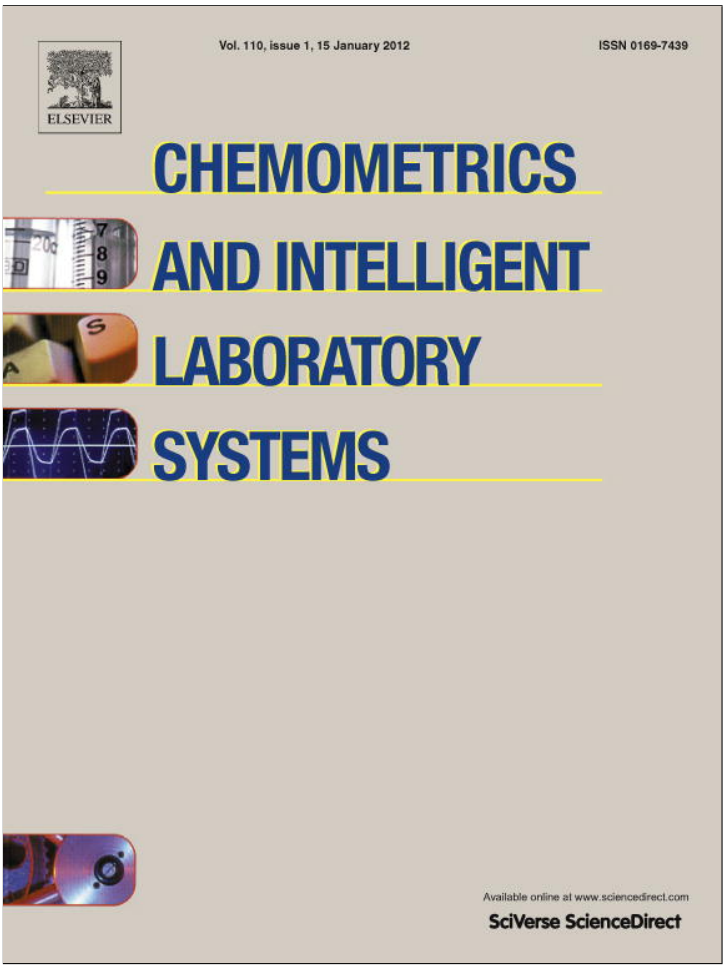

Chemometrics and Intelligent Laboratory Systems 110 (2012) 177–181

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Short Communication

CORAL: QSAR modeling of toxicity of organic chemicals towards *Daphnia magna*

A.P. Toropova [a], A.A. Toropov [a,*], S.E. Martyanov [b], E. Benfenati [a], G. Gini [c], D. Leszczynska [d], J. Leszczynski [e]

[a] *Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano, Italy*
[b] *Teleca OOO, 603093, 23, Rodionova st, Nizhny Novgorod, Russia*
[c] *Department of Electronics and Information, Politecnico di Milano, piazza Leonardo da Vinci 32, 20133 Milano, Italy*
[d] *Interdisciplinary Nanotoxicity Center, Department of Civil and Environmental Engineering, Jackson State University, 1325 Lynch St, Jackson, MS 39217-0510, USA*
[e] *Interdisciplinary Nanotoxicity Center, Department of Chemistry and Biochemistry, Jackson State University, 1400 J. R. Lynch Street, P.O. Box 17910, Jackson, MS 39217, USA*

ARTICLE INFO

Article history:
Received 24 August 2011
Received in revised form 4 October 2011
Accepted 7 October 2011
Available online 15 October 2011

Keywords:
QSAR
SMILES
Molecular graph
CORAL software
Toxicity to *Daphnia magna*

ABSTRACT

Convenient to apply and available on the Internet, CORAL software (http://www.insilico.eu/CORAL) has been used to build up quantitative structure–activity relationships (QSAR) for prediction of toxicity to *Daphnia magna*. The QSARs developed in this study are one-variable models based on the optimal descriptors calculated with the Monte Carlo method. The toxicity has been modeled with the following representations of the molecular structure: (i) by hydrogen-suppressed graph (HSG); (ii) by simplified molecular input line entry system (SMILES); and (iii) by hybrid representation, i.e. the HSG together with SMILES. Four random splits into the sub-training, calibration, and test sets were examined. The hybrid version of the representation of the molecular structure provided the best accuracy of the prediction for the considered endpoint.

