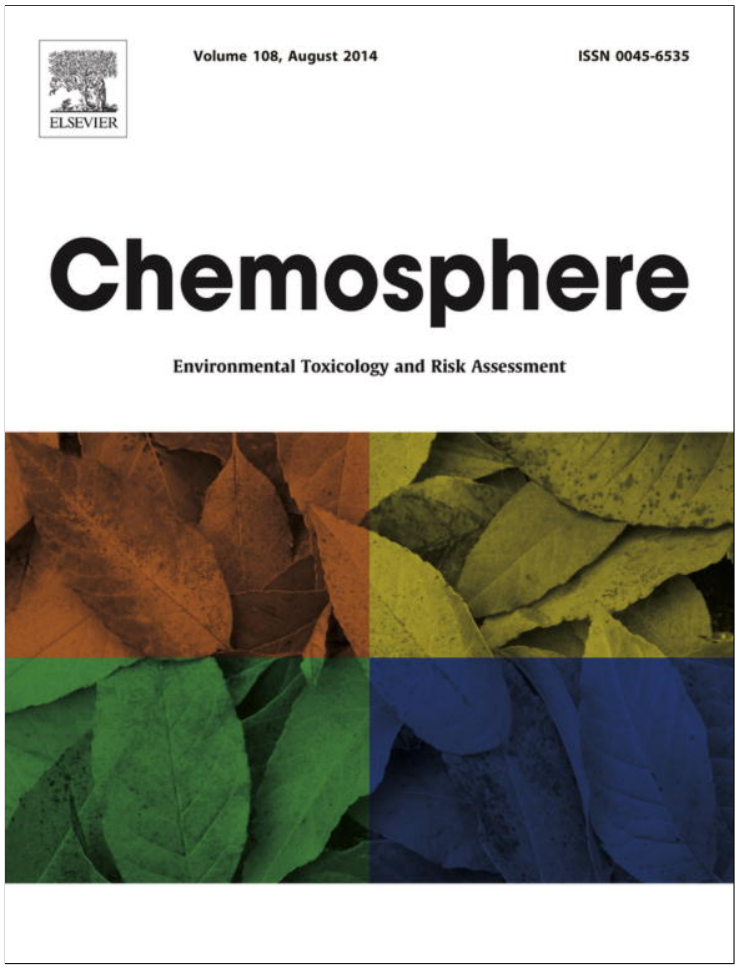


Chemosphere 108 (2014) 10–16

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

A new *in silico* classification model for ready biodegradability, based on molecular fragments

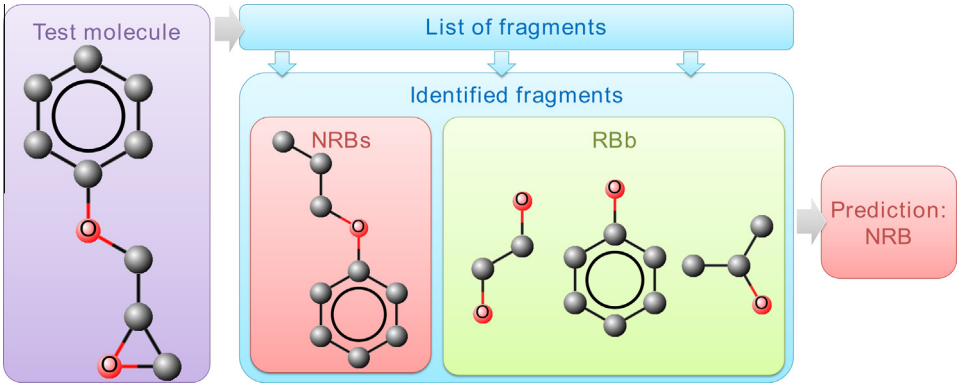
Anna Lombardo [a], Fabiola Pizzo [a], Emilio Benfenati [a,*], Alberto Manganaro [a], Thomas Ferrari [b], Giuseppina Gini [b]

[a] *IRCCS – Istituto di Ricerche Farmacologiche Mario Negri, Via G. La Masa 19, 20156 Milano, Italy*
[b] *Politecnico di Milano, Dipartimento di Elettronica, Informazione e Bioingegneria, Piazza L. da Vinci 32, 20133 Milano, Italy*

HIGHLIGHTS

- A new fragment-based model to predict ready biodegradability was developed.
- A new software to extract fragments was used: SARpy.
- Statistical and expert-based fragments were used to build the new model.
- The model is freely available and useful for regulatory purposes.
- The model has performance comparable to other existing models.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 27 November 2013
Accepted 22 February 2014
Available online 6 April 2014

Handling Editor: S. Jobling

Keywords:
Ready biodegradability
REACH
QSAR
Fragment-based model
SARpy

ABSTRACT

Regulations such as the European REACH (Registration, Evaluation, Authorization and restriction of Chemicals) often require chemicals to be evaluated for ready biodegradability, to assess the potential risk for environmental and human health. Because not all chemicals can be tested, there is an increasing demand for tools for quick and inexpensive biodegradability screening, such as computer-based (*in silico*) theoretical models. We developed an *in silico* model starting from a dataset of 728 chemicals with ready biodegradability data (MITI-test Ministry of International Trade and Industry). We used the novel software SARpy to automatically extract, through a structural fragmentation process, a set of substructures statistically related to ready biodegradability. Then, we analysed these substructures in order to build some general rules. The model consists of a rule-set made up of the combination of the statistically relevant fragments and of the expert-based rules. The model gives good statistical performance with 92%, 82% and 76% accuracy on the training, test and external set respectively. These results are comparable with other *in silico* models like BIOWIN developed by the United States Environmental Protection Agency (EPA); moreover this new model includes an easily understandable explanation.

