Prioritisation of Polybrominated Diphenyl Ethers (PBDEs) by Using the QSPR-THESAURUS Web Tool

Igor V. Tetko,^{1,2} Pantelis Sopasakis,¹ Prakash Kunwar,¹ Stefan Brandmaier,¹ Sergii Novoratskyi,² Larysa Charochkina,³ Volodymyr Prokopenko³ and Willie J.G.M. Peijnenburg^{4,5}

¹Institute of Structural Biology, Helmholtz-Zentrum München — German Research Centre for Environmental Health (GmbH), Munich, Germany; ²eADMET GmbH, Neuherberg, Munich, Germany; ³Institute of Bioorganic Chemistry and Petrochemistry, National Academy of Sciences of Ukraine, Kyir, Ukraine; ⁴National Institute of Public Health and the Environment (RIVM), Laboratory for Ecological Risk Assessment, Bilthoven, The Netherlands; ⁵Leiden University, Institute of Environmental Sciences (CML), Department of Conservation Biology, Leiden, The Netherlands

Summary — The prioritisation of chemical compounds is important for the identification of those chemicals that represent the highest threat to the environment. As part of the CADASTER project (http://www.cadaster.eu), we developed an online web tool that allows the calculation of the environmental risk of chemical compounds from a web interface. The environmental fate of compounds in the aquatic compartment is assessed by using the SimpleBox model, while adverse effects on the aquatic compartment are assessed by the Species Sensitivity Distribution approach. The main purpose of this web tool is to exemplify the use of quantitative structure–activity relationships (QSARs) to support risk assessment. A case study of QSAR integrated risk assessment of 209 polybrominated diphenyl ethers (PBDEs) demonstrates the treatment and influence of uncertainty in the predicted physicochemical and toxicity parameters in probabilistic risk assessment.

Key words: fate assessment, hazard assessment, multimedia models, online tools, risk assessment.

Address for correspondence: Igor Tetko, Institute of Structural Biology, Helmholtz-Zentrum München — German Research Centre for Environmental Health (GmbH), Neuherberg D-85764, Munich, Germany. E-mail: itetko@vcclab.org

Introduction

The estimation of the risk of chemical compounds is an important part of the European Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) system. The risk presented by chemical compounds depends on their toxicity and emission scenario, as well as on their physicochemical properties, which govern their distribution, metabolism and degradation in the environment. In general, the evaluation of risk is a very complex problem, since chemical compounds, depending on their physicochemical properties, can potentially impact the environment in very different ways. For example, some compounds can be very persistent and can cause their effects over tens and even thousands of years. Examples of such chemicals are perfluorinated compounds, some of which can have half-lives for atmospheric degradation of thousands of years (1). Other chemicals are easily degraded or have an extremely low solubility, and will therefore not be of concern.

Several multimedia models, such as the Multimedia Environmental Pollutant Assessment System (MEPAS; 2), the Multimedia Contaminant Fate, Transport, and Exposure Model (MMSOILS) provided by the US Environmental Protection Agency (http://www.epa.gov/ceampubl/mmedia), and the SimpleBox tool (3) developed by the Dutch National Institute of Public Health and the Environment (RIVM), have been used to provide a means of simulating the accumulation and degradation of chemicals in different media, based on a number of explicit mathematical models for the transfer and degradation of molecules.

In this contribution to the field, as part of the EU CADASTER project, the estimation of the Predicted Environmental Concentration (PEC) and the Predicted No-Effect Concentration (PNEC) were performed by using SimpleBox (3) for the former, and the Species Sensitivity Distribution (SSD) approach (4, 5) for the latter. Both methods require the input of various physicochemical properties of the chemicals, as well as data on their ecotoxicity. Since such data are usually not readily available, risk assessors frequently face the dilemma of having to find experimental values for these properties, or accepting their predicted values based on the properties of similar molecules.

QSAR (quantitative structure-activity relationship) methods permit learning from available experimental data. They can be used to substitute experimental measurements, thus decreasing cost, time and, when in vivo toxicological studies are conducted to generate effect data, also animal use. Like experimental measurements, QSAR predictions also contain some uncertainty. In previous work (6), the fate analysis of five polybrominated diphenyl ethers (PBDEs) was performed, based on QSARpredicted uncertainties. Compounds of this class have received a lot of attention, due to their persistence and toxicity. Therefore, the commercial production of some of these chemicals is restricted according to the Stockholm Convention on Persistent Organic Pollutants. The purpose of this work was to demonstrate how one can perform risk estimation and prioritisation of chemical compounds by using QSAR predictions, and how the uncertainty of the QSAR predictions employed can be included into the environmental risk assessment.

