Environ. Sci. Technol. **2008,** *42,* 491–496

Regulatory Perspectives in the Use and Validation of QSAR. A Case Study: DEMETRA Model for *Daphnia* Toxicity

CHIARA PORCELLI, ELENA BORIANI, ALESSANDRA RONCAGLIONI, ANTONIO CHANA, AND EMILIO BENFENATI*

Istituto di Ricerche Farmacologiche Mario Negri, via La Masa 19, 20156 Milan, Italy

Received June 14, 2007. Revised manuscript received October 29, 2007. Accepted October 30, 2007.

The DEMETRA acute toxicity model toward the water flea (*Daphnia magna)* was used as a case study to outline a validation method compatible with regulatory use. Reliability, predictive power, uncertainty, and applicability were verified with an external test set of pesticides. Predictions for this external set using the DEMETRA model, developed *ad hoc* for pesticides, were compared with the results using ECOSAR and TOPKAT as benchmarks. The evaluation considered statistical parameters and the presence of errors, with their size and sign. DEMETRA gave good statistical predictions, and the maximum error of the outliers was lower than that with the other two models. DEMETRA gave a limited number of false negatives, and the use of defined rules indicated the level of uncertainty was acceptable.

Introduction

Quantitative structure–activity relationship (QSAR) studies are increasingly challenged to evaluate huge numbers of compounds. The European legislation REACH specifies the use of QSAR models to predict properties of industrial chemicals (*1*). QSAR studies are based on the premise that biological activities may be related to certain chemical properties. A large number of relationships have been reported, where the biological effect, mathematically determined as the output of the model, was defined in relation to some chemical parameters, identified as the model inputs (*2*). The partition coefficient between octanol and water, given as a logarithm called logP or logK_{ow}, has been used in many equations, sometimes in combination with other properties such as electronic factors (*2, 3*). For example, logP is the key factor in programs commonly used to predict aquatic toxicity, such as in the U.S. EPA software ECOSAR (*4*).

Another program giving toxicity predictions is TOPKAT (*5*), which evaluates the so-called "optimal prediction space" (OPS). The identification of boundaries characterizing the model's validity has been studied and debated (*6*).

It is widely accepted that, in the case of aquatic toxicity, models mainly based on logP are quite reliable for compounds acting through a narcotic effect; however, they fail, in particular for more reactive compounds, because of the need to detect more varied mechanisms, and this is a serious drawback if models are required to play a role in protecting the environment.

The EC-funded project DEMETRA, Development of Environmental Modules for Evaluation of Toxicity of pesticide Residues in Agriculture, was aimed at developing predictive models for ecotoxicity of pesticides (*7*). Compared to industrial chemicals, pesticides are quite complex because they typically contain several functional groups, and are intended to have a toxic effect on specific targets through a variety of biological mechanisms, not all fully identified. The DEMETRA models are publicly available through the Internet (*8*). Here we report a further validation test of the model for *Daphnia magna* acute toxicity, on a very large set of more than 100 compounds. We also report the values predicted with ECOSAR and TOPKAT. These two programs were chosen because they give an automatic prediction of toxicity using QSAR, similarly to DEMETRA and therefore usable by regulators. We examine the results in terms of practical utility, unambiguity, and applicability domains of the models.

Materials and Methods

Biological Data and Structure Availability. The new test set was organized by collecting data from the HAIR ecotoxicity database (*9*). This contains data for about 242 pesticides extracted from the German database of the Federal Biological Research Centre for Agriculture and Forestry (BBA). The main sources of BBA data are EU reports on active substances used in plant protection products, published before 2006. However, some data come from various other protocols.

An initial screening was done in order to avoid polymers, inorganic compounds, mixtures of molecules, and mismatch between CAS number and name. Out of the remaining compounds we only collected data on water flea (*Daphnia magna*) LC_{50} 48 h, which is the dose that kills 50% of the fleas after 48 h exposure. Finally, we divided the subset into a new test set (135 compounds) and a set of compounds already used for DEMETRA modeling (74 chemicals). Acute toxicity values were converted to the negative of the logarithm of LC_{50}. For each new compound the chemical structure was checked and downloaded from the ChemIDplus Web site (*10*), then saved as an MDL mol file.

