Full length article

Predicting the acute neurotoxicity of diverse organic solvents using probabilistic neural networks based QSTR modeling approaches

Nikita Basant[a], Shikha Gupta[b], Kunwar P. Singh[a,*]

[a] ETRC, Gomtinagar, Lucknow 226010, India
[b] Environmental Chemistry Division, CSIR-Indian Institute of Toxicology Research, Post Box 80, Mahatma Gandhi Marg, Lucknow 226001, India

ABSTRACT

Organic solvents are widely used chemicals and the neurotoxic properties of some are well established. In this study, we established nonlinear qualitative and quantitative structure-toxicity relationship (STR) models for predicting neurotoxic classes and neurotoxicity of structurally diverse solvents in rodent test species following OECD guideline principles for model development. Probabilistic neural network (PNN) based qualitative and generalized regression neural network (GRNN) based quantitative STR models were constructed using neurotoxicity data from rat and mouse studies. Further, interspecies correlation based quantitative activity–activity relationship (QAAR) and global QSTR models were also developed using the combined data set of both rodent species for predicting the neurotoxicity of solvents. The constructed models were validated through deriving several statistical coefficients for the test data and the prediction and generalization abilities of these models were evaluated. The qualitative STR models (rat and mouse) yielded classification accuracies of 92.86% in the test data sets, whereas, the quantitative STRs yielded correlation (R^2) of >0.93 between the measured and model predicted toxicity values in both the test data (rat and mouse). The prediction accuracies of the QAAR (R^2 0.859) and global STR (R^2 0.945) models were comparable to those of the independent local STR models. The results suggest the ability of the developed QSTR models to reliably predict binary neurotoxicity classes and the endpoint neurotoxicities of the structurally diverse organic solvents.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Organic solvents are widely used in various applications including emulsion and micro-emulsion formulation, shoe making, degreasing, detergents, cosmetics, paint, metal processing, auto manufacturing, aeronautical maintenance and manufacturing, and pharmaceutical industries. Moreover, solvents may be used in liquid-liquid extraction and absorption processes, as a reaction medium and as a carrier, to deliver chemical compounds in solutions in the required amounts (Gani et al., 2005; Al-Malah, 2012). Many organic solvents are low molecular weight compounds and are volatile, thus transferring a fraction of their volume to the atmospheric environment at room temperature. Inhalation of solvent vapors is the most frequent type of occupational exposure (Dick, 2006). The ability of various solvents to evoke acute neurotoxic symptoms and signs is one crucial parameter for the assessment of the hazard of the solvents for adverse human health effects. This warrants assessing of risk of solvents in a systematic manner. Although, experimental test protocols for assessing the neurotoxicity of solvents in rodents have been developed (OECD, 1997), these are tedious and time and resource intensive. On the other hand, computational toxicology continues to be an attractive, viable approach to reduce the amount of effort and cost of experimental toxicity assessment (Chandler et al., 2011) and provides a method for the early evaluation in the development of new solvents (Cronin et al., 2003; Jaworska et al., 2003). The European Union (EU) regulation "Registration, Evaluation, Authorization and Restriction of Chemicals (REACH, 2015) advocates the use of non-animal testing methods and in particular quantitative structure-toxicity/activity relationship (QSTR/QSAR) approaches. The OECD has provided a set of guidelines for development of QSARs (OECD, 2007). A qualitative QSAR model may be useful in classifying solvents into relative neurotoxicity classes (high or low) and quantitative QSAR is expected to be a useful tool in predicting the neurotoxicity potential of chemicals. A few attempts have been made to develop QSAR models for the neurotoxicity of solvents in rodents (Cronin, 1996; Estrada et al., 2001). However, both of these QSAR studies, based on linear modeling methods, reported low

* Corresponding author.
E-mail addresses: kpsingh_52@yahoo.com, kunwarpsingh@gmail.com (K.P. Singh).

prediction accuracies. Moreover, none of the studies attempted to develop qualitative QSARs. Poor performance of the QSAR model may be due to the selection of inappropriate modeling method or irrelevant descriptors. Experimental toxicity data generally have nonlinear structure and linear methods failing to capture nonlinear dependence. Further, interspecies quantitative activity–activity relationships (QAARs) (Cronin, 2010; Cassani et al., 2013; Furuhama et al., 2015), which extrapolate data for one toxicity endpoint to those for another toxicity endpoint, can be used to determine the species-specific toxicity of a chemical. When the toxicity values of defined chemicals for one endpoint correlate well with the values for another endpoint, the chemicals can be expected to have similar modes of action with respect to both endpoints.

Probability density function (PDF) based neural networks, such as probabilistic neural networks (PNNs), and generalized regression neural networks (GRNNs) capable of capturing the nonlinearities in the data have successfully been used in various qualitative (classification) and quantitative (regression) QSAR studies (Mosier and Jurs, 2002; Panaye et al., 2006; Singh et al., 2013, 2014). These methods learn quickly and produce reproducible outputs without any risk for a local minimum of the error surface (Walzack and Massart, 2000).

