European Journal of Medicinal Chemistry 45 (2010) 4399–4402

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Short communication

A new bioconcentration factor model based on SMILES and indices of presence of atoms

A.P. Toropova [a], A.A. Toropov [a,*], A. Lombardo [a], A. Roncaglioni [a], E. Benfenati [a], G. Gini [b]

[a] Istituto di Ricerche Farmacologiche Mario Negri, Via La Masa 19, 20156 Milano, Italy
[b] Department of Electronics and Information, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy

ARTICLE INFO

Article history:
Received 19 April 2010
Received in revised form
31 May 2010
Accepted 10 June 2010
Available online 17 June 2010

Keywords:
QSAR
SMILES
Bioconcentration factor
Balance of correlations
Index of presence of atoms

ABSTRACT

Indices of the presence of atoms (IPA) encode the presence or absence of atoms, such as nitrogen, oxygen, sulphur, phosphorus, fluorine, chlorine, and bromine in a molecule. They are calculated with the simplified molecular input line entry system (SMILES). Using the Monte Carlo method for correlation weights of these indices, one can improve the predictive ability of optimal SMILES-based descriptors in quantitative structure–activity relationships (QSAR) for bioconcentration factor. The model without IPA gave the following results: $n = 503$, $r^2 = 0.6803$, $q^2 = 0.6781$, $s = 0.759$, $F = 1066$ (subtraining set); $n = 322$, $r^2 = 0.8181$, $r^2_{pred} = 0.8159$, $s = 0.565$, $F = 1439$ (calibration set); $n = 105$, $r^2 = 0.6703$, $r^2_{pred} = 0.6577$, $R^2_m = 0.6628$, $s = 0.728$, $F = 209$ (test set); $n = 106$, $r^2 = 0.6624$, $r^2_{pred} = 0.6502$, $R^2_m = 0.6212$, $s = 0.757$, $F = 204$ (validation set) The model with IPA gave: $n = 503$, $r^2 = 0.7082$, $q^2 = 0.7062$, $s = 0.725$, $F = 1216$ (subtraining set); $n = 322$, $r^2 = 0.8401$, $r^2_{pred} = 0.8383$, $s = 0.528$, $F = 1682$ (calibration set); $n = 105$, $r^2 = 0.7489$, $r^2_{pred} = 0.7402$, $R^2_m = 0.7252$, $s = 0.637$, $F = 307$ (test set); $n = 106$, $r^2 = 0.7306$, $r^2_{pred} = 0.7217$, $R^2_m = 0.7010$, $s = 0.680$, $F = 282$ (validation set).

1. Introduction

Quantitative structure–property/activity relationships (QSPR/QSAR) models are often classified as theory. However, they make it possible to formulate a new type of experiment. Instead of direct work with substances, one can employ computational treatment of available experimental data to gain fresh knowledge [1–7].

Data on the bioconcentration factor (BCF) is important from an ecological point of view [8–12]. It is also important for regulatory purposes. The European regulation on chemical substances REACH [13] requires BCF for all compounds. Experimental definition of the BCF is complex technical task. This leads to an increase of QSAR studies dedicated to models of BCF of organic compounds. BCF can be represented as a mathematical function of topological indices ($n = 16$, $r^2 = 0.5673$, biphenyls) [8]. A hybrid model (i.e. a consensus of a group of model) quite well predicted BCF (training set: $n = 378$, $r^2 = 0.83$, test set $n = 95$, $r^2 = 0.80$) [9]. The molecular electronegativity-distance vector (MEDV) gave a model for a set of organic pollutants with $n = 236$, $r^2 = 0.8080$, $q^2 = 0.7873$ [10]. Finally, the super structure–activity relationship for the BCF of biphenyl was gave $n = 58$, $r_{cv} = 0.958$ [11].

Unfortunately, it is hard to compare these models, since they are not standardized: they use different sets of compounds, some sets are numerous, others quite small, and the algorithms and validation procedures are different.