1. Introduction

Toxicity of a compound towards *Daphnia magna* represents well-known and important ecological indicator of potential environmental hazards of chemicals [1–8]. Experimental evaluations of environmental properties (e.g. toxicity to *D. magna*) for all new synthetic substances which are being used in everyday life (marketing, cosmetics, medicine, industry, etc.) become impossible owing to increasing number of these substances. Under such circumstances, the development of efficient quantitative structure–activity relationships (QSAR) represents the only compelling alternative to the experiment.

European Union REACH (Registration, Evaluation and Authorisation of Chemicals) explicitly encourages the use of computational methods for estimation of environmental parameters of all new and existing chemicals. Obviously, QSAR will play an important role in addressing of this task [9–14].

Recently, CORAL software has been suggested as an efficient tool for the QSAR analysis [15]. The CORAL models represent one-variable correlations between an endpoint and optimal descriptors. The optimal descriptors are calculated with special coefficients related to presence of various molecular features (molecular fragments and physicochemical characteristics of molecules). These coefficients (correlation weights) are obtained by the Monte Carlo method. One can use as the representation of the molecular structure for the optimal descriptors hydrogen-suppressed molecular graph (HSG) [16], simplified molecular input line entry system (SMILES) [17–19], or a hybrid representation which includes both the HSG and SMILES.

The comparison of aforementioned three representations of the molecular structure in the development of QSAR approaches devoted to toxicity towards *D. magna* is the aim of the present study.

2. Methods

2.1. Data

The descriptions of organic chemicals related to 48 h *D. magna*-toxicity expressed in negative decimal logarithm of the dose that kills 50% of organisms i.e. pLC50 were taken from the literature [1]. The data set covers range of octanol/water partition coefficient from −2 to 8. The range of toxicity (daphnia) is from 0.46 to 10.09. In regard to the chemical domain, the data set includes hydrocarbons, aliphatic alcohols, phenols, ethers, and esters; anilines, amines, nitriles, nitroaromatics, amides, and carbamates; urea and thiourea derivatives; iso-thiocyanates; thiols; phosphorothionate and phosphate esters; and halogenated derivatives. The list of compounds represented by CAS numbers and SMILES with their daphnia toxicity values are shown in Supplementary Materials.

2.2. Molecular descriptors

CORAL software can generate three kinds of optimal descriptors: graph-based, SMILES-based, and hybrid descriptors which are calculated

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

0169-7439/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2011.10.005

with both graph and SMILES. Accordingly, the CORAL software can generate three kinds of molecular graphs: the above-mentioned HSG, hydrogen filled graph (HFG), and graph of atomic orbitals (GAO).

The graph-based optimal descriptors are calculated as the following:

$$^{Graph}\mathrm{DCW}\left(\mathrm{Threshold}, N_{\mathrm{epoch}}\right) = \sum \mathrm{CW}(A_k) + \alpha \sum \mathrm{CW}\left({}^{0}EC_k\right) + \beta \sum \mathrm{CW}\left({}^{1}EC_k\right) + \gamma \sum \mathrm{CW}\left({}^{2}EC_k\right) + \delta \sum \mathrm{CW}\left({}^{3}EC_k\right) \quad (1)$$

where A_k is chemical element, such as, C, N, O, etc., for HSG and HFG; or atomic orbitals, such as $1s^1$, $2p^3$, $3\,d^{10}$, etc., for GAO; ${}^{0}EC_k$, ${}^{1}EC_k$, … ${}^{3}EC_k$ represents the hierarchy of the Morgan extended connectivity; α, β, γ, and δ can be 1 or 0: combinations of their values gives possibility to define various versions of the graph-based optimal descriptor; CW(x) is the correlation weight of a molecular feature (encoded by A_k or ${}^{x}EC_k$).

