Science of the Total Environment 889 (2023) 164337

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Predicting the ecotoxicity of endocrine disruptive chemicals: Multitasking in silico approaches towards global models

Amit Kumar Halder [a,b,*], Ana S. Moura [a], M. Natalia D.S. Cordeiro [a,**]

[a] LAQV@REQUIMTE/Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal
[b] Dr. B. C. Roy College of Pharmacy and Allied Health Sciences, Dr. Meghnad Saha Sarani, Bidhannagar, Durgapur 713212, West Bengal, India

HIGHLIGHTS

- Novel in silico ecotoxicity predictive tool for endocrine disruptor chemicals.
- First moving average–multitasking models for EDCs environmental risk assessment.
- Detection of the determining factors for eliciting higher ecotoxicity of EDCs.
- N-S and C-N distances are among the ecotoxicity determining factors for EDCs.
- Open-access tools foster reproducibility and screening for faster policy decisions.

GRAPHICAL ABSTRACT

ARTICLE INFO

Editor: Jay Gan

Keywords:
Endocrine disrupting chemicals
QSTR
Multitasking models
Machine learning
Consensus modeling

ABSTRACT

Manufactured substances known as endocrine disrupting chemicals (EDCs) released in the environment, through the use of cosmetic products or pesticides, can cause severe eco and cytotoxicity that may induce trans-generational as well as long-term deleterious effects on several biological species at relatively low doses, unlike other classical toxins. As the need for effective, affordable and fast EDCs environmental risk assessment has become increasingly pressing, the present work introduces the first moving average-based multitasking quantitative structure-toxicity relationship (MA-mtk QSTR) modeling specifically developed for predicting the ecotoxicity of EDCs against 170 biological species belonging to six groups. Based on 2,301 data-points with high structural and experimental diversity, as well as on the usage of various advanced machine learning methods, the novel most predictive QSTR models display overall accuracies > 87% in both training and prediction sets. However, maximum external predictivity was achieved when a new multitasking consensus modeling approach was applied to these models. Additionally, the developed linear model provided means to investigate the determining factors for eliciting higher ecotoxicity by the EDCs towards different biological species, identifying several factors such as solvation, molecular mass and surface area as well as the number of specific molecular fragments (e.g.: aromatic hydroxy and aliphatic aldehyde). The resource to non-commercial open-access tools to develop the models is a useful step towards library screening to speed up regulatory decision on discovery of safe alternatives to reduce the hazards of EDCs.

1. Introduction

Endocrine disrupting chemicals (EDCs) or endocrine disruptors, have been evidenced to disturb the functionalities of hormones leading to detrimental health hazards (Lauretta et al., 2019; De Coster and van Larebeke, 2012; Kumar et al., 2020). Even though regulation of hormone receptors

* Correspondence to: A.K. Halder, LAQV@REQUIMTE/Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal.
** Corresponding author.
E-mail addresses: amit-halder@fc.up.pt (A.K. Halder), ncordeir@fc.up.pt (M.N.D.S. Cordeiro).

http://dx.doi.org/10.1016/j.scitotenv.2023.164337
Received 2 March 2023; Received in revised form 10 May 2023; Accepted 17 May 2023
Available online 19 May 2023

remains the main mechanisms of EDCs, other mechanisms such as alteration in metabolism and transport of hormones or the inhibition of biosyntheses of hormones have also been reported (Vafeiadi et al., 2015; Schug et al., 2011). Moreover, the wide array of receptors classes through which EDCs act, such as the molecular mechanisms of the reproductive system, may be contributing to infertility and subfertility (De Coster and van Larebeke, 2012). The situation is the more worrisome as some of the mechanisms upon which EDCs act via this plethora of receptors are unknown (Kumar et al., 2020).

As EDCs comprise in their category a diverse and wide array of chemicals, such as pharmaceuticals, cosmetic products, pesticides, herbicides and even some natural chemicals (Marty, 2014), there is a significant probability of their release in high amounts to the environment, which presents a severe hazardous effect on various biological species including humans, something that raised increasing awareness in recent years (Velmurugan et al., 2017). In addition, EDCs, unlike other classical toxins, may induce trans-generational as well as long-term deleterious effects on several biological species at relatively low doses (Kumar et al., 2020; Velmurugan et al., 2017).

As it become clear the urgent need for regular, if not continuous, environmental assessment of ECDs release, the biological testing of EDCs against different species, similar to other hazardous materials released in the environment, is expensive, time-consuming, and demands usage of laboratory animals (He et al., 2018; Khan et al., 2019). The last decade has witnessed the advocating of many regulatory agencies worldwide for the application of in silico modeling for prediction of hazardous effects for newly developed chemicals in order to address these shortcomings. Such recommendations come with no surprise as different machine learning techniques, thanks to the development of highly accurate models, provide a highly effective cost-effective strategy for prediction of the environmental toxicity of new chemicals (Heo et al., 2019; Sheffield and Judson, 2019).

In fact, there have been recent works attempting to develop quantitative structure toxicity relationship (QSTR) models with diverse classes of EDCs. In 2018, one-target conventional QSTR models on the toxicity of EDCs have been developed for targeting eight different types of fish (He et al., 2018), while in 2019, two works proposed toxicity prediction for EDCs, one through deep learning driven classification based QSTR models for 125 and 114 chemicals with inhibition against the sex hormone-binding globulin and estrogen receptor (Heo et al., 2019), and the other with not only the development of one-target QSTR models characterizing the environmental toxicity EDCs for fourteen species that belong to nine different species groups, but also with interspecies quantitative structure–toxicity–toxicity relationships (i-QSTTR) models (Khan et al., 2019). Similarly, the Monte Carlo methodology has been successfully implemented to develop QSTR models for determining the activity of a range of endocrine disrupting chemicals (Toropova et al., 2015).

Notwithstanding, these works did not resource to the so-called multi target or multi-tasking (mtk) QSTR modeling approaches, which allow to develop single QSTR models that can probe the bio-physicochemical profiles of chemicals or materials in multiple experimental and/or theoretical conditions, providing more feasible and versatile in silico tools.

In the present study, we opted for a moving average-based multitasking modeling (MA-mtk) approach to develop effective in silico classification tools for the future toxicity prediction and environmental risk assessment of EDCs versus numerous environmental species. Though the details of this classification approach and its latest advances have been extensively described and discussed lately (Speck-Planche and Cordeiro, 2015; Speck-Planche and Cordeiro, 2017; Halder et al., 2022a), a brief summary for introductory purposes will be made.

MA-mtk modeling takes advantage of the Box-Jenkins moving average algorithm (Box et al., 2015) to derive models with as many experimental assay and/or theoretical conditions as possible, with the resulting models being able to predict outcomes against those conditions simultaneously (Speck-Planche and Cordeiro, 2015; Speck-Planche and Cordeiro, 2017; Halder et al., 2022a; Kleandrova et al., 2021; Ambure et al., 2019; Halder and Cordeiro, 2021a). Such an approach thus allows to overcome the conventional 'one target' QSTR models' handicap, i.e., overcome the one target – one bioactivity that characterizes the later conventional strategy.

Essentially, the Box-Jenkins based moving average approach applies simple formula to produce deviation descriptors, $\Delta(D_i)c_j$, from the original descriptors (D_i), incorporating both structural and experimental/theoretical variations:

$$\Delta(D_i)c_j = D_i - avg(D_i)c_j \quad (1)$$

where the $avg(D_i)$ c_j stands for the arithmetic mean of active/positive datapoints of a specific element of the experimental and/or theoretical conditions (ontology), c_j.

To assure the dynamic nature of the ontological variable, thus enabling to adjust to any experimental conditions, c_j is usually defined as dependent on the types of measures of toxicity (e.g. LC_{50}), different bio-targets (e.g. bacteria), assay times (e.g. duration of exposure), and/or any other variable that will build a unique identity for the toxicological profile of the problem. Specifically here, to define the variations in experimental assay conditions of the data retrieved from the ECOTOX database (https://cfpub.epa.gov/ecotox/), three experimental elements were considered, namely: the species name (i.e., *spn*) against which the chemical have been assayed; the species group (i.e., *spg*) specifying under which group the species belongs to; and the concentration type (Standardized) (i.e., *co*), where *co* can be 'total', 'formulation' and 'active ingredient'. Following the ECOTOX terminology, the term 'active ingredient' refers to the chemical substance within a product that causes a toxic response. The 'formulation' term is specifically used in the context of commercial products that have been prepared for practical use, such as pesticides. Finally, the term 'total' is used to refer to the overall, un-ionized or dissolved concentration of chemicals present.

The MA-mtk QSTR classification models can then be established by implementing a particular modeling technique (e.g. LDA) as follows:

$$\mathrm{Tox}_{c_j} = a_0 + \sum_k b_k \cdot \Delta(D_i)_{c_j} \quad (2)$$

in which a_0 stands for the constant term, b_k for the coefficients of the input independent variables and Tox_{c_j} for the toxicological endpoint (in this case, a categorical binary variable) to be predicted for any chemical *i* under a specific set of experimental conditions c_j.

