

QSAR models for fish toxicity: what works and what doesn't

Klaus L.E. Kaiser
TerraBase Inc.

1063 King St. W., Suite 130, Hamilton, Ontario L8S 4S3, Canada

Abstract

The recent assessment of six QSAR models for aquatic (fish) toxicity by the European Commission is considered from a practical point of view. The primary criterion here is not how well a model is defined in statistical terms, but if it is applicable and useful to "real life" problems, such as for the Domestic Substances List (Canada) of over 24,000 chemicals in use. In this writer's opinion, only broadly applicable models, which can predict the effects of chemicals without any other *a priori* knowledge, such as mode of action, etc., are of any practical use.

Introduction

The European Commission [1] recently undertook a comparative assessment of six quantitative structure-activity (QSAR) models for aquatic (fish) toxicity. The six models assessed were the (i) non-polar narcosis model (NPN) [2], (ii) the polar narcosis model (PN) [3], (iii) the global narcosis model (N) [4], (iv) the mixed mode of action model (MIXED) [5], (v) an electrotopological index model (E-State) [6], and (vi) the probabilistic neural network model TerraQSAR-FHM [7]. For the model assessments, a comparison of the experimental testing data with the estimates derived from each QSAR model was made. In total, 177 test chemicals were considered and this list of compounds, commonly referred to as the OECD Screening Information Data Set (SIDS) is given in [1], together with their experimental fish toxicity values and other information. However, only 120 of the 177 compounds have any measured toxicity values [1]. The assessment results for the QSAR-1 to QSAR-3 models were described in an earlier publication by the same authors [4]. These models consist of linear relationships with the octanol/water partition coefficient (K_{ow}) and are given below:

$$\begin{array}{ll} \text{QSAR-1, (NPN):} & pT = 0.862 \log K_{ow} + 1.330 \\ \text{QSAR-2, (PN):} & pT = 0.723 \log K_{ow} + 2.159 \\ \text{QSAR-3 (N):} & pT = 0.810 \log K_{ow} + 1.744 \end{array}$$

While fulfilling the OECD criteria for regulatory acceptance [8,9], none of these three linear models provide estimates for all SIDS compounds, as these models are only applicable to compounds acting by narcosis.

The following outlines the main steps and results of the SIDS model assessments from a practical, *i.e.* user point of view. For details of the model assessments, see Pavan and coworkers [1,4].

Method outline

1. Preliminary analysis of SIDS acute fish toxicity data.
2. Generation of molecular structure files for the SIDS chemicals (Smiles, mol files), for further calculation of both two-dimensional molecular descriptors and three-dimensional descriptors. An excel file containing chemical names, CAS numbers and SMILES for 177 chemicals was kindly provided by Eva Wedebye (DK).
3. Development of a list of literature-based models to make predictions of SIDS endpoints. The focus was on models for fish toxicity.
4. Selection of transparent and reproducible models: recovery of the training set used to develop the models and checking of the test method used to generate the model and identification of the molecular descriptors used and assessment of the transparency of the algorithm.
5. Estimation of predictive ability by internal validation techniques (cross-validation, bootstrap, response randomization).
6. Evaluation of QSAR applicability domains by making predictions of SIDS test data: checking the domain of applicability with respect to descriptor ranges and any structural rules defining the group of substances for which the models are valid.
7. Application of the models to the SIDS chemicals.
8. Evaluation of predictive performance in terms of explained variance (Q^2_{ext}) and the prediction reliability (order of magnitude between estimated and experimental data). Predictive performance was assessed for the full set of SIDS substances, and for subsets based on different hypotheses about the applicability domain.
9. Comparative analysis of the model quality.