The aim of the present QSAR analysis was to assess the balance of correlations [14,15] as a tool for modelling the BCF. The basic idea of the balance of correlations is to split the training set into subtraining and the calibration sets. The calibration set serves as a preliminary estimate of the model for substances which are then used to optimize the model parameters in order to avoid overtraining. The calibration set is a preliminary test set.

In addition to local SMILES attributes [16–18] we used global SMILES attributes [19,20], which are indices of presence of atoms (IPA). IPA use information on the presence or absence of atoms such as nitrogen, oxygen, sulphur, phosphorus, fluorine, chlorine, and bromine as additional components for the QSAR modelling. The IPA represents some global constituents of the molecules, while local attributes [16–18] are representations of molecular fragments.

2. Method

2.1. Data

The numerical data for BCF were selected from the literature [21–25]. Salt substances were used after removing the cation, and

* Corresponding author.
E-mail address: andrey.toropov@marionegri.it (A.A. Toropov).

0223-5234/$ – see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2010.06.019

Table 1
An example of preparation of SMILES attributes: SMILES = "Clc1ccc(cc1)C(c2ccc(Cl)cc2)C(Cl)(Cl)Cl"; CAS = 50-29-3; DCW(3) = 44.7234623NOSP of this structure should be defined as 'NOSP0000' (Table 2), correlation weight for this value is CW('NOSP0000') = 9.3732049; HALO of this structure should be defined as 'HALO010' (Table 3), correlation weight for this value is CW('HALO010') = 2.3172002. Numerical data on the correlation weights were obtained by the Monte Carlo method. Each attribute is represented by string of 12 symbols. Each SMILES element is represented by a zone of four symbols; S_k involves only one zone; SS_k involves two zones; and SSS_k involves three zones.

