22

QSAR Modeling of Aquatic Toxicity of Cationic Polymers

Hans Sanderson[1], Pathan M. Khan[2], Supratik Kar[3], Kunal Roy[4], Anna M.B. Hansen[1], Kristin Connors[5], and Scott Belanger[5]

[1] *Department of Environmental Science, Aarhus University, Aarhus, Denmark*
[2] *Drug Theoretics and Cheminformatics (DTC) Lab, Department of Pharmaceutical Technology, & Jadavpur University, National Institute of Pharmaceutical Education and Research, Kolkata, India*
[3] *Department of Chemistry, Physics & Atmospheric Science, Interdisciplinary Center for Nano-toxicity, Jackson State University, Jackson, MS, USA*
[4] *Drug Theoretics and Cheminformatics (DTC) Lab, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India*
[5] *Global Product Stewardship, Procter and Gamble Company, Cincinnati, OH, USA*

22.1 Introduction

Polymers are increasingly drawing the attention from national and international chemical regulatory agencies regarding their environmental safety. Polymers were assessed for inclusion in the regulations by regulatory agencies in the 1990s, prominently by the United States of America Environmental Protection Agency (USEPA). The USEPA developed a framework for environmental assessment of polymers under the U.S. Toxic Substances Control Act (TSCA) [1]. Polymers have been regarded to be of low environmental concern since they frequently have high molecular weights (MW), which limits their bioavailability, as well as their toxicity [2]. They have hence been subject to exemption or reduced regulatory requirements globally (e.g. currently exempt in the registration, evaluation, authorization and restriction of chemicals [REACH] regulation), however, this is expected to be lifted in the coming years [3].

Polymers are produced in millions of tons annually worldwide and are used in a wide range of different products from plastics, to water soluble consumer products, to functional pharmaceutical and nanomaterials. They are large macromolecules composed of repeating subunits with MWs typically ranging from the 100s to several 1000s or even millions of Daltons. The size limits their ability to pass membranes according to the Lipinski rule and therefore also their toxicity as realized by the regulatory agencies. Hence, polymers have remained understudied with regards to their environmental properties and potential hazard profile during the past decades [3].

Quantitative structure–activity relationships (QSARs) are an integral part of most chemical legislation and agencies work. In order to address the increased regulatory scrutiny of polymers, QSARs need to be revised and/or new models developed. There is generally limited property and toxicity data publicly available for most polymers. The availability of environmental toxicological QSARs are also limited. The most comprehensive data collection and attempt to develop environmental toxicity quantitative structure–toxicity relationships (QSTRs) for polymers was conducted by the USEPA in the mid-1990s [1]. More recently, Nolte et al. [4] attempted to develop QSARs for algae toxicity for a diverse group of polymers ($n = 43$). The USEPA in 1996 [1] found a set of relevant descriptors and published the data, and Nolte et al. [4] confirmed the findings that charge density (%a-N) was a strong descriptor of algae toxicity and developed an acute algae toxicity QSAR of cationic polymers:

Chemometrics and Cheminformatics in Aquatic Toxicology, First Edition. Edited by Kunal Roy.

$$-\log EC_{50}\ \text{mg}\,\text{l}^{-1} = -1.73 * 10^{-2}\left(\%\text{a} - \text{N}\right) + 5.78$$
$$\left(n = 9;\ r^2 = 0.75;\ p = 3e^{-3}\right)$$

The USEPA in 2013 [5] also developed models based on the same dataset. They found that with regards to aquatic toxicity, charge density (%a-N) was the only quantitative descriptor, and that at high %a-N (above 3.5% for algae and fish and 4.3% for daphnids), the toxicity and charge density no longer correlate but plateaued and remained essentially constant. The backbone also influence polymer toxicity, with the rank toxicity potency being: Carbon-based backbone ≥ Si-backbone = natural backbone. They have developed acute structure–activity relationships (SARs) as evident in the report below showing the most conservative acute SARs for carbon-based backbone cationic polymers with a %a-N ≤ 3.5:

$$\text{Log fish 96 h}\,LC_{50}\ \text{mg}\ \text{l}^{-1} = 1.209 - 0.462 * \%\text{a} - \text{N}$$
$$\text{Log Daphnid 48 h}\,LC_{50}\ \text{mg}\ \text{l}^{-1} = 2.839 - 1.194 * \%\text{a} - \text{N}$$
$$\text{Log Green algae 96 h}\,EC_{50}\ \text{mg}\,\text{l}^{-1} = 1.569 - 0.97 * \%\text{a} - \text{N}$$

The models by the USEPA [1] and Nolte et al. [4] are based on univariate regression models. In this chapter, we test and improve the QSAR models for cationic polymers, the class of polymers of greatest environmental concern, by applying multivariate methods and interspecies analysis to the available data employing the guidelines of Organization for Economic Co-operation and Development (OECD) [6] for QSAR model development.

22.2 Materials and Methods

22.2.1 Polymers

Polymers are a diverse group of macromolecules. They can have many different sizes, charges, backbone, functionalities, etc. [3]. In this analysis, we focus on cationic polymers as these have been identified as being of potential environmental concern [1]. Below is a depiction of a commonly used cationic polymer, polydiallyldimethylammonium chloride (polyDADMAC) and also known as polyquaterium 6 (PQ6), CAS# 26062-79-3. PQ6 is widely used as a flocculant in waste water treatment where its sorptive properties, due to its charge density, enable the capture of particles and removes them from the water phase (Figure 22.1).

Due to its use within wastewater treatment, PQ6 is the most studied polyquaterium material. A recent review by Pecquet et al. [7] collected and critically evaluated the available polymers relevant to household cleaning, including PQ6 [7]. PQ6 is the most well-described polyquaterium cationic polymer but as evident from the distribution of the toxicity values below in Figure 22.2, there are large variations in toxicity among species, indicating potentially both different sensitivities and uncertainties in the tests as large charged and sorptive materials such as PQ6 are difficult to test.

PQ6 has been assessed by the Canadian authorities and has been found to be inherently toxic to aquatic organisms based on a *Ceriodaphnia dubia* acute 48 h LC_{50} value of 0.32 mg l^{-1}. The annual PQ6 usage in Canada is >1000 tons year^{-1} [8]. Other polymers, including cationic polymers, are much less researched as Pecquet et al. [7] documented. We found that in a SciFinder analysis with the query combination: cationic + polymer + aquatic + toxicity, resulted in only seven references.

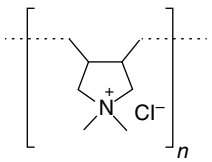


Figure 22.1 PQ6 structure.

22.2.2 Dataset

As evident from the section above, limited publicly available environmental toxicity data for polymers exists, which both impairs and highlights the need for QSARs for polymers. The USEPA dataset on polymers in the 1990s [1] identified

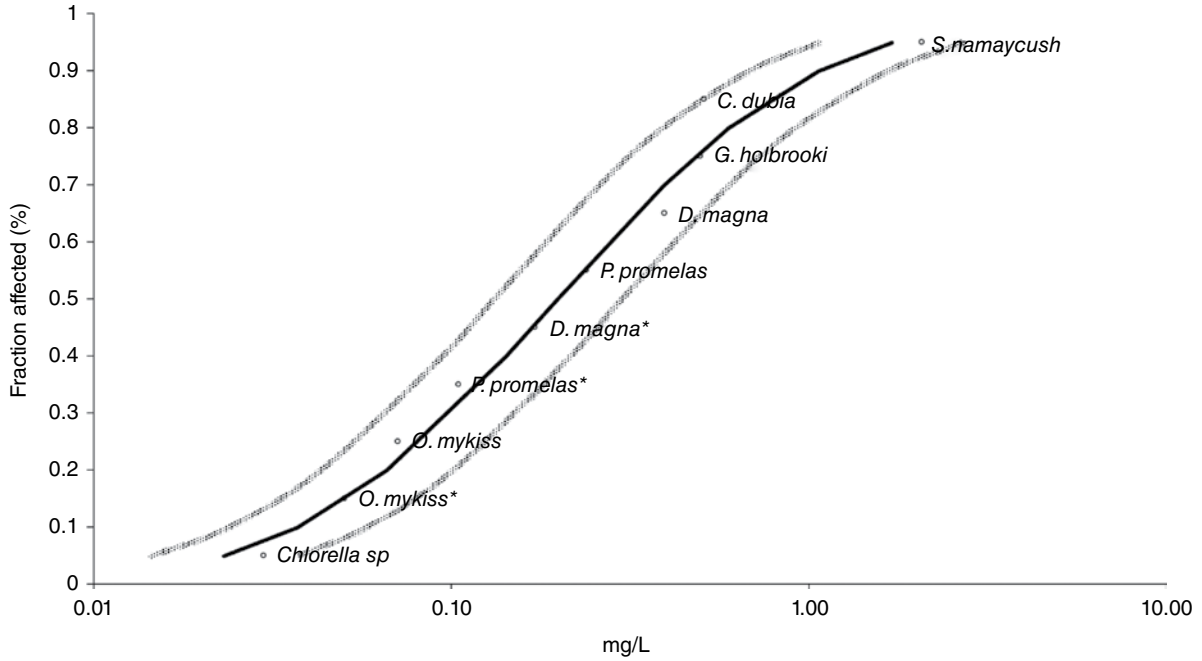


Figure 22.2 Distribution of PQ6 aquatic toxicities.

relevant toxicological descriptors for modeling purposes – but did not develop robust models. The dataset is confidential and therefore blinded, but information is provided pertaining to key properties and measured toxicities in the published report [1]. This is the most comprehensive publicly available dataset for cationic polymers. Table 22.1 lists cationic polymers, properties, and aquatic toxicity data from the USEPA report from 1996 [1].

