Development of a novel mathematical model using a group contribution method for prediction of ionic liquid toxicities

M. Ismail Hossain[a], Brahim Belhaouari Samir[b], Mohanad El-Harbawi[a, 1], Asiah Nusaibah Masri[a], M.I. Abdul Mutalib[a], Glenn Hefter[c], Chun-Yang Yin[c,]

[a] Chemical Engineering Department, Universiti Teknologi PETRONAS, 31750 Bandar Seri Iskandar, Tronoh, Perak, Malaysia

[b] Fundamental & Applied Sciences Department, Universiti Teknologi PETRONAS, 31750 Bandar Seri Iskandar, Tronoh, Perak, Malaysia

[c] School of Chemical and Mathematical Sciences, Murdoch University, Murdoch, 6150 WA, Australia

Abstract

A new mathematical model has been developed that expresses the toxicities (EC_{50} values) of a wide variety of ionic liquids (ILs) towards the freshwater flea *Daphnia magna* by means of a quantitative structure–activity relationship (QSAR). The data were analyzed using summed contributions from the cations, their alkyl substituents and anions. The model employed multiple linear regression analysis with polynomial model using the MATLAB software. The model predicted IL toxicities with R^2 = 0.974 and standard error of estimate of 0.028. This

model affords a practical, cost-effective and convenient alternative to experimental ecotoxicological assessment of many ILs.

Keywords: Group contribution method; QSAR; MATLAB; Ecotoxicity; Ionic liquids; *Daphnia magna*

1. Introduction

Ionic liquids (ILs) are low-melting point salts typically containing bulky organic cations and/or anions. Since ILs are essentially non-volatile, they present a low hazard with regard to atmospheric pollution and therefore offer potential as 'green' substitutes for volatile organic compounds in process chemistry. On the other hand, due to their often high solubility in water, the future possible industrial discharge of spent or waste ILs into natural water bodies may have significant detrimental ecotoxicological consequences for aquatic organisms.

Standard experimental ecotoxicological assessments employ bioassays on specified organisms to evaluate the ecological risk posed by a test substance towards the aquatic environment (Pretti et al., 2009). Indeed, the biological effects of a number of ILs have recently been reported on various aquatic organisms including the marine bacterium *Vibrio fischeri* (Ranke et al., 2004 and Docherty and Kulpa, 2005), algae (Cho et al., 2008), the freshwater crustacean *Daphnia magna* (Bernot et al., 2005a) and the freshwater snail *Physa acuta* (Bernot et al., 2005b). However, such experimental investigations are time-consuming and require significant physical and chemical resources. In principle, there are well over one

million ILs which can be synthesized, though only about one thousand have been reported to date (Luis et al., 2007). There is, therefore, a need to provide a faster and more cost-effective approach to assess the aquatic toxicity of ILs.

An alternative to experimental assessment is to predict the toxicity of ILs via first-principles calculations or structure–property correlations (Wilkes, 2004). The first of these approaches is generally not viable with current theories. On the other hand, the latter method, in the form of quantitative structure–activity [or property] relationships (QSARs or QSPRs), is relatively straightforward and has been applied widely to the modeling and prediction of many physicochemical and biological properties. In particular, QSARs based on group contribution methods have been frequently used to estimate thermodynamic and other properties from summative contributions of molecular structures. Properties evaluated to date have included boiling points, melting points, aqueous solubilities, toxicities and retention indices, the last with particular focus on medicinal drugs (Schultz et al., 2003 and Cao et al., 2010).

In toxicology, for example, Luis et al. (2007) have utilized a group contribution method to develop a QSAR based on multiple linear regression (MLR) analysis using the Polymath 5.0 software. This method was able to reliably assess the toxicity of ILs towards the marine bacterium *V. fischeri* using a database of log EC_{50} values (where EC_{50} is the effective concentration of IL required for 50% toxicity within a specified exposure time) for various imidazolium-, pyridinium- and pyrrolidinium-based ILs. Development of a QSAR for the convenient and rapid evaluation of IL toxicity towards more widely used species (Cho et al., 2008 and Docherty et al., 2010) would therefore be of considerable value. While

experimental toxicity studies on the freshwater flea *D. magna* have been widely reported (Cho et al., 2008, Pretti et al., 2009 and Docherty et al., 2010), development of group contribution method for toxicity prediction for this species has been limited. To date, there is only one identified QSAR study for assessment of IL toxicity towards *D. magna* (Couling et al., 2006) with $R^2 = 0.862$.

