ACS Sustainable Chemistry & Engineering

Research Article

Cite This: ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

pubs.acs.org/journal/ascecg

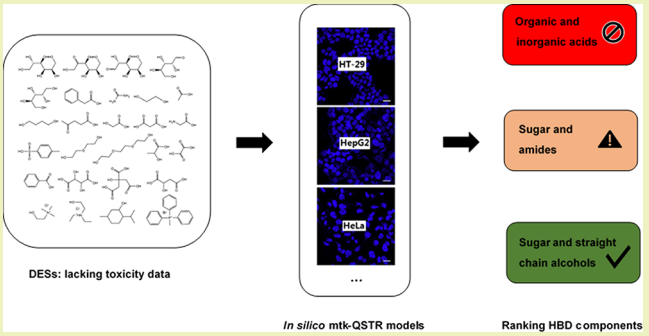
Probing the Environmental Toxicity of Deep Eutectic Solvents and Their Components: An In Silico Modeling Approach

Amit Kumar Halder and M. Natália D. S. Cordeiro*

LAQV@REQUIMTE/Department of Chemistry and Biochemistry, University of Porto, Porto 4169-007, Portugal

Supporting Information

ABSTRACT: Because of the increasing demand of greener solvents, deep eutectic solvents (DES) have just emerged as low-cost alternative solvents for a broad range of applications. However, recent toxicity assay studies showed a non-negligible toxic behavior for these solvents and their components. Alternative in silico-based approaches such as the one proposed here, multitasking-Quantitative Structure Toxicity Relationships (mtk-QSTR), are increasingly used for risk assessment of chemicals to speed up policy decisions. This work reports a mtk-QSTR modeling of 572 DES and their components under multiple experimental conditions. To set up a reliable model from such data, we examined here the use of 0D−2D descriptors along with classification analysis, and the Box−Jenkins approach. This procedure led to a final mtk-QSTR model with high overall accuracy and predictivity (ca. 90%). The model highlights also the crucial role that polarizability, electronegativity, hydrogen-bond donor (HBD), and topological properties play into the DES toxicity. Furthermore, with the help of the derived mtk-QSTR model, 30 different HBD components were ranked on the basis of their toxic contributions to DES. More importantly, the proposed in silico modeling approach is shown to be a valuable tool to mine relevant STR information, therefore guiding the rational design of potentially safe DES.



KEYWORDS: *Green solvents, Deep eutectic solvents, Toxicity, mtk-QSTR, Box−Jenkins approach, Linear discriminant analysis*

INTRODUCTION

Green chemistry involves applications of more environmentally friendly (or less hazardous) chemicals and their synthetic routes. It encourages chemists for minimum usage of chemicals, which are toxic to the environment, or for choosing alternate (and more environmentally friendly) chemicals, media, reaction conditions, and energy resources.[1−3] Among the 12 principles of green chemistry, one is the use of environmentally benign solvents and other substances. Therefore, the use of environmentally friendly solvents has gained much attention in recent years.[4] Ionic liquids (ILs) are considered as one of the most important class of green solvents and have been applied in various chemical processes and technologies.[5,6] However, most of the ILs are difficult to prepare and expensive. In addition, several of these were found to be highly toxic toward the environment due to their harmful effects on various biological species.[5−8] Because of this, replacing ionic liquids with more environmentally friendly solvents has become a necessity. One of the most valuable replacing options of IL are now considered to be deep eutectic (DES) solvents. Indeed, different studies have shown that DES are extremely useful solvents with many characteristics similar to IL.[9] In addition, they are comparatively cheaper, easy to prepare, and more environmentally friendly than IL.[4,10,11] DES are prepared by mixing multiple low-cost components (with hydrogen-bond formation characteristics) to produce a eutectic mixture, the melting point of which is much lower than those of either of their individual components.[9,10,12,13] In most of the cases, DES are prepared by mixing a salt with a hydrogen-bond donor (HBD) molecule in different molar fractions.[9] An example is the choline chloride (ChCl)/urea mixture (in a 1:2 molar ratio) that has a melting temperature of 12 °C. Therefore, this DES is a liquid at room temperature, and its melting point is far below the melting points of its components (the melting points of ChCl and urea are 302 and 133 °C, respectively).[4,9] DES were first reported in 2001 and were immediately used in several applications,[14,15] such as in materials or biomolecular syntheses,[16−19] extraction, separation, electrochemistry,[9,20,21] nanotechnology,[22] biotransformation,[23] lubrication,[24] drug discovery,[25−27] and so forth. DES may also be obtained from natural sources, such as organic acids, amino acids, urea, sugars, etc., and these are then called natural deep eutectic solvents (NADES).[28]

Recent studies have shown however that potential toxicity and cytotoxicity may be imparted by various DES.[4,9] These studies indeed refute the idea that DES are completely nontoxic in nature.[29,30] What is more, the biological effects of any DES are often found to be significantly different from

Received: March 5, 2019
Revised: May 13, 2019
Published: May 24, 2019

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

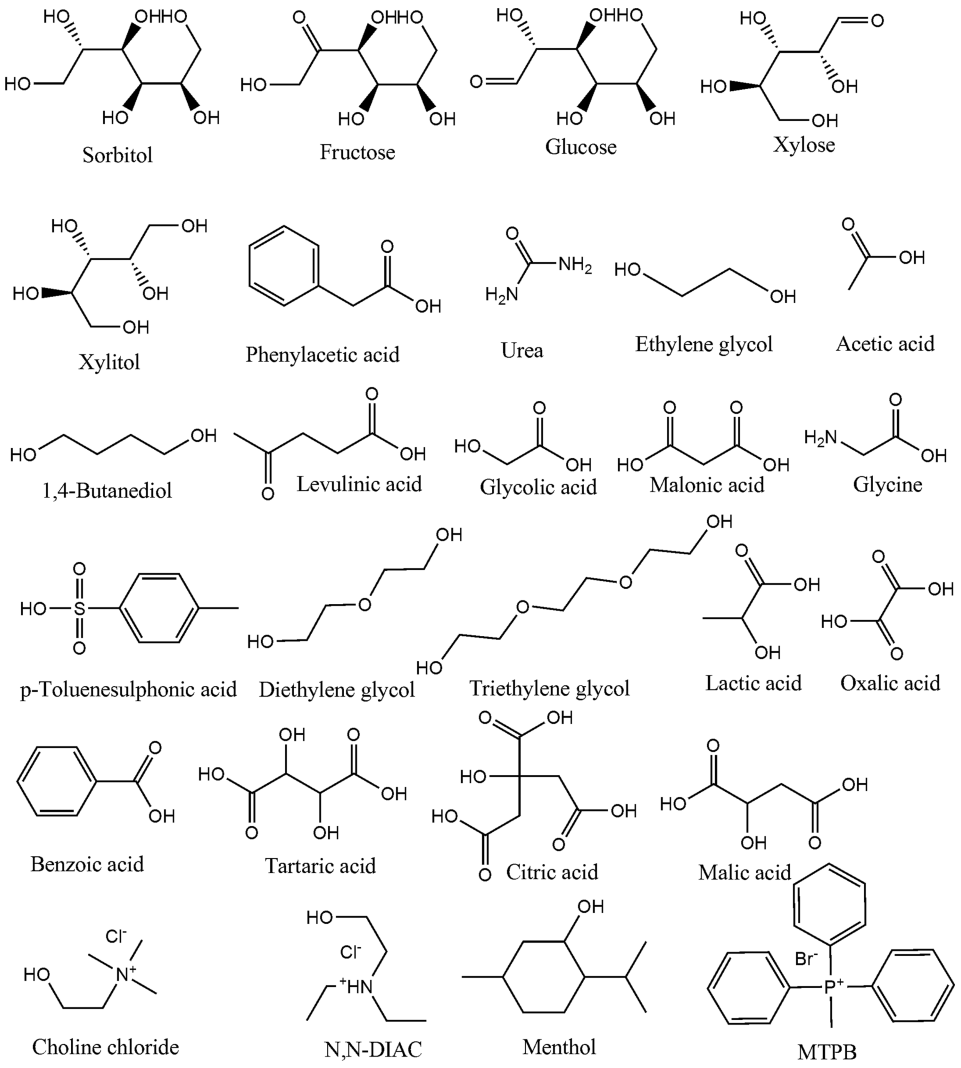


Figure 1. Chemical structures of several important components included in the modeling data set.

those of its individual components. Therefore, even though the components of a DES are found to be nontoxic, the final mixture often exhibits potential toxicity to different environmental biological species due to synergistic effects.[7] For instance, Hayyan et al.[31−34] carried out a series of toxicity and cytotoxicity studies of ChCl-based DES. They observed that the cytotoxicity of these DES is higher than their individual components and that their overall toxicity varies depending on the structures of the latter. Similarly, other research studies suggested that toxic effects may be elicited by DES toward many biological species.[7,35−42] Unfortunately, so far only a few experimental investigations have been reported and focused on the environmental toxicity of these solvents. Nevertheless, the REACH (Registration, Evaluation and Authorization of Chemicals) legislation from the European Union and other initiatives worldwide put emphasis on the risk assessment of all new and existing chemicals.[43] Considering the increasing number of chemical and biological applications of DES, including, for example, pharmaceutical and drug discovery endeavors, problems with their environmental hazard must be addressed more efficiently.