The experimental toxicity data of a series of cationic polymers against three different species *Daphnia magna*, fish, and green algae were collected [1]. The dataset consists of three types of polymers namely carbon-based backbone, silicon-based backbone, and natural-based backbone. The acute toxicity was measured in fish after a 96 h exposure, in *D. magna* after 48 h, and in green algae after 96 h. All acute toxicity values were expressed in EC_{50} ($mg\,l^{-1}$). In addition, chronic toxicity to green algae was also measured and was expressed as the chronic value (ChV, or geometric mean of the no-observed-effect concentration and lowest-observed-effect-concentration). In the dataset, many polymers had multiple toxicity data against the same taxonomic group. To consider a specific polymer once in the modeling study for a particular endpoint, we have averaged the toxicity data for the specific polymer against same species. For *in silico* modeling purpose, the EC_{50} values were transformed into molar scale (dividing the EC_{50} value with the reported average molecular weight (avMW) 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. Few polymers were removed from the modeling study due to absence of data related to associated properties. Therefore, for modeling purpose, the number of polymers considered with toxicity data to Fish (96 h), *D. magna* (48 h), green algae (96 h), and chronic green algae were 38, 20, 17, and 16, respectively.

22.2.3 Descriptor Calculation

Polymers are class 2 substances as they have variable compositions. The chemistry of polymers focuses on the monomeric units a polymer is composed of and their mole percentages, MW distributions including the avMW

Table 22.1 Blinded cationic polymer dataset [1].

Comp #	%a-N	AvMW (Da)	MW < 1000 Da%	MW < 500 Da%	Cat Pos	Cat type	Fish EC_{50} (mg l^{-1})	Fish pEC_{50} (EC_{50} in molar)	*D. magna* EC_{50} (mg l^{-1})	*D. magna* pEC_{50} (EC_{50} in molar)	Algae EC_{50} (mg l^{-1})	Algae pEC_{50} (EC_{50} in molar)	Algae ChV EC_{50} (mg l^{-1})	Algae ChV pEC_{50} (EC_{50} in molar)
Carbon-based backbone polymers														
27	0.69	6.20E + 03	0	0	11	3	4.6	6.130	NA	NA	1.3	6.678	0.16	7.588
28	0.7	1.80E + 03	NA	NA	NA	NA	9.2	5.291	300	3.778	2.2	5.913	0.88	6.311
28	0.7	1.80E + 03	NA	NA	NA	NA	8.5	5.326	310	3.764	NA	NA	NA	NA
28	0.7	1.80E + 03	NA	NA	NA	NA	3.9	5.664	NA	NA	NA	NA	NA	NA
29	0.7	8.00E + 06	0	0	4	4	3.3	9.385	NA	NA	360	7.347	130	7.789
30	0.7	8.00E + 06	0	0	4	4	10.61	8.877	28.2	8.453	NA	NA	NA	NA
31	0.7	8.00E + 06	0	0	4	4	10	8.903	40	8.301	NA	NA	NA	NA
32	1.1	6.00E + 03	1	1	0	3	30	5.301	NA	NA	NA	NA	NA	NA
33	1.4	2.40E + 03	13	2	0	2	17	5.150	16	5.176	0.52	6.664	0.27	6.949
34	2	2.50E + 06	NA	NA	2	3	0.97	9.411	1.7	9.167	NA	NA	NA	NA
34	2	2.50E + 06	NA	NA	1	3	2.3	9.036	NA	NA	NA	NA	NA	NA
35	2	1.10E + 06	0	0	5	3	0.64	9.235	NA	NA	NA	NA	NA	NA
35	2	1.10E + 06	0	0	5	3	1.2	8.962	NA	NA	NA	NA	NA	NA
36	2.1	1.90E + 07	0	0	4	3	0.84	10.354	NA	NA	NA	NA	NA	NA
37	3	1.00E + 02	NA	NA	2	3	0.9	5.046	NA	NA	NA	NA	NA	NA
38	3.1	1.80E + 05	0	0	1	4	0.32	8.750	NA	NA	NA	NA	NA	NA
39	3.3	5.00E + 06	0	0	3	2	1.5	9.523	0.09	10.745	0.035	11.155	0.006	11.921
40	3.4	5.00E + 04	0	0	4	4	0.6	7.921	NA	NA	NA	NA	NA	NA
41	3.4	5.00E + 04	0	0	4	4	0.3	8.222	NA	NA	NA	NA	NA	NA
42	3.6	2.50E + 05	0	0	4	4	0.73	8.535	0.18	9.143	0.025	10.000	0.013	10.284
43	4.5	1.00E + 06	0	0	4.5	4	0.31	9.509	NA	NA	NA	NA	NA	NA
44	4.6	1.00E + 06	0	0	5	4	0.99	9.004	NA	NA	NA	NA	NA	NA
45	5.1	2.20E + 05	0	0	4	4	0.76	8.462	NA	NA	NA	NA	NA	NA
46	6	5.00E + 03	NA	NA	0	4	0.15	7.523	NA	NA	NA	NA	NA	NA
46	6	5.00E + 03	NA	NA	0	4	0.16	7.495	NA	NA	NA	NA	NA	NA

46	6	5.00E+03	NA	NA	0	4	0.29	7.237	NA	NA	NA	NA	NA	NA
47	6.4	1.00E+03	38	12	0	4	0.72	6.143	0.073	7.137	0.014	7.854	0.006	8.222
48	7.8	1.00E+03	32	23	0	2	1	6.000	2.9	5.538	0.015	7.824	0.007	8.155
49	8	5.00E+03	NA	NA	0	4	0.13	7.585	0.34	7.167	0.016	8.495	NA	NA
49	8	5.00E+03	NA	NA	0	4	0.22	7.357	NA	NA	NA	NA	NA	NA
50	8.1	1.50E+03	20	5	0	2	0.73	6.313	NA	NA	0.047	7.504	0.03	7.699
51	8.1	5.80E+04	1	1	2.5	4	0.18	8.508	NA	NA	0.02	9.462	0.008	9.860
52	9.2	4.00E+03	49	19	0	3	0.5	6.903	NA	NA	NA	NA	NA	NA
53	11	1.80E+03	55	26	0	3	0.22	6.913	0.58	6.492	0.07	7.410	0.034	7.724
54	11	1.50E+03	44	26	0.1	1	0.2	6.875	NA	NA	NA	NA	NA	NA
55	12	8.00E+03	NA	NA	0	2	1.9	6.624	1.2	6.824	NA	NA	NA	NA
56	14	2.00E+03	20	3	0	2	0.1	7.301	0.69	6.462	0.061	7.516	0.03	7.824
56	14	2.00E+03	20	3	0	2	0.16	7.097	0.38	6.721	0.052	7.585	0.01	8.301
56	14	2.00E+03	20	3	0	2	0.1	7.301	NA	NA	NA	NA	NA	NA
57	14	2.00E+03	NA	NA	0	2	0.3	6.824	0.78	6.409	0.08	7.398	0.044	7.658
58	14	2.00E+03	25	5	0	2	0.072	7.444	0.276	6.860	0.032	7.796	0.017	8.071
58	14	2.00E+03	25	5	0	2	0.084	7.377	2.9	5.839	0.08	7.398	0.05	7.602
58	14	2.00E+03	25	5	0	2	0.16	7.097	NA	NA	NA	NA	NA	NA
59	15	2.50E+04	NA	NA	0	2	0.26	7.983	0.26	7.983	NA	NA	NA	NA
59	15	2.50E+04	NA	NA	0	2	0.24	8.018	NA	NA	NA	NA	NA	NA
60	17	5.00E+04	1	1	0	2	0.45	8.046	NA	NA	NA	NA	NA	NA
61	20	2.50E+04	NA	NA	0	2	0.32	7.893	0.17	8.167	NA	NA	NA	NA
61	20	2.50E+04	NA	NA	0	2	0.32	7.893	NA	NA	NA	NA	NA	NA
61	20	2.50E+04	NA	NA	0	2	0.23	8.036	NA	NA	NA	NA	NA	NA
61	20	2.50E+04	NA	NA	0	2	0.2	8.097	NA	NA	NA	NA	NA	NA
Silicone-based backbone polymers														
62	0.4	1.10E+03	23	19	4.5	2	300	3.564	370	3.473	0.11	7	0.09	7.087
63	0.5	3.00E+04	0	0	3	1		NA	44	5.834	NA	NA	NA	NA
64	0.7	7.60E+03	NA	NA	5.5	1		NA	1	6.881	NA	NA	NA	NA

(*Continued*)

Table 22.1 (Continued)

