New Journal and we have not received input yet 20 (2021) 100181

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.sciencedirect.com/journal/computational-toxicology

QSTR and interspecies-QSTR modelling for aquatic toxicity data gap filling of cationic polymers☆

Pathan Mohsin Khan [a], Hans Sanderson [b], Kunal Roy [c,*]

[a] National Institute of Pharmaceutical Education and Research-Kolkata, Kolkata 700054, India
[b] Aarhus University, Department of Environmental Science, 4000 Roskilde, Denmark
[c] Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

ARTICLE INFO

Keywords:
QSTR
i-QSTR
Cationic polymers
PLS
Fish toxicity
Algae toxicity
D. magna toxicity

ABSTRACT

Polymers are extensively used in several fields representing a growing multi-billion dollar industry covering several thousands of materials and billions of kilos used globally every year. The widespread use of cationic polymers may significantly discharge such compounds into the aquatic environment, which may cause potential toxic effects on aquatic organisms. The amount of publicly available, high-quality environmental toxicity data for industrial polymers such as cationic polyquaterniums is low. We have developed here individual quantitative structure–toxicity relationship (QSTR) models for toxicity prediction against fish and algae. These models against fish and algae showed optimistic statistical quality in terms of several internal and external quality and validation metrics such as determination coefficient R^2 (0.703 and 0.676), cross-validated leave-one-out Q^2 (0.638 and 0.516) and predictive R^2_{pred} or Q^2_{ext} (0.776 and 0.703) for fish ($N_{train} = 72$, $N_{test} = 23$) and algae ($N_{train} = 40$, $N_{test} = 14$) toxicity datasets, respectively. The study revealed that higher charge density increases the toxicity against both the response endpoints. However, a higher percentage of oligomers with a molecular mass of lower than 1000 Daltons results in a decreased toxicity towards both the studied endpoints. Similarly, primary amines in the molecular building block result in a reduction in the toxicity against the algal species. However, acceptable individual QSTR models against *D. magna* could not be generated with the limited feature information obtained from the United States Environmental Protection Agency and additional data provided by Environmental Climate Change Canada (ECCC). Therefore, we have also proposed interspecies quantitative structure–toxicity relationship (i-QSTR) models among three species (*D. magna*, fish and algae species) to bridge the toxicity data gap for cationic polymers. The mechanistic interpretation of i-QSTR models revealed several important characteristic features of polymers along with the experimental response of one species which are also helpful for toxicity prediction of other species, ultimately helping to reduce the experimental testing for toxicity prediction. Finally, the proposed QSTR and i-QSTR models can be helpful to compute the toxicity of polymers in the early stages of screening for regulatory purposes and data gap filling for new or untested polymers falling within the applicability domain of the models.

1. Introduction

Polymers are diverse classes of synthetic organic substances that are made up of a sequence of one or more types of small molecules (monomers) linked together by covalent bonds [1–3]. They are also known as macromolecules because they are made up of repeating subunits with MWs ranging from 100s to several 1000s or even millions of Daltons. Polymers are extensively used in several fields such as packaging, electrical and electronic equipment, transportation, controlled drug delivery systems, medical implants, agricultural products, personal care products, household cleaning products such as conditioners or softeners and as flocculants in drinking water treatment plants [4,5]. They represent a growing multi-billion dollar industry covering several 1000s of materials and billions of kilos used globally every year. However, polymers have until recently been considered of low environmental concern due to their high molecular weight, lower bioavailability, and inadequate reaction potential under different ecological circumstances. Therefore, they have been subjected to

☆ Presented in QSAR2021 virtual conference (Posters 223 and 224) (https://www.ascctox.org/qsar2021).

* Corresponding author.
E-mail address: kunal.roy@jadavpuruniversity.in (K. Roy).

https://doi.org/10.1016/j.comtox.2021.100181
Received 23 June 2021; Received in revised form 16 August 2021; Accepted 17 August 2021
Available online 19 August 2021

exemptions or reduced regulatory requirements in many jurisdictions worldwide [6–8]. However, polymers are now drawing the attention of national and international chemical regulatory agencies regarding their environmental safety, and these exemptions are expected to be revised in the coming years in several geographies; for example, as per Article 138(2) of the REACH Regulation [9], the exemption of polymers from REACH registration is under consideration, with the goals of analyzing the human health and environmental risk of polymers in comparison to other organic chemicals and assessing the need for the registration of specific types of polymers. The EU has prepared three reports to address the upper mentioned objective; the first report [10] compares REACH's registration requirements to those of other regulations. The second study [7] proposes a method for identifying polymers of low concern (PLC) based on the criteria such as low content of low MW material, restricted presence of reactive functional groups, and nil or low cationic charge density [11,12], which is comparable to the US EPA's approach. Other countries have introduced similar PLC requirements [7,8,10,13]. The European Commission's third report [14] summarised the criteria for identifying polymers that may need to be registered under REACH in the future and adapted the information requirements for these polymers. The ECETOC Polymers Task Force is also looking at polymers, having presented a conceptual framework for polymer risk assessment [8] and looked into the applicability of chemical-analytical methodologies, standardized test procedures, and prediction models [13].

Polymers are macromolecules with various physicochemical properties that are important in determining how the polymer will behave in the environment or environmental fate of polymers. These properties include a monomeric composition of polymers, size of polymers, molecular weight distribution, an average molecular weight of polymeric compounds, oligomer content of the polymer with molecular mass less than 1000 and 500 Daltons, the equivalent weight of reactive functional groups present in the polymers, cationic charge density, type of cation, the position of cation in backbone/pendant chain, and water solubility etc.[12] However, such basic information about polymers was missing or not reported in the ways that are useful for risk assessment of polymeric compounds in the registration reports[12]. Similarly, the amount of publicly available, high-quality environmental toxicity data on industrial polymers such as cationic polyquaterniums is minimal. In the present study, we have employed the Quantitative Structure-Toxicity relationship (QSTR) modelling for the aquatic toxicity data gap filling of polymers. To predict activity, toxicity, and property of organic chemicals, the Quantitative Structure-Activity Relationship (QSAR) methodology is regarded as one of the most commonly utilized and widely acknowledged computational approaches by numerous chemical legislations and organizations [15,16]. The experimental toxicity data of a series of cationic polymers against different aquatic species for individual QSTR and interspecies-Quantitative Structure-Activity Relationship (i-QSTR) modelling were obtained from the USEPA 1996 report [12] and Environment and Climate Change Canada (ECCC) [17].

Only a few reports have been published so far using the QSAR technique to estimate cationic polymer toxicity. The USEPA proposed univariate QSTR models for predicting acute and chronic toxicity of cationic polymers against fish, *D. magna*, and algal species in the 1990s. The final models were based on the cationic charge density (it is generally based on the percent of amines nitrogen (% A-N) because more than 99.9% of all polymers that have been submitted under section 5 of TSCA have had their cationic group based on the nitrogen) of polymers, implying that the cationic charge density of polymeric compounds governs polymer toxicity [12]. More recently, Nolte et al. [18] used a diverse set of polymers (n = 43) to create QSARs for algal toxicity. They proposed four separate QSTR models, each using a different set of polymer classes (cationic, non-ionic, and anionic polymers) and one model from a combined dataset of different polymer classes. The QSTR model for cationic polymer was developed using the dataset of only nine compounds using the multiple linear regression techniques. The final model for predicting algal toxicity of cationic polymers was based on a single descriptor, i.e., #CN (number of carbon-nitrogen bonds of the nitrogen in the central amine (or amidine) functional group) that was normalized for polymer charge density.

In the current study, we have first proposed individual QSTR models for algae and fish species, followed by interspecies-QSTR models among *D. magna*, fish and algae species. The following are the primary objectives of the current work:

1. To develop and evaluate individual QSTR models against fish and algae species to deal with the polymer toxicity data gap.
2. Similarly, novel robust interspecies-QSTR models among *D. magna*, fish, and algae are developed and proposed. This will aid in extrapolating toxicity data from one species to another. These methods will aid in reducing experimental efforts as well as animal testing.
3. The proposed individual QSTR and interspecies-QSTR models allowed us to identify multiple polymers' structural features, which have significant effects on toxicity prediction.
4. The proposed models will be used to assess the risks of new or untested polymeric compounds within the applicability domain of stated models.
5. This toxicity prediction strategy can also help to reduce experiment time, expense, and animal testing by several folds.

2. Materials and methods

2.1. Dataset

The experimental toxicity data of a series of cationic polymers were obtained from the USEPA 1997 report [12] and Environment and Climate Change Canada [17] against different aquatic species. The USEPA 1997 report comprises toxicity data of 73 polymers against three different species *D. magna,* fish and green algae. The final dataset was reduced to *D. magna (n = 20)*, fish (n = 38) and green algae (n = 17) toxicity data after a data pre-treatment based on the available information regarding different property and response endpoints. Similarly, Environment and Climate Change Canada's additional data comprises toxicity data of 242 polymers against different species. The final dataset comprises *D. magna (n = 78)*, fish (*O. mykiss, n* = 47 and *P. promelas*, n = 29) and algae (*P. subcapitata, n* = 36 and *S. capricornutm*, n = 25) toxicity data after data pre-treatment based on the available information regarding different property and response endpoints. To obtain the final QSTR models, we have combined the data points obtained from two sources. The final datasets for fish and algae toxicity modelling were composed of 112 and 78 polymers, respectively. All the toxicity values were expressed in $1/EC_{50}$ (mg/L). In the composite dataset, many polymers had multiple toxicity data against the same taxonomic group. To consider a specific polymer once in the modelling study for a particular endpoint, we have averaged the specific polymer toxicity data against the same species. For *in silico* modelling purposes, the EC_{50} values were transformed into the molar scale (dividing the EC_{50} value with the reported average molecular weight (MW_n) of the individual polymers) followed by conversion to the negative logarithmic form, i.e., pEC_{50} (EC_{50} in molar). Note that pEC_{50} values are directly proportional to the toxicity.