Abbreviations: AD, Applicability Domain; BOD, Biological Oxygen Demand; ECHA, European Chemicals Agency; FN, False Negative; FP, False Positive; MCC, Matthews Correlation Coefficient; (N)RB, (Non)Readily Biodegradable; OECD TG, Organisation for Economic Co-operation and Development-Test Guideline; PBT, Persistent, Bioaccumulative, Toxic; QSAR, Quantitative Structure–Activity Relationships; REACH, Registration, Evaluation, Authorization and restriction of Chemicals; SAR, Structure–Activity Relationships; SMARTS, SMiles Arbitrary Target Specification; SMILES, Simplified Molecular Input Line Entry System; TN, True Negative; TP, True Positive; vPvB, very Persistent very Bioaccumulative.

* Corresponding author. Tel.: +39 02 39014420; fax: +39 02 39014735.
E-mail address: emilio.benfenati@marionegri.it (E. Benfenati).

http://dx.doi.org/10.1016/j.chemosphere.2014.02.073
0045-6535/

1. Introduction

With their multiple possibilities of prolonged contact with sensitive targets, chemicals that are stable in the environment arouse concern because their potential harmful effects may last longer and become chronic. Generally if a chemical is labile it is easier to investigate its exposure scenarios and the chronic effects may be less important. It is therefore important to assess whether a chemical is persistent in the environment.

REACH legislation (Registration, Evaluation, Authorization and restriction of Chemicals) (REACH, 2006) aims to raise the level of protection for human health and the environment against the risk of exposure to chemicals. Persistence is addressed under REACH and ready biodegradability is a screening test for persistence. All chemicals produced or imported for more than one ton/year must be tested for ready biodegradability (REACH, 2006) (Annex VII of REACH). Persistent and Non-Readily Biodegradable (NRB) are not synonymous: the definitions and thresholds are different. A compound is defined as persistent if it resists degradation and remains in the environment for a long time (ECHA, 2008a). It is considered persistent if its degradation half-life reaches the thresholds of 60 d in marine water, 40 d in fresh or estuarine water, 180 d in marine sediment and 120 d in fresh or estuarine water sediment and in soil, as in the new Annex XIII of REACH (REACH, 2011).

Ready biodegradability is defined as a screening test in which a high concentration of the test substance is used and ultimate biodegradation is measured by non-specific parameters under aerobic conditions. A substance is considered Readily Biodegradable (RB) when it degrades by 60% within 28 d (OECD, 1992). This means that a RB compound is also considered non-persistent but a NRB one is not necessarily considered persistent without further tests.

The reference test for ready biodegradability is the OECD TG 301 (Organisation for Economic Co-operation and Development-Test Guideline; OECD, 1992). Besides the European Community, USA, Canada, and Japan have adopted the OECD TG 301C test for evaluating ready biodegradability (OPPTS, 2008; CEPA, 1999; Yoshioka, 2007).

Within REACH the use of Structure–Activity Relationships (SAR) and Quantitative Structure–Activity Relationships (QSAR) models is encouraged. These examine the compound's properties starting from its chemical structure, exploiting the principle that similar compounds should have similar biological activities (ECHA, 2008b). SAR focuses on the rule determining the relationship, as a classifier, while QSAR quantitatively assesses of the effect (regression model).

We used SARpy (Ferrari et al., 2011) to build up a classifier for ready biodegradation. This new general software automatically extracts knowledge from a dataset and detects the molecular structural fragments associated with the activity of interest. The model we developed, based on ready biodegradability data for the OECD TG 301C – modified MITI – I test, predicts whether a compound is RB or not, to screen its persistence for the PBT (Persistent, Bioaccumulative, Toxic)/vPvB (very Persistent very Bioaccumulative) assessment.

2. Materials and methods

2.1. Data

The dataset described in (Toropov et al., 2012) was used. Two compounds were eliminated, one inorganic and one tautomer. The final dataset of 728 compounds was split into a training set (582 compounds) and a test set (146 compounds), amounting to respectively 80% and 20% of the total maintaining the same proportions of classes as the original set in both subsets.

After the development of the model a new data set was available (Cheng et al., 2012), so their continuous and binary data were extracted and combined in a single dataset. The doubtful compounds (or data), compounds with a percentage of BOD > 100% and duplicates were eliminated. If multiple data were available for the same compounds, the arithmetic mean was maintained if the data were consistent, otherwise the compound was eliminated. From this extended dataset we used the compounds not present in the training or the test set of the model presented here, for a total of 874 new compounds, as the external set.

2.2. Software

SARpy takes in input a set of chemical structures paired with their experimental activity label and produces as output a set of structural fragments associated with the property under investigation. The input and the output structures of SARpy are all expressed as Simplified Molecular Input Line Entry System (SMILES); a SMILES is a string of characters that provides a compact representation of the structure of a molecule (http://www.daylight.com/dayhtml/doc/theory/).

SARpy applies to the input structures (the training set) a fragmentation process to extract all the substructures, within a customizable size range, expressed as the number of atoms (usually 2–18). Then, the software mines for correlations between the incidence of any molecular substructure and the activity of the molecules containing it. Finally, a subset of fragments is selected and proposed to the user in the form of rules "IF fragment THEN activity".