Comp #	%a-N	AvMW (Da)	MW < 1000 Da%	MW < 500 Da%	Cat Pos	Cat type	Fish EC_{50} ($mg\,l^{-1}$)	Fish pEC_{50} (EC_{50} in molar)	*D. magna* EC_{50} ($mg\,l^{-1}$)	*D. magna* pEC_{50} (EC_{50} in molar)	Algae EC_{50} ($mg\,l^{-1}$)	Algae pEC_{50} (EC_{50} in molar)	Algae ChV EC_{50} ($mg\,l^{-1}$)	Algae ChV pEC_{50} (EC_{50} in molar)
65	0.7	7.60E+03	NA	NA	4.5	1		NA	0.5	7.182	NA	NA	NA	NA
66	0.7	7.80E+03	NA	NA	4.5	4		NA	0.1	7.892	NA	NA	NA	NA
67	0.7	7.80E+03	NA	NA	4.5	1	6.1	6.107	1	6.892	NA	NA	NA	NA
67	0.7	7.80E+03	NA	NA	4.5	1		NA	1.1	6.851	NA	NA	NA	NA
67	0.7	7.80E+03	NA	NA	4.5	1		NA	1.4	6.746	NA	NA	NA	NA
68	0.78	1.00E+04	4	2	8.5	1	28	5.553	15	5.824	NA	NA	NA	NA
69	2.6	1.70E+03	17	5	7	4	0.65	6.418	0.16	7.026	9.6	5.248	NA	NA
70	5.5	1.10E+03	5	0	7	2	1.17	5.973	1.04	6.024	NA	NA	NA	NA
Natural-based backbone polymers														
71	0.07	2.30E+06	0	0	4	3	850	6.432	NA	NA	1000	6.362	1000	6.362
72	0.2	3.00E+07	0	0	3	4	1000	7.477	177	8.229	NA	NA	NA	NA
73	0.29	3.00E+07	0	0	3.3	3	1000	7.477	1000	7.477	NA	NA	NA	NA
74	0.4	1.00E+07	0	0	3	3	1000	7.000	1000	7.000	NA	NA	NA	NA
75	0.93	2.20E+06	0	0	5	4	1.18	9.271	NA	NA	NA	NA	NA	NA
75	0.93	2.20E+06	0	0	5	4	0.37	9.774	NA	NA	NA	NA	NA	NA
76	0.93	9.80E+05	NA	NA	10	4	NA	NA	130	6.877	NA	NA	NA	NA
77	1.7	7.30E+04	20	10	4	4	0.86	7.929	NA	NA	NA	NA	NA	NA
78	3.4	1.70E+03	30	25	4	4	1.5	6.054	26	4.815	0.62	6.438	0.39	6.639
79	4.3	5.00E+05	0	0	0	1	570	5.943	11	7.658	480	6.018	480	6.018

%a-N, cationic charge density in percent amine nitrogen; Cat pos, position of cation in polymer; cat type, type of cation i.e. primary amine (1°), secondary amine (2°), tertiary amines (3°), and quaternary amines (4°).

Source: Boethling, R. S.; Nabholz, J. V., Environmental assessment of polymers under the US Toxic Substances Control Act. United States Environmental Protection Agency: 1996.

and the distribution of oligomers, physical properties, etc. In the present dataset [1], the names as well as explicit chemical structures of the polymers are not revealed, although the number of polymers with corresponding toxicity against different species (fish, *D. magna*, and green algae), and six different 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. The initial idea was to use these properties as descriptors for the modeling. However, the number of these properties is limited. Hence, we included the backbone structure of the polymers as additional descriptors. In this way, we got three more descriptors namely: C-backbone, Si-backbone, and Natural-backbone. For example, when a polymer has a C-backbone, the assigned value is 1 for the C-backbone descriptor and 0 for Si-backbone and Natural- backbone descriptors. Similarly, we defined other two indicator variables, cation position and cation type. In case of cation position properties, based on the available position for at least one occurrence, we have defined nine descriptors namely: cation position 0, 1, 2, 3, 4, 5, 7, 8, and 11. Therefore, a total set of 17 descriptors were prepared for modeling. Cation position is explored by computing nine descriptors, but not cation type. We explored the cation and the types of cations present. Many cation types are present in a polymer but were considered as single descriptor, as including the different types did not have any effect nor improvement of the models.

22.2.4 Dataset Division

To validate each individual QSTR strategically, we divided the fish dataset into training and test sets by randomly employing a 3 : 1 ratio. The training set compounds were involved in model generation, while the test set compounds were used for validation of the final models. However, for the remainder of the datasets, we have used all the polymers in model generation due to the small size of datasets and determined the robustness of the generated models using the leave-many-out cross-validation approach as it is wasteful to keep aside some molecules as test compounds in case of small datasets.

22.2.5 Model Development

The dataset of 17 descriptors with toxicity endpoints was subjected to development of a genetic algorithm (GA) using software tools freely available at https://dtclab.webs.com/software-tools. As per the size of the dataset, the length of QSTR equation was decided based on the specified criteria of 5 : 1 ratio of number of observations to the number of descriptors and a GA run was made repeatedly to obtain a number of different combinations of genetic algorithm derived multiple linear regression (GA-MLR) models [9]. In most of the cases, the resulting models were obtained by partial least squares (PLS) [2] regression applied to the selected descriptors of the GA-MLR models except the green algae 96 h dataset.

Similarly, interspecies quantitative structure–toxicity relationship (i-QSTR) models were prepared between fish, *D. magna*, and green algae to explore whether the toxicity data of one species could be helpful for the prediction of toxicity of another species. The response endpoint of one species acts as the dependent variable and the response endpoint of another species acts as one of the independent variables along with calculated descriptors [3]. The obtained model can be used for the prediction of the toxicity of untested and new polymers when toxicity to one species available and toxicity to another species is missing. Thus, these i-QSTR models can also serve as toxicity data gap filling. In this sense, these models are structurally similar to those interspecies extrapolation tools developed and used by US EPA in WebICE (https://www3.epa.gov/webice/index.html). These models are capable of extrapolating data for one toxicity endpoint to another toxicity endpoint when the data for the second species are unavailable [10]. i-QSTR can overcome the cost of multiple toxicity tests and improve the understanding of the mechanism of toxic action (MOA) of chemicals for different organisms. Such models may be more reliable compared to a single endpoint QSTR models and can thus be used in order to fill the data gaps where toxicity value for a particular compound is absent for a specific endpoint.

22.2.6 Model Validation

The most crucial step in QSTR modeling is model validation. In the current work, the generated models were validated by employing universally acceptable internal (all cases) and external validation metrics (in case of fish QSTR modeling). The model quality was ascertained by evaluating the determination coefficient (R^2), leave-one-out determination coefficient (Q^2_{LOO}), external set prediction variance (R^2_{pred}), and mean absolute error (MAE) of predictions [11, 12] for each model. The robustness of the generated models was also examined by considering the leave-many-out approach (Q^2_{LMO}). The applicability domain (AD) study was performed employing the distance to model of chemical space (DModX) approach, and a *Y*-randomization study was performed generating 100 random models using the soft independent modeling of class analogy (SIMCA)-P software tool [13]. The AD of each model was defined in the chemical space, within the domain of predictions of any compound being considered as reliable. In the DModX approach, if the compound has higher DModX score than the D-critical value, then the compound is considered as an outlier (training set) and outside the AD (test set). The *Y*-randomization study was performed to prove that the developed models were not obtained by random chance.

22.3 Results and Discussion

22.3.1 QSTR Modeling for Fish Toxicity 96 h Dataset

An initial analysis suggested that out of 38 polymers, three showing high prediction residuals might be influential observations for the QSTR model. Thus, the identified three polymers (ID#: 27, 36, 71) were removed from the dataset. The final curated dataset of 35 polymers was used for toxicity modeling against fish by dividing the dataset into two individual sets (i.e. training and test sets) using a random approach. The training set comprises 27 polymers which were used for development of the QSTR model, while the test set consists of 14 polymers used for extensive validation of the developed model. The resulting robust four-descriptor model was generated employing GA-MLR [4] followed by PLS using three latent variables (LVs) [2]. The final model was considered significant and acceptable based on the quality of predictions such as 76% predictive variance (R^2), 61% leave one out variance ($Q^2_{(LOO)}$), and 81% external predictive variance (R^2_{pred}).

$$pEC_{50(EC50\text{ in Molar})} = 8.21 + 0.142\ \text{Charge Density}(\text{a} - \text{N}\%) - 2.71\ \text{Si backbone} - 0.0577\%\text{MW less than 500 Da} - 2.31\ \text{cat pos 0}$$

$$n_{train} = 27;\ n_{test} = 8;\ \text{LV} = 3;\ R^2 = 0.767;\ Q^2 = 0.614;\ R^2_{pred} = 0.814;\ \overline{rm^2_{LOO}} = 0.493;\ \Delta rm^2_{LOO} = 0.170;$$
$$\overline{rm^2_{test}} = 0.720;\ \Delta rm^2_{LOO} = 0.143;\ \text{MAE}_{Test100\%} = 0.448;\ \text{Quality}_{Test} = \text{Moderate}$$

To determine the relative importance of each variable appearing in the final QSTR model, we performed a variable importance projection (VIP) plot analysis [14] using the SIMCA-P software tool, and found that out of the four appearing variables, three variables (Si-backbone, %MW less than 500 Da and Cat pos 0 resulted in higher VIP scores) were the most important ones, while charge density (a-N%) was considered less significant compared to the other three (Figure 22.3).

Further, we have performed a loading plot analysis to find out the most influential descriptors in the QSTR. Charge density (a-N%), Si-backbone, %MW less than 500 Da were considered the most influential descriptors as they are situated far away from the origin and Cat pos 0 is somewhat less influential variable as it is located near to the center of plot (Figure 22.4).

For the randomization study, 100 random models were generated by shuffling the value of the response variable, while the independent variables were used as it is. For the random models, the value of R^2Y intercept and Q^2Y intercept of the generated models should not exceed 0.3 and 0.05, respectively. In our case, it was found that the value R^2Y intercept and Q^2Y intercept of generated models were below than the stated limits, and the developed model was considered as non-random (not obtained by chance) (Figure 22.5).