Methods

The use of the SimpleBox model required a number of physicochemical properties of molecules, which are listed in Table 1.

When experimental or predicted values are not available, the user can employ pre-defined default values, which are usually selected to provide a conservative estimation of the fate assessment. Some of these parameters, such as "Standard mass fraction organic carbon in soil/sediment", "Gas phase diffusion coefficient", "Water phase diffusion coefficient", and "Mineral density sediment and soil", are likely to be the same for different compounds. Several other parameters can be calculated by using predefined formulae (3) from other physicochemical parameters — for example, the "Gas/water partition

 Table 1: List of physicochemical and biological properties required for risk assessment in

 SimpleBox

Property name	Abbreviation	Units	QSAR model
Gas phase DIFFUSION coefficient	DIFFgas	$[m^2.s^{-1}]$	
Water phase DIFFUSION coefficient MOLECULAR WEIGHT	DIFFwater Molweight	$[{ m m}^2.{ m s}^{-1}]$ [g.mol ⁻¹]	
Soil Organic Carbon/Water Partition Coefficient	Koc	[—]	Y
Solid/water PARTITION COEFFICIENT for standard solids Octanol/water PARTITION COEFFICIENT Standard mass FRACTION organic carbon in soil/sediment Mineral DENSITY sediment and soil	Kp Kow CORG RHOsolid	[–] [–] [kg.m ^{–3}]	Y
Gas/water PARTITION COEFFICIENT at 25°C VAPOUR PRESSURE at 25°C ENTHALPY of vaporisation Water SOLUBILITY at 25°C ENTHALPY of dissolution Junge's constant Melting point	Kh Pvap25 H0vap Sol25 H0sol JungeCons Tm	[-] [Pa] [kJ.mol ⁻¹] [mg.L ⁻¹] [kJ.mol ⁻¹] [Pa.m] [°C]	Y Y Y
Gas phase degradation RATE CONSTANT at 25°C OH radical CONCENTRATION FREQUENCY FACTOR OH radical reaction ACTIVATION ENERGY OH radical reaction	kdeg.air C.OHrad k0.OHrad Ea.OHrad	${f [s^{-1}]}\ [cm^{-3}]\ [cm^3.s^{-1}]\ [kJ.mol^{-1}]$	Ya Y
Dissolved phase degradation RATE CONSTANT at 25°C Biodegradability test result CONCENTRATION BACTERIA in test water RATE INCREASE factor per 10°C	kdeg.water biodeg BACT.test Q.10	${f [s^{-1}]}\ [r/r-/i/p]\ [CFU.ml^{-1}]\ [-]$	Y
Bulk degradation RATE CONSTANT standard sediment at 25°C	kdeg.sed	[s ⁻¹]	Yb
Bulk degradation RATE CONSTANT standard soil at 25°C	kdeg.soil	$[s^{-1}]$	Yb

^akdeg.air was considered to be a sum of photolytic and OH degradation. ^bThe degradation in soil and sediments were considered to be two and nine times higher than in water, as considered by the US Environmental Protection Agency (http://www.epa.gov/pbt/tools/toolbox.htm) and also in the previous study (6).

coefficient at 25°C" is approximated, based on the solubility of chemical compounds in water and their vapour pressure. Of course, the user can also provide experimental values for all parameters and import more-confident estimations. However, in the absence of both experimental and predicted values, the use of the default, or calculated, values is the only feasible solution.

In the previous study (6), QSAR models for seven physicochemical parameters were employed (see Table 1). In addition to these properties, we also considered the QSAR model for the octanol/water partition coefficient, which was developed by the CADASTER project participants (7). All of the models from the previous study (6) were considered for implementation in the web tool. The models for melting point (Tm, °C; 7), octanol/water partition coefficient (log K_{ow}; 7), vapour pressure (Vp, log [1/Pa]; 7), photolysis lifetime in air (log [$t_{photo_degrada}$. tion], hour; 8) were developed exclusively with 2-D descriptors, and thus were easily reproduced. The model for water solubility (log S, log [mol/L]) that was used in the previous study was based only on *n* = 12 molecules (7), and it was not validated. Instead of this model, we evaluated solubility with the ALOGPS program, which has been demonstrated to have a very good accuracy in several studies (9, 10). We advanced this program to version 3.01c by extending the training set to 8102 compounds. The ALOGPS 3.01c predictive error for the twelve PBDE molecules was a root mean squared error (RMSE) = 0.48, which is lower than its RMSE = 0.69 for the whole training set. ALOGPS 3.01c thus provided good accuracy for the PBDE compounds and was used to predict this property.