Assessment Results

Table 1, below, is an excerpt of the “Comparative Assessment of QSAR Models for Aquatic Toxicity” [1] comparing six different QSAR models for the “Danish dataset” of 177 SIDS compounds. Table 1 provides a summary of the number of chemicals in the training set (N. Train), the number of SIDS chemicals in the training set (SIDS Train), the average predictive power calculated by leave-one-out validation (Q^2_{LOO}), the average predictive power calculated by boot-strapping validation ($Q^2_{bootstrap}$), the standard deviation error of prediction (SDEP), the coefficient of determination (R^2), the number of known SIDS chemicals for which predictions are made (N. Test, [xxx/177]), the number of unknown SIDS chemicals for which predictions are made (Unknown SIDS predictions, [xxx/177]), the total number of SIDS chemical for which predictions are made (Total SIDS predictions, [xxx/177]), and the explained variance in prediction calculated by external validation (Q^2_{ext}). The results obtained by the European Commission [1] are summarized in Table 1 (Table XI in [1]):

Table 1. Model performance comparison, see text for abbreviations.

| Model | N. Train | SIDS Train | Q^2_{LOO} | $Q^2_{bootstrap}$ | SDEP | R^2 | Test MOA | N.Test | Unknown SIDS predictions | Total SIDS predictions | Q^2_{ext} |
|-------------------|------------|------------|-------------|-------------------|------------|--------------|--------------|-----------|--------------------------|------------------------|--------------|
| NPN | 58 | 8 | 91.51 | 91.66 | 0.421 | 92.18 | NPN | 14 | 37 | 51 | 89.06 |
| | | | | | | | Mixed | 28 | 97 | 125 | 90.86 |
| PN | 86 | 5 | 89.59 | 89.64 | 0.336 | 90.07 | PN | 2 | 4 | 6 | N.A. |
| | | | | | | | Mixed | 25 | 98 | 123 | 86.66 |
| N | 144 | 13 | 87.06 | 87.11 | 0.461 | 87.55 | NPN + PN | 13 | 41 | 54 | 92.18 |
| | | | | | | | Mixed | 24 | 97 | 121 | 91.63 |
| MIXED | 114 | 9 | 75.94 | 75.83 | 0.495 | 77.57 | Mixed | 22 | 51 | 73 | 87.10 |
| E-State | 121 | 8 | 68.28 | 9.30 | 0.505 | 84.04 | Mixed | 17 | 69 | 86 | 89.43 |
| Terra QSAR | 886 | N.A | N.A | N.A | N.A | 94.56 | Mixed | 57 | 120 | 177 | 99.39 |

As is evident from Table 1, the TerraQSAR™ - FHM program delivers estimates for all (177) test compounds versus 125 compounds by the nearest competitor. As it is based on a “mixed mode” principle, *i.e.* does not require *a priori* knowledge of the mode of action, it also enjoys the largest set of chemicals in the training set. The result is convincing. It has the best performance as measured by the number of chemicals for which it makes predictions (177/177) and it has the highest explained variance in prediction calculated by external validation ($Q^2_{ext} = 99.39$).

Applicability Domain

In their conclusions [1], the authors remark that “the applicability domain of the TerraQSAR-FHM model was not estimable, since the identification of the training set chemicals is missing, together with the descriptors used to train the net”. In fact, the entire list of 886 fathead minnow toxicity data used in the training of the TerraQSAR-FHM model was provided and they can be found in the Table IX of the report [1], however without compound identifiers. In addition, Table X of the report [1] gives the TerraQSAR-FHM estimates for each SIDS compounds.

The “applicability domain” of a model is an issue of importance to regulatory agencies. A detailed analysis of this with respect to the SIDS chemicals is found in the preliminary assessment report [4]. It is worth repeating some of the important findings here. In essence, a chemical falls within a given applicability domain if its “mode of action” (MOA) is covered by the model. Initially, the 177 SIDS substances were classified by four distinct methods, including one human expert. Their classifications resulted in 4, 17, 17, and 20 different MOA classes, respectively. Following that, a consensus based on the majority principle (Consensus-1), resulted in a classification of all SIDS chemicals into 10 MOA classes, as given in Table 2, where 62 of the 177 chemicals are deemed to have an unknown MOA.