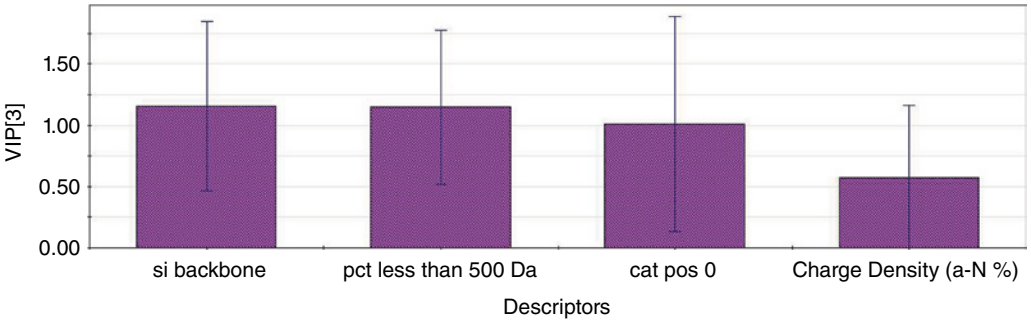
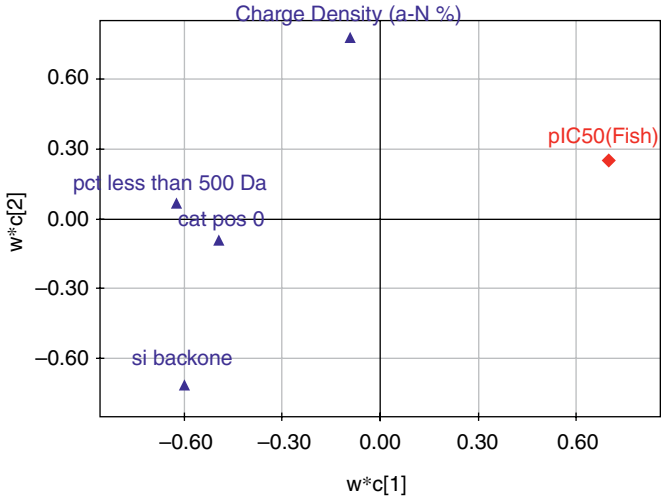


Figure 22.3 VIP plot of the PLS model for prediction of cationic polymer acute toxicity to fish.

Figure 22.4 Loading plot of the PLS model for prediction of cationic polymer acute toxicity to fish.



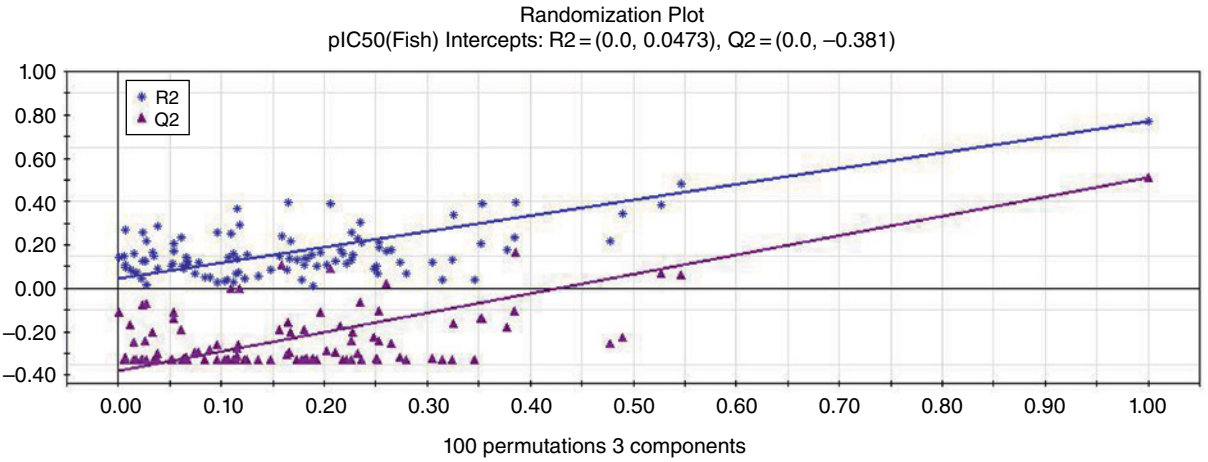


Figure 22.5 Randomization plot (unitless) of the PLS model for prediction of cationic polymer acute toxicity to fish (R^2 and Q^2 values for the random models [*Y*-axis]) are plotted against correlation coefficient between the original *Y* values and the permuted *Y* values (*X* axis).

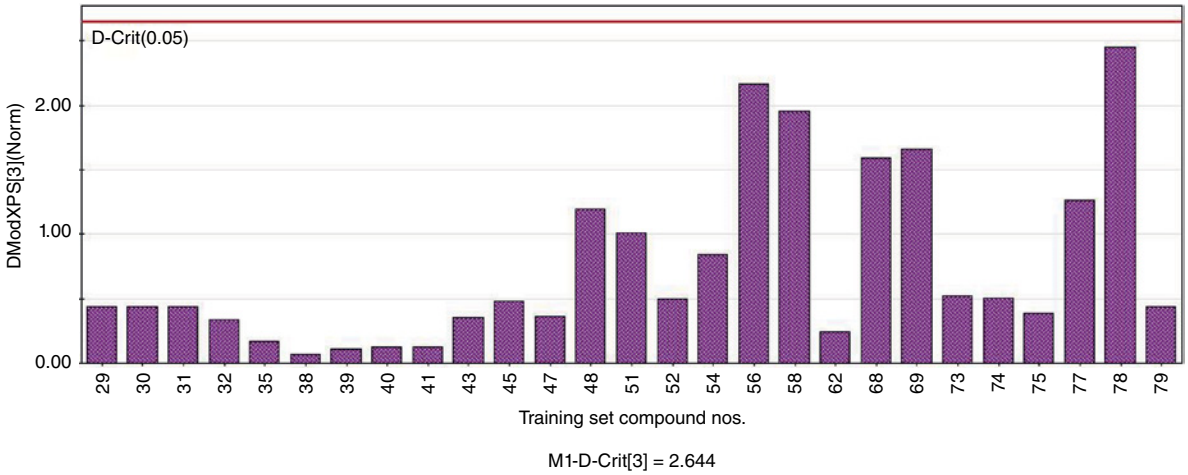


Figure 22.6 AD plot of training set compounds of the PLS model generated for prediction of cationic polymer acute toxicity to fish at 95% confident level.

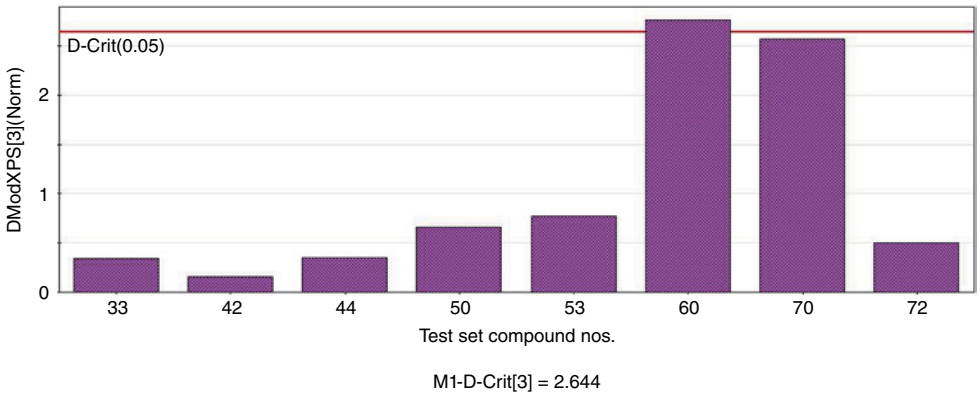


Figure 22.7 AD plot of test set compounds of the PLS model generated for prediction of cationic polymer acute toxicity to fish at 95% confidence level.

Lastly, we performed the AD study and found that none of the polymers was considered as an outlier for training set while one polymer (ID# 60) was identified outside the AD for the test set (Figures 22.6 and 22.7).

All the variables appearing in the final model demonstrate negative contributions (decreased toxicity) toward fish except charge density, suggesting that a polymer with silicone-based backbones, presence of cation at position 0 and %MW less than 500 Da leads to a decrease in the fish toxicity (pEC_{50}) (an increase in the EC_{50} value of the polymer against fish) and vice versa. It is stated that the position of cation in the polymer backbone as well as the type of polymer backbone were the most important factors to influence the aquatic toxicity of polymers [1]. For example, compounds 68 (presence Si-backbone in the chemical structure), 79 (due to presence of the cationic charge at position 0) showed a decrease in the fish toxicity (higher LC_{50} values). On the other hand, variables with a positive impact for the toxicity is cationic charge density of molecules, indicating that higher charge density results into higher toxic polymer and vice versa. For example, polymers 58 and 56 show higher charge density which results in higher toxicity to fish.