In silico modeling approaches gained enormous attention in the last few decades due to their ability to rationalize diverse types of experimental outcomes. Reliable predictive in silico models allow one to promptly identify potential hazardous chemical, saving costs, as well as diminishing experimental efforts.[44,45] Among different in silico techniques, quantitative structure−toxicity relationship (QSTR) approaches provide important insights into the toxic behavior of a wide range of chemicals.[44,46−48] Indeed, the REACH regulation has long been encouraging the application of QSTR modeling to determine the potential toxic effects of different chemical entities.[43] According to the Organization for Economic Co-operation and Development (OECD) guidelines,[49] five principles must be obeyed for a successful QSTR modeling. These are (a) a definite data set with well-defined end points, (b) an explainable model development strategy, (c) a well-defined domain of applicability, (d) an appropriate validation strategy, and (e) a proper mechanistic interpretation (if possible). Previously, a number of QSTR modeling studies have targeted the toxicity of IL against various biological species.[6,44,50−58] To the best of our knowledge, only one study has been reported so far on the QSTR modeling of the toxic effects of DES. In that study, Ahmadi et al. developed a conventional QSTR model aimed at relating the cytotoxicity data and the structural characteristics of a set of 28 ChCl-based DES.[59]

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

The available toxicity data of DES are considerably less as compared to those of IL, and whereas IL are single molecules, DES are mixtures of multiple components. Therefore, modeling the latter is much more complicated and challenging. In the past few years, multitasking quantitative structure−toxicity relationship (mtk-QSTR) modeling approaches have been successfully implemented by different researchers to establish a number of validated predictive in silico models for various target objectives.[60−67] Moreover, whereas conventional QSTR approaches rely on a specific toxicity measure, that is, pertaining to a single biological species and specific assay conditions, mtk-QSTR allows one to include multiple assay conditions, biological targets, and many other parameters while setting up the model.[62] The mtk-QSTR approach is therefore remarkably useful when the lack of available data hampers the development of statistically reliable in silico models.[66,68] Here, it should be noticed that the mtk-QSTR modeling approach is closely related to the Perturbation-Theory Machine Learning (PTML) modeling one, which also resorts to the Box−Jenkins operators to quantify perturbations related to small variations of the structure or external factors or conditions of the systems under study.[69,70]

The present work aims at predicting with the mtk-QSTR strategy the toxic and antimicrobial properties of several DES against various biological species. This strategy afforded a final mtk-QSTR model with particularly good overall accuracy and predictivity on an external validation set. Emphasis is also given to rank the different contributions of the components on inducing the overall toxicity of the final DES.

MATERIALS AND METHODS

Data Set and Molecular Descriptors. The toxicity data of 498 DES and their components were collected from the literature.[7,31−42] Details of this modeling data set (samples 1−498) are provided in the Supporting Information, and the structures of some of its DES components are shown in Figure 1. The decision for including individual components along with DES was mainly guided by the fact that the toxic behavior of DES varies considerably with their components.[32,33,41] Furthermore, in so doing, more useful information could be extracted by the statistical analysis due to the increased chemical and biological space. These samples were experimentally assayed by using at least 1 out of 5 different measures of toxic effects (m_e) against at least 1 out of 27 types of biological targets (b_t). The latter comprised 11 different mammalian cell lines (A375, AGS, B16F10, CaOV3, CCO, HELA, HepG2, HT29, MCF7, OKF6, and PC3) and 12 different microbial organisms (*Aspergillus niger*, *Bacillus subtilis*, *Candida cylindracea*, *Escherichia coli*, *Listeria monocytogenes*, *Lentinus tigrinus*, *Pseudomonas aeruginosa*, *Phanerochaete chrysosporium*, *Staphylococcus aureus*, *Salmonella enteritidis*, *Escherichia coli*, and *Vibrio fischeri*).

As a final essential element, the presence of water (w_t) was considered. In fact, as it has been observed in previous studies,[33,71] water may considerably alter the biological functions of DES. Some of the DES of our data were assayed with water; that is, these were either solvated in water, or water was used as a component of the mixture. Therefore, depending on the presence of water, two types of conditions (i.e., water/dry) were used for the last element w_t.

Here, it should be mentioned that the combination of these elements defines a specific experimental condition that can be expressed as an ontology of the form $c_j \rightarrow (m_e, b_t, w_t)$. It should be noticed also that the data set samples were not assayed against all of the possible experimental conditions c_j. Similarly, some of the data set samples were tested against more than one experimental condition. Further, each sample was assigned to one out of two possible categories, that is, toxic $[Tox_i(c_j) = +1]$ and nontoxic $[Tox_i(c_j) = -1]$, based on specific cutoff values chosen for the different toxicological measurements (see Table 1). One may argue that a 30 mM cutoff value for the inhibitory concentration (IC_{50}) is too high. Yet this and the other cutoff values were chosen for the following reasons. First, an effective or inhibitory concentration higher than 1000 mg/Lit (approximately 20 mM considering the lowest molecular weight of the data set chemicals) is generally considered as "relatively nontoxic".[72] We have thus adjusted this value to 30 mM to maintain an overall consistency for all inhibitory or effective concentrations (i.e., EC_{50}, IC_{50}, MIC) employed. Furthermore, the increase of threshold may also limit the inclusion of toxic agents as nontoxic in the data set. Second, the selected cutoffs prevented an excessive imbalance between the number of samples assigned as toxic and as nontoxic. Indeed, a well-balanced ratio of toxic/nontoxic cases (=231/222) was found by applying these cutoff values.

Table 1. Cutoff Values Employed for the Diverse Measurements of Toxicological Effects

type of measurement	meaning	cutoff values[a]
IC_{50}	50% inhibitory concentration against human cell lines	>30 mM
EC_{50}_c	50% effective concentration against human cell lines	>10 mM[b]
EC_{50}_m	50% effective concentration against microbial cells	>30 mM
MIC	minimum inhibitory concentration against microbial cells	>30 mM

[a]Values adopted for assigning cases as nontoxic. [b]Undefined responses (>10 nM or >5 nM) were taken as nontoxic.

Here, it should be remarked that, in the current modeling, two data samples (mixtures/single components) should be considered as duplicates when these have the same composition (i.e., ratio of the constituent(s)) as well as the same experimental elements (i.e., m_e, b_t, and w_t). Yet, during our data curation, we did not find any duplicate samples, and no sample was therefore removed.

Typically, DES are a mixture of hydrogen-bond donors and hydrogen-bond acceptors (e.g., salts) and other additional components (e.g., water), and these are mixed in different molar fractions.[9,73] Therefore, the calculation of molecular descriptors for DES is not straightforward and requires a special treatment. Previously, two strategies have been followed for setting up the descriptors of DES: (i) All of the DES systems were first optimized using density functional theory (DFT) calculations, and then their descriptors were obtained from a topological analysis on the electronic density under the formalism of Bader's Quantum Theory of Atoms in Molecules (QTAIM[74]).[75] (ii) Only descriptors for one of the DES components (e.g., the HBD ones) were computed from their optimized structures, because the influence of the other component (choline chloride in all cases) could be disregarded.[59,73] As can be judged, the first strategy is very time-consuming and complex especially for large data sets, while the second one is not applicable for the present data set. In this work, we therefore adopted a different strategy for computing the DES descriptors, similar to others previously suggested for mixtures.[76−80] That is, because the properties of DES change with the concentrations of the components, their descriptors (D_{mix}) were computed as weighted sums (by molar fractions X_i) of the individual components' descriptors (D_i), as shown below:

$$D_{mix} = \sum_i X_i D_i \quad (1)$$

The major challenge here is to obtain descriptors that would allow one to sufficiently discriminate between two closely related DES, such as those comprising the same components but in different molar fractions. The strategy thus followed for calculating the DES descriptors appears to be the most appropriate, straightforward, and least time-consuming. Moreover, because single descriptor values are to be obtained for the single components, that allowed us to

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

additionally include toxicity data of all of the DES components into our data set, making more reliable the following statistical analysis.

