology, and database design, liaising within the project in an integrative working environment.

After establishing an effective data sharing intellectual property (IP) protection within an "honest broker" approach (*see* Fig. 11), the project was able to compile a unique, well-curated dataset of currently more than 6,000 study reports, corresponding to ca. 1800 test compounds. The concept to divide the results from the legacy reports of the pharmaceutical companies in different "confidentiality classes" was fundamental to facilitate data sharing and overcome IP and legal hurdles. Public data (class 1) are accessible to the public on request, nonconfidential data (class 2) are open for eTOX consortium members, confidential data (class 3) are only accessible within the consortium with an additional secrecy agreement, and private data (class 4) are only for EFPIA data owners, but can be shared for model generation on request.

Treatment-related findings have been classified within the database, reflecting the interpreted study outcome of every report. A suite of ontologies, built through OntoBrowser now released by eTOX to the public domain, enables the user to directly compare observed effects or toxicities of chemically similar structures (read-across).

A new in silico tool—eTOXsys—has been developed with a single user interface, which manages search queries on the high-quality preclinical database and organizes requests to a steadily growing collection of independent prediction models. Aspects of IP rights for data sharing, definition of ontologies, design of database structure, development of in silico models, data analysis, validation, and sustainability are key aspects of the eTOX project.

References

1. Muller L, Alexander B, Christoph F, Wolfgang M, Axel P (2008) Strategies for using computational toxicology methods in pharmaceutical R&D. In: Ekins S (ed) Computational toxicology: risk assessment for pharmaceutical and environmental chemicals. Wiley, Hoboken, NY, pp 545–579
2. Muster WG et al (2008) Computational toxicology in drug development. Drug Discov Today 13(7–8):303–310
3. Cases M et al (2014) The eTOX data-sharing project to advance in silico drug-induced toxicity prediction. Int J Mol Sci 15(11): 21136–21154
4. Kavlock R (2009) The future of toxicity testing—the NRC vision and the EPA's ToxCast program national center for computational toxicology. Neurotoxicol Teratol 31(4):237–237
5. Kohonen P et al (2013) The ToxBank Data Warehouse: supporting the replacement of in vivo repeated dose systemic toxicity testing. Mol Inform 32(1):47–63
6. Arrowsmith J (2011) Trial watch: phase III and submission failures: 2007–2010. Nat Rev Drug Discov 10:87
7. Arrowsmith J, Miller P (2013) Trial watch: phase II and Phase III attrition rates 2011–2012. Nat Rev Drug Discov 12:569
8. Arrowsmith J (2011) Trial watch: phase II failures: 2008–2010. Nat Rev Drug Discov 10:328–329
9. Hillebrecht A et al (2011) Comparative evaluation of in silico systems for Ames test mutagenicity prediction: scope and limitations. Chem Res Toxicol 24:843–854
10. Sanderson DM, Earnshaw CG (1991) Computer prediction of possible toxic action from chemical structure; the DEREK system. Hum Exp Toxicol 10:261–273
11. Greene N, Judson PN, Langowski JJ, Marchant CA (1999) Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. SAR QSAR Environ Res 10:299–314
12. Judson PN (2006) Using computer reasoning about qualitative and quantitative information to predict metabolism and toxicity. In: Testa B, Kramer SD, Wunderli-Allespach H, Volkers G (eds) Pharmacokinetic profiling in drug research: biological, physicochemical, and computational strategies. Wiley, New York, pp 183–215
13. Derek Nexus (2015) http://www.lhasalimited.org/products/derek-nexus.htm
14. ToxTree version 2.6.6 (2015) https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxtree
15. CASE Ultra version 1.5.2.0 (2015) http://www.multicase.com/case-ultra
16. Leadscope Expert Alerts version 3.2.4-1 (2015) http://www.leadscope.com/expert_alerts/
17. Limited L (2015) Derek Nexus: negative predictions for bacterial mutagenicity. http://www.lhasalimited.org/products/negative-predictions-for-bacterial-mutagenicity.htm
18. Pavan M, Worth AP (2008) Publicly-accessible QSAR software tools developed by the Joint Research Centre. SAR QSAR Environ Res 19:785–799
19. Benigni R, Bossa C (2008) Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. Mutat Res 659:248–261
20. Leadscope® Genetox Expert Alerts White paper (2014) http://www.leadscope.com/white_papers/Leadscope_alerts_white_paper.pdf
21. Tropsha A (2010) Best practices for QSAR model development, validation, and exploitation. Mol Inform 29:476–488
22. Sarah Nexus (2015) http://www.lhasalimited.org/products/sarah-nexus.htm
23. Leasdcope Model Appliers (2015) http://www.leadscope.com/model_appliers/
24. van Leeuwen K, Schultz TW, Henry T, Diderich B, Veith GD (2009) Using chemical categories to fill data gaps in hazard assessment. SAR QSAR Environ Res 20:207–220
25. Bioclipse (2015) http://www.bioclipse.net/
26. Prous Institute Symmetry (2015) http://symmetry.prousresearch.com/about-symmetry/

27. Hanser T, Barber C, Rosser E, Vessey JD, Webb SJ, Werner S (2014) Self organising hypothesis networks: a new approach for representing and structuring SAR knowledge. J Cheminform 6:21
28. Sarah Nexus Methodology (2015) http://www.lhasalimited.org/products/methodology-confidence-and-interpretation-of-predictions.htm
29. Klopman G (1992) A hierarchical computer automated structure evaluation program. Quant Struct Act Relat 11:176–184
30. Klopman G (1984) Artificial intelligence approach to structure-activity studies. Computer automated structure evaluation of biological activity of organic molecules. J Am Chem Soc 106:7315–7321
31. Chakravarti SK, Saiakhov RD, Klopman G (2012) Optimizing predictive performance of CASE ultra expert system models using the applicability domains of individual toxicity alerts. J Chem Inf Model 52:2609–2618
32. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(379–423):379
33. The OECD QSAR Toolbox (2015) http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm
34. OECD Toolbox guidance document (2015)
35. OpenTox (2015) http://www.opentox.org/
36. Kazius J, McGuire R, Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem 48:312–320
37. Kuhn T, Willighangen EL, Zielesny A, Steinbeck C (2010) CDK-Taverna: an open workflow environment for cheminformatics. BMC Bioinformatics 11:159
38. Steinbeck C, Hoppe C, Kuhn S, Floris M, Guha R, Willighagen EL (2006) Recent developments of the chemistry development kit (CDK)—an open-source java library for chemo- and bioinformatics. Curr Pharm Des 12:2111–2120
39. Ekins S (2014) Progress in computational toxicology. J Pharmacol Toxicol Methods 69:115–140
40. Cheng A, Li W, Zhou Y, Shen J, Wu Z, Liu G et al (2012) admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. J Chem Inform Model 52:3099–3105
41. admetSAR (2015) http://lmmd.ecust.edu.cn:8000/
42. Valencia A, Prous J, Mora O, Sadrieh N, Valerio LG Jr (2013) A novel QSAR model of Salmonella mutagenicity and its application in the safety assessment of drug impurities. Toxicol Appl Pharmacol 273(3):427–434
43. Guideline E (2006) Guideline on the limits of genotoxic impurities. http://www.emea.europa.eu/pdfs/human/swp/519902en.pdf
44. Brigo A, Müller L (2011) Development of the threshold of toxicological concern concept and its relationship to duration of exposure. In: Teasdale A (ed) Genotoxic impurities. Wiley, New York, pp 27–63
45. ICH, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (1997) Topic Q3C. Impurities: residual solvents
46. Kroes R, Renwick AG, Cheesemann M, Kleiner J, Mangelsdorf I, Piersma A, Schilter B, Schlatter J, van Schothorst F, Vos JG, Wurtzen G (2004) Structure-based thresholds of toxicological concern (TTC): guidance for application to substances present at low levels in the diet. Food Chem Toxicol 42:65–83
47. ICH, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2002) Topic Q3A(R). Impurities testing guideline: impurities in new drug products (Revision)
48. ICH, International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) (2002) Topic Q3A(R). Impurities testing guideline: impurities in new drug substances (revision)
49. Müller L, Mauthe RJ, Riley CM, Andino MM, Antonis DD, Beels C, DeGeorge J, De Knaep AG, Ellison D, Fagerland JA, Frank R, Fritschel B, Galloway S, Harpur E, Humfrey CD, Jacks AS, Jagota N, Mackinnon J, Mohan G, Ness DK, O'Donovan MR, Smith MD, Vudathala G, Yotti L (2006) A rationale for determining, testing, and controlling specific impurities in pharmaceuticals that possess potential for genotoxicity. Regul Toxicol Pharmacol 44:198–211
50. Kasper P, Müller L (2015) Genotoxic impurities in pharmaceuticals. In: Graziano MJ, Jacobson-Kram D (eds) Genotoxicity and carcinogenicity testing of pharmaceuticals. Springer, Switzerland
51. M7 I. Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenicity risk (2014)
52. Sutter A, Amberg A, Boyer S, Brigo A, Contrera JF, Custer LL, Dobo KL, Gervais V, Glowienke S, van Gompel J, Greene N, Muster W, Nicolette J, Reddy MV, Thybaud V, Vock E, White AT, Müller L (2013) Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic

impurities. Regul Toxicol Pharmacol 67(1): 39–52

53. Brigo A, Muster W, Singer T (2015) Comparative assessment of several in silico systems and models to predict the outcome of the Ames mutagenicity assay. In: Toxicology SO (ed) Society of Toxicology Annual Meeting 2015. San Diego, California, USA
54. Ames BN, Durston WE, Yamasaki E, Lee FD (1973) Carcinogens are mutagens: a simple test system combining liver homogenate for activation and bacteria for detection. Proc Natl Acad Sci U S A 70:2281–2285
55. Escobar PA, Kemper RA, Tarca J, Nicolette J, Kenyon M, Glowienke S, Sawant SG, Christensen J, Johnson TE, McKnight C, Ward G, Galloway SM, Custer L, Gocke E, O'Donovan MR, Braun K, Snyder RD, Mahadevan B (2013) Bacterial mutagenicity screening in the pharmaceutical industry. Mutat Res 752:99–118
56. Hansen K, Mika S, Schroeter T, Sutter A, ter Laak A, Steger-Hartmann T, Heinrich N, Müller KR (2009) Benchmark data set for in silico prediction of Ames mutagenicity. J Chem Inf Model 49:2077–2081
57. Reuters T (2015) Metacore—data-mining and pathway analysis http://thomsonreuters.com/en/products-services/pharma-life-sciences/pharmaceutical-research/metacore.html
58. Fischer H, et al (2001) Prediction of in vitro phospholipidosis of drugs by means of their amphiphilic properties. In: Rational approaches to drug design, p 286–289
59. Kruhlak NL et al (2007) Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. Adv Drug Deliv Rev 59(1):43–55
60. CompuDrug. MetabolExpert (2015) http://www.compudrug.com/metabolexpert
61. Discovery M (2015) MetaSite http://www.moldiscovery.com/software/metasite/
62. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57:4977–5010
63. Piatetsky-Shapiro G (2012) Big data hype (and reality). https://hbr.org/2012/10/big-data-hype-and-reality
64. Sciences PL (2015) http://www.point-crosslifesciences.com/
65. James LP, Mayeuy PR, Hinson JA (2003) Acetaminophen-induced hepatotoxicity. Drug Metab Dispos 31(12):1499–1506
66. FDA (2014) http://www.fda.gov/downloads/Drugs/Guidances/UCM292334.pdf
67. Briggs K et al (2012) Inroads to predict in vivo toxicology—an introduction to the eTOX project. Int J Mol Sci 13(3):3820–3846

Chapter 21

The Consultancy Activity on In Silico Models for Genotoxic Prediction of Pharmaceutical Impurities

Manuela Pavan, Simona Kovarich, Arianna Bassan, Lorenza Broccardo, Chihae Yang, and Elena Fioravanzo

Abstract

The toxicological assessment of DNA-reactive/mutagenic or clastogenic impurities plays an important role in the regulatory process for pharmaceuticals; in this context, in silico structure-based approaches are applied as primary tools for the evaluation of the mutagenic potential of the drug impurities. The general recommendations regarding such use of in silico methods are provided in the recent ICH M7 guideline stating that computational (in silico) toxicology assessment should be performed using two (Q)SAR prediction methodologies complementing each other: a statistical-based method and an expert rule-based method.

Based on our consultant experience, we describe here a framework for in silico assessment of mutagenic potential of drug impurities. Two main applications of in silico methods are presented: (1) support and optimization of drug synthesis processes by providing early indication of potential genotoxic impurities and (2) regulatory evaluation of genotoxic potential of impurities in compliance with the ICH M7 guideline. Some critical case studies are also discussed.

Key words Genotoxic impurities, In silico methods, (Q)SAR, Statistical-based methods, Expert rule-based methods, ICH M7

1 Introduction

In silico modeling, such as (quantitative) structure-activity relationships ((Q)SARs) and molecular modeling, have been widely used in drug discovery, drug development, and regulatory purposes. In the current chapter, the focus will be primarily on the use of (Q)SARs for the evaluation of the genotoxic potential of drug impurities.

Drug impurities are defined as any component of the drug substance or drug product that is not the drug substance or an excipient (i.e., inactive constituent) and that can arise from drug synthesis or subsequent degradation, as well as from external

Emilio Benfenati (ed.), *In Silico Methods for Predicting Drug Toxicity*, Methods in Molecular Biology, vol. 1425,
DOI 10.1007/978-1-4939-3609-0_21, © Springer Science+Business Media New York 2016

contamination. In the regulatory framework for pharmaceuticals, specific guidelines exist for the qualification and control of the majority of the impurities, e.g., the International Conference on Harmonisation (ICH) Quality Guidelines Q3A ("Impurities in New Drug Substances") [1] and Q3B ("Impurities in New Drug Products") [2] and the ICH Multidisciplinary Guideline M3 ("Nonclinical Safety Studies for the Conduct of Human Clinical Trials and Marketing Authorizations for Pharmaceuticals") [3]. Recently, a new guideline (ICH M7) was introduced for the identification, categorization, qualification, and control of DNA-reactive (mutagenic) impurities to limit the potential carcinogenic risk of drugs [4]. The ICH M7 guideline outlines recommendations on the use of in silico structure-based methods for genotoxicity assessment of drug impurities. According to ICH M7, computational (in silico) toxicology assessment should be performed using two (Q)SAR prediction methodologies complementing each other: a statistical-based method and an expert rule-based method. The employed (Q)SAR models should follow the internationally recognized principles for QSAR validation as defined by the Organisation for Economic Co-operation and Development (OECD) [5, 6]. According to the OECD principles, a QSAR model should (1) provide predictions for a defined endpoint; (2) be based on an unambiguous algorithm; (3) have a defined domain of applicability; (4) be internally and externally validated by applying appropriate measures of goodness of fit, robustness, and predictivity; and (5) provide a mechanistic interpretation of the prediction, when possible. The guideline recommendations state also that the outcome of any computer system-based analysis should be reviewed with the use of expert knowledge in order to provide additional supportive evidence on relevance of any positive or negative prediction and to elucidate underlying reasons in case of conflicting results. The crucial role of the expert in the final assessment is also highlighted in the literature [7–9].