2.2. Descriptor calculation

The dataset employed in the present study were blinded in terms of the names as well as explicit chemical structures of the polymers, although the number of polymers with corresponding toxicity against different species (fish, *D. magna* and algae) and six other associated properties (Charge Density (a-N %), avMw, %MW less than 1000 Da, % MW less than 500 Da, Cation position and Cation type (i.e. primary (1°), secondary (2°), tertiary (3°) and quaternary (4°) cation) have been reported [12]. The initial idea was to use these properties as descriptors for the modelling. However, the number of these properties is limited.

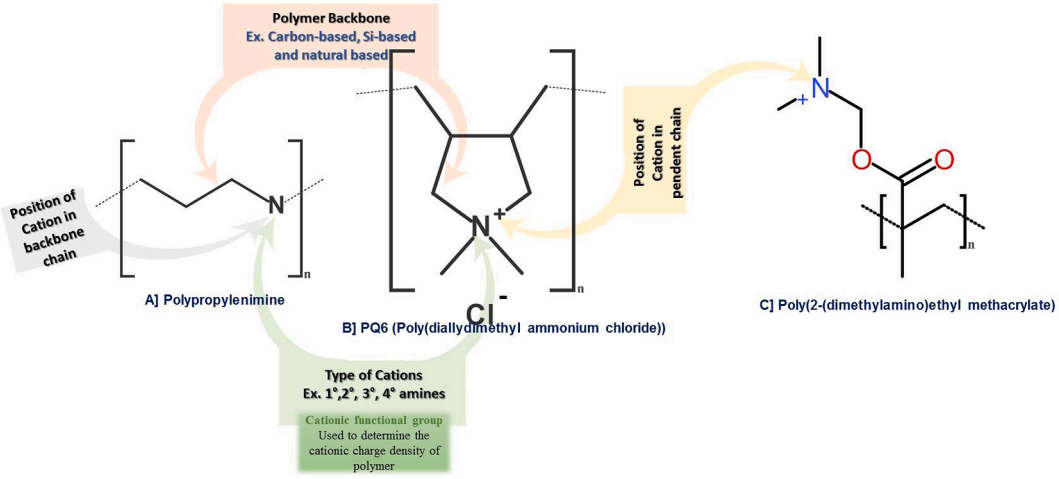

Fig. 1. General representation of different descriptors by considering three cationic polymers i.e. A] Polypropylenimine, B] PQ6 and C] Poly(2-(dimethylamino)ethyl methacrylate).

Hence, we defined the cation type into four descriptors: primary, secondary, tertiary, and quaternary, which will help explore the cation types in the polymers. Similarly, we defined another indicator variable, cation position. In the case of cation position properties, based on the available position for at least one occurrence, we have defined nine descriptors, named as cation positions 0, 1, 2, 3, 4, 5, 7, 8 and 11. Therefore, an entire set of 17 descriptors was prepared for modelling. Fig. 1 represents the general overview of different calculated descriptors by considering three cationic polymers.

2.3. Dataset division

To strategically validate each QSTR model, we divided each fish and algae dataset separately into two subsets (training and test sets) using three separate data division approaches employing a dataset division software that can be downloaded for free from https://dtclab.webs.com/software-tools [19–21]. The compounds in the training set were used to specifically train the model, while the compounds in the test set were used to validate the model. For fish and algae QSTR modelling, the training sets comprise 72 ($N_{train} = 72$) and 40 polymers ($N_{train} = 40$) and the validation sets comprise 23 ($N_{test} = 23$) and 14 polymers ($N_{test} = 14$), respectively. On the other hand, for i-QSTR modelling, the dataset was distributed into two subsets, i.e., training and test sets (only in the case of i-QSTR modelling between *D. magna* and fish species).

2.4. Model building and validation

This research aims to develop robust and accurate QSTR models for the prediction of toxicity of cationic polymers against fish and algae species (defined endpoint as per OECD Principle 1). Only the training set compounds were used in the model building process, while the test set compounds were only used to validate the best model that has been chosen. Since the current study involves a small number of features, we ran genetic algorithm-multiple linear regression (GA-MLR) repeatedly

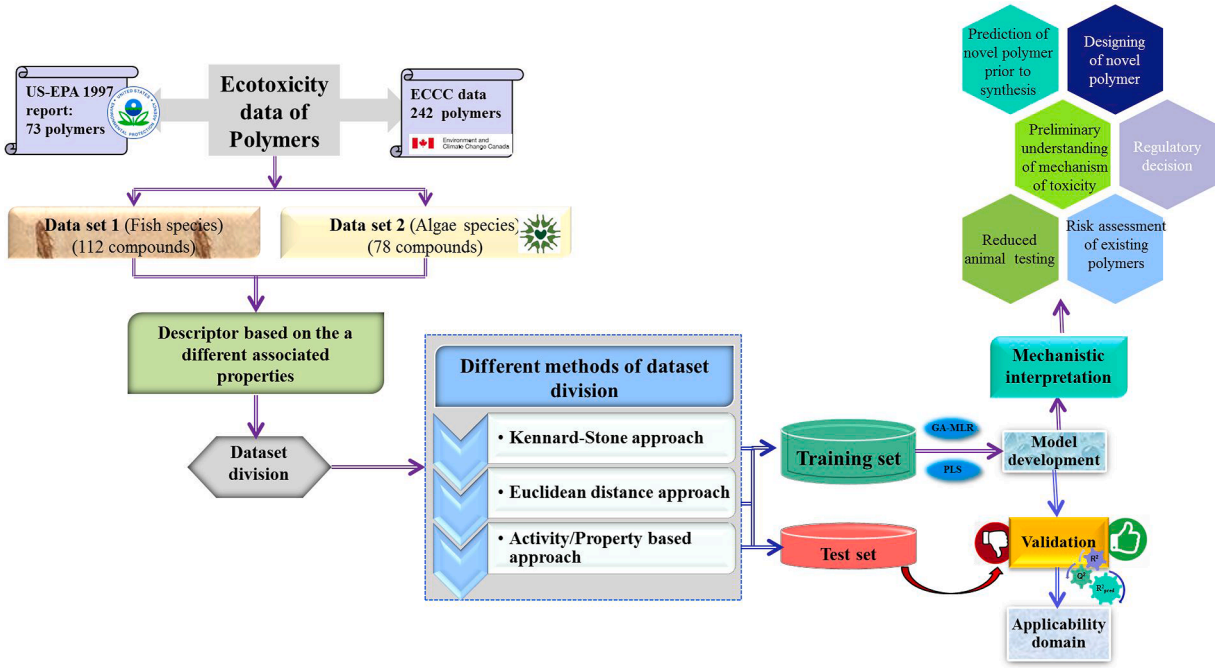

Fig. 2. Schematic overview of the detailed methodology used in the present study.

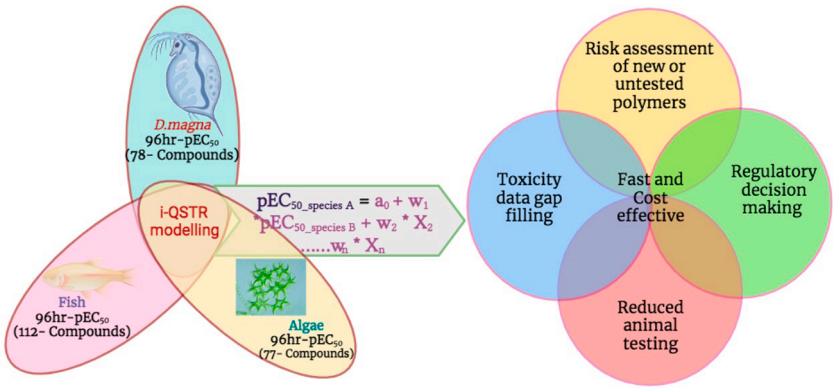

Fig. 3. Basic architecture and advantages of i-QSTR models among *D. magna,* fish and algae species.

to find the best MLR model [22]. Finally, the descriptors selected in the best MLR models were subjected to the partial least squares (PLS) regression technique [23] to derive more robust models at various latent variables (unambiguous algorithm as per OECD Prnciple 2). The final best models were submitted to a comprehensive validation procedure following OECD principle 4, which states that "appropriate metrics of goodness-of-fit, robustness, and predictability" must be established for any QSTR model. These parameters may be calculated based on a variety of internal and external validation metrics. The internal quality/validation metrics include determination coefficient, Q^2_{LOO} (Q^2 leave one out), and r_m^2 metrics for the training set [16,24,25]. However, any model's prediction ability depended on its performance in predictions for external compounds, i.e. external set predictive variance (R^2_{pred}) [16,23,24]. We have also checked the mean absolute error of the test set ($MAE_{100\%\ test}$), and subsequently, we have also determined the r_m^2 metric for the test set [26]. Fig. 2 depicts a schematic overview of the detailed methodology used in the present study.

The finally selected best QSTR models were further subjected to Y-randomization using SIMCA-P [27] to decide whether the final models were obtained by chance or not. The study was carried out by creating 100 new models for an original model by shuffling the values of the Y-variable while keeping the values of the X-variable unchanged [28]. Similarly, we established the applicability domain of the final models in the chemical space (OECD Principle 3) using the DModX approach [28] available in the SIMCA-P software at 99% confidence level, which is helpful to determine whether the predictions obtained from a particular model are reliable or not.