S_k			$CW(S_k)$	SS_k			$CW(SS_k)$	SSS_k			$CW(SSS_k)$
Zone 1	Zone 2	Zone 3		Zone 1	Zone 2	Zone 3		Zone 1	Zone 2	Zone 3	
Clxxxxxxxxxx			1.3730003								
cxxxxxxxxxxx			0.0628097	cxxxClxxxxxx			3.3772763				
1xxxxxxxxxxx			−0.9370994	cxxx1xxxxxxx			1.5594321	Clxxcxxx1xxx			1.8763542
cxxxxxxxxxxx			0.0628097	cxxx1xxxxxxx			1.5594321	cxxx1xxxcxxx			0.3170920
cxxxxxxxxxxx			0.0628097	cxxxcxxxxxxx			0.2495392	cxxxcxxx1xxx			−0.9995273
cxxxxxxxxxxx			0.0628097	cxxxcxxxxxxx			0.2495392	cxxxcxxxcxxx			0.7490967
(xxxxxxxxxxx			−1.3758277	cxxx(xxxxxxx			−0.3772863	cxxxcxxx(xxx			1.7504048
cxxxxxxxxxxx			0.0628097	cxxx(xxxxxxx			−0.3772863	cxxx(xxxcxxx			1.7521419
cxxxxxxxxxxx			0.0628097	cxxxcxxxxxxx			0.2495392	cxxxcxxx(xxx			1.7504048
1xxxxxxxxxxx			−0.9370994	cxxx1xxxxxxx			1.5594321	cxxxcxxx1xxx			−0.9995273
(xxxxxxxxxxx			−1.3758277	1xxx(xxxxxxx			−1.0578391	cxxx1xxx(xxx			0.1268784
Cxxxxxxxxxxx			−0.6234328	Cxxx(xxxxxxx			−0.3130233	Cxxx(xxx1xxx			1.6889031
(xxxxxxxxxxx			−1.3758277	Cxxx(xxxxxxx			−0.3130233	(xxxCxxx(xxx			1.4383448
cxxxxxxxxxxx			0.0628097	cxxx(xxxxxxx			−0.3772863	cxxx(xxxCxxx			2.3733259
2xxxxxxxxxxx			0.4980521	cxxx2xxxxxxx			1.1901563	2xxxcxxx(xxx			−0.8102675
cxxxxxxxxxxx			0.0628097	cxxx2xxxxxxx			1.1901563	cxxx2xxxcxxx			−1.2545720
cxxxxxxxxxxx			0.0628097	cxxxcxxxxxxx			0.2495392	cxxxcxxx2xxx			1.0012945
cxxxxxxxxxxx			0.0628097	cxxxcxxxxxxx			0.2495392	cxxxcxxxcxxx			0.7490967
(xxxxxxxxxxx			−1.3758277	cxxx(xxxxxxx			−0.3772863	cxxxcxxx(xxx			1.7504048
Clxxxxxxxxxx			1.3730003	Clxx(xxxxxxx			0.8757681	cxxx(xxxClxx			2.0577519
(xxxxxxxxxxx			−1.3758277	Clxx(xxxxxxx			0.8757681	(xxxClxx(xxx			−0.5642011
cxxxxxxxxxxx			0.0628097	cxxx(xxxxxxx			−0.3772863	cxxx(xxxClxx			2.0577519
cxxxxxxxxxxx			0.0628097	cxxxcxxxxxxx			0.2495392	cxxxcxxx(xxx			1.7504048
2xxxxxxxxxxx			0.4980521	cxxx2xxxxxxx			1.1901563	cxxxcxxx2xxx			1.0012945
(xxxxxxxxxxx			−1.3758277	2xxx(xxxxxxx			−2.0033907	cxxx2xxx(xxx			−0.6220122
Cxxxxxxxxxxx			−0.6234328	Cxxx(xxxxxxx			−0.3130233	Cxxx(xxx2xxx			3.1830874
(xxxxxxxxxxx			−1.3758277	Cxxx(xxxxxxx			−0.3130233	(xxxCxxx(xxx			1.4383448
Clxxxxxxxxxx			1.3730003	Clxx(xxxxxxx			0.8757681	Cxxx(xxxClxx			1.8082129
(xxxxxxxxxxx			−1.3758277	Clxx(xxxxxxx			0.8757681	(xxxClxx(xxx			−0.5642011
(xxxxxxxxxxx			−1.3758277	(xxx(xxxxxxx			−1.1226275	Clxx(xxx(xxx			3.1878825
Clxxxxxxxxxx			1.3730003	Clxx(xxxxxxx			0.8757681	Clxx(xxx(xxx			3.1878825
(xxxxxxxxxxx			−1.3758277	Clxx(xxxxxxx			0.8757681	(xxxClxx(xxx			−0.5642011
Clxxxxxxxxxx			1.3730003	Clxx(xxxxxxx			0.8757681	Clxx(xxxClxx			−1.2515411

neutralising them. Duplicates were removed using ChemFinder Ultra v10.0. The endpoint we have examined is the decimal logarithm, log BCF. Different BCF values were available for the same substance. We eliminated values from experiments done at pH outside the limits in guidelines defined by the REACH legislation [13]. The arithmetic mean was then used as experimental value of reference.

For the modelling purposes, we split the chemicals randomly into subtraining ($n = 503$), calibration ($n = 322$), test ($n = 105$), and validation ($n = 106$) sets. We verified that the ranges of log BCF values for these sets were approximately the same ones, according to OECD principles [26].

Table 2
Calculation of the NOSP. This index relates to the presence or absence of four chemical elements: nitrogen, oxygen, sulphur, and phosphorus

N	O	S	P	Comments
0	0	0	0	Nitrogen, oxygen, sulphur, and phosphorus are absent
0	0	0	1	The molecule only contains phosphorus
0	0	1	0	The molecule only contains sulphur
0	0	1	1	The molecule contains sulphur and phosphorus
0	1	0	0	The molecule only contains oxygen
0	1	0	1	The molecule contains oxygen and phosphorus
0	1	1	0	The molecule contains oxygen and sulphur
0	1	1	1	The molecule contains oxygen, sulphur, and phosphorus
1	0	0	0	The molecule only contains nitrogen
1	0	0	1	The molecule contains nitrogen and phosphorus
1	0	1	0	The molecule contains nitrogen and sulphur
1	0	1	1	The molecule contains nitrogen, sulphur, and phosphorus
1	1	0	0	The molecule contains nitrogen and oxygen
1	1	0	1	The molecule contains nitrogen, oxygen and phosphorus
1	1	1	0	The molecule contains nitrogen, oxygen, and sulphur
1	1	1	1	The molecule contains nitrogen, oxygen, sulphur, and phosphorus

