Development of predictive QSAR models for *Vibrio fischeri* toxicity of ionic liquids and their true external and experimental validation tests†

Rudra Narayan Das,[a] Tânia E. Sintra,[b] João A. P. Coutinho,[b] Sónia P. M. Ventura,*[b] Kunal Roy*[a,c] and Paul L. A. Popelier*[c]

Despite possessing an interesting chemical nature and tuneable physicochemical properties, ionic liquids (ILs) must have their ecotoxicity tested in order to be commercialized. The water solubility of ILs allows their easy access to the aquatic compartment of the ecosystem creating a potential hazard to aquatic organisms. Hence, it is relevant to design ionic liquids with lower toxicity while keeping the desired properties of interest. Considering the possibility of an enormous number of combinations of different cations and anions, a rational guidance for the structural design of ionic liquids is essential in order to prioritize the synthesis as well as testing of selected molecules only. Predictive *in silico* models, such as quantitative structure–activity relationship (QSAR) models, can play a pivotal role in exploring the important chemical attributes contributing to the response activity. These models may then lead to the design of novel ionic liquids. The present study aims at developing predictive QSAR models for the ecotoxicity of ionic liquids using the bacteria *Vibrio fischeri* as an indicator response species. Instead of a single model, here we have used multiple models to capture more complete structural information of ionic liquids for toxicity towards *Vibrio fischeri*. The derived chemical attributes have been implemented in designing new analogues, some of which have been synthesized and had their ecotoxicity tested for the same model organism. The predictive QSAR models reported here can be used for ecotoxicity prediction of new IL chemicals and for data-gap filling. Moreover, the synthesized low-toxic ILs could be considered for evaluation as well as for application in suitable processes serving the purpose of industry and academia.

1. Introduction

Ionic liquids are molten salts exhibiting the liquid state at or close to room temperature.[1] These compounds are a novel class of solvents that have drawn industrial attention due to their interesting properties such as thermal and electrochemical stability, negligible vapor pressure, non-inflammability, satisfactory solvation behavior, high density, high thermal conductivity, a wide temperature range for the liquid state, high selectivity and tuneable properties.[2,3] Due to these advantages, ILs have found diverse uses as electrolytic media,

in synthesis, catalysis and separation, as solvents and in engineering and biological applications.[4] Actually, ILs are no longer confined to academia because they are already present in a range of different commercial products and industrial processes.[5] Ionic liquids can be used for multi-tasking because of their easily tuneable nature arising from their intrinsic properties. It may be noted that approximately one billion (10^{12}) binary and one trillion (10^{18}) ternary combination systems of ionic liquids would be possible only by using 1 million simple systems (cations and anions).[6] It would be time consuming, expensive and even impossible to obtain different physical, chemical and biological (toxicological) properties data by measuring all possible ILs in order to screen for an optimum IL given a special purpose. Due to the huge number of potential ILs, experimental data for different properties are currently available only for a small fraction of these ILs. This lack of data can be a major drawback, especially in systematic screening to find the best-suited solvent for a particular task. In order to design any new process involving ionic liquids on an industrial scale, it is necessary to have knowledge of various properties as well as an understanding of the molecular struc-

[a]Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India.
E-mail: kunalroy_in@yahoo.com
[b]CICECO – Aveiro Institute of Materials, Department of Chemistry, University of Aveiro, 3810-193 Aveiro, Portugal. E-mail: spventura@ua.pt
[c]Manchester Institute of Biotechnology, 131 Princess Street, Manchester M1 7DN, UK. E-mail: paul.popelier@manchester.ac.uk
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c6tx00180g

ture of the compounds. Therefore, it is necessary to develop mathematical models to predict the various property endpoints of ILs. Quantitative structure–property/activity relationship (QSPR/QSAR) methods[7–9] would help to develop quantitative models capable of predicting the properties directly from molecular structure information.

Ionic liquids could theoretically be designed to have a desired property by combining different pairs of ions. To explore the "tuneability" and "designability" features of ILs, predictive QSPR models have to be developed to relate the properties to the chemical structure or other physicochemical properties.[4,9] Such models should also be rigorously validated in order to prove their predictive capacity and applicability to a new set of ILs.

Although ILs were originally promoted as green solvents, studies have also shown that ionic liquids, as any organic solvent, may have some degree of toxicity to the various organisms of the ecosystem.[10,11] Predictive QSAR models can explore the structural attributes of ILs towards various physicochemical and toxicological endpoints, thereby leading to the design of "greener" analogues with higher process selectivity. This approach can lead to filling large data gaps, since the toxicity data are only available for a limited number of ILs against different indicator organisms of the ecosystem. The QSAR approach for toxicity predictions is also encouraged in the REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) legislation of the European Union.[12] However, the *in silico* models should be developed in accordance with the guidelines of the Organization for Economic Cooperation and Development (OECD).[13]

Vibrio fischeri (*V. fischeri*) is a Gram-negative, rod-shaped bacterium, and considered as an important member in a marine ecosystem.[14] It can be easily cultured and bred in the laboratory. Thus, *V. fischeri* can be easily applied as a test organism for assessment of toxicity of chemicals in the aquatic systems. Bioassays based on *V. fischeri* (such as Microtox®) involve a simple procedure, short testing times and are cost-effective. Such assays are recommended by international standards to monitor the toxicity of environmental contaminants.[15] There have been a few reports in the literature on modelling the toxicity of ILs towards *V. fischeri*.[16–22] In the present work, we have developed QSAR models for *V. fischeri* toxicity using the largest available set of ionic liquids with the experimental toxicity data using Microtox®. We have also applied these models for prediction of toxicity of a recently available set of ionic liquids for a true external validation of the developed models. In order to experimentally validate the models, a set of IL compounds with low predicted toxicity values was designed, subsequently synthesized and experimentally tested for their toxicity towards *V. fischeri*. Note that this is the first attempt to perform both true external validation and experimental validation of QSPR models for toxicity of ionic liquids to *V. fischeri*. We have also given serious attention to the applicability domain of the developed models during the prediction of external compounds as recommended by the OECD guidelines.[13]

2. Materials and methods

2.1 The dataset and descriptors

We have assembled from the literature a large dataset of 305 ionic liquids with their ecotoxicity values on Microtox® based on the luminescence inhibition of *V. fischeri*.[16,17,20,22–36] It should be mentioned that we have considered the median effective concentration (EC_{50}) data determined at 15 min or 30 min exposure. When both data are available, their average values were used considering the insignificant impact of the exposure time[23] on *V. fischeri* toxicity for ionic liquids. For maintaining uniformity, the EC_{50} data obtained from different literature sources were converted to the molar unit (mol L^{-1}) followed by their transformation into a negative logarithmic scale *i.e.*, pEC_{50} (EC_{50} in mol L^{-1}). The cationic composition of the dataset varies within ammonium, cholinium, imidazolium, morpholinium, melanimium, phosphonium, tropinium, piperidinium, pyridinium, pyrrolidinium, quinuclidinium, and sulphonium in suitable combination of various inorganic as well as organic anions. The predictor variables employed in this study were computed for both the cations and anions, and include various one- and two-dimensional descriptors involving constitutional features, connectivity parameters, information indices, extended topochemical atom (ETA) indices,[37] functional group counts, atom-centred fragments, molecular profiles, 2D-atom pair based parameters, *etc.*,[38] in addition to quantum chemical attributes namely Quantum Topological Molecular Similarity (QTMS) parameters,[39,40] and computed lipophilicity measures. The detailed categorical list of the descriptors can be found in Table S1 of the ESI.† While the other descriptors were obtained without the need of any geometry optimization process, the QTMS parameters were derived from the *ab initio* based optimized geometry at the HF/6-31G(d) level of theory, and were limited to only cations. The log k_0 values were computed using QTMS and ETA indices as proposed by Roy and Popelier.[41]

Finally, we have employed an additional external dataset of eight compounds (not used for developing the models), but for judging the true external predictivity of the models.[42]

2.2 Development of predictive QSTR models

2.2.1 Dataset division and descriptor pre-treatment. Variance and correlation based criteria were implemented for the thinning of the descriptor pool giving predictor variables with a variance >0.0001 and an inter-correlation (r) among descriptors <0.95. The dataset was divided into a training set and a test set of compounds using the k-means clustering algorithm.[43] A total of six clusters were derived for the whole data followed by random selection of approximately 70% of compounds in the training set ($n_{training}$ = 213) and the remaining 30% compounds in the test set (n_{test} = 92) from each cluster. We preferred to choose the k-means clustering technique over a mere random method in order to achieve a rational and uniform division of the dataset so that the training set can encompass the entire structural domain with the test set chemicals lying in the vicinity of one or more training set

molecules. Note that QSAR models make predictions that are based on the similarity principle and will thus perform better when the test set molecules are structurally similar to the training set compounds and are thus within the applicability domain of the models. The information on the k-clusters can be found in Table S2 of the ESI.†

2.2.2 Employed statistical analyses, chemometric tools and validation parameters for *in silico* modeling. Multiple linear regression (MLR)[44] and partial least squares (PLS)[45] techniques have been used as the statistical methods for the derivation of the QSAR models while the selection of features has been performed by employing chemometric tools, namely, genetic function approximation (GFA)[46,47] and a stepwise based method[48] coupled with Fischer value (F-value) based criteria. In the present study, we have used the GFA technique for the identification of most occurring descriptors, which were subsequently used for stepwise based MLR analysis using the stepping criteria of F to enter = 4.0 and F to remove = 3.9.[44] The best equation obtained from the latter was then subjected to a PLS run considering PLS to be a more robust method for avoiding the problems of multicollinearity.[45] The PLS model was also optimized considering a 5% rise in the Q^2 value as the indicator. Thus, a three layered treatment, that is, GFA followed by a stepwise-based MLR followed by PLS regression was applied for the development of QSAR models.