In the present chapter, a practical approach for in silico assessment of mutagenic potential of drug impurities is described. The focus is on two main applications: (1) support and optimization of drug synthesis processes by providing early indication of potential genotoxic impurities and (2) regulatory evaluation of genotoxic potential of impurities in compliance with the ICH M7 guideline. Different approaches are proposed according to the specific application of the in silico assessment, and some critical case studies are discussed based on our experience.

2 Materials

In the toxicity framework, in silico predictions can be obtained by three main computational approaches: QSAR statistical-based methodologies, (Q)SAR expert rule-based methodologies, and

grouping approaches, which include read-across and chemical category formation. A brief description of the three approaches, including the underlying theory and examples of tools implementing these methods, is described in the following paragraphs.

2.1 QSAR Statistical-Based Methodology

The statistical-based QSAR method is a quantitative (mathematical) relationship between a numerical representation of the chemical structure (i.e., molecular descriptors) and a biological activity, physicochemical or fate property. Statistical-based QSARs are models based on experimental data, which extract the knowledge directly through a process of data mining and knowledge engineering. Thousands of molecular descriptors encoding for mono-, bi-, or tridimensional structural features (e.g., atom counters, topological descriptors, symmetry and steric descriptors) or chemical properties (e.g., LogP or electronic properties) have been proposed and derived from different theories and approaches, with the aim to provide an "exhaustive" description of the chemical structure. At the same time, a wide range of algorithms are now available to identify the quantitative relationship between the structure and the studied property/activity and to build statistically robust and predictive QSAR models (e.g., multiple linear regression (MLR), partial least squares (PLS) regression, artificial neural networks (ANN), etc.). It follows that the majority of statistical-based QSARs are characterized by robust validation techniques and high predicting performances, and can provide predictions also when the mechanism of action is unknown. Additionally, several mathematical/chemometrical metrics have been developed to define model applicability domain and to measure the level of extrapolation. On the other hand, in some cases, their predictions could miss a mechanistic reasoning and a clear interpretation, especially when based on complex algorithms and molecular descriptors, thus resulting "nontransparent" to the end user.

Nowadays, several tools (both commercial and freeware) are available coding QSAR statistical models for the prediction of mutagenic/genotoxic potential [10–14]. We routinely use an array of commercial and freely available tools in a weight of evidence approach. All the predictors we use fulfill the OECD principles for QSAR validation and are characterized by (1) wide and heterogeneous training set collected from valid sources (e.g., FDA—US Food and Drug Administration), (2) high robustness and external predictivity, (3) wide applicability domain, and (4) defined parameters for reliability assessment. Additionally they allow the user to visualize structure and experimental data of structural analogues, thus providing supporting information to further assess the prediction. A brief description of these tools is as follows:

- *ACD/Percepta Impurity Profiling* [15, 16] provides a battery of in silico models to accurately assess the genotoxic

and carcinogenic potential of chemicals. The impurity profiling module is a result of the collaboration between ACD/Labs and FDA Center for Food Safety and Applied Nutrition (CFSAN). This module includes probabilistic predictive models for 21 different endpoints that cover various mechanisms of hazardous activity (including mutagenicity, clastogenicity, DNA damage mechanisms, carcinogenicity, and endocrine disruption mechanisms) and that are based on experimental data obtained from FDA. Probabilistic predictive models were developed using GALAS modeling methodology [17]. Each GALAS model consists of two parts: (1) a global (baseline) model, built using binomial PLS method based on fragmental descriptors, that reflects a "cumulative" mutagenicity potential, and (2) local corrections that are applied to baseline predictions using a special similarity-based routine, after performing an analysis for the most similar compounds used in the training set. The reliability of prediction is assessed in terms of reliability index (RI), which ranges from 0 to 1 and takes into account the similarity of the target with the training set compounds and the consistency of experimental values for similar compounds. A "positive" or "negative" call is then provided if the compound can be reliably classified on the basis of p-value (i.e., probability that a compound will result in a positive test in the respective assay) and RI values ("undefined" otherwise).

- *ChemTunes Studio* is a knowledge base software consisting of experimental in vitro and in vivo toxicity information (QC'ed by experts) and in silico models for a series of human health toxicity endpoints, comprising the key genetic toxicity endpoints (i.e., Ames mutagenicity, chromosome aberration, and in vivo micronucleus). The software is made of multiple components, including genotoxic chemotypes (structural alerts); mechanistically informed (mode-of-action driven) QSAR models, i.e., an approach used at US FDA CERES (Chemical Evaluation and Risk Estimation System) [18, 19]; and comparison of the prediction results to structural analogues. A mathematically rigorous and quantitative weight of evidence (WoE) decision theory approach is used to obtain the final overall assessment and to provide a quantitative estimation of the uncertainty associated with the prediction. All ChemTunes Studio QSAR models consist of chemical mode-of-action category models as well as a general global model. The computational modeling approach is a hybrid of partial least squares (PLS)/ordinal logistic regression methods. For model building, global molecular and shape descrip-

tors (from CORINA Symphony [20]) and quantum-mechanic parameters are used. The models return probabilistic predictions (positive and negative probabilities plus a quantitative estimate of the associated uncertainty) and an overall prediction (positive/negative/equivocal). Applicability domain analysis reports whether the target compound is out of domain. QSARs for bacterial reverse mutagenesis (Ames mutagenicity) are based on selected studies for more than 2200 structures, compiled from various sources, and including *S. typhimurium* and *E. coli* strains with and without metabolic activation.

- *Leadscope Model Applier/Genetox QSAR Statistical Suite* [21] is a chemoinformatic platform that provides QSARs for the prediction of potential toxicity and adverse human clinical effects, including the microbial in vitro *Salmonella* mutagenicity model that is used by the US FDA (Food and Drug Administration) in their testing under the ICH M7 Guidance for impurities [22–24]. The in vitro *Salmonella* mutagenicity QSAR model was constructed by the FDA scientists based on a training set of over 3500 compounds (including both proprietary and nonproprietary data). The model is based on a wide set of molecular descriptors, including 369 substructural features and seven calculated properties, and on partial logistic regression (PLS) modeling technique. Model predictions consist of four possible results, i.e., "positive," "negative," "indeterminate," or "not in domain," and probability of a positive result. Predictions are provided together with several parameters, which can be used to assess the prediction in terms of applicability domain (e.g., the presence in the target compound of model training set structural features and the presence of structural analogues in the training set).
- *VEGA/CAESAR Mutagenicity* model is a QSAR model predicting mutagenicity developed under the EU project CAESAR [25] and implemented in the VEGA platform [26]. The QSAR model is based on a dataset of 4225 compounds and consists of an integrated model made of two complementary techniques: a machine learning algorithm (SVM), to build an early model with the best statistical accuracy, equipped with an expert facility for false negative removal based on known structural alerts, to refine its predictions. The reliability of predictions is assessed using an Applicability Domain Index (ADI) that ranges from 0 to 1 and is calculated by grouping several other indices, each one taking into account a particular issue of the applicability domain (i.e., the presence of similar compounds in the training set, the consistency of their experimental data and

their prediction accuracy, the presence in the target of structural fragments possessed by training set compounds, and the range of values of modeling descriptors).

