



Quantitative structure–activity relationship methods for the prediction of the toxicity of pollutants

Raghunath Satpathy¹

Received: 15 June 2018 / Accepted: 13 July 2018 / Published online: 20 July 2018
© Springer Nature Switzerland AG 2018

Abstract

Continuous flow of toxic and persistent compounds to the environment is a global health issue. However, assessing the toxic effects of compounds is a difficult task, because some compounds may possess a combined effect during exposure. Moreover, toxicity evaluation by animal testing is long and costly. Alternatively, modeling of quantitative structure–activity relationships (QSAR) can be used to predict the acute toxicity of molecules. Properties of toxic compounds are computed and correlated using softwares and databases. Recently, this method has found potential applications for the risk assessment of several untested, toxic chemicals. This review focuses on quantitative structure–activity relationship modeling methods for the analysis of toxic compounds. Computational tool and databases are presented.

Keywords Biodegradation · Computational tools · Database · Descriptors · Environmental toxicity · Quantitative structure–activity relationship modeling · Validation methods · Risk assessment · Toxic compounds

Introduction

The diversity in the chemistry of the several environmental toxic compounds is one of the important considerations to study their toxicity (Raies et al. 2016; Chen et al. 2013; Basant et al. 2016). Some of these toxic substances are extensively used in day-to-day life as well as many industrial processes as solvents, as precursors for another compound such as dyes, plasticizers, pharmaceuticals, pesticides. Every year, millions of tons of chemicals are released into the environment (Settivari et al. 2015; Awaleh and Soubaneh 2014; Perocco et al. 1983). In order to protect human health and the environment, risk assessments should be conducted for all toxic chemicals, especially for those already found in the environment. Several tons of these toxic compounds produced from industry and their accumulation in the environmental habitats represent a global ecological danger. Also, many of these substances have also been reported to be persistent in nature (Halling-Sørensen et al. 1998; Battaglin and Kolpin 2009; Li et al. 2011; Luo et al. 2015; Satpathy et al. 2015a, b). Toxicity tests are usually used to predict

the harmful effect of these compounds by considering the animal models (Raies et al. 2016). However, this is a time-consuming process; also it is not economically feasible to conduct toxicity tests on each chemical individually and also difficult to study the combinatorial effects of the toxic substances on organisms. Therefore, in the field of environmental toxicology and environmental chemistry, quantitative structure–activity relationships (QSARs) are currently used to predict toxicity for a number of toxic chemicals (Doke and Dhawale 2015; Raunio 2011; Boobis et al. 2002).

In the QSAR analysis, the relationship between physico-chemical properties and toxicological effects can be represented as mathematical models, termed quantitative structure–activity relationships models (Fig. 1).

Furthermore, by a proper selection of the training compounds, QSARs can cover large series of similar compounds and thus provide the basis for an efficient and comparative screening assessment of toxic substances (Tunkel et al. 2005; Hulzebos et al. 2005; Satpathy et al. 2015a, b). This article is an abridged version of the chapter published by Satpathy (2018), chapter 10 in the series Environmental Chemistry for a Sustainable World (<http://www.springer.com/series/11480>).

✉ Raghunath Satpathy
rnsatpathy@gmail.com

¹ Department of Biotechnology, MITS Engineering College, Rayagada, Odisha 765017, India