Several other models used in the previous study (6) were based on 3-D descriptors, which were calculated by using DRAGON, as well as MOPAC programs. The chemical structures were optimised by using the AM1 semi-empirical method in the HYPERCHEM program (Version 7.03 for Windows, 2002), and descriptors were calculated by using *.hin files. An attempt to re-implement them within the QSPR-THESAURUS (http://gspr-thesaurus.eu) did not succeed, presumably due to differences in the structure optimisation protocols. Indeed, QSPR-THESAURUS uses CORINA (11) for the generation of 3-D structures that are further processed by using the AM1 method provided by MOPAC 7.1, developed by Stewart (12). Therefore, these models were recalculated by using the standard protocol available at the QSPR-THESAURUS website. We used the neural network method, which was applied to two sets of descriptors: a) E-State indices (13) and predicted solubility; and b) octanol/water partition coefficients predicted with the ALOGPS 2.1 program (10) for all the properties analysed. The performances of the models developed were evaluated by using the standard five-fold cross-validation protocol. The newly-developed models had similar performances in comparison to the models used in the previous study (6).

The biodegradation half-life in water was estimated in the previous study (6) by using a model developed by Aronson et al. (14). Aronson et al. mapped Biowin3 category predictions (from EPI Suite) to a set of experimental half-lives, which were observed for each category. The categories were formed according to the Biowin3 quantitative predictions (Biowin3 "primary" predictions). The accuracy of this model was not very high. We tested it for the prediction of n = 249 molecules used in a Molcode model Q8-10-30-265 available at the JRC website (http://qsardb.jrc.it). The Aronson et al. model (14) performance of RMSE = 0.76 was significantly lower than the $RMSE = 0.42 \log units$ calculated by the Molcode model. Unfortunately, the Molcode model was developed by using 3-D descriptors calculated with the CODESSA program (15), which were absent from the QSPR-THESAURUS database. Therefore, this model could not be reproduced. The use of our standard neural network-based approach, yielded a model with RMSE = 0.40, by using a five-fold cross-validation protocol for the same set of molecules. The use of the same training set protocol as in the Molcode model provided a calculated coefficient of determination of $Q^2 = 0.86$ for the testing set (n =83) compounds. This coefficient was higher than the $Q^2 = 0.82$ reported for the same compounds by the authors of the Molcode model. Thus, the newly implemented model provided the predictions of the highest accuracy compared to the previous studies, and it was used to estimate the degradation of the chemical compounds in water.

All of the models developed and the data used, as well as the statistical parameters of the models, are accessible on the QSPR-THESAURUS database webpage (http://qspr-thesaurus.eu/).

Emission scenario

As in the previous work (6), the compound air emission scenario was considered. The parameters analysed in the previous work (persistency and long-range transport potential) do not depend on the factual emission volume. In the current project, an emission rate of 10 tonnes/year to the air with an uncertainty of 10% was used. These numbers are of the same order of magnitude as the estimation of 22–31 tonnes peak emission of PBDE-47 in 1997 (16).

Species Sensitivity Distribution (SSD)

The SSD approach was used to estimate the probability distribution of PNEC values. The theoretical background of SSD has been explained elsewhere (4, 5). For the purpose of this analysis, we used the 5% percentile at the 50% confidence value as the estimate of PNEC.

Aquatic toxicity

The toxicity models for *Daphnia* and two fish species were used to predict the toxicity of chemical compounds to aquatic species. Unfortunately, there are only limited aqueous toxicity data for PBDEs, which are not sufficient to build any models for this class of compounds. Therefore, in addition to PBDEs, we also used diphenyl ethers and aryl bromides. This allowed us to create acceptable training sets for three species, which were used to model the data. Nonetheless, the models developed should be considered as more suitable for demonstration purposes rather than being state-of-theart models. The newly-developed models, including the underlying data, are publicly available at http://qspr-thesaurus.eu/article/11457.