As outcome SARpy lists the SMILES fragments paired with an activity label (e.g., positive, negative), ordered by descending precision in identifying the property under investigation. The statistical measure used for the precision is a likelihood ratio that is computed for each fragment from the ratio of positive (True Positives, TP) to negative predicted as positive (False Positive, FP) elements in the subset of molecules containing the fragment, and the ratio of negative to positive elements in the whole training set.

$$\text{likelihood ratio} = (\text{TP}/\text{FP}) \times (\text{negatives}/\text{positives}) \quad (1)$$

The likelihood can be used as a quantitative attribute of the fragment. Thus, the first fragments in the list identify the molecules with the desired activity label with almost no errors, then come the fragments with a higher misclassification rate. A more detailed description of SARpy is in (Ferrari et al., 2011, 2013); its code is available from the authors.

SARpy can be customized to improve the specificity of the model, or in a more balanced way to improve the accuracy. We obtained different series of fragments (called rule-sets) considering as active the RB compounds (and inactive the non-ready biodegradable ones). Each rule-set was obtained using the settings specified in "Supporting Information A".

3. Results and discussion

3.1. The procedure for obtaining the rules

The fragments for this model derive both from a statistical part and an expert-based part. The modeling has been done in three steps (Fig. 1). Initially, four rule-sets of fragments were generated with SARpy: NRB fragments with high specificity (rule-set 1), NRB fragments with balanced performance (rule-set 2), RB fragments with high specificity (rule-set 3) and RB fragments with

balanced performance (rule-set 4). Fragments with high specificity appear only in one category of compounds (RB or NRB), while fragments with balanced performance are present in both classes, but prevalent in one. The statistical performance of the fragments extracted was checked and some fragments were eliminated (see "The SARpy fragments" section).

In the second step, compounds that could not be classified using the four rule-sets were used to generate other rule-sets. In this way, we obtained another rule-set: RB fragments extracted from chemicals which were labelled unknown when processed together with the others (rule-set 5). For details see "The SARpy fragments" section.

Not all the fragments listed in these five rule-sets can always be assigned to an unambiguous mode of action. Finally, in the third step, we grouped some fragments on the basis of their inherent chemical meaning. In this case we manually added more general rules, partly as a generalization of the fragments extracted using SARpy and partly as formulated by experts. The rules are expressed as SMARTS, which are an extension of SMILES notation including, for instance, wildcards characters. Considering both the generalized and expert-knowledge based fragments, four new fragments were added in rule-sets 6 and 7. For details see "The expert-based fragments" section. These rules made the model more general and consequently more able to correctly classify new compounds. The seven rule-sets are reported in the "Supporting Information B".

On the basis of the chemical and statistical meaning of each of the seven rule-sets, we built a decision tree (Fig. 2) illustrating the workflow of the final model. Fragments related to non-ready biodegradability are initially checked; if some are found in rule-set 1, the compound is predicted as NRB: if only fragments from rule-sets 2 or 6 (but not 1) are found it is predicted as possible NRB. Thus, also a degree of reliability is provided, coming from the statistical quality of the fragments found. If no fragments related to non-biodegradability are found, but there are some related to biodegradability, a similar prediction is provided: RB (if fragments from set 3 are found) or possible RB (if only fragments from sets 4, 5 or 7 are found). If no matching fragments are found at all, the compound is considered non-predictable (not assignable).

The overall model is conservative, and if conflicting fragments are present the prediction is for non-ready biodegradability. The logic of the model comes directly from chemical reasoning: a substance is always considered non-biodegradable if at least one fragment related to non-biodegradability is found, even if there are easily biodegradable fragments too, because this means that part of the compound is anyway persistent.

3.2. The SARpy fragments

As mentioned, four rule-sets were extracted with SARpy (using the settings reported in the "Supporting Information A"). For each fragment we took into account the number of TPs (correctly predicted as active) and False Positives (FP, wrongly predicted as active), considering as matched compounds only those matched first by the fragment under examination (e.g. if a compound was matched for instance by fragments 2 and 16, we considered it matched only by fragment 2). We removed the fragments that