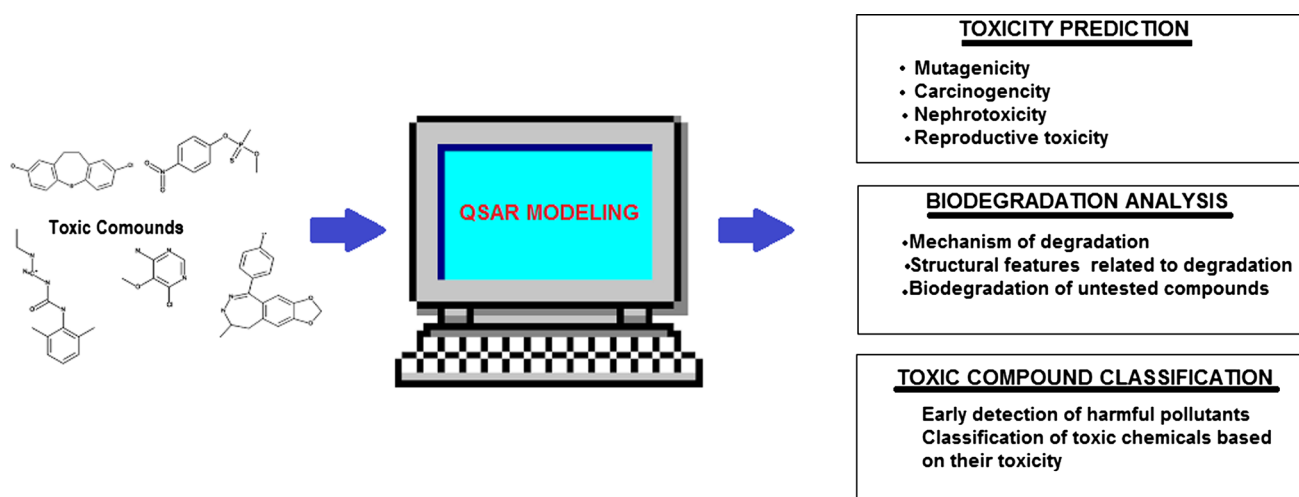


Fig. 1 Overview of quantitative structure–activity relationships (QSAR)-based analysis on toxic compounds

Quantitative structure–activity relationship analytical methods

The QSAR method begins by taking a suitable group of compounds known as training dataset. The required molecular features of a suitable training set are data maybe taken from the literature, various databases and sometimes computed by the software tools. The parameter to be studied and correlated with the molecular features may be either a physicochemical property. After the proper selection of training set, the actual QSAR modeling starts. The modeling procedure is to statistically correlate the molecular feature along with the biological/environmental properties of consideration (Fig. 2). Finally, the measure of goodness-of-fit of the QSAR models (validation) is performed by determination coefficient (R^2) or adjusted determination coefficient (R_{adj}^2), summary square error (SSE) (reflects the deviation of predicted values from observed values), root mean square error (RMSE), used to indicate the precision of prediction and F -test (a variance test method of the overall significance level).

Computational tools

QSAR-based toxicology research utilizes a wide variety of computational tools (Pirhadi et al. 2016; Liao et al. 2011), such as databases for storing data about chemicals, their toxicity and chemical properties, software for generating molecular descriptors and simulation tools to generate the QSAR equation and validation (Tables 1

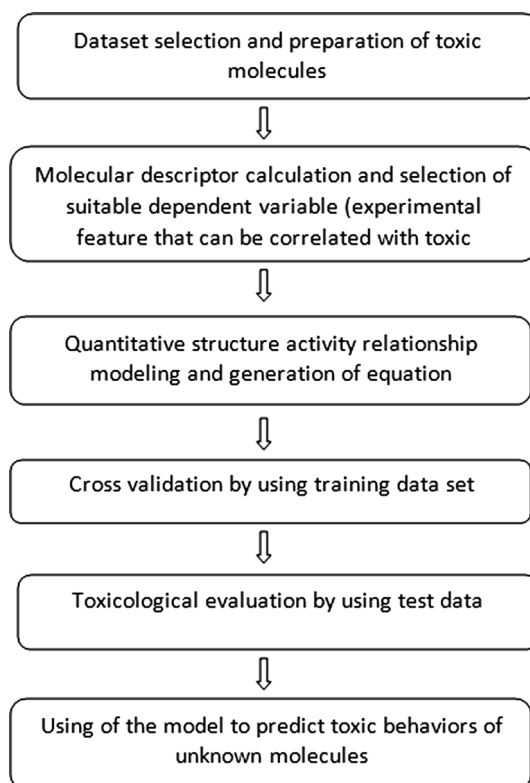


Fig. 2 Steps of quantitative structure–activity relationship modeling methods for prediction of toxic compounds

and 2). However, the good predictive models for toxicity parameters depend crucially on selecting the right mathematical approach and the right molecular descriptors for the particular toxicity endpoint.