22.3.2 QSTR Modeling for *Daphnia magna* Toxicity 48 h Dataset

An initial analysis suggested that out of 20 polymers, two were showing high prediction residuals which might have a negative influence on the quality of the resulting QSTR model. Thus, the identified two polymers (ID#: 33, 39) were removed from the dataset. The QSTR model was generated by employing the final curated dataset of 18 polymers with reported experimental toxicity against *D. magna* using GA-MLR [9] followed by the PLS algorithm [10]. As the dataset is too small for division into training and test sets, instead of dataset division, we performed leave-many-out cross-validation analysis to examine the quality and robustness of the final model using MINITAB software tool using the entire dataset. The final PLS model consists of three descriptors and 2LVs.

$$pEC_{50(EC50\,\text{in Molar})} = 6.505 - 0.0990\%\text{MW less than 500 Da} - 1.231\,\text{Si backbone} + 0.446\,\text{Cation type}$$

$$n_{\text{train}} = 18,\ \text{LV} = 2,\ R^2 = 0.809,\ Q^2_{\text{LOO}} = 0.654,\ Q^2_{\text{L3}O} = 0.663,\ \overline{rm^2_{\text{LOO}}} = 0.543,\ \Delta rm^2_{\text{LOO}} = 0.124,$$
$$\text{MAE}_{\text{Training100\%}} = 0.625$$

As with the fish toxicity model, we have performed different plot analyses of the generated QSTR model against *D. magna*. The most important and influential descriptors in the final model were identified using the VIP and loading plot analyses, respectively. The plots reveal that %MW less than 500 Da was the most important as well as most influential descriptor in the final model although rest of the two descriptors were of almost similar importance and influence in the generated model (Figures 22.8 and 22.9).

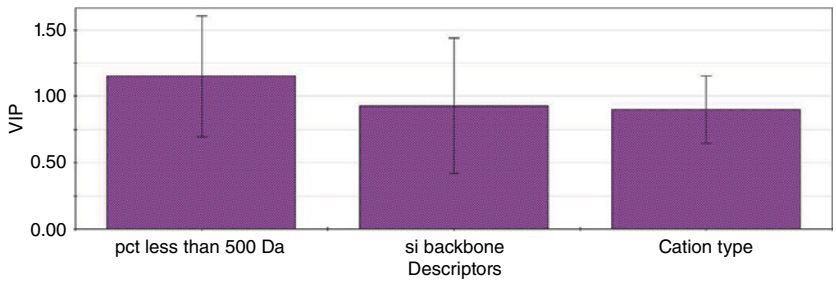


Figure 22.8 VIP plot of the PLS model for prediction of cationic polymer acute toxicity to *D. magna*.

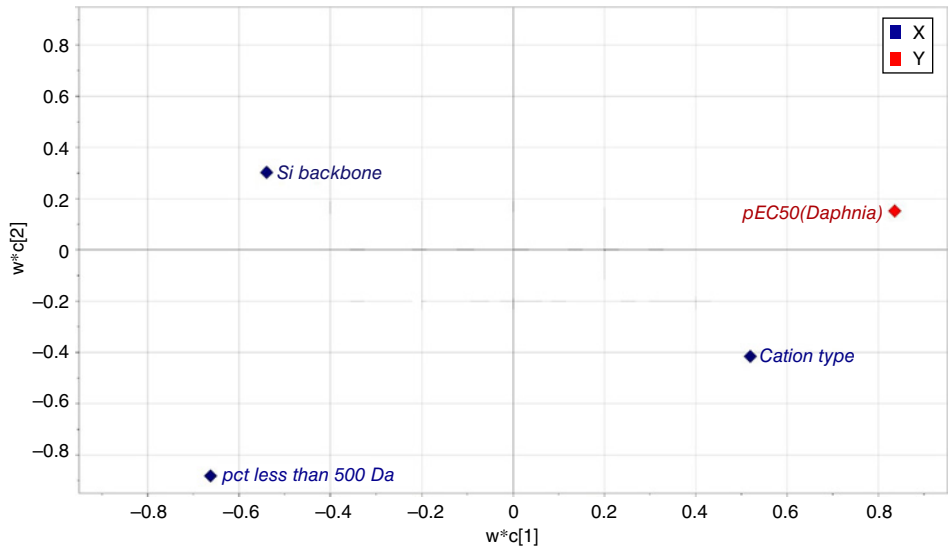


Figure 22.9 Loading plot of the PLS model for prediction of cationic polymer acute toxicity to *D. magna*.

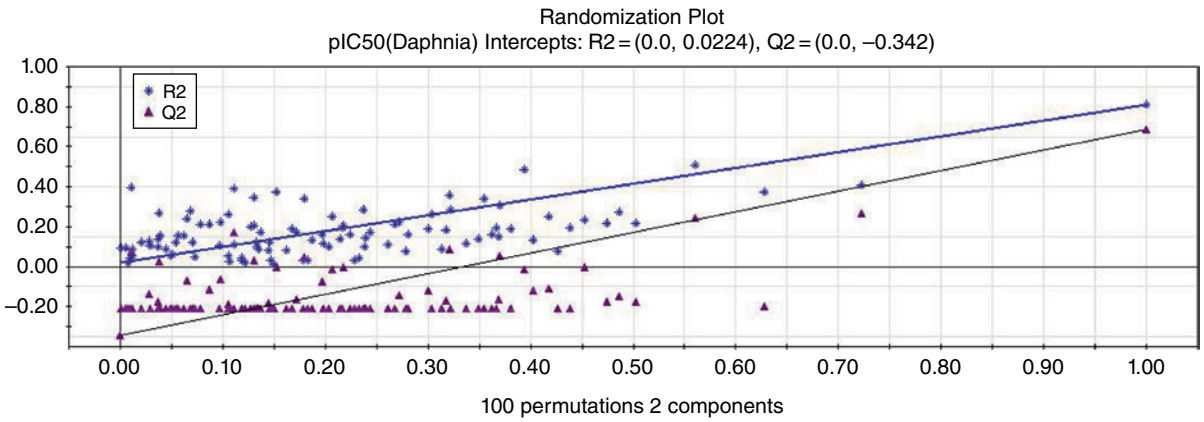


Figure 22.10 Randomization plot of the PLS model for prediction of cationic polymer acute toxicity to *D. magna* (R^2 and Q^2 values for the random models [*Y*-axis]) are plotted against correlation coefficient between the original *Y* values and the permuted *Y* values (*X* axis).

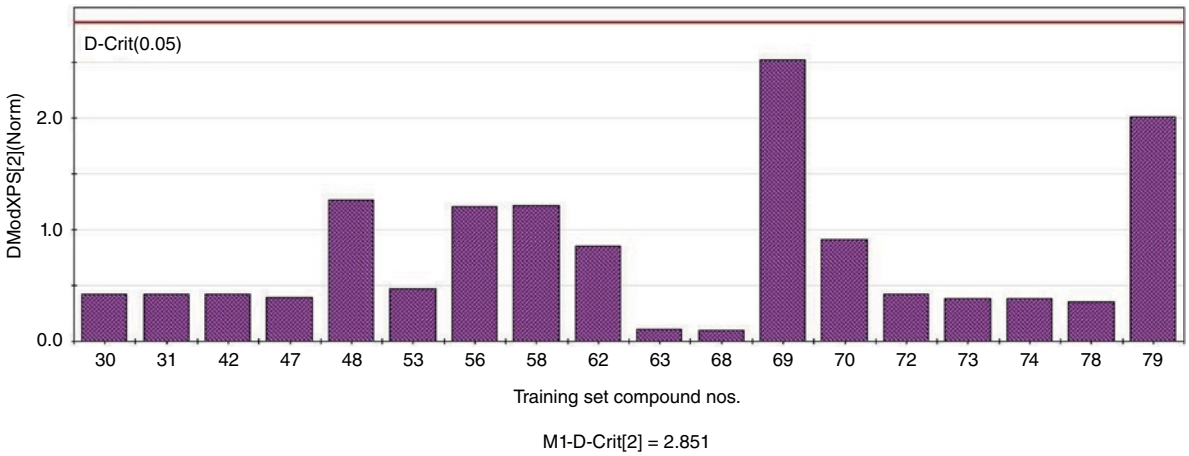


Figure 22.11 AD plot of the PLS model generated for prediction of cationic polymer acute toxicity to *D. magna* at 95% confidence level.

The *Y*-randomization study also suggested that based on the stated criteria the model was not obtained by chance (Figure 22.10).

Lastly, we checked the AD of the obtained QSTR model in the chemical space using DModX approach employing the SIMCA-P, and the AD revealed that none of the polymers were outliers (Figure 22.11).

The final PLS model comprises three unique variables with either positive or negative contributions toward toxicity against *D. magna*. The variables with a negative contribution (inversely proportional to the toxicity) include polymers with molecular mass less than 500 Da and silicone-based polymeric backbone, indicating that higher values of these variables result in a decrease in the toxicity of the polymer and vice versa. For example, compound 62 shows lower toxicity due to a high value of %MW less 500 Da variable as well as presence of silicone-based backbone. On the other hand, cation type polymers had a positive contribution. From a closer analysis of the dataset, it was importantly revealed that a polymer with a quaternary cation (compound ID# 30) in its chemical structure resulted in higher toxicity toward *D. magna*.