2.2 (Q)SAR Expert Rule-Based Methodology

The (Q)SAR expert rule-based (or knowledge-based) method relies on rules derived from toxicological knowledge, which are likely to have strong mechanistic basis, used to make predictions about a defined adverse effect. In the expert rule-based systems, human experts identify structural fragments related to the studied effect. The examination of a series of chemicals sharing the same fragment ("structural alert"—SA) is used to detect the toxic effect (e.g., genotoxic or not); the chemical information is simply the fragment and the algorithm is, in this case, the rule. The expert rule-based systems have several advantages, e.g., they are mechanistically connected to the predicted activity, provide reasoning for the predictions, and in many cases support the prediction with literature references or expert knowledge. On the other side, applicability domain measures for expert systems are not well defined [27], and usually it is not possible to discriminate active from inactive chemicals bearing the same structural alert. The accuracy in prediction is mostly comparable to statistical-based QSARs; however, expert systems tend to exhibit a higher sensitivity at the cost of a lower specificity (SAs are conservative), whereas the statistical-based QSARs show the opposite behavior [28].

Several tools (both commercial and freeware) are now available coding expert rule-based systems [10–14]. In some tools, expert systems are combined with statistical-based models (the so-called hybrid systems), in order to provide supporting knowledge-based evidence to QSAR predictions. For our consultant activities, we routinely use an array of commercial and freely available tools in a weight of evidence approach. The predictors in use are based on wide sets of chemicals and alerts and provide means to assess the reliability of predictions. A brief description of these tools is as follows:

- *ACD/Percepta Impurity Profiling* [15, 16] is supplemented with a knowledge-based expert system that identifies potentially hazardous structural fragments that could be responsible for genotoxic and/or carcinogenic activity of the compound of interest. The expert system contains a list of 70 alerting groups of toxicophores, of which 33 represent mutagens, 24 clastogens, and 13 epigenetic carcinogens (androgens, peroxisome proliferators, etc.). The alert list is not limited to directly acting substructures, such as planar polycyclic arenes, aromatic amines, quinones, and N-nitro and N-nitroso groups, but also includes various fragments that may undergo biotransformation to reactive intermediates. Each hazardous fragment is provided with a

description of its mechanism of action, literature references, and *z*-scores. *z*-Scores show whether the presence of the fragment leads to a statistically significant increase in the proportion of compounds with a positive test result for a particular assay. The identified alerting groups are highlighted on the structure of the molecule and the five most structurally similar structures from the training set, along with experimental results, are shown.

- *ChemTunes Studio* includes, in addition to QSAR statistical-based models, genotoxic chemotypes (structural alerts), developed from mechanistic hypothesis; each alert is provided with likelihood prioritization, so that alerts can be used when combining the different information at the WoE stage. A knowledgebase was built and curated for a large dataset (over 8000 compounds) of Ames mutagenicity data from public sources. The reliability of each alert is determined by exploring the ability of the alert to hit positive compounds in a large training set. Different training sets were used for the QSAR models and the alerts, so that predictions from these are independent.
- *Leadscope Model Applier/Genetox Expert Alerts Suite* is implemented as part of the Leadscope Model Applier (in addition to the existing statistical-based QSAR model) [21]. To develop this system, an initial library of mutagenicity structural alerts was identified from the literature. Information on plausible mechanisms was collected as well as the structural definitions. Factors that deactivate the alerts were also identified from the literature and through an analysis of the corresponding data using the Leadscope data mining software. Over 200 distinct alerts are encoded in the system. These alerts were further validated against a reference database of over 7000 chemicals with known bacterial mutagenesis results. A confidence score based upon information collected for each alert is provided alongside the positive or negative call. Up to ten structurally similar structures from the alert reference set, along with experimental results, are provided.
- *Toxtree* [29] is a flexible and user-friendly open-source application that places chemicals into categories and predicts various kinds of toxic effects by applying decision tree approaches. The decision tree for estimating mutagenicity is based on discriminant analysis and structural rules as described in Benigni et al. [30]. It estimates in vitro (Ames test) mutagenicity, based on a list of 30 structural alerts (SAs). As one or more SAs embedded in a molecular structure are recognized, the system flags the potential mutagenicity of the chemical. The use of Toxtree Benigni-Bossa

decision tree implemented in VEGA platform [26] allows the user to assess the reliability of predictions by means of the Applicability Domain Index (ADI) calculated in VEGA and to visualize chemical structure and experimental data for the most similar structures in Toxtree alert training set.

2.3 Grouping Approaches: Read-Across Methodology

Chemical grouping approaches are based on the formation of chemical "categories" or "analogues," composed by groups of chemicals whose physicochemical, (eco-)toxicological, and/or environmental fate properties are likely to be similar or follow a regular pattern. This can be the result of a structural similarity or other similarity characteristics (e.g., common mechanism of action). In principle, the chemical category is composed by several members, enabling the detection of trends across endpoints, while the grouping by analogue approach is based on a limited number of chemicals, where trends in properties are not apparent [31]. In these cases, predictions are generated by applying the "read-across" method. In the read-across technique, the endpoint information for one chemical is used to predict the same endpoint for another chemical, which is considered "similar" in some way (usually based on structural similarity). The chemical(s) being used to make an estimate is commonly referred to as a "source chemical(s)," whereas the chemical for which the endpoint is being estimated is referred to as a "target chemical." The read-across methodology is currently accepted to fill data gaps in the regulatory framework, basically for the transparency and interpretability of the approach and of the final outcome. However, read-across is not a formalized approach (i.e., it is not based on a defined and reproducible algorithm), and the obtained predictions strongly depend on the expert judgment. For these reasons, specific guidelines on how to perform a read-across study in order to be accepted for regulatory purposes (e.g., REACH) have been developed [32]. According to this guideline, any read-across analysis should be supported by a detailed documentation to be provided according to the defined read-across reporting formats [31, 33].

The OECD QSAR Toolbox [34] is the main tool we use to perform read-across predictions [35]. It was developed by the OECD to use (Q)SAR methodologies to group chemicals into categories and to fill data gaps by read-across and trend analysis. It is currently recommended and released by the European Chemicals Agency (ECHA) in collaboration with OECD. The Toolbox incorporates information and tools from various sources into a logical workflow, which supports the user to carry out read-across studies through the identification of relevant structural characteristics and potential mechanism or mode of action of a target chemical, the identification of other chemicals that have the same structural characteristics, and/or mechanism or mode of action and the use of existing experimental data to fill the data gaps. Another freely avail-

able software useful to assist users for read-across evaluations is ToxRead [36]. ToxRead was recently developed by IRCCS (Istituto di Ricerche Farmacologiche Mario Negri), Politecnico di Milano, and KODE within a joint collaboration between the LIFE projects CALEIDOS and PROSIL and offers a workflow to generate read-across predictions with high reproducibility.

2.4 Weight of Evidence Approach

Any predictive model is by definition a simulation of reality, and therefore it will never be completely accurate. The same applies to (Q)SARs. As discussed in the previous paragraphs, each computational approach, i.e., statistical-based, expert rule-based, or read-across approach, has its own advantages and weaknesses. Likewise, each (Q)SAR model is characterized by distinctive predictive performances (e.g., sensitivity versus specificity) and a defined applicability domain (i.e., no QSAR model can be applied to every chemical of interest), thus providing different partial "views" of the whole picture. Thus, the most reasonable way to get the best out of several views and achieve accurate predictions is to combine predictions from different models and approaches in a weight of evidence approach [37–39]. A weight of evidence (WoE) approach involves an assessment of the values and relative weights of different pieces of available information [40]; in our case, it implies an assessment of different in silico predictions taking into account the reliability of each prediction and the concordance among different predictions. This can be achieved either in an objective way by using a formalized procedure or by using expert judgment. Some tools, such as ChemTunes and Leadscope Model Applier, provide algorithms for the calculation of WoE (or consensus) predictions based on the combination of predictions from statistical- and expert rule-based models as well as experimental data. It has been broadly demonstrated that the complementary use of statistical-based and expert-based approaches, supplemented by expert knowledge, improves prediction accuracy [8, 11, 14, 41].