2.5. Interspecies-QSTR modelling (i-QSTR)

Interspecies modelling is a chemometric data modelling technique in which experimental response data from one species is used to extrapolate the response data from another species. The basic interspecies model provides a statistical relationship between the responses of two species. An i-QSTR model consists of an experimental response value of a specific species, which serves as a predictor variable (independent variable) along with other numerical descriptors (structural as well as physicochemical properties) to predict the response value for another species, which is referred to as interspecies quantitative structure-toxicity relationships model (interspecies-QSTR) [29,30]. This approach will be helpful to fill toxicity data gaps for multiple species as well as for some kind of preliminary understanding of the chemical's mechanism of action (MoA) to a particular species [31–34]. Simultaneously, i-QSTR models significantly decrease the expense and duration of laboratory studies and animal testing. Fig. 3 depicts the simple architecture and advantages of i-QSTR models among *D. magna,* fish and algae species.

Although i-QSTR modelling for small molecule toxicity has been reported in various publications, we have only reviewed here a few studies that have studied the interspecies relationship of chemical ecotoxicity. For example, Hao et al. [35] reported an individual QSTR models for acute oral toxicity of nitroaromatic chemicals to rats and i-QSTR models between rat and mouse. The individual QSTR modelling was performed using the toxicity data of 128 nitroaromatic compounds against rats, and model training was accomplished by using the 101 nitroaromatic compounds, while 27 molecules were used for model validation. They have stated that the van der Waals surface area, the presence of C-F at topological distance 6, high frequency of C-N at topological distance nine primarily contribute to the acute toxicity of NACs against rats. In contrast, the heteroatom content and the presence of N-O at a topological distance of ten mainly result in the decrease of toxicity. On the other hand, they have developed the rat-mouse and mouse-rat interspecies i-QSTR models using toxicity datasets of 100 and 102 nitroaromatic compounds, respectively; both the i-QSTR models were based on single descriptors, i.e. experimental toxicity value of nitroaromatic compounds against rat and mouse vice versa. Finally, they have used the rat-mouse and mouse-rat interspecies QSTR (i-QSTR) models employed in toxicity prediction for true external sets consisting of 67 and 265 compounds, respectively. Similarly, Guohui Sun et al. [36] reported an individual QSTR model for acute oral toxicity of polycyclic aromatic hydrocarbons (PAHs) to rats, as well as the i-QSTR models between rat and mouse. Individual QSTR modelling was performed using the toxicity data of 276 PAHs against rats, and the best model obtained from the training set of 102 PAHs compounds and subsequently validated using the test set of 20 PAHs compounds was based on the eight molecular descriptors. They have stated that the toxicity of PAHs increases with an increase in lipophilicity and decrease with a reduction in the polarity of compounds. Simultaneously, they have developed rat-mouse and mouse-rat i-QSTR models using the training set of 44 PAHs, and final models were validated using the test set of 14 PAHs compounds. Lastly, they have used the rat-mouse and mouse-rat interspecies i-QSTR models for toxicity prediction of true external sets consisting of 61 and 68 compounds, respectively. Kar et al. [37] have published a detailed overview on the idea and application of interspecies quantitative structure toxicity relationship modelling (i-QSTR modelling). It is stated that the i-QSTR technique is beneficial when toxicity data for one species is missing since it is possible to extrapolate toxicity from the toxicity endpoint of another species. This approach can also overcome the cost of many toxicity tests, improve understanding of the mechanism of toxic action (MOA) of chemicals for different organisms and endpoints, and fill data gaps where toxicity value for a particular chemical compound is absent for a specific endpoint. For more details about the i-QSTR modelling, please refer to an article published by Kar et al. [37].

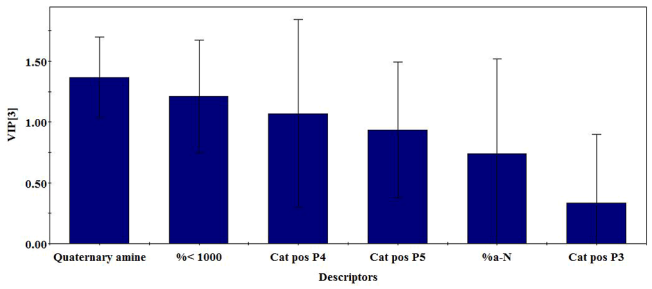

Fig. 4. VIP plot for combined *fish* species toxicity model of polymers.

3. QSTR modelling of the combined data of ECCC and USEPA

3.1. QSTR modelling for toxicity against combined fish species

In this case, we have tried to develop a QSTR model using the combined dataset of three different fish species (*O. mykiss* and *P. promelas* toxicity data obtained from the ECCC dataset and fish data extracted from USEPA dataset). The initial dataset comprising 112 compounds with toxicity against three different fish species was employed for QSTR modelling. The initial results indicate that the seventeen compounds (polymer id: 48, 53, 55, 56, 64, 83, 216, 217, 218, 238, 35A, 36A, 38A, 39A, 62A, 73A and 75A) show high prediction residuals, influencing the final model quality. Therefore, the identified seventeen compounds were removed from the initial dataset, and the rest of the compounds were used for modelling. The final dataset of 95 compounds was divided into two different subsets (i.e., training and test sets) employing dataset division software tool freely available from https://dtclab.webs.com/software-tools. The training set (N_{train} = 72 polymers) was explicitly used for model building, while the test set was used for rigorous validation purpose (N_{test} = 23 polymers). The best model was obtained from the splitted data obtained from the Euclidean distance [20] dataset division approach. Models were obtained employing the GA-MLR technique followed by PLS regression [23], leading to a final model with three latent variables. The selected model was robust and acceptable as per the internationally acceptable internal and external validation metrics, as shown below in Eq 1:

$$n_{train} = 72,\ n_{test} = 23,\ LV = 3,\ R^2 = 0.703,\ Q^2 = 0.638,\ R^2_{pred} = 0.776,\ \overline{rm^2_{LOO\_train}} = 0.516,\ \Delta rm^2_{LOO\_train} = 0.194,\ MAE_{Training100\%} = 0.649,\ \overline{rm^2_{test}} = 0.711,\ \Delta rm^2_{test} = 0.044,\ MAE_{Test100\%} = 0.522$$

We have performed a VIP plot analysis, representing the importance of each variable present in the final PLS QSTR model. The VIP score was used to differentiate between higher and lower statistically significant variables. If the VIP score of any particular descriptor is higher than one, it is considered a more significant variable for QSTR modelling. In the plot, the descriptors were arranged in descending order of importance from left to right. Out of six variables, three variables were considered more important, i.e., quaternary amine (presence of quaternary amine in the molecule), %<1000 (percentage of oligomers with molecular mass less than 1000 Daltons) [12] and cat pos P4 (presence of the cationic functional group in the pendant chain at position four) with VIP scores higher than one.

On the other hand, cat pos 5, cat pos P3 (presence of the cationic functional group in the pendant chain at positions five and three respectively) [12] and %a-N (charge density) [12] were considered as less essential descriptors in the final model (Fig. 4). It is also evident from the loading plot analysis that quaternary amine, cat pos P4, %< 1000 were the most influential descriptors among all based on their locations in the plot (Fig. 5). A scatter plot of the observed vs predicted training and test set compounds was shown in Fig. S1 in the supporting information file, which represents the goodness of fit and predictions obtained by the QSTR model against different fish species.

The QSTR equation for the prediction of combined fish toxicity comprises six unique variables. Out of the six variables, only one descriptor (<%1000) contributes negatively to polymers toxicity. In contrast, the rest of the descriptors result in positive contributions towards the toxicity against fish species, which indicates that presence of quaternary amines in the molecular building block, presence of cation atom in the pendant chain at three, four and five positions and higher charge density of polymeric compounds result in higher toxicity against fish and vice versa. For example, compound **22** (presence of quaternary amines in the molecular building block), **29A and 43A** (presence of cation atom in the pendant chain at fourth and fifth position), and compound **103** (higher charge density) result in higher toxicity against fish species. The close analysis of data revealed that a compound with only higher charge density results in lower toxicity than a compound

$$pEC_{50\_fish} = 5.70 + 0.126 \times \%a - N + 0812 \times Quaternaryamines + 1.074 \times CatposP3 + 1.230 \times CatposP4 + 1.260 \times CatposP5 - 0.030 \times \% < 1000$$

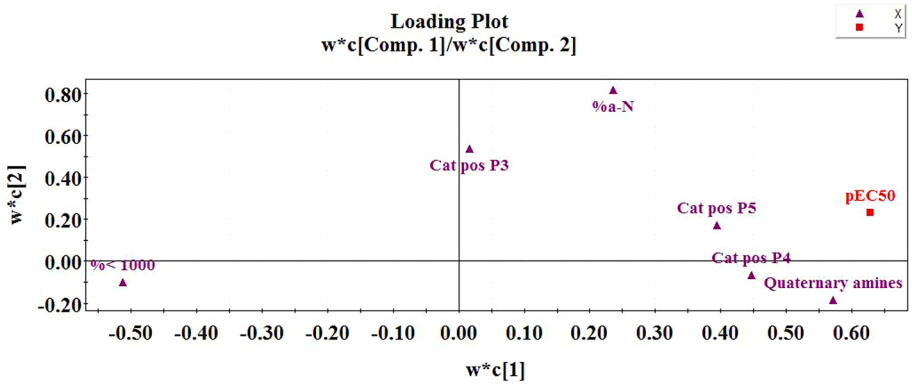

Fig. 5. Loading plot for combined *fish* species toxicity model of polymers.