The developed models were subjected to sufficient statistical validation tests using various metrics to denote model fitness as well as predictivity. Multiple validation strategies involving quality parameters (R^2, R_a^2, F-value),[44] internal (Q^2_{LOO}) and external ($Q^2_{ext(F1)}$, $Q^2_{ext(F2)}$, $Q^2_{ext(F3)}$) validation metrics have been adopted. The chemical domain of applicability[49] of the developed models was determined using the distance to model X (DModX) based approach.[45] Some details of different validation metrics are provided in Table S3 in the ESI.†

2.2.3 Used software tools for QSTR analysis. The chemical structures of the cations and anions were drawn using MarvinSketch (version 15.12.7) software,[50] while Dragon (version 6)[51] and PaDEL-Descriptor (version 2.11) software[52] were employed for the computation of various two-dimensional variables. The determination of an estimated geometry and *ab initio* optimization of the cations were respectively carried out using the GUI GaussView[53] and the program GAUSSIAN03[54] followed by derivation of the QTMS indices using the in-house computer program MORPHY.[55] The k-means cluster based division was performed using SPSS (version 9.0.0) software.[56] The GFA analysis was performed using Cerius[2] (version 4.10) software,[57] while the stepwise based MLR and PLS operations were respectively carried out by employing MINITAB (version 14.13)[58] and SIMCA-P (version 10.0)[59] software, which was also used for the determination of DModX values.

2.3 Synthesis and toxicological assessment of morpholinium-based ILs

2.3.1 Materials. Bromoethane (98.0 wt% of purity), 1-bromopropane (99.0 wt% of purity), dimethylsulfate (99.8 wt% of purity), iodomethane (99.0 wt% of purity), morpholine (99.0 wt% of purity), 4-methylmorpholine (99.0 wt% of purity), 4-(2-hydroxyethyl)morpholine (99.0 wt% of purity), potassium acetate (99.0 wt% of purity), acetic acid glacial (99.9 wt% of purity) and toluene (99.8 wt% of purity) were acquired from Sigma-Aldrich. Ethyl acetate (99.0 wt% of purity) and ethanol (99.9 wt% of purity) were purchased from Carlo Herba. Silver nitrate (99.8 wt% of purity) and formic acid (91.0 wt% of purity) were bought from Panreac. 4-Ethylmorpholine (97 wt% of purity), potassium hydroxide (pure) and acetone (HPLC grade) were acquired from Fluka, Pronalab and VWR, respectively. The water used was double distilled, passed by using a reverse osmosis system and further treated with a Milli-Q plus 185 water purification apparatus. Seven morpholinium-based ILs were synthesized, namely *N*-methyl-*N*-ethylmorpholinium bromide, [Mor12]Br; *N*-methyl-*N*-ethylmorpholinium acetate, [Mor12][Acetate]; *N*-methyl-*N*-ethylmorpholinium formate, [Mor12][For]; *N*-methyl-*N*-ethylmorpholinium methylsulfate, [Mor12][CH$_3$SO$_4$]; *N*-methyl-*N*-propylmorpholinium bromide, [Mor13]Br, *N*-methyl-*N*-hydroxyethylmorpholinium iodide, [Mor12OH]I; and morpholinium acetate, [Mor][Acetate]. Their respective acronyms and chemical structures are depicted in Fig. 1. The structure of all compounds synthesized was confirmed by ^{1}H and ^{13}C NMR spectroscopy, showing the high purity level of all the ionic structures after their synthesis. Due to the quadrupole moment of the ^{14}N nucleus, ^{1}H–^{14}N and ^{13}C–^{14}N, couplings were observed in the NMR spectra of the [Mor$_{12}$] cation, which are in accordance with the literature.[60]

2.3.2 Synthesis and characterization of morpholinium-based ILs. • *N*-Methyl-*N*-ethylmorpholinium bromide, [Mor12]Br, was prepared by dropwise addition of 6 mL of bromoethane (80.4 mmol) to a solution of 4-methylmorpholine (72.3 mmol, 7.32 g) in ethyl acetate, at room temperature. The reaction mixture was refluxed, stirred at 55 °C, and protected from light overnight. A solid was formed, which was filtered off and washed with ethyl acetate (3 × 15 mL). Finally, the residual solvent was removed under reduced pressure and the obtained compound was dried under high vacuum for at least 48 h.[60] [Mor12]Br was obtained as a white solid (44% yield, 6.67 g). ^{1}H NMR (D$_2$O, 300 MHz, [ppm]): δ 1.39 (tt, J_{HH} = 7.3 Hz and J_{NH} = 1.9 Hz, 3H, NCH$_2$CH_3), 3.18 (s, 3H, NCH_3), 3.39–3.65 (m, 6H, N(CH_2)$_3$), 3.98–4.12 (m, 4H, O(CH_2)$_2$). ^{13}C NMR (D$_2$O, 75.47 MHz, [ppm]): δ 6.57, 46.01 (t, J_{CN}, NCH$_3$), 59.06 (t, J_{CN}, NCH$_2$CH$_3$), 60.40, 60.81.

• *N*-Methyl-*N*-ethylmorpholinium acetate, [Mor12][Acetate]. As a first step, potassium acetate in a water solution (5.2 mmol, 0.51 g) was added to an aqueous solution of silver nitrate (4.7 mmol, 0.80 g). The solution was stirred at room temperature for 3 h, leading to the formation and precipitation of silver acetate. After filtration, the solid was washed with water (3 × 10 mL). The residual water was removed under reduced pressure. Silver acetate was obtained with 90% yield (0.71 g). In the second stage, a stoichiometric amount of silver acetate (1.2 mmol, 0.20 g) was added to an aqueous solution of *N*-methyl-*N*-ethylmorpholinium bromide (1.2 mmol, 0.25 g). The reaction mixture was stirred at room temperature, and pro-
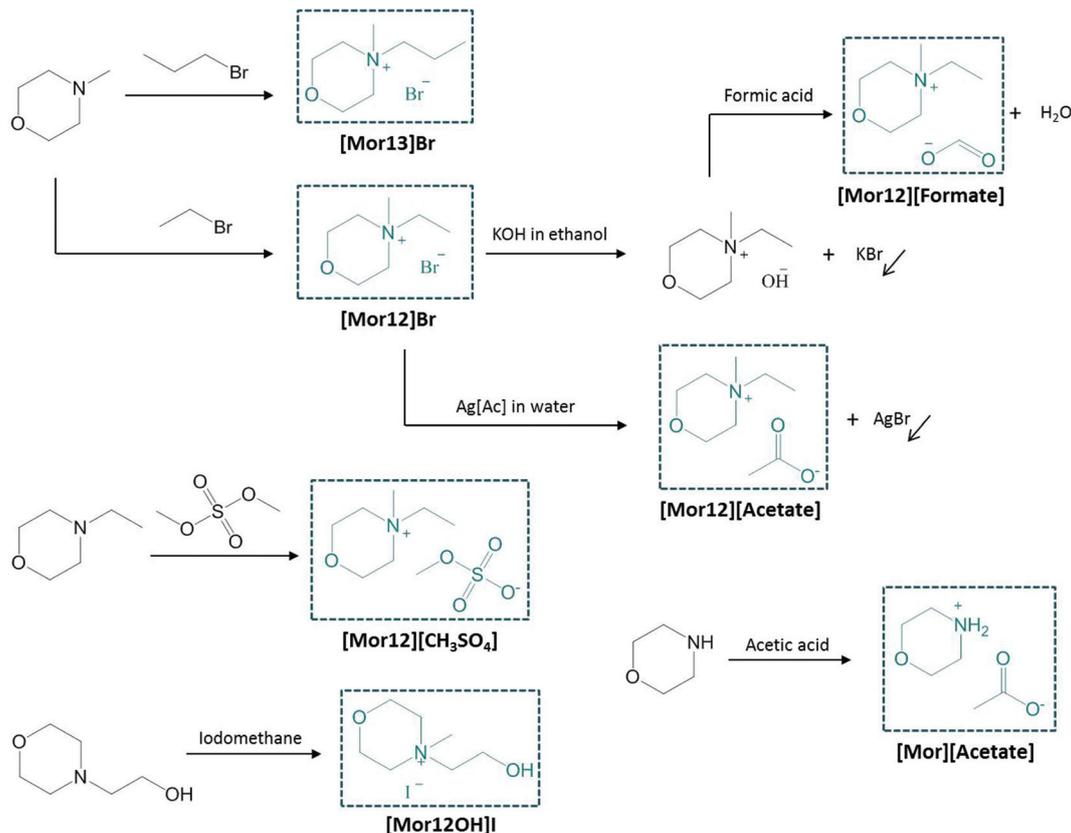
Fig. 1 Synthesis scheme and chemical structure of the morpholinium-based ILs synthesized specifically for this study.