In this work, both linear and non-linear MA-mtk models were developed by applying a range of feature selection algorithms and machine learning tools, as well as by taking in consideration the guidelines of the Organization for Economic Co-Operation and Development (OECD) (Cumming et al., 2013) for their successful implementation in regulatory assessments. In so doing, not only the developed models in the present work were employed for more than 2,000 toxicity data recorded on 170 environmental species belonging to six different species groups, but also several consensus models were derived to pick the one with maximum predictive accuracy. In fact, consensus prediction is often preferred to improve the predictivity as an ensemble of models can provide better results than one single model (Roy et al., 2018; Valsecchi et al., 2020). Moreover, the open-access *in-house* tool QSAR-Co-X (Halder and Cordeiro, 2021a) was used here for the calculation of deviation descriptors as well as for setting up most of the linear and non-linear models. The other employed open-access tools will be discussed in the appropriate sections of the paper.

This MA-mtk in silico modeling approach to the environmental risk assessment of EDCs is not only novel but it provides mechanistic interpretations towards higher ecotoxicity of EDCs, which are also discussed in detail in the present work. Furthermore, besides providing identification of the determining factors at the root of high ecotoxicity in the EDCs, it also delivers a model landscape for future and subsequent in silico approaches aiming to obtain safe alternatives to reduce the hazards of EDCs.

2. Materials and methods

2.1. Dataset collection, curation and descriptor calculation

The dataset used for setting up the QSTR models consist of 139 chemicals collected from recent work (Khan et al., 2019), with a specific toxicological end point LC_{50} (50% lethal concentration; duration: 4 days) against 170 species taken from the ECOTOX database.

Placing cut-off value for the data points regarding the measure of toxicity LC_{50} follows the guidelines of the globally harmonized system of classification and labelling of chemicals (GHS) (*Globally Harmonized System of Classification and Labelling of Chemicals (GHS)*, 2021), i.e., a chemical shall pass from Category 4 ('Warning') to Category 3 ('Danger') if the inhalation values are 10 mg/l < LC_{50} < 20 mg/l for vapor, and 1 mg/l < LC_{50} < 5 mg/l for dust/mist. Also in GHS, it is mentioned that for aquatic animals, acute toxicity Level 2 (which is close to 'Warning') is in fact 1 mg/l < LC_{50} (96 hr.) < 10 mg/l. As such, fixing the dose as 5 mg/l and considering that both the mean and median molecular weights of the dataset are approximately 250 mg, the cut-off value was set to 0.02 mM (=5/250). Therefore, data-points with LC_{50} value less than 0.02 mM were considered as toxic ($Tox_{c_j} = +1$) and otherwise non-toxic ($Tox_{c_j} = -1$). Yet some data-points were found to be expressed in ">" (greater than) or "<" (less than) notations. Thanks to that, 83 toxicity data-points with LC_{50} values expressed as $> 7.5 \times 10^{-5}$ mM to > 0.019 mM were grouped as toxic ($Tox_{c_j} = +1$), whereas 5 toxicity data-points with LC_{50} values expressed as < 0.07 mM to < 1.95 mM were treated as non-toxic ($Tox_{c_j} = -1$).

The removal of duplicated molecules ensued, taking in consideration that for MA-mtk modeling, duplicate samples are not merely defined by duplicate structures since the dataset may consist of duplicate structures if their experimental assay conditions are varied. After the analysis and subsequent removal 2,301 datapoints were obtained.

The Simplified Molecular Input Line Entry System (SMILES) (Weininger, 1988) notation of the chemicals was retrieved from ECOTOX, and after being converted to .sdf format using the Discovery Studio Visualizer tool (BIOVIA, 2021), these were further processed by the Standardizer tool (Standardizer, 2010) as described in our previous work (Halder et al., 2022b). Calculation of descriptors for the standardized structures then followed using the Dragon software (Mauri et al., 2006) under the OCHEM webserver (Sushko et al., 2021). For the calculation of 3D descriptors, geometrical optimization of the chemical structures was performed using the Corina software (Sadowski et al., 2002) under the OCHEM web platform. Extensive details of the dataset used in the current work can be found in Table S1.

2.2. Dataset division and descriptor pre-treatment

The dataset division plays a particularly important role in MA-mtk QSTR modeling. The QSAR-Co-X tool was first used to divide the data into a training set and a validation set by employing the *k*-means cluster analysis (*k*-MCA) technique (Gore Jr., 2000), where 30% data was placed in the test set. The allocation of this percentage of data to the test set was provided by the *k*-MCA rational data-distribution strategy in which after both the response and independent variables being clustered into ten groups, 30% of the data was extracted randomly (random seed 3 in QSAR-Co-X) from each group to build the validation set. After dividing the starting dataset into a training set (70%) and validation set (30%), the Box-Jenkins based moving average technique was applied to the training set data to calculate the deviation descriptors, after which this dataset was randomly (random seed 4 in QSAR-Co-X) divided into a sub-training set (80%) and a test set (20%). In addition, the average descriptors values calculated for the training set were deployed to calculate the deviation descriptors for the validation set.

The models developed in the current work were developed with the sub-training set data, while their external predictivity was assessed with both the test and the validation sets. As the validation set does not participate either in the descriptor calculation or in the model development, it serves as a 'true' prediction set. Therefore, the external predictivity against this prediction set plays a major role in the selection of the best QSTR models.

Moreover, the test set, which participates in the calculation of the deviation descriptors serves multiple purposes. Though similarly to the validation set, it is also used to assess the models' external predictivity, it may be used as a calibration set for judging the most predictive model to be further analyzed with the validation set (Halder et al., 2022a, 2022b; Halder and Cordeiro, 2021a). Notice that if a large difference is found regarding the predictivities obtained against these two prediction sets, this may indicate that the developed model is slightly biased towards the moving average technique. As such, the predictivity towards the validation set becomes more important as far as the model selection is concerned for future predictions.

Before developing the models, a descriptor pre-treatment was carried out to remove highly correlated and invariant descriptors by setting a correlation cut-off value of 0.95 and a variation cut-off of 0.001, narrowing down the initial more than 15,000 deviation descriptors to the final 3246 deviation descriptors considered for model development.

2.3. Linear model development

In this work, we opted for the linear discriminant analysis (LDA) technique (Tinsley and Brown, 2000) to find classification-based QSTR linear models that best describe the target toxicological endpoint (i.e.: LC_{50}). As the development of linear models largely depends on the feature selection algorithm applied, multiple LDA models were developed using three different procedures, namely: (a) fast-stepwise selection (FS), (b) stepwise forward selection (SFS) and (c) genetic algorithm (GA) (Halder and Cordeiro, 2019).

A maximum of ten descriptors was allowed in all linear QSTR models for comparative analyses of their statistical parameters. In FS, the descriptors with maximum significance with the response variable, i.e., the lowest *p*-value, are included step-by-step and these are retained/dropped depending on its concordance with other descriptors of the model. For this feature selection technique, both '*p*-value to enter' and '*p*-value to remove' were set as 0.05 (Halder and Cordeiro, 2021a). Regarding SFS, the LDA models were generated by varying different scoring functions, i.e., through the accuracy and scores obtained from the area under the receiver operating characteristic curve (ROC-AUC score) and selected by cross-validation (CV) techniques as follows: no CV, 5-fold CV and 10-fold CV. Note that the Python-based Mlxtend library (http://rasbt.github.io/mlxtend/) was utilized to produce the SFS-LDA models with the help of QSAR-Co-X (Halder and Cordeiro, 2021a). For setting up the GA-LDA models, we used another in-house Java-based open access tool (i.e.: QSAR-Co) with the default parameters (Ambure et al., 2019; Halder and Cordeiro, 2019).

The most predictive models obtained from these feature selection algorithms were subjected to an analysis named 'post-selection similarity search-based modification', or simply PS3M, with the help of the open-access *in-house* tool named PS3M_v2 (available at https://github.com/ncordeirfcup/PS3M_v2) (Halder and Cordeiro, 2020). A full in-depth description of this methodology can be found elsewhere (Halder et al., 2022b; Halder and Cordeiro, 2021b) as well as in Supplementary materials.

2.4. Non-linear model development

A range of non-linear classification QSTR models was developed in this work with different types of descriptors, namely: the descriptors of the most predictive linear model, 3,246 deviation descriptors obtained after pre-treatment, and the descriptors selected from differential Shannon entropy (*dSe*). The IMMAN software (Urias et al., 2015) was utilized for performing the *dSe* based feature selection as descriptors selected by this technique have been reported to be highly significant in non-linear MA-mtk modeling efforts carried out in recent years (Speck-Planche, 2020). Additionally, the feature extraction technique so-called genetic algorithm-based *k*-nearest neighbors (GA-*k*NN) (Cover and Hart, 1967; Gunturi et al., 2010) extracted the most significant descriptors from the pre-treated 3246 descriptors, by

employing the Python based *sklearn-genetic* program (https://github.com/manuel-calzolari/sklearn-genetic) that utilizes a 'deap' function for running GA. An adequate choice of the parameters used for running the GA (population: 100; mutation probability: 0.2; cross-over probability: 0.5; number of generations: 100; number of generations no change: 10; cross-over independent probability: 0.1; mutation independent probability: 0.05; scoring: accuracy; cross-validation: 5) and the *k*NN (number of neighbors: 12; weights: uniform; algorithm: auto) allowed extraction of the key number of descriptors.