Thus, the aim of the present work is to develop a novel mathematical model using a group contribution method that is capable of predicting the toxicity of a wide variety of ILs towards *D. magna.* To this end, EC_{50} data have been collected from the literature for as many ILs as practicable, including alkyl-substituted imidazolium, pyridinium, ammonium, phosphonium and pyrrolidinium-based salts, together with a limited (restricted by the available data) range of anions. The model was developed using a combination of MLR and a polynomial model with the code written using the MATLAB software which, to the best of our knowledge, has not been previously reported.

2. Model development

The EC_{50} values for *D. magna* were collected from reported studies (Garcia et al., 2001, Garcia et al., 2005, Bernot et al., 2005a, Couling et al., 2006, Wells and Coombe, 2006, Samorì et al., 2007, Pretti et al., 2009 and Yu et al., 2009). MLR was used to quantify the toxicity contribution of the cations, their alkyl substituents and the anions. Such contributions were represented in the form of an equation that related the dependant variables and one or more explanatory or predictor variables (Chatterjee and Hadi, 2006). In this study, the response variable was denoted as Y while the predictor variables were denoted as X_1,

$X_2,\cdots,X_p$, where p represents the total number of predictor variables. The true relationship between Y and X_1, X_2,..., X_p is approximated by a regression model (Chatterjee and Hadi, 2006) expressed as:

$$Y = f(X_1 + X_2, \cdots X_p) + \varepsilon \quad (1)$$

where ε is defined as a normal random error expressing the discrepancy in the approximation. The linear form of Eq. (1) can be expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \quad (2)$$

where β_o, $\beta_1,\cdots,\beta_p$, are defined as regression coefficients, i.e., constants to be derived from the data. Eq. (2) was integrated with a polynomial model (Angelov, 2002) and the interaction between the two models (ωX_i) was described by Eq. (3):

$$Y = \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i + \sum_{i=1}^{m} \gamma_i \varpi^i + \sum_{i=n+1}^{k} \delta_i \varpi X_i \quad (3)$$

where $\alpha_i = \beta_i$ for $i \leqslant n$; $\gamma_i = \beta_{i+n}$ for $i \leqslant m$; $\delta_i = \beta_{i+n+m}$ for $i \leqslant k$; $n + m + k = p$. The MATLAB software version 7.8.0.347 was used to develop the code and estimate log EC_{50} by means of the algorithm illustrated in Fig. 1.

Results and discussion

The database of IL toxicities towards *D. magna* is shown in Table S1 (electronic Supplementary material). Listed anions were hexafluorophosphate (PF^-_6), N-cyanocyanamide [$N(CN)^-_2$], bromide (Br^-), chloride (Cl^-), tetrafluoroborate (BF^-_4) and bis(trifluoromethyl)sulfonylimide (Tf_2N^-). Cations included imidazolium, pyridinium, ammonium, phosphonium, pyrrolidinium, morpholinium, and thiophenium (denoted as

Imida, *Pyrid*, *Ammo*, *Phos*, *Pyrrol*, *Morp* and *thio*, respectively). Substituent groups included $–CH_3$, $–CH_2–$, $–NCH_3$, $–NCH_2–$, $–OCH_3$, $–OCH_2–$, $–SCH_2–$, $–PCH_2–$ and Bz (benzene ring). The molecular-level descriptors for a given IL have a non-zero value when the particular group is present in the IL (Luis et al., 2007). Fig. 2 shows the chemical structures of the ILs covered, together with the substituents used in the group contribution method.

The database in Table S1 was fitted to the QSAR using the MATLAB software and the obtained contributions (α_i, γ_i *and* δ_i) of each group are shown in Table 1. A good fit was achieved using Eq. (3) which gives $R^2 = 0.974$ and a standard error of 0.0283 at the 95% confidence interval (CI). The CIs for each contribution to the molecular structure are presented in Table 1. This implies consistency between the experimental toxicity values and those predicted by the QSAR. These values are comparable to QSAR models developed by other researchers (Table 2). The ILs used in the initial training set for model development by Couling et al. (2006) encompassed results from three separate IL toxicity studies (Bernot et al., 2005a, Docherty and Kulpa, 2005 and Couling et al., 2006) while the EC_{50} values for *D. magna* used in our model were collected from eight separate studies (see above). Incorporation of a wide range of IL toxicities, determined by independent research groups, should produce a more reliable model (which is consistent with the high R^2 value).