The SMILES-based optimal descriptors are calculated as the following:

$$^{SMILES}\mathrm{DCW}(\mathrm{Threshold}, N_{\mathrm{epoch}}) = \alpha \sum \mathrm{CW}(S_k) + \beta \sum \mathrm{CW}(SS_k) + \gamma \sum \mathrm{CW}(SSS_k) + x \cdot \mathrm{CW}(NOSP) + y \cdot \mathrm{CW}(\mathrm{HALO}) + z \cdot \mathrm{CW}(\mathrm{BOND}) \quad (2)$$

where S_k, SS_k, and SSS_k are one-, two-, and three-component SMILES attributes, respectively; the component of SMILES represents one symbol (e.g. C, c, N, n, =, #, etc.) or two symbols which cannot be separated (e.g. Cl, Br, @@, etc.); NOSP, HALO, and BOND are indices calculated according to presence or absence of chemical elements: nitrogen, oxygen, sulfur, and phosphorus (NOSP); fluorine, chlorine, and bromine (HALO). The BOND symbolizes a mathematical function related to the presence or absence of double (=), triple (#), or stereo chemical bonds (@ or @@); α, β, γ, x, y, and z can be 1 or 0: combinations of their values provide possibility to define various versions of the SMILES-based optimal descriptor. CW(x) is the correlation weight of a molecular feature (encoded by $S_kS_kS_k$ or ${}^{x}EC_k$).

The hybrid optimal descriptors are calculated with taking into account both representations of the molecular structure by graph and by SMILES.

$$^{Hybrid}\mathrm{DCW}(\mathrm{Threshold}, N_{\mathrm{epoch}}) = {}^{SMILES}\mathrm{DCW}(\mathrm{Threshold}, N_{\mathrm{epoch}}) + {}^{Graph}\mathrm{DCW}(\mathrm{Threshold}, N_{\mathrm{epoch}}) \quad (3)$$

Threshold and N_{epoch} (in Eqs. (1)–(3)) are parameters of the Monte Carlo optimization. Threshold is criterion for classification of components of the representation of the molecular structure into two classes: rare (noise) and active (not rare). The correlation weight of a rare component is fixed as zero; hence rare component is not involved in the building up of the model. N_{epoch} is the number of epochs of the Monte Carlo optimization. Fig. 1 shows the theoretical influence of the threshold and of the number of epochs of the Monte Carlo method optimization for the correlation coefficient between the experimental and calculated values of an endpoint.

One can see (Fig. 1) that the increase of the threshold is accompanied by decrease of the correlation coefficient between experimental and calculated values of an endpoint for the sub-training and calibration set, whereas the correlation coefficient for the external test set has a maximum (Threshold = 2). The increase of the number of epochs of the Monte Carlo method optimization is accompanied by