Simple descriptors (0D−2D) were considered to codify the molecular structure of each DE component. 0D-descriptors refer to, for example, atom and bond type counts, whereas 1D-descriptors comprise, for example, functional groups and fragments counts, and both can be computed without any information about molecular structures. Regarding 2D-descriptors, these are topological descriptors derived from molecular graphs, and, although conformationally independent, they may be weighted by specific atomic chemical properties. Notice that we refrained from using 3D or upper descriptors as these do require a careful selection of the most important three-dimensional conformations of the DES components accounting for temperature effects, as well as to avoid typical problems related to the choice of the most reliable quantum mechanical method used to compute those.[81] Moreover, 0D−2D descriptors are easily interpretable, and especially they are mostly suited for probing a toxicophore in a particular series of molecules.[60,82−86] Thus, 2D structures of the individual component molecules were initially drawn by ChemDraw,[87] and, afterward, 0D−2D descriptors for the individual components were calculated using the DRAGON software package.[88]

Even though the generated descriptors (D_{mix}) can differentiate among closely related data set samples, they are unable to characterize the different elements (b_t, m_e, and w_t) related to the assay conditions under which the samples were tested. To overcome this, we applied the Box−Jenkins approach.[61−63,84,89,90] The details of this approach have been extensively reported and discussed,[60−62] so we will limit ourselves to a brief outline underlining only the most important aspects. Initially, arithmetic means for a specific experimental condition c_j, $avg(D_i,D_{mix})c_j$, are calculated as follows:

$$avg(D_i, D_{mix})c_j = \frac{1}{n(c_j)} \sum_{i=1}^{n(c_j)} (D_i, D_{mix}) \quad (2)$$

In this equation, $n(c_j)$ is the number of data set samples that were assayed by the same element of the experimental condition (ontology) c_j. For example, in the case of the element b_t, $n(c_j)$ will be the number of compounds considered as active cases [$Tox_i(c_j) = +1$] that were tested against the same measure of toxic effect. Finally, new deviation descriptors are calculated by the expression below:

$$\Delta(D_i, D_{mix})c_j = (D_i, D_{mix})c_j - avg(D_i, D_{mix})c_j \quad (3)$$

The descriptors $\Delta(D_i,D_{mix})c_j$ are an adaption of the Box−Jenkins moving averages[61,62] and characterize both the chemical structures and the combination of the different elements of the experimental condition c_j. Only these descriptors were used for setting up the mtk-QSTR model. It is important to note here that, for monosaccharide carbohydrates, that is, D-fructose, D-glucose, and D-xylose, six-membered ring structures (i.e., pyranose forms) were used for the descriptors' calculation. However, it is well-known that these carbohydrates may exist in multiple structural forms (i.e., five-membered ring, furanose; six-membered rings, pyranose and straight-chain) that are in equilibrium with each other. One may thus calculate the descriptors of sugars as an average value of all of these forms. Because 2D descriptors are used in this work, whatever descriptor calculation technique is adopted, no significant differences are to be observed in the values of the final descriptors as well as in the statistical quality of the final model.

As usual, the modeling data set (n = 498) was split into two series: training and validation sets. The training set was used to derive the mkt-QSTR model, while the validation (or test) set was employed to assess the predictive ability of the final derived model. For that purpose, we employed the k-means cluster analysis (k-MCA) technique[91] to ensure that the validation set covers the same chemical/biological space as the training one.[92] After computing the DES descriptors (D_{mix}), the input descriptor matrix (comprising 5610 descriptors) along with the $Tox_i(c_j)$ values were subjected to k-MCA using the STATISTICA software[93,94] and considering Euclidian distances as metric. Five clusters were generated from 500 iterations, and the validation set samples were then gathered by ensuring that at least three samples were chosen from each cluster. In so doing, we ended up with 373 training set samples (178 toxic and 195 nontoxic) and 125 validation set samples (69 toxic and 56 nontoxic). Therefore, 20% of the data set samples were used as the validation set (test set 1).

Multitasking Model Setup and Evaluation. Linear discriminant analysis (LDA), specifically the LDA technique implemented in the STATISTICA software,[93] was employed here to find the classification model (eq 4), which best describes the toxicity $Tox_i(c_j)$, as a linear combination of the predictor X-variables (the $\Delta(D_i,D_{mix})c_j$ descriptors) weighted by the a_k coefficients:

$$Tox_i(c_j) = a_1X_1 + a_2X_2 + ... + a_kX_k + a_0 \quad (4)$$

The forward stepwise procedure was applied to select the descriptors (X-variables) with the highest influence on the toxicity. In so doing, highly correlated descriptors ($-0.7 < r < 0.7$, r being Pearson's correlation coefficient[95]) were disregarded, and the default values of the Fisher's statistics (F) were used. Thus, only descriptors with $F \geq 1$ and a p-value ≤ 0.05 were finally selected.

Several diagnostic statistical tools were employed for evaluating the performance of the developed mtk-QSTR classification model. Goodness-of-fit of the model was assessed by analyzing standard statistical indices for the training set, such as the Wilks' lambda (λ), chi-squared (χ^2), the square of Mahalanobis distance (D^2), Fisher's statistic (F), and the corresponding p-value (p). In addition, a Y-randomization test was carried out on the training set to check whether the model is unique and not developed by chance correlations.[82,96,97] To do so, the values of the dependent variable were randomly scrambled 100 times without changing the independent X-variables, and the Wilk's lambda (λ) of the original model was then compared to the average Wilk's lambda (λ_{rand}) of the randomized models.

Goodness-of-prediction for both the training and the test sets was evaluated by computing the following statistical measures: sensitivity (correct classification of the toxic cases), specificity (correct classification of nontoxic cases), accuracy (overall correct classification), Matthews correlation coefficient (MCC), and the area under receiver operating characteristic (ROC) curve,[98] plus the two ROC parameters based on distances proposed by Pérez-Garrido et al.:[99] the ROC graph Euclidean distance (ROCED), and the ROC graph Euclidean distance corrected with Fitness Function (FIT(λ)) (ROCFIT).

Finally, a careful assessment of the true predictivity of the final derived mtk-QSTR model was performed by determining its applicability domain using the confidence estimation approach proposed by Roy et al.[100] This approach is based on a posteriori probability values, and we set a cutoff value of 0.25 to identify possible outliers of the model. That is, all samples with differences in a posteriori probability (Prob.) values (i.e., |Prob.(+1) − Prob.(−1)|) less than 0.25 were considered as possible outliers of the model.

RESULTS AND DISCUSSION

Mtk-QSTR Model. The best mtk-QSTR model derived from the training set (an eight-variable equation, eq 5), by combining the LDA and FS techniques, is given below together with the statistical parameters of the LDA, while the meanings of the selected descriptors are shown in Table 2, and their calculated values are in Supporting Information, S2−S4.

