

QSARs for predicting physicochemical and biochemical properties of industrial chemicals

Dimosthenis A. Sarigiannis^{1,2*}, Krystalia Papadaki¹, Spyros P. Karakitsios^{1,2}

¹ Aristotle University of Thessaloniki, Department of Chemical Engineering, Environmental Engineering Laboratory, 54124, Thessaloniki, Greece

² Centre for Research and Technology Hellas, Chemical Process and Energy Resources Institute, Natural and Renewable Resource Exploitation Laboratory, 57001, Thessaloniki, Greece

ABSTRACT

A current limitation to further introducing physiology based biokinetic PBBK models in the risk assessment arena is the lack of generic character of these models. A critical limiting factor of describing ADME (absorption, distribution, metabolism and elimination) processes for a large chemical space is the proper parameterization for “data poor” compounds. In order to expand the applicability of PBBK models to cover as much as possible the chemical space, model parameterization for data poor chemicals is done using advanced quantitative structure-activity relationships (QSARs). In silico approaches, including QSARs, are widely used for the estimation of physicochemical and biochemical properties, biological effects as well as understanding the physicochemical features governing a biological response. QSARs are described as regression or classification models, which form a relationship between the biological effects and chemistry of each chemical and comprise the activity data to be modeled, the data with which to model and a method to formulate the model.

In order to expand the applicability of toxicokinetic models to cover as wide as possible a chemical space, parameterization of the models for data poor chemicals was done using advanced quantitative structure-activity relationships (QSARs). This pertains to all necessary parameters such as blood-tissue partition coefficients, clearance and elimination kinetics. Several QSAR modeling approaches have been investigated, including (a) the Peyret, Poulin and Krishnan algorithm, which is based on the fractional content of cells, interstitial fluid in tissue, plasma in blood, erythrocyte in blood, tissue lipids and the lipophilicity of the compound of interest; (b) the molecular fractions algorithm proposed by Béliveau et al. that takes into account the frequency of occurrence of the several molecular fragments of the compounds and (c) Abraham’s solvation equation for estimating biological properties, which takes into account the excess molar refraction that can be determined simply from knowledge of the compound refractive index, the compound dipolarity/polarizability, the solute effective or summation hydrogen-bond acidity, the solute effective or summation hydrogen-bond basicity and the McGowan characteristic volume that can trivially be calculated for any solute simply from a knowledge of its molecular structure. The most efficient of all the examined methods was Abraham’s solvation equation combined with Artificial Neural Networks (ANN). The whole data set was divided into training and test sets (85% and 15% respectively of the total number of compounds) and the model developed from training set is externally validated using the test set. The data sets of 33 and 29 industrial chemicals are used in order to predict the partition coefficients of seven tissues and the metabolic constants,

respectively. The examined QSAR model was analyzed with Multi-Layer Perceptron (MLP) followed by BFGS Quasi-Newton. The Multi-layer network contains a single input layer, which consists of the values of molecular descriptors, one or more hidden layers, which process the descriptors into internal representations and an output layer utilizing the internal representation to produce the final prediction.

A comparison between the two statistical methods used for the analysis of the equation shown that the ANN model is a much better predictor than the non-linear regression model. The square of the correlation coefficient in the case of the partition coefficients ranges between 0.76 and 0.94 while for maximal velocity and Michaelis – Menten constant equals to 0.70 and 0.67, respectively. The model was, then, applied on a test set of compounds, categorized into various chemical groups, in order to examine its applicability domain. Results show that Abraham's solvation equation, estimated by ANN modeling, results in an efficient method for the prediction of industrial chemical physicochemical and biochemical properties.

The significantly improved performance of Abraham's equation and ANN combination can be ascribed to its capacity to represent mathematically the complex interactions of biochemical micro-processes, which are lumped into the metabolic and physicochemical parameters. Generally, the predictive capacity of the Abraham's model for new chemical entities is influenced by chemical nature of the training set molecules used for development of the model. In actual case, the test set molecules can be predicted well when these molecules are structurally very similar to the training set molecules. The reason is that the model has captured all features common to the training set molecules. Nonetheless not even robust and validated QSARs can be expected to reliably predict the response for the entire chemical space.

Overall, the proposed methodology, offers two major advantages:

- It helps filling the data gaps of "data poor" chemicals, allowing the use of internal dosimetry metric for a wide array of chemical classes.
- It effectively supports the "safe by design" concept for industrial chemicals, by allowing the successful prediction of toxicokinetic behavior based on molecular parameters, thus avoiding chemical structures resulting in bioaccumulative and non-rapidly metabolized chemicals.