A first analysis compared data in the new database with those already used for the DEMETRA modeling. Only 16 of the 74 common compounds had identical LC_{50} values in the two databases (probably the same experiment) while the other 58 showed a correlation coefficient, R^2, of 0.89 between the log scale values of the two series. Although this indicates a good correlation between the two databases, proving agreement between values and protocol, 15 of these figures differed by more than a factor of 4, and 6 differed by more than 1 order of magnitude. This is a major limit in building predictive QSAR models: the experimental values, i.e. the model input, are an intrinsic source of uncertainty.

Molecular Descriptors. The three models require descriptors calculated on the basis of the bidimensional structure. OpenBabel v 1.0.0.1 was used to create the MDL mol files (*12*). OpenBabel was also used to generate the SMILES codes needed by the TOPKAT and ECOSAR models. The 16 chemical descriptors needed for the DEMETRA model (*8*) were calculated with DRAGON free version 3.0 (*11*). All logP values reported in the HAIR database for the new test set are within the range of logP for DEMETRA training compounds.

DEMETRA Model. The DEMETRA model for *Daphnia* is based on a training set of 220 compounds. The software was built through a hybrid model approach: the final model is composed of three individual models (one based on partial

* Corresponding author e-mail: benfenati@marionegri.it; tel: +390239014420; fax: +390239014735.

10.1021/es071430t CCC: $40.75
Published on Web 11/30/2007

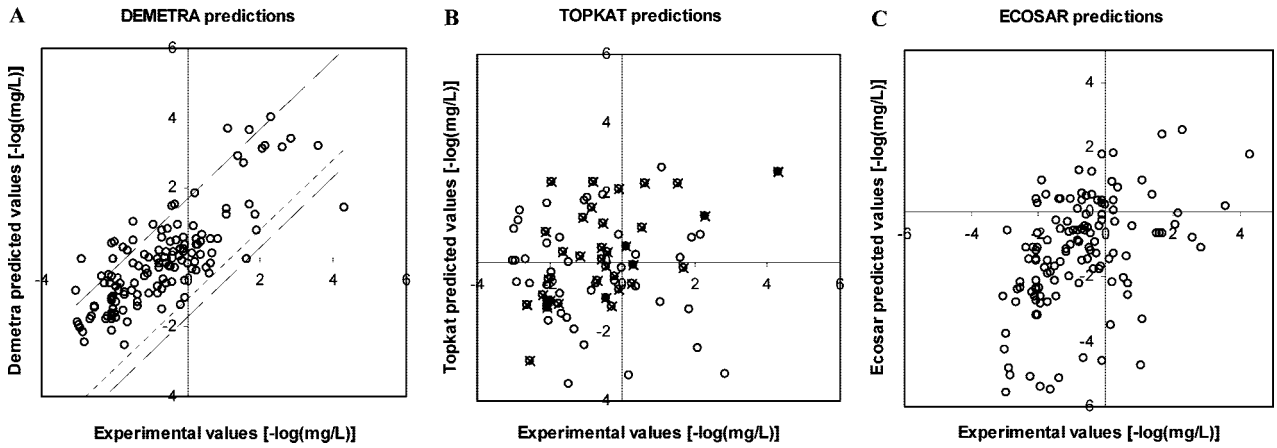

FIGURE 1. Experimental and predicted values of HAIR compounds. (A) DEMETRA prediction (135 compounds, $R^2 = 0.63$). Dashed lines represent an error of a factor of 50 between experimental and predicted values, dotted lines represent a factor of 15 in case of false negatives. (B) TOPKAT prediction of compounds inside OPS/OPS limits without considering the "Coverage Examination" (circles, 78 compounds, $R^2 = 0.02$) with predictions of compounds inside the OPS whose fragments were considered well represented in the training set by the "Coverage Examination" (crosses, 37 compounds, $R^2 = 0.30$). The six compounds in the training set are shown as black circles. (C) ECOSAR prediction (127 compounds, $R^2 = 0.21$).

least-squares and two based on neural networks) joined in a mathematical function that leads them toward a single predictive value. A hybrid model integrates the results of the individual models in an intelligent optimized way, achieving better prediction. The software integrating the individual models combines the strengths of each QSAR model and reduces false negatives (compounds with lower predicted than experimental toxicity, while false positives are those with higher predicted toxicity).