tected from light for 1.5 h. The reaction flask was then placed into a water-ice bath in order to ensure the complete precipitation of silver bromide, which was later removed by filtration. Finally, the water was removed under reduced pressure and the obtained compound was dried under high vacuum for at least 48 h.[61] [Mor12][Acetate] was obtained as a white solid (92% of yield, 0.21 g). ^{1}H NMR (D$_2$O, 300 MHz, [ppm]): δ 1.39 (tt, J_{HH} = 7.3 Hz and J_{NH} = 1.9 Hz, 3H, NCH$_2$CH$_3$), 1.91 (s, 3H, COOCH$_3$), 3.17 (s, 3H, NCH$_3$), 3.42–3.61 (m, 6H, N(CH$_2$)$_3$), 3.97–4.10 (m, 4H, O(CH$_2$)$_2$). ^{13}C NMR (D$_2$O, 75.47 MHz, [ppm]): δ 6.45, 23.23, 45.94 (t, J_{CN}, NCH$_3$), 59.00 (t, J_{CN}, NCH$_2$CH$_3$), 60.33, 60.75, 181.26.

• *N*-Methyl-*N*-ethylmorpholinium formate, [Mor12] [Formate]. Firstly, a stoichiometric amount of potassium hydroxide was added to a solution of *N*-methyl-*N*-ethylmorpholinium bromide (2.4 mmol) prepared in ethanol. The solutions were stirred at room temperature for 0.5 h, after which the precipitated potassium bromide was removed by filtration. Then, a stoichiometric amount of formic acid was added to the filtrate. Again, the solutions were stirred overnight at room temperature, then the reaction flask was placed in a water-ice bath and the remaining inorganic salt was removed. Finally, the ethanol was removed under reduced pressure and the obtained compound was dried under high vacuum for at least 48 h.[62] [Mor12][Formate] was obtained as a white solid (95% of yield, 0.40 g). ^{1}H NMR (D$_2$O, 300 MHz, [ppm]): δ 1.37 (tt, J_{HH} = 7.3

J_{NH} = 1.8 Hz, 3H, NCH$_2$CH$_3$), 3.16 (s, 3H, NCH$_3$), 3.35–3.64 (m, 6H, N(CH$_2$)$_3$), 3.94–4.15 (m, 4H, O(CH$_2$)$_2$), 8.41 (s, 1H, CO$_2$H). ^{13}C NMR (D$_2$O, 75.47 MHz, [ppm]): δ 6.45, 45.96 (t, J_{CN}, NCH$_3$), 59.01 (t, J_{CN}, NCH$_2$CH$_3$), 60.34, 60.77, 170.18.

• *N*-Methyl-*N*-ethylmorpholinium methylsulfate, [Mor12][CH$_3$SO$_4$], was prepared by slow dropwise addition of 2.0 mL of dimethylsulfate (21.1 mmol) to a solution of 4-ethylmorpholine (19.0 mmol, 2.19 g) in toluene, at 0 °C, under a nitrogen atmosphere. The reaction mixture was stirred at room temperature under a nitrogen atmosphere and protected from light for 4 h. The obtained white solid was isolated by filtration and washed with ethyl acetate (2 × 15 mL). Finally, the residual solvent was removed under reduced pressure and the obtained compound was dried under high vacuum for at least 48 h.[63] [Mor12][CH$_3$SO$_4$] was obtained as a white solid (68% of yield, 3.10 g). ^{1}H NMR (D$_2$O, 300 MHz, [ppm]): δ 1.37 (tt, J_{HH} = 7.3 Hz and J_{NH} = 1.8 Hz, 3H, NCH$_2$CH$_3$), 3.15 (s, 3H, NCH$_3$), 3.37–3.61 (m, 6H, N(CH$_2$)$_3$), 3.73 (s, 3H, OCH$_3$), 3.97–4.11 (m, 4H, O(CH$_2$)$_2$). ^{13}C NMR (D$_2$O, 75.47 MHz, [ppm]): δ 6.43, 45.93 (t, J_{CN}, NCH$_3$), 55.38, 59.01 (t, J_{CN}, NCH$_2$CH$_3$), 60.34, 60.76.

• *N*-Methyl-*N*-propylmorpholinium bromide, [Mor13]Br, was prepared by dropwise addition of 4.5 mL of 1-bromopropane (49.5 mmol) to a solution of 4-methylmorpholine (44.6 mmol, 4.51 g) in ethyl acetate, at room temperature. The reaction mixture was refluxed and stirred at 55 °C, and protected from light overnight. After cooling, a solid was formed,

then filtered off and washed with ethyl acetate (3 × 15 mL). Finally, the residual solvent was removed under reduced pressure and the obtained compound was dried under high vacuum for at least 48 h.[60] [Mor13]Br was obtained as a white solid (43% of yield, 4.30 g). ^{1}H NMR (D$_2$O, 300 MHz, [ppm]): δ 1.00 (t, J_{HH} = 7.3 Hz, 3H, NCH$_2$CH$_2$CH_3), 1.74–1.92 (m, 2H, NCH$_2$CH_2CH$_3$), 3.19 (s, 3H, NCH_3), 3.35–3.64 (m, 6H, N(CH_2)$_3$), 3.98–4.13 (m, 4H, O(CH_2)$_2$). ^{13}C NMR (D$_2$O, 75.47 MHz, [ppm]): δ 9.73, 14.63, 46.74, 59.51, 60.34, 66.43.

• *N*-Methyl-*N*-hydroxyethylmorpholinium iodide, [Mor12OH]I, was prepared by dropwise addition of 2.0 mL of iodomethane (32.1 mmol) to a solution of 4-(2-hydroxyethyl)morpholine (21.6 mmol, 2.83 g) in ethyl acetate, at room temperature, under a nitrogen atmosphere. The reaction mixture was refluxed and stirred at 45 °C, under a nitrogen atmosphere and protected from light overnight. A solid was formed, which was filtered off and washed with ethyl acetate (3 × 25 mL). Finally, the residual solvent was removed under reduced pressure and the obtained compound was dried under high vacuum for at least 48 h.[60] [Mor12OH]I was obtained as a white solid (88% of yield, 5.18 g). ^{1}H NMR (d_6-DMSO, 300 MHz, [ppm]): δ 3.24 (s, 3H, NCH_3), 3.40–3.69 (m, 6H, N(CH_2)$_3$), 3.80–4.03 (m, 6H, O(CH_2)$_2$ and CH_2OH), 5.21 (t, J_{HH} = 4.6 Hz, 1H, OH). ^{13}C NMR (d_6-DMSO, 75.47 MHz, [ppm]): δ 48.30, 54.94, 60.21, 60.31, 65.08.

• Morpholinium acetate, [Mor][Acetate], was prepared by dropwise addition of 2 mL of acetic acid (35.0 mmol) to a solution of morpholine (35.0 mmol, 3.05 g) in ethyl acetate, at 0 °C. The reaction mixture was stirred at room temperature overnight. A solid was formed, which was filtered off and washed with ethyl acetate (2 × 20 mL). Finally, the residual solvent was removed under reduced pressure and the obtained compound was dried under high vacuum for at least 48 h.[64] [Mor][Acetate] was obtained as a white solid (97% of yield, 4.99 g). ^{1}H NMR (D$_2$O, 300 MHz, [ppm]): δ 1.92 (s, 3H, CH_3), 3.23–3.35 (m, 4H, N(CH_2)$_2$), 3.90–4.01 (m, 4H, O(CH_2)$_2$). ^{13}C NMR (D$_2$O, 75.47 MHz, [ppm]): δ 23.18, 42.95, 63.49, 181.17.