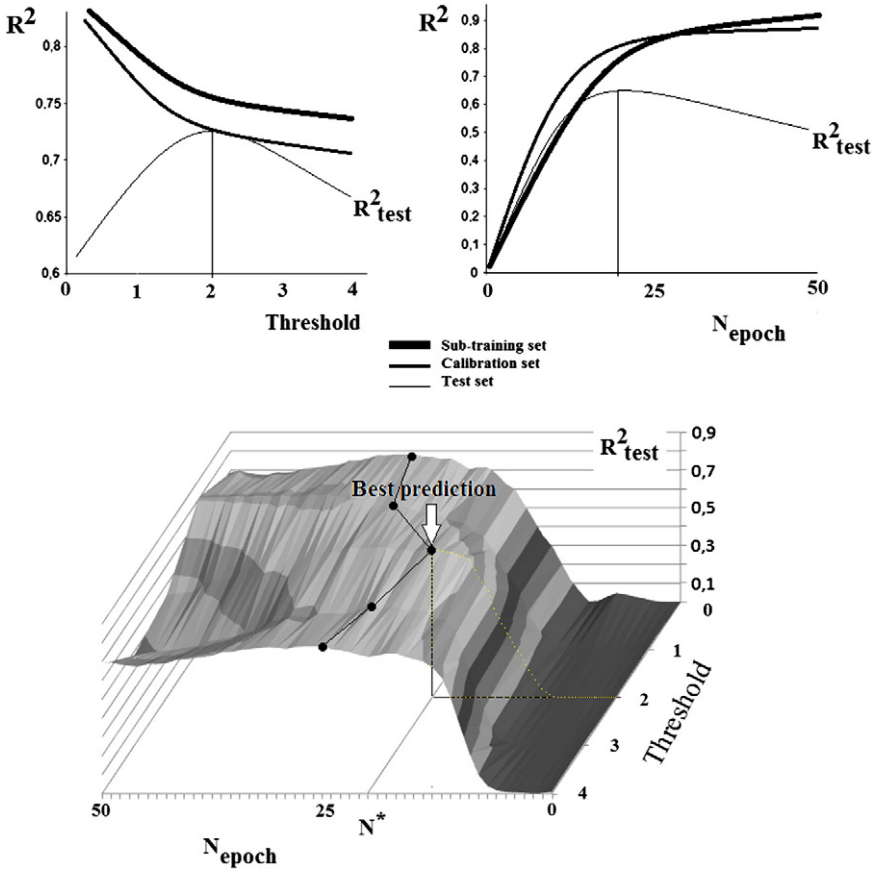

Fig. 1. Correlation coefficient between experimental and predicted values of an endpoint for the external test set as a mathematical function of the threshold and the number of epochs of the Monte Carlo method optimization.

increase of the correlation coefficient for the sub-training and calibration sets, but again the correlation coefficient for the test set has a maximum ($N_{epoch} = N^* = 21$). Thus, the preferable model, i.e. preferable Threshold and N_{epoch} (denoted N^*) can be obtained from analysis of the surface

$$R^2_{test} = F\left(\text{Threshold}, N_{epoch}\right) \quad (4)$$

The principles of CORAL software are the following:

1. Molecular graph and/or SMILES are representations of the molecular structure.
2. These representations (by graph or SMILES) are not identical.
3. Significance of various components of these representations depends on endpoint, and on split of substances into the training and test sets. For instance, SMILES attributes which are important for QSAR model of carcinogenic potential can be uninformative for QSAR model of toxicity towards *D. magna*.
4. Correlation weights of the molecular features which produce satisfactory correlation between optimal descriptor (Eq. (1), or (2), or (3)) and an endpoint can also give the satisfactory correlation for external test set.
5. Predictability of a selected version of the optimal descriptor should be checked with several random splits into the training and test sets.

3. Results and discussion

Table 1 contains statistical quality of models for toxicity towards *D. magna*. The versions of descriptors ^{Graph}DCW(Threshold,N_{epoch}) calculated here with Eq. (1) (HSG) are the following: $\alpha = 0$, $\beta = 1$, $\gamma = 1$, and $\delta = 0$. The SMILES-based descriptors ^{SMILES}DCW(Threshold,N_{epoch}) applied in our work with Eq. (2) are as follows: $\alpha = 1$, $\beta = 1$, $\gamma = 0$, $x = 0$, $y = 0$, and $z = 1$. Upon the analysis of all obtained data one can see from Table 1 that hybrid version of the descriptors gives improvement of the accuracy of the model (Table 1).

Table 2 contains descriptions of the statistical quality of models for toxicity towards *D. magna* which are built up with the same hybrid descriptor for three random splits. The balance of correlations with ideal slopes has been used in the Monte Carlo method optimization.

The aforementioned models are the following (n is the number of compounds in a set; r is correlation coefficient; q^2 and r^2_{pred} are leave-one-out correlation coefficients; s is standard error of estimation; F is Fischer F-ratio; novel validation metrics $\overline{R^2_m}$ [20]: $\overline{R^2_m}$ is mean of R^2_m and R'^2_m values and ΔR^2_m is the absolute difference

Table 1
Comparison of predictability of optimal descriptors calculated with HSG, SMILES, and optimal descriptors which are calculated with both HSG and SMILES: the case of split taken from the literature [1]. Best models are indicated by bold.