Some restriction rules apply to identify outliers (*7*). These rules, proposed by DEMETRA and generated by visual inspection, were taken into account and new considerations were added to define the two confidence limits better: compounds with a ratio of the observed to predicted values in mg/L either more than 50 or 15 (see Supporting Information). The DEMETRA model works with mmol/L, then transforms the results to mg/L because these are the units used by regulators.

TOPKAT Model. TOPKAT (version 6.1) uses a series of individual models for assessing specific toxicological end points. In the case of *Daphnia magna* (DAPHNIA EC50 v3.1 model), four submodels are available: alcohols, single benzene, other aromatics, and aliphatics; the original models are based on a training set of, respectively, 66, 101, 37, and 34 compounds. Only one model is automatically associated with a new compound, considering some distinctive fragments.

TOPKAT performs and merges two kinds of preprocessing analysis: univariate analysis, the "Coverage Examination", to establish whether all the structural fragments of the query structure are well represented in the training set; and multivariate analysis, called "Optimum Prediction Space -OPS- Examination", to assess whether the query structure falls into the ensemble of good prediction, OPS, or into a border space called the OPS limit. The two steps are summarized in a percentage confidence limit that reflects the information about the two analyses (*5*). The software always gives a prediction of toxicity for a new molecule, but may in some cases be accompanied by an error message.

Five compounds of the new test set were discarded; for three the error was because the SMILES input was not recognized by the software, and for the others the descriptors could not be calculated because the compounds contained Sn.

The output text file indicates whether the compound belongs to the initial database used for training and the results of the two preprocessing analyses; finally the assessment is reported with the confidence limit (*13*).

The manual specifies that a QSAR model is applicable only to query structures that fall inside or in the vicinity of the OPS (*5*). Therefore, we analyzed the prediction of all the compounds inside the OPS and OPS limits (without considering the "Coverage Examination") and, for further details, we also considered the results of the "Coverage Examination" (Figure 1B).

ECOSAR Model. The ECOSAR program v0.99 h is the computer version of the ECOSAR procedure used by the U.S. EPA Office of Pollution Prevention and Toxics (OPPT) for assessing new chemicals. The program automatically classifies the query compound and associates individual QSAR models for each class. The 62 chemical classes considered by ECOSAR are identified by the presence of distinctive functional groups. Whenever more than one class is found, human expert evaluation is required to associate the query structure with the right class, and consequently the correct QSAR model. Eight chemicals from the HAIR data set were discarded because the software could not process the SMILES code or the chemicals were classified in a class for which no QSAR was available, as in the case of Imides.

Since ECOSAR is based on local models, the training set of each class contains fewer chemicals than the training set of DEMETRA (a few dozen at most) and in many cases compounds contain just one functional group, the one identifying the chemical class.

The predictions of toxicity for new compounds rely on linear correlations of experimental toxicity values with the compound's octanol/water partition coefficient, K_{ow}. K_{ow} for the test set was computed by KOWWIN, a companion program contained in the integrated tool EPI suite v3.12 (*14*). The range of K_{ow} values valid for estimating the toxicity is given for each chemical class. If the log K_{ow} is above a certain cutoff the developers suggest that the lower solubility might affect the validity of the prediction and the QSAR models for longer exposure should be used to determine the LC_{50} (*15*). This was taken into consideration in evaluating the models but gave no real improvement; thus the results presented did not distinguish compounds with high log K_{ow}.

Applicability Domain. The applicability domain (AD) is a sensitive issue and DEMETRA and TOPKAT methods each have their own approach. Further methods were investigated independently of the development of the model. AMBIT Discovery v0.04 is a software developed for the identification

of the QSAR applicability domain (*16*). The AD assessment relies on projection of the training set into the model descriptors space or on structural similarity assessed by fingerprints. All techniques in AMBIT Discovery were used to explore the AD of the DEMETRA model for *Daphnia* toxicity and were compared to an *a posteriori* rule base approach. This second approach, developed by DEMETRA (*7*), uses the toxicity prediction values jointly with the identification of structural alerts, by visual inspection. The recognition of some chemical fragments characteristic for the outliers of the model identifies compounds outside the model domain. Visual inspection was done, on the basis of the training set and checked on the test sets, highlighting two sets of rules related to the error measured as the ratio of the max(observed value [mg/L]; predicted value [mg/L]) and min(observed value [mg/L]; predicted value [mg/L]): (1) fragments that give a possible error higher than 50, such as carbamates (only false negatives were considered for this rule); and (2) fragments that give a possible error higher than 15 for false negatives and higher than 50 for false positives, such as benzyl-ethers.