2.3.3 Standard Microtox® liquid-phase assays. In order to validate the predictive QSAR models herein reported, the ecotoxicity of the morpholinium-based ILs synthesized was evaluated using the Standard Microtox® liquid-phase assay. Microtox® is a bioluminescence inhibition method based on the bacterium *V. fischeri* (strain NRRL B-11177) luminescence after its exposure to each sample solution at 15 °C. The bacterium was exposed to a range of diluted aqueous solutions (from 0 to 81.9 wt%) of each tested compound, where 100% corresponds to a known concentration of a stock solution.[65] The light output of *V. fischeri* was measured after 15 and 30 minutes of exposure to each morpholinium based-IL, and compared with the light output of a blank control sample. Then, the corresponding 15 min- and 30 min-EC$_{50}$ values (estimated concentration yielding a 50% of inhibition effect) plus the corresponding 95% confidence intervals were estimated for each compound tested by non-linear regression, using the least-squares method to fit the data to the logistic equation.

3. Results and discussion

3.1 Developed QSAR models

The QSAR models developed in the present study follow the OECD guidelines as characterized by a uniformly defined response data (principle 1), explicitly described methodology (principle 2), suitable chemical domain of applicability (principle 3), statistical measures defining fitness, predictivity and robustness (principle 4), as well as interpretation of the captured chemical information (principle 5).[13] A total of six PLS models have been developed. Here, we have used the spline option of the GFA algorithm in order to account for the presence of any non-linear relationship along with the linear variables.

Table 1 presents the statistical quality of the developed equations and their external predictivity on the same test set, which were not used during model development. All models show acceptable quality in terms of fitness, stability and classical predictivity measures. Recently, it has been shown by Roy *et al.*[56] that the R^2 based criteria for model validation might be insufficient and misleading in some cases. Instead, mean absolute error (MAE) based criteria[66] have been proposed for a better understanding of the quality of predictions. Here, using the MAE-based judgement of model external predictivity, the external predictive quality of the first three models was characterized as 'moderate' while the remaining three as 'bad'. Interestingly, classical validation metrics such as $Q^2_{ext(F1)}$ and $Q^2_{ext(F2)}$ show acceptable quality of external predictions for all the models on the same test set data, but the MAE-based criteria penalize the last three models employing an error threshold using a range of training set responses.[66] Here, both $Q^2_{ext(F1)}$ and $Q^2_{ext(F2)}$ show overpredictivity for models 4, 5, and 6, since the response values of the test set observations lie away from the mean value of the training and test set responses. The summed predicted residual values obtained from these models were found to be higher than those obtained using models 1, 2 and 3 portraying relatively poor model performances by the models 4, 5, and 6. We have omitted the latter models from consensus predictions in order to obtain more precise prediction values based on models 1, 2 and 3 (Table S4 in the ESI†). However, we have retained models 4, 5, 6 in Table 1 as these models might still provide useful information about the structure–toxicity relationships in the form of uncommon descriptors not present in models 1, 2 and 3, and may hence be useful for designing new chemicals.

The number of latent variables (LVs) reported in Table 1 varies from three to seven, which is encouraging, considering the training set size comprising 213 chemicals.[67] Here, we explain the chemical information captured by models 1, 2, and 3, which are characterized as 'moderate' by the MAE based criteria. The reported models have captured information of both the cations and anions for the toxicity of ionic liquids towards *V. fischeri*. The repeated occurrence of the descriptor CATS2D_03_LL$_{(cation)}$ in models 1, 2, and 3 emphasizes its greater importance in encoding the toxicity of ionic liquids. This parameter is characterized by a spline function with a

Table 1 Predictive QSAR models developed using the ecotoxicity values of ionic liquids towards *Vibrio fischeri*. Here, $n_{\text{training}} = 213$, $n_{\text{test}} = 92$

Sl. no.	Predictive models	LVs	R^2	R_a^2	Q^2_{LOO}	$Q^2_{\text{ext (F1)}}$	$Q^2_{\text{ext (F2)}}$	MAE based criteria	Consensus prediction $Q^2_{\text{ext(F1)}}$	$Q^2_{\text{ext(F2)}}$	MAE based
1	$pEC_{50} = 5.390 - 0.354 \times < 6 - \text{CATS2D\_03\_LL} >_{(cation)}$ $- 1.083 \times \text{B05[C-O]}_{(cation)} + 0.007 \times \text{ZM1V}_{(cation)}$ $- 0.003 \times < 435.42 - \text{MW} >_{(total)}$ $- 0.304 \times \text{CATS2D\_05\_PL}_{(cation)}$ $- 0.827 \times \text{B02[C-S]}_{(anion)} - 17.292 \times < 0.328 - {}^2\chi_{avg} >_{(cation)}$ $+ 0.000 \times \text{GMTIV}_{(cation)} + 0.164 \times n\text{Cs}_{(anion)}$ $- 1.631 \times n\text{RCN}_{(cation)} + 0.474 \times \text{Cl} - 102_{(anion)}$ $- 0.534 \times \text{MCD}_{(cation)}$	5	0.746	0.740	0.711	0.696	0.696	Moderate	0.704	0.704	Moderate
2	$pEC_{50} = 5.665 - 0.315 \times < 6 - \text{CATS2D\_03\_LL} >_{(cation)}$ $- 2.305 \times \Sigma\beta'_{(cation)} + 0.903 \times I\text{arom}_{(cation)}$ $+ 0.875 \times < 1 - n\text{O} >_{(cation)} + 0.136 \times na_{\text{AromAtom(cation)}}$ $- 8.839 \times \Delta\varepsilon_{C(cation)} - 0.003 \times < 435.42 - \text{MW} >_{total}$ $+ 0.445 \times < \text{C} - 005 - 2 >_{(cation)}$ $- 16.397 \times < 0.328 - {}^2\chi_{avg} >_{(cation)} - 1.166 \times \lambda_{3(cation)}$	7	0.703	0.693	0.664	0.645	0.645	Moderate			
3	$pEC_{50} = 4.796 - 0.326 \times < 6 - \text{CATS2D\_03\_LL} >_{(cation)}$ $- 0.008 \times \text{ZM1V}_{(cation)} - 0.899 \times \text{B05[C-O]}_{(cation)}$ $- 0.0001 \times \text{GMTIV}_{(cation)} - 0.003 \times < 435.42 - \text{MW} >_{total}$ $+ 0.688 \times \text{Cl-102}_{(anion)} + 0.206 \times n\text{Cs}_{(anion)}$ $- 0.683 \times \text{B02[C-S]}_{(anion)} - 13.595 \times < 0.328$ $- {}^2\chi_{avg} >_{(cation)} - 0.557 \times \text{MCD}_{(cation)}$ $+ 0.234 \times V_{\text{index (anion)}} - 1.468 \times n\text{RCN}_{(cation)}$	5	0.748	0.742	0.710	0.656	0.656	Moderate			
4[a]	$pEC_{50} = 0.563 + 0.118 \times \text{H} - 046_{(cation)} - 0.122 \times (\log k_0)^2_{(cation)}$ $+ 0.419 \times \text{MSD}_{(cation)} + 1.601 \times \eta'_{F(cation)}$ $- 0.573 \times n\text{BR}_{(anion)} + 0.267 \times n\text{HM}_{(anion)}$ $+ 0.225 \times n\text{Cs}_{(anion)} - 0.325 \times \text{SdssC}_{(anion)}$ $- 0.598 \times {}^4\chi^v_{avg(cation)} - 4.054 \times {}^5\chi_{avg(cation)}$ $- 1.340 \times n\text{RCN} - 0.101 \times \text{C} - 006_{(anion)}$	7	0.697	0.687	0.655	0.635	0.635	Bad	—	—	—
5[a]	$pEC_{50} = 5.317 - 0.270 \times < 6 - \text{CATS2D\_03\_LL} >_{(cation)}$ $- 0.060 \times < 19.549 - (\text{ALOGP})^2 >_{(cation)}$ $+ 0.572 \times < \text{C-005} - 2 >_{(cation)} + 0.409 \times \text{N} - 075_{(cation)}$ $- 0.002 \times < 435.42 - \text{MW} > -0.626 \times n\text{Cq}_{(cation)}$ $- 1.037 \times n\text{I}_{(anion)} - 0.319 \times n\text{BR}_{(anion)} - 1.446 \times n\text{ROR}_{(anion)}$	3	0.705	0.701	0.677	0.609	0.609	Bad			
6[a]	$pEC_{50} = 5.172 - 0.361 \times < 6 - \text{CATS2D\_03\_LL} >_{(cation)}$ $- 0.067 \times (\log k_0)^2_{(cation)} - 0.004 \times < 435.42 - \text{MW}_{(total)} >$ $+ 0.034 \times \text{H} - 046_{(cation)} + 0.975 \times \text{Cl} - 102_{(anion)}$ $+ 1.339 \times \text{lap}_{(cation)} - 0.349 \times \text{H} - 053_{(cation)}$ $- 0.289 \times \text{SdssC}_{(anion)} + 0.260 \times V_{\text{index(anion)}}$ $+ 0.135 \times n\text{CS}_{(anion)} + 0.001 \times \text{MW}_{(cation)}$	5	0.718	0.711	0.689	0.609	0.609	Bad			