22.3.3 QSTR Modeling for Green Algae Toxicity 96 h Dataset

An initial analysis suggested that out of 17 polymers, three were showing high prediction residuals and these might be influential observations for the QSTR model. Thus, the identified three polymers (ID:27, 39, 42) were removed from the dataset. The resulting MLR model for toxicity prediction of polymers toward green algae was generated using the entire dataset of 14 polymers. The best model was based on only two descriptors. For the present endpoint too, we have tried to generate a PLS model, but due to the drastic reduction of quality of Q^2_{LOO} parameter, we have reported the following acceptable MLR model for toxicity prediction of polymers against green algae [9].

$$pEC_{50(EC50\,\text{in Molar})} = 5.91 + 1.61\,\text{C backbone} + 1.52\,\text{cat pos}.$$

$$n_{\text{train}} = 14;\ R^2 = 0.826;\ Q^2_{\mathrm{LOO}} = 0.631;\ R^2_{\text{Adjusted}} = 0.663;\ \overline{rm^2_{\mathrm{LOO}}} = 0.517 \Delta rm^2_{\mathrm{LOO}} = 0.145;\ \mathrm{MAE}_{\text{Training}95\%} = 0.616$$

The final MLR model based on two variables with positive contributions toward toxicity against green algae with good quality of predictions explained 82% variance for the training set (R^2) and 63% in terms of LOO variance (Q^2). The descriptors provide information about the type of backbone (carbon based) as well as position of cation (cat pos 5) in the polymer backbone. A closer analysis of the data revealed that all the compounds with carbon-based polymer backbones show average toxicity of $pEC_{50} = 7.69$ (EC_{50} in molar) and only one compound throughout the dataset with presence of cation at position 5 shows toxicity of $pEC_{50} = 7$ (EC_{50} in molar).

22.3.4 QSTR Modeling for Chronic Toxicity Against Green Algae

An initial analysis suggested that out of 16 polymers, three were showing high prediction residuals which might be influential observations for the QSTR model. Thus, the identified three polymers (ID#: 39, 42, 51) were removed from the dataset. The dataset of 13 polymers was used to predict the chronic toxicity of polymers against green algae. In this case, the final model was obtained by GA-MLR [9] followed by PLS regression [10] with one LV, while the robustness of the selected model was examined by employing the Q^2_{LMO} approach.

$$pEC_{50(EC50\,\text{in Molar})} = 6.361 + 0.013\%\text{MW less than 500 Da} + 1.318\,\text{C backbone}$$

$$n_{\text{train}} = 13;\ \mathrm{LV} = 1;\ R^2 = 0.781;\ Q^2_{\mathrm{LOO}} = 0.631;\ Q^2_{\mathrm{L3O}} = 0.606;\ \overline{rm^2_{\mathrm{LOO}}} = 0.468,\ \Delta rm^2_{\mathrm{LOO}} = 0.276;$$
$$\mathrm{MAE}_{\text{Training}100\%} = 0.343$$

From the analysis of VIP and loading plots, we have found that the most important variable in the resulting model was C-backbone, while %MW less than 500 Da was considered as the least significant descriptor (Figures 22.12 and 22.13).

The *Y*-randomization study also suggested that based on the stated criteria the model was not obtained by chance (Figure 22.14).

Finally, we performed the AD study and found that none of the polymers were outliers (Figure 22.15).

Similar to the acute green algae toxicity model, the final PLS model was based on the two variables with positive contributions toward chronic green algae toxicity with good quality of predictions explained (78% variance for the training set (R^2), 63% in terms of LOO variance (Q^2)). The descriptors provide information about the type of backbone (carbon based) as well as average number MW of polymers in thousand with percent less than 500 Da. A closer analysis of the data revealed that all the compounds with carbon-based polymeric backbone showed an average toxicity of $pEC_{50} = 7.76$ M while compounds with absence of carbon-based polymeric backbone result in an average toxicity of $pEC_{50} = 6.52$ M.

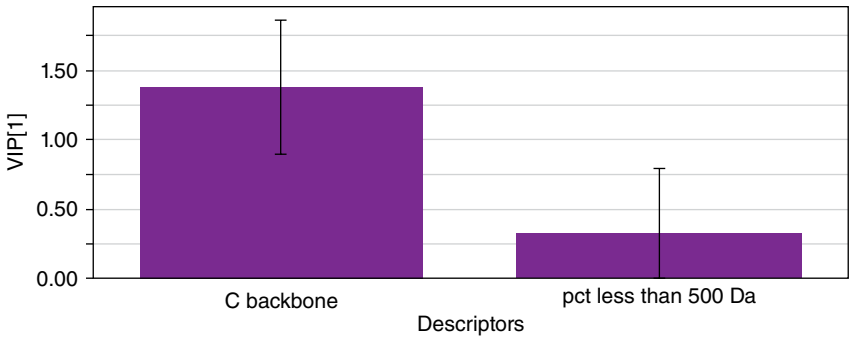


Figure 22.12 VIP plot of the PLS model for prediction of cationic polymer chronic toxicity to green algae.

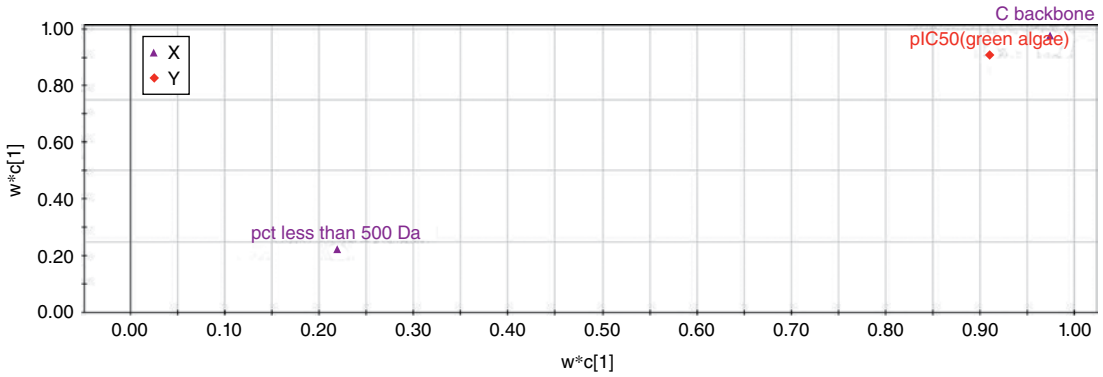


Figure 22.13 Loading plot of the PLS model for prediction of the of cationic polymer chronic toxicity to green algae.

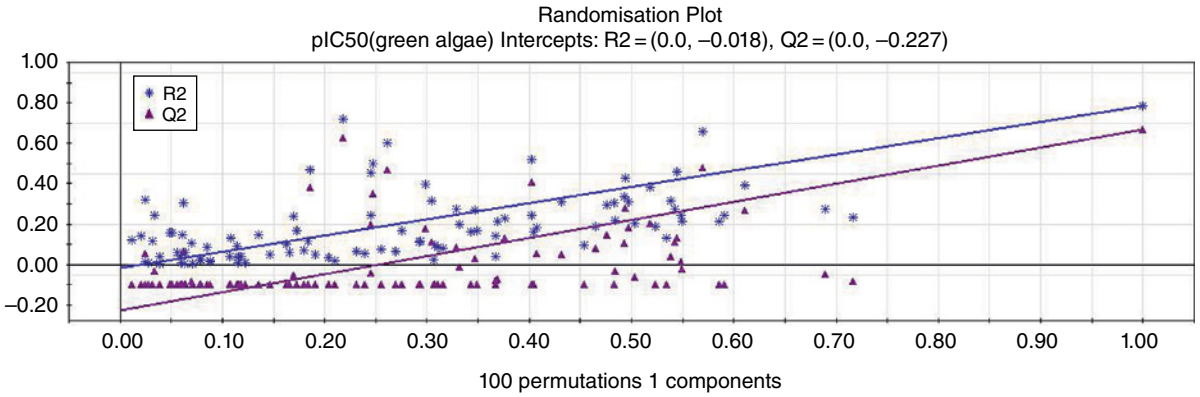


Figure 22.14 Randomization plot of the PLS model for prediction of cationic polymer chronic toxicity to green algae (R^2 and Q^2 values for the random models [*Y*-axis]) are plotted against correlation coefficient between the original *Y* values and the permuted *Y* values (*X* axis).

22.3.5 Interspecies Modeling of Polymers

22.3.5.1 i-QSTR Modeling Between *D. magna* (48 h) and Fish (96 h)

Out of 38 polymers in the fish dataset, 19 polymers were found to have their reported pEC_{50} value against *D. magna*. These 19 polymers had toxicity data for both fish and *D. magna* and were used for i-QSTR model development using the GA-MLR [9] followed by the PLS technique [10].