Fig. 3 shows that the model predicts toxicities reliably since the plotted values remain close to the Log $EC_{50predicted}$ = Log $EC_{50experimental}$ line throughout and there are no extreme outliers. It can be seen (Fig. 4), that there are about 45 ILs with ‘zero’ difference between the predicted and experimental log EC_{50} values. The total sum of all contributions (Table 1) except for the constant α_0 (total mean) was calculated and the percentage of every descriptor

was measured. The toxicity percentage was obtained by considering the individual contributions for the substituents, cations and anions. These results indicated that the ammonium, imidazolium, pyridinium and morpholinium cations contribute approximately 48%, 32%, 31% and 12% of the total toxicity, respectively (Table S2). The contribution of the anions to the overall toxicity was negative with the exception of BF^-_4 (7% of total toxicity) suggesting that the inclusion of Br^-, Cl^-, TF_2N^-, and PF−6 in the chemical structures of the ILs reduces their toxicity to *D. magna*. This trend is consistent with the published toxicity data for *V. fischeri* (Luis et al., 2007). A recent report by Wang et al. (2011) indicated that some IL anions (TF_2N^-, PF^-_6, BF^-_4) exhibit toxicity towards gram positive bacteria *Clostridium* sp. This suggests that anion toxicity mostly depends on the susceptibility of the particular organisms concerned.

The order of toxicity of ILs towards *D. magna* based on the present findings is summarized in Fig. 5. The higher toxicity of quaternary ammonium compared to imidazolium and pyridinium is in agreement with findings of Wells and Coombe, 2006 and Pretti et al., 2009.

Of course the present findings are not an all-encompassing representation of the impact of ILs on the aquatic environment. Similar data and models will be required for other representative species and, as the syntheses of ILs continue to expand, a much wider range of components, especially of substituents and anions, will need to be investigated. Other factors such as biodegradation and bioaccumulation will also need to be studied to attain an overview of the likely environmental impact of ILs (Couling et al., 2006).

4. Conclusions

The proposed group contribution method is capable of reliably predicting the toxicity of a variety of ILs towards the freshwater flea *D. magna*. This model provides a practical, cost-effective, convenient and reliable alternative to experimental ecotoxicological assessment of ILs. The use of the group contribution method allows flexibility to “add” or “omit” specific chemical groups based on the requirements of the model. Future studies in this area will need to cover a wider spectrum of ILs with respect to their toxicity towards algae, fish and other higher-order organisms.

Acknowledgment

The authors acknowledge the PETRONAS Ionic Liquid Center (PILC) and Chemical Engineering Department, Universiti Teknologi PETRONAS for their support in conducting this study.

References

Angelov, P.P., 2002. Evolving Rule-Based Models: A Tool for Design of Flexible Adaptive Systems. Springer-Verlag, London, United Kingdom.

Bernot, R.J., Brueseke, M.A., Evans-White, M.A., Lamberti, G.A., 2005a. Acute and chronic toxicity of imidazolium based ionic liquids on *Daphnia magna*. Environ. Toxicol. Chem. 24, 87–92.

Bernot, R.J., Kennedy, E.E., Lamberti, G.A., 2005b. Effects of ionic liquids on the survival, movement, and feeding behavior of the fresh water snail *Physa acuta*. Environ. Toxicol. Chem. 24, 1759–1765.

Cao, D.S., Liang, Y.Z., Xu, Q.S., Li, H.D., Chen, X., 2010. A new strategy of outlier detection for QSAR/QSPR, J. Comput. Chem. 31, 592–602.

Chatterjee, S., Hadi, A.S., 2006. Regression Analysis by Example, fourth ed. John Wiley, New York.

Cho, C.W., Jeon, Y.C., Pham, T.P.T., Vijayaraghavan, K., Yun, Y.S., 2008. The ecotoxicity of ionic liquids and traditional organic solvents on microalga *Selenastrum capricornutum.* Ecotoxicol. Environ. Safety 71, 166–171.

Couling, D.J., Bernot, R.J., Docherty, K.M., Dixon, J.K., Maginn, E.J., 2006. Assessing the factors responsible for ionic liquid toxicity to aquatic organisms via quantitative structure–property relationship modeling. Green Chem. 8, 82–90.

Docherty, K.M., Joyce, M.V., Kulacki, K.J., Kulpa, C.F., 2010. Microbial biodegradation and metabolite toxicity of three pyridinium-based cation ionic liquids. Green Chem. 12, 701–712.