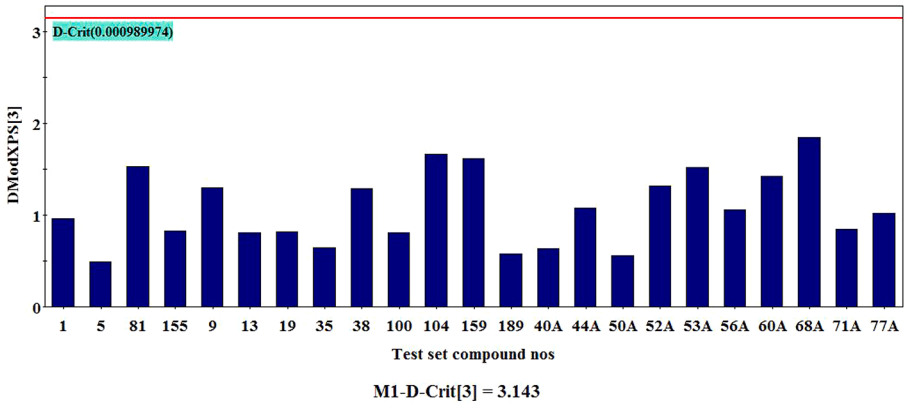

Fig. 6. AD analysis of test set compounds at a 99% confidence level (QSTR model for fish toxicity).

with significant charge density and presence of quaternary amines in the molecular building block (e.g. Compound **215** results in higher toxicity than **103)**. The previous study has also evidenced that increased charge density raises toxicity against fish [12]. On the other hand, compound **172** (due to a high value of <%1000, which stands for the percentage of oligomers with molecular mass less than 1000 Daltons) results in a lower toxicity towards fish.

We have also performed the Y-randomization study using SIMCA-P with an objective to determine whether the final selected model was obtained by chance (random) or not (non-random). The analysis revealed that the final model was not obtained by chance, i.e., the model is non-random. (as shown in Fig. S2 of supporting information file). Furthermore, finally, we have performed applicability domain analysis of the QSTR model to define its domain in chemical space, which will be helpful to determine whether the prediction of a particular compound obtained using the model is reliable or not. From Fig. 6, we can observe that all the test set compounds are within the final QSTR model's applicability domain at a 99% confidence level (D-critical = 0.0009999).

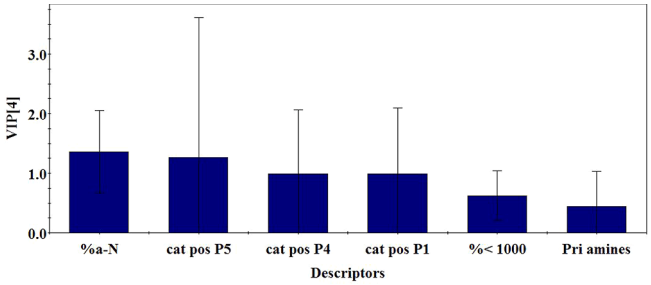

Fig. 7. VIP plot for combined *algae* species toxicity model of polymers.

3.2. QSTR modelling for toxicity against combined algae species

In this case, we have tried to develop a QSTR model using the combined dataset of three different algal species (*P. subcapitata* and *S. caricornutum* toxicity data obtained from the ECCC dataset and green algae data extracted from the USEPA dataset). The initial dataset comprising 78 compounds with toxicity against three different algae species was employed for QSTR modelling. The initial results indicate that the twenty-four compounds show high prediction residuals, which may influence the final model quality. Therefore, the identified twenty-four compounds were removed from the initial dataset, and the rest of the compounds were used for modelling. The final dataset of 54 compounds was divided into two subsets (i.e., training and test sets) employing a dataset division software tool that was freely available from https://dtclab.webs.com/software-tools. The training set (N_{train} = 40 polymers) was explicitly used for model building, while the test set was used for rigorous validation purpose (N_{test} = 14 polymers). The best model was obtained from the split data derived from the Kennard-Stone-based [19] dataset division approach. Models were obtained employing the GA-MLR technique followed by PLS regression [23], leading to a final model with four latent variables. The selected model was robust and acceptable as per the internationally acceptable internal and external validation metrics, as shown below in Eq 2:

$$pEC_{50\_algae} = 5.855 + 0.137 \times \%a - N - 0.779 Priamines + 0.641 \times catposP1 + 1.018 \times catposP4 + 1.854 \times catposP5 - 0.013 \times \% < 1000$$

$n_{train} = 40,\ n_{test} = 14,\ LV = 4,\ R^2 = 0.676,\ Q^2 = 0.516,\ R^2_{pred} = 0.703,\ \overline{rm^2_{LOO\_train}} = 0.393,\ \Delta rm^2_{LOO\_train} = 0.159,\ MAE_{Training100\%} = 0.592,\ \overline{rm^2_{test}} = 0.401,\ \Delta rm^2_{test} = 0.295,\ MAE_{Test100\%} = 0.437$

Here also, we have performed the VIP plot analysis. Out of six variables, two variables were considered as more important, i.e., %<a-N (charge density) and cat pos P5 (presence of the cation in the pendant chain at positions five) [12] and with a VIP score of more than one [27,28] (Fig. 7). On the other hand, cat pos P4 and P1 (presence of the cationic functional group in the pendant chain at positions four and one, respectively), presence of primary amines in the molecular building block were less important descriptors in the final model. It is also evident from the loading plot analysis that cat pos P5 and %a-N were more influential descriptors among all based on their locations in the plot. The rest of the four descriptors were considered less influential because they are situated near the plot origin (Fig. 8). A scatter plot of the observed vs predicted training and test set compounds was shown in Fig. S3 in the supporting information file, which represents the goodness of fit and predictions obtained by the QSTR model against different algae species.

The QSTR equation for the prediction of the combined algal toxicity comprises six unique variables. Out of the six variables, two descriptors (primary amines and %<1000) show negative contributions towards

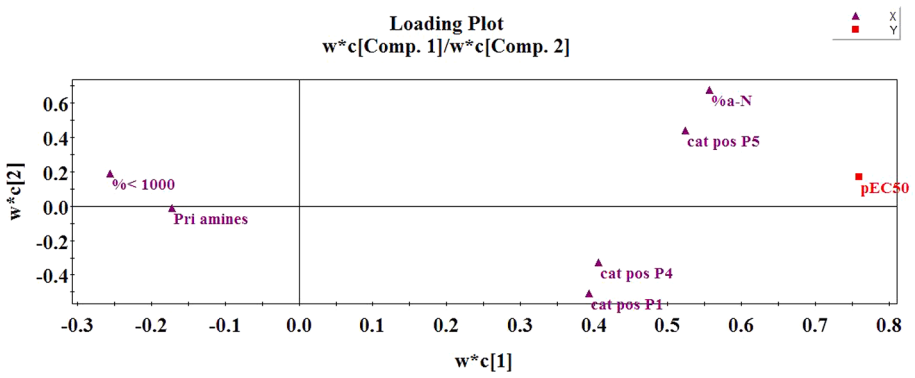

Fig. 8. Loading plot for combined *algae* species toxicity model of polymers.

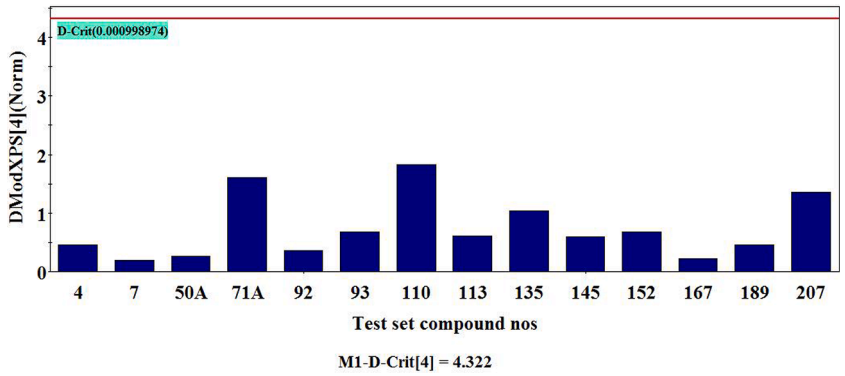

Fig. 9. AD analysis of test set compounds at a 99% confidence level (QSTR model for algae toxicity).

polymers toxicity, indicating that the presence of primary amines in the molecular building block and higher value of the percentage of oligomers with molecular mass less than 1000 Daltons descriptors result into low toxicity to algal species and vice versa. For example, compounds **191** and **169** show lower toxicity due to the presence of primary amines in their chemical structure and higher value of the percentage of oligomers with molecular mass less than 1000 Daltons descriptors respectively Simultaneously, other descriptors result in positive contributions towards the toxicity of polymers against algal species, indicating that higher charge density of polymeric compounds and cation presence at one, four and five positions result in higher toxicity of a particular polymer against algae and vice versa. For example, compounds **104 (**due to higher charge density), **84, 29A** and **51A** (presence of cation at one, fourth and fifth positions) show higher toxicity against the algal species.

Here also, we have also performed the Y-randomization study using SIMCA-P with an objective to determine whether the final selected model was obtained by chance (random) or not (non-random). The analysis revealed that the final model was not obtained by chance, i.e., the model is non-random (as shown in Fig. S4 of supporting information file). Furthermore, we have performed an AD study using the DModX software tool [28]. From Fig. 9, we can easily understand that all the test set compounds are within the final QSTR model's applicability domain at 99% confidence level (D-critical = 0.0009999) (Fig. 8).