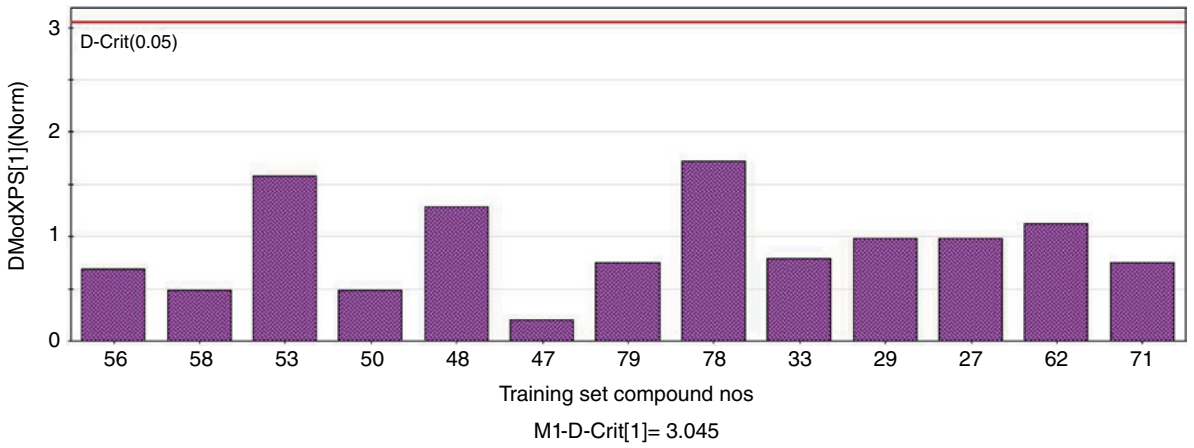


Figure 22.15 AD plot of the PLS model generated for prediction of the polymer chronic toxicity to green algae at 95% confidence level.

$$pEC_{50\_D.\ magna(EC50\text{ in Molar})} = 1.215 - 0.0402\%\text{MW less than 500 Da} + 0.339\,\text{cat pos3} + 0.860\,pEC_{50\_\text{Fish}}$$

$$n_{\text{train}} = 19;\ \text{LV} = 2;\ R^2 = 0.842;\ Q^2_{\text{LOO}} = 0.753;\ Q^2_{\text{L3O}} = 0.715;\ \overline{rm^2_{\text{LOO}}} = 0.660;\ \Delta rm^2_{\text{LOO}} = 0.147,$$
$$\text{MAE}_{\text{Training100\%}} = 0.694;\ \text{Quality}_{\text{Train}} = \text{Moderate}$$

The final model was obtained with two LVs, and subsequently the robustness was determined by the leave-many-out cross-validation approach. The final PLS model was used for prediction of *D. magna* toxicity of rest of 19 compounds of the fish dataset whose toxicity data was missing for *D. magna*. The bar plot (Figure 22.16) depicts the experimental fish toxicity and predicted *D. magna* toxicity for 19 polymers using the i-QSTR model.

We can conclude that the polymers had similar pattern of toxicity for fish and *D. magna* which is an important observation for toxicity data gap filling and risk assessment of the studied polymers. This supports observations using WebICE where *D. magna*, for example, predicts an acute toxicity of $1700\,\mu g l^{-1}$ to rainbow trout when *Daphnia* is measured to have an acute toxicity of $1000\,\mu g l^{-1}$. In WebICE, the toxicants are agnostic, being built from trends of sensitivity and tolerance to a wide array of toxicants. Importantly, insights for cationic polymers are not found in the WebICE database.

22.3.5.2 i-QSTR Modeling Between Fish (96 h) and *D. magna* (48 h) Toxicities

Out of 20 polymers in the daphnia dataset, 19 polymers were found to have their experimental pEC_{50} value against fish. These 19 compounds were used for model development using the GA-MLR [9] followed by the PLS technique [10]. The best model was based on the two components (obtained by extracting the vital information from three individual descriptors for i-QSTR modeling).

$$pEC_{50(\text{Fish})(EC50\text{ in Molar})} = 2.205 - 0.399\,\text{Si backbone} - 0.505\,\text{cat pos5} + 0.690\,pEC_{50(D.magna)}$$

$$n_{\text{train}} = 19;\ \text{LV} = 2;\ R^2 = 0.823;\ Q^2_{\text{LOO}} = 0.776;\ Q^2_{\text{L3O}} = 0.778;\ \overline{rm^2_{\text{LOO}}} = 0.689;\ \Delta rm^2_{\text{LOO}} = 0.140;$$
$$\text{MAE}_{\text{Training100\%}} = 0.509$$

The final PLS model was used for prediction of fish toxicity of single compound (ID# 63) with reported *D. magna* toxicity value.

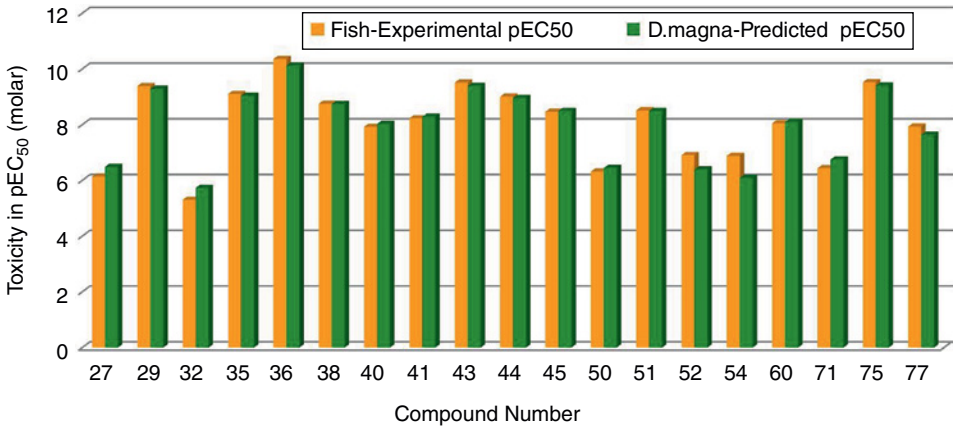


Figure 22.16 Experimental acute fish toxicity and predicted acute *D. magna* toxicity.

22.3.5.3 i-QSTR Modeling Between Acute Green Algae (96 h) and Acute Fish (96 h) Toxicities

Comparing acute green algae (96 h; EC_{50}) and fish (96 h) datasets, 17 polymers were found to have their reported pEC_{50} value for both species. These 17 polymers were employed for model development using the GA-MLR [9].

$$pEC_{50(\text{algae})(EC50\,\text{in Molar})} = 4.22 + 1.36\,\text{C backbone} + 2.75\,\text{cat pos}3 + 1.54\,\text{cat pos}5 + 0.297\,pEC_{50(\text{Fish})(EC50\,\text{in molar})}$$

$$n_{\text{train}} = 19;\ R^2 = 0.802;\ Q^2_{\text{LOO}} = 0.649;\ R^2_{\text{Adjusted}} = 0.737;\ \overline{rm^2_{\text{LOO}}} = 0.548;\ \Delta rm^2_{\text{LOO}} = 0.107$$

The resulting MLR model was employed for the prediction of green algae acute toxicity of remaining 21 polymers of the fish dataset which had experimental toxicity data to fish but acute toxicity to green algae was missing. The bar plot (Figure 22.17) depicts the experimental fish toxicity and predicted acute green algae toxicity of the above mentioned 21 polymers.

Analyzing the bar plot, we can conclude that the polymers had similar pattern of acute toxicity for fish and green algae except polymers (ID#: 32, 68, 72, 73, 74) for which green algae showed much higher acute toxicity than fish, which is an important observation of toxicity data gap filling and risk assessment of these studied polymers. Importantly, it appears that the inter-species sensitivity relationships are somewhat more variable than fish-*Daphnia* relationships, which is expected based on taxonomic distance [15].

22.3.5.4 i-QSTR Modeling Between Fish (96 h) and Acute Green Algae (96 h) Toxicities

Out of 17 polymers of the green algae (96 h) dataset, 17 polymers were found to have their reported pEC_{50} value against fish. These 17 common polymers with toxicity data for both species were used for model development using the GA-MLR [9] followed by PLS [10]. The final PLS model was obtained with one LV and acceptable values of validation metrics.

$$pEC_{50(\text{Fish})(EC50\,\text{in Molar})} = 1.42 + 0.539\,pEC_{50(\text{G.Algae})(EC50\,\text{in Molar})} + 0.00000031\,\text{av Mw} + 0.365\,\text{Cation type}$$

$$n_{\text{train}} = 17;\ \text{LV} = 1;\ R^2 = 0.705;\ Q^2_{\text{LOO}} = 0.610;\ Q^2_{\text{L3O}} = 0.56;\ \overline{rm^2_{\text{LOO}}} = 0.480;\ \Delta rm^2_{\text{LOO}} = 0.218;$$
$$\text{MAE}_{\text{Training100\%}} = 0.706$$

22.3.5.5 i-QSTR Modeling Between *D. magna* (48 h) and Acute Green Algae (96 h) Toxicities

Out of 17 polymers of the acute green algae (96 h) set, 12 polymers were found to have their reported toxicity value against *D. magna* too. These 12 compounds were used for model development using the GA-MLR [9] followed by PLS [10].

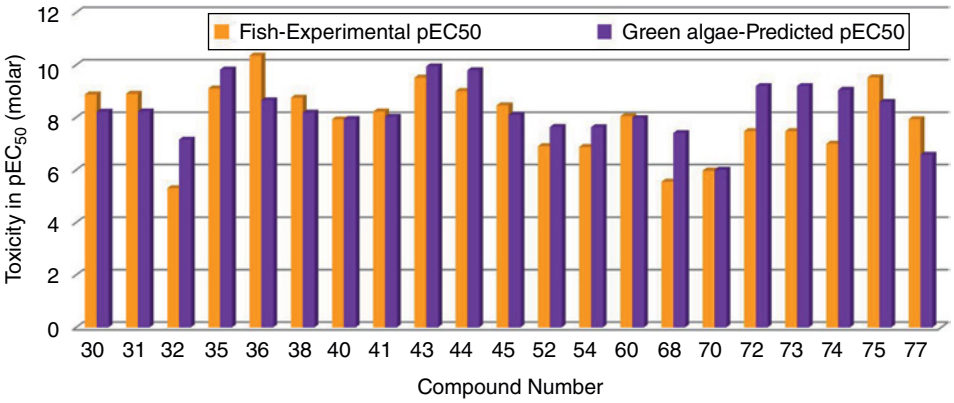
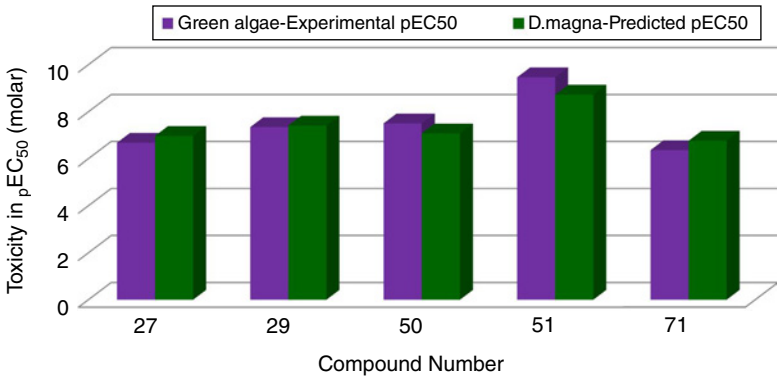


Figure 22.17 Experimental fish toxicity and predicted acute green algae acute toxicity employing the i-QSTR model.