The non-linear models were developed by employing six machine learning (ML) classifier tools, namely: (a) *k*NN; (b) Random Forests (RF); (Breiman, 2001) (c) Support Vector Classifier (SVC) (Boser et al., 1992); (d) Multilayer Perception (MLP) (Guang-Bin and Babri, 1998); (e) Gradient Boosting (GB) (Friedman, 2001); and (f) Deep Neural Networks (DNN) (Muratov et al., 2020; Sosnin et al., 2018). Save for DNN, the ML classifier tools were established using QSAR-Co-X (Module 2) through hyperparameter optimization, i.e., following a 5-fold cross-validation carried out with the sub-training set to determine the optimized parameters (Halder and Cordeiro, 2021b). A detailed description of the parameters varied during this ML optimization process has been provided before (Halder and Cordeiro, 2021a). Wherever applicable, a random seed value of 42 was set in the QSAR-Co-X tool.

The DNN models were built using the Python-based Keras module (https://www.tensorflow.org/api_docs/python/tf/keras). The basic architecture of this DNN (*vide* Table S2) was adopted from the literature (Sosnin et al., 2018) and it consists of six hidden layers with a gradual decrement in the number of neurons and activation function Relu. Hyperparameter optimization was performed with the sub-training set to optimize the learning rate, batch size and epochs using 5-fold CV. Details regarding the implementation of these tasks are provided in the exemplary Jupyter notebook file: DNN_M3PS3M.ipynb, available at https://github.com/ncordeirfcup/Endocrine_disruptors. The external predictivity of the optimized models was first established using the test set and finally with the validation set.

2.5. Model evaluation

The statistical robustness and predictivity of the various models was made by resorting to the QSAR-Co-X tool as reported in our previous work (Halder and Cordeiro, 2021a). Full details about the statistical parameters used for assessing the quality of the models can be found in Supplementary materials.

No matter how validated any QSTR the model might be, it is also important to assess its applicability domain (AD), that is, the scope within which it makes reliable predictions for new chemicals. For such a goal, the applicability domain of both the linear and non-linear models were determined by using the standardization (Roy et al., 2015) and confidence estimation approaches (Ambure et al., 2018), respectively.

2.6. Consensus modeling

Finally, we resort to consensus modeling for checking if all combinations of the most predictive models will eventually lead to improved models. It should be noted that, in consensus modeling, using an odd number of models can foster a more straightforward classification, given that the prediction appearing maximum times can be taken as the consensus prediction result. In contrast, for an even number of models a conflict might appear when prediction of 50% models is found to be opposite. To overcome this, the following rules are proposed and implemented:

(a) An odd number of models implies the dependence of the consensus prediction on the predictions maximum voting, as these are obtained from a maximum number of models.
(b) When an even number of models are involved, then there is either consensus or conflict. In case of conflict, resource to the posterior probability of the models should follow and the prediction with maximum value posterior probability was counted.

A novel tool, integrated to QSAR-Co-X, was developed to facilitate fast calculations of consensus prediction with several models for a large number of data-points. In this new module, the users may provide as many as models in Excel format data and results of all possible combinations are produced according to the rules described above.

3. Results and discussion

3.1. Linear MA-mtk model

As referred to above, we began by seeking mtk-QSTR classification-based linear models derived from the sub-training set, by combining the LDA and the three different feature selection algorithms (i.e.: FS, SFS and GA) along with DRAGON descriptors for the structure representation of the chemicals understudy. A summary of the statistical performance for these eight models (models M1-M8) is given in Table S3. As can be seen, the statistical results clearly indicate that non-stochastic feature selection procedures such as FS and SFS are more successful for achieving more accurate LDA models compared to stochastic ones like GA. Moreover, M6, M2, M7 and M3 are found as the four most predictive models, with MCC (MCC can take values from −1 to +1, in which −1 stands for total inconsistency between predictions and observations, 0 for random and +1 for perfect predictions) values of 0.716, 0.715, 0.700 and 0.693, respectively (Boughorbel et al., 2017). These four models were then subjected to PS3M analysis in search for more predictive models, having all save for model M6 yielded models with a higher statistical quality. The statistical results of the PS3M-based models are shown in Table S4. The M3p model in Table S4, with the highest average MCC, was developed starting from M3 after ten steps of PS3M, as shown in Fig. S1. In each PS3M step, a new model was created by replacing one descriptor of the existing model with another similar descriptor with low Euclidean distance from it. After each step, the new model was accepted only when its internal predictivity was improved or stayed the same. When the accuracy towards the sub-training set remained unaltered (cf. steps 6–8), the test set was used as a calibration set. Fig. S1 demonstrates that in each step, the overall predictivity of the models kept on increasing to a steady state. Regarding models M7 and M2, the PS3M analysis yielded better models after three steps and one step, respectively.

The resulting best-predictive model found by PS3M (a ten-variable equation, model M3p) is given below, whereas its selected descriptors are shown in Table S5.

$$\begin{aligned}\mathrm{Tox}_{c_j} = {} & +0.498\Delta(\mathrm{X2sol})_{spg} + 5.225\Delta(\mathrm{R5m+})_{spn} \\ & - 0.047\Delta(\mathrm{P\_VSA\_LogP\_4})_{co} + 3.487\Delta(\mathrm{nArOH})_{spg} \\ & + 11.589\Delta(\mathrm{nRCHO})_{spn} - 15.796\Delta(\mathrm{VE2\_B(m)})_{spn} \\ & - 3.587\Delta(\mathrm{Eig07\_EA(dm)})_{spn} - 1.240\Delta(\mathrm{Mor11v})_{spg} \\ & - 0.074\Delta(\mathrm{T(N\ldots S)})_{spn} + 0.110\Delta(\mathrm{F06[C-N]})_{spn} + 2.874\end{aligned} \quad (3)$$

A summary of the statistical performance of this model is given in Table 1. As can be seen, in fact, M3p is a very balanced LDA model in terms of its internal and external predictivity (i.e., with MCC values of 0.727, 0.739 and 0.711 against the training, test and validation sets, respectively). Furthermore, M3p provides a satisfactory goodness-of-fit with a Wilks' λ value of 0.500 (λ can take values from zero, ideal discrimination, to one, no discrimination) while affording approximately an overall accuracy > 87% (Wilks, 1932). More significantly, this model delivers an accuracy of 86.8% for the validation set. One can also see that the model is a truly statistically significant classifier, since the area under the ROC curve for all sets (AUROC values in Table 1) are higher than the one of a random classifier (AUROC = 0.5), lying all points in the upper left triangle space (see Fig. S2) (Fawcett, 2006; Hanczar et al., 2010).

Further analysis of the M3p model should only be carried out after checking the non-multicollinearity among its descriptors. Following the observation that the model has a maximum intercollinearity (absolute value)

Table 1
Statistical results for the most predictive linear model M3p.[a]

LDA statistics	Sub-training set	Test set	Validation set
$\lambda = 0.500$	TP = 813	TP = 183	TP = 421
$F = 126.95$	TN = 316	TN = 97	TN = 183
$p < 10^{-16}$	FP = 121	FP = 34	FP = 70
	FN = 34	FN = 7	FN = 22
	Sn = 72.3	Sn = 74.0	Sn = 72.3
	Sp = 96.0	Sp = 96.3	Sp = 95.0
	Acc = 87.9	Acc = 87.2	Acc = 86.8
	F1-score = 91.3	F1-score = 89.9	F1-score = 90.2
	MCC = 0.727	MCC = 0.739	MCC = 0.711
	AUROC = 0.841	AUROC = 0.852	AUROC = 0.837

[a] Model refined from the original linear model M3 found (see Table S3) by applying the PS3M technique. The abbreviations for the statistical metrics presented are as follows: the Wilks' lambda (λ), the Fisher ratio (*F*) and the corresponding *p*-value; true positives (TP), true negatives (TN), false positives (FP); false negatives (FN); sensitivity (Sn); specificity (Sp); accuracy (Acc), F1-score, the Matthews correlation coefficient (MCC), and the area under the curve for the receiver operating characteristic curve (AUROC). The metrics Sn, Sp, Acc, and F1-score are given in percentage.

of 0.727, meaning that its descriptors can be considered independent, the Y_c randomization test (Halder and Cordeiro, 2021a) was then performed. The produced randomized models afforded an average Wilks' λ of 0.993 and accuracy of 65.0%, which shows that both the goodness-of-fit and accuracy of such models are considerably poorer than those of the original model. The statistical uniqueness of model M3p is thereby established. Finally, the applicability domain of M3p, using the standardization approach as suggested by Roy et al. (2015), yielded 66, 23 and 39 data-points as outliers for the sub-training set, test and validation sets respectively, though most of these (120 out of 128) were found to be correctly predicted by the model. Therefore, in order to retain the structural and experimental diversity in the dataset, we refrained from removing these structural outliers and what is more, the whole dataset was subsequently used for development of the consensus models (see below).

3.2. Non-linear MA-mtk model

Even though we found a very acceptable linear model, previous investigations revealed that non-linear models generally afford predictive accuracies much higher than linear models, yet somehow at the expense of their mechanistic interpretation (Halder et al., 2022a). Therefore, non-linear models were generated by applying different advanced machine learning techniques essentially to check how these perform compared to the proposed linear model M3p. To do so, non-linear MA-mtk models were set up by applying the six already referred ML techniques (i.e.: *k*NN, RF, GB, MLP, SVC and DNN) after hyperparameter optimization, and based on using the following:

(a) Descriptors that appear in the best linear model, i.e., the descriptors of M3p.
(b) Descriptors selected from Differential Shannon entropy (*dSe*).
(c) Descriptors selected by applying the GA-*k*NN technique, which provided 36 descriptors (see details in Table S6) after 4 consecutive steps.
(d) All kind of descriptors obtained, after applying pre-treatment with a variance cut-off value 0.001 and an intercorrelation cut-off of 0.95.