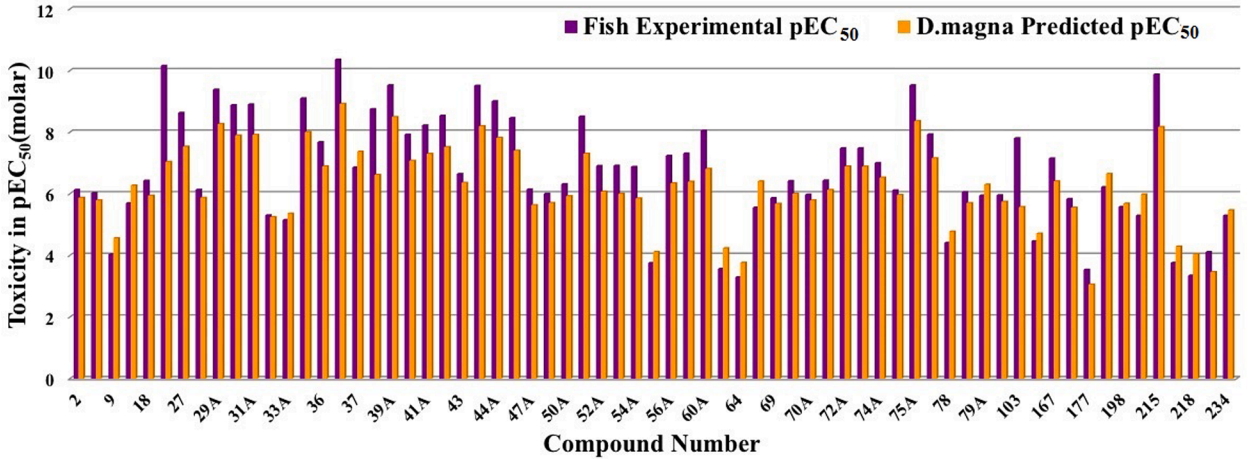

Fig. 10. Bar chart of experimental fish toxicity and predicted *D. magna* toxicity employing the i-QSTR model.

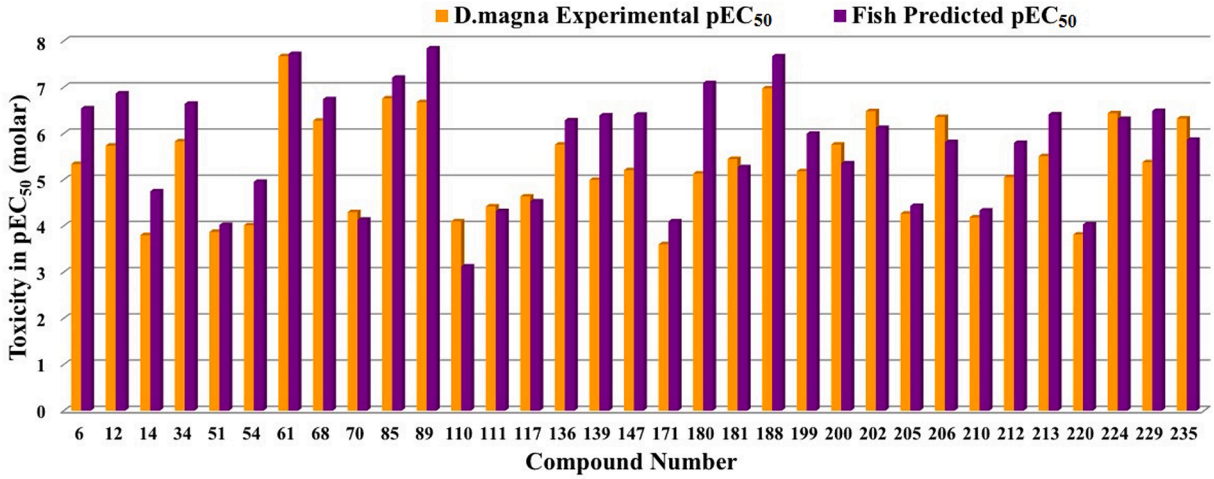

Fig. 11. Bar diagram of experimental *D. magna* toxicity and predicted *fish* toxicity employing the i-QSTR model.

4. Interspecies toxicity modelling of polymers

In this study, we have successfully developed individual species-species correlation models as well as i-QSTR models using multiple linear regressions techniques, as shown in Table S1 (supporting information file), but all of the final i-QSTR models were generated using the partial least squares regression modelling algorithm, as detailed below:

4.1. i-QSTR modelling between toxicities against D. magna and fish

To identify the possible use of the existing experimental toxicity of different fish species data to estimate the *D. magna* toxicity, we have performed the interspecies quantitative toxicity relationship between *D. magna* and Fish. Out of 78 polymers in the *D. magna* data set, 45 polymers were found to have their reported pEC_{50} values against different fish species. These 45 polymers had toxicity data for both Fish and *D. magna* and were used for i-QSTR model development. The set of 45 common compounds was initially divided into training and test sets. The training set of 34 polymeric compounds was used for model development, while 11 test set compounds were explicitly used for model validation purpose. The final i-QSTR model was based on the three components (extracted meaningful information for model development from an original set of descriptors) and obtained using the PLS regression analysis technique. The final model was considered the most promising and predictive based on external set prediction quality and different internationally accepted internal and external metrics metrics, as shown below in Eq 3.

$$pEC_{50\_Daphnia} = 1.390 - 0.0436 \times \%a - N + 0.720 \times Priamines + 0.236 \times Secamines - 1.079 \times PoscatP1 + 0.736 \times pEC_{50\_Fish}$$

$$pEC_{50\_Fish} = 1.661 + 0.111 \times \%a - N - 0.992 \times Primary\ amine - 0.978 \times cation\ pos\ backbone + 0.411 \times catposP5 + 0.797 \times pEC_{50\_Daphnia}$$

$n_{train} = 34,\ n_{test} = 11,\ LV = 3,\ R^2 = 0.695,\ Q^2 = 0.577,\ R^2_{pred} = 0.731,\ \overline{rm^2_{LOO\_train}} = 0.443,\ \Delta rm^2_{LOO\_train} = 0.211,\ MAE_{Training100\%} = 0.538,\ \overline{rm^2_{test}} = 0.690,\ \Delta rm^2_{test} = 0.074,\ MAE_{Test100\%} = 0.457$

The final i-QSTR model consists of five independent variables, one of which is experimental fish toxicity value (pEC_{50_fish}). The other four variables are the presence of primary and secondary amines in the polymeric molecules, charge density, and the cationic functional group in the pendant chain. Equation **3** suggests that the presence of primary and secondary amine and higher experimental fish toxicity result in an increase in *D. magna* toxicity and vice versa. For example, compounds **40** and **8** show higher toxicity due to primary and secondary amines in the polymeric compounds and higher experimental fish toxicity values. Again, compounds **15** and **169** resulted in lower *D. magna* toxicity due to the higher charge density of molecules and a cationic functional group in the pendant chain.

The final i-QSTR model was used to predict *D. magna* toxicity of the rest of the 67 compounds of the fish dataset whose toxicity data was missing. The bar plot below (Fig. 10) depicted the experimental fish toxicity and predicted *D. magna* toxicity of the above mentioned 67 polymers.

4.2. i-QSTR modelling between toxicities against fish and D. magna

Out of 112 polymers in the fish toxicity data set, 45 polymers were found to have their reported pEC_{50} values against different *D. magna* species. These 45 polymers had toxicity data for both Fish and *D. magna* and were used for i-QSTR model development. Prior to model development, the final data set was divided into two subsets (i.e., training and test set) using the data set division software tool. Only training set compounds were employed in the model development, while the test set compounds were used to validate the developed QSTR model. The best model was based on four latent variables and obtained using the best subset selection, followed by the PLS technique. The developed model was considered robust and predictive based on their statistical internal and external parameters, as shown below in Eq 4:

$n_{train} = 34,\ n_{test} = 11,\ LV = 4,\ R^2 = 0.764,\ Q^2 = 0.693,\ R^2_{pred} = 0.803,\ \overline{rm^2_{LOO\_train}} = 0.585,\ \Delta rm^2_{LOO\_train} = 0.177,\ MAE_{Training100\%} = 0.574,\ \overline{rm^2_{test}} = 0.612,\ \Delta rm^2_{test} = 0.187,\ MAE_{Test100\%} = 0.570$

Eq 4 comprises five unique independent variables, with a 3:2 ratio of positive and negative contributions respectively towards predicting fish toxicity. A close analysis suggests that higher experimental *D. magna*

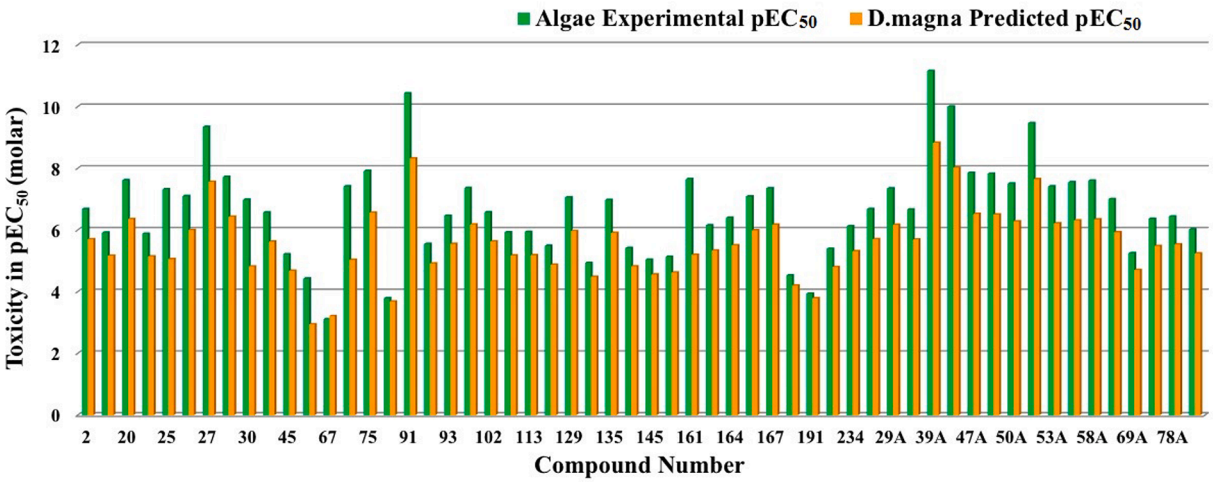

Fig. 12. Bar diagram of experimental *algae* toxicity and predicted *daphnia* toxicity employing the i-QSTR model.