Table 2. Consensus-1 MOA classification of 177 SIDS chemicals.

| Consensus MOA | Description | Number of chemicals |
|----------------------|---|----------------------------|
| AChE | Acetylcholinesterase (AChE) inhibition | 1 |
| CNS | Central nervous system seizure action | 2 |
| EN | Ester narcosis | 5 |
| MTA | Michael-type addition | 16 |
| NPN | Non polar narcosis | 75 |
| PE | Electrophile and proelectrophile reactivity | 2 |
| PN | Polar narcosis | 12 |
| SB | Schiff-base formation | 1 |
| SN2 | SN2 reaction | 1 |
| UNK | Unknown mode of action | 62 |

A second consensus classification (Consensus-2) resulted in a further reduction of the number of MOAs to only five. Its breakdown is given in Table 3.

Table 3. Consensus-2 MOA classification of 177 SIDS chemicals.

| CONS2- MOA | Description | N. Chemicals |
|-------------------|--------------------------|---------------------|
| N | Narcosis | 97 |
| N* | Narcosis modeled by LogD | 18 |
| R | Reactive | 44 |
| S | Specifically acting | 3 |
| UNK | Unknown mode of action | 15 |

It is obvious that the issue of the applicability domain is critical to the use of any QSAR for regulatory purposes. What is much less obvious though, is information on how to determine whether or not a compound falls inside or outside the domain of any of the linear models and can be modeled by it. In fact, many modern widely-used substances, including textile processing substances, pesticidal or herbicidal compounds, and most modern drugs would be found outside the applicability domain for any of the linear models QSAR-1 to QSAR-5. Therefore, the even more important question becomes what to do with a substance that falls outside the domain of the chosen model(s). This conundrum is central to the practical use of any model and it will not be solved by ignoring it. This problem has been pointed out previously [9,10], and is worth repeating here. While there are some mathematical concepts to determine if a compound falls within or outside the applicability domain of a model [11] they are of little practical use for most potential users of models requiring a MOA predetermination. Furthermore, it is worth noting here that none of the linear models (QSAR-1 to QSAR-5) can provide predictions for all SIDS substances, despite their supposed applicability to unknown MOA compounds. The only model which provides estimates for all SIDS substances and which is not limited by any MOA, is the TerraQSAR-FHM (QSAR-6) model.

Complexity of substances / Representativity

An issue entirely overlooked in the assessment document [1] and its precursor [4] is the problem of SIDS substance complexity and representativity, *i.e.* their relevance to the world of substances for which environmental and human health effect assessments may be required. In this regard, even a cursory inspection of the list will reveal that it contains a preponderance of relatively small, mostly one-ring molecules. In fact, the molecular weight (MW) range of the SIDS substances ranges from MW 30 to 959, but only eight out of the 177 substances have a MW above 400 and three out of these eight substances contain four or more of the (relatively heavy) atom bromine. Furthermore, most or all of the compounds with a MW >400 are outliers to one or more of the models QSAR-1 to QSAR-5 (see, for example, Table XIV in [1]). Therefore, it would appear that the SIDS list of substances is more representative of the world of chemicals which can be modeled by simple linear QSAR relationships rather than of what is found in the environment or what is really of concern. For example, the most recent “Domestic Substances List (DSL)” of chemicals in production/use in Canada, covers 24,603 compounds [12]. It would be most revealing if the OECD would undertake an assessment of the DSL substances, or of another similarly representative list of substances in use, as to their applicability domain for any of the tested models.

On the Probabilistic Neural Network

Neural networks (NNs) come in many different colors. While every model needs training, common NN models do not have a unique solution or an optimum which can be derived at without some trial and error. This is not the case for the Probabilistic Neural Network (PNN), used in TerraQSAR programs. Therefore, the notion that “PNNs do not provide the same numerical solution to a problem when repeated” is utterly mistaken. The PNN optimization process provides exactly the same answer to any given set of data, independent on any other circumstances. During the model training process, there is a complete convergence of the model variables to result in a unique, fully repeatable model. The mathematical concept of the TerraQSAR system is also described in the EU assessment document [1].