The predictive accuracy of these non-linear models, summarized in Table S7, demonstrates that irrespectively of the descriptors employed for model development, RF and GB stand out as the more successful classifiers. Indeed, the overall predictive accuracy of the RF and GB non-linear models steadily improved with an increasing number of descriptors. The models generated with the descriptors of M3p (i.e., models M8-M13 in Table S7) provided results considerably better than those developed with the same number of descriptors selected from differential Shannon entropy. Interestingly, no matter what ML tool was applied, the descriptors of M3p yielded predictive models in a consistent manner, further confirming their importance in proper shaping the toxicity for diverse EDCs. However, that was not the case for the descriptors picked by differential Shannon entropy since the predictivity of its models varied considerably.

SVC produced a non-linear model with the descriptors of M3p (i.e., M12 in Table S7), the overall predictive accuracy of which is close to that of M3p. Table S7 also reveals that the GA-*k*NN feature extraction technique was particularly useful in extracting the most crucial descriptors for developing non-linear models. The predictive accuracy of the GA-*k*NN-based RF model (i.e., M20 in Table S7) should be considered as one of the most predictive models in terms of overall accuracy. In addition, as far as the external validation is concerned, this model depicts the highest MCC value for the external validation set, i.e., 0.726.

Worth mentioning here is that the significant differences obtained between predictivities against two prediction sets (i.e., test and validation sets) for models such as M20, M23, M26 and M29, indicated that with a larger number of descriptors, the ML tools showed tendency to generate more overfitted models, either based on the descriptors or the moving average technique. However, the descriptors selected by GA-*k*NN were able to produce predictive models with each ML tool but that did not happen when all descriptors were employed. For example, *k*NN and SVC failed to generate predictive models with all descriptors. Nevertheless, a maximum overall predictivity was gathered from M26, which is a GB-based model derived with all descriptors. Notwithstanding, this is also one of the most overfitted models, with MCC values for the sub-training set, test and validation sets of 0.779, 0.819 and 0.712, respectively. More importantly, no non-linear model did have an external predictivity considerably higher towards the external validation set as compared to the best linear model - i.e., to M3p. Moreover, the latter itself provides mechanistic interpretations in a far more straightforward way than any of these predictive non-linear models. Yet, on the basis of the overall predictivity as well as on the predictive accuracy towards the external validation set, four models, namely M20, M23, M26 and M29 were chosen for further refinement, that is, for consensus prediction.

Focusing now on the other statistical parameters of the two most predictive non-linear models, i.e., M23 and M26, presented in Table 2 and with their ROC plots shown in Fig. S3, the model M23 has accuracy values of 88.4%, 91.3% and 87.5% against the sub-training, test and validation sets, respectively, whereas model M26 affords > 90% accuracy against the sub-training and test sets but the accuracy of the validation set is restricted to 86.6%. Another noteworthy observation is the significant difference between the sensitivity and specificity for both prediction sets of M23 and M26. However, among these three models, M26 provides much more balanced results from this aspect, with lesser differences between the values of false positives and false negatives. These results indicate that model M26 may be different from the other two models and so, consensus prediction with these models may result in higher overall predictive accuracies

Table 2
Statistical parameters of the two most predictive non-linear models (M23 and M26).[a]

Parameters	M23			M26		
	Sub-training	Test	Validation	Sub-training	Test	Validation
TP	790	184	416	801	180	396
TN	345	109	193	357	113	207
FP	92	22	60	80	18	46
FN	57	6	27	46	10	47
Sn	93.3	83.2	76.3	94.6	86.3	81.8
Sp	79.0	96.8	93.9	81.7	94.7	89.4
Acc	88.4	91.3	87.5	90.2	91.3	86.6
F1-score	91.4	92.9	90.53	92.7	92.8	89.5
MCC	0.738	0.820	0.726	0.779	0.819	0.712
AUROC	–	0.900	0.851	–	0.905	0.856

[a] For the meaning of the statistical parameters, please check the footnotes of Table 1. The results presented for the sub-training sets are from a 5-fold cross validation procedure. The metrics Sn, Sp, Acc, and F1-score are given in percentage.

Table 3
Main in silico results versus mechanistic insight and industrial value for green EDCs manufacture.

Descriptor significance	Associated experimental element/dimension	Result	Mechanistic Insight	Industrial value of descriptors
1	Species group/2D	$\Delta(X2sol)_{spg}$: Based on the connectivity index X2sol	Likely relation between EDCs environmental toxicity and the solvation entropy and dispersion interactions of solved EDCs (as function of EDCs topological structures)	• ↑ $\Delta(X2sol)_{spg}$|↑ environmental toxicity
2	Concentration/2D	$\Delta(P\_VSA\_LogP\_4)_{co}$: Based on descriptor projection of the van der Waals surface area of EDCs	Possible EDC toxicity affected by atomic lipophilicity (or LogP values), in a specific topological range (4 in this instance)	• ↓ $\Delta(P\_VSA\_LogP\_4)_{co}$|↑ environmental toxicity • ↓ $\Delta(P\_VSA\_LogP\_4)_{co}$ or ↑ toxicity|↑ number of non-polar atoms
3	Species name/2D	$\Delta(VE2\_B(m))_{spn}$: Average coefficient of the last eigenvector (absolute values) of atomic masses weighted Burden matrix	It could indicate the importance of specific topological contributions of the molecules towards higher toxicity *per* atomic masses	• ↓ $\Delta(VE2\_B(m))_{spn}$|↑ environmental toxicity • ↓ molecular mass|↑ $\Delta(VE2\_B(m))_{spn}$
4	Species group/2D	$\Delta(nArOH)_{spg}$: Presence/number of aromatic hydroxyls	Likelihood of higher presence of aromatic hydroxyls imply higher EDC toxicity	• ↑ $\Delta(nArOH)_{spg}$|↑ environmental toxicity
5	Species name/2D	$\Delta(nRCHO)_{spn}$: Presence/number of aldehydes (aliphatic)	Higher presence of aldehydes (aliphatic) induces higher EDC toxicity	• ↑ $\Delta(nRCHO)_{spn}$|↑ environmental toxicity
6	Species name/2D	$\Delta(Eig07\_EA(dm))_{spn}$: Relates atomic dipole moments with topology	Influence of physicochemical properties like the atomic dipole moments (dm) in lowering EDC toxicity	• ↓ $\Delta(Eig07\_EA(dm))_{spn}$|↑ environmental toxicity
7	Species group/3D	$\Delta(R5m+)_{spg}$: Probes the intermolecular influence as function of molecular spatial coordinates in a chosen conformation	Likelihood of physicochemical properties such as the atomic masses (m) to positively trigger the EDC toxicity	• ↑ 5.225 $\Delta(R5m+)spn$|↑ environmental toxicity
8	Species name/3D	$\Delta(Mor11v)_{spn}$: Probes the dependence of the 3D geometry of the compounds	Likelihood of physicochemical properties such as the atomic van der Waal volumes (v) to negatively trigger the EDC toxicity	• ↓ $\Delta(Mor11v)spg$|↑ environmental toxicity
9	Species name/2D	$\Delta(T(N\ldots S))_{spn}$: Sum of topological distances between nitrogen (N) and sulfur (S) atoms	Lower values for the sum of topological distance between nitrogen and sulfur atoms in EDCs might imply higher EDC toxicity	• ↓ $\Delta(T(N\ldots S))spn$|↑ environmental toxicity
10	Species name/2D	$\Delta(F06[C\text{-}N])_{spn}$: Assess the importance of the frequency of carbon and nitrogen separated at a given topological distance (7 in this case)	High frequency of carbon and nitrogen separated at topological distance might elicit higher EDC toxicity	• ↑ $\Delta(F06[C\text{-}N])spn$|↑ environmental toxicity

towards the prediction sets. Finally, M23 model has 75 and 166 outliers in the test and validation set whereas M26 contained only 9 outliers in the test set and 40 outliers in the validation sets.

3.3. Mechanistic interpretation of linear model

After confirming statistical similarity between the non-linear models and the linear models and assuring that the former does not increase much their predictive accuracy towards the validation set, focus is once more on the best linear model M3p, because it affords simpler mechanistic interpretability. In fact, an in-depth analysis of the descriptors from this linear MA-mtk QSTR model will allow to disclose the relevant features of the EDCs molecular structure under the targeted experimental conditions that affect their end-point toxicity. The main results of that analysis are displayed in Table 3, *per* in silico result and its subsequent mechanistic insight, with clarification of the industrial value implication in developing more environmental-friendly EDCs.

The M3p model included ten descriptors, whose description is provided in Table S5 and relative contributions are shown in Fig. 1, grouped by one of three experimental elements that were considered to be important in the beginning of the present MA-mtk QSAR modeling. The experimental element *spn*, or species name, appeared in six deviation descriptors, followed by the *spg*, or species group, found in three deviation descriptors. The third experimental element *co*, or concentration, was found only in one descriptor, namely $\Delta(P\_VSA\_LogP\_4)_{co}$, which is one of the most significant descriptors of the model (*vide* Fig. 1). Interestingly, the frequency at which these experimental elements were found are proportional to their variations in the modeling datasets.