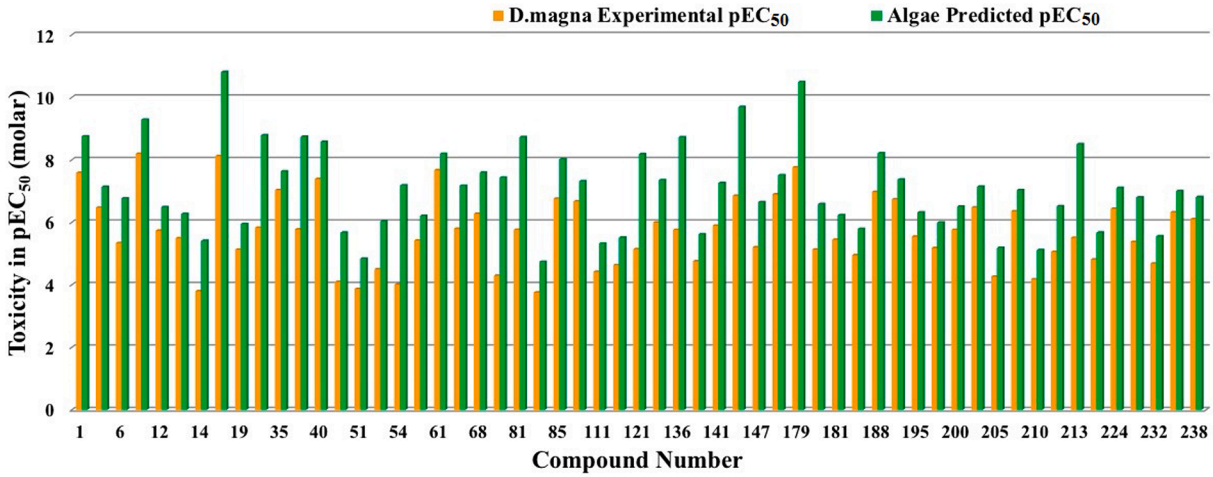

Fig. 13. Bar diagram of experimental *D. magna* toxicity and predicted *algae* toxicity employing the i-QSTR model.

toxicity value, presence of a cationic functional group in the pendant chain and higher charge density result in higher fish toxicity values. Hall and Mirenda et al. [38] also showed an increased toxicity with increasing charge density of METAC and AETAC polymers against fish [39]. However, if we compare Eq 3 and Eq 4, the analysis shows that the presence of primary amine and charge density act oppositely for the prediction of fish and *D. magna toxicity*. This means that cationic groups and charge density are the inversely controlling factors in the toxicity of the cationic polymers to *D. magna*, while toxicity to fish seems directly proportional to the charge density of polymeric compounds.

The final i-QSTR model was used to predict fish toxicity for the rest of the 33 compounds of the *D. magna* dataset whose toxicity data was missing for fish. The bar plot below (Fig. 11) depicted the experimental *D. magna* toxicity and predicted *fish* toxicity of the above mentioned 33 polymers.

4.3. i-QSTR modelling between toxicities against D. magna and algae

In this current study, we have explored whether experimental acute toxicity data of different algal species will be helpful in model development and prediction of toxicity against the *D. magna* or not. Out of 78 polymers in the *D. magna* data set, 21 polymers were found to have their reported pEC_{50} values against different algal species. These 21 polymers had toxicity data for both algae and *D. magna* and were used for i-QSTR model development using the GA-MLR technique in Eq 5 below.

$$pEC_{50\_Daphnia} = 1.040 + 0.698 \times pEC_{50\_Algae} - 1.184 \times PoscatP6 - 1.095 \times PoscatP1$$

$$n_{train} = 21,\ LV = 2,\ R^2 = 0.778,\ Q^2 = 0.701,\ \overline{rm^2_{LOO_{train}}} = 0.595,\ \Delta rm^2_{LOO_{train}} = 0.193,\ MAE_{Training100\%} = 0.406$$

The final model was based on the experimental algal toxicity value (pEC_{50_algae}) as an essential independent variable and two other essential descriptors: the presence of the cationic functional group in the pendant chain at positions one and six. Eq. 5 proposes that *D. magna* toxicity is directly proportional to the algal toxicity (higher pEC_{50_fish} results in higher toxicity towards *D. magna)*. On the other hand, the presence of cationic functional groups in the pendant chain of polymer leads to a decrease in the *D. magna* toxicity and vice versa.

The final i-QSTR model was used to predict the *D. magna* toxicity of the rest of the 56 compounds of the algae dataset whose toxicity data was missing for *D. magna*. The bar plot below (Fig. 12) depicted the experimental *algae* toxicity and predicted *D. magna* toxicity of the above mentioned 56 polymers.

4.4. i-QSTR modelling between toxicities against algae and D. magna

Out of 77 polymers in the algal data set, 21 polymers were found to

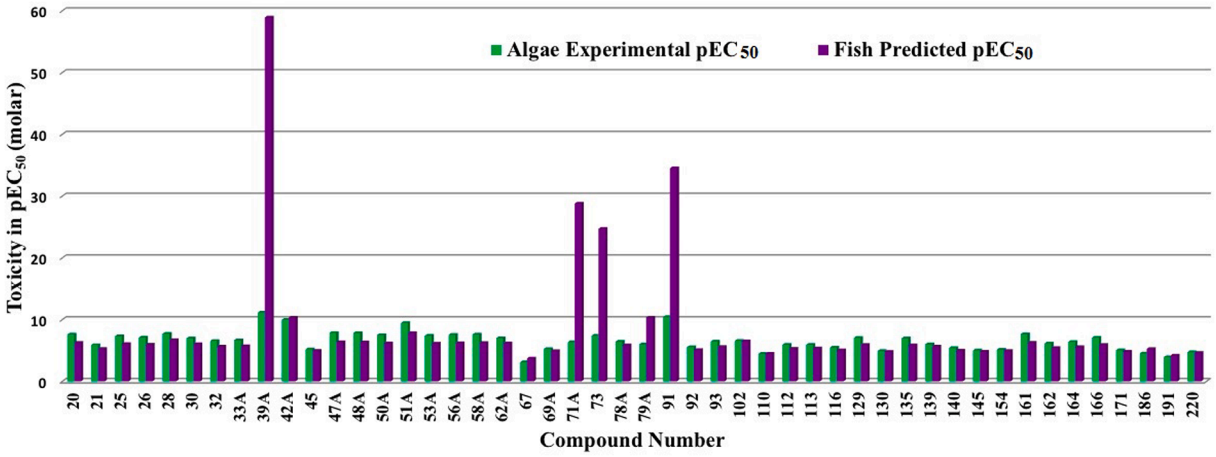

Fig. 14. Bar diagram of experimental *algae* toxicity and predicted *fish* toxicity employing the i-QSTR model.

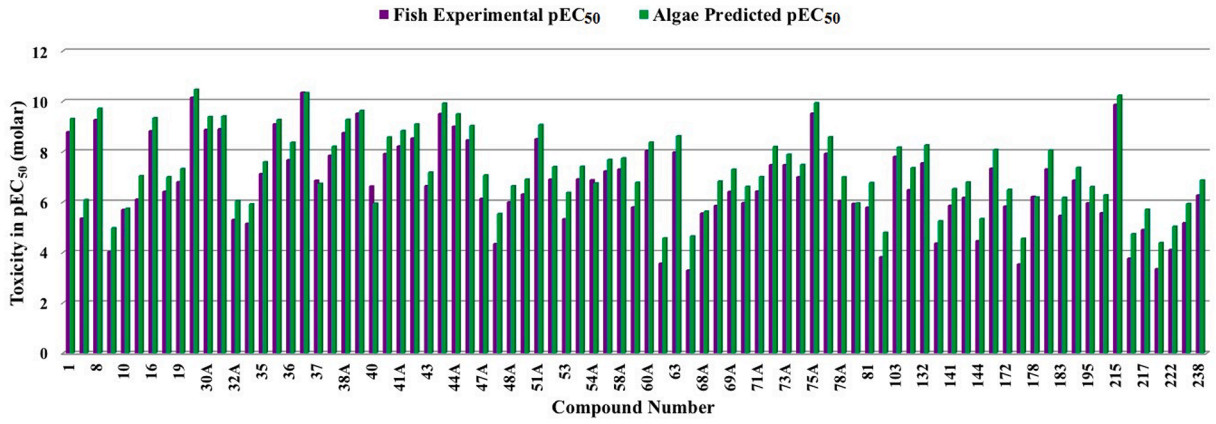

Fig. 15. Bar diagram of experimental fish toxicity and predicted *algae* toxicity employing the i-QSTR model.

have their reported pEC_{50} values against *D. magna*. These 21 polymers had toxicity data for both Fish and *D. magna* and were used for i-QSTR model development using the GA-MLR technique in Eq 6 below.

$$pEC_{50\_Algae} = 2.055 + 0.882 \times pEC_{50\_Daphnia} - 0.633 \times catposbackbone + 1.586 \times catpos4$$

$$n_{train} = 21,\ LV = 2,\ R^2 = 0.811,\ Q^2 = 0.763,\ \overline{rm^2_{LOO\_train}} = 0.683,\ \Delta rm^2_{LOO\_train} = 0.134,\ MAE_{Training100\%} = 0.526$$

The final i-QSTR model was used to predict algae toxicity of the rest of the 57 compounds of the *D. magna* dataset for which the toxicity data was missing algae. The bar plot below (Fig. 13) depicts the experimental *D. magna* toxicity and predicted *algae* toxicity of the above mentioned 57 polymers.