3 Methods

3.1 Early Indication of Potential Genotoxic Impurities

In silico methods can be efficiently employed in the early stages of drug development for the screening and identification of potential genotoxic impurities, thus providing useful information to optimize the design of the synthesis scheme. When in silico methods are used for screening purposes, the integration of statistical-based and knowledge-based approaches is not mandatory, and a less detailed documentation of the burden of proof is required. Our procedure for an early indication, by means of in silico methods, of the potential genotoxicity of impurities is described and summarized in Fig. 1.

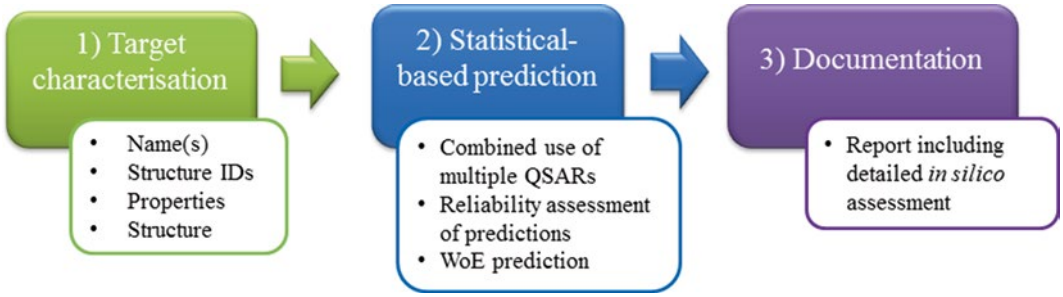


Fig. 1 Workflow for early indication of potential genotoxic impurities

1. Characterization of the target impurity by means of chemical names, registry number, structure identifiers (e.g., SMILES, InChI), chemical structure, and properties (e.g., molecular weight, molecular formula).
2. QSAR statistical-based prediction of bacterial mutagenicity:
   (a) Combined use of multiple tools for the prediction of genotoxicity as microbial in vitro *Salmonella* (Ames test). For screening purposes, statistical-based QSAR models are usually preferred than knowledge-based approaches because of their higher accuracy and wider applicability [42]. Among the available predictors based on a statistical approach, we are currently using ACD/Labs Percepta, Leadscope Model Applier, and the CAESAR Mutagenicity model implemented in VEGA, while ChemTunes is going to be integrated. These predictors are particularly indicated for screening purposes since they are characterized by wide and heterogeneous training set (including drug substances), external predictivity, and wide applicability domains.
   (b) Assessment of the prediction reliability taking into account multiple issues, e.g., (i) whether the target impurity falls within the applicability domain of the model, (ii) whether and how the target impurity is represented in the training set by analyzing the structural analogues included in the training sets, (iii) prediction accuracy of the identified analogues, and (iv) consistency between the analogues' experimental test results (Ames test) and the prediction for the target impurity. Identification of the proper analogues is a critical step and depends on the methodology used to measure chemical similarity. Defining chemical similarity measures to infer mutagenic potential as well as approaches to assess the reliability of predictions is still an open challenge [43].
   (c) Generation of a WoE prediction, i.e., positive/negative for microbial in vitro *Salmonella*, taking into account only reli-

able predictions. If different predictors, based on different training molecules, molecular descriptors, and modeling approaches, lead to consistent results, then a higher level of confidence in the in silico prediction is achieved. If equally reliable but not consistent results are provided by different predictors, then the most conservative outcome, i.e., positive, should be concluded. Examples on how to deal with critical case studies, e.g., not consistent and/or unreliable predictions, are commented in Subheading 4 (**Notes 1–5**).

3. Documentation of the results. The predictions provided by the different tools together with the performed WoE analysis are described in a detailed report.

3.2 Regulatory Evaluation of Genotoxic Potential of Impurities (ICH M7 Guideline)

According to ICH M7 guideline, hazard assessment of genotoxic impurities first involves an analysis of actual and potential impurities, based on experimental carcinogenicity and bacterial mutagenicity data available from database and literature. If such data are not available, in silico (Q)SAR assessment of the impurities should be performed to provide predictions for bacterial mutagenicity. As a result of the hazard assessment, drug impurities are assigned to one of the five classes summarized in Fig. 2, and specific control actions are suggested [4].

The ICH M7 guideline states that the computational toxicology assessment should be performed by using two (Q)SAR

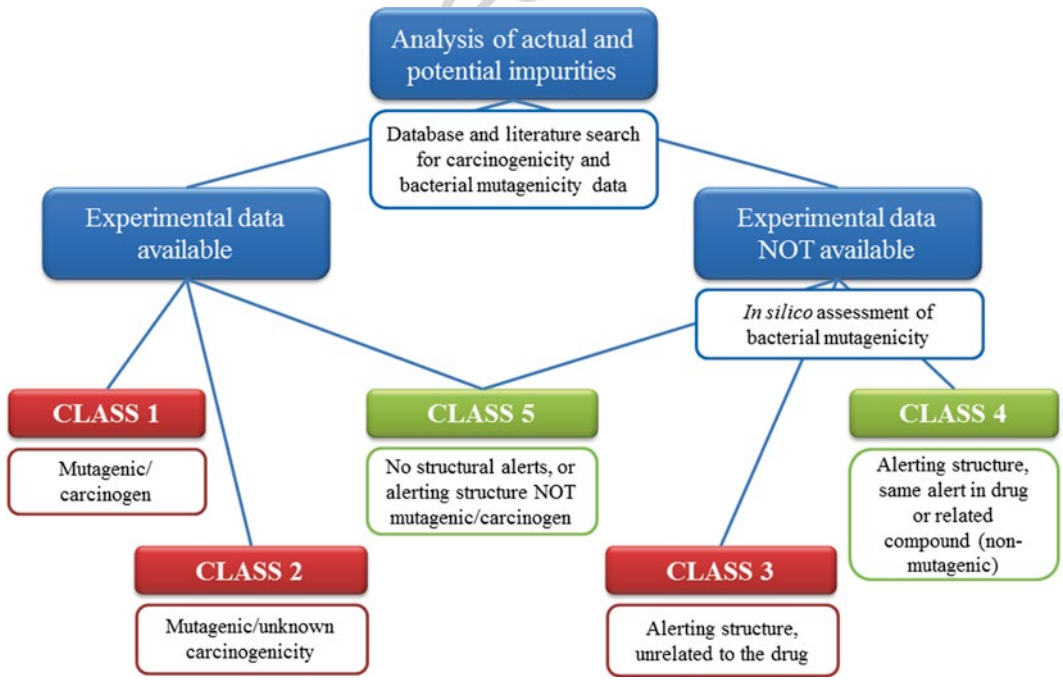


Fig. 2 Impurities classification with respect to mutagenic and carcinogenic potential

Table 1
Examples of critical case studies for in silico assessment of genotoxic impurities