Table 1 Software tools and servers details for calculating variables in toxicity prediction in case of compounds

| Sl. no. | Software | Availability | Application |
|---------|---|---|--|
| 1 | ADMET Predictor | http://www.simulations-plus.com/ | Quantitative prediction of estrogen receptor toxicity in rats |
| 2 | ACD ToxSuite (ToxBboxes) | http://www.acdlabs.com/products/admet/tox/ | Prediction of endoplasmic reticulum (ER) binding affinity prediction |
| 3 | CAESAR | http://www.lhasalimited.org/ | Two classification models for developmental toxicity |
| 4 | Derek | https://www.lhasalimited.org/ | Different levels of classification models (based on developmental toxicity) |
| 5 | Leadscope | http://www.leadscope.com/ | Classification models for developmental toxicity in the rodent fetus |
| 6 | MolCode toolbox | http://molcode.com/ | Quantitative prediction of rat ER binding affinity and AhR binding affinity |
| 7 | MultiCASE | http://www.multicase.com/ | Classification models for developmental toxicity associated with drugs |
| 8 | OSIRIS property explorer | http://www.organic-chemistry.org/prog/peo/ | Predicts mutagenicity, tumorigenicity, irritating effects and reproductive effects |
| 9 | PASS | http://ibmc.p450.ru/PASS/ | Classification models giving the probability of reproductive toxic effects. |
| 10 | T.E.S.T.: the toxicity estimation software tool | http://oasis-lmc.org/ | Developmental toxicity estimation |
| 11 | TOPKAT | http://www.accelrys.com | Classification model for developmental toxicity of pesticides, industrial chemical |
| 12 | Toxbboxespharma algorithms | http://pharma-algorithms.com/tox_boxes.htm | A classification model for the prediction of ER binding |
| 13 | VirtualToxLa | http://www.biograf.ch | Classification model for endocrine disrupting potential |
| 14 | HAZARD EXPERT | http://www.compudrug.com/hazardexpertpro | Human carcinogenicity and genotoxicity prediction |
| 15 | Toxline | https://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE | Human neurotoxicity prediction |
| 16 | BCABAF | https://www.epa.gov/tsca-screening-tools/epi-suite-tm-estimation-program-interface | Prediction of bio-concentration of toxic substances |
| 17 | PCKOCWIN | cpas.mtu.edu/cencitt/oppt/tsld019.htm | Prediction of soil sorption with the toxic chemicals |
| 18 | BIOWIN | envirosim.com/products/biowin | Prediction of biodegradability of toxic substances |
| 19 | KOWWIN | https://www.epa.gov/tsca-screening-tools/epi-suite-tm-estimation-program-interface | Estimates octanol–water partition coefficient of toxic chemicals |
| 20 | AMBIT | http://ambit.sourceforge.net/intro.html | Chemical structure search, and predictive model can be obtained |
| 21 | PreADMET | https://preadmet.bmdrc.kr/ | Prediction of ADME data and toxicity prediction |

Applications for the prediction of toxic compounds

In the field of environmental toxicology, quantitative structure–activity relationships (QSARs) methods have been used as robust tools for predicting the toxicity of chemicals whenever no or little amount of data are available. The QSAR analysis has been used to predict a specific chemical class that behaves in a toxicologically similar manner (Pavan et al. 2008; Chen et al. 2004; Li et al. 2006). In the case of human being, the prediction of the acute toxicity of a compound is an important task in order to justify the in the regulatory assessment of particular compounds. However, mostly this information is obtained from the animal studies that is

related to animal ethics and cost considerations. Therefore, the method alternative to animal experiments is preferable (Lapenna et al. 2010; Raies and Vladimir 2016). As a powerful technique, the QSAR methods have been widely applied in toxicology by many researchers. Cronin et al. (2003) emphasized that the application of QSAR techniques to predict ecologic effects and environmental fate of chemicals for facilitating the regulatory agencies and authorities will find them to be acceptable alternatives to chemical testing. Roberts (1991) studied the acute lethal toxicity data for a range of anionic and nonionic surfactants by QSAR modeling to predict by calculated $\log P$ (octanol/water) values. The toxicity prediction about active ingredients in pharmaceutical products and their importance and mechanism have been