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

Table 2. Molecular Descriptors Used in the Final mtk-QSTR Model[a]

symbol	meaning	type
ChiA_X	average Randić-like index derived from the Chi matrix	2D matrix-based
DLS_05	modified drug-like score from Zheng et al.[101]	0D drug-like index
SpAD_EA(ed)	spectral weighted absolute deviation from edge adjacency matrix	2D matrix-based
nHDon	number of donor atoms for H-bonds (N and O)	1D counts
B05[O−O]	presence/absence of O−O at topological distance 5	2D atom pairs
B03[C−N]	presence/absence of C−N at topological distance 3	2D atom pairs
Chi_D_z(p)	Randić-like index from Barysz matrix weighted by polarizability (p)	2D matrix-based
MATS1e	Moran autocorrelation of lag 1 weighted by Sanderson electronegativity (e)	2D autocorrelation

[a]Equation 5.

$$Tox_i(c_j) = +16.137\Delta(\text{ChiAo\_X})_{me} - 3.423\Delta(\text{DLS\_05})_{bt} + 0.086\Delta[\text{SpAD\_EA(ed)}]_{bt} - 1.759\Delta(\text{nHDON})_{me} + 1.325\Delta(\text{B05[O-O]})_{wt} + 4.808\Delta(\text{B03[C-N]})_{wt} - 4.929\Delta[\text{Chi\_D}_z\text{(p)}]_{me} - 9.462\Delta(\text{MATS1e})_{me} + 1.512 \quad (5)$$

where $N = 373$, $\lambda = 0.329$, $\chi^2 = 407.84$, $D^2 = 8.126$, $F(8364) = 92.740$, p-value $<10^{-16}$, ROCED = 0.204, and ROCFIT = 0.120. As can be seen, the large sample size, large χ^2, large F-index, high value of D^2, and the very low p-value are indicative of the model's statistical internal quality. Further, the small value of the Wilks' λ statistic (λ can take values from zero, perfect discrimination, to one, no discrimination) shows that the model displays an adequate ability for differentiating both groups. The latter is also confirmed by the classification results (see Table 3). In fact, the model correctly classifies 91.6% of the 178 toxic solvents and 95.4% of the 195 nontoxic ones, giving rise to an overall 93.6% effective discrimination for the 373 training set DES. Regarding the validation set (i.e., test set 1), 92.8% of the 69 toxic solvents and 94.6% of the 56 nontoxic ones were predicted correctly. These in turn show that the model was built with satisfactory external predictivity as well. Moreover, the high and well-balanced MCC values obtained for both the training and the validation sets (=0.871) suggest that a solid correlation was found between the observed and predicted values of the categorical variable $Tox_i(c_j)$.

Table 3. Overall Performance of the Final mtk-QSTR LDA Model

classification[a]	training set	test set 1
$NDES_{total}$	373	125
$NDES_{toxic}$	178	69
$CCDES_{toxic}$	163	64
sensitivity (%)	91.6	92.8
$NDES_{nontoxic}$	195	56
$CCDES_{nontoxic}$	186	53
specificity (%)	95.4	94.6
accuracy (%)	93.6	93.6
MCC	0.871	0.871

[a]NDES, number of DES cases; CCDES, correct classified DES cases; sensitivity, percentage of correct classified toxic DES cases; specificity, percentage of correct classified nontoxic DES cases; accuracy, percentage of overall correct classified cases; MCC, Mathews correlation coefficient.

Another aspect that deserves special attention is the degree of multicollinearity among the model' variables, but that may easily be diagnosed by examining the cross-correlation matrix. As seen in Table 4, the highest value observed for the Pearson correlation coefficient is 0.58. Therefore, the lack of high collinearity among the included variables (descriptors) indicates that the developed model is not redundant in nature.

Figure 2 shows the ROC curves obtained for the training (5-fold cross validation) and validation (test set 1) sets. One can easily see that the model is not a random, but a truly statistically significant classifier, because the area under the ROC curve for both sets (training, area = 0.969; test set 1, area = 0.957) sets is significantly higher than the area under the random classifier curve (diagonal line; area = 0.5).

The applicability domain (AD) of any QSTR model is the chemical and structural space within which the model's predictions are to be considered reliable.[96] As described previously, the AD of the model was defined using the confidence estimation (CE) approach with a threshold of 0.25.[100] Application of the latter, however, pinpointed that 22 compounds of the data set may be regarded as possible outliers (see Supporting Information, S5). As can be noticed in Table 5, after removal of these outliers from the data set, the new attained model is slightly statistically improved with respect to the previous model (cf., Table 3). Indeed, the average overall accuracy from both the training and the validation sets increased ca. 2% as well as the average MCC (from 0.871 to 0.888) when outliers are removed from the data set.

In addition, the Y-randomization test[82,96] was performed to check the uniqueness of the developed model. In so doing, the $Tox_i(c_j)$ response values were randomized 100 times and new models were built. The λ values of these randomized models were compared to that of the original model (eq 5). The average λ value for the randomized models was 0.977, which is much higher than the λ value (i.e., 0.258) of the original model. It justifies that our model was not generated by chance, and it is unique in nature. The randomized data and corresponding λ values of these new models are provided in the Supporting Information, S6.

As referred to before, in this work, we have applied the k-MCA technique for setting up the validation set, basically with the aim of ensuring that this set covers the same chemical/biological space as the training one. For mixtures, Oprisiu et al.[102] and Muratov et al.[103] have suggested however the following alternative validation approaches: (a) "points out", (b) "mixtures out", and (c) "compounds out". To judge the quality of our setting up approach, we have also applied such validation strategies to understand how the developed QSTR model performs (see eq 5). A detailed description of these validation techniques as well as of the statistical results obtained are provided in the Supporting Information. Overall, similar to eq 5, around 30 compounds of the modeling set were found to be wrongly predicted in each of these validation techniques. Therefore, one may infer that the statistical quality and predictivity of the developed mtk-QSTR model remain almost intact when different validation techniques (i.e., points out, mixtures out, and compounds out) are adopted.

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

Table 4. Intercorrelation among the Eight Descriptors Selected as Statistically Significant by the mtk-QSTR LDA Model[a]

	$\Delta(\text{MATS1e})_{me}$	$\Delta(\text{ChiA\_X})_{me}$	$\Delta(\text{B03[C−N]})_{wt}$	$\Delta[\text{SpAD\_EA(ed)}]_{bt}$	$\Delta[\text{Chi\_D}_z\text{(p)}]_{me}$	$\Delta(\text{B05[O−O]})_{wt}$	$\Delta(\text{nHDON})_{me}$	$\Delta(\text{DLS\_05})_{bt}$
$\Delta(\text{MATS1e})_{me}$	1.00	0.35	−0.03	0.32	−0.46	0.05	−0.18	0.34
$\Delta(\text{ChiA\_X})_{me}$	0.35	1.00	0.00	0.03	0.23	−0.02	0.03	0.17
$\Delta(\text{B03[C−N]})_{wt}$	−0.03	0.00	1.00	0.02	0.01	0.18	0.15	−0.04
$\Delta[\text{SpAD\_EA(ed)}]_{bt}$	0.32	0.03	0.02	1.00	−0.14	0.38	0.36	0.12
$\Delta[\text{Chi\_D}_z\text{(p)}]_{me}$	−0.46	0.23	0.01	−0.14	1.00	−0.17	−0.27	−0.39
$\Delta(\text{B05[O−O]})_{wt}$	0.05	−0.02	0.18	0.38	−0.17	1.00	0.58	0.06
$\Delta(\text{nHDON})_{me}$	−0.18	0.03	0.15	0.36	−0.27	0.58	1.00	0.12
$\Delta(\text{DLS\_05})_{bt}$	0.34	0.17	−0.04	0.12	−0.39	0.06	0.12	1.00

[a]Equation 5.

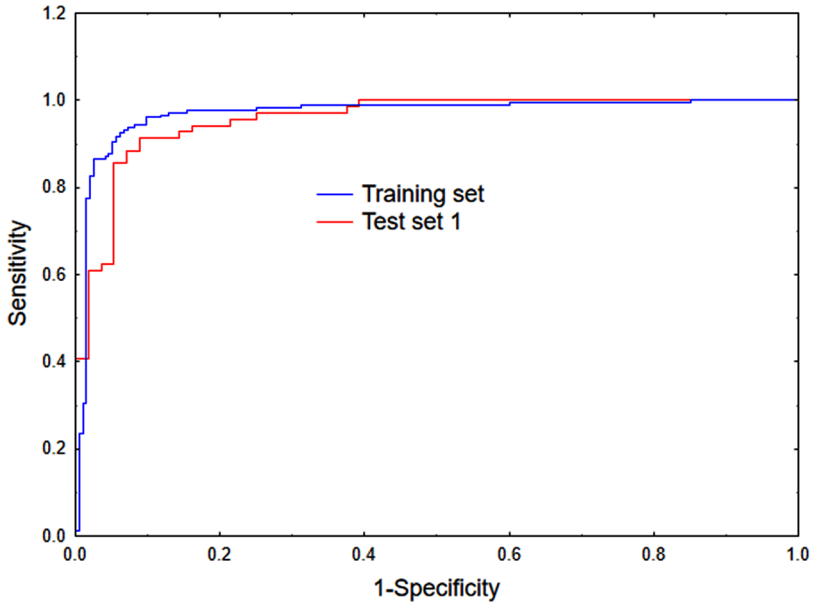


Figure 2. ROC curves for the training and validation (test set 1) sets.