The complete set of rules for outliers is given as Supporting Information.

The results of the two approaches were compared in terms of number of compounds excluded, correlation coefficient (R^2) of the compounds in the AD, and root-mean-square error (RMSE) for the false negatives within the AD. A consensus AD was achieved comprising the AMBIT best result and DEMETRA rules.

Results and Discussion

Validation Methods. The first classical QSAR studies were mainly interested in verifying whether some chemical features were related to a given biological effect. Today there is more interest in the model's predictive power than in simply unveiling such relationships. As a result, a battery of statistical tools has been introduced within the past few years to assess this predictive power. The classical QSAR models indicated mainly the fitting property of the model, given as R^2. Nowadays it is accepted that the model's predictive power has to be measured and reported (*17*). Different tools should be used depending on the model and the number of chemicals used (*18*). Generally, leave-one-out validation is not considered suitable, and good statistical results based on an external set of compounds, never used in the model building steps, are considered a more rigorous test of the predictive power of the model (*19*).

DEMETRA models have been thoroughly validated by internal validation techniques, such as y-scrambling, leave-one-out, leave-more-out. Furthermore, DEMETRA adopted the criteria indicated by Tropsha et al. for the test set (43 compounds) prediction ($Q^2 > 0.5$, slope between 0.85 and 1.15, $(R^2 - R^2_0)/R^2 < 0.1$) to evaluate the model robustness (*19*). The results on the training and test set, with all these criteria satisfied, were R^2training = 0.74, R^2test = 0.70. More details of the validation are given elsewhere (*7*).

Performance. In this second phase we used the 135 HAIR compounds as a new test set. Figure 1A shows the results of the DEMETRA model for *Daphnia* toxicity. The R^2 on the new test set is 0.63 (without applying any AD rule). These results show that the DEMETRA model is a predictive model for pesticides. Overall, the number of compounds used for this exercise was about 80% of the training set—a particularly demanding percentage. Normally the validation uses a smaller proportion.

For a comparison, Figure 1B and C show the results with TOPKAT and ECOSAR. R^2 is 0.21 using ECOSAR and 0.30 using TOPKAT.

The number of compounds outside the OPS and the OPS limits with TOPKAT is 51; 6 were inside the training database and 2 presented further warnings indicating critical features (e.g., "Computed LogP Value Outside the Range Spanned by the Training Set"). ECOSAR detected 19 different chemical classes in the test set but almost half the compounds were classified as Neutral Organics. The classification was univocal for most of the chemicals though for 18 compounds the program assigned more than one class. When there are multiple residues it is up to users to decide the most appropriate model, but this may result in lack of reproducibility. For these 18 chemicals with more than one possible model, we analyzed the predictions of the models for the different classes and found no big differences.

The main difficulty in the use of ECOSAR is the lack of specific QSAR for most classes of pesticides because ECOSAR is designed for industrial chemicals (*15*). For instance, no specific QSAR exist for carbamates and, without any warning, they were classified as esters, amines, or in the more general class Neutral Organic. The limited number of compounds in the training set of the ECOSAR models, and the dependence on a single factor (logP) may explain the difficulty of modeling a large variety of compounds (*20*). Regressions based on logP are not predictive for pesticides, even considering only chemicals which should act through narcosis, which is the theoretical assumption of the models based on logP (*21, 22*).