Figure 22.18 Experimental acute green algae toxicity and predicted daphnia toxicity employing the i-QSTR model.



$$pEC_{50(D.magna)(Ec50\,\text{in Molar})} = 2.726 - 0.0911\%\text{MW less than } 500\,\text{Da} + 0.643\,pEC_{50(\text{G.Algae})(EC50\,\text{in Molar})}$$

$$n_{\text{train}} = 12;\ \text{LV} = 1;\ R^2 = 0.664;\ Q^2_{\text{LOO}} = 0.444;\ Q^2_{\text{L3O}} = 0.45;\ \overline{rm^2_{\text{LOO}}} = 0.324;\ \Delta rm^2_{\text{LOO}} = 0.155;$$
$$\text{MAE}_{\text{Training100\%}} = 1.148$$

The developed model was further employed for the perdition of toxicity of five polymers to *D. magna* which had toxicity to green algae but toxicity to *D. magna* were missing. The bar plot depicts the experimental acute toxicity to green algae and predicted toxicity to *D. magna* for these five polymers (Figure 22.18).

Analyzing the bar plot, we can conclude that the polymers had similar pattern of toxicity for fish and *D. magna* which is an important observation for toxicity data gap filling and risk assessment of these studied polymers.

22.3.5.6 i-QSTR Modeling Between Acute Green algae (96 h) *and D. magna* (48 h) Toxicities

In this case, after repeated runs of GA with changing equation length, no model was obtained with *D. magna* toxicity as the dependent variable, as the dataset was too limited and small.

22.4 Conclusions

In the present study, we have generated different QSTR models to predict acute toxicity toward three individual species; fish, *D. magna*, and green algae, using either PLS regression or MLR technique [9, 10]. The resulting model for fish toxicity prediction was validated using different internationally accepted internal and external metrics, while other models were validated using the leave-many-out approach.

The study revealed that presence of silicone based polymeric backbone and having a high %MW less than 500 Da resulted in a decreased toxicity (high EC_{50}) for fish and *D. magna*. On the other hand, higher cationic charge density (%a-N) and type of cation (especially quaternary cation) lead to an increase in toxicity to fish and *D. magna* (lower LC_{50}), respectively. In case of green algae, the toxicity of polymeric compounds are enhanced in the presence of a carbon-based polymeric backbone with cation at position 5 and with higher value for %MW less than 500 Da.

This work expands and updates the work by the USEPA [1, 5] and Nolte et al. [4]. Our work confirms that charge density (%a-N) is an important toxicity descriptor; however, we also demonstrate that the type of cation and its position impacts the toxicity significantly for fish and *D. magna*. For fish, several factors influence toxicity including: *a-N*%, Si backbone, %MW < 500 Da, and Cat position 0. For *D. magna*, %MW < 500 Da, Si backbone and Cat type influence toxicity. In the case of acute algal toxicity, the carbon-based backbone and cat position 5 govern acute toxicity. The chronic algal toxicity was influenced by the %MW < 500 Da and carbon based backbone.

The USEPA [1] found the following overall backbone driven toxicity rank: Carbon-based backbone ≥ Si-backbone = Natural-backbone across species. We have further detailed this assessment as evident from the above findings.

For the first time in science, we have developed i-QSTR models for polymers, with an objective to assess whether the response toxicity data (pIC_{50}) of one species could be helpful for prediction of toxicity of another species. From the i-QSTR modeling, we found that polymers had similar pattern of acute toxicity among fish, *D. magna*, and green algae. The resulting i-QSTR model were used for prediction of toxicity of untested compounds, which is important for toxicity data gap filling.

In regulatory application, development of sound strategies to support read across, especially for data poor categories, is very important [14]. Decision support for identifying domains of applicability and physical–chemical drivers for toxicity are necessary to qualitatively defend the establishment of chemical categories. Strong, defensible QSARs assist in quantitative hazard assessment and identify utility for both interpolation and extrapolation beyond the boundaries of the empirical data. As described earlier in this chapter, QSARs for data poor categories can become powerful additions to the assessment arsenal. European Center for Ecotoxicology and Toxicology of Chemicals (ECETOC) [15, 16] in 2019 and 2020 summarized risk assessment needs, the framework to assess human and environmental safety of polymers, and the applicability of many previously standardized test methodologies to polymer compounds. These should prove useful in the coming years as harmonized datasets from which additional polymer QSARs can be developed.

There are still many aspects pertaining to modeling the toxicity of cationic polymers let alone all polymers. For the cationic polymers, further details about the structures would help the model development, as would be possible from increasing the data base with newer datasets since the early 1990s. In addition, there are challenges to be clarified in terms of the toxicity testing of polymers in the lab and characterization of toxic mechanism of action that needs to be further analyzed. The iTAP project (http://cefic-lri.org/projects/eco-46-improved-aquatic-testing-and-assessment-of-cationic-polymers-itap) is pursuing these and other objectives to enhance the modeling of this important class of materials entering the regulatory frameworks in these years globally.

Acknowledgments

HS, AMBH, KC, and SEB appreciate CEFIC LRI ECO46 project iTAP for funding of the research behind this paper. PMK thanks the Department of Pharmaceuticals, Ministry of Chemicals and Fertilizers, Govt. of India for a fellowship. KR thanks Science and Engineering Research Board (SERB), New Delhi for financial assistance under the MATRICS scheme (File number MTR/2019/000008).

References

1 Boethling, R.S. and Nabholz, J.V. (1996). *Environmental Assessment of Polymers under the US Toxic Substances Control Act*. United States Environmental Protection Agency.

2 OECD (2009). Data analysis of the identification of correlations between polymers characteristics and potential health or ecotoxicologocal concern. ENV/JM/MONO https://www.oecd.org/env/ehs/risk-assessment/42081261.pdf

3 Sanderson, H., Khan, K., Hansen, A.M.B. et al. (2020). Environmental toxicity (Q) SARs for polymers as an emerging class of materials in regulatory frameworks, with a focus on challenges and possibilities regarding cationic polymers. In: *Ecotoxicological QSARs* (ed. K. Roy), 681–705. New York: Springer.

4 Nolte, T.M., Peijnenburg, W.J., Hendriks, A.J., and van de Meent, D.J.C. (2017). Quantitative structure-activity relationships for green algae growth inhibition by polymer particles. *Chemosphere 179*: 49–56.

5 USEPA (2013). Interpretive assisstance document for assessment of polymers. https://www.epa.gov/sites/production/files/2015-05/documents/06-iad_polymers_june2013.pdf (accessed 10 October 2020).

6 OECD (2007). *Guidance Document on the Validation of QSAR*. OECD http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono(2007)2 (accessed 10 October 2020).

7 Pecquet, A., McAvoy, D., Pittinger, C., and Stanton, K. (2019). Polymers used in US household cleaning products: assessment of data availability for ecological risk assessment. *Integrated Environmental Assessment and Management 15* (4): 621–632.

8 ECCC (2020). Environment and climate change Canada (2020). https://canadachemicals.oecd.org/ChemicalDetails.aspx?ChemicalID=4A5FBDDC-FF99-4AF2-9B01-6207A2693BE7 (accessed 10 October 2020).

9 Ambure, P., Aher, R.B., Gajewicz, A. et al. (2015). "NanoBRIDGES" software: open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and Intelligent Laboratory Systems 147*: 1–13.

10 Wold, S., Sjöström, M., and Eriksson, L.J.C. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems 58* (2): 109–130.

11 Kar, S., Das, R.N., Roy, K., and Leszczynski, J. (2016). Can toxicity for different species be correlated? The concept and emerging applications of interspecies quantitative structure-toxicity relationship (i-QSTR) modeling. *International Journal of Quantitative Structure-Propperty Relationships* 1 (2): 23–51.8.

12 Roy, K., Kar, S., and Das, R.N. (2015). *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*. New York: Springer.

13 SIMCA-P, v. U. U., Sweden (2002). SIMCA. info@ umetrics.com, www.umetrics.com (accessed 10 October 2020).

14 ECHA (2017). European Chemicals Agency. Read-across assessment framework (RAAF). ECHA-17-R-01-EN.

15 ECETOC (2019) European Centre for Ecotoxicology and Toxicology of Chemicals. Technical Report No. 133-1; Version 1. The ECETOC Conceptual Framework for Polymer Risk Assessment (CF4Polymers). Brussels, Belgium. p. 131p.

16 ECETOC (2020) European Centre for Ecotoxicology and Toxicology of Chemicals. Technical Report No. 133-2; Version 1. Applicability of Analytical Tools, Test Methods and Models for Polymer Risk Assessment. Brussels, Belgium. p. 150p.