Table 5. Performance of the mtk-QSTR LDA Model Derived after Removal of the Outliers

	after removal of outliers	
classification[a]	training set	test set 1
$\text{NDES}_{\text{total}}$	356	120
$\text{NDES}_{\text{toxic}}$	168	64
$\text{CCDES}_{\text{toxic}}$	158	60
sensitivity (%)	94.1	93.8
$\text{NDES}_{\text{nontoxic}}$	188	56
$\text{CCDES}_{\text{nontoxic}}$	179	53
specificity (%)	95.2	94.6
accuracy (%)	94.7	94.2
MCC	0.893	0.883

[a]Notation is the same as that in Table 4.

To further assess the predictive ability of the model, an external validation set comprising 74 new DES samples (test set 2) was assembled from three recently published works.[71,104,105] Here, it should be pointed out that, even though test set 2 included the same experimental elements, that is, m_e, b_t, and w_t, some of its DES components (e.g., tetramethylammonium chloride, tetraethylammonium chloride, tetrapropylammonium chloride, betaine, 1-propanol, and proline) were not present in our modeling data set. Therefore, the data set contained 15 DES mixtures formed by two new components, whereas 20 DES mixtures of the data set contained at least one new component.

The descriptors for the test set 2 DES samples were calculated following eq 3 using the $avg(D_i,D_{\text{mix}})c_j$ values for all samples of our modeling data set tested against the same element of c_j (see Supporting Information, S3). Details of this external data set, calculated descriptors, as well as the results of the predictions can be found in Supporting Information, S7−S10. For this whole external set, the percentage of overall discrimination is 87.8% (65 out of 73), and, simultaneously, the developed model correctly classifies 88.2% (15 out of 17) of toxic and 87.7% (50 out of 57) of nontoxic samples, respectively. Yet once more, by establishing the AD for this new data set, 11 nontoxic DES were found to be possible outliers, and, after removal of those, the model managed to correctly predict 44 out of 46 nontoxic samples, thus increasing its overall discrimination (95.7%) and specificity (93.7%) over external data. Given the diversity of such set and the complexity of the toxicological response being modeled, altogether this reveals the very good predictive ability of the present model.

To judge the quality of our mtk-QSTR model (eq 5), a comparative analysis with some previously derived mtk-QSAR/QSTR and PTML-QSTR models published in the last five years is shown in Table 6. As can be seen, all of these previously multitasking models were developed mainly for targeting single organic compounds or inorganic nanomateri-

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

Table 6. Comparison of the Derived mtk-QSTR Model with Previously Reported mtk-QSAR and Perturbation Theory Machine Learning (PTML) Models

model	data set	type	no. of cases[a]	no. of c_j[b]	accuracy (%) Tr[c]	accuracy (%) Ts[d]	ML[e]	ext. valid.[f]
mtk-QSTR[g]	DES toxicity	mixtures	499	3	93.60	93.60	LDA	yes
mtk-QSAR[61]	breast cancer-related protein inhibitors	single compounds	24285	3	93.24	93.16	LDA	no
mtk-QSAR[62]	antibacterial agents	single compounds	46229	4	97.15	97.25	LDA	no
mtk-QSAR[60]	antimicrobial peptides	peptides	3592	2	97.64	97.16	LDA	no
mtk-QSAR[64]	antimicrobial peptides	peptides	2488	1	94.37	94.77	LDA	no
PTML-QSTR[82]	nanoparticles (NPs) toxicity	NPs	54371	6	98.50	98.20	ANN	yes
PTML-QSTR[47]	NPs toxicity	NPs	36488	5	98.35	98.73	LDA	yes
PTML-QSTR[106]	NPs cytotoxicity	NPs	1681	4	93.58	93.57	LDA	yes

[a]Number of cases. [b]Number of conditions. [c]Training set. [d]Test set. [e]Machine learning techniques. [f]External validation. [g]This work.

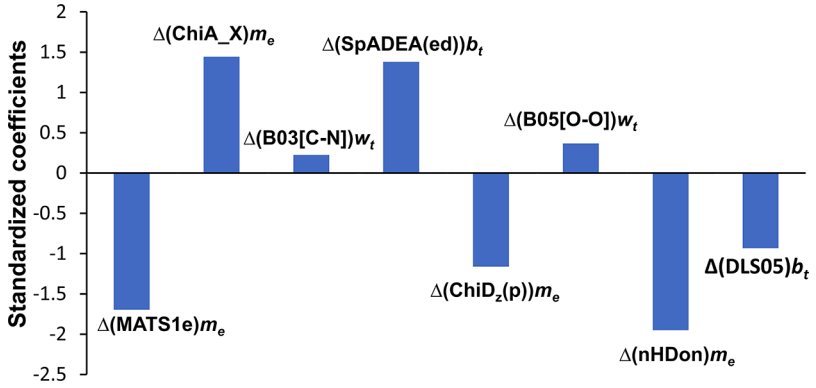


Figure 3. Standardized coefficients versus descriptors in the mtk-QSTR LDA model.

als. Even though our model targets mixtures and is based on a data set containing a comparatively smaller number of cases (as compared to the previously developed multitasking models), its development conditions as well as its statistical quality are similar to such previously developed models. Moreover, in contrast to the latter, the predictivity of our model is not only justified internally (test set) but also based on an external validation set.

Finally, our group has recently developed a stand-alone tool named QSAR-Co to assist both the development and the validation of mtk-QSAR/QSTR models. The tool is available in the weblink https://sites.google.com/view/qsar-co along with its manual. A detailed description of this tool is communicated elsewhere.[107]

Interpretation of the Model Descriptors. Interpreting a QSTR model in terms of the specific contribution of substituents and other molecular features plus the targeted experimental conditions to the modeled toxicity is always a difficult task. The present mtk-QDTR LDA model is based on eight descriptors of the 0D−2D type. As seen from the selected descriptors (Δ(ChiA_X)m_e, Δ(DLS_05)b_t, Δ(SpAD_EA-(ed))b_t, Δ(nHDon)m_e, Δ(B05[O−O])w_t, Δ(B03[C−N])w_t, Δ[Chi_D$_z$(p)]m_e, Δ(MATS1e)m_e), the kinds of measures of toxic effects (m_e), the biological targets (b_t), and water content (w_t) do have a balanced influence on the toxicity of this set of DES. Further, atomic polarizability (p) and electronegativity (e) also have influence on their toxicity. Moreover, the larger is the absolute value of a standardized LDA coefficient, the greater is the weight of the variable in the model, which therefore leads to the following ranking of contributions to $Tox_i(c_j)$ (see Figure 3): Δ(nHDon)m_e > Δ(MATS1e)m_e > Δ(ChiA_X)m_e ≈ Δ(SpAD_EA(ed))b_t > Δ[Chi_D$_z$(p)]m_e > Δ(DLS_05)b_t > Δ(B05[O−O]w_t > Δ(B03[C−N])w_t). All of these things considered, we will focus the analysis on the weighting variables of our mtk-QSTR model, because those are the most meaningful descriptors.

Interestingly, the most influential descriptor was found to be Δ(nHDon)m_e. nHDon is a 1D functional group counting descriptor pertaining to the number of hydrogen-bond donors (HBDs: N and O atoms). As seen, the higher is the number of H-bond donors, the lesser is the toxic nature of the DES. The same applies for the second most important descriptor found in the model, that is, the Δ(MATS1e)m_e descriptor that contributes negatively to toxicity. MATS1e is a Moran 2D autocorrelation descriptor of lag 1 weighted by the Sanderson electronegativity (e), which provides information about the interaction between close atom pairs. Nevertheless, the following significant descriptors, Δ(ChiA_X)m_e and Δ-(SpAD_EA(ed))b_t, make a positive contribution to the DES toxicity, while Δ[Chi_D$_z$(p)]m_e contributes negatively also. The latter are 2D matrix-based descriptors, calculated on the basis of graph-theoretical matrices by applying certain algebraic operators. In particular, Δ[Chi_D$_z$(p)] being weighted by atomic polarizability (p) shows the importance of this property to the underlying DES toxic effects. Following on, the most important descriptor is the drug like index Δ(DLS_05)b_t, depending on the biological targets (b_t) and which contributes also negatively to the toxicity. Zheng et al.[101] proposed the DLS_05 descriptor to assess the ADMET (absorption, distribution, metabolism, excretion, and toxicity) profile of drug-like compounds, and it is based on two rules. The first rule dictates that the index NO_C3 (ratio of total number of nitrogen and oxygen with sp^3 hybridized carbon atoms) should be in the range of 0.1−1.8, whereas according to the second

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

rule, the unsaturation index should be equal to or less than 0.43. Therefore, this descriptor mainly considers the characteristics of heteroatoms and unsaturation of the mixtures,[101] and the higher those are the lesser is the DES toxicity. Finally, the last important descriptors of the current mtk-QSTR model (eq 5) are $\Delta(B05[O-O])w_t$ and $\Delta(B03[C-N])w_t$, both depending on the water content (w_t) and contributing positively to the toxicity. These are 2D atom pairs descriptors, B03[C−N] accounting for the presence/absence of C−N at topological distance 3 and B05[O−O] for the presence/absence of O−O at topological distance 5.