Another important issue in case of a QSAR for regulatory purposes is transparency. DEMETRA is fully transparent: the toxicity data, chemical structures, descriptors, and algorithms are publicly available, as is the source code (*8*). A detailed description of the modeling procedure has been published (*7*). The transparency of the model and the data availability are important issues according to the OECD principles for validation of QSAR models for regulatory purposes (*23*). Unfortunately, it is not always easy to obtain the data and the models, but this information is essential for a correct evaluation of the model. For instance, in the case of TOPKAT 6 chemicals were present in the training set of the model. These are presented in Figure 1B (filled circles); TOPKAT appears to give better predictions for these chemicals than for the others in Figure 1B. This may be interpreted as overfitting of the model, meaning that for other compounds the model is not able to give similarly accurate results.

Modeling Approaches. There are some differences between DEMETRA, ECOSAR, and TOPKAT. All three address heterogeneous chemical classes. DEMETRA actually more specifically addresses complex chemicals, such as pesticides and their metabolites, TOPKAT includes some pesticides, whereas ECOSAR, not designed for most classes of pesticides, includes mainly simpler chemicals scattered in individual models.

The descriptors the models need can be generated from the 2D structure for all three cases, and this is a good feature because optimization of the 3D structure can involve a variable and time-consuming procedure. The difference is the nature and number of the descriptors: the chemical parameters of ECOSAR are fewer and TOPKAT and DEMETRA introduce more sophisticated parameters, which can be an advantage in describing the different components of the query structure better.

DEMETRA was designed to develop a battery of QSAR models combined within a hybrid model that uses the outputs (the predicted values) of the individual QSAR models as inputs. ECOSAR and TOPKAT contain series of possible QSAR models, which work alternatively. Thus, these two do not integrate the results from their multiple modules. These modules are structured following the same approach. ECOSAR and TOPKAT theoretical reasoning is based on human expertise initially to assign a query chemical to a chemical class based on the presence of key fragments. Then local class-based QSAR models are applied. The models therefore encode explicit knowledge to identify chemical