Docherty, K.M., Kulpa, C.F., 2005. Toxicity and antimicrobial activity of imidazolium and pyridinium ionic liquids. Green Chem. 7, 185–189.

Garcia, M.T., Gathergood, N., Scammells, P.J., 2005. Biodegradable ionic liquids. Part II: effect of the anion and toxicology. Green Chem. 7, 9–14.

Garcia, M.T., Ribosa, I., Guindulain, T., Sanchez-Leal, J., Vives-Rego, J., 2001. Fate and effect of monoalkyl quaternary ammonium surfactants in the aquatic environment. Environ. Pollut. 111, 169–175.

Garcia-Lorenzo, A., Tojo, E., Tojo, J., Teijeira, M., Rodriguez-Berrocal, F.J., Gonzalez, M.P., Martinez-Zorzano, V.S., 2008. Cytotoxicity of selected imidazolium-derived ionic liquids in the human Caco-2 cell line sub-structural toxicological interpretation through a QSAR study. Green Chem. 10, 508–516.

Lacrama, A.-M., Putz, M.V., Ostafe, V., 2007. A spectral-SAR model for the anioniccationic interaction in ionic liquids: application to *Vibrio fischeri* ecotoxicity. Int. J. Mol. Sci. 8, 842–863.

Luis, P., Garea, A., Irabien, A., 2010. Quantitative structure–activity relationships (QSARs) to estimate ionic liquids ecotoxicity EC50 (*Vibrio fischeri).* J. Mol. Liq. 152, 28–33.

Luis, P., Ortiz, I., Aldaco, R., Irabien, A., 2007. A novel group contribution method in the development of a QSAR for predicting the toxicity (*Vibrio fischeri* EC50) of ionic liquids. Ecotoxicol. Environ. Safety 67, 423–429.

Pretti, C., Chiappe, C., Baldetti, I., Brunini, S., Monni, G., Intorre, L., 2009. Acute toxicity of ionic liquids for three fresh water organisms: *Pseudokirchneriella subcapitata, Daphnia magna* and *Danio rerio*. Ecotoxicol. Environ. Safety 72, 1170–1176.

Ranke, J., Moelter, K., Stock, F., Bottin-Weber, U., Poczobutt, J., Hoffmann, J., Ondruschka, B., Filser, J.B., Jastorff, 2004. Biological effects of imidazolium ionic liquids with varying chain lengths in acute *Vibrio fischeri* and WST-1cell viability assays. Ecotoxicol. Environ. Safety 58, 396–404.

Ranke, J., Müller, A., B-Weber, U., Stock, F., Stolte, S., Arning, J., Störmann, R., Jastorff, B., 2007. Lipophilicity parameters for ionic liquid cations and their correlation to in vitro cytotoxicity. Ecotoxicol. Environ. Safety 67, 430–438.

Samorì, C., Pasteris, A., Galletti, P., Tagliavini, E., 2007. Acute toxicity of oxygenated and nonoxygenated imidazolium based ionic liquids to Daphnia magna and *Vibrio fischeri*. Environ. Toxicol. Chem. 26, 2379–2382.

Schultz, T.W., Cronin, M.T.D., Walker, J.D., Aptula, A.O., 2003. Quantitative structure–activity relationships (QSARs) in toxicology: a historical perspective. J. Mol. Struct. Theochem. 622, 1–22.

Torrecilla, J.S., Garcia, J., Roja, E., Rodriguez, F., 2009. Estimation of toxicity of ionic liquids in leukemia rat cell line and acetylcholinesterase enzyme by principal component analysis, neural networks and multiple linear regressions. J. Hazard. Mater. 164, 182–194.

Wang, H., Malhotra, S.V., Francis, A.J., 2011. Toxicity of various anions associated with methoxyethyl methyl imidazolium-based ionic liquids on *Clostridium sp*. Chemosphere 82, 1597–1603.

Wells, A.S., Coombe, V.T., 2006. On the freshwater ecotoxicity and biodegradation properties of some common ionic liquids. Org. Proc. Res. Dev. 10, 794–798.

Wilkes, J.S., 2004. Properties of ionic liquid solvents for catalysis. J. Mol. Catal. A Chem. 214, 11–17.

Yu, M., Wang, S.-H., Luo, Y.-R., Han, Y.-W., Li, X.-Y., Zhang, B.-J., Wang, J.-J., 2009. Effects of the 1-alkyl-3-methylimidazolium bromide ionic liquids on the antioxidant defense system of *Daphnia magna*. Ecotoxicol. Environ. Safety 72,1798–1804.