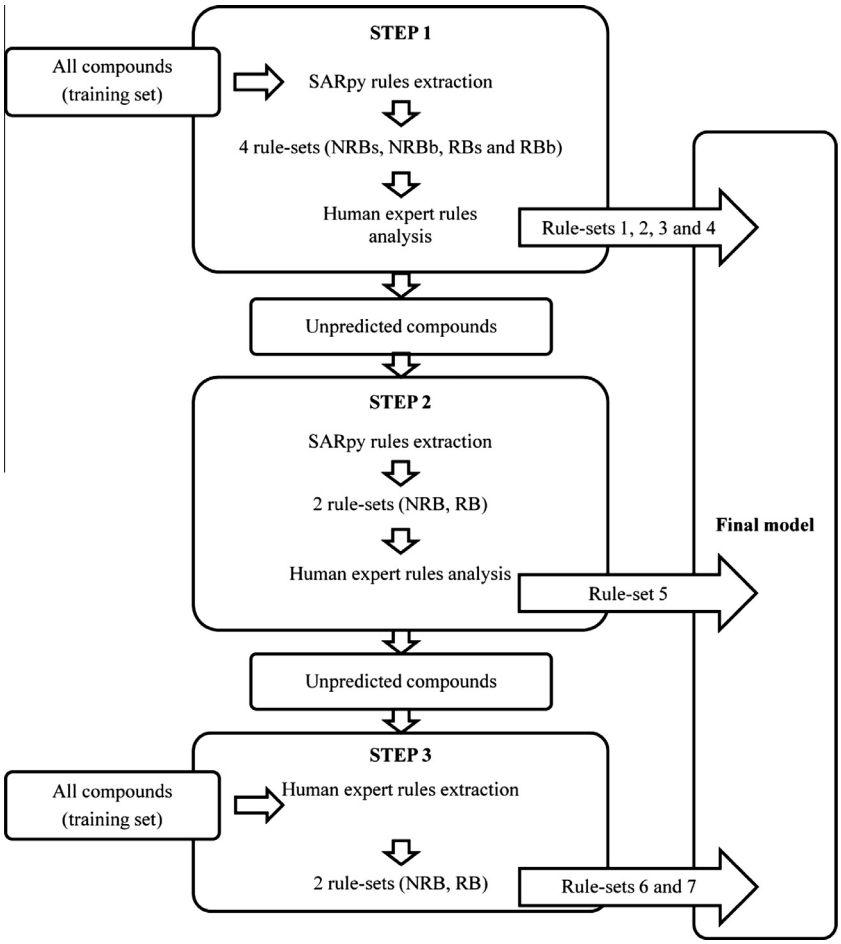


Fig. 1. The procedure for generating the model. (N)RBs and (N)RBb stand for (N)RB specific (s) and balanced (b) fragments respectively.

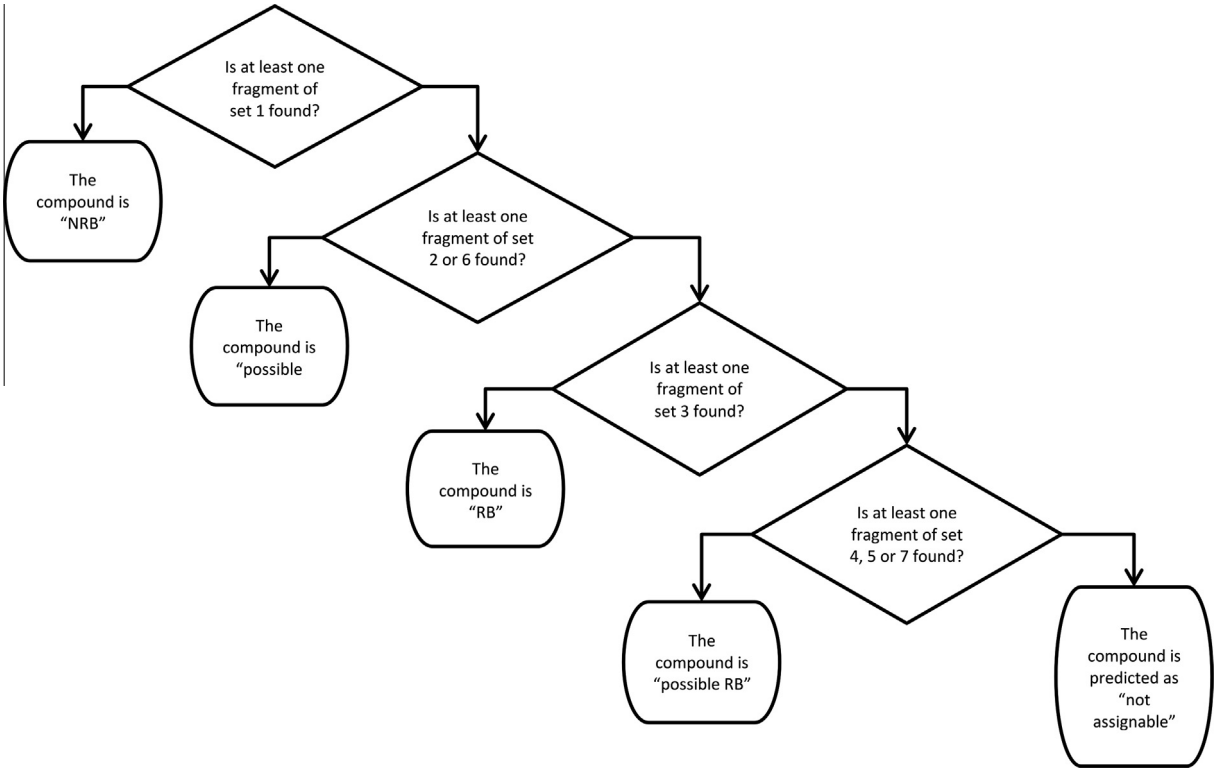


Fig. 2. The model tree. It shows how a compound is classified on the basis of the fragments found.

did not reach an arbitrary threshold of 70% of well predicted compounds out of the total. Then, we removed fragments with likelihood ratio (as calculated by SARpy) less than 2, as the ratio of active/inactive compounds was too low.

SARpy was used again to process the 154 compounds which had no matching fragment. We extracted two new rule-sets (one for NRB and one for RB) (see "Supporting Information A" for setting). The fragments were analysed with the same procedure explained above. We selected fragments that could be used to classify some of the 154 compounds, with a low (below 30%) percentage of errors in the remaining compounds, and added them to the original rule-sets. We added six fragments related to ready biodegradability (rule-set 5), while no fragments related to non-ready biodegradability passed the check.

3.3. The expert-based fragments

As SARpy builds rule-sets on a statistical basis, the last step involved human expertise. The remaining compounds were analysed and some new expert-based fragments were extracted. Each new fragment was verified, studying the literature to check whether there was any general behaviour (i.e. to see why it was statistically related to a positive or a negative activity). This analysis showed that some fragments could be grouped and expressed in a more generic form on the basis of their common chemical meaning. We produced two sets of new rules (rule-set 6 for ready biodegradability and rule-set 7 for non-ready biodegradability) mostly written as SMARTS (SMiles Arbitrary Target Specification) strings (http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html) which describe the chemical structure in a more general way.