[a] Models showing 'bad' external predictivity as identified by the MAE based criteria.

knot value of 6 signifying that the spline term will vanish if the value of the descriptor CATS2D_03_LL$_{(cation)}$ is more than or equal to 6. The descriptor belongs to the pharmacophoric point pair based CATS2D group and designated as the 'lipophilic–lipophilic at lag 03' portraying the presence of two lipophilic atoms at topological distance 3 in cations. Considering the nature of the present dataset, the value of the descriptor was found to increase with the alkyl chain length of the cationic substituent. Now, the cationic substituents consist of carbon atoms that are considered as lipophilic in the CATS2D formalism. Hence, the parameter portrays information on cationic lipophilicity. An additional descriptor of the pharmacophore point pair type, namely CATS2D_05_PL$_{(cation)}$ in model 1, signifying the presence of a positively charged atom and a lipophilic atom at a distance of 5, also gives information on cationic lipophilicity.

In Table 1, a spline form of the descriptor MW$_{total}$ coding for the total molecular weight of an ionic liquid (considering both cations and anions) is present in the three models (i.e., models 1, 2 and 3). The spline function $<435.42 - $ MW$>_{total}$ determines that ionic liquids with a molecular weight value less than 435.42 will influence the toxicity response. The molecular weight gives a measure of the bulk of analyzed chemicals, which can be correlated with the hydrophobic behavior. The spline function is appended with a negative coefficient in the equations signifying that compounds with a lower molecular weight value will have a higher value of the spline term and a lower value of toxicity thereof compared to those having a higher MW$_{total}$ value. Hence, considering a knot value of 435.42, ionic liquids with a lower molecular weight will be less toxic than the ones with a higher molecular weight value. The compound [IM1,2]Cl (Sl. no. 137, pEC$_{50}$ = 1.450) with a $<435.42 - $ MW$_{total}>$ value of 288.78 is less toxic than [IM1,16] Cl (Sl. no. 211, pEC$_{50}$ = 5.770) having a value of 92.36 for the same parameter. The descriptor nCs$_{(anion)}$ encoding the total number of secondary sp^3 hybridized carbon atoms, i.e., a –CH$_2$ group in anions is present in models 1 and 3 and highlights the contribution of non-halogenated organic anions towards toxicity.

The higher toxicity value of the compound [2-HDEA][Pe] (Sl. no. 120, pEC$_{50}$ = 2.772) compared to that of [2-HDEA][Pr] (Sl. no. 113, pEC$_{50}$ = 2.440) is due to the possession of a higher number of –CH$_2$ groups in the pentanoate anion ([Pe]) of the former compound than propionate ([Pr]) in the latter. Another anionic descriptor Cl-102$_{(anion)}$ present in models 1 and 3 emphasizes the contribution of the chloride anion towards toxicity. The occurrence of the cationic descriptor nRCN$_{(cation)}$ with a negative coefficient in models 1 and 3 suggests a negative impact of the count of aliphatic nitrile groups in cations on the toxicity. The presence of a nitrile group in cations incorporates H-bonding behavior to the molecule and hence reduced toxicity due to a decreased lipophilic nature. The compound [PYR4,3CN]Br (Sl. no. 280, pEC$_{50}$ = 0.670) is less toxic than [PYR4]Cl (Sl. no. 282, pEC$_{50}$ = 1.700) since the former contains a nitrile functionality in its cationic moiety. Another cationic descriptor with repeated occurrence is $<0.328 - $

$^2\chi_{avg}>_{(cation)}$ (see models 1, 2, and 3), which encodes information on branching. The descriptor $^2\chi_{avg}$ is the average second-order connectivity index defining the importance of the cationic branchedness in a non-linear i.e., spline fashion. A few other descriptors describing the cationic features viz. MCD$_{(cation)}$, B05[C–O]$_{(cation)}$, ZM1V$_{(cation)}$, GMTIV$_{(cation)}$, <C-005 − 2>$_{(cation)}$, <1 − nO>$_{(cation)}$, $\lambda_{3(cation)}$, $\Delta\varepsilon_{C(cation)}$, $\sum\beta'_{(cation)}$ as well as anionic attributes viz. B02[C–S]$_{(anion)}$ and V$_{index(anion)}$ are observed to be present in models 1, 2, and 3.

The descriptor ZM1V is the valence-vertex-degree-based first Zagreb index while GMTIV is the valence vertex degree based Gutman Molecular Topological Index, both of which define the molecular branchedness. MCD is the molecular cyclized degree denoting the importance of acyclic groups and substituents along with cyclic moieties. The descriptor C-005 codes for the atom type fragment of CH$_3$X type (X is any heteroatom), i.e., for the given dataset it provides information on the role of methyl groups attached to heteroatoms, e.g., N of imidazolium, ammonium, pyrrolidinium and other systems. Another cationic descriptor B05[C–O] shows the presence or absence of C–O at the topological distance of 5 bonds and thereby it provides information on the H-bonding nature of the oxygenated substituents on cations leading to reduced toxicity of ILs principally by modifying the system lipophilicity. The impact of cationic oxygen atoms towards the toxicity of ILs is further emphasized by the spline parameter <1 − nO>$_{(cation)}$ where nO represents the number of oxygen atoms. Model 2 additionally captures information on cationic aromaticity (I_{arom}, $na_{AromAtom}$), and electronic distribution for the ecotoxicity of ionic liquids to V. fischeri. The information on the electronic distribution of cations is also shown by the ETA indices $\Delta\varepsilon_C$ $_{(cation)}$, $\sum\beta'_{(cation)}$ and the QTMS variable λ_3 (model 2). Among the anionic descriptors, V_{index} is the Balaban V index, which defines the role of anionic branchedness and B02[C–S] depicts the role of sulfated and sulfonated anions for the ecotoxicity of ILs towards V. fischeri.

Hence, a chemical interpretation of the descriptors portrays that the toxicity of ionic liquids towards V. fischeri is monitored by features such as lipophilicity, hydrogen bonding propensity, branching, aromaticity, and electronegativity. While a parameter such as MW$_{total}$ shows the impact of lipophilicity as a whole, the descriptors viz. CATS2D_03_LL, CATS2D_05_PL, MCD, ZM1V, GMTIV, etc. provide further insight into the pattern of side chain substituents as well as branching of cations. The lipophilicity attribute of anions was also observed to play a major role in the descriptors nCs and Cl-102. The presence of hydrogen bonding groups on cationic side chains, e.g., groups containing O, N, etc. were also observed to influence ionic liquid toxicity towards V. fischeri as encoded by the descriptors B05[C–O], nRCN, nO, C-005, etc.

3.2 True external validation

The true external prediction was performed on a separate dataset of new eight ionic liquids (which are not common to the 305 ionic liquids used for the development of the QSAR models) using models 1, 2 and 3, which showed encouraging

Table 2 Predictive quality of the models for the true external validation set (n = 8) by employing classic metrics and MAE based criteria

	$Q^2_{ext(F1)}$	$Q^2_{ext(F2)}$	MAE based criteria
Model 1	0.605	−1.738	Moderate
Model 2	0.883	0.190	Good
Model 3	0.652	−1.408	Moderate
Predictions from the best model[a]	0.736	−0.826	Good
Consensus predictions	0.743	−0.783	Good

[a] The best model corresponds to the one showing the lowest value of DModX with respect to a particular query compound.

values of the classical external validation metrics as well as the MAE based judgment criteria as portrayed in Table 2. The reliability of predictions for these eight molecules was also verified using the chemical domain of the models (models 1, 2, and 3) by employing the DModX approach.