Summary

Only the TerraQSAR-FHM program is applicable to all types of substances, irrespective of their mode of action, use, presence or absence of certain basic structures, presence or absence of certain types of substituents, and so forth. Only the TerraQSAR-FHM model does not require the use of any other measured or computed values for its use. This can be summarized in the following statements:

- **The TerraQSAR-FHM program has the highest coefficient of determination ($R^2 = 94.56$).**
- **The TerraQSAR-FHM program has the highest explained variance in prediction calculated by external validation ($Q^2_{ext} = 99.39$).**
- **The TerraQSAR-FHM program has the lowest standard deviation error of prediction by external validation ($SDEP_{ext} = 0.116$).**
- **The TerraQSAR-FHM program does not require prior knowledge of the mode of action of any substance.**
- **The TerraQSAR-FHM program is the only model capable of estimating values for all 177 SIDS testing substances.**
- **The TerraQSAR-FHM program is not limited to certain structures.**
- **The TerraQSAR-FHM program does not require the knowledge of another parameter, such as the octanol/water partition coefficient ($\log K_{ow}$) or the lowest unoccupied molecular orbital energy (E_{LUMO}).**
- **The TerraQSAR-FHM program has the broadest applicability domain.**
- **The TerraQSAR-FHM program has the largest training set.**

References

- [1] Pavan M, Worth AP, Netzeva TI, 2005. Comparative assessment of QSAR models for aquatic toxicity. European Commission, EUR 21750 EN, http://www.terrabase-inc.com/EUR_21750_EN.pdf.
- [2] Veith GD, Call DJ, Brooke LT, 1983. Structure-toxicity relationship for the fathead minnow, *Pimephales promelas*: Narcotic industrial chemicals. *J. Fish. Aquat. Sci.* 40: 743-748.
- [3] Verhaar HJM, Mulder W, Hermens JLM, 1995. QSARs for ecotoxicity. An overview of structure-activity relationships for environmental endpoints, Part I: general outline and procedure. In: Report on QSAR for predicting fate and effect of chemicals in the environment, Hermens JLM (ed.),
- [4] Pavan M, Worth AP, Netzeva TI, 2005. Preliminary analysis of an aquatic toxicity dataset and assessment of QSAR models for narcosis. European Commission, EUR 21749 EN.
- [5] Veith GD, Mekenyan OG, 1993. A QSAR approach for estimating the aquatic toxicity of soft electrophiles. *QSAR* 12: 349-356
- [6] Huuskonen J, 2003. QSAR modeling with the electrotopological state indices: predicting the toxicity of organic chemicals. *Chemosphere* 50: 949-953.
- [7] TerraBase Inc., 2003. TerraQSAR-FHM, fathead minnow 96-hr LC50 estimation software, vs. 1.1, <http://www.terrabase-inc.com>.
- [8] Jaworska JS, Comber M, van Leeuwen CJ, 2003. Summary of a workshop on regulatory acceptance of (Q)SAR for human health and environmental endpoints. *Environ. Health Persp.* 111: 1358-1360.
- [9] Kaiser KLE, 2004. Correspondence re validation of (Q)SAR models. *Environ. Health Persp.* 112: A86.
- [10] Jaworska JS, Cronin M, Eriksson L, Comber M, 2004. Validation of (Q)SAR models: Jaworska et al.'s Reply. *Environ. Health Persp.* 112: A86-A87.
- [11] Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P, 2002. Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs. *Environ. Health Persp.* 111: 1361-1375.
- [12] Environment Canada, 2005. Domestic Substances List, http://www.ec.gc.ca/substances/nsb/HTML/dsl_e.HTM, last updated: December 15, 2005.

TerraBase Inc. Statement

TerraBase Inc. founded in 1996, is a private Canadian corporation. Its main activities are the development of specialized databases and computational programs for use with personal computers under the Windows platform. For a detailed list of databases and programs, see the web site at <http://www.terrabase-inc.com>. TerraBase Inc. has not received any grants from any government or non-government agency to develop any of its databases or computational programs. Therefore, TerraBase Inc. cannot provide such resources free to scientists or managers in the field. However, in order to foster cooperation and development in the field, we offer free time-limited evaluation copies of certain company products, subject to certain conditions. Details are available on request.