Threshold	Probe 1	Probe 2	Probe 3	Average	Dispersion
HSG					
R^2_{test} 1	0.6481	0.6455	0.6523	0.6486	0.0028
2	0.6947	0.6928	0.6854	**0.6910**	0.0040
3	0.6727	0.6746	0.6735	0.6736	0.0008
4	0.6584	0.6521	0.6621	0.6575	0.0041
5	0.6545	0.6535	0.6522	0.6534	0.0009
6	0.6461	0.6436	0.6424	0.6440	0.0015
7	0.6202	0.6182	0.6204	0.6196	0.0010
N^* 1	10	10	10	10.00	0.00
2	13	14	14	**13.67**	0.47
3	16	15	16	15.67	0.47
4	16	15	17	16.00	0.82
5	17	16	18	17.00	0.82
6	17	18	17	17.33	0.47
7	18	18	16	17.33	0.94
SMILES					
R^2_{test} 1	0.6794	0.6814	0.6795	0.6801	0.0009
2	0.6996	0.6979	0.6950	**0.6975**	0.0019
3	0.6674	0.6633	0.6655	0.6654	0.0017
4	0.6617	0.6567	0.6579	0.6588	0.0021
5	0.6538	0.6503	0.6512	0.6518	0.0015
6	0.6471	0.6470	0.6474	0.6472	0.0001
7	0.6353	0.6434	0.6380	0.6389	0.0034
N^* 1	14	16	14	14.67	0.94
2	16	16	16	**16.00**	0.00
3	15	16	15	15.33	0.47
4	16	16	18	16.67	0.94
5	15	16	17	16.00	0.82
6	16	16	15	15.67	0.47
7	17	16	17	16.67	0.47
HSG and SMILES					
R^2_{test} 1	0.7439	0.7412	0.7362	0.7404	0.0032
2	0.7716	0.7768	0.7732	**0.7739**	0.0022
3	0.7517	0.7530	0.7539	0.7529	0.0009
4	0.7328	0.7353	0.7354	0.7345	0.0012
5	0.7224	0.7193	0.7204	0.7207	0.0013
6	0.7125	0.7113	0.7142	0.7127	0.0012
7	0.7057	0.6982	0.7050	0.7030	0.0033
N^* 1	14	14	14	14.00	0.00
2	17	18	16	**17.00**	0.82
3	16	15	15	15.33	0.47
4	16	17	18	17.00	0.82
5	15	16	17	16.00	0.82
6	19	17	17	17.67	0.94
7	17	18	18	17.67	0.47

Table 2
Statistical quality of models calculated with optimal descriptors calculated with both HSG and SMILES for three random splits into sub-training, calibration, and test sets. Best models are indicated by bold.

Threshold	Probe 1	Probe 2	Probe 3	Average	Dispersion
SPLIT 1					
R^2_{test} 1	0.7620	0.7617	0.7595	0.7611	0.0011
2	0.7643	0.7661	0.7668	**0.7657**	0.0011
3	0.7605	0.7605	0.7581	0.7597	0.0012
4	0.7511	0.7461	0.7508	0.7493	0.0023
5	0.7562	0.7557	0.7500	0.7540	0.0028
6	0.7464	0.7469	0.7445	0.7459	0.0010
7	0.7286	0.7345	0.7292	0.7307	0.0026
N^* 1	13	12	14	13.00	0.82
2	12	13	13	**12.67**	0.47
3	15	14	13	14.00	0.82
4	15	13	14	14.00	0.82
5	14	15	16	15.00	0.82
6	15	15	15	15.00	0.00
7	15	15	15	15.00	0.00
SPLIT 2					
R^2_{test} 1	0.7754	0.7805	0.7744	0.7767	0.0027
2	0.7792	0.7762	0.7766	0.7774	0.0013
3	0.7833	0.7821	0.7826	**0.7827**	0.0005
4	0.7804	0.7822	0.7819	0.7815	0.0008
5	0.7699	0.7674	0.7717	0.7696	0.0018
6	0.7804	0.7713	0.7798	0.7771	0.0042
7	0.7686	0.7704	0.7725	0.7705	0.0016
N^* 1	11	13	12	12.00	0.82
2	11	12	13	12.00	0.82
3	12	11	12	**11.67**	0.47
4	11	12	12	11.67	0.47
5	18	13	13	14.67	2.36
6	19	20	18	19.00	0.82
7	18	16	18	17.33	0.94
SPLIT 3					
R^2_{test} 1	0.8120	0.8143	0.8138	0.8134	0.0010
2	0.8157	0.8139	0.8188	0.8162	0.0020
3	0.8158	0.8175	0.8204	**0.8179**	0.0019
4	0.8028	0.8023	0.8045	0.8032	0.0009
5	0.7932	0.7918	0.7936	0.7929	0.0008
6	0.7758	0.7829	0.7818	0.7802	0.0031
7	0.7761	0.7773	0.7722	0.7752	0.0022
N^* 1	12	13	12	12.33	0.47
2	12	12	14	12.67	0.94
3	13	13	12	**12.67**	0.47
4	13	12	12	12.33	0.47
5	14	14	13	13.67	0.47
6	12	12	13	12.33	0.47
7	13	15	14	14.00	0.82