2.2. Optimal descriptors

We studied two versions of SMILES-based optimal descriptors [27–29]:

$$DCW(T) = \sum CW(S_k) + \sum CW(SS_k) + \sum CW(SSS_k) \quad (1)$$

$$DCW(T) = CW(NOSP) + CW(HALO) + \sum CW(S_k) + \sum CW(SS_k) + \sum CW(SSS_k) \quad (2)$$

The T is threshold and it is used to divide SMILES attributes into two categories: rare or active (not rare). The attribute is rare if the

Table 3
Calculation of the HALO. This index relates to the presence or absence of three chemical elements: fluorine, chlorine, and bromine.

F	Cl	Br	Comments
0	0	0	Fluorine, chlorine and bromine are absent
0	0	1	The molecule only contains bromine
0	1	0	The molecule only contains chlorine
0	1	1	The molecule contains chlorine and bromine
1	0	0	The molecule only contains fluorine
1	0	1	The molecule contains fluorine and bromine
1	1	0	The molecule contains fluorine and chlorine
1	1	1	The molecule contains fluorine, chlorine, and bromine

number of SMILES of the subtraining set which contain this attribute is less than *T*. Active SMILES attributes were used to calculate of the BCF models, and rare SMILES attributes were blocked, fixing their value at zero, so rare attributes were not involved in the modelling process. We used a threshold value of three after preliminary experiments. This threshold (three) gives the best statistical quality of the model for the external (validation and test) set of substances. Also, we tested thresholds of 2 and 4, and obtained very similar (but slightly worst) results.

S_k, SS_k, and SSS_k are SMILES attributes involving one, two, and three elements respectively [16–18]. The SMILES element is a fragment of the SMILES that cannot be separated into parts, e.g., 'Cl', 'Br', etc. CW(S_k), CW(SS_k), and CW(SSS_k) are the correlation weights for the S_k, SS_k, and SSS_k. Table 1 shows the definition of S_k, SS_k, and SSS_k. Table 2 shows the definition of the NOSP index involving nitrogen, oxygen, sulphur, and phosphorus. Table 3 shows the definition of the HALO index which involves fluorine, chlorine, and bromine. CW(NOSP) and CW(HALO) are correlation weights for the NOSP and HALO, respectively.

The CW(^xS_k) were calculated by the Monte Carlo optimization method. The target function was:

$$IS = R + R' - \mathrm{abs}(R - R') \times dR_{\mathrm{weight}} - \mathrm{abs}\,(C0 + C0' + C1 - C1') \times dC_{\mathrm{weight}}, \quad (3)$$

where R and R' are correlation coefficient between log BCF and optimal descriptor for the subtraining and calibration sets; $C0$ and $C0'$ are intercepts for the subtraining and calibration sets; $C1$ and $C1'$ are slopes for the subtraining set and calibration set. dR_{weight} and dC_{weight} are empirical parameters. Thus the target function calculated with Eq. (3) is a modification of the balance of correlations [14–16] where the slopes (experimental versus predicted) are taken into account: the maximum IS should lead not only to a minimal difference between correlation coefficients for the subtraining and calibration sets, but also to minimal differences between the slopes for these sets.

Thus, each set (subtraining, calibration, test, and validation) has special function. Subtraining set is used for building up a model (values of correlation weights for active attributes). Calibration set is used to avoid the overtraining. The statistical characteristics of the test set are an indicator of predictive potential of a model. Finally, the statistical characteristics of the validation set serve as the second additional checking of the model: one can speak about the ideal situation if statistical characteristics for the test and validation sets are similar.