Uncertainty in input parameters

A Gaussian distribution was assumed to best fit the uncertainties in the predictions. While researchers often use the Student *t*-distribution for linear models, its basic assumptions about identical, independent, and normally distributed model errors are frequently not valid. In addition, the correlations of input variables and, moreover, the process of variable selection cannot be easily accounted for, and could result in an erroneous estimation of standard errors of predictions (SEPs). The standard error of the predictions is estimated as:

$$[SEP(\mathbf{Y}_p)]^2 = s^2 (1 + \mathbf{X}_p^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_p) = s^2 (1 + \mathbf{h})$$

[Equation 1]

where s^2 is the model error, **X** is the matrix of the explanatory variables (descriptors) for the training molecules, \mathbf{X}_p is the same matrix for the *p* test molecules and **h** is the vector of calculated leverage values for the latter molecules. As it was shown in the previous benchmarking study, the leverage (17) provided one of the lowest performances for the estimation of the prediction errors for molecules. Moreover, in practice, **h** has values that are usually less than one and, thus, the prediction errors are dominated by the average errors of the model. Therefore, we assumed that prediction errors for linear models followed a normal distribution, whose standard deviation is estimated by *s*.

For neural network models, we estimated the prediction errors based on the standard deviation of ensemble predictions. This approach provided the best estimation in benchmarking studies both for regression and classification models (17).

Risk estimation

After calculating the distribution of the predicted values for fate (PEC, as provided by SimpleBox) and effect (PNEC, as provided by SSD analysis) assessment, one could easily estimate the risk by the numerical integration of the overlap of both distributions. However, since the number of steps in Monte Carlo iterations was limited to N = 10,000, the estimation of risks with p < 1/N was not possible. In order to do this, we also calculated median values for both distributions, and used their difference to rank molecules and rule out the ones that exhibit a high risk of toxicity.

Results and Discussion

The software that was developed in this project is available at http://qspr-thesaurus.eu/risk. It can be accessed through the menu item "Risk Assessment" of the web tool. The QSPR-THESAURUS website itself was developed as an extension of the Online CHEmical Modeling environment (OCHEM; 18). Any modern web browser that supports JavaScript can be used to run it.

Figure 1 demonstrates the graphical interface that allows users to specify data for the fate assessments. The user can either select experimental values, which are available in the QSPR database, use model predictions, or provide their values for some of the parameters based on expert judgement. All of the intermediate results from the Monte Carlo simulation runs (both for fate and effect assessments) can be downloaded and used for offline analysis.

The online supplementary information (Table S1) contains the risk estimations for all 209 compounds from the PBDE data set. A graphical representation of the calculated results is shown in Figure 2. It is clear that compounds with larger numbers of Br atoms are more likely to pose a significant risk to the environment. Indeed, all PBDE congeners with more than five bromine atoms were predicted to yield a non-zero risk and could be toxic to aquatic species under the emission scenario considered. An example of a risk assessment for PBDE-177 is given in Figure 3.

The web tool developed also allows a visual inspection of the impact of individual physicochemical parameters on the fate assessment of chemical compounds via sensitivity analysis. The sensitivity is evaluated by using the Spearman's rank correlations between each individual parameter and the calculated PEC values, provided by SimpleBox. The use of Monte Carlo simulations Figure 1: Graphical interface to introduce properties of a PBDE molecule (PBDE-183) for the fate assessment

là
sumed to be emitted
ission ar 10 1
Mions 0
operues and Uncertai Description
Relative Molecula
Water SOLUBILITY
Melting point
VAPOR PRESSURE
il Organic Carbon-Wat Coefficient

The user can use values stored in the database (white area), calculate properties by using models (blue area) and also provide them (red area) as expert knowledge.

-3.30±0.700 log(mg/L) 170±20.0 °C (in AD)

queous Solubility Point (BFR) [...] Pressure (BFR)

Ξ

Exp. St. Dev.

Exp. Value

Provide Values

Prediction

Model

¥.

narked as *. The parameters in the advanced options section are optional.)

.....

.....

0 0 0 0

-3.30±0.600 -log(s) (in AD)

-7.60±0.700 lg(1/s) 5.20±0.600 log10 log(Pa) (in AD) -5.90±0.200

> HMGU: Abiotic degradation HMGU: Photolysis rate [...]

in water [...]

•

.

lg([s-1]) ([1-s])6 ([-])5

Dissolved phase degradation RATE CONSTANT at 25 oC Photolysis rate

kdeg.water

kdeg.phot Kow

3U: Koc [...]

Ξ

.....

.....

log(cm^(3)/(molecule*sec)) 8.10±0.200 log10 (in AD)

Rate [...]