Notwithstanding, the most significant descriptor of the model is $\Delta(X2sol)_{spg}$, based on the connectivity index X2sol (Kier and Hall, 2000). Even though this descriptor is topological in nature and computed from H-depleted molecular graphs, it tends to explain the solvation entropy and dispersion interactions of the compounds in solution with respect to the topology of their structures. In this work, compounds with higher value of $\Delta(X2sol)_{spg}$ depicted higher environmental toxicity, as can be inspected for example in Fig. 2 that shows six randomly selected data-points with variations in the $\Delta(X2sol)_{spg}$ values.

The second most significant deviation descriptor of the model is $\Delta(P\_VSA\_LogP\_4)_{co}$, which is based on the descriptor P_VSA_LogP_4 representing the P_VSA-like descriptor on LogP with bin 4 (Labute, 2000; Chavan et al., 2014). In fact, this descriptor tends to project the van der Waals surface area of compounds, estimated with the help of atomic lipophilicity (or LogP values), in a specific topological range (4 in this instance) specified in the connection table (Labute, 2000). Here, a low value of $\Delta(P\_VSA\_LogP\_4)_{co}$ was found in the compounds with high toxicity, as depicted in Fig. 3 for some typical examples. From the latter, it can be observed that compounds with low $\Delta(P\_VSA\_LogP\_4)_{co}$ values or with high toxicity comprise a greater number of non-polar atoms compared to those with low $\Delta(P\_VSA\_LogP\_4)_{co}$ values (or with low toxicity).

The 2D-matrix descriptor VE2_B(m) follows as the third most influential deviation descriptor of the model, though it was found as significant as $\Delta(P\_VSA\_LogP\_4)_{co}$ since both have highly similar absolute standardized coefficients and are negatively correlated to the response variable. VE2_B (m) means the average coefficient of the last eigenvector (absolute values) calculated with the Burden matrix weighted by atomic masses (Todeschini and Consonni, 2009). Being a topological descriptor based on the 2D-adjacency matrix, this descriptor is less prone to mechanistic interpretability though it points out the importance of specific topological contributions of the chemicals towards higher toxicity along with the masses (m) of their atoms. Fig. 4 depicts some toxic data-points (with low $\Delta(VE2\_B(m))_{spn}$ values) along with a few non-toxic data-points (with high $\Delta(VE2\_B(m))_{spn}$ values). It may readily be understood from this figure that molecular mass plays an important role in determining $\Delta(VE2\_B(m))_{spn}$ since compounds with low molecular mass possess higher values for this deviation descriptor.

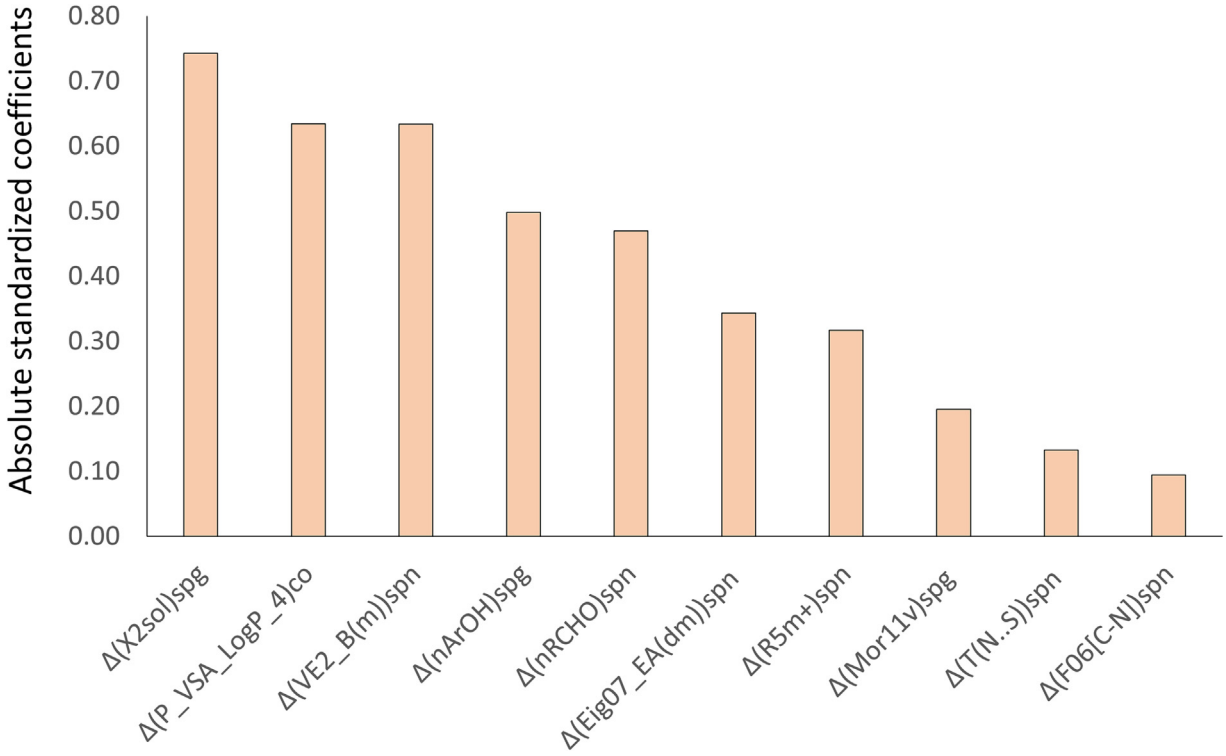

Fig. 1. Relative importance of the descriptors appearing in the linear model M3p.

Unlike the former descriptor, the present model includes some descriptors that are comparatively easier to interpret (Todeschini and Consonni, 2009). For example, $\Delta(nArOH)_{spg}$ and $\Delta(nRCHO)_{spn}$ are two of such descriptors. Being positively correlated with the toxicity, these descriptors highlight the importance of aromatic hydroxyl and aliphatic aldehyde groups for inducing higher environmental toxicity of the EDCs. In Fig. S4, some compounds with high values of $\Delta(nArOH)_{spg}$ or $\Delta(nRCHO)_{spn}$ are shown. Irrespective of the experimental conditions, these compounds consistently exhibit high environmental toxicity.

In addition, the model contains another topological descriptor, namely $\Delta(Eig07\_EA(dm))_{spn}$, that shows that physicochemical properties like the atomic dipole moments (dm) play a negative role in sparking the environmental toxicity of EDCs. The only 3D-based descriptors in this model are $\Delta(Mor11v)_{spn}$ and $\Delta(R5m+)_{spg}$, which also illustrate that physicochemical

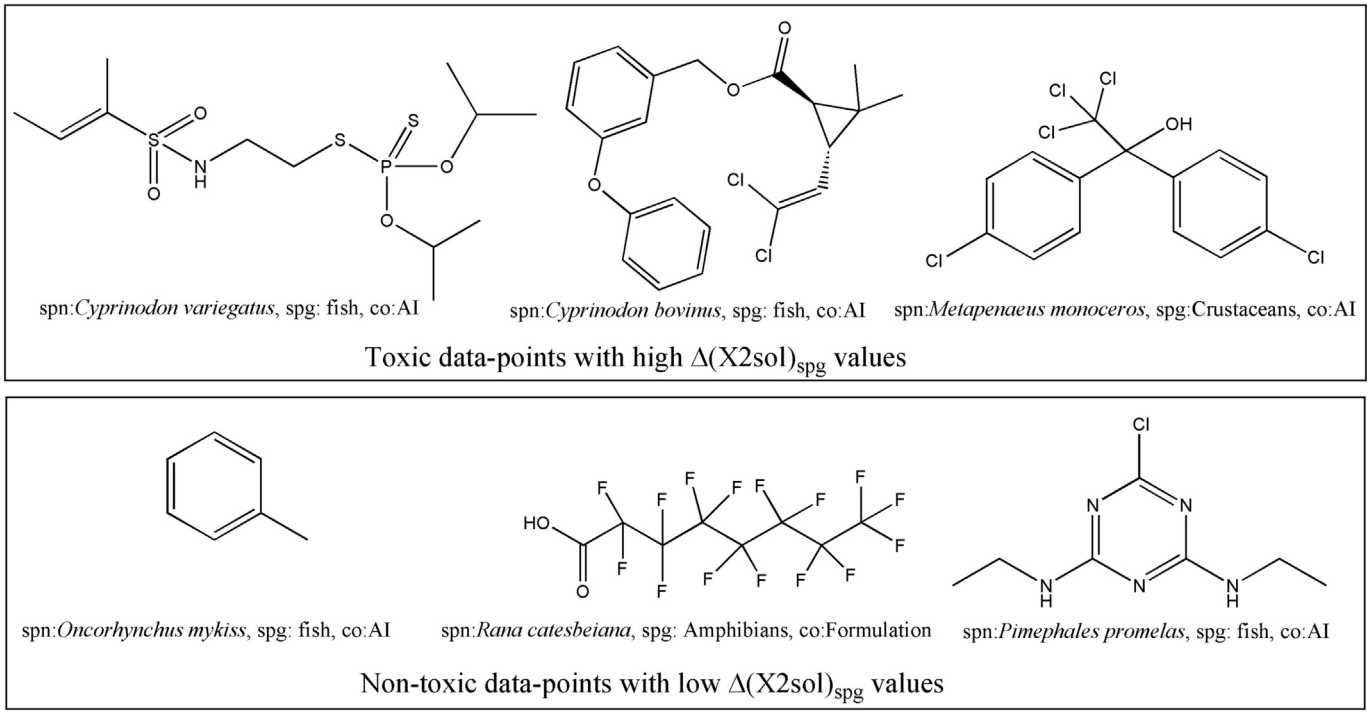

Fig. 2. Typical examples taken from the current dataset with varying values of $\Delta(X2sol)_{spg}$, which is the most significant descriptor of the linear model M3p.