between R^2_m and R'^2_m[20]. $\overline{R^2_m}$ should be more than 0.5 and ΔR^2_m should be smaller than 0.2):

Split from Ref. [1]

$$pLD50 = 1.6993(\pm 0.0209) + 0.0888(\pm 0.0005) * DCW(2,17) \quad (5)$$

$n = 107, r^2 = 0.7251, q^2 = 0.7145, s = 0.889, F = 277$(sub – training set)

$n = 115, r^2 = 0.8030, r^2_{pred} = 0.7958, s = 0.878, F = 461$(calibration set)

$n = 75, r^2 = 0.7675, r^2_{pred} = 0.7547, s = 0.905, F = 241$(test set)

$R_m^2 = 0.7212, R'^2_m = 0.5142, = 0.6177, \Delta R_m^2 = 0.2050$

Random split 1

$$pLD50 = 1.5779(\pm 0.0164) + 0.0774(\pm 0.0004) * DCW(2,13) \quad (6)$$

$n = 149, r^2 = 0.7006, q^2 = 0.6920, s = 1.04, F = 344$(sub – training set)

$n = 59, r^2 = 0.8855, r^2_{pred} = 0.8754, s = 0.600, F = 441$(calibration set)

$n = 89, r^2 = 0.7680, r^2_{pred} = 0.7564, s = 0.878, F = 288$(test set)

$R_m^2 = 0.7413, R'^2_m = 0.6109, \overline{R^2_m} = 0.6761, \Delta R_m^2 = 0.1304$

Random split 2

$$pLD50 = 1.9378(\pm 0.0184) + 0.0667(\pm 0.0004) * DCW(3,12) \quad (7)$$

$n = 138, r^2 = 0.6506, q^2 = 0.6400, s = 1.06, F = 253$(sub – training set)

$n = 82, r^2 = 0.8493, r^2_{pred} = 0.8388, s = 0.745, F = 451$(calibration set)

$n = 77, r^2 = 0.7838, r^2_{pred} = 0.7705, s = 0.865, F = 272$(test set)

$R_m^2 = 0.7310, R'^2_m = 0.5342, \overline{R^2_m} = 0.6326, \Delta R_m^2 = 0.1968$

Random split 3

$$pLD50 = 1.9381(\pm 0.0162) + 0.0744(\pm 0.0003) * DCW(3,13) \quad (8)$$

$n = 152, r^2 = 0.6885, q^2 = 0.6796, s = 1.03, F = 332$(sub – training set)

$n = 57, r^2 = 0.8389, r^2_{pred} = 0.8257, s = 0.808, F = 286$(calibration set)

$n = 88, r^2 = 0.8138, r^2_{pred} = 0.8004, s = 0.765, F = 376$(test set)

$R_m^2 = 0.7308, R'^2_m = 0.5434, \overline{R^2_m} = 0.6371, \Delta R_m^2 = 0.1875$

Fig. 2 displays the model of pLD50 calculated with Eq. (5).

There have been earlier studies related to our work. Statistical characteristics of the model for toxicity towards *D. magna* described in the literature [1] are the following: $n = 222$, $r^2 = 0.695$ (training set), and $n = 75$, $r^2_{pred} = 0.741$, $R^2_m = 0.707$ (test set). Thus, Eq. (5) can be estimated as at least equivalent to the above-mentioned model [1]. Models calculated with Eq. (6)–(8) also have good statistical quality. Thus the predictions which are calculated with the CORAL software can be estimated as robust.