The first fragment (rule-set 6) is the double bond nitrogen-nitrogen (azo group). This bond is related to a low biodegradation rate. The azo group is mainly present in azo dyes and there is ample literature about the biodegradation of single dyes. In general the group is reported in the literature as non-readily degradable under natural conditions (Rajaguru et al., 2000). Indeed, the initial step of the biodegradation of azo dyes is cleavage of the azo group. This reaction is catalyzed by the enzyme azoreductase, which is inhibited by molecular oxygen (Sandhya et al., 2005). Since in this study we considered aerobic biodegradation (through the modified MITI-I test), it is reasonable to consider this group as a NRB fragment. The azo group is also included in the list of persistent functional groups of the Canadian Guidance Manual (2003).

The second fragment (rule-set 6) comprised all compounds containing halogen atoms (chlorine, bromine, fluorine and iodine). As several fragments of this type are present in the first two rule-sets (non-biodegradable, specific and balanced), we tried to find out how the halogens influenced biodegradability beyond the particular fragments found. All these halogenated fragments could be summarized in a more general rule, with a clearer chemical sense: chemicals containing a halogen-substituted ring structure. Indeed, the presence of halogenated organic compounds in the environment in significant quantities is the result of human activities over the past 50 years or so. As a result the enzymes that have evolved to metabolize these compounds are considered to be in a relatively early stage of development (Allpress and Gowland, 1998). The initial conversion of non-toxic compounds yields toxic products (e.g. the monooxygenase-catalyzed oxidations of xenobiotics performed by various microorganisms) (Van Hylckama Vlieg et al., 2000). Halogenated compounds are included in the list of persistent compounds of the Canadian Guidance Manual (2003) like: aromatic-I, aromatic-F, aromatic-Cl, aliphatic-Cl, aliphatic-Br, trifluoromethyl group –CF3, two halogen substitutions on unbranched, non-cyclic and one or more halogen substitutions on branched, non-cyclic or cyclic chemicals. Some of the Canadian rules for halogenated compounds overlap ours (both extracted with SARpy and expert-based).

The third fragment (rule-set 7) comprises aromatic aldehydes (defined as a carbonyl group linked to any aromatic ring) which is linked to ready biodegradability. Several species of bacteria oxidize aromatic aldehydes to aromatic acids (Crawford et al., 1982). A broad range of peripheral reactions convert a huge variety of

aromatic compounds into a restricted set of central intermediates, which are subject to ring-cleavage and subsequent funnelling into the Krebs cycle (Pérez-Pantoja et al., 2004). For supporting the decision to include aromatic aldehydes in the list of fragments linked to ready biodegradability in the model presented here, the Canadian Guidance Manual (2003) considers both aromatic aldehydes and acids as easily biodegradable.

The fourth expert-based fragment (rule-set 7) is the nitrile group which is also included in the Canadian Guidance Manual list with easily biodegraded structural features (2003). Nitriles are readily biodegraded by several strains of bacteria (common in sewage sludge, natural water and soil), fungi and plants under aerobic conditions (Ebbs, 2004; Bhalla et al., 2012).

3.4. Results with the model

The model gave good statistical performance. The accuracy (i.e. the correctly predicted compounds), the sensitivity (percentage of correctly predicted NRB compounds), specificity (percentage of correctly predicted RB compounds) and the Matthews Correlation Coefficient (MCC, Matthews, 1975) (Table 1) were calculated considering the possible NRB or RB output values for the RB or NRB classes respectively. We did not consider the unknown compounds for this general evaluation since they are like non-predicted values. Table 1 also lists the numbers of correctly and wrongly predicted compounds. FP are the ones wrongly predicted as RB, FN those wrongly predicted as NRB; TP are those correctly predicted as RB and True Negatives (TN) the ones correctly predicted as NRB. These results are comparable with other classifiers such as BIOWIN 5 and 6 (Tunkel et al., 2000), built to estimate the modified MITI-I test for ready biodegradability (Table 1).

Figs. 3–5 show details of the performance of the model. The distribution of the experimentally RB and NRB compounds in the five predicted classes for each set (Fig. 3) gives few wrongly predicted compounds considering the NRB and possible NRB predicted classes. For the compounds predicted as RB and possible RB, there were several errors in prediction for the external set (about 50%), mainly in the possible RB predictions. A possible explanation,

Table 1
Performance of the SARpy model for ready biodegradability for the training, test, test and external in AD sets. The statistics published (Tunkel et al., 2000) for BIOWIN 5 and 6 (both training and test sets) are reported, so they are not the same training and tests set used in this work.