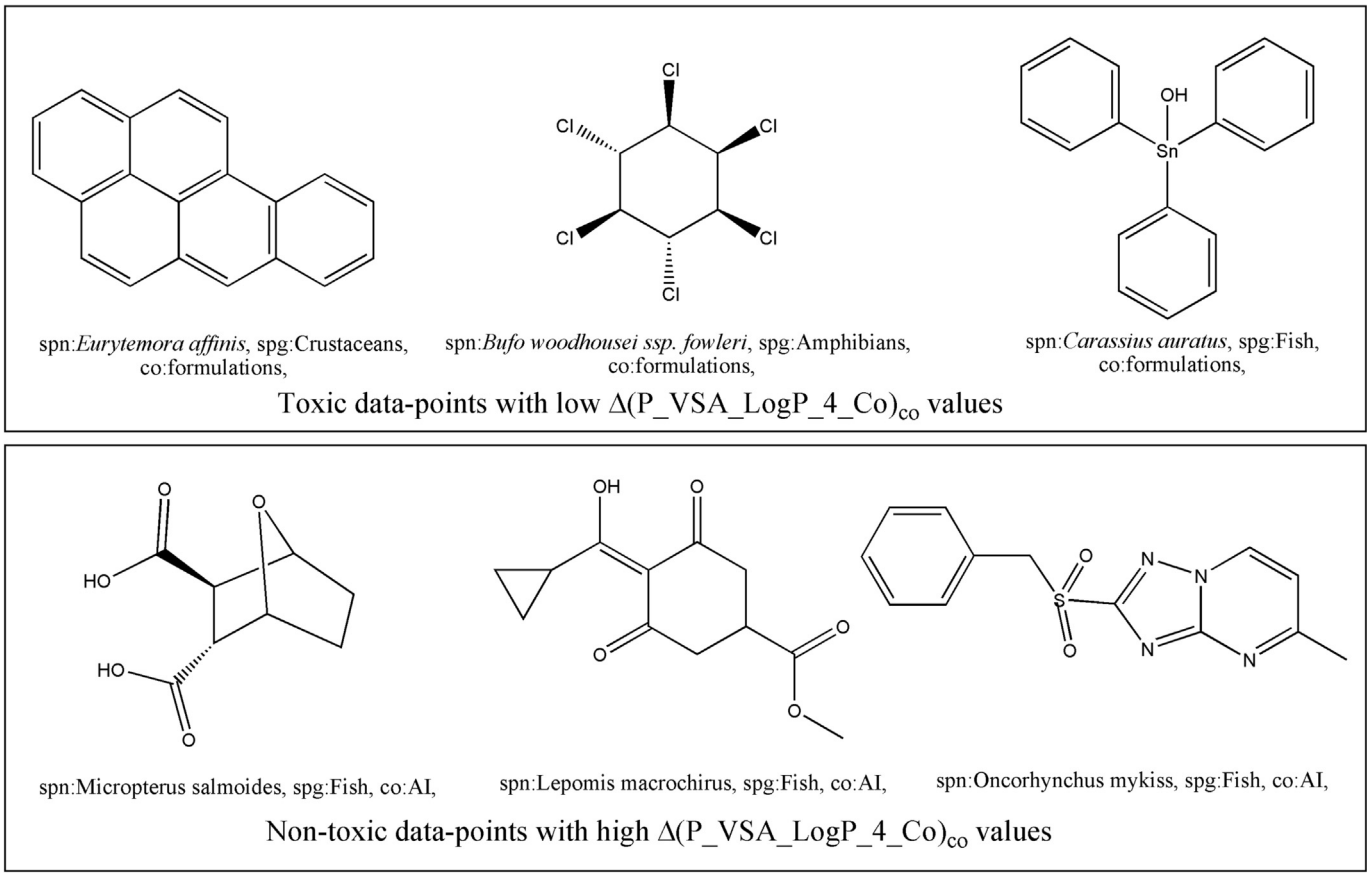

Fig. 3. Typical examples taken from the current dataset with varying values of $\Delta(P\_VSA\_LogP\_4)_{co}$, which is the second most significant descriptor of the linear model M3p.

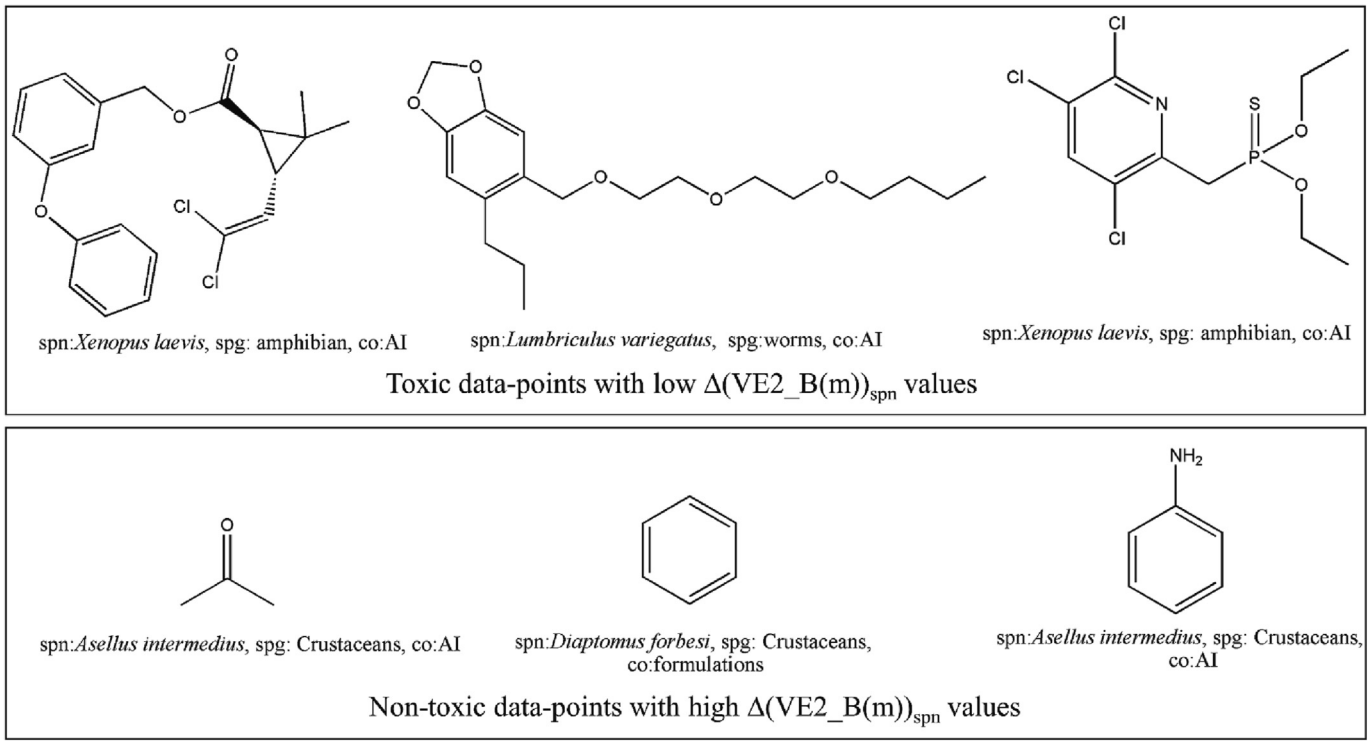

Fig. 4. Typical examples taken from the current dataset with varying values of $\Delta(VE2\_B(m))_{spn}$, which is the third most significant descriptor of the linear model M3p.

properties such as the atomic van der Waal volumes and masses trigger either negatively (v) or positively (m) the environmental toxicity of EDCs. The Mor11v belongs to the 3D-Morse descriptors that are computed from equations based on electron diffraction studies and so, these descriptors are highly dependent on the 3D geometry of the compounds (Schuur et al., 1996; Devinyak et al., 2014). Similarly, R5m+ belongs to the R-type GEometry, Topology, and Atom-Weights AssemblY (GETAWAY) descriptors, which are based on the molecular influence matrix (*MIM*) (Consonni et al., 2002), being the latter employed as molecular representation computed from the spatial coordinates of the molecule atoms in a chosen conformation. The R-GETAWAY descriptors particularly are derived by combining the *MIM* information with the geometric interatomic distances in the molecule, being therefore able of differentiating closely related structures with the help of the 3D-shape of molecules.

The remaining descriptors contributing to this model are the 2D atom pair descriptors T(N…S) and F06[C-N] (Todeschini and Consonni, 2009) The T (N…S) descriptor stands for the sum of topological distances between nitrogen (N) and sulfur (S) atoms. Finally, F06[C-N] is another descriptor that emphasizes the importance of the frequency of carbon and nitrogen separated at a topological distance of 7. A higher value of this parameter was observed in the compounds exhibiting high toxicity to various environmental species.