3. Results

The computational experiment showed the following preferable values of the options: the number of epochs of the training (i.e., the Monte Carlo optimization [27]) is 15; $dR_{\mathrm{weight}} = 0.1$; $dC_{\mathrm{weight}} = 0.01$; coefficients for the optimization procedure are $D_{\mathrm{start}} = 0.5$, $d_{\mathrm{precession}} = 0.1$.

Table 4 shows the statistical quality of the models obtained with the threshold of 3 in the case of the DCW(3) calculated with Eqs. (1) and (2), with the statistical characteristics on three probes of the Monte Carlo optimization. The first probe of the model based on DCW(3) calculated with Eq. (1) is the following:

$$\log \mathrm{BCF} = 0.6920 + 0.0928\mathrm{DCW}(3) \quad (4)$$

$n = 503$, $r^2 = 0.6803$, $q^2 = 0.6781$, $s = 0.759$, $F = 1066$ (subtraining set);
$n = 322$, $r^2 = 0.8181$, $r^2_{\mathrm{pred}} = 0.8159$, $s = 0.565$, $F = 1439$ (calibration set);
$n = 105$, $r^2 = 0.6703$, $r^2_{\mathrm{pred}} = 0.6577$, $R^2_{\mathrm{m}} = 0.6628$, $s = 0.728$, $F = 209$ (test set);
$n = 106$, $r^2 = 0.6624$, $r^2_{\mathrm{pred}} = 0.6502$, $R^2_{\mathrm{m}} = 0.6212$, $s = 0.757$, $F = 204$ (validation set).

The first probe of the model based on DCW(3) calculated with Eq. (2) is the following:

$$\log \mathrm{BCF} = -0.0316 + 0.0898\mathrm{DCW}(3) \quad (5)$$

$n = 503$, $r^2 = 0.7082$, $q^2 = 0.7062$, $s = 0.725$, $F = 1216$ (subtraining set);
$n = 322$, $r^2 = 0.8401$, $r^2_{\mathrm{pred}} = 0.8383$, $s = 0.528$, $F = 1682$ (calibration set);
$n = 105$, $r^2 = 0.7489$, $r^2_{\mathrm{pred}} = 0.7402$, $R^2_{\mathrm{m}} = 0.7252$, $s = 0.637$, $F = 307$ (test set);
$n = 106$, $r^2 = 0.7306$, $r^2_{\mathrm{pred}} = 0.7217$, $R^2_{\mathrm{m}} = 0.7010$, $s = 0.680$, $F = 282$ (validation set).

The R^2_{m} is the measure of predictive potential for a QSPR/QSAR model suggested by Roy P.P. and Roy K. [2]. According to Ref. [2], a model is satisfactory if $R^2_{\mathrm{m}} > 0.5$. Fig. 1 shows the model calculated with Eq. (5) graphically.

4. Discussion

In assessing the quality of a QSAR model, it is important define its predictive ability. Besides internal validation, external validation with a set of compounds never used in building up the model is recommended for regulatory use of QSAR models [26].

A robust principle may be formulated as the following: the test and validation set must be interchangeable. In other words, the model must give similar (identical) statistical quality for the test

Table 4
Statistical characteristics of the model based on the DCW(3) calculated with Eqs. (1) and (2). N_{act} is the number of SMILES attributes which are involved in the modelling process (i.e., which are not rare, if the threshold used is 3).

Probe	N_{act}	Subtraining set, $n = 503$			Calibration set, $n = 322$			Test set, $n = 105$			Validation set, $n = 106$		
		r^2	s	F	r^2	s	F	r^2	s	F	r^2	s	F
Eq. (1), i.e., without NOSP and HALO													
1	513	0.68	0.76	1066	0.82	0.57	1439	0.67	0.73	209	0.66	0.76	204
2		0.68	0.76	1044	0.81	0.57	1395	0.68	0.72	215	0.66	0.76	206
3		0.68	0.76	1056	0.82	0.57	1438	0.66	0.73	201	0.65	0.77	193
Eq. (2), i.e., with NOSP and HALO													
1	530	0.71	0.73	1216	0.84	0.53	1682	0.75	0.64	307	0.73	0.68	282
2		0.70	0.73	1178	0.84	0.53	1689	0.74	0.65	295	0.73	0.67	285
3		0.70	0.73	1199	0.84	0.53	1698	0.71	0.68	258	0.74	0.66	297