	Training set	Test set	External set	External set in AD	BIOWIN 5 (training)	BIOWIN 5 (test)	BIOWIN 6 (training)	BIOWIN 6 (test)
No. of compounds	582	146	874	491	–	–	–	–
No. of TN	228	50	385	249	–	–	–	–
No. of TP	221	48	173	147	–	–	–	–
No. of FN	12	7	34	15	–	–	–	–
No. of FP	26	15	142	80	–	–	–	–
No. of unknown	95	26	140	–	–	–	–	–
Accuracy %	92.2	81.7	76.0	80.7	82.2	81.4	82.7	80.7
Sensitivity %	94.8	87.3	73.1	75.6	–	–	–	–
Specificity %	89.8	76.9	83.6	90.7	–	–	–	–
MCC	0.85	0.64	0.51	0.63	–	–	–	–

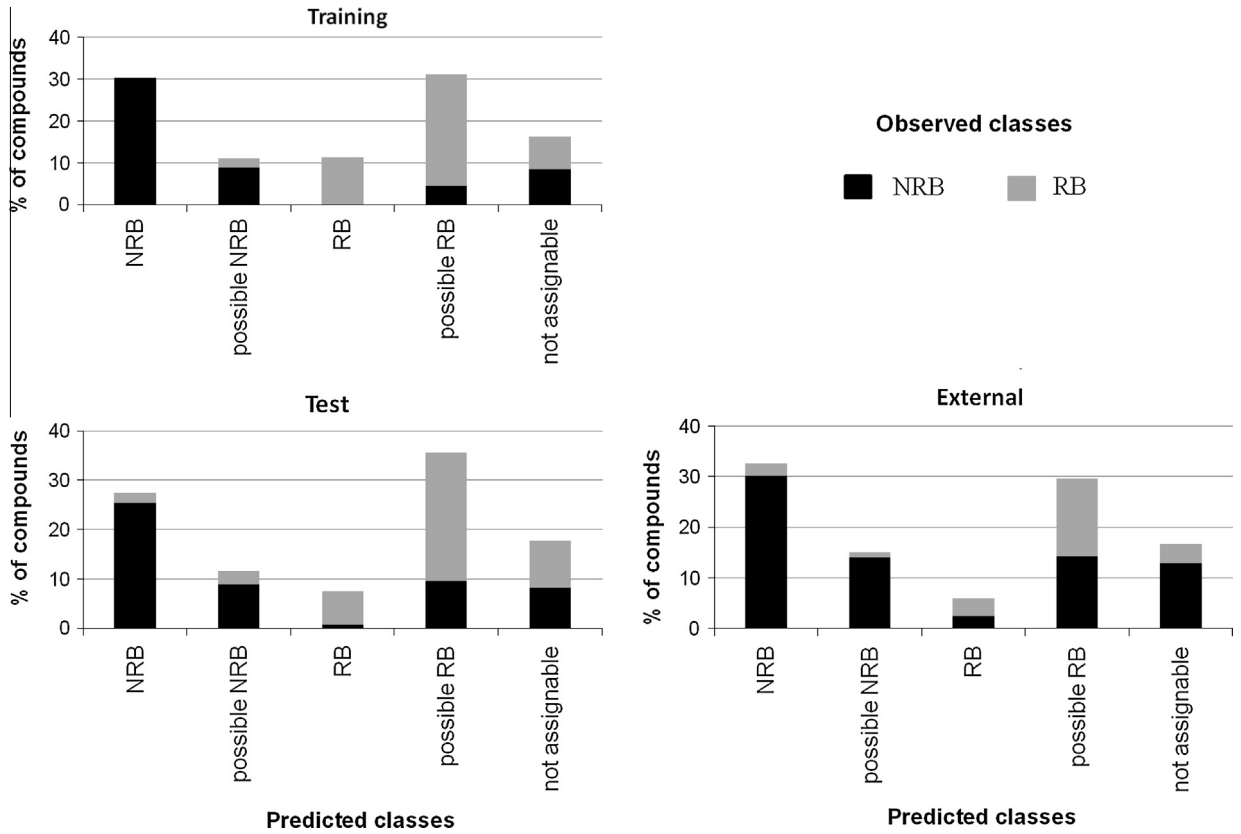


Fig. 3. Percentages of compounds (calculated on the entire set, i.e. including the compounds predicted as "not assignable") experimentally RB and NRB for each class of prediction for the three sets.

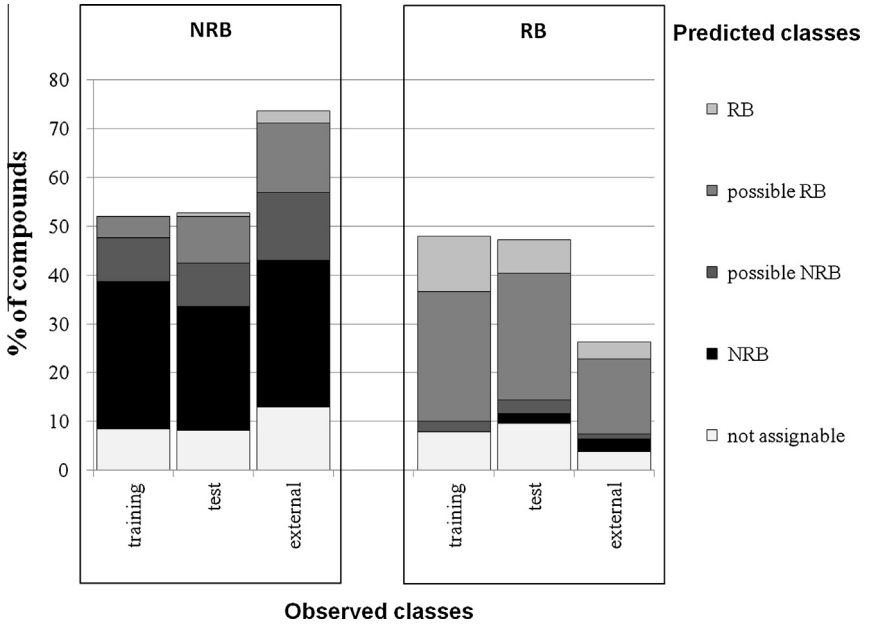


Fig. 4. Percentages of compounds (calculated on the entire set, i.e. including the compounds predicted as "not assignable") for each predicted class that are experimentally RB and NRB for the three sets.