The exact mechanism of action for the toxicity of DES is yet to be elucidated. However, it has been hypothesized earlier that the toxic effects of DES may be attributed to the charge delocalization occurring due to the H-bonding between the anion and the HBDs, which in turn tends to lower the binding affinity of the anion to the cellular membranes.[40,41] Similarly, the derived mtk-QSTR model depicted the importance of three key factors, HBD, polarizability, and electronegativity, all of these pertaining to charge delocalization facets that may exist in various DES mixtures. In addition, the lipophilicity and Hofmeister effects of the DES components, which depend on the kosmotropic or order-making and chaotropic or disorder-making nature of salts, have been reported to contribute to the overall toxicity of these mixtures. Therefore, it is not surprising that many complex topochemical 2D matrix-based descriptors, such as ChiA_X, SpAD_EA(ed), and Chi_D_z(p), were selected in the setup of our mtk-QSTR model. Another important descriptor of our model is DLS_05, originally proposed to assess ADMET properties,[99] and thus this descriptor can be considered also as a crucial factor to characterize the toxic nature of the DES.

Predicting the General Toxicity Behavior of DES Components. One of the most important aspects of the developed mtk-QSTR model is that it can help us understand the behavior of the DES individual components in eliciting toxicity. For such purpose, the model was used to quantitatively estimate the relative influence of the different HBD components under the diverse experimental conditions (i.e., a total of 53 experimental conditions depending on the combinations of the m_o, b_t, and w_t elements). Thirty different HBD components from our modeling data set were considered, and with the help of eq 5 we calculated the $Tox_i(c_j)$ scores for 1590 (=53 × 30) cases in the various experimental conditions tested (see details in Supporting Information, S11). From these, 1545 cases were retained after removal of the outliers detected by establishing the AD based on the CE approach.[94] When the average $Tox_i(c_j)$ score for each HBD component as well as its classification (toxic, +1; nontoxic, −1) were considered, an overall estimation of its contributions to the toxicity with respect to the multiple experimental conditions as well as to the type of systems (single or mixture) may be obtained. The classification results for these DES HBD components are shown in Table 7, along with their toxicity likelihood. The latter can be confirmed by examining the calculated average scores (score; see Table 7), because HBD components with negative average scores are to be regarded as less toxic than those with positive average scores.

As seen in Table 7, sucrose, xylitol, glycerol, and sorbitol may be regarded as the least toxic HBD components of the DES. Among other HBDs, acetamide, ethylene glycol, urea, diethylene glycol, lactic acid, glycine, and 1,4-butanediol have less tendency to induce toxicity. On the opposite, oxalic acid, phenylacetic acid, zinc chloride, benzoic acid, citric acid, tartaric acid, malic acid, malonic acid, *p*-toluenesulfonic acid, glycolic acid, levulinic acid, and acetic acid are likely to have the most positive contributions to the DES toxicity. Uncertain classifications are nevertheless obtained for glucose, maltose, fructose, triethylene glycol, and xylose. Interestingly, it is observed that organic and inorganic acids are comparatively more toxic than other HBD components. Save for lactic acid, all other acids are predicted to have toxic contributions. Most of the alcohols (e.g., xylitol, sorbitol, 1,2-propanediol, and ethylene glycol), on the contrary, are predicted to be relatively nontoxic. Natural sugars (e.g., sucrose, glucose, maltose, fructose, etc.) are also ranked as less toxic. Although acetic acid is found to deliver a toxic contribution, acetamide and urea are predicted to present nontoxic contributions to most of the DES systems.

To sum, sugar alcohols (e.g., sorbitol, xylitol, glycerol, and ethylene glycol) as well as alcohols with a straight chain (e.g., 1-propanol, 1,2-propanediol, and 1,4-butanediol) are the least toxic HBD components of the DES. Further, amides like urea and acetamide can also be considered as HBD components in DES due to their low contributions for toxicity. Yet sugars such as glucose, sucrose, fructose, and xylose may have intermediate levels of toxicity. Finally, all kinds of acids display the most positive contributions for the DES toxicity (see Figure 4).

Table 7. Classification of the DES HBD Components under the Various Experimental Conditions and Their Calculated Average $Tox_i(c_j)$ Scores (score)

component	toxic (+1)	nontoxic (−1)	score	likelihood of toxicity
oxalic acid	53	0	+5.673	high
phenylacetic acid	53	0	+4.635	high
zinc chloride	53	0	+4.289	high
benzoic acid	53	0	+4.207	high
citric acid	53	0	+3.628	high
tartaric acid	50	0	+3.576	high
malic acid	53	0	+3.477	high
p-toluenesulfonic acid	50	0	+3.472	high
malonic acid	50	3	+2.886	high
glycolic acid	49	4	+2.322	high
levulinic acid	47	4	+2.135	high
acetic acid	42	9	+1.439	high
xylose	29	22	+0.717	intermediate
fructose	22	24	+0.360	intermediate
triethylene glycol	21	28	+0.251	intermediate
maltose	21	31	+0.018	intermediate
glucose	21	31	+0.016	intermediate
1,4-butanediol	19	31	−0.068	low
glycine	18	32	−0.116	low
lactic acid	18	32	−0.174	low
diethylene glycol	18	32	−0.350	low
urea	18	35	−0.685	low
ethylene glycol	14	35	−1.020	low
acetamide	14	35	−1.026	low
1-propanol	14	39	−1.201	low
1,2-propanediol	14	39	−1.220	low
glycerol	14	39	−1.613	low
xylitol	14	39	−1.709	low
sorbitol	0	53	−2.481	low
sucrose	0	53	−2.481	Low

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

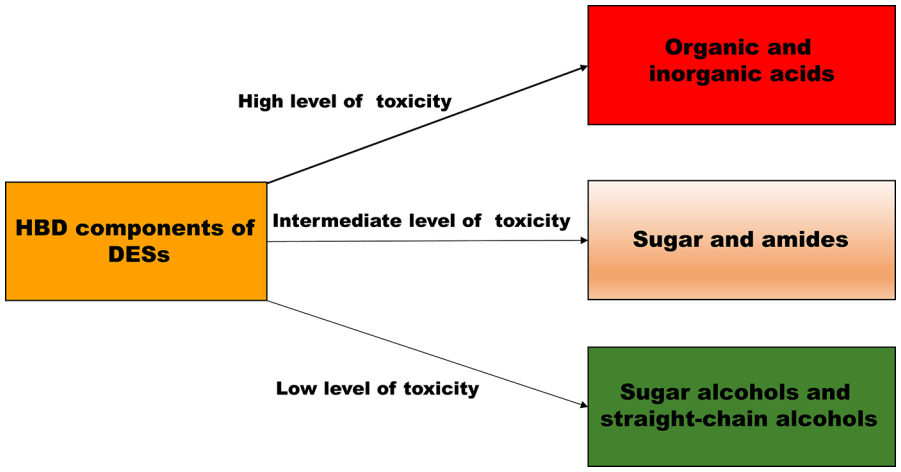


Figure 4. Likelihood for toxicity imparted by different groups of HBDs as DES components, according to the derived mtk-QSTR LDA model.