Table 2 Description of nine major types of databases to compute the descriptor calculation in case of toxic compounds

| Sl. no. | Databases | Availability | Application |
|---------|--|---|---|
| 1 | Toxicology Data Network (Toxnet) Developmental and Reproductive Toxicology Database (DART) | http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?DARTETIC | References to the developmental and reproductive toxicology |
| 2 | Endocrine Disruptor Knowledge Base (EDKB) | http://www.fda.gov/ScienceResearch/BioinformaticsTools/EndocrineDisruptorKnowledgebase | Contains in vitro and in vivo experimental data for toxic chemicals |
| 3 | Endocrine Active Substances information system (EASIS) | https://eurl-ecvam.jrc.ec.europa.eu/databases/eas_database | Provides information for chemical structure, toxicity |
| 4 | NureXbase | http://www.nursa.org | Information on chemical structure, properties and mechanism of endocrine action |
| 5 | OECD (Q)SAR Toolbox | http://www.oecd.org/env/ehs/riskassessment/theoecdqsartoolbox.htm | Contains reproductive toxicity data |
| 6 | Acute Toxicity Database | https://www.cerc.usgs.gov/data/acute/acute.html | Database of aquatic acute toxicity test results for chemicals |
| 7 | Distributed Structure-Searchable Toxicity (DSSTox) Database | https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dssto-database | Resources for supporting improved predictive toxicology |
| 8 | TerraBase | http://www.terrabase-inc.com/ | Quick search of compounds with specific biological effects |
| 9 | EXTOXNET | http://extoxnet.orst.edu/pips/ghindex.html | Pesticide information including experimental toxicity value |

reviewed by Kruhlak et al. (2007). The successful prediction of genotoxicity of the compounds like 2-amino-9H-pyrido [2, 3-b] indole (A α C) and 2-aminoacetophenone (2-AAP) by QSAR has been studied by Worth et al. (2013). Comparative evaluation and prediction of mammalian acute toxicity, by considering lethal dose (LD₅₀) as a dependent variable, have been studied by Gonella Diaza et al. (2015) in a dataset of 7417 toxic chemical compounds. Also, to predict *no observed effect level* (NOEL), developmental and reproductive toxicities have been successfully predicted by Hisaki et al. (2015) from a total of 892 toxic chemicals.

Advantages of quantitative structure–activity relationship-based studies

The QSAR-based study of toxic compounds having advantages summarized below:

- Prediction of the environmental fate of the toxic compounds such as bio-concentration, soil sorption, biodegradation and so on.
- Since the prediction methods are computer based, they provide a rapid assessment of toxicity of these compounds.
- Further, they have the capability of reducing, and even replacing, animal tests for toxicological assessment of the pollutant compounds.

- Industrial users can apply these models to screen new compounds and to assist in the process of designing out toxic features of new chemical entities. Ultimately, this information can be used by the regulatory agencies to impose the regulation of new and existing chemical compounds.
- Prediction of toxicity can be applied to environmental risk assessments for common pollutants.

Challenges

Although QSAR-based methods have enormous potential for analyzing the toxic profile of compounds, certain challenges should be overcome.

- Problems in Biological dataset

The foremost important thing in QSAR analysis is the data. One of the limitations is that very little amount of biochemical data are available in terms of mechanism of toxic action. Therefore, for validation purpose, it faces a problem, thereby causing inconsistency in prediction.

- Better selection of dependent and independent variables and domain applicability

Predictions of toxicity should be made within the domain of applicability of an appropriately validated QSAR.

Appropriately choosing descriptor variables and dependent variable leads to a good prediction of the model. Also, the number of independent variables in the case of specific chemicals is important for model generation and prediction.

- Variability in toxicity action of compounds

Usually, it is expected that similar toxic chemicals possess similar mechanism of action in toxicity, but any of the compounds that do not possess the same mechanism of action will show up as outliers; that is, they will not be well modeled by the QSAR.

Conclusion

The development of models for quantitative structure–activity relationships (QSARs) and its application for toxicological effects are of great importance. A million number of chemicals from industry and other sources are released and exposed to the environment, but little is known about the toxicity of them. It would be difficult, expensive and time-consuming to test all such chemicals for toxicity. Due to this, QSAR methods have been used to interpret the toxicity and also many regulatory agencies are beginning to accept toxicities predicted by QSAR. The basic principles behind QSAR analysis are a prediction of biological activities from chemical structures that are closely related. However, while performing the analysis many of the aspects are to be considered such as avoiding the false correlation of the data and perfect experimental design. Therefore, validation process for the experiment must be observed carefully.