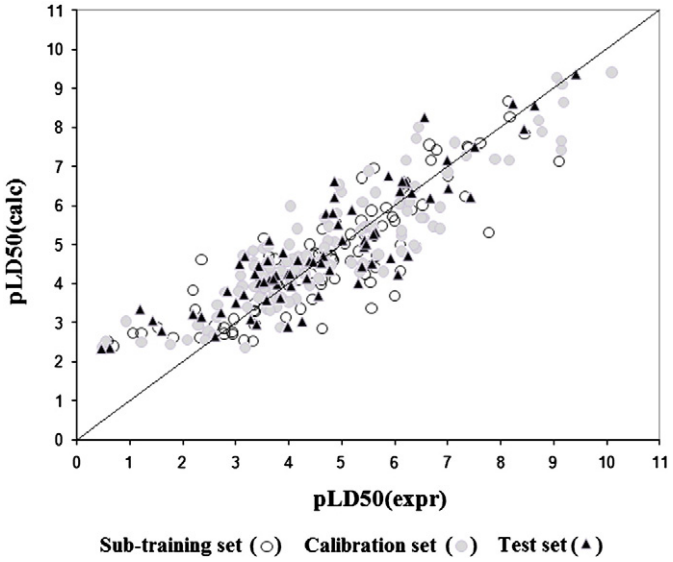

Fig. 2. Correlation between experimental and predicted toxicity towards *Daphnia magna* calculated using Eq. (5).

According to REACH [21], the determination of whether a QSAR result may be used to replace an experimental test result can be broken down into the following three main steps:

1. an evaluation of the scientific validity (relevance and reliability) of the model;
2. an assessment of the applicability of the model to the chemical of interest and the reliability of the individual model prediction;
3. an assessment of the adequacy of the information for making the regulatory decision, including an assessment of completeness, i.e. whether the information is sufficient to make the regulatory decision, and if not, what additional (experimental) information is needed.

To be used as a full replacement of an experimental test, all three conditions need to be fulfilled. Even in cases where some information elements are missing, QSAR results may still be used in the research aspect.

The evaluation of the scientific validity (relevance and reliability) of the CORAL model can be done from probabilistic point of view. Each character of SMILES is a part of information on the molecular structure. Each fact related to the molecular structure can be informative or uninformative for a given model. Consequently, each run of the CORAL with different threshold and number of epochs of the Monte Carlo optimization represents typical experiment. The result can be positive (good model) or negative (poor model). An additional checking up of a good model is the reproducibility of the statistical quality in series of runs of the Monte Carlo optimization.

The applicability domain can be defined on basis of the analysis of correlation weights of molecular attributes extracted from SMILES and graph: there are attributes with stable positive values and attributes with stable negative values of their correlation weights. It is also possible the presence of attributes which are characterized by mixed values of the correlation weights in the series of the optimization (i.e. both positive and negative values of the correlation weights). Finally, for a CORAL model with threshold more than zero, there are rare attributes which have correlation weights equal to zero. We deem that a first approximation of criterion for selection of the compound which falls in the

applicability domain is the presence of attributes with stable positive or stable negative values of the correlation weights in the series of the Monte Carlo optimization, because their influence is apparent: positive correlation weights are promoters of the increase and negative correlation weights are promoters of decrease of an endpoint. The list of stable promoters (increase or decrease) may be used to formulate mechanistic interpretation of the endpoint.

Most probably, the CORAL models cannot be used as a full replacement of an experimental test if data set used for the building up those models is not large. Even in the case of large data set, precision of the CORAL model will be worse than precision of the experimental measurement. However, the precision of the CORAL model can be reliably estimated by means of performance of series of probes with different splits into the sub-training, calibration, and test sets. Taking into account all aforementioned circumstances, one can conclude that the CORAL model can be useful for praxis.

Supplementary materials section contains details of the three splits into the sub-training, calibration, and test sets.