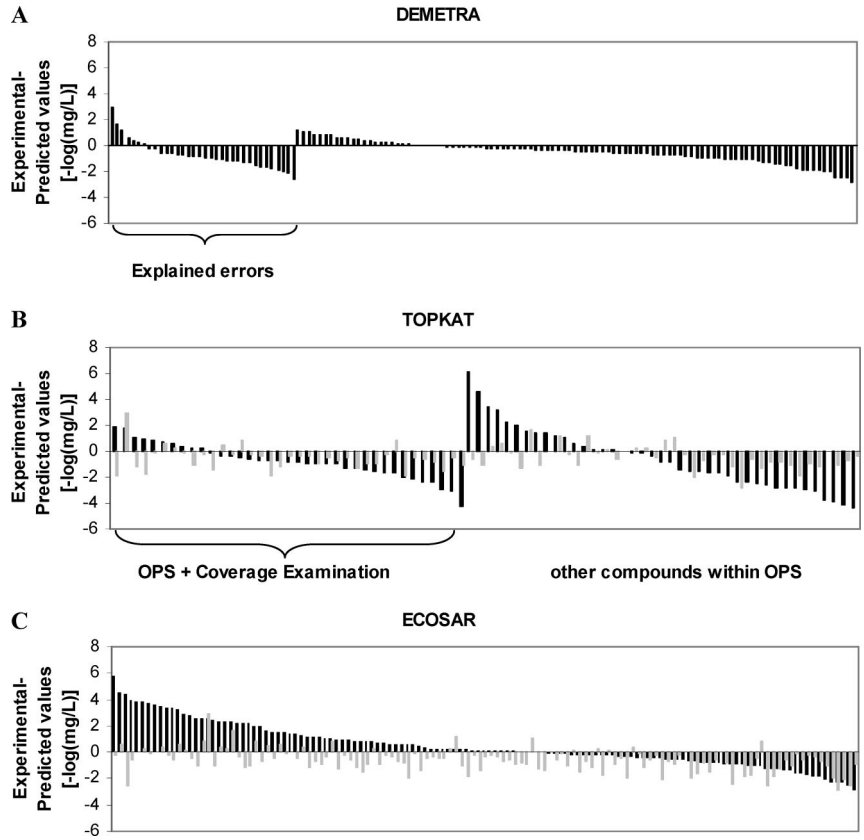

FIGURE 2. Errors for the predictions of the three models expressed as experimental minus predicted value [−log(mg/L)]. (A) DEMETRA errors for each of the 135 compounds from the HAIR data set. The 34 explained errors are compounds outside the applicability domain using the DEMETRA rule-based approach. Range of not-explained errors 1.18/−2.87. (B) TOPKAT errors (black) compared with DEMETRA errors (grey) of the 78 compounds in the OPS. Over brackets, 37 compounds predicted with the maximum degree of confidence for TOPKAT (OPS and "Coverage Examination"). Range of errors 6.08/−4.31 (1.85/−4.28 in the bracketed subset). (C) ECOSAR errors (black) for 127 compounds compared with DEMETRA errors (grey). Range of errors 5.79/−2.86.

classes, thus biasing the final results and introducing a degree of uncertainty.

DEMETRA does not introduce any human-based scheme but exploits modern knowledge discovery techniques. The assumption is that there is implicit knowledge in the data, and suitable information technology tools may extract this in an automatic and reproducible way. Various models have been developed using different approaches to produce as large as possible a basis for the final hybrid model.

Instead of focusing on a specific approach, we preferred to screen large series of chemical and mathematical tools. For wider discussion of the use and basis of hybrid systems in QSAR see Benfenati (*7*). DEMETRA hybrid models, as typical of hybrid systems, produced better results than any single individual model (*7*). Furthermore, if different models produce different values, as with ECOSAR, the results may be conflicting, and the user may be confused. A specifically optimized hybrid model can cope with this outcome better, assigning different weights to different models.

Nature of Errors. Besides the statistical parameters given above, it is important to assess the exact nature of the error given by the model (*7*). Figure 2 shows the errors for DEMETRA, ECOSAR, and TOPKAT. The errors with DEMETRA are much smaller than with the other two. Large errors may pose a serious problem for the use of a model within a risk assessment procedure. Indeed, if such an error cannot be explained by rules that can be used for a new compound, the regulator would prefer to apply a safety limit as large as the maximum error of a given model. The use of safety limits is typical in risk assessment. For instance, experiments might be done for one animal species, and then species sensitivity might be applied to transfer the result to other species. To protect the different species, safety factors are used.

QSARs are not widely used for regulatory purposes, at least in Europe, and on the basis of the precautionary criterion a cautious safety factor may be applied for their predictions. Thus, within DEMETRA we developed models with a limited error. This further criterion is distinct from the common statistical parameters. In addition we introduced different safety limits on the basis of the chemical structure. Our errors were larger for some chemical classes, such as carbamates (5 compounds in this class belong to the new test set, 3 of which predicted with an error higher than 15), so we set restrictive rules to warn the user in case of predictions for these chemicals. A further facility is that users can see the range of values predicted by the individual models that are components of the hybrid model. Thus, they can choose not the final toxicity of the hybrid model, but the most conservative value of the individual component models.

Besides the size of the error, it is interesting to look at its sign. Typically, QSAR models are evaluated using squared parameters, such as R^2 or RMSE, so these positives or negatives have the same importance. For toxicity prediction in a regulatory setting, false negatives, which falsely imply safety, are of much more concern than false positives. For this reason, in training the DEMETRA hybrid model the reduction of false negatives was optimized, and great attention was paid to identifying rules explaining them. This introduces a new perspective in the QSAR models since the issue of false negatives and false positives is addressed in classification models rather than in models predicting continuous values. Figure 2 shows that ECOSAR and TOPKAT

give more false negatives. This is probably because the basic mechanism that is modeled best is narcosis, and deviations from this mechanism are not adequately codified. The DEMETRA model was developed designing the hybrid model to avoid false negatives. Different strategies have been developed and tested, using different mathematical tools, as explained elsewhere (*7*). The results are shown in figure 2A. DEMETRA was focused closely on the regulatory use of the models, considering both the extent and the sign of the errors, besides the statistical validation tools which disregard this. We believe these parameters must always be considered in models for evaluating toxicity.