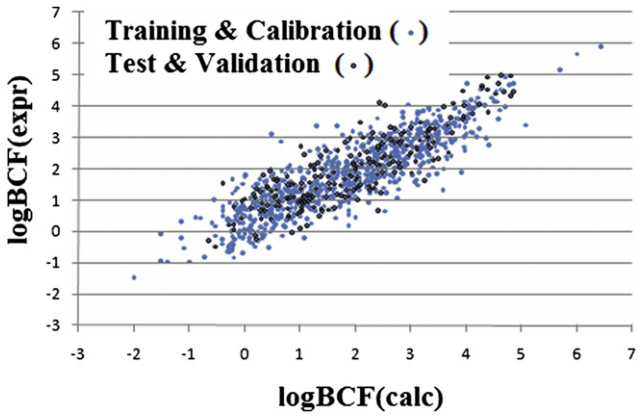

Fig. 1. Graphical representation of log BCF model calculated with Eq. (5).

and validation set. In fact both test and validation set are actually independent and give no information for models calculated with Eqs. (4) and (5).

The model calculated with Eq. (5) can be considered preferable, since its statistical quality is better than the model calculated with Eq. (4) (Table 3).

The statistical characteristics of the log BCF model from Ref. [29] are $n = 105$, $r^2 = 0.805$, $R^2_{pred} = 0.797$, $s = 0.528$, $F = 427$ (test set); Ref [30] are $n = 131$, $r^2 = 0.871$, $s = 0.978$, $F = 213$. The model from Ref. [31] gives $n = 93$, $r^2 = 0.88$ and Ref. [32] gives $n = 511$, $r^2 = 0.84$. Comparison of these models and the one calculated with Eq. (5) indicates that our approach gives similar satisfactory prediction for log BCF. However the present study used considerably larger number of substances than in Refs. [29–32].

The correlation weights for HALO and NOSP indices show they are statistical contributors to the BCF model. In case when there are no halogens (the index is 'HALO00000000'), correlation weights in three probes of the Monte Carlo optimization are 6.8100887, 6.1827120, 5.8742920; this index is present in subtraining set 279 times, in calibration set 184, and in test set 53. The chlorine (HALO01000000) correlation weights in three probes of the optimization are 2.3172002, 1.2516364, 1.3703599, in the subtraining, calibration and test sets are 147, 106, and 41, respectively. The most significant promoter of NOSP indices is the presence of nitrogen with oxygen (NOSP11000000) correlation weights are 1.8782653, 1.5038785, 2.3713518; and the subtraining, calibration and test sets have 122, 85, and 14. For nitrogen, oxygen and sulphur (NOSP11100000), correlation weights are 4.0613332, 3.6904595, 4.5024961, present in subtraining, calibration and test sets 44, 29, 8.

Supplementary materials section contain correlation weights for calculating the DCW(3), experimental log BCF and calculated with Eq. (5), and split into the subtraining, calibration, test, and validation sets.

5. Conclusions

Correlation weights for IPA which are a mathematical function of the presence of nitrogen, oxygen, sulphur, phosphorus (Table 1), fluorine, chlorine, and bromine (Table 2) have improved statistical characteristics of the prediction for the bioconcentration factor (log BCF). Statistical characteristics of models calculated with Eqs. 4 and 5 are reproduced in three probes of the Monte Carlo optimization (Table 3).

Acknowledgement

The authors express gratitude to OSIRIS for financial support.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.ejmech.2010.06.019.