No.	Statistical-based WoE	Expert rule-based WoE	Read-across study	Conclusive in silico assessment
1	NEGATIVE	OUT OF DOMAIN/ INCONCLUSIVE	NEGATIVE based on negative source chemical(s) (e.g., the API or structural related impurities)	NEGATIVE
2	OUT OF DOMAIN/ INCONCLUSIVE	NEGATIVE	NEGATIVE based on negative source chemical(s)	NEGATIVE
3	OUT OF DOMAIN/ INCONCLUSIVE	POSITIVE based on alert X	NEGATIVE based on negative source chemical(s) possessing the same alert X	NEGATIVE
4	NEGATIVE	POSITIVE based on alert X	NEGATIVE based on negative source chemical(s) possessing the same alert X	NEGATIVE
5	NEGATIVE	POSITIVE based on alert X	NOT FEASIBLE/ POSITIVE positive source chemical(s) possessing the same alert X	POSITIVE

prediction methodologies that complement each other, i.e., a statistical-based and an expert rule-based methodology. In addition, expert analysis including read-across is applied to provide additional supportive evidence on the predictions and/or to solve conflicting results. It is here described our stepwise procedure for regulatory in silico assessment of genotoxic impurities. The procedure is also summarized in the workflow of Fig. 3.

1. Characterization of the target impurity (i.e., chemical names, structure identifiers, chemical structure, and properties)
2. QSAR statistical-based prediction of bacterial mutagenicity:
   (a) Combined use of multiple statistical-based QSAR models for the prediction of genotoxicity as microbial in vitro *Salmonella* (Ames test).
   (b) Assessment of the reliability of the predictions provided by the individual statistical-based tools as described in Subheading 3.1 (**step 2b**).
   (c) Computation of the statistical-based WoE prediction, i.e., positive/negative for microbial in vitro *Salmonella*, based

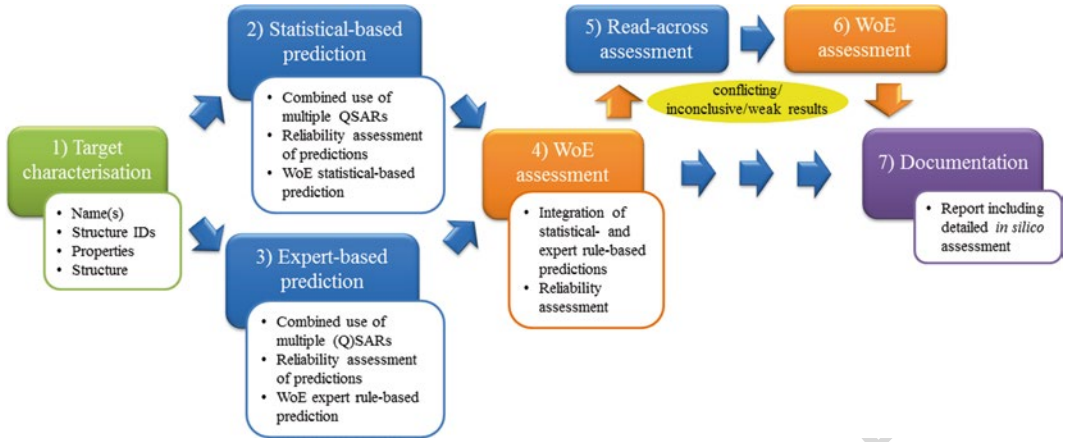


Fig. 3 Workflow for regulatory evaluation of potential genotoxic impurities

on the employed statistical-based tools. The level of confidence of the WoE prediction (e.g., unreliable, borderline, moderate, or highly reliable) is defined taking into account the reliability and consistency of the predictions obtained by the individual employed statistical-based tools.

3. (Q)SAR expert rule-based prediction of bacterial mutagenicity:

   (a) Combined used of multiple expert rule-based methods for the prediction of genotoxicity as microbial in vitro *Salmonella* (Ames test). Among the available knowledge-based tools, we are currently using ACD/Labs Percepta, Leadscope Model Applier, and the Toxtree in vitro mutagenicity (Benigni-Bossa) decision tree implemented in VEGA. The novel expert system implemented in ChemTunes based on genotoxic chemotypes is going to be integrated in our in silico assessment. These tools provide a positive, negative, or inconclusive prediction based on the identification of one or more structural alerts for mutagenicity, as well as the means to assess the reliability of the prediction (as discussed in the next step). Particular attention is paid to negative ("non-genotoxic") predictions based on the absence of structural alerts. In fact, the absence of any known structural alerts is NOT a sufficient evidence for a lack of effect, and there is the possibility that the target impurity may act through an unknown mechanism of action, for which structural alerts have not been developed yet.

   (b) Assessment of the reliability of the predictions provided by the expert SA-based tools. Although structural alerts often lack an adequately defined applicability domain [27], the level of confidence of the predictions can be assessed focusing on the following issues: (i) whether the target impurity

is sufficiently represented in the training set, in terms of structural similarity, chemical fragments, or other structural features represented in the training set; (ii) relevance of the identified alert, i.e., the alert is characterized by a statistically significant higher frequency in genotoxic compounds compared to non-genotoxic (from the training set); (iii) precision of the identified alert, i.e., accuracy of the alert in the correctly predicted genotoxic compounds (i.e., true positive rate); and (iv) consistency between the experimental test results (Ames test) of the identified analogues (particularly those sharing the same alert(s)) and the predicted outcome of the target impurity. If no structural alerts for genotoxicity are identified, a proper reliability assessment is not applicable. In these cases, a detailed analysis of the structural analogues with no alerts and the precision of the expert system toward training compounds with no alerts is recommended [13].

(c) Generation of the expert rule-based WoE prediction, i.e., positive/negative for microbial in vitro *Salmonella*, based on the employed expert rule-based tools. The level of confidence of the WoE prediction is defined taking into account the reliability and consistency of the predictions obtained by individual tools.

4. Generation of the final WoE prediction, i.e., positive/negative for microbial in vitro *Salmonella*, based on the integration of the outcome of the statistical-based and expert rule-based WoE predictions. The level of confidence of the WoE prediction is defined taking into account the reliability and consistency of the predictions obtained by the two approaches. In case of conflicting results and/or weak WoE assessment (i.e., low reliability), either we conclude for a predicted genotoxic potential (conservative scenario) or, preferably, we integrate the in silico assessment with a read-across study (as described in **step 5**). It is important to highlight that the WoE approach is not an automatic procedure, rather an assessment based on expert judgment performed on a case-by-case analysis of the predictions. Examples on how to deal with some critical case studies, e.g., not consistent and/or unreliable predictions, are commented in Subheading 4 (**Notes 1–5**).

5. Read-across study to provide additional supportive evidence on the predictions and/or to solve conflicting results. From our consultancy experience, the source chemical(s) is often suggested by the commissioner and could be either the API (active pharmaceutical ingredient), compounds related to the drug substance (e.g., process intermediates), or structurally related impurities, for which the commissioner already conducted an experimental Ames test. Alternatively, an extensive

search in the literature and in open databases (e.g., DSSTox [44], ECHA CHEM [45], NTP [46], GENE-TOX [47], etc.) is performed to identify the most appropriate source(s) for the target impurity. The read-across study is performed and documented according to the guidance document on the grouping of chemicals (including read-across and chemical categories) [31–33]. The OECD QSAR toolbox is employed to identify the functional groups (by applying the Organic Functional Groups (OFG) system) and to profile the source and target chemicals by describing their foreseen mechanism of action relevant for mutagenic activity. Two general mechanistic profilers, namely, DNA binding by OECD and DNA binding by OASIS v.1.2, and three endpoint-specific profilers, namely, DNA alerts for AMES, MN, and CA by OASIS v.1.2, in vitro mutagenicity (Ames test), and in vivo mutagenicity (micronucleus) alerts by ISS, are used being the most meaningful profilers for genotoxicity available in the toolbox [48].