Table 3 gives the predicted toxicity values (and the experimental values) of the true external set based on the lowest DModX value out of the three models. We have also determined the consensus prediction values for the compounds, which are reported in Table 3. It may be noted that predicted toxicity values obtained using the best model (model with the lowest DModX value with respect to a specific observation) as well as the consensus approach are characterized as 'good' by the MAE based criteria (Table 2) although the classical external validation metric $Q^2_{ext(F2)}$ fails to portray the acceptable predictivity. Here, we have found that the absolute values of the predicted residuals for all the eight observations obtained from models 1, 2 and 3 are less than 0.2 times the training set range (where 0.2 × training range = 1.206) signifying good predictions. Only one observation showed a predicted residual (absolute) value more than 0.15 times the training set range (where 0.15 × training range = 0.905) with respect to models 1 and 3. Accordingly, the predictive quality was judged 'moderate' by the MAE based criteria for these two models. However, the $Q^2_{ext(F2)}$ metric renders the models underpredictive because of the low range of the response of the true external

Table 3 Experimental and predicted ecotoxicity values for the true external set of ionic liquids obtained from the best model (with the lowest DModX value) and consensus approach

Sl. no.	Ionic liquids	Expt. pEC$_{50}$	Predicted pEC$_{50}$		
			Consensus approach[a]	Using the best model	Best model no.
1	2-HEAA	1.8100	2.4376	2.5593	1
2	2-HEAPr	1.8900	2.5651	2.7654	1
3	2-HEAiB	1.9700	2.4824	2.6435	1
4	2-HEAPe	2.0400	2.8688	2.3434	2
5	2-HTEAF	1.7400	2.3491	2.1686	1
6	2-HTEAA	2.6500	2.3561	2.2107	1
7	2-HTEAPr	2.7300	2.4836	2.4910	3
8	2-HTEAiB	2.8100	2.4010	2.3229	3

[a] Based on models 1, 2 and 3.
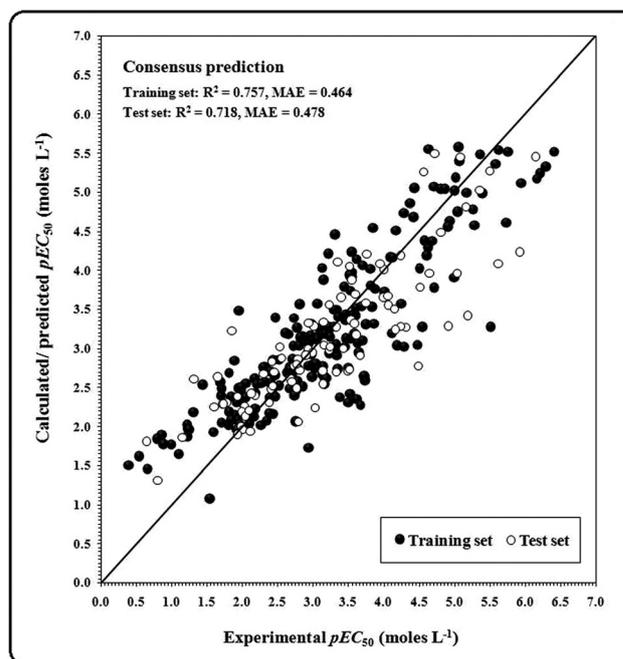
Fig. 2 Scatter plot of the observed *versus* computed (consensus from models 1, 2, 3) toxicity values of ILs.

set (1.07 log unit) where most of the compounds are close to the mean response of the set, which means that the mean can perform better than the model. A scatter plot of the observed *versus* computed (consensus from models 1, 2, and 3) toxicity values of ILs is shown in Fig. 2. In a recent study, some of us[68] have shown that in any predictive modeling analysis, the observations lying close to the mean are more reliably predicted than those lying towards extremities, which may suffer from the trouble of over- or under-predictive attributes. Here, the average pEC$_{50}$ for the training set compounds is 3.221 and the observed pEC$_{50}$ values of the designed compounds lie towards the lower range. The experimental toxicity values of the designed compounds being low, the corresponding predicted values appear to be somewhat higher because of the relatively high value of the mean response (3.221) of the training set, on the basis of which the models have been developed. However, the predicted response values are reliable considering the wide response domain of the training set, which has been used for the development of models and this reliability is also evident from the fact that the MAE-based criteria are satisfactorily met.

3.3 Design, synthesis, and evaluation of new ionic liquids

Based on the derived chemical information from the predictive *in silico* modeling analysis, we have designed a series of twenty new "low toxic" or harmless ionic liquids within the chemical domain of the developed models (models 1, 2 and 3) which is reported in Table 4. Table 4 also shows the predicted toxicity values of all the designed ionic liquids obtained using DModX based best model predictions and consensus predictions.

Table 4 Predicted ecotoxicity values of the designed ionic liquids towards *V. fischeri* determined using the best model (model with the lowest DModX value) and the consensus approach

Sl. no.	Designed ionic liquids		Predicted pEC_{50}		
	Cation	Anion	Consensus approach	Using best model	Best model no.
1	[MOR]	[Acetate]	2.0733	1.7458	2
2	[MOR1,2]	[Formate]	1.3866	1.3544	2
3	[MOR1,2]	[Acetate]	1.3935	1.3965	2
4	[MOR1,2]	[CH$_3$SO$_4$]	1.0161	1.5527	2
5	[MOR1,2]	Br	1.3319	1.4591	2
6	[MOR1,2OH]	I	1.8729	1.8645	1
7	[MOR1,3]	Br	1.5160	1.6001	2
8	[MOR1,2]	[SCN]	1.4258	1.3936	2
9	[MOR1,3]	[Formate]	1.5706	1.4954	2
10	[MOR1,3]	[Propionate]	1.7051	1.5796	2
11	[MOR1,3]	[CH$_3$SO$_4$]	1.2001	1.6937	2
12	[IM1,1OH]	[Acetate]	1.7109	1.7907	2
13	[IM1,1OH]	[i-Butyrate]	1.7558	1.8749	2
14	[IM1,1OH]	[CH$_3$SO$_4$]	1.3334	1.9469	2
15	[N1,1,2OH]	[Butyrate]	2.6916	1.9335	2
16	[PYD2OH]	[Formate]	1.7451	1.6763	2
17	[PYD2OH]	[Acetate]	1.7521	1.7183	2
18	[PYD2OH]	[Propionate]	1.8796	1.7604	2
19	[PYD2OH]	[SCN]	1.7843	1.7155	2
20	[PYD2OH]	[CH$_3$SO$_4$]	1.3746	1.8745	2

Out of these twenty ionic liquids, seven were synthesized and experimentally tested for their toxicity potential to *V. fischeri*. The toxicities for these seven new compounds are reported in Table 5. According to the obtained results (Table 5), it is possible to categorize these compounds as belonging to the Category "Acute III" according to the European Classification,[69] as (1) "practically harmless" ([MOR1,2][CH$_3$SO$_4$] and [MOR1,2][Formate] with 100 mg L^{-1} < EC$_{50}$ < 1000 mg L^{-1}) and as (2) "harmless" ([MOR1,2OH]I, [MOR1,2][Acetate], [MOR1,2]Br, [MOR1,3]Br and [MOR1,2] [Acetate] with EC$_{50}$ > 1000 mg L^{-1}). Thus, all the seven designed and subsequently synthesized compounds portrayed an ecotoxicity potential categorizable as "harmless" or "practically harmless" as expected from the developed QSAR models. It is sometimes more relevant for a good model to appropriately categorize the test chemicals as toxic or non-toxic, and to

Table 5 Experimental ecotoxicity values of the selected synthesized ionic liquid compounds

Sl. no.	Designed ionic liquids		Expt. EC$_{50}$ (mg L^{-1})		
	Cation	Anion	5 min	15 min	30 min
1	[MOR]	[Acetate]	3091.67	3221.95	2599.86
2	[MOR1,2]	[Formate]	224.85	231.18	236.41
3	[MOR1,2]	[Acetate]	5564.58	5322.99	4459.70
4	[MOR1,2]	[CH$_3$SO$_4$]	669.32	659.30	653.13
5[a]	[MOR1,2]	Br	—	—	—
6	[MOR1,2OH]	I	18659.24	17502.09	16631.66
7	[MOR1,3]	Br	—	34633.57	23985.45

[a] The EC$_{50}$ value was not measurable using a stock solution of 61760 mg L^{-1} showing the very low toxicity of [MOR12]Br to the bacterium *V. fischeri*.

maintain a correct rank order prediction, rather than to deliver quantitatively precise predictions.[70] In our present study, the developed models could successfully predict the designed chemicals as "harmless" or "practically harmless".

Summing up, this work allowed the development of predictive models with good predictability performance considering the ionic liquids' chemical structure and their associated toxicity. In the near future this work will benefit the industrial dissemination of safer ionic liquids.

4. Conclusions

Ionic liquids are neoteric solvents with wide industrial applicability. However, comprehensive assessment of their hazardous outcome is necessary to assure their safe use. Considering the ethical issues associated with biological experimentation on living beings, predictive *in silico* modeling provides a rational alternative strategy for prioritizing the chemicals. The present study involves *in silico* modelling of the largest toxicity dataset of ionic liquids to *Vibrio fischeri* currently available. Here, we have developed predictive PLS models using topological and quantum chemical descriptors. The chief aim of this study has been to develop multiple models capturing chemical information, enabling us to design and prepare new ionic liquids with reduced toxicity profile. The whole study has been performed in consonance with the OECD guidelines in terms of dataset selection, model development, applicability domain determination, model validation, and mechanistic interpretation of the diagnosed chemical attributes.