4.5. i-QSTR modelling between toxicities against different fish and algal species data

Out of 112 polymers in the fish species data set, 27 polymers were found to have their reported pEC_{50} values against different algal species. These 27 polymers had toxicity data for both Fish and *D. magna* and were used for i-QSTR model development in Eq 7 below.

$$pEC_{50\_Fish} = 1.989 + 0.278 \times catpos5 + 0.557 \times pEC_{50\_Algae} + 0.00001 \times Mn + 0.738 \times catpos4$$

$$n_{train} = 27,\ LV = 2,\ R^2 = 0.826,\ Q^2 = 0.780,\ \overline{rm^2_{LOO\_train}} = 0.700,\ \Delta rm^2_{LOO\_train} = 0.114,\ MAE_{Training100\%} = 0.447$$

The final QSTR model was used to predict fish toxicity for the rest of the 48 compounds of the *algae* dataset whose toxicity data was missing for fish. The bar plot below (Fig. 14) depicted the experimental *algae* toxicity and predicted *fish* toxicity of the above mentioned 48 polymers.

4.6. i-QSTR modelling between toxicities against different algal and fish species data

Out of 77 polymers in the algal data set, 27 polymers were found to have their reported pEC_{50} values against different fish species. These 27 polymers had toxicity data for both Fish and *D. magna* and were used for i-QSTR model development using the GA-MLR technique (in Eq 8 below).

$$pEC_{50\_Algae} = 1.543 + 0.849 \times pEC_{50\_Fish} - 0.632 \times Primaryamines - 0.593 \times catposP6 + 0.306 \times Quaternaryamines$$

$$n_{train} = 27,\ LV = 3,\ R^2 = 0.789,\ Q^2 = 0.730,\ \overline{rm^2_{LOO\_train}} = 0.634,\ \Delta rm^2_{LOO\_train} = 0.143,\ MAE_{Training100\%} = 0.477$$

The final i-QSTR model was used to predict the algal toxicity of the rest of the 83 compounds of the *different fish species* dataset for which toxicity data was missing for algae. The bar plot below (Fig. 15) depicted the experimental fish toxicity and predicted *algae* toxicity of the above

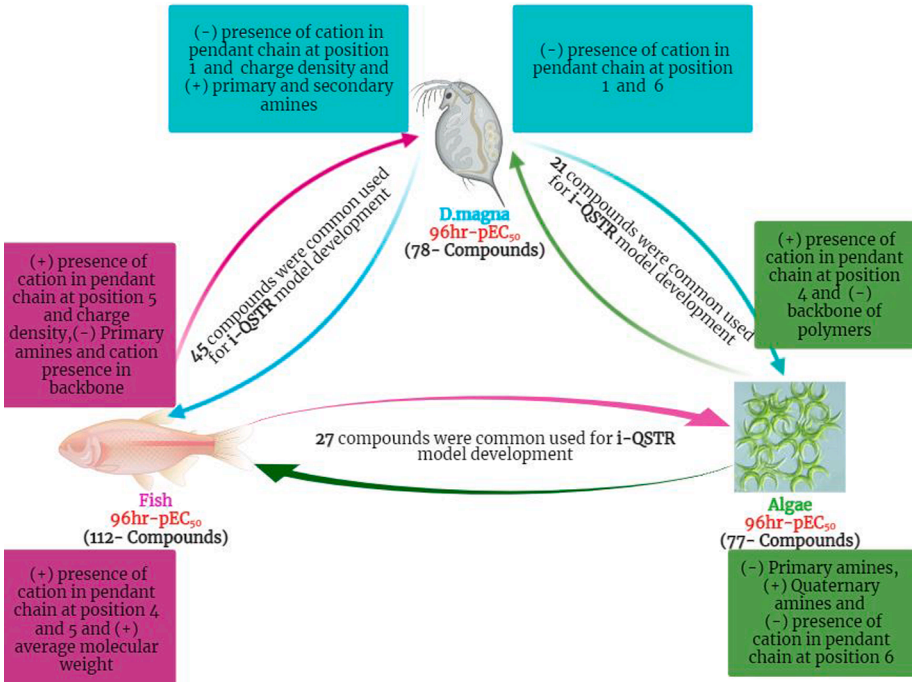

Fig. 16. Schematic overview of i-QSTR modelling among *D. magna*, fish and algae, and mechanistic interpretation from the final i-QSTR models. The "+" notation signifies a positive impact on the toxicity, and "-" defines a negative impact on toxicity. Similarly, the "Sky blue" arrow represents the i-QSTR model against *D. magna*, the "Orange" arrow represents the i-QSTR model against fish, and the "Green" arrow represents the i-QSTR models against algae species. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

mentioned 83 polymers.

Lastly, the individual species-species experimental toxicity values relationship were shown by plotting the experimental toxicity values of cationic polymers among fish, *D. magna*, and algae as shown in Fig. S5 (supporting information). Fig. S5 depicts scatter plots of experimental toxicity values of cationic polymers. A] Experimental Fish toxicity values vs experimental *D. magna* toxicity values of cationic polymers, B] Experimental algae toxicity values vs experimental Fish toxicity values of cationic polymers, and C] Experimental algae toxicity values vs experimental *D. magna* toxicity values of cationic polymers.

5. Conclusion

In the present study, we have generated different individual QSTR models for toxicity prediction against fish and algae and i-QSTR models among three species, which are *D. magna*, fish and algae species to bridge the toxicity data gap for cationic polymers. All the proposed models were generated by employing the partial least squares regression technique at different latent variables. The presented study revealed that a cationic functional group in the pendant chain at positions one, three, four and five enhances the toxicity towards different fish and algal species involved in the study. Similarly, higher charge density increases the toxicity against both the response endpoints. Quaternary amines in the molecular building block of polymers also result in higher toxicity towards different fish species. On the other hand, higher values of pct less than 1000 Da result in a decreased toxicity towards both the studied endpoints. Similarly, the presence of primary amines in the molecular building block reduces the toxicity against the algal species. On the other hand, the mechanistic interpretation of the i-QSTR models revealed several critical characteristic features of polymers along with the experimental response of one species which are also helpful for toxicity prediction of other species, ultimately helping to reduce the experimental efforts as well as animal studies (as shown in Fig. 16). Subsequently, the final i-QSTR models were successfully used for toxicity prediction of polymers with the absence of experimental toxicity for one species but with available toxicity data for other species and vice versa. For example, *D. magna* toxicity prediction of 67 and 56 polymers was done using fish and algae experimental toxicity data.

Similarly, prediction of fish toxicity of 33 and 48 chemicals was done using *D. magna* and algae experimental toxicity data. Finally, prediction of algae toxicity of 57 and 83 chemicals was done with *D. magna* and fish experimental toxicity data. Last but not least, mechanistic interpretation of the i-QSTR models revealed that the toxicity among these three species is directly proportional to each other, i.e., if polymer A shows high toxicity against fish, the same polymer A results in higher toxicity against *D. magna* and algae and vice versa.

Inspired by the encouraging results from this study, we are in the process of gathering additional cationic polymer toxicity data from the US-EPA and private industries to refine the reported models. We also plan to develop (a) category-specific QSAR models using a smaller group of polymers, i.e., based on C backbone, Si backbone and natural backbone polymers; (b) read across models to overcome the data scarcity; (c) different classification models to categorize polymers into different classes as per the EPA guidelines with an ultimate objective to supplement/replace the existing regulatory models of US-EPA for cationic polymer toxicity.

CRediT authorship contribution statement

Pathan Mohsin Khan: Data curation, Methodology, Validation, Investigation, Writing – original draft. **Hans Sanderson:** Conceptualization, Funding acquisition. **Kunal Roy:** Conceptualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

PMK thanks to National Institute of Pharmaceutical Education and Research Kolkata, the Ministry of Chemicals & Fertilizers, Department of Pharmaceuticals, Government of India for providing financial assistance in the form of a fellowship, HS appreciates CEFIC LRI ECO46 project iTAP for funding; KR thanks SERB, New Delhi for financial assistance under the MATRICS scheme (File number MTR/2019/000008).

Disclosure

This work was presented in QSAR2021 virtual conference (Posters 223 and 224) (https://www.ascctox.org/qsar2021).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.comtox.2021.100181.

References

[1] OECD, Organisation for Economic Co-operation and Development: Definition of polymer. Paris, 1991, https://www.oecd.org/env/ehs/oecddefinitionofpolymer.htm.