3.4. Consensus modeling approach

Recently consensus in silico modeling drew considerable attention since one single model may not be able to provide maximum predictivity for any given dataset. When the predictions of multiple models are combined, higher predictivity may be achieved. In fact, Roy et al. recently developed the idea of intelligent consensus prediction for multilinear regression models (Roy et al., 2018). Furthermore, consensus modeling may also help in widening the applicability domain of the models. Also recently, Valsecchi et al. proposed different consensus modeling approaches, such as majority voting and the Bayes consensus with discrete probability distributions, to improve the predictive accuracy of classification models (Valsecchi et al., 2020). Even if the consensus prediction does not assist much in improving the predictivity to a considerable extent regarding the individual models, such predictions may at least justify that maximum predictivity has been achieved by those individual models, which in turn may improve the reliability towards proposing the best individual models, as well as the methodology adopted to set up the latter.

In the present work, the simplest of the rationales was employed for developing consensus models from the most predictive linear and non-linear models. Basically, consensus modeling was applied for the following three different groups:

(i) Linear models - The top five linear models were chosen based on having an average MCC score greater than 0.70, i.e.: M3p, M2p, M7p, M2 and M6. The latter yielded a total of 26 combinations that were subjected to consensus prediction.
(ii) Non-linear models - The top four non-linear models were chosen, namely: M20, M23, M26 and M29, for testing a total of 11 combinations.
(iii) Linear and non-linear models - The top two linear and non-linear models were selected, i.e.: M3p, M2p, M20 and M26, resulting in a total of 11 combinations.

The best results obtained for each group are summarized in Table S8. As can be seen, the most promising model was achieved when the linear model M3p was combined with the non-linear models M23 and M26. The variations observed in the MCC values for the two prediction sets of the third group consensus models are shown in Fig. S5. The predictive accuracy of the best consensus model found (i.e.: CM3) was slightly improved when compared to most individual models, especially for the external validation set. In fact, model CM3 afforded an accuracy of 88.4% with a MCC of 0.745 for the validation set. The other statistical parameters for this model regarding the test set are given in Table 4. This model highlights the significance of the most predictive linear model found, i.e.: M3p. Similarly, M23 and M26 pertain also to the most predictive non-linear models developed in this work. What is more, since the consensus model fails to improve the external predictivity to a large extent, the individual linear and non-linear models proposed in this work reached almost a maximum predictivity with the current set of descriptors. For example, compared to M3p, CM3 provides less than 2% improvement in the predictive accuracy towards the external validation set. Therefore, it can be concluded that the linear model M3p can be used for predicting the ecotoxicity for novel structures. However, for slightly higher predictive accuracy, one might resource to the consensus model CM3.

Table 4
Statistical parameters of the models obtained from the most predictive consensus model CM3.[a]

Parameters	Test set	Validation set
TP	184	418
TN	111	197
FP	20	56
FN	6	25
Sn	84.7	77.9
Sp	96.8	94.4
Acc	91.9	88.4
F1-score	93.4	91.2
MCC	0.833	0.745
AUROC score	0.908	0.861

[a] For the meaning of the statistical parameters, please check the footnotes of Table 1. The metrics Sn, Sp, Acc, and F1-score are given in percentage.

4. Conclusions

Exploring MA-mtk in silico modeling techniques to develop global models for ecotoxicity prediction of EDC towards a large number of biological species of various species groups was the main objective of this work. Recent years have witnessed the emergence of moving average-based multitasking modeling as a useful strategy for developing mtk-QSTR models due to its simplicity and ability to jointly predict outcomes against multiple experimental conditions. As many as 2,301 data-points with high structural and experimental diversity were used to set up linear and non-linear models, using various advanced machine learning tools. Overall, the most predictive linear mtk-QSTR model afforded an overall accuracy of 87%, with the recently proposed PS3M feature selection technique playing a crucial role in finding this ten-descriptor model from a large number of deviation descriptors. The obtained results thus emphasize the significance of the linear model both as a predictive QSTR model as well as a model that highlights crucial structural attributes responsible for higher toxicity. Consensus models were also generated and helped improve the predictive accuracy towards external validation sets. Presence of number of fragments such as aromatic hydroxyl and aliphatic aldehyde, solvation, van der Waals volumes, atomic dipole moments plus masses, along with the distance between nitrogen-sulfur and carbon-nitrogen were found to be determining factors for eliciting higher ecotoxicity by the EDCs towards different biological species. Additionally, it was revealed that the specific 2D- and 3D-geometry of these chemical compounds are also responsible for their toxicological nature.

Finally, all models were developed and validated using non-commercial open-access tools such as the second version of the QSAR-Co-X tool (available at https://github.com/ncordeirfcup/QSAR-Co-X_v2), which now includes the newly introduced consensus prediction method. The latter are thus easily reproducible by the end-users and may be used for library screening to speed up regulatory decision for gathering safe alternatives that will reduce the hazards of EDCs.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2023.164337.

CRediT authorship contribution statement

Amit Kumar Halder: Conceptualization, Methodology, Formal analysis, Software, Investigation, Visualization, Writing - Original draft

preparation. **Ana S Moura:** Investigation, Writing - Reviewing & Editing. **Maria Natalia D. S. Cordeiro:** Investigation, Writing - Reviewing & Editing, Supervision, Funding acquisition. All authors discussed the results and commented on the manuscript. All authors read and approved the final manuscript.

Data availability

Data are available as supplementary materials or in Github links provided in the manuscript.

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

This work was supported by UIDB/50006/2020 with funding from FCT/MCTES through national funds. The authors are thankful to ChemAxon for providing the academic license of Standardizer. Ana S. Moura further acknowledges FCT/MECS for the contract IF CEECIND/03631/2017.

References

Ambure, P., Bhat, J., Puzyn, T., Roy, K., 2018. Identifying natural compounds as multi-target-directed ligands against Alzheimer's disease: an in silico approach. J. Biomol. Struct. Dyn. 37, 1282–1306. https://doi.org/10.1080/07391102.2018.1456975.

Ambure, P., Halder, A.K., Gonzalez-Diaz, H., Cordeiro, M.N.D.S., 2019. QSAR-Co: an open source software for developing robust multitasking or multitarget classification-based QSAR models. J. Chem. Inf. Model. 59, 2538–2544. https://doi.org/10.1021/acs.jcim.9b00295.

BIOVIA, 2021. Dassault Systèmes, Discovery Studio Visualizer 2021 Client. Dassault Systèmes, San Diego.

Boser, B.E., et al., 1992. A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory ACM 144–152, Pittsburgh, PA, USA, 27–29 July.

Boughorbel, S., Jarray, F., El-Anbari, M., 2017. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS One 12, e0177678. https://doi.org/10.1371/journal.pone.0177678.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M., 2015. Time series analysis: forecasting and control. Wiley Series in Probability and Statistics, 5th ed. ISBN: 978-1-118-67491-8.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Chavan, S., Nicholls, I.A., Karlsson, C.G., Rosengren, A.M., Ballabio, D., Consonni, V., 2014. Towards global QSAR model building for acute toxicity: Munro Database case study. Int. J. Mol. Sci. 15, 18162–18174. https://doi.org/10.3390/ijms151018162.

Consonni, V., Todeschini, R., Pavan, M., 2002. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. J. Chem. Inf. Comput. Sci. 42, 682–692. https://doi.org/10.1021/ci015504a.

Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. IEEE Trans. Inf. Theory 13, 21–27. https://doi.org/10.1109/tit.1967.1053964.

Cumming, J.G., Davis, A.M., Muresan, S., Haeberlein, M., Chen, H., 2013. Chemical predictive modelling to improve compound quality. Nat. Rev. Drug Discov. 12, 948–962. https://doi.org/10.1038/nrd4128.

De Coster, S., van Larebeke, N., 2012. Endocrine-disrupting chemicals: associated disorders and mechanisms of action. J. Environ. Public Health 2012, 1–52. https://doi.org/10.1155/2012/713696.

Devinyak, O., Havrylyuk, D., Lesykb, R., 2014. 3D-MoRSE descriptors explained. J. Mol. Graphics Model. 54, 194–203. https://doi.org/10.1016/j.jmgm.2014.10.006.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29. https://doi.org/10.1214/aos/1013203451.

Globally Harmonized System of Classification and Labelling of Chemicals (GHS) 9th revised edition. United Nations, New York and Geneva. https://unece.org/sites/default/files/2021-09/GHS_Rev9E_0.pdf.

Gore Jr., P.A., 2000. Cluster analysis. In: Tinsley, H.E.A., Brown, S.D. (Eds.), Handbook of Applied Multivariate Statistics and Mathematical Modeling. Academic Press, pp. 297–321 https://doi.org/10.1016/B978-012691360-6/50012-4.

Guang-Bin, H., Babri, H.A., 1998. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. IEEE Trans. Neural Net. 9, 224–229. https://doi.org/10.1109/72.655045.

Gunturi, S.B., Theerthala, S.S., Patel, N.K., Bahl, J., Narayanan, R., 2010. Prediction of skin sensitization potential using D-optimal design and GA-kNN classification methods. SAR QSAR Environ. Res. 21, 305–335. https://doi.org/10.1080/10629361003773955.

Halder, A.K., Cordeiro, M.N.D.S., 2019. Development of multi-target chemometric models for the inhibition of class i pi3k enzyme isoforms: A case study using QSAR-Co tool. Int. J. Mol. Sci. 20, 4191. https://doi.org/10.3390/ijms20174191.