In this study, the PNN and GRNN based QSTR models were established for the qualitative (neurotoxicity classes) and quantitative neurotoxicity predictions of structurally diverse organic solvents in rodents (Cronin, 1996) following the OECD guidelines for QSAR validation. The predictive and generalization abilities of the proposed QSTR models constructed here were evaluated using several statistical criteria. The external predictive power of the QSTR model was evaluated using the OECD recommended external validation tests. Moreover, the possibility finding interspecies correlations (ISC) for the experimental data for rat and mouse has been investigated in order to derive a quantitative activity–activity relationship (QAAR) model able to predict rat neurotoxicity from the experimental data measured in mouse.

2. Materials and methods

2.1. Datasets

The rodent neurotoxicity data (pEC_{30} μM) of 47 organic solvents were collected from the literature (Cronin, 1996). This database contained experimental values for the neurotoxicities of organic solvents in rats and mice. A detailed methodology for the experimental measurements of the neurotoxicities of solvents is provided elsewhere (Frantik et al., 1994). In brief, the experimental values refer to a whole body exposure for 2 h in mouse and 4 h in rats. Inhibition of propagation and maintenance of the electrically evoked seizure discharge was used as a criterion of the acute neurotropic effect. Out of a range of concentrations of solvents, an effective concentration amounting to 30% of the maximum possible effect (EC_{30}) was reported. The selected solvent database includes aromatic and aliphatic hydrocarbons, chlorinated hydrocarbons, alcohols, ketones, and acetates. The neurotoxicity values (pEC_{30}, μM) of the solvents in rat and mouse vary between -2.94 and -0.57; -2.98 and -0.82, respectively (Table S1, Supplementary material).

2.2. Molecular descriptors and data processing

For calculating the descriptors, the SMILES (simplified molecular input line entry system) codes of the solvent molecules were obtained using Chemspider (2015). The SMILES codes were then used for the geometry optimization of the molecules using PM7 semi empirical method (ChemMop, 2015). PM7 is a parameterized

Hartree-Fock method, which is able to capture any specific chemical interaction (Stewart, 2013). The optimized molecular structures were transferred to Chemopy (2015) for the descriptor calculation. A total of 1135 molecular descriptors were calculated for each chemical that includes 1D, 2D (constitutional, connectivity, Basak, topology, Kappa, Burden, E-state, autocorrelations, molecular property, charge, MOE-type) and 3D (geometrical, charged partial surface area, Randic molecular profiles from the geometrical matrix, MoRSE) descriptors. To reduce redundant and useless information, descriptors with constant and near constant values (variance < 0.5) were removed. Finally, 262 descriptors were retained to undergo subsequent descriptor selection for QSTR analysis. The most relevant parameters were then selected using the model-fitting approach. Prior to model construction, the neurotoxicity datasets (rat and mouse) were split into respective training (70%) and test (30%) subsets using a random distribution method. Using this approach, the samples are selected randomly with a uniform distribution. For the training subset T_{tr}: $p(X \in T_{tr}) = n_{tr}/n, n = |T|$, $n_{tr} = |T_{tr}|$—each sample has an equal probability of selection. This method leads to low bias of the model performance (Reitermanova, 2010). For determining the optimal values of the model parameters, the models were trained (training set) with the retained pool of descriptors through a 5-fold cross-validation (CV) and computing the scoring function (mean squared error, MSE) to rank the contribution of the descriptors in the current set. The lowest ranked descriptors ($< 10\%$ contribution) were then removed in the successive steps (Singh et al., 2015). The most significant descriptors were then retained and the corresponding prediction accuracies were computed. The descriptor selection process was performed separately for each modeling method (PNN and GRNN). Finally retained descriptors for the qualitative and quantitative QSTRs in both the test species (rat and mouse) are presented in Table S2 (Supplementary material). For qualitative QSTR modeling, the solvents were categorized as high neurotoxic ($EC_{30} > 50 \mu$M) and low neurotoxic ($EC_{30} > 50 \mu$M), rendering a total 25 compounds in the high neurotoxicity (class = 1) and remaining 22 compounds in the low neurotoxicity (class = 2) categories.

2.3. Model development, validation and applicability domain analysis

In this study, PNN and GRNN based QSTR models (quantitative and quantitative) were established for predicting the class and neurotoxicity of structurally diverse organic solvents in rats and mice. An ISC based linear QAAR model was also constructed using the rat and mouse neurotoxicity datasets. A brief account of these methods is provided here.

2.3.1. QSTR modeling

PNN estimates the probability density function (PDF) of the features of each class from the available training samples using the Gaussian kernel function, which are then used in a Bayes decision rule to perform the classification (Gelman et al., 2003). PNN uses a nonparametric technique known as the Parzen window to construct the class-dependent PDF for each classification category required by Bayes' theory. This allows determination of the chance a given vector pattern lies within a given category. If the jth training pattern for category C_1 is x_j, then the Parzen estimate of the PDF for category C_1 is;

$$F_1(x) = \frac{1}{2(\pi)^{m/2}\sigma^m n} \sum \exp\left[\frac{(x - x_j)^T(x - x_j)}{2\sigma^2}\right],$$

where n is the number of training patterns, m is the input space dimension, j is the pattern number, and σ is the adjustable smoothing parameter (Goh, 2002). A PNN consists of a node in