References

[1] J.T. Leonard, K. Roy, Eur. J. Med. Chem. 43 (2008) 81–92.
[2] P.P. Roy, K. Roy, QSAR Comb. Sci. 27 (2008) 302–313.
[3] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, G. Kollias, O. Igglessi-Markopoulou, Mol. Divers. 13 (2009) 301–311.
[4] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, QSAR Comb. Sci. 27 (2008) 432–436.
[5] E. Vicente, P.R. Duchowicz, E.A. Castro, A. Monge, J. Mol. Graphics. Modell. 28 (2009) (2009) 28–36.
[6] P.R. Duchowicz, E.A. Castro, Int. J. Mol. Sci. 10 (2009) 2558–2577.
[7] T. Puzyn, A. Mostrag, J. Falandysz, Y. Kholod, J. Leszczynski, J. Hazard. Mater. 170 (2009) 1014–1022.
[8] P.V. Khadikar, S. Singh, D. Mandloi, S. Joshi, A.V. Bajaj, Bioorg. Med. Chem. 11 (2003) 5045–5050.
[9] C. Zhao, E. Boriani, A. Chana, A. Roncaglioni, E. Benfenati, Chemosphere 73 (2008) 1701–1707.
[10] S. Cui, J. Yang, S. Liu, L. Wang, Sci. China, Ser. B: Chem. 50 (2007) 587–592.
[11] K. Roy, I. Sanyal, P.P. Roy, SAR QSAR Environ. Res. 17 (2006) 563–582.
[12] T. Ivanciuc, O. Ivanciuc, D.J. Klein, Mol. Divers. 10 (2006) 133–145.
[13] http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm.
[14] A.A. Toropov, B.F. Rasulev, J. Leszczynski, Bioorg. Med. Chem. 16 (2008) 5999–6008.
[15] A.A. Toropov, A.P. Toropova, E. Benfenati, Int. J. Mol. Sci. 10 (2009) 3106–3127.
[16] A.A. Toropov, A.P. Toropova, I. Raska, E. Benfenati, Eur. J. Med. Chem. 45 (2010) (2010) 1639–1647.
[17] A.A. Toropov, A.P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, J. Comput. Chem. 31 (2010) 381–392.
[18] A.A. Toropov, A.P. Toropova, E. Benfenati, Cent. Eur. J. Chem. 7 (2009) 846–856.
[19] A.A. Toropov, E. Benfenati, Eur. J. Med. Chem. 42 (2007) 606–613.
[20] A.A. Toropov, E. Benfenati, Comput. Biol. Chem. 31 (2007) 57–60.
[21] D.B. Arnot, J.A. Arnot, F.A.P.C. Gobas, Environ. Rev. 14 (2006) 257–297.
[22] D.B. Dimitrov, S. Dimitrov, N. Dimitrova, T. Parkerton, M. Comber, M. Bonnell, O. Mekenyan, SAR QSAR Environ. Res. 16 (2005) 531–554.
[23] http://www.euras.be/eng/project.asp?ProjectID92.
[24] http://www.eu-footprint.org/index.html.
[25] D.B. Ionizable, W. Fu, A. Franco, S. Trapp, Environ. Toxicol. Chem. 28 (2009) 1372–1379.
[26] http://www.oecd.org/dataoecd/33/37/37849783.pdf.
[27] Istituto di Ricerche Farmacologiche Mario Negri. Available from: http://www.insilico.eu/coral/, 2010.
[28] A.A. Toropov, A.P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, J. Math. Chem. 47 (2010) 647–666.
[29] A.A. Toropov, A.P. Toropova, E. Benfenati, Eur. J. Med. Chem. 44 (2009) 2544–2551.
[30] V.K. Sahu, R.K. Singh, Clean – Soil, Air, Water 37 (2009) 850–857.
[31] S.H. Jackson, C.E. Cowan-Ellsberry, G. Thomas, J. Agric. Food Chem. 57 (2009) 958–967.
[32] S. Dimitrov, N. Dimitrova, T. Parkerton, M. Comber, M. Bonnell, O. Mekenyan, SAR QSAR Environ. Res. 16 (2005) 531–554.