Figure 1. Algorithm for the estimation of log EC_{50} using MATLAB 7.8.0.347.

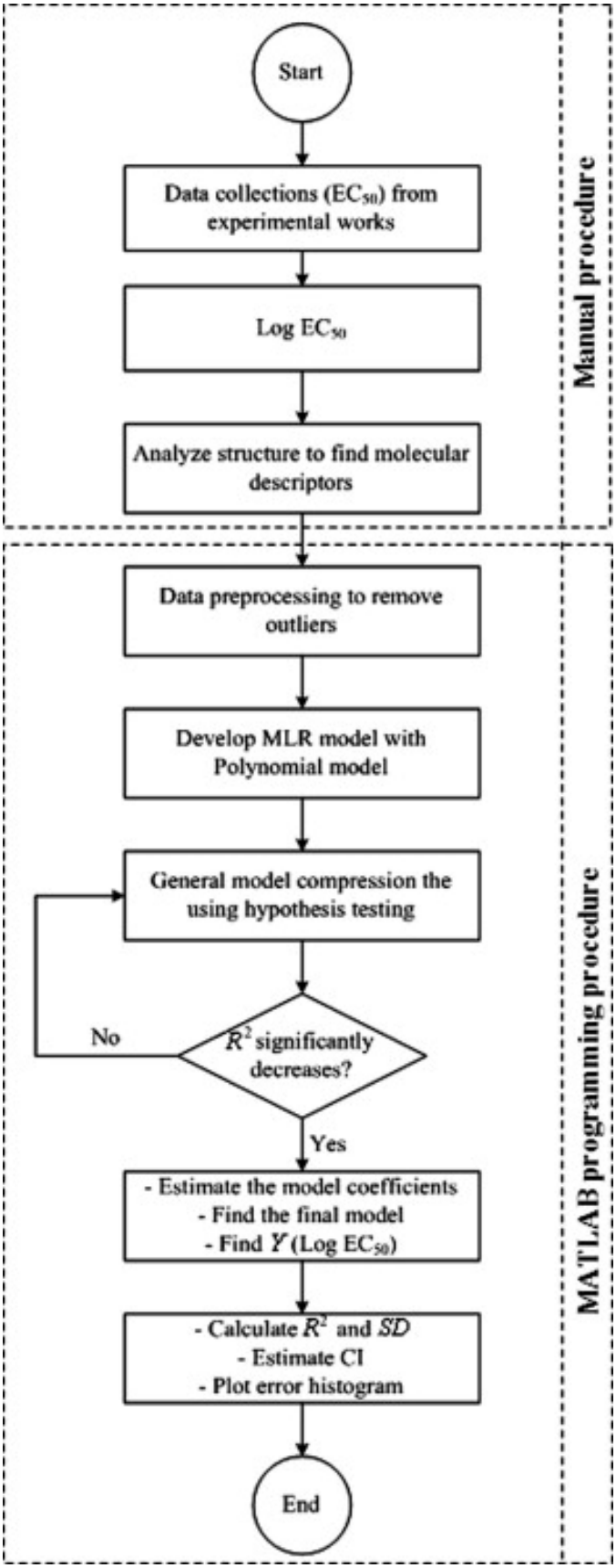

Figure 2. Chemical structures of ILs together with substitution groups used in the group contribution method.