References

- Awaleh MO, Soubaneh YD (2014) Waste water treatment in chemical industries: the concept and current technologies. *Hydrol Current Res* 5(1):1–12. <https://doi.org/10.4172/2157-7587.1000164>
- Basant N, Gupta S, Singh KP (2016) *In silico* prediction of the developmental toxicity of diverse organic chemicals in rodents for regulatory purposes. *Toxicol Res* 5(3):773–787. <https://doi.org/10.1039/C5TX00493D>
- Battaglin WA, Kolpin DW (2009) Contaminants of emerging concern: introduction to a featured collection I. *J Am Water Resour Assoc* 45(1):1–3. <https://doi.org/10.1111/j.1752-1688.2008.00299.x>
- Boobis A, Gundert-Remy U, Kremers P, Macheras P, Pelkonen O (2002) *In silico* prediction of ADME and pharmacokinetics: report of an expert meeting organised by COST B15. *Eur J Pharm Sci* 17(4):183–193. [https://doi.org/10.1016/S0928-0987\(02\)00185-9](https://doi.org/10.1016/S0928-0987(02)00185-9)
- Chen J, Harner T, Ding G, Quan X, Schramm KW, Ketrup A (2004) Universal predictive models on octanol–air partition coefficients at different temperatures for persistent organic pollutants. *Environ Toxicol Chem* 23(10):2309–2317. <https://doi.org/10.1897/03-341>
- Chen L, Lu J, Zhang J, Feng KR, Zheng MY, Cai YD (2013) Predicting chemical toxicity effects based on chemical–chemical interactions. *PLoS ONE* 8(2):e56517. <https://doi.org/10.1371/journal.pone.0056517>
- Cronin MT, Jaworska JS, Walker JD, Comber MH, Watts CD, Worth AP (2003) Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environ Health Perspect* 111(10):1391. <https://doi.org/10.1289/ehp.5760>
- Doke SK, Dhawale SC (2015) Alternatives to animal testing: a review. *Saudi Pharm. J* 23(3):223–229. <https://doi.org/10.1016/j.jsps.2013.11.002>
- Gonella Diaza R, Manganelli S, Esposito A, Roncaglioni A, Manganaro A, Benfenati E (2015) Comparison of *in silico* tools for evaluating rat oral acute toxicity. *SAR QSAR Environ Res* 26(1):1–27. <https://doi.org/10.1080/1062936X.2014.977819>
- Halling-Sørensen B, Nielsen SN, Lanzky PF, Ingerslev F, Lützhøft HH, Jørgensen SE (1998) Occurrence, fate and effects of pharmaceutical substances in the environment—a review. *Chemosphere* 36(2):357–393. [https://doi.org/10.1016/S0045-6535\(97\)00354-8](https://doi.org/10.1016/S0045-6535(97)00354-8)
- Hisaki T, Née Kaneko MA, Yamaguchi M, Sasa H, Kouzuki H (2015) Development of QSAR models using artificial neural network analysis for risk assessment of repeated-dose, reproductive, and developmental toxicities of cosmetic ingredients. *J Toxicol Sci* 40(2):163–180. <https://doi.org/10.2131/jts.40.163>
- Hulzebos E, Walker JD, Gerner I, Schlegel K (2005) Use of structural alerts to develop rules for identifying chemical substances with skin irritation or skin corrosion potential. *QSAR Comb Sci* 24(3):332–342. <https://doi.org/10.1002/qsar.200430905>
- Kruhlik NL, Contrera JF, Benz RD, Matthews EJ (2007) Progress in QSAR toxicity screening of pharmaceutical impurities and other FDA regulated products. *Adv Drug Deliv Rev* 59(1):43–55. <https://doi.org/10.1016/j.addr.2006.10.008>
- Lapenna S, Gatnik MF, Worth AP (2010) Review of QSA R models and software tools for predicting acute and chronic systemic toxicity. *Publ Off Eur Union, Luxembourg*. <https://doi.org/10.2788/60766>
- Li X, Chen J, Zhang L, Qiao X, Huang L (2006) The fragment constant method for predicting octanol–air partition coefficients of persistent organic pollutants at different temperatures. *J Phys Chem Ref Data* 35(3):1365–1384. <https://doi.org/10.1063/1.2203356>
- Li J, Lu Y, Shi Y, Wang T, Wang G, Luo W, Yan F (2011) Environmental pollution by persistent toxic substances and health risk in an industrial area of China. *J Environ Sci* 23(8):1359–1367. [https://doi.org/10.1016/S1001-0742\(10\)60554-2](https://doi.org/10.1016/S1001-0742(10)60554-2)
- Liao C, Sitzmann M, Pugliese A, Nicklaus MC (2011) Software and resources for computational medicinal chemistry. *Future Med Chem* 3(8):1057–1085. <https://doi.org/10.4155/fmc.11.63>
- Luo J, Hu J, Wei X, Fu L, Li L (2015) Dehalogenation of persistent halogenated organic compounds: a review of computational studies and quantitative structure–property relationships. *Chemosphere* 131:17–33. <https://doi.org/10.1016/j.chemosphere.2015.02.013>
- Pavan M, Netzeva TI, Worth AP (2008) Review of literature-based quantitative structure–activity relationship models for bioconcentration. *QSAR Comb Sci* 27(1):21–31. <https://doi.org/10.1002/qsar.200710102>
- Perocco P, Bolognesi S, Alberghini W (1983) Toxic activity of seventeen industrial solvents and halogenated compounds on human lymphocytes cultured in vitro. *Toxicol Lett* 16(1):69–75. [https://doi.org/10.1016/0378-4274\(83\)90012-7](https://doi.org/10.1016/0378-4274(83)90012-7)
- Pirhadi S, Sunseri J, Koes DR (2016) Open source molecular modeling. *J Mol Graphics Modell* 69:127–143. <https://doi.org/10.1016/j.jmgm.2016.07.008>
- Raies AB, Vladimir BB (2016) *In silico* toxicology: computational methods for the prediction of chemical toxicity. *Wiley Interdiscip Rev Comput Mol Sci* 6(2):147–172. <https://doi.org/10.1002/wcms.1240>
- Raunio H (2011) *In silico* toxicology—non-testing methods. *Front Pharmacol* 2:33. <https://doi.org/10.3389/fphar.2011.00033>