Halder, A.K., Cordeiro, M.N.D.S., 2021a. QSAR-Co-X: an open source toolkit for multitarget QSAR modelling. J. Cheminform. 13, 29. https://doi.org/10.1186/s13321-021-00508-0.

Halder, A.K., Cordeiro, M.N.D.S., 2021b. Multi-target in silico prediction of inhibitors for mitogen-activated protein kinase-interacting kinases. Biomolecules 11, 1670. https://doi.org/10.3390/biom11111670.

Halder, A.K., Moura, A.S., Cordeiro, M.N.D.S., 2022a. Moving average-based multitasking in silico classification modeling: where do we stand and what is next? Int. J. Mol. Sci. 23, 4937. https://doi.org/10.3390/ijms23094937.

Halder, A.K., Delgado, A.H.S., Cordeiro, M.N.D.S., 2022b. First multi-target QSAR model for predicting the cytotoxicity of acrylic acid-based dental monomers. Dent. Mat. 38, 333–346. https://doi.org/10.1016/j.dental.2021.12.014.

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., Doughterty, E.R., 2010. Small-sample precision of ROC-related estimates. Bioinformatics. 26, 822–830. https://doi.org/10.1093/bioinformatics/btq037.

He, J., Peng, T., Yang, X., Liua, H., 2018. Development of QSAR models for predicting the binding affinity of endocrine disrupting chemicals to eight fish estrogen receptor. Ecotoxicol. Environ. Saf. 148, 211–219. https://doi.org/10.1016/j.ecoenv.2017.10.023.

Heo, S., Safder, U., Yoo, C., 2019. Deep learning driven QSAR model for environmental toxicology: Effects of endocrine disrupting chemicals on human health. Environ. Pollut. 253, 29–38. https://doi.org/10.1016/j.envpol.2019.06.081.

Khan, K., Roy, K., Benfeneti, E., 2019. Ecotoxicological QSAR modeling of endocrine disruptor chemicals. J. Hazard. Mater. 369, 707–718. https://doi.org/10.1016/j.jhazmat.2019.02.019.

Kier, L.B., Hall, L.H., 2000. Intermolecular accessibility: The meaning of molecular connectivity. J. Chem. Inf. Comput. Sci. 40, 792–795. https://doi.org/10.1021/ci990135s.

Kleandrova, V.V., Scotti, L., Bezerra Mendonça Junior, F.J., Muratov, E., Scotti, M.T., Speck-Planche, A., 2021. QSAR modeling for multi-target drug discovery: Designing simultaneous inhibitors of proteins in diverse pathogenic parasites. Front. Chem. 9. https://doi.org/10.3389/fchem.2021.634663.

Kumar, M., Sarma, D.K., Shubham, S., Kumawat, M., Verma, V., Prakash, A., et al., 2020. Environmental endocrine-disrupting chemical exposure: role in non-communicable diseases. Front. Public Health 8, 553850. https://doi.org/10.3389/fpubh.2020.553850.

Labute, P., 2000. A widely applicable set of descriptors. J. Mol. Graphics Model. 18, 464–477. https://doi.org/10.1016/s1093-3263(00)00068-1.

Lauretta, R., Sansone, A., Sansone, M., Romanelli, F., Appetecchia, M., 2019. Endocrine disrupting chemicals: effects on endocrine glands. Front. Endocrinol. 10, 178. https://doi.org/10.3389/fendo.2019.00178.

Marty, S., 2014. Introduction to "Screening for endocrine activity-experiences with the US EPA's endocrine disruptor screening program and future considerations". Birth Defects Res. B Dev. Reprod. Toxicol. 101, 1–2. https://doi.org/10.1002/bdrb.21100.

Mauri, A., Consonni, V., Pavan, M., Todeschini, T., 2006. Dragon software: An easy approach to molecular descriptor calculations. Match-Commun. Math. Co. 56, 237–248.

Muratov, E.N., Bajorath, J., Sheridan, R.P., Tetko, I.V., Filimonov, D., Poroikov, V., et al., 2020. QSAR without borders. Chem. Soc. Rev. 49, 3525–3564. https://doi.org/10.1039/d0cs00098a.

Roy, K., Kar, S., Ambure, P., 2015. On a simple approach for determining applicability domain of QSAR models. Chemom. Intell. Lab. Syst. 145, 22–29. https://doi.org/10.1016/j.chemolab.2015.04.013.

Roy, K., Ambure, P., Kar, S., Ojha, P.K., 2018. Is it possible to improve the quality of predictions from an "intelligent" use of multiple QSAR/QSPR/QSTR models? J. Chemom. 32, e2992. https://doi.org/10.1002/cem.2992.

Sadowski, J., Gasteiger, J., Klebe, G., 2002. Comparison of automatic three-dimensional model builders using 639 X-ray structures. J. Chem. Inf. Comput. Sci. 34, 1000–1008. https://doi.org/10.1021/ci00020a039.

Schug, T.T., Janesick, A., Blumberg, B., Heindel, J.J., 2011. Endocrine disrupting chemicals and disease susceptibility. J. Steroid Biochem. Mol. Biol. 127, 204–215. https://doi.org/10.1016/j.jsbmb.2011.08.007.

Schuur, J., Selzer, P., Gasteiger, J., 1996. The coding of the three-dimensional structure of molecules by molecular transforms and its application to structure-spectra correlations and studies of biological activity. J. Chem. Inf. Comput. Sci. 36, 334–344. https://doi.org/10.1021/ci950164c.

Sheffield, T.Y., Judson, R.S., 2019. Ensemble QSAR modeling to predict multispecies fish toxicity lethal concentrations and points of departure. Environ. Sci. Technol. 53, 12793–12802. https://doi.org/10.1021/acs.est.9b03957.

Sosnin, S., Karlov, D., Tetko, I.V., Fedorov, M.V., 2018. Comparative study of multitask toxicity modeling on a broad chemical space. J. Chem. Inf. Model. 59, 1062–1072. https://doi.org/10.1021/acs.jcim.8b00685.

Speck-Planche, A., 2020. Multi-scale QSAR approach for simultaneous modeling of ecotoxic effects of pesticides. In: Roy, K. (Ed.), Ecotoxicological QSARs, Methods in Pharmacology and Toxicology. Humana, New York, NY, pp. 639–660.

Speck-Planche, A., Cordeiro, M.N.D.S., 2015. Multitasking models for quantitative structure–biological effect relationships: current status and future perspectives to speed up drug discovery. Expert Opin. Drug Discov. 10, 245–256. https://doi.org/10.1517/17460441.2015.1006195.

Speck-Planche, A., Cordeiro, M.N.D.S., 2017. Advanced In silico approaches for drug discovery: Mining information from multiple biological and chemical data through mtk-QSBER and pt-QSPR strategies. Curr. Med. 24, 1–18. https://doi.org/10.2174/0929867324666170124152746.

Standardizer, 2010. Version 15.9.14.0 Software. ChemAxon, Budapest, Hungary.

Sushko, I., Novotarskyi, S., Korner, R., Pandey, A.K., Rupp, M., Teetz, W., et al., 2021. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. J. Comput. Aided Mol. Des. 25, 533–554. https://doi.org/10.1007/s10822-011-9440-2.

Tinsley, H.E.A., Brown, S.D., 2000. Handbook of Applied Multivariate Statistics and Mathematical Modeling. Academic Press, San Diego, pp. 209–235.
Todeschini, R., Consonni, V., 2009. Molecular Descriptors for Chemoinformatics. 2nd ed. Wiley-VCH, Weinheim, Germany.
Toropova, A.P., Toropov, A.A., Benfenati, E., 2015. CORAL: Prediction of binding affinity and efficacy of thyroid hormone receptor ligands. Eur. J. Med. Chem. 101, 452–461. https://doi.org/10.1016/j.ejmech.2015.07.012.
Urias, R.W., Barigye, S.J., Marrero-Ponce, Y., Garcia-Jacas, C.R., Valdes-Martini, J.R., Perez-Gimenez, F., 2015. IMMAN: Free software for information theory-based chemometric analysis. Mol. Divers. 19, 305–319. https://doi.org/10.1007/s11030-014-9565-z.
Vafeiadi, M., Georgiou, V., Chalkiadaki, G., Rantakokko, P., Kiviranta, H., Karachaliou, M., et al., 2015. Association of prenatal exposure to persistent organic pollutants with obesity and cardiometabolic traits in early childhood: The Rhea Mother–Child Cohort (Crete, Greece). Environ. Health Perspect. 123, 1015–1021. https://doi.org/10.1289/ehp.1409062.
Valsecchi, C., Grisoni, F., Consonni, V., Ballabio, D., 2020. Consensus versus individual QSARs in classification: Comparison on a large-scale case study. J. Chem. Inf. Model. 60, 1215–1223. https://doi.org/10.1021/acs.jcim.9b01057.
Velmurugan, G., Ramprasath, T., Gilles, M., Swaminathan, K., Ramasamy, S., 2017. Endocrine-disrupting chemicals, and the diabetes epidemic. Trends Endocrinol. Metab. 28, 612–625. https://doi.org/10.1016/j.tem.2017.05.001.
Weininger, D., 1988. SMILES, a chemical language and information system: 1: Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36. https://doi.org/10.1021/ci00057a005.
Wilks, S.S., 1932. Certain generalizations in the analysis of variance. Biometrika. 24, 471–494. https://doi.org/10.1093/biomet/24.3-4.471.