-13.0±0.900

UI: logKow (PBDE) [...] HMGU: Atmospheric OH

000

-5.17-log(s) 8.27log10 .

00

lg([cm3.s-1])

FREQUENCY FACTOR OH radical reaction Octanol/water PARTITION COEFFICIENT

k0.0Hrad

Advanced options...

therefore enables us to identify which parameters have the highest impact on the resulting concentration of the chemicals in the environment.

Figure 4 indicates that, for the fate assessment of PBDE-177 (shown in Figure 3), the highest impact is provided by the model predicting degradation in air via photolysis. If this mechanism is not considered, the degradation in air due to hydroxyl radical oxidation, the rate of which is about five-fold smaller than the rate of photolysis, does not have a major effect on the accumulation of chemicals in water. In the latter case, the dominating effect is due to the vapour pressure. The absence of photolytic degradation in the air also increases the environmental risk of the compound. In itself, this is obvious, as in the absence of degradation in air, the predicted concentration in the environment increases, thus contributing to a higher estimated risk.

Conclusions

We have developed a web tool for the exemplification of the use of QSAR models for the fate, effect and risk assessments of chemical compounds. The web tool developed allows an easy and interactive analysis of the fate and effect assessments of chemical compounds by using the SimpleBox and SSD approaches, respectively. The PEC and PNEC values calculated with both of these approaches can be used to provide a risk assessment of chemicals under the described emission scenario. We have also demonstrated its use for the risk assessment of 209 PBDEs. The use of the tool developed is limited by the availability of experimental values or QSAR models for the classes of molecules analysed. The QSPR-THESAURUS database contains a number of such models described in multiple publications (6, 7, 19, 20), which were developed during the CADASTER project for the four classes of compounds. However, in case of experimental values and/or compound classes not yet being available in the database, the users can provide the required properties according to their own expert knowledge. Thus, the approach developed can be easily used for the estimation of risk of new classes of molecules beyond those analysed in the CADASTER project.

Online Supplementary Information

Table S1 contains the predicted risk values for 209 PBDEs and is available at www.frame.org.uk.

Acknowledgements

This study was partly supported by the EU through the CADASTER project (FP7-ENV-2007-212668) and the FP7 MC ITN project Environ-

Figure 2: Environmental risk for PBDEs as function of the number of Br atoms in the molecule



The environmental risk was calculated as a numerical integration of the overlap of PEC and PNEC distributions (see Figure 3).



Environmental fate assessment





log10 concentration



Figure 4: Sensitivity analysis of the most important properties determining the fate of the PBDE-177

a) If photolysis in air is considered (kdeg.phot is non-zero), the variability of the QSPR model to predict this property provides the highest impact on the fate of this chemical compound. b) If photolysis is not taken into account, vapour pressure, which determines the volatility of molecules, becomes the most influential parameter for the accumulation of this compound in water.

mental Chemoinformatics (ECO), grant agreement No. 238701, and the GO-Bio 1B BMBF project iPRIOR, grant agreement No. 315647. The authors thank Dr Ullrika Sahlin (Linnaeus University) for providing the R-code used in this publication, and Professor Mark Huijbregts (Radboud University) and Mr Sarfraz Iqbal (Linnaeus University) for their helpful remarks.

References

- Key, B.D., Howell, R.D. & Criddle, C.S. (1997). Fluorinated organics in the biosphere. *Environmental Science & Technology* 31, 2445–2454.
- Strenge, D.L. & Smith, M.A. (2006). Multimedia Environmental Pollutant Assessment System (MEPAS): Human Health Impact Module Description, 19pp. Richland, WA, USA: Pacific Northwest National Laboratory.
- den Hollander, H.A. & Van de Meent, D. (2004). SimpleBox 3.0: A Multimedia Mass Balance Model for Evaluating the Environmental Fate of Chemicals. RIVM Report 601200003, 155pp. Bilthoven, The Netherlands: RIVM, National Institute of Public Health and the Environment.
- Posthuma, L., Suter II, G.W. & Traas, T.P. (2002). Species Sensitivity Distributions in Ecotoxicology. Boca Raton, FL, USA: Lewis Publishers.
- Aldenberg, T. & Rorije, E. (2013). Species Sensitivity Distribution estimation from uncertain (QSARbased) effects data. ATLA 41, 19–31.