Applicability Domain. In ECOSAR the AD is not defined explicitly, but relies on the outline of the chemical classes. However, there are ambiguities for conflicting class assignments and the role of multiple functionalities has not been addressed. Different tools now exist to assess the AD, mainly based on measures of chemical similarity (*6*). OPS and checking fragment are examples of these tools in TOPKAT. However, these two preprocessing analyses are not consistently able to identify known active chemicals. In fact, we found several compounds not inside the OPS with acceptable prediction; then too, many compounds were accepted as within the model domain (defined by the "OPS Examination" and the "Coverage Examination"), but the model gave unsatisfactory predictions (Figure 2B).

The approach used in DEMETRA was to identify outliers *a posteriori*, by visual inspection, on the basis of the predicted values, then to explain them in terms of chemical components. Thus, we identify uncertainty values which were larger for some chemical classes. The chemical residues/features characterizing outliers, particularly false negatives, has been discussed in detail (*7*) and the list is given as Supporting Information. Generating different models for these outliers with apparent similarity should be another option but our approach is cautious. Since only a few chemicals are outliers, we decided not to model them, because the statistical basis would be weak, and simply warn users that there may be greater uncertainty in some cases. In other words, the user can decide to apply a more restricted chemical domain in order to use a lower safety factor. In this case some chemical classes may be excluded.

Whether the error is sufficient to ensure safe use of the chemical must be decided considering the expected scenario, which means the presumable exposure level. For risk assessment of ecotoxicological properties, the typical situation involves a toxicity/exposure ratio, whereas in the case of labeling the evaluation is based on the toxicity class. Thus, depending on the application context, users have to decide what degree of uncertainty is acceptable. The errors explained by the DEMETRA rules (strictest case) are marked in Figure 2A for the new test set. Different rules can be applied, with different uncertainties, both for false positives and negatives (see Supporting Information).

In our evaluation of DEMETRA outliers we also used AMBIT, testing its different tools. The Supporting Information (Table S2) shows the best results of the AMBIT procedure compared with the DEMETRA rule-based approach. It was not possible to differentiate outliers without excluding a large number of compounds. A possible reason is that methods based on chemical similarity work on the chemical space, but do not use the toxicity values, i.e. they do not evaluate areas covered well by descriptor space but characterized by a weak model.

In terms of R^2 of the compounds considered in the domain, the results were comparable to the AMBIT approach. DEMETRA was more efficient, however, from a regulatory point of view because the rules we identified explain most of the outliers, particularly false negatives.

Within DEMETRA we identified chemical-based rules, but other types, such as mechanistic rules, can be obtained too (*7*). Some outliers of the DEMETRA model can be explained by human experts on a biochemical basis, e.g. because they affect electron transfer in the cell. But in this case the rule, even if sound because it is based on a known mechanism of toxic action, cannot be used to predict outliers because at the moment it is not possible to predict whether a new chemical will affect the same mechanism. Thus, this kind of rule is not applicable from the regulatory perspective.

Recommendations. Neither ECOSAR nor TOPKAT is a specific model for pesticides so their predictions are not as good as those obtained with the DEMETRA model. This does not mean these models are not appropriate for predicting other end points or compounds (*24*); for example, TOPKAT performs well for bacterial mutagenicity (*25*). The DEMETRA model can be improved by further exploring the chemical space where this model has greater uncertainty. One possibility, for example, is to integrate in the hybrid model a QSAR developed *ad hoc* for carbamates, using further data on these compounds. This could overcome the uncertainty remaining for this class of chemicals.

Acknowledgments

We gratefully acknowledge the financial contribution of the EC projects CAESAR, contract SSPI-022674, and HAIR, contract SSPE-CT-2003-501997 and thank EPA-OPP for kindly providing the database used within DEMETRA project.

Supporting Information Available

DEMETRA restriction rules applied to identify outliers; results of the methods used to assess the AD of DEMETRA model; list of DRAGON descriptors used in DEMETRA model; list of compounds in the new test set. This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) EC 1907/2006. Regulation of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/45/EC and repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC. OJ L 396, 30.12.2006, 1–849.

(2) Cronin, M. T. D.; Livingstone, D. J. *Predicting Chemical Toxicity and Fate*; CRC Press: Boca Raton, FL, 2004.

(3) Pavan, M.; Worth, A. P.; Netzeva, T. I. *Comparative Assessment of QSAR Models for Aquatic Toxicity*; JRC Report EUR 21750 EN; European Commission, Joint Research Centre: Ispra, Italy, 2005.