It was very interesting to observe that the classical external validation metrics were unable to portray poor model performance in three cases. By using our newly developed MAE based judgment criteria, we have selected three suitable models, which have been explored further for true external validation as well as the design of new analogues. The synthesis and experimental determination of toxicity of the newly designed ionic liquids were carried out following standard protocols. Note that this is the first attempt to perform both true external validation and experimental validation of QSPR models for toxicity of ionic liquids to *V. fischeri*. The designed ionic liquids were experimentally confirmed to be "harmless" or "practically harmless" as defined in the acute toxicity determination criteria by the European Commission.[71] Hence, these newly designed and synthesized ionic liquids can be considered as 'greener' analogues, beneficial for industrial use.

Conflict of interest

Declared none.

Acknowledgements

References

1 F. Yan, Q. Shang, S. Xia, Q. Wang and P. Ma, Application of topological index in predicting ionic liquids densities by the quantitative structure property relationship method, *J. Chem. Eng. Data*, 2015, **60**, 734–739.

2 S. Khooshechin, Z. Dashtbozorgi, H. Golmohammadi and W. E. Acree Jr., QSPR prediction of gas-to-ionic liquid partition coefficient of organic solutes dissolved in 1-(2-hydroxyethyl)-1-methylimidazolium tris(pentafluoroethyl) trifluorophosphate using the replacement method and support vector regression, *J. Mol. Liq.*, 2014, **196**, 43–51.

3 M. Sattari, A. Kamari, A. H. Mohammadi and D. Ramjugernath, Prediction of refractive indices of ionic liquids - a quantitative structure-property relationship based model, *J. Taiwan Inst. Chem. Eng.*, 2015, **52**, 165–180.

4 F. Yan, S. Xia, Q. Wang, Q. Shang and P. Ma, Predicting the glass transition temperature of ionic liquids by the quantitative structure property relationship method using a topological index, *Fluid Phase Equilib.*, 2013, **358**, 166–171.

5 N. V. Plechkova and K. R. Seddon, Applications of ionic liquids in the chemical industry, *Chem. Soc. Rev.*, 2008, **37**, 123–150.

6 Y. Zhao, Y. Huang, X. Zhang and S. Zhang, A quantitative prediction of the viscosity of ionic liquids using S[sigma]-profile molecular descriptors, *Phys. Chem. Chem. Phys.*, 2015, **17**, 3761–3767.

7 K. Roy, S. Kar and R. N. Das, *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*, Academic Press, 2015.

8 J. C. Dearden, The history and development of quantitative structure-activity relationships (QSARs), *Int. J. Quant. Struct.–Property Relationships (IJQSPR)*, 2016, **1**, 1–44.

9 J. A. P. Coutinho, P. J. Carvalho and N. M. C. Oliveira, Predictive methods for the estimation of thermophysical properties of ionic liquids, *RSC Adv.*, 2012, **2**, 7322–7346.

10 R. Das and K. Roy, Advances in QSPR/QSTR models of ionic liquids for the design of greener solvents of the future, *Mol. Diversity*, 2013, **17**, 151–196.

11 T. P. T. Pham, C.-W. Cho and Y.-S. Yun, Environmental fate and toxicity of ionic liquids: a review, *Water Res.*, 2010, **44**, 352–372.

12 N. Fjodorova, M. Novich, M. Vrachko, V. Smirnov, N. Kharchevnikova, Z. Zholdakova, S. Novikov, N. Skvortsova, D. Filimonov and V. Poroikov, Directions in QSAR modeling for regulatory uses in OECD member countries, EU and in Russia, *J. Environ. Sci. Health, Part C*, 2008, **26**, 201–236.

13 OECD, *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*, OECD Series on Testing and Assessment, No. 69, OECD Publishing, Paris, 2014, DOI: 10.1787/9789264085442-en.

14 M. Pattard and H. Moser, Luminescent bacteria test, in *Ecotoxicological Characterization of Waste*, ed. H. Moser and J. Römbke, Springer, New York, 2009, pp. 105–115.

15 C. Ledda, V. Rapisarda, M. Bracci, L. Proietti, M. Zuccarello, R. Fallico, M. Fiore and M. Ferrante, Professional exposure to basaltic rock dust: assessment by the *Vibrio fischeri* ecotoxicological test, *J. Occup. Med. Toxicol.*, 2013, **8**, 23.

16 F. Yan, Q. Shang, S. Xia, Q. Wang and P. Ma, Topological study on the toxicity of ionic liquids on Vibrio fischeri by the quantitative structure–activity relationship method, *J. Hazard. Mater.*, 2015, **286**, 410–415.

17 S. Ma, M. Lv, F. Deng, X. Zhang, H. Zhai and W. Lv, Predicting the ecotoxicity of ionic liquids towards Vibrio fischeri using genetic function approximation and least squares support vector machine, *J. Hazard. Mater.*, 2015, **283**, 591–598.

18 C. W. Cho, J. Ranke, J. Arning, J. Thöming, U. Preiss, C. Jungnickel, M. Diedenhofen, I. Krossing and S. Stolte, *In silico* modelling for predicting the cationic hydrophobicity and cytotoxicity of ionic liquids towards the Leukemia rat cell line, *Vibrio fischeri* and *Scenedesmus vacuolatus* based on molecular interaction potentials of ions, *SAR QSAR Environ. Res.*, 2013, **24**, 863–882.

19 R. N. Das and K. Roy, Development of classification and regression models for *Vibrio fischeri* toxicity of ionic liquids: green solvents for the future, *Toxicol. Res.*, 2012, **1**, 186–195.

20 S. Bruzzone, C. Chiappe, S. E. Focardi, C. Pretti and M. Renzi, Theoretical descriptor for the correlation of aquatic toxicity of ionic liquids by quantitative structure-toxicity relationships, *Chem. Eng. J.*, 2011, **175**, 17–23.

21 Á. Irabien, A. Garea and P. Luis, Hybrid molecular QSAR model for toxicity estimation: application to ionic liquids, *Comput. – Aided Chem. Eng.*, 2009, **26**, 63–67.

22 P. Luis, I. Ortiz, R. Aldaco and A. Irabien, A novel group contribution method in the development of a QSAR for predicting the toxicity (*Vibrio fischeri* EC 50) of ionic liquids, *Ecotoxicol. Environ. Saf.*, 2007, **67**, 423–429.

23 M. Alvarez-Guerra and A. Irabien, Design of ionic liquids: an ecotoxicity (*Vibrio fischeri*) discrimination approach, *Green Chem.*, 2011, **13**, 1507–1516.

24 P. J. Carvalho, S. P. M. Ventura, M. L. S. Batista, B. Schröder, F. Gonçalves, J. Esperança, F. Mutelet and J. A. P. Coutinho, Understanding the impact of the central atom on the ionic liquid behavior: phosphonium vs ammonium cations, *J. Chem. Phys.*, 2014, **140**, 064505.

25 K. A. Kurnia, T. E. Sintra, C. M. S. S. Neves, K. Shimizu, J. N. C. Lopes, F. Gonçalves, S. P. M. Ventura, M. G. Freire, L. M. N. B. F. Santos and J. A. P. Coutinho, The effect of the cation alkyl chain branching on mutual solubilities with water and toxicities, *Phys. Chem. Chem. Phys.*, 2014, **16**, 19952–19963.

26 P. Luis, A. Garea and A. Irabien, Quantitative structure–activity relationships (QSARs) to estimate ionic liquids ecotoxicity EC 50 (*Vibrio fischeri*), *J. Mol. Liq.*, 2010, **152**, 28–33.

27 F. A. e Silva, F. Siopa, B. F. H. T. Figueiredo, A. M. M. Gonçalves, J. L. Pereira, F. Gonçalves, J. A. P. Coutinho, C. A. M. Afonso and S. P. M. Ventura, Sustainable design for environment-friendly mono and di-cationic cholinium-based ionic liquids, *Ecotoxicol. Environ. Saf.*, 2014, **108**, 302–310.

28 M. Taha, F. A. e Silva, M. V. Quental, S. P. M. Ventura, M. G. Freire and J. A. P. Coutinho, Good's buffers as a basis for developing self-buffering and biocompatible ionic liquids for biological research, *Green Chem.*, 2014, **16**, 3149–3159.