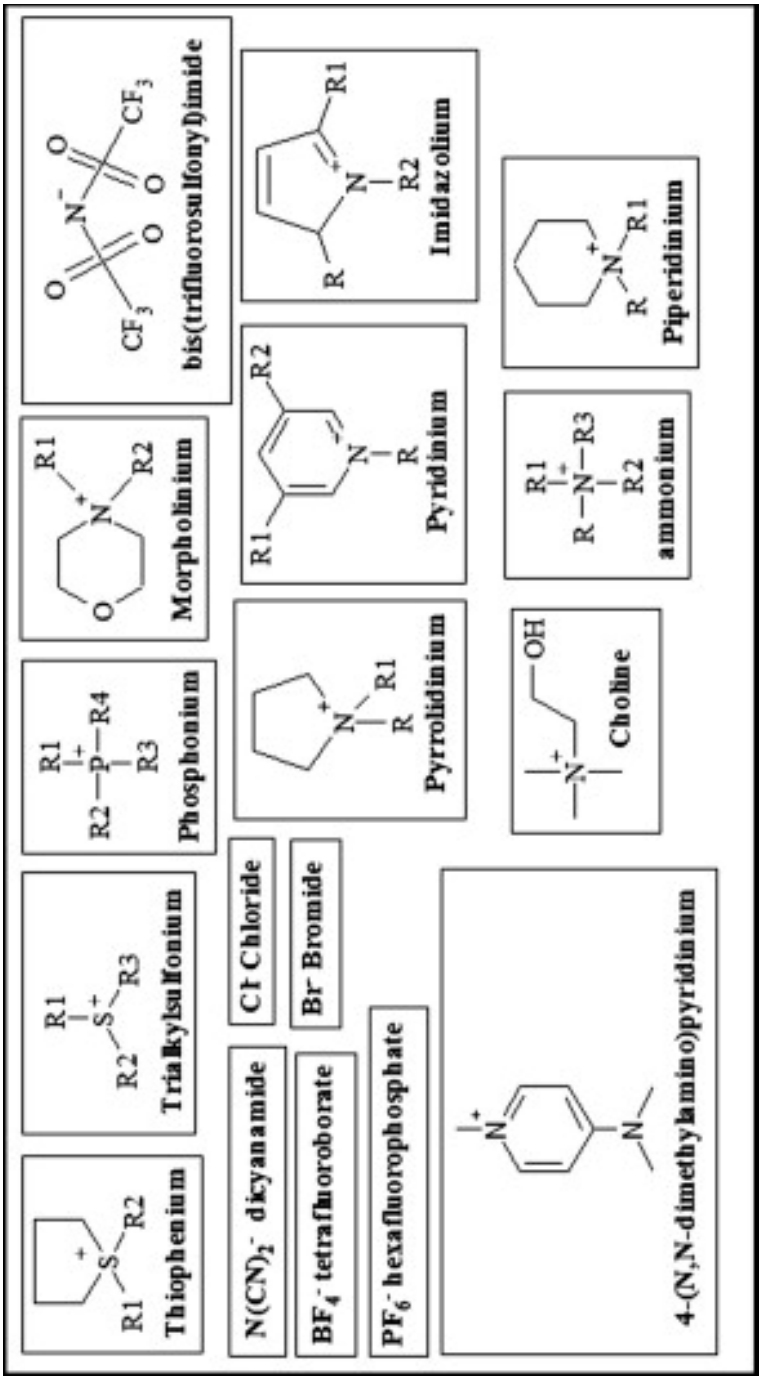

Table 1. Quantified contributions of molecular descriptors to the overall toxicity.

Coefficients No.	Quantified descriptors[a]	Coefficients	Contribution/ Coefficients	Upper CI	Lower CI	Coefficient percent
1	1	α_0	-0.44	-10.28	9.40	-
2	X_1	α_1	-2.55	-2.91	-2.18	28.44
3	X_2	α_2	-2.02	-2.25	-1.78	22.53
4	X_3	α_3	-0.46	-0.68	-0.25	5.19
5	X_8	α_8	0.45	-9.59	10.50	-5.08
6	X_{10}	α_{10}	-2.91	-4.81	-1.01	32.52
7	X_{11}	α_{11}	-2.82	-4.65	-0.98	31.49
8	X_{12}	α_{12}	-4.33	-8.27	-0.38	48.33
9	X_{15}	α_{15}	-1.16	-2.81	0.50	12.93
10	X_{17}	α_{17}	0.30	-10.61	11.20	-3.33
11	X_{18}	α_{18}	0.87	-9.40	11.13	-9.67
12	X_{19}	α_{19}	-0.66	-11.51	10.19	7.36
13	X_{21}	α_{21}	0.58	-10.87	12.02	-6.42
14	X_{23}	α_{23}	0.39	-8.69	9.47	-4.37
15	X_{25}	α_{25}	1.15	-6.49	8.79	-12.85
16		γ_1	0.74	-1.61	3.08	-8.26
17		γ_2	-0.16	-0.17	-0.14	1.78
18	X_1	δ_1	0.13	0.13	0.13	-1.49
19	X_2	δ_2	0.09	0.09	0.09	-0.97
20	X_3	δ_3	0.01	0.01	0.01	-0.11
21	X_6	δ_6	-0.14	-0.15	-0.13	1.57
22	X_7	δ_7	-0.07	-0.07	-0.07	0.81
23	X_8	δ_8	-0.01	-0.08	0.07	0.07
24	X_9	δ_9	0.26	0.21	0.31	-2.91
25	X_{10}	δ_{10}	0.26	0.19	0.32	-2.86
26	X_{11}	δ_{11}	0.24	0.18	0.31	-2.72
27	X_{12}	δ_{12}	0.40	0.35	0.45	-4.44
28	X_{13}	δ_{13}	0.10	0.07	0.14	-1.17
29	X_{15}	δ_{15}	0.19	0.13	0.25	-2.10
30	X_{17}	δ_{17}	0.27	0.20	0.35	-3.06
31	X_{18}	δ_{18}	0.24	0.17	0.31	-2.70
32	X_{19}	δ_{19}	0.33	0.26	0.41	-3.73
33	X_{20}	δ_{20}	0.30	0.25	0.35	-3.35
34	X_{21}	δ_{21}	0.21	0.13	0.28	-2.30
35	X_{22}	δ_{22}	0.28	0.22	0.34	-3.12
36	X_{23}	δ_{23}	0.24	0.18	0.30	-2.66
37	X_{24}	δ_{24}	0.30	0.28	0.32	-3.36