- Roberts DW (1991) QSAR issues in aquatic toxicity of surfactants. *Sci Total Environ* 109:557–568. [https://doi.org/10.1016/0048-9697\(91\)90209-W](https://doi.org/10.1016/0048-9697(91)90209-W)
- Satpathy R (2018) Quantitative structure–activity modelling of toxic compounds. In: *Nanotechnology, food security and water treatment*. Springer, Cham, pp 313–331. https://doi.org/10.1007/978-3-319-70166-0_10
- Satpathy R, Konkimalla VB, Ratha J (2015a) Application of bioinformatics tools and databases in microbial dehalogenation research: a review. *Appl Biochem Microbiol* 51(1):11–20. <https://doi.org/10.1134/S0003683815010147>
- Satpathy R, Konkimalla VB, Ratha J (2015b) *In silico* based prediction and correlation of dehalogenase enzyme activity of some haloalkane compounds by quantitative structure biodegradation relationship (QSBR) analysis. *Curr. Environ. Eng.* 2(2):122–126. <https://doi.org/10.2174/221271780202160122192808>
- Settivari RS, Ball N, Murphy L, Rasoulpour R, Boverhof DR, Carney EW (2015) Predicting the future: opportunities and challenges for the chemical industry to apply 21st-century toxicity testing. *J Am Assoc Lab Anim Sci* 54(2):214–223
- Tunkel J, Mayo K, Austin C, Hickerson A, Howard P (2005) Practical considerations on the use of predictive models for regulatory purposes. *Environ Sci Technol* 39(7):2188–2199. <https://doi.org/10.1021/es049220t>
- Worth AP, Lapenna S, Serafimova R (2013) QSAR and metabolic assessment tools in the assessment of genotoxicity. *Comput Toxic II*:125–162. https://doi.org/10.1007/978-1-62703-059-5_6