Similarly, when four HBAs of our current data set, that is, choline chloride, menthol, *N*,*N*-diethylethanol ammonium chloride (DEAC), and methyltriphenyl phosphonium bromide (MTPB), were analyzed in a similar manner, *N*,*N*-diethylethanol ammonium chloride was found to be the least toxic component followed by choline chloride. The other two components, MTPB and DEAC, were found to impart toxicity toward most of the systems. However, more experimental data are required to confirm their overall toxic contributions to the targeted biological systems.

CONCLUSIONS

DES are a new class of the so-called green solvents that have been attracting increasing attention due to their unique properties, low cost, and broad range of applications. Concerns over the potential hazardous effect of DES as well as of their components have recently been raised. However, so far only a few experimental works have targeted the toxicological properties of DES, and the data set reported is too scarce and complex for setting up reliable structure−toxicity relationships with conventional modeling approaches. Herein, we have examined the ability of a mtk-QSTR approach to provide a discriminant model for probing the toxicity of DES over multiple biological targets. Indeed, the combination of LDA in conjunction with a 0D−2D structure representation along with an efficient variable selection algorithm was found to produce a final classification model with very good accuracy as well as predictivity ability. In particular, the derived mtk-QSTR model provided new insights about the key structural requirements for developing more environmentally friendly DES. Further, the model was proved to be a reliable in silico tool for the assessment of the DES components effects contributing to their overall toxicity. To conclude, the physicochemical interpretation of the model descriptors, taken together with its statistical quality and ability for virtual screening, shows that our approach can play a key role in understanding the mechanisms of toxicity of DES. Furthermore, that can be productively applied toward the future design of novel nontoxic DES.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssuschemeng.9b01306.

Results of the additional validation of the mtk-QSTR LDA model, using the points out, mixtures out, and compounds out approaches (PDF)

Chemical and toxicological data of the DES under study as well as of their HBD components along with their classification (XLSX)

Chemical and toxicological data of the DES under study as well as of their HBD components along with their classification (XLSX)

Chemical and toxicological data of the DES under study as well as of their HBD components along with their classification (XLSX)

AUTHOR INFORMATION

Corresponding Author

*E-mail: ncordeir@fc.up.pt.

ORCID

Amit Kumar Halder: 0000-0002-4818-9047

M. Natália D. S. Cordeiro: 0000-0003-3375-8670

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by UID/QUI/50006/2019 with funding from FCT/MCTES through national funds.

REFERENCES

(1) Anastas, P. T.; Warner, R. *Green Chemistry, Theory and Practice*; Oxford University Press: New York, 1998; p 30.

(2) Tobiszewski, M.; Mechlinska, A.; Namiesnik, J. Green analytical chemistry-theory and practice. *Chem. Soc. Rev.* **2010**, *39* (8), 2869−2878.

(3) Keith, L. H.; Gron, L. U.; Young, J. L. Green analytical methodologies. *Chem. Rev.* **2007**, *107* (6), 2695−2708.

(4) Kudlak, B.; Owczarek, K.; Namiesnik, J. Selected issues related to the toxicity of ionic liquids and deep eutectic solvents-a review. *Environ. Sci. Pollut. Res.* **2015**, *22* (16), 11975−11992.

(5) Bubalo, M. C.; Radosevic, K.; Redovnikovic, I. R.; Halambek, J.; Srcek, V. G. A brief overview of the potential environmental hazards of ionic liquids. *Ecotoxicol. Environ. Saf.* **2014**, *99*, 1−12.
(6) Das, R. N.; Roy, K. Advances in QSPR/QSTR models of ionic liquids for the design of greener solvents of the future. *Mol. Diversity* **2013**, *17* (1), 151−196.
(7) Juneidi, I.; Hayyan, M.; Ali, O. M. Toxicity profile of choline chloride-based deep eutectic solvents for fungi and Cyprinus carpio fish. *Environ. Sci. Pollut. Res.* **2016**, *23* (8), 7648−7659.
(8) Pham, T. P.; Cho, C. W.; Yun, Y. S. Environmental fate and toxicity of ionic liquids: a review. *Water Res.* **2010**, *44* (2), 352−372.
(9) Smith, E. L.; Abbott, A. P.; Ryder, K. S. Deep eutectic solvents (DESs) and their applications. *Chem. Rev.* **2014**, *114* (21), 11060−11082.
(10) Shishov, A.; Bulatov, A.; Locatelli, M.; Carradori, S.; Andruch, V. Application of deep eutectic solvents in analytical chemistry. A review. *Microchem. J.* **2017**, *135* (Supplement C), 33−38.
(11) Płotka-Wasylka, J.; Rutkowska, M.; Owczarek, K.; Tobiszewski, M.; Namiesnik, J. Extraction with environmentally friendly solvents. *TrAC, Trends Anal. Chem.* **2017**, *91* (Supplement C), 12−25.
(12) Carriazo, D.; Serrano, M. C.; Gutierrez, M. C.; Ferrer, M. L.; del Monte, F. Deep-eutectic solvents playing multiple roles in the synthesis of polymers and related materials. *Chem. Soc. Rev.* **2012**, *41* (14), 4996−5014.
(13) Francisco, M.; van den Bruinhorst, A.; Kroon, M. C. Low-transition-temperature mixtures (LTTMs): a new generation of designer solvents. *Angew. Chem., Int. Ed.* **2013**, *52* (11), 3074−3085.
(14) Xu, P.; Zheng, G. W.; Zong, M. H.; Li, N.; Lou, W. Y. Recent progress on deep eutectic solvents in biocatalysis. *Bioresour. Bioprocess.* **2017**, *4* (1), 34.
(15) Wagle, D. V.; Zhao, H.; Baker, G. A. Deep eutectic solvents: sustainable media for nanoscale and functional materials. *Acc. Chem. Res.* **2014**, *47* (8), 2299−2308.
(16) Atilhan, M.; Aparicio, S. Deep eutectic solvents on the surface of face centered cubic metals. *J. Phys. Chem. C* **2016**, *120* (19), 10400−10409.
(17) González-Martínez, D.; Gotor, V.; Gotor-Fernández, V. Application of deep eutectic solvents in promiscuous lipase-catalysed aldol reactions. *Eur. J. Org. Chem.* **2016**, *2016* (8), 1513−1519.
(18) Lindberg, D.; Revenga, M. F.; Widersten, M. Deep eutectic solvents (DESs) are viable cosolvents for enzyme-catalyzed epoxide hydrolysis. *J. Biotechnol.* **2010**, *147* (3−4), 169−171.
(19) Wang, A.; Zheng, X.; Zhao, Z.; Li, C.; Zheng, X. Deep eutectic solvents to organic synthesis. *Prog. Chem.* **2014**, *26*, 784−795.
(20) Paiva, A.; Craveiro, R.; Aroso, I.; Martins, M.; Reis, R. L.; Duarte, A. R. C. Natural deep eutectic solvents − solvents for the 21st century. *ACS Sustainable Chem. Eng.* **2014**, *2* (5), 1063−1071.
(21) Li, X.; Row, K. H. Development of deep eutectic solvents applied in extraction and separation. *J. Sep. Sci.* **2016**, *39* (18), 3505−3520.
(22) Abo-Hamad, A.; Hayyan, M.; AlSaadi, M. A.; Hashim, M. A. Potential applications of deep eutectic solvents in nanotechnology. *Chem. Eng. J.* **2015**, *273* (Supplement C), 551−567.
(23) Guajardo, N.; Müller, C. R.; Schrebler, R.; Carlesi, C.; de María, P. D. Deep eutectic solvents for organocatalysis, biotransformations, and multistep organocatalyst/enzyme combinations. *ChemCatChem* **2016**, *8* (6), 1020−1027.
(24) Lawes, S. D. A.; Hainsworth, S. V.; Blake, P.; Ryder, K. S.; Abbott, A. P. Lubrication of steel/steel contacts by choline chloride ionic liquids. *Tribol. Lett.* **2010**, *37* (2), 103−110.
(25) Faggian, M.; Sut, S.; Perissutti, B.; Baldan, V.; Grabnar, I.; Dall'Acqua, S. Natural deep eutectic solvents (nades) as a tool for bioavailability improvement: pharmacokinetics of rutin dissolved in proline/glycine after oral administration in rats: possible application in nutraceuticals. *Molecules* **2016**, *21* (11), 1531.
(26) Serrano, M. C.; Gutierrez, M. C.; Jimenez, R.; Ferrer, M. L.; del Monte, F. Synthesis of novel lidocaine-releasing poly(diol-co-citrate) elastomers by using deep eutectic solvents. *Chem. Commun. (Cambridge, U. K.)* **2012**, *48* (4), 579−581.
(27) Morrison, H. G.; Sun, C. C.; Neervannan, S. Characterization of thermal behavior of deep eutectic solvents and their potential as drug solubilization vehicles. *Int. J. Pharm.* **2009**, *378* (1−2), 136−139.
(28) Dai, Y.; van Spronsen, J.; Witkamp, G. J.; Verpoorte, R.; Choi, Y. H. Natural deep eutectic solvents as new potential media for green technology. *Anal. Chim. Acta* **2013**, *766*, 61−68.
(29) Abbott, A. P.; Boothby, D.; Capper, G.; Davies, D. L.; Rasheed, R. K. Deep eutectic solvents formed between choline chloride and carboxylic acids: versatile alternatives to ionic liquids. *J. Am. Chem. Soc.* **2004**, *126* (29), 9142−9147.
(30) Singh, B. S.; Lobo, H. R.; Shankarling, G. S. Choline chloride based eutectic solvents: Magical catalytic system for carbon−carbon bond formation in the rapid synthesis of *β*-hydroxy functionalized derivatives. *Catal. Commun.* **2012**, *24* (Supplement C), 70−74.
(31) Hayyan, M.; Hashim, M. A.; Al-Saadi, M. A.; Hayyan, A.; AlNashef, I. M.; Mirghani, M. E. Assessment of cytotoxicity and toxicity for phosphonium-based deep eutectic solvents. *Chemosphere* **2013**, *93* (2), 455−459.
(32) Hayyan, M.; Hashim, M. A.; Hayyan, A.; Al-Saadi, M. A.; AlNashef, I. M.; Mirghani, M. E.; Saheed, O. K. Are deep eutectic solvents benign or toxic? *Chemosphere* **2013**, *90* (7), 2193−2195.
(33) Hayyan, M.; Looi, C. Y.; Hayyan, A.; Wong, W. F.; Hashim, M. A. In vitro and in vivo toxicity profiling of ammonium-based deep eutectic solvents. *PLoS One* **2015**, *10* (2), No. e0117934.
(34) Hayyan, M.; Mbous, Y. P.; Looi, C. Y.; Wong, W. F.; Hayyan, A.; Salleh, Z.; Mohd-Ali, O. Natural deep eutectic solvents: cytotoxic profile. *SpringerPlus* **2016**, *5* (1), 913.
(35) Aroso, I. M.; Silva, J. C.; Mano, F.; Ferreira, A. S.; Dionisio, M.; Sa-Nogueira, I.; Barreiros, S.; Reis, R. L.; Paiva, A.; Duarte, A. R. Dissolution enhancement of active pharmaceutical ingredients by therapeutic deep eutectic systems. *Eur. J. Pharm. Biopharm.* **2016**, *98*, 57−66.
(36) de Morais, P.; Gonçalves, F.; Coutinho, J. A. P.; Ventura, S. P. M. Ecotoxicity of cholinium-based deep eutectic solvents. *ACS Sustainable Chem. Eng.* **2015**, *3* (12), 3398−3404.
(37) Juneidi, I.; Hayyan, M.; Hashim, M. A. Evaluation of toxicity and biodegradability for cholinium-based deep eutectic solvents. *RSC Adv.* **2015**, *5* (102), 83636−83647.
(38) Mbous, Y. P.; Hayyan, M.; Wong, W. F.; Looi, C. Y.; Hashim, M. A. Unraveling the cytotoxicity and metabolic pathways of binary natural deep eutectic solvent systems. *Sci. Rep.* **2017**, *7*, 41257.
(39) Radosevic, K.; Bubalo, M. C.; Srcek, V. G.; Grgas, D.; Dragicevic, T. L.; Redovnikovic, I. R. Evaluation of toxicity and biodegradability of choline chloride based deep eutectic solvents. *Ecotoxicol. Environ. Saf.* **2015**, *112*, 46−53.
(40) Radosevic, K.; Zeleznjak, J.; Bubalo, M. C.; Redovnikovic, I. R.; Slivac, I.; GaurinaSrcek, V. Comparative in vitro study of cholinium-based ionic liquids and deep eutectic solvents toward fish cell line. *Ecotoxicol. Environ. Saf.* **2016**, *131*, 30−36.
(41) Wen, Q.; Chen, J. X.; Tang, Y. L.; Wang, J.; Yang, Z. Assessing the toxicity and biodegradability of deep eutectic solvents. *Chemosphere* **2015**, *132*, 63−69.
(42) Zhao, B.-Y.; Xu, P.; Yang, F.-X.; Wu, H.; Zong, M.-H.; Lou, W.-Y. Biocompatible deep eutectic solvents based on choline chloride: characterization and application to the extraction of rutin from Sophora japonica. *ACS Sustainable Chem. Eng.* **2015**, *3* (11), 2746−2755.
(43) Worth, A. P.; Bassan, A.; De Bruijn, J.; Gallegos Saliner, A.; Netzeva, T.; Patlewicz, G.; Pavan, M.; Tsakovska, I.; Eisenreich, S. The role of the European Chemicals Bureau in promoting the regulatory use of (Q)SAR methods. *SAR QSAR Environ. Res.* **2007**, *18* (1−2), 111−125.
(44) Cho, C.-W.; Stolte, S.; Yun, Y.-S. Comprehensive approach for predicting toxicological effects of ionic liquids on several biological systems using unified descriptors. *Sci. Rep.* **2016**, *6*, 33403.
(45) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discovery* **2010**, *9* (3), 203−214.

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

(46) Singh, K. P.; Gupta, S.; Basant, N.; Mohan, D. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. *Chem. Res. Toxicol.* **2014**, *27* (9), 1504−1515.
(47) Kleandrova, V. V.; Luan, F.; Gonzalez-Diaz, H.; Ruso, J. M.; Speck-Planche, A.; Cordeiro, M. N.D.S. Computational tool for risk assessment of nanomaterials: novel QSTR-perturbation model for simultaneous prediction of ecotoxicity and cytotoxicity of uncoated and coated nanoparticles under multiple experimental conditions. *Environ. Sci. Technol.* **2014**, *48* (24), 14686−14694.
(48) Asadollahi-Baboli, M. Exploring QSTR analysis of the toxicity of phenols and thiophenols using machine learning methods. *Environ. Toxicol. Pharmacol.* **2012**, *34* (3), 826−831.
(49) OECD; Environment Health and Safety Publications Series on Testing and Assessment No. 69. Guidance document on the validation of (quantitative) structure-activity relationship [(Q) SAR] models, 2017.
(50) CvjetkoBubalo, M.; Radošević, K.; GaurinaSrček, V.; Das, R. N.; Popelier, P.; Roy, K. Cytotoxicity towards CCO cells of imidazolium ionic liquids with functionalized side chains: Preliminary QSTR modeling using regression and classification based approaches. *Ecotoxicol. Environ. Saf.* **2015**, *112* (Supplement C), 22−28.
(51) Roy, K.; Das, R. N. QSTR with extended topochemical atom (ETA) indices. 16. Development of predictive classification and regression models for toxicity of ionic liquids towards Daphnia magna. *J. Hazard. Mater.* **2013**, *254*−*255*, 166−178.
(52) Zhao, Y.; Zhao, J.; Huang, Y.; Zhou, Q.; Zhang, X.; Zhang, S. Toxicity of ionic liquids: database and prediction via quantitative structure-activity relationship method. *J. Hazard. Mater.* **2014**, *278*, 320−329.
(53) Yan, F.; Shang, Q.; Xia, S.; Wang, Q.; Ma, P. Topological study on the toxicity of ionic liquids on Vibrio fischeri by the quantitative structure-activity relationship method. *J. Hazard. Mater.* **2015**, *286*, 410−415.
(54) Roy, K.; Das, R. N.; Popelier, P. L. Quantitative structure-activity relationship for toxicity of ionic liquids to Daphnia magna: aromaticity vs. lipophilicity. *Chemosphere* **2014**, *112*, 120−127.
(55) Peric, B.; Sierra, J.; Marti, E.; Cruanas, R.; Garau, M. A. Quantitative structure-activity relationship (QSAR) prediction of (eco)toxicity of short aliphatic protic ionic liquids. *Ecotoxicol. Environ. Saf.* **2015**, *115*, 257−262.