6. Conclusion from the in silico assessment on the potential genotoxicity of the target impurity, based on results of the two QSAR prediction methodologies, i.e., a statistical-based method and an expert rule-based method, and the supporting evidence coming from the read-across study.
7. Documentation of the results. The predictions provided by the different tools and approaches, together with the performed WoE analysis, are described in a detailed report.

4 Notes

The interpretation of results from a (Q)SAR assessment of genotoxic impurities is not always straightforward, and several issues are commonly encountered. Thus, the role of the expert is crucial to build up a WoE prediction by an integrated approach, which considers information gained by various techniques, to provide additional supportive evidence on relevance of any positive or negative prediction and to elucidate underlying reasons in case of conflicting or inconclusive results. Some examples of critical and real case studies are reported and illustrated in Table 1. In all cases, three statistical-based models, i.e., ACD/Percepta Impurity Profiling (in vitro *Salmonella* model), Leadscope Model Applier/Genetox QSAR Statistical Suite (microbial in vitro *Salmonella* model), and VEGA/CAESAR Mutagenicity model, were employed together with three expert rule-based systems, i.e., ACD/Percepta Impurity Profiling (in vitro *Salmonella* expert system), Leadscope Model Applier/Genetox QSAR Expert Suite (Bacterial Mutation), and the Toxtree in vitro mutagenicity (Benigni-Bossa) decision tree implemented in VEGA platform.

1. Case study 1: The target impurity is reliably predicted as negative by the statistical-based approach, while the prediction obtained by the expert rule-based approach is not reliable ("out of domain") or inconclusive. In this case, it is not possible to derive a robust WoE prediction, since two approaches are required by the ICH M7 regulation, and the read-across approach is suggested to provide further evidence of the negative prediction.
2. Case study 2: The prediction obtained from the statistical-based approach is not reliable ("out of domain"), or inconclusive, while the outcome of the expert rule-based approach is negative, based on the absence of structural alerts for genotoxicity. Again, a read-across study is suggested to provide further evidence of the negative prediction.
3. Case study 3: The prediction obtained from the statistical-based approach is not reliable, or inconclusive, while the outcome of the expert rule-based approach is a reliable positive prediction, based on the detection of one or more structural alerts for genotoxicity. In this case, it is not possible to derive a robust WoE prediction, and the read-across approach is suggested to verify whether the presence of the alert induces (or not) a positive effect. If the identified source chemical (e.g., the API or structural related impurities) shares with the target impurity the same structural alert (e.g., same structural alert in the same position and environment in the impurity and the source) and the source chemical is non-mutagenic, then the target impurity is predicted negative by the read-across (Class 4 according to ICH M7). In this case, in agreement with the ICH M7 guideline, the read-across study overturns the expert rule-based prediction, and the final in silico assessment concludes for a negative prediction.
4. Case study 4: Conflicting predictions are obtained applying the two different methodologies, e.g., negative outcome obtained with the statistical-based approach and positive outcome obtained with the expert rule-based system. The WoE assessment, based on a precautionary approach, would conclude for a positive prediction, leading possibly to a false positive. The read-across approach is thus suggested to solve conflicting results. As discussed in case study 3, if the impurity shares with the source chemical the same structural alert and the source chemical is non-mutagenic, then the target impurity is predicted negative by the read-across (Class 4 according to ICH M7). Thus, the read-across study overturns the WoE assessment based on statistical-based and expert rule-based predictions, and the final in silico assessment concludes for a negative prediction.

5. Case study 5: Conflicting predictions are obtained applying the two different methodologies, e.g., negative outcome obtained with the statistical-based approach and positive outcome obtained with the expert rule-based system. As discussed in case study 4, the target impurity is predicted as suspect positive following a precautionary approach, and the read-across approach is suggested. If no structural analogues justifying the read-across study can be identified or if the source chemical(s) possessing the structural alert identified in the target impurity shows positive experimental Ames test results, then the in silico assessment concludes for a positive prediction. Hence, the target impurity must be submitted for experimental assessment of mutagenicity.

References

1. ICH Harmonised Tripartite Guideline (2006) Impurities in new drug substances – Q3A(R2). Current Step 4 version. http://www.ich.org/products/guidelines.html. Accessed 6 May 2015
2. ICH Harmonised Tripartite Guideline (2006) Impurities in new drug products – Q3B(R2). Current Step 4 version. http://www.ich.org/products/guidelines.html. Accessed 6 May 2015
3. ICH Harmonised Tripartite Guideline (2009) Guidance on nonclinical safety studies for the conduct of human clinical trials and marketing authorization for pharmaceuticals – M3(R2). Current Step 4 version. http://www.ich.org/products/guidelines.html. Accessed 6 May 2015
4. ICH Harmonised Tripartite Guideline (2014) Assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals to limit potential carcinogenic risk – M7. Current Step 4 version. http://www.ich.org/products/guidelines.html. Accessed 6 May 2015
5. OECD – Organization for Economic Co-operation and Development (2006) Report on the regulatory uses and applications in oecd member countries of (quantitative) structure-activity relationship [(Q)SAR] models in the assessment of new and existing chemicals. OECD Environment Health and Safety Publications, Series on testing and assessment no. 58. ENV/JM/MONO(2006)25. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2006)25&doclanguage=en. Accessed 11 May 2015
6. OECD – Organization for Economic Co-operation and Development (2007). Guidance document on the validation of (quantitative) structure activity relationship [(Q)SAR] Models. OECD Environment Health and Safety Publications, series on testing and assessment no. 69. ENV/JM/MONO(2007)2. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)2&doclanguage=en. Accessed 11 May 2015
7. Dobo KL, Greene N, Fred C et al (2012) In silico methods combined with expert knowledge rule out mutagenic potential of pharmaceutical impurities: an industry survey. Regul Toxicol Pharmacol 62(3):449–455
8. Sutter A, Amberg A, Boyer S et al (2013) Use of in silico systems and expert knowledge for structure-based assessment of potentially mutagenic impurities. Regul Toxicol Pharmacol 67(1):39–52
9. Powley MW (2014) (Q)SAR assessments of potentially mutagenic impurities: a regulatory perspective on the utility of expert knowledge and data submission. Regul Toxicol Pharmacol 71(2):295–300
10. Serafimova R, Gatnik MF, Worth A (2010) Review of QSAR models and software tools for predicting genotoxicity and carcinogenicity. JRC Scientific and Technical Reports EUR 24427 EN – 2010. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_24427_EN.pdf. Accessed 6 May 2015
11. Worth AP, Lapenna S, Serafimova R (2013) QSAR and metabolic assessment tools in the assessment of genotoxicity. Methods Mol Biol 930:125–162
12. Roncaglioni A, Toropov AA, Toropova AP, Benfenati E (2013) In silico methods to predict drug toxicity. Curr Opin Pharmacol 13(5):802–806
13. Fioravanzo F, Bassan A, Pavan M et al (2012) Role of in silico genotoxicity tools in the regulatory

assessment of pharmaceutical impurities. SAR QSAR Environ Res 23(3-4):257–277. doi:10.1080/1062936X.2012.657236