[2] L. Sperling, Introduction to Physical Polymer Science, fourth ed., John Wiley & Sons, Hoboken (NJ), 2006.

[3] A. Ravve, Principles of Polymer Chemistry, Springer Science & Business Media, Platinum Press, New York, USA, 2013.

[4] ECHA. 2012a. European Chemicals Agency. Guidance for monomers and polymers. Guidance for the implementation of REACH. ECHA-12-G-02-EN, version 2.0, April 2012. https://echa.europa.eu/documents/10162/23036412/polymers_en.pdf/9a74545f-05be-4e10-8555-4d7cf051bbed.

[5] H. Sanderson, P. M. Khan, S. Kar, K. Roy, A.M.B. Hansen, K. Connors and S. Belanger QSAR Modeling of Aquatic Toxicity of Cationic Polymers, Chemometrics and Cheminformatics in Aquatic Toxicology, first ed. John Wiley & Sons, Inc. 2021.

[6] OECD. 2009. Organisation for Economic Co-operation and Development. Data analysis of the identification of correlations between polymer characteristics and potential for health or ecotoxicological concern. ENV/JM/MONO (2009)1. Paris, France; https://www.oecd.org/env/ehs/risk-assessment/42081261.pdf.

[7] A. de Toni, N. Saïdi, L.R. Santos, S. Mudgal, Technical assistance related to the review of REACH with regard to the registration requirements on polymers. Final report, prepared for the European Commission, DG Environment. Bio By Deloitte., 2015. https://ec.europa.eu/environment/chemicals/reach/pdf/FINAL%20REPORT%20POLYMER%20SI671025.pdf.

[8] O.D. Matos, The ECETOC Conceptual Framework for Polymer Risk Assessment (CF4Polymers) EUROPEAN CENTRE FOR ECOTOXICOLOGY AND TOXICOLOGY OF CHEMICALS, European Centre for Ecotoxicology and Toxicology of Chemicals Rue Belliard 40, B-1040 Brussels, Belgium., 2019. https://www.ecetoc.org/wp-content/uploads/2019/06/ECETOC-TR133-1CF4Polymers.pdf.

[9] EC(2006), REGULATION (EC) No 1907/2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency, amending Directive 1999/4, 2006 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006R1907&from=EN.

[10] M. Postle, P. Holmes, M. Camboni, A. Footitt, N. Tuffnell, M. Blainey, G. Stevens, A. Pye, Review of REACH with regard to the Registration Requirements on Polymers and 1 to 10 Tonne Substances, Risk & Policy Analysts Limited, Farthing Green House, 1 Beccles Road, Loddon, Norfolk, NR14 6LT, 2012. https://ec.europa.eu/environment/chemicals/reach/pdf/studies_review2012/report_study10.pdf.

[11] USEPA(1997), Polymer exemption guidance manual EPA 744-B-97–001. United States Environmental Protection Agency, Office of Pollution Prevention and Toxics, Washington, 1997 https://www.epa.gov/sites/default/files/2015-03/documents/polyguid.pdf.

[12] R.S. Boethling, J. v. Nabholz, Environmental assessment of polymers under the US Toxic Substances Control Act, United States Environmental Protection Agency. (1996) 1–64. https://nepis.epa.gov/Exe/ZyPDF.cgi/901A0Q00.PDF?Dockey=901A0Q00.PDF.

[13] O.D. Matos, Applicability of Analytical Tools, Test Methods and Models for Polymer Risk Assessment, Rue Belliard 40, B-1040 Brussels, Belgium., 2020. https://www.ecetoc.org/wp-content/uploads/2020/03/ECETOC-TR133-2.Polymers-Risk-Asessment.pdf.

[14] K. Bougas, C. Corden, M. Crookes, G. Federici, P. Fisk, Scientific and technical support for the development of criteria to identify and group polymers for registration/evaluation under REACH and their impact assessment , Wood E&IS GmbH Weserstrasse, 4 60329 Frankfurt Am Main, 2020. https://op.europa.eu/en/publication-detail/-/publication/1cc811ff-d5fc-11ea-adf7-01aa75ed71a1.

[15] K. Roy, Quantitative Structure-Activity Relationships in Drug Design, Predictive Toxicology, and Risk Assessment, IGI Global, 2015.

[16] K. Roy, S. Kar, R.N. Das, A Primer on QSAR/QSPR Modeling: Fundamental Concepts, Springer, 2015.

[17] https://www.canada.ca/en/environment-climate-change/services/evaluating-existing-substances/screening-assessment-poly-amines-group.html.

[18] T. Nolte, W. Peijnenburg, A.-H. Chemosphere, undefined, Quantitative structure-activity relationships for green algae growth inhibition by polymer particles, Chemosphere 179 (2017) (2017) 49–56.

[19] R.W. Kennard, L.A. Stone, Computer aided design of experiments, Technometrics 11 (1) (1969) 137–148.

[20] H. Golmohammadi, Z. Dashtbozorgi, W.E. Acree, Quantitative structure–activity relationship prediction of blood-to-brain partitioning behavior using support vector machine, Eur. J. Pharm. Sci. 47 (2) (2012) 421–429.

[21] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, A. Tropsha, Does rational selection of training and test sets improve the outcome of QSAR modeling? J. Chem. Inf. Model. 52 (10) (2012) 2570–2578.

[22] P. Ambure, R.B. Aher, A. Gajewicz, P. Puzyn, K. Roy, "NanoBRIDGES" software: open access tools to perform QSAR and nano-QSAR modeling, Chemom. Intell. Lab. Syst. 147 (2015) 1–13.

[23] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2) (2001) 109–130.

[24] K. Roy, I. Mitra, On various metrics used for validation of predictive QSAR models with applications in virtual screening and focused library design, Comb. Chem. High Throughput Screen. 14 (2011) 450–474.

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

[26] K. Roy, I. Mitra, P.K. Ojha, S. Kar, R.N. Das, H. Kabir, R.N. Das, H. Kabir, Introduction of rm2 (rank) metric incorporating rank-order predictions as an additional tool for validation of QSAR/QSPR models, Chemom. Intell. Lab. Syst. 118 (2012) 200–210.

[27] Z. Wu, D. Li, J. Meng, H. Wang, Introduction to SIMCA-P and its application, in: V. Esposito Vinzi, W.W. Chin, Jörg Henseler, H. Wang (Eds.), Handbook of Partial Least Squares, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, pp. 757–774, https://doi.org/10.1007/978-3-540-32827-8_33.

[28] S. Wold, J. Trygg, A. Berglund, H. Antti, Some recent developments in PLS modeling, Chemom. Intell. Lab. Syst. 58 (2) (2001) 131–150.

[29] S. Kar, K. Roy, First report on development of quantitative interspecies structure–carcinogenicity relationship models and exploring discriminatory features for rodent carcinogenicity of diverse organic chemicals using OECD guidelines, Chemosphere 87 (2012) 339–355.

[30] S. Raimondo, C.R. Jackson, M.G. Barron, Influence of taxonomic relatedness and chemical mode of action in acute interspecies estimation models for aquatic species, Environ. Sci. Technol. 44 (19) (2010) 7711–7716.

[31] S. Cassani, S. Kovarich, E. Papa, P.P. Roy, Daphnia and fish toxicity of (benzo) triazoles: Validated QSAR models, and interspecies quantitative activity–activity modelling, J. Hazard. Mater. 258 (2013) 50–60.

[32] N. Basant, S. Gupta, Multi-target QSTR modeling for simultaneous prediction of multiple toxicity endpoints of nano-metal oxides, Nanotoxicology 11 (3) (2017) 339–350.

[33] S. Kar, K. Roy, First report on interspecies quantitative correlation of ecotoxicity of pharmaceuticals, Chemosphere 81 (2010) 738–747.

[34] S.D. Dyer, D.J. Versteeg, S.E. Belanger, J.G. Chaney, F.L. Mayer, Interspecies correlation estimates predict protective environmental concentrations, Environ. Sci. Technol. 40 (9) (2006) 3102–3111.

[35] Y. Hao, G. Sun, T. Fan, X. Tang, J. Zhang, Y. Liu, N. Zhang, L. Zhao, R. Zhong, Y. Peng, In vivo toxicity of nitroaromatic compounds to rats: QSTR modelling and interspecies toxicity relationship with mouse, J. Hazard. Mater. 399 (2020) 122981, https://doi.org/10.1016/j.jhazmat.2020.122981.

[36] G. Sun, Y. Zhang, L. Pei, Y. Lou, Y. Mu, J. Yun, F. Li, Y. Wang, Z. Hao, S. Xi, S. and C. Li, Chemometric QSAR modeling of acute oral toxicity of Polycyclic Aromatic Hydrocarbons (PAHs) to rat using simple 2D descriptors and interspecies toxicity modeling with mouse. Ecotoxicol. Environ. Saf., 222(2021), p.112525.

[37] S. Kar, R.N. Das, K. Roy, J. Leszczynski, Can toxicity for different species be correlated?: the concept and emerging applications of interspecies quantitative structure-toxicity relationship (i-QSTR) modeling, Int. J. Quant. Struct. Prop. Rela. 1 (2016) 23–51.

[38] W.S. Hall, R.J. Mirenda, Acute toxicity of wastewater treatment polymers to Daphnia pulex and the fathead minnow (Pimephales promelas) and the effects of humic acid on polymer toxicity, J. Water Pollut. Control Fed. 63 (1991) 895–899.

[39] J. Cumming, Environmental Fate, Aquatic Toxicology and Risk Assessment of Polymeric Quaternary Ammonium Salts from Cosmetic Uses, QLD, Australia, 2008.