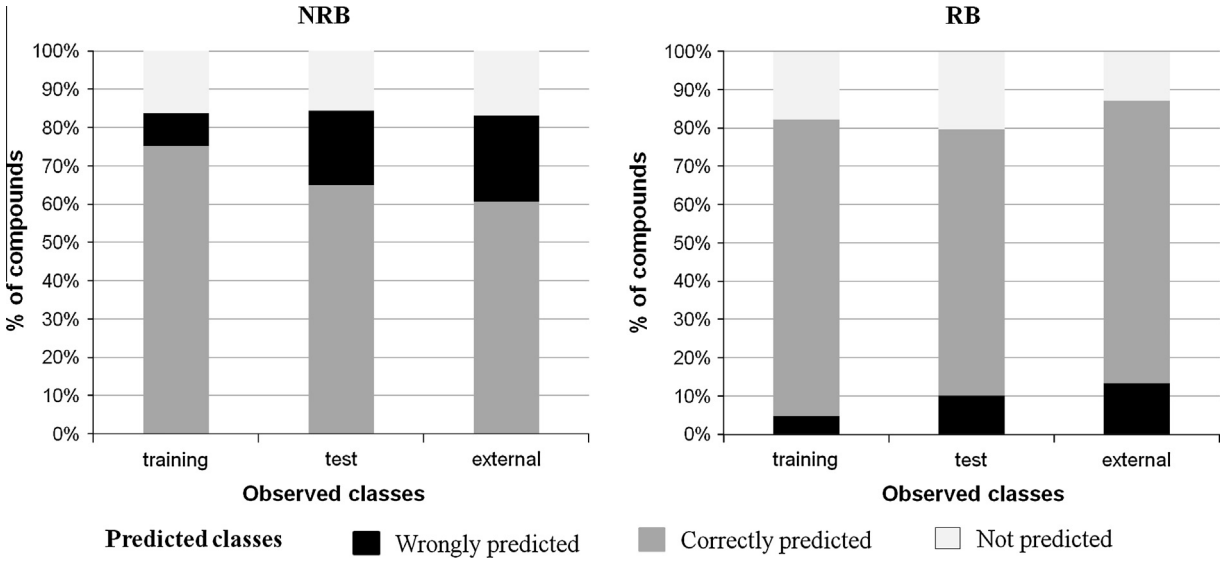


Fig. 5. Percentages of compounds (calculated on the entire set, i.e. including the compounds predicted as "not assignable") correctly predicted, wrongly predicted or predicted as not assignable for experimentally NRB and RB compounds for each set.

taking into consideration the good general performance of the model on the external set, is the unbalanced distribution of the dataset: more than the 70% of the compounds are NRB (both the training and the test set had 52% of NRB compounds).

Fig. 4 shows the distribution of the five predicted classes for the experimentally RB and NRB compounds. Apart from the small loss of performance passing from the training set to test set or external set, which is normal performance was comparable for each set of compounds. For both RB and NRB compounds, the errors were mostly among the compounds predicted as possible (N)RB. Thus, the predictions can be considered more reliable if the label "possible" does not appear, which is reasonable.

Considering separately the sets of RB and NRB compounds, the percentages of correctly predicted, wrongly predicted and not assignable compounds are shown in Fig. 5. There is an increase in the errors for the external set, but they remain comparable to those in the training and test sets. There were 176 wrongly predicted compounds: 20 (11% of the total) NRB compounds were predicted as RB, 122 (69%) NRB as possible RB, 24 (14%) RB as NRB and 10 (6%) RB as possible NRB. So the errors among NRB compounds were mainly due to possible RB fragments.

A detailed output (like the indication of "possible" compounds) gives the user more useful information for analysing and understanding the prediction, thus making it more valuable than the net statistics. The majority of the errors, involved the balanced fragments, but as they gave a prediction of possible ready (or non-ready) biodegradability, they illustrate the uncertainty of the predicted value.

The percentage of NRB compounds correctly predicted ("NRB" and "possible NRB") was high (75.2% for the training set, 64.9%

for the test set and 60.7% for the external set). This was slightly higher for the correctly predicted RB compounds: 79.2% for the training set, 69.6% for the test set and 72.1% for the external set

These percentages were calculated including the compounds predicted as not assignable, to give the real distribution of the classes.

For this model, which is available on the VEGA web site (http://www.vega-qsar.eu/index.php), an Applicability Domain (AD) is available as explained in (Pizzo et al., 2013). The AD evaluates the prediction, assigning a reliability score. If we consider chemicals with a high AD value (greater than 0.65, that indicates a more reliable prediction), the performance of the 491 compounds of the external set that are inside the AD increases (see Table 1).

The model is a combination of computer modeling and human skill. The first is used to make certain chemical features related to the property of interest more transparent. Recursive use of the model improved the extraction of knowledge from the data, establishing a dialogue between the computer and the human experts. The computer program can give the expert valuable facilitating data analysis. The information extracted from the data was carefully evaluated by the human expert, who smoothed certain rules linked inevitably to the specific data. Generalization was done using expert knowledge based on the chemistry and the environmental sciences. SARpy proved to be efficient and helpful. The model supports the user in assessing the prediction, because fragments with lower accuracy classify the compounds as possible (N) RB. It can therefore be used for regulatory purposes, particularly in Europe (i.e. for REACH), USA, Japan and Canada where a ready biodegradability assessment is required for registration of a compound. It also fulfills the OECD principles: it is scientifically valid since it was tested on new data sets (the test and external sets), and it has a well-defined AD when used through the VEGA platform, because this platform implements and combines several ways to define the AD.