14. Cassano A, Raitano G, Mombelli E et al (2014) Evaluation of QSAR models for the prediction of ames genotoxicity: a retrospective exercise on the chemical substances registered under the EU REACH regulation. J Environ Sci Health C Environ Carcinog Ecotoxicol Rev 32(3):273–298
15. ACD/Percepta, release 2014. Advanced Chemistry Development, Inc., Toronto, ON, Canada. www.acdlabs.com
16. Lanevskij K, Juska L, Dapkunas J et al (2012) In silico test battery for rapid evaluation of genotoxic and carcinogenic potential of chemicals. Poster presented at 243rd ACS national meeting, San Diego, California, 25–29 Mar 2012
17. Didziapetris R, Dapkunas J, Sazonovas A, Japertas P (2010) Trainable structure-activity relationship model for virtual screening of CYP3A4 inhibition. J Comput Aided Mol Des 24(11):891–906
18. Leis M, Lidbury BA, Yang C et al (2012) Novel technologies and an overall strategy to allow hazard assessment and risk prediction of chemicals, cosmetics, and drugs with animal-free methods. Altex 29(4):373–388
19. Cherkasov A, Muratov EN, Fourches D et al (2014) QSAR modeling: where have you been? Where are you going to? J Med Chem 57(12):4977–5010
20. CORINA Symphony, Molecular Networks GmbH, Erlangen, Germany. www.molecular-networks.com. Accessed 6 May 2015
21. Leadscope Model Applier, version 2.0 (2015) Leadscope Inc., Columbus, OH. http://www.leadscope.com
22. Dearfield KL, Benz RD (1999) Can the new genetic toxicology tests be used for regulatory safety decisions? Environ Mol Mutagen 33(2):91–93
23. Matthews EJ, Kruhlak NL, Cimino MC et al (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: I. Identification of carcinogens using surrogate endpoints. Regul Toxicol Pharmacol 44(2):83–96
24. Matthews EJ, Kruhlak NL, Cimino MC et al (2006) An analysis of genetic toxicity, reproductive and developmental toxicity, and carcinogenicity data: II. Identification of genotoxicants, reprotoxicants, and carcinogens using in silico methods. Regul Toxicol Pharmacol 44:97–110
25. Ferrari T, Gini G (2010) An open source multistep model to predict mutagenicity from statistical analysis and relevant structural alerts. Chem Cent J 4(Suppl 1):S2
26. VegaNIC Application (Virtual Models for evaluating the properties of chemicals within a global architecture), ver 1.1.0. Laboratory of Environmental Chemistry and Toxicology of Mario Negri Institute of Pharmacological Research. http://www.vega-qsar.eu/download.html. Accessed 6 May 2015
27. Ellison CM, Sherhod R, Cronin MT et al (2011) Assessment of methods to define the applicability domain of structural alert models. Chem Inf Model 51(5):975–985. doi:10.1021/ci1000967
28. Hillebrecht A, Muster W, Brigo A et al (2011) Comparative evaluation of in silico systems for ames test mutagenicity prediction: scope and limitations. Chem Res Toxicol 24(6):843–854
29. Toxtree (Toxic Hazard Estimation by decision tree approach), ver. 2.6.13. Ideaconsult Ltd. http://toxtree.sourceforge.net/index.html. Accessed 6 May 2015
30. Benigni R, Bossa C, Jeliazkova N et al (2008) The Benigni/Bossa rulebase for mutagenicity and carcinogenicity – a module of Toxtree. JRC Scientific and Technical Reports, EUR 23241 EN – 2008. https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/doc/EUR_23241_EN.pdf. Accessed 6 May 2015
31. ECHA – European Chemicals Agency (2008) Guidance on information requirements and chemical safety assessment. Chapter R.6: QSARs and grouping of chemicals. http://echa.europa.eu/documents/10162/13632/information_requirements_r6_en.pdf. Accessed 6 May 2015
32. OECD – Organization for Economic Co-operation and Development (2007) Guidance on grouping of chemicals. OECD Environment Health and Safety Publications, series on testing and assessment no. 80. ENV/JM/MONO(2007)28. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=env/jm/mono(2007)28&doclanguage=en. Accessed 11 May 2015
33. ECHA – European Chemicals Agency (2012) Practical guide 6: how to report read-across and categories. http://echa.europa.eu/documents/10162/13655/pg_report_readacross_en.pdf. Accessed 6 May 2015
34. OECD (Q)SAR Application Toolbox, version 3.3.2 (2015) Organization for economic co-operation and development. http://www.oecd.org/chemicalsafety/risk-assessment/theoecdqsartoolbox.htm. Accessed 6 May 2015

35. OECD – Organization for Economic Co-operation and Development (2009) Guidance document for using the OECD (Q) SAR application toolbox to develop chemical categories according to the OECD guidance on grouping of chemicals. OECD Environment Health and Safety Publications, series on testing and assessment no. 102. ENV/JM/MONO(2009)5. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2009)5. Accessed 11 May 2015
36. ToxRead, v. 0.9 BETA (2015) Istituto di Ricerche Farmacologiche Mario Negri – Laboratory of Environmental Toxicology and Chemistry. http://www.toxgate.eu/. Accessed 6 May 2015
37. Judson P (2009) Combining predictions. In: Royal Society of Chemistry (ed) Knowledge-based expert systems in chemistry: not counting on computers, 1st edn. Royal Society of Chemistry, London
38. Cronin MTD (2010) Prediction of harmful human health effects of chemicals from structure. In: Puzyn T, Leszczynski J, Cronin MTD (eds) Recent advances in QSAR studies: methods and applications. Springer, New York, NY
39. Ellison CM, Madden JC, Judson P, Cronin MTD (2010) Using in silico tools in a weight of evidence approach to aid toxicological assessment. Mol Inf 26(1-2):97–110
40. ECHA – European Chemicals Agency (2010) Practical guide 2: how to report weight of evidence. https://echa.europa.eu/documents/10162/13655/pg_report_weight_of_evidence_en.pdf Accessed 8 June 2015
41. Nendza M, Gabbert S, Kühne R et al (2013) A comparative survey of chemistry-driven in silico methods to identify hazardous substances under REACH. Regul Toxicol Pharmacol 66(3):301–314
42. Roy PP, Kovarich S, Gramatica P (2011) QSAR model reproducibility and applicability: a case study of rate constants of hydroxyl radical reaction models applied to polybrominated diphenyl ethers and (benzo-)triazoles. J Comput Chem 32:2386–2396. doi:10.1002/jcc.21820
43. Naven RT, Greene N, Williams RV (2012) Latest advances in computational genotoxicity prediction. Expert Opin Drug Metab Toxicol 8(12):1579–1587
44. DSSTox (Distributed Structure-searchable Toxicity) database. U.S. Environmental Protection Agency. http://www.epa.gov/ncct/dsstox. Accessed 6 May 2015
45. ECHA CHEM database. European Chemical Agency (ECHA). http://echa.europa.eu/chem_data_en.asp. Accessed 6 May 2015
46. NTP (National Toxicology Program) database. U.S. National Institute of Environmental Health Sciences, National Institutes of Health. http://ntp.niehs.nih.gov. Accessed 6 May 2015
47. GENE-TOX (Genetic Toxicology Databank) database. U.S. National Library of Medicine, National Institutes of Health. http://toxnet.nlm.nih.gov/newtoxnet/genetox.htm Accessed 6 May 2015
48. OECD – Organization for Economic Co-operation and Development (2013), User Manual based on QSAR Toolbox version 3.0. Strategies for grouping chemicals to fill data gaps to assess genetic toxicity and genotoxic carcinogenicity. http://www.oecd.org/env/ehs/risk-assessment/genetic%20toxicity.pdf. Accessed 7 May 2015