[a] Quantified descriptors represent the structural fragments, $X_1 = -CH_3$ ..... $X_{22} = N(CN)_2$, $X_{23} = \mathrm{Sum}(X_1{:}X_4)$, $X_{24} = \mathrm{Sum}(X_5{:}X_8)$, $X_{25} = \mathrm{Sum}(X_9{:}X_{22})$, $\omega = \mathrm{Sum}(X_1{:}X_{25})$.

Table 2. Comparison between several QSAR models and their correlation coefficients.a

Method	Organism(s)	Number of ILs used	Correlation Coefficient, R^2	Software used	Error	References
Genetic function approximation (GFA)	*Vibrio fischeri*; *Daphnia magna*	25 and 17	0.78–0.88	MOPAC	na	Couling et al. (2006)
MLR	*Vibrio fischeri*	43	0.925	Polymath 5.0	0.0051	Luis et al. (2007)
Linear regression model	Leukemia rat cell line, IPC-81	74	0.78	R 2.12.1	0.35	Ranke et al. (2007)
Spectral-SAR vectorial model	*Vibrio fischeri*	22	na	HyperChem Program	na	Lacrama et al. (2007)
Topological sub-structural molecular design (TOPS-MODE)	Caco-2 cells	15	0.98	MODESLAB 1.5	na	Garcia-Lorenzo et al. (2008)
MLR, Radial Basis (RB) multilayer perception (MLP), neural network (NN)	Leukemia rat cell line, IPC-81; Acetylcholinesterase (AChE)	153	MLR, R^2 = 0.867 (IPC) MLR, R^2 = 0.814 (AChE) NN, R^2 = 0.982 (IPC) NN, R^2 = 0.973 (AChE)	MATLAB and Statgraphics Plus	0.022 and 0.038	Torrecilla et al. (2009)
MLR	*Vibrio fischeri*	96	0.924	Polymath 5.0	na	Luis et al. (2010)
MLR and polynomial methods	*Daphnia magna*	64	0.974	MATLAB 7.8.0.347	0.0283	This study

[a] na denotes "not available".

Figure 3. Comparison between predicted and experimental toxicity values.