Conclusions

CORAL software is able to be an efficient tool to build up a model toxicity towards *D. magna* for the set of diverse substances. The predictive potential of the applied approach was tested with four random splits into the sub-training, calibration, and test sets. The best results were obtained using the hybrid version of the representation of the molecular structure i.e. taking into account representation of the molecular structure by both molecular graph and SMILES.

Acknowledgements

We thank ANTARES (the project number LIFE08-ENV/IT/00435), the NSF CREST Program (Grant # HRD-0833178), Department of Defense (DoD) for the HPCDNM project (Contract # W912HZ-09-C-0108) and CMCM project (Contract # W912HZ-1–2-0045) through the U.S. Army/Engineer Research and Development Center (Vicksburg, MS) and the National Science Foundation (NSF/CREST HRD-0833178), EPSCoR Award #:362492-190200-01\NSFEPS-0903787 for financial support. Also, the authors express their gratitude to Dr. L. Cappellini, Dr. G. Bianchi and Dr. R. Bagnati for valuable consultations on the computer use. Finally, the authors express gratitude J. Baggott for English edition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.chemolab.2011.10.005.

References

[1] S. Kar, K. Roy, Journal of Hazardous Materials 177 (2010) 344–351.
[2] S. Kar, K. Roy, Chemosphere 81 (2010) 738–747.
[3] A.R. Katritzky, S.H. Slavov, I.S. Stoyanova-Slavova, I. Kahn, M. Karelson, J. Toxicol, Environmental Health—Part A 72 (2009) 1181–1190.
[4] S.P. Niculescu, M.A. Lewis, J. Tigner, SAR and QSAR in Environmental Research 19 (2008) 735–750.
[5] C. Porcelli, E. Boriani, A. Roncaglioni, A. Chana, E. Benfenati, Environmental Science and Technology 42 (2008) 491–496.
[6] A.A. Toropov, E. Benfenati, Bioorganic & Medicinal Chemistry 14 (2006) 2779–2788.
[7] E. Lo Piparo, F. Fratev, F. Lemke, P. Mazzatorta, M. Smiesko, J.I. Fritz, E. Benfenati, Journal of Agricultural and Food Chemistry 54 (2006) 1111–1115.
[8] R. Todeschini, M. Vighi, R. Provenzani, A. Finizio, P. Gramatica, Chemosphere 32 (1996) 1527–1545.
[9] 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, Office for Official Publications of the European Communities (OPOCE), Luxembourg, 2006, L 396/850 of 30.12.2006.
[10] P.R. Duchowicz, A.G. Mercader, F.M. Fernández, E.A. Castro, Chemometrics and Intelligent Laboratory Systems 90 (2008) 97–107.
[11] T. Ivanciuc, O. Ivanciuc, D.J. Klein, Molecular Diversity 10 (2006) 133–145.
[12] M. Fernarndez, J. Caballero, A.M. Helguera, E.A. Castro, M.P. Gonzalezd, Bioorganic & Medicinal Chemistry 13 (2005) 3269–3277.
[13] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, QSAR and Combinatorial Science 25 (2006) 928–935.
[14] T. Puzyn, N. Suzuki, M. Haranczyk, J. Rak, Journal of Chemical Information and Modeling 48 (2008) 1174–1180.
[15] CORAL, http://www.insilico.eu/CORAL(Accessed June 15, 2011).
[16] A.A. Toropov, A.P. Toropova, I. Gutman, Croatica Chemica Acta 78 (2005) 503–509.
[17] D. Weininger, Journal of Chemical Information and Computer Sciences 28 (1988) 31–36.
[18] D. Weininger, A. Weininger, J.L. Weininger, Journal of Chemical Information and Computer Sciences 29 (1989) 97–101.
[19] D. Weininger, Journal of Chemical Information and Computer Sciences 30 (1990) 237–243.
[20] P.K. Ojha, I. Mitra, R.N. Das, K. Roy, Chemometrics and Intelligent Laborary Systems 107 (2011) 194–205.
[21] REACH Regulation (EC) No 1907/2006, http://guidance.echa.europa.eu/docs/guidance_document/information_requirements_r6_en.pdf.