(4) U.S. EPA. *Ecological Structure Activity Relationships*; http://www.epa.gov/oppt/newchems/tools/21ecosar.htm (accessed 1/12/2006).

(5) TOPKAT, version 6.1. Accelrys Inc.: San Diego, CA, 2001.

(6) Netzeva, T. I.; Worth, A. P.; Aldenberg, T.; Benigni, R.; Cronin, M. T. D.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D. W.; Schultz, T. W.; Stanton, D. T.; Van de Sandt, J. J. M.; Tong, W.; Veith, G.; Yang, C. Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *Atla Altern. Lab. Anim.* **2005**, *33*, 155–173.

(7) Benfenati, E. *Quantitative Structure-Activity Relationship (QSAR) for Pesticide Regulatory Purposes*; Elsevier: Amsterdam, 2007.

(8) DEMETRA project (Development of Environmental Modules for Evaluation of Toxicity of pesticide Residues in Agriculture); http://www.DEMETRA-tox.net (accessed 1/12/2006).

(9) HAIR project (Harmonised environmental indicators for pesticide risk); http://www.rivm.nl/rvs/overige/risbeoor/Modellen/HAIR.jsp (accessed 1/12/2006).

(10) ChemIDplus Advanced. Free database of 350,000 chemical compounds; http://chem.sis.nlm.nih.gov/chemidplus/ (accessed 1/12/2006).

(11) DRAGON. http://www.talete.mi.it (accessed 1/12/2003), software for Molecular Descriptor Calculations, v 3.0. Talete srl.
(12) Open Babel. http://openbabel.sourceforge.net/ (accessed 1/12/2006).
(13) Gombar, V. K.; Enslein, K.; Blake, B. W. Assessment of developmental toxicity potential of chemicals by quantitative structure-toxicity relationship models. *Chemosphere* **1995**, *31*, 2499–2510.
(14) U.S. EPA. *Exposure Assessment Tools and Models*; http://www.epa.gov/opptintr/exposure/pubs/episuite.htm (accessed 1/12/2006).
(15) Clements, R. G.; Nabholz, J. V.; Zeeman, M. Estimating Toxicity of Industrial Chemicals to Aquatic Organisms Using Structure-Activity Relationships, 2nd ed. ClementsR. G., Ed.; Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency: Washington, DC, 1996.
(16) AMBIT. http://AMBIT.acad.bg (accessed 1/12/2006), software for chemoinformatic data management.
(17) Tropsha, A.; Gramatica, P.; Gombar, V. K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.
(18) Hawkins, D. M.; Basak, S. C.; Mills, D. Assessing Model fit by Cross-validation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579–586.
(19) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graph. Model.* **2002**, *20*, 269–276.
(20) Kaiser, K. L. E.; Dearden, J. C.; Klein, W.; Schultz, T. W. A note of caution to users of ECOSAR. *Water Qual. Res. J. Can.* **1999**, *34*, 179–182.
(21) Hansen, O. C. *Quantitative Structure-Activity Relationships (QSAR) and Pesticides*; Pesticides Research 94; Danish Environmental Protection Agency: Denmark, 2004.
(22) Sinclair, C. J.; Boxall, A. B. A. Assessing the Ecotoxicity of Pesticide Transformation Products. *Environ. Sci. Technol.* **2003**, *37*, 4617–4625.
(23) OECD Organisation for Economic Cooperation and Development. *The Report from the Expert Group on (Quantitative) Structure-Activity Relationships [(Q)SARs] on the Principles for the Validation of (Q)SARs*; ENV/JM/TG(2004)27/REV; Paris, 2004.
(24) Moore, D. R. J.; Breton, R. L.; MacDonald, D. B. A comparison of model performance for six quantitative structure-activity relationship packages that predict acute toxicity to fish. *Environ. Toxicol. Chem.* **2003**, *22*, 1799–1809.
(25) Cariello, N. F.; Wilson, J. D.; Britt, B. H.; Wedd, D. J.; Burlinson, B.; Gombar, V. Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity. *Mutagenesis* **2002**, *17*, 321–329.

ES071430T