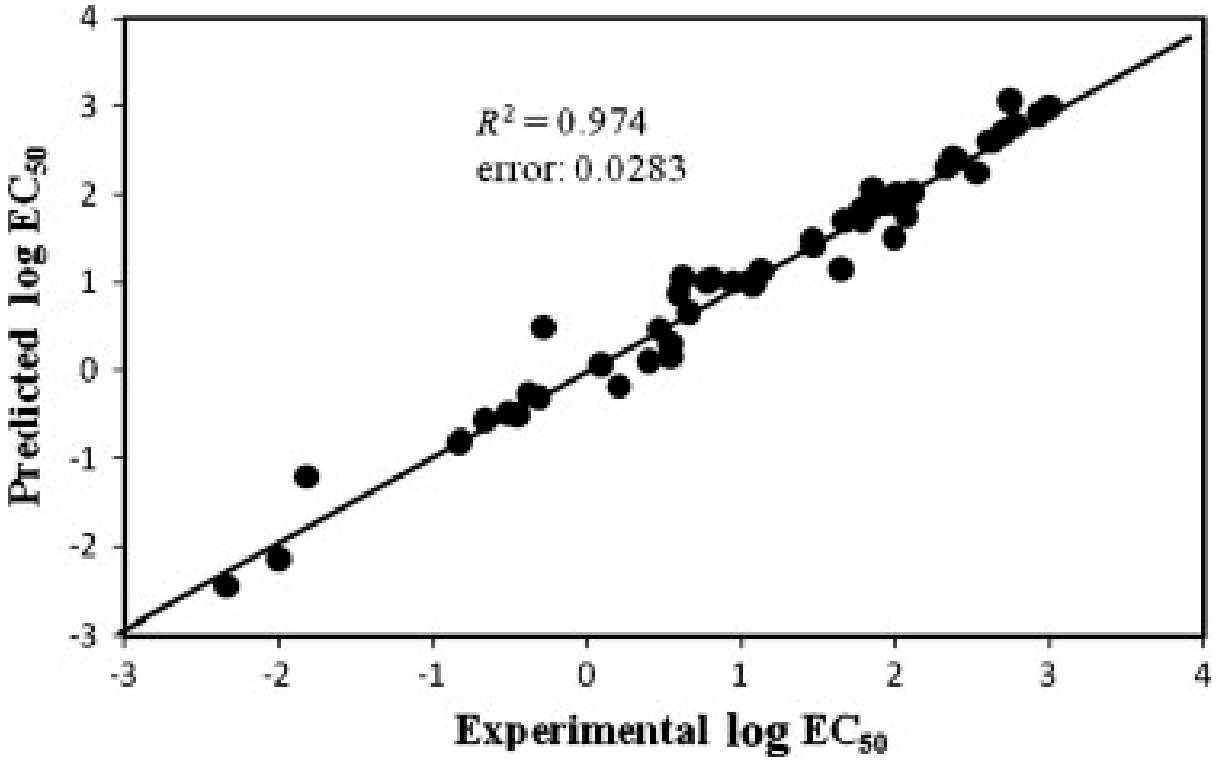

Figure 4. Results error histogram

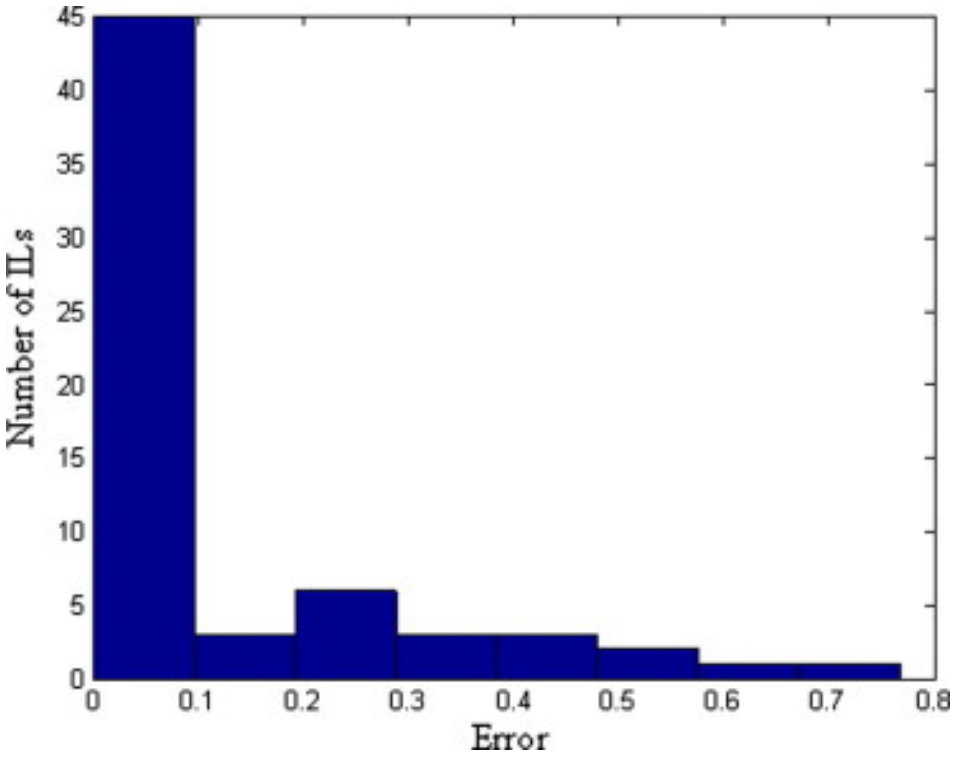

Figure 5. Order of toxicity of ILs towards *Daphnia magna.* Note that the toxicity sensitivity differs between columns (see text).