29 S. P. M. Ventura, R. L. Gardas, F. Gonçalves and J. A. P. Coutinho, Ecotoxicological risk profile of ionic liquids: octanol-water distribution coefficients and toxicological data, *J. Chem. Technol. Biotechnol.*, 2011, **86**, 957–963.

30 S. P. M. Ventura, A. M. M. Gonçalves, F. Gonçalves and J. A. P. Coutinho, Assessing the toxicity on [C3mim][Tf2N] to aquatic organisms of different trophic levels, *Aquat. Toxicol.*, 2010, **96**, 290–297.

31 S. P. M. Ventura, C. S. Marques, A. A. Rosatella, C. A. M. Afonso, F. Gonçalves and J. A. P. Coutinho, Toxicity assessment of various ionic liquid families towards *Vibrio fischeri* marine bacteria, *Ecotoxicol. Environ. Saf.*, 2012, **76**, 162–168.

32 S. P. M. Ventura, A. M. M. Gonçalves, T. Sintra, J. L. Pereira, F. Gonçalves and J. A. P. Coutinho, Designing ionic liquids: the chemical structure role in the toxicity, *Ecotoxicology*, 2013, **22**, 1–12.

33 S. P. M. Ventura, M. Gurbisz, M. Ghavre, F. M. M. Ferreira, F. Gonçalves, I. Beadham, B. Quilty, J. A. P. Coutinho and N. Gathergood, Imidazolium and pyridinium ionic liquids from mandelic acid derivatives: Synthesis and bacteria and algae toxicity evaluation, *ACS Sustainable Chem. Eng.*, 2013, **1**, 393–402.

34 S. P. M. Ventura, F. A. e Silva, A. M. M. Gonçalves, J. L. Pereira, F. Gonçalves and J. A. P. Coutinho, Ecotoxicity analysis of cholinium-based ionic liquids to *Vibrio fischeri* marine bacteria, *Ecotoxicol. Environ. Saf.*, 2014, **102**, 48–54.

35 S. Viboud, N. Papaiconomou, A. Cortesi, G. Chatel, M. Draye and D. Fontvieille, Correlating the structure and composition of ionic liquids with their toxicity on *Vibrio fischeri*: a systematic study, *J. Hazard. Mater.*, 2012, **215–216**, 40–48.

36 C. Wang, Z. Wei, L. Wang, P. Sun and Z. Wang, Assessment of bromide-based ionic liquid toxicity toward aquatic organisms and QSAR analysis, *Ecotoxicol. Environ. Saf.*, 2015, **115**, 112–118.

37 K. Roy and R. N. Das, in *Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment*, ed. K. Roy, IGI Global, USA, 2015, pp. 180–211.

38 R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics*, John Wiley & Sons, 2009.

39 S. E. O'Brien and P. L. A. Popelier, Quantum topological molecular similarity. Part 4. A QSAR study of cell growth inhibitory properties of substituted (E)-1-phenylbut-1-en-3-ones, *J. Chem. Soc., Perkin Trans. 2*, 2002, 478–483.

40 U. A. Chaudry and P. L. A. Popelier, Ester hydrolysis rate constant prediction from quantum topological molecular similarity descriptors, *J. Phys. Chem. A*, 2003, **107**, 4578–4582.

41 K. Roy and P. L. A. Popelier, Chemometric modeling of the chromatographic lipophilicity parameter $logk_0$ of ionic liquid cations with ETA and QTMS descriptors, *J. Mol. Liq.*, 2014, **200 Part B**, 223–228.

42 B. Peric, J. Sierra, E. Martí, R. Cruañas and M. A. Garau, Quantitative structure–activity relationship (QSAR) prediction of (eco)toxicity of short aliphatic protic ionic liquids, *Ecotoxicol. Environ. Saf.*, 2015, **115**, 257–262.

43 B. S. Everitt, S. Landau and M. Leese, *Cluster analysis*, Arnold, London, 2001.

44 G. W. Snedecor and W. G. Cochran, *Statistical methods*, Oxford and IBH Publishing Co, New Delhi, 1967.

45 S. Wold, M. Sjöström and L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109–130.

46 D. Rogers and A. J. Hopfinger, Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 854–866.

47 J. Friedman, *Multivariate adaptive regression splines, technical report No. 102, Laboratory for Computational Statistics*, Department of Statistics, Stanford University, Stanford, CA, 1988 (revised August 1990).

48 R. B. Darlington, *Regression and linear models*, McGraw-Hill, New York, 1990.

49 G. Domenico, M. Giuseppe Felice, C. Marco, C. Angelo and N. Orazio, Applicability domain for QSAR models: where theory meets reality, *Int. J. Quant. Struct.–Property Relationships (IJQSPR)*, 2016, **1**, 45–63.

50 *MarvinSketch, version 15.12.7*, ChemAxon Ltd, 2016, accessible at http://www.chemaxon.com.

51 *Dragon, version 6*, TALETE srl, Italy, 2010, accessible at http://www.talete.mi.it/products/dragon_description.htm.

52 C. W. Yap, PaDEL-Descriptor: an open source software to calculate molecular descriptors and fingerprints, *J. Comput. Chem.*, 2011, **32**, 1466–1474.

53 *GaussView, 4.1*, Semichem Inc.,Gaussian Inc., Pittsburgh, PA, USA, 2003.

54 M. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr.,

T. Vreven, K. N. Kudin and J. C. Burant, *Gaussian 03, revision D. 01*, Gaussian Inc., Wallingford, CT, 2004, vol. 26.

55 P. L. A. Popelier, MORPHY, a program for an automated "atoms in molecules" analysis, *Comput. Phys. Commun.*, 1996, **93**, 212–240.

56 *SPSS, version 9.0.0*, SPSS Inc., USA, 1998, accessible at http://www.spss.co.in/.

57 *Cerius², version 4.10*, Accelrys Inc., San Diego, CA, USA, 2005, accessible at http://accelrys.com/.

58 *Version 14.13*, Minitab, Inc., USA, 2004, accessible at http://www.minitab.com/en-US/default.aspx.

59 *SIMCA-P, version 10.0*, Umetrics, UMEA, Sweden, 2002, accessible at http://umetrics.com/.

60 M. Lorenzo, M. Vilas, P. Verdia, M. Villanueva, J. Salgado and E. Tojo, Long-term thermal stabilities of ammonium ionic liquids designed as potential absorbents of ammonia, *RSC Adv.*, 2015, **5**, 41278–41284.

61 J. S. Wilkes and M. J. Zaworotko, Air and water stable 1-ethyl-3-methylimidazolium based ionic liquids, *J. Chem. Soc., Chem. Commun.*, 1992, 965–967.

62 J. Pernak, N. Borucka, F. Walkiewicz, B. Markiewicz, P. Fochtman, S. Stolte, S. Steudte and P. Stepnowski, Synthesis, toxicity, biodegradability and physicochemical properties of 4-benzyl-4-methylmorpholinium-based ionic liquids, *Green Chem.*, 2011, **13**, 2901–2910.

63 J. D. Holbrey, W. M. Reichert, R. P. Swatloski, G. A. Broker, W. R. Pitner, K. R. Seddon and R. D. Rogers, Efficient, halide free synthesis of new, low cost ionic liquids: 1,3-dialkylimidazolium salts containing methyl-

and ethyl-sulfate anions, *Green Chem.*, 2002, **4**, 407–413.

64 C. K. Lee, H. W. Huang and I. J. B. Lin, Simple amphiphilic liquid crystalline-alkylimidazolium salts. A new solvent system providing a partially ordered environment, *Chem. Commun.*, 2000, 1911–1912.

65 *Environmental, Azur MicrotoxOmni Software for Windows 95/98 NT-User Manual*, 1999.

66 K. Roy, R. N. Das, P. Ambure and R. B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, *Chemom. Intell. Lab. Syst.*, 2016, **152**, 18–33.

67 J. G. Topliss and R. P. Edwards, Chance factors in studies of quantitative structure–activity relationships, *J. Med. Chem.*, 1979, **22**, 1238–1244.

68 K. Roy, P. Ambure and R. B. Aher, Neglecting systematic error in predictions and statistical applicability domain: A recently emerged misconception in the QSAR literature, *Chemom. Intell. Lab. Syst.*, 2016 (under review).

69 D. R. M. Passino and S. B. Smith, Acute bioassays and hazard evaluation of representative contaminants detected in great lakes fish, *Environ. Toxicol. Chem.*, 1987, **6**, 901–907.

70 D. L. J. Alexander, A. Tropsha and D. A. Winkler, Beware of R^2: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models, *J. Chem. Inf. Model.*, 2015, **55**, 1316–1322.

71 EU, Environmental Hazards 10 March 2011, http://www.unece.org/fileadmin/DAM/trans/danger/publi/ghs/ghs_rev01/English/04e_part4.pdf, (accessed 22.04.16).