- Iqbal, S., Golsteijn, L., Öberg, T., Sahlin, U., Papa, E., Kovarich, S. & Huijbregts, M.A.J. (2013). Understanding QSPR uncertainty in environmental fate modeling. *Environmental Toxicology & Chemistry* [in press.]
- Papa, E., Kovarich, S. & Gramatica, P. (2009). Development, validation and inspection of the applicability domain of QSPR models for physicochemical properties of polybrominated diphenyl ethers. QSAR & Combinatorial Science 28, 790– 796.
- Raff, J.D. & Hites, R.A. (2007). Deposition versus photochemical removal of PBDEs from Lake Superior air. *Environmental Science & Technology* 41, 6725–6731.
- Tetko, I.V., Tanchuk, V.Y., Kasheva, T.N. & Villa, A.E. (2001). Estimation of aqueous solubility of chemical compounds using E-state indices. *Journal* of Chemical Information & Computer Sciences 41, 1488–1493.
- Tetko, I.V. & Tanchuk, V.Y. (2002). Application of associative neural networks for prediction of lipophilicity in ALOGPS 2.1 program. *Journal of Chemical Information & Computer Sciences* 42, 1136–1145.
- Sadowski, J., Gasteiger, J. & Klebe, G. (1994). Comparison of automatic three-dimensional model builders using 639 X-ray structures. *Journal of Chemical Information & Computer Sciences* 34, 1000-1008.
- Stewart, J.J. (1990). MOPAC: A semiempirical molecular orbital program. Journal of Computer-aided Molecular Design 4, 1–105.
- 13. Hall, L.H. & Kier, L.B. (1995). Electrotopological state indices for atom types a novel combination

of electronic, topological, and valence state information. *Journal of Chemical Information & Computer Sciences* **35**, 1039–1045.

- Aronson, D., Boethling, R., Howard, P. & Stiteler, W. (2006). Estimating biodegradation half-lives for use in chemical screening. *Chemosphere* 63, 1953– 1960.
- Karelson, M., Maran, U., Wang, Y.L. & Katritzky, A.R. (1999). QSPR and QSAR models derived using large molecular descriptor spaces. A review of CODESSA applications. Collection of Czechoslovak Chemical Communications 64, 1551–1571.
- Prevedouros, K., Jones, K.C. & Sweetman, A.J. (2004). Estimation of the production, consumption, and atmospheric emissions of pentabrominated diphenyl ether in Europe between 1970 and 2000. *Environmental Science & Technology* 38, 3224– 3231.
- Tetko, I.V., Sushko, I., Pandey, A.K., Zhu, H., Tropsha, A., Papa, E., Oberg, T., Todeschini, R., Fourches, D. & Varnek, A. (2008). Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: Focusing on applicability domain and overfitting by variable selection. *Journal of Chemical Information & Modeling* 48, 1733–1746.
- Sushko, I., Novotarskyi, S., Korner, R., Pandey, A.K., Rupp, M., Teetz, W., Brandmaier, S., Abdelaziz, A., Prokopenko, V.V., Tanchuk, V.Y., Todeschini, R.,

Varnek, A., Marcou, G., Ertl, P., Potemkin, V., Grishina, M., Gasteiger, J., Schwab, C., Baskin, I.I., Palyulin, V.A., Radchenko, E.V., Welsh, W.J., Kholodovych, V., Chekmarev, D., Cherkasov, A., Aires-de-Sousa, J., Zhang, Q.Y., Bender, A., Nigsch, F., Patiny, L., Williams, A., Tkachenko, V. & Tetko, I.V. (2011). Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *Journal of Computer-aided Molecular Design* **25**, 533–554.

- Cassani, S., Kovarich, S., Papa, E., Roy, P., Rahmberg, M., Nilsson, S., Sahlin, U., Jeliazkova, N., Kochev, N., Pukalov, O., Tetko, I.V., Brandmaier, S., Durjava, M.K., Kolar, B., Peijnenburg, W.J.G.M. & Gramatica, P. (2013). Evaluation of CADASTER QSAR models for the aquatic toxicity of (benzo)triazoles and prioritisation by consensus prediction. *ATLA* 41, 49–64.
- Bhhatarai, B. & Gramatica, P. (2011). Prediction of aqueous solubility, vapor pressure and critical micelle concentration for aquatic partitioning of perfluorinated chemicals. *Environmental Science & Technology* 45, 8120–8128.
- Van Straalen, N.M. (1990). New methodologies for estimating the ecological risk of chemicals in the environment. In *Proceedings 6th International IAEG Congress* (ed. D.G. Price), pp. 165–173. Rotterdam, The Netherlands: Balkema.