Acknowledgment

We are grateful to the project Prioritization of chemicals: a methodology embracing PBT parameters into a unified strategy (PROMETHEUS) research ID: 3713 63 414, business ID: Z 6 - 80 710/5 by Umweltbundesamt (UBA) for financial support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chemosphere.2014.02.073.

References

Allpress, J.D., Gowland, P.C., 1998. Dehalogenases: environmental defence mechanisms and model of enzyme evolution. Biochem. Educ. 26, 267–276.

Bhalla, T.C., Sharma, N., Bhatia, R.K. (Eds.), 2012. Microbial degradation of cyanides and nitrilesSatyanarayana, T., Narain, J.B., Prakash, A. (Eds.). Microorganisms in environmental management: microbes and environment. Springer Science+Business Media B.V., pp. 569–587.

Canadian Guidance Manual, 2003. Guidance Manual for the Categorization of Organic and Inorganic Substances on Canada's Domestic Substances List. Determining Persistence, Bioaccumulation Potential, and Inherent Toxicity to Non-human Organisms. Existing Substances Branch Environment Canada.

CEPA New Substances Notification Regulations (Chemicals and Polymers) of the Canadian Environmental Protection Act (CEPA), 1999.

Cheng, F., Ikenaga, Y., Zhou, Y., Yu, Y., Li, W., Shen, J., Du, Z., Chen, L., Xu, C., Liu, G., Lee, P.W., Tang, Y., 2012. In silico assessment of chemical biodegradability. J. Chem. Inf. Model. 52, 655–669.

Crawford, D.L., Sutherland, J.B., Pometto III, A.L., Miller, J.M., 1982. Production of an aromatic aldehyde oxidase by Streptomyces viridosporus. Arch. Microbiol. 131, 351–355.

Ebbs, S., 2004. Biological degradation of cyanide compounds. Curr. Opin. Biotechnol. 15, 231–236.

ECHA European Chemicals Agency (ECHA), 2008a. Guidance on information requirements and chemical safety assessment, Chapter R.7B: Endpoint specific guidance, May.

ECHA European Chemicals Agency (ECHA), 2008b. Guidance on information requirements and chemical safety assessment, Chapter R.6: QSARs and grouping of chemicals, May.

Ferrari, T., Gini, G., Golbamaki Bakhtyari, N., Benfenati, E., 2011. Mining structural alerts from SMILES: a new way to derive structure-activity relationships. In: Proc. IEEE SSCI 2011: Symposium Series on Computational Intelligence – CIDM 2011, pp. 120–127.

Ferrari, T., Cattaneo, D., Gini, G., Golbamaki Bakhtyari, 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.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta 405, 442–451.

OECD Organization for Economic Cooperation and Development (OECD), 1992. OECD Guideline for the Testing of Chemicals No. 301, Paris, France, July.

OPPTS United States Environmental Protection Agency: Fate, 2008. Transport and Transformation Test Guidelines OPPTS 835.0001 Principles and Strategies Related to Biodegradation Testing of Organic Chemicals under the Toxic Substances Control Act (TSCA), October.

Pérez-Pantoja, D., Donoso, R., Agulló, L., Córdova, M., Seeger, M., Pieper, D.H., González, B., 2004. Genomic analysis of the potential for aromatic compounds biodegradation in Burkholderiales. Environ. Microbiol. 14, 1091–1117 (OPPTS, 2008).

Pizzo, F., Lombardo, A., Manganaro, A., Benfenati, E., 2013. In silico models for predicting ready biodegradability under REACH: a comparative study. Sci. Tot. Env. 463–464, 161–168.

Rajaguru, P., Kalaiselvi, K., Palanivel, M., Subburam, V., 2000. Biodegradation of azo dyes in a sequential anaerobic–aerobic system. Appl. Microbiol. Biotechnol. 54, 268–273.

REACH: Registration, 2006. Evaluation, Authorisation and restriction of Chemicals (REACH) Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December.

REACH: Registration, 2011. Evaluation, Authorisation and restriction of Chemicals (REACH) Regulation (EU) No 253/2011 of the European Parliament and of the Council of 15 March.

Sandhya, S., Padmavathy, S., Swaminathan, K., Subrahmanyam, Y.V., Kaul, S.N., 2005. Microaerophilic-aerobic sequential batch reactor for treatment of azo dyes containing simulated wastewater. Process Biochem. 40, 885–890.

Toropov, A.A., Toropova, A.P., Lombardo, A., Roncaglioni, A., De Brita, N., Stella, G., Benfenati, E., 2012. CORAL: the prediction of persistence biodegradation of organic compounds with optimal SMILES-based descriptors. Cent. Eur. J. Chem. 10, 1042–1048.

Tunkel, J., Howard, P.H., Boethling, R.S., Stiteler, W., Loonen, H., 2000. Predicting ready biodegradability in the Japanese Ministry of International Trade and Industry set. Environ. Toxicol. Chem. 19 (10), 2478–2485.

Van Hylckama Vlieg, J.E.T., Poelarends, G.J., Mars, A.E., Janssen, D.B., 2000. Detoxification of reactive intermediates during microbial metabolism of halogenated compounds. Curr. Opin. Microbiol. 3, 257–262.

Yoshioka, Y., 2007. Regulations and Guidelines on Human Health Products in Japan. Drug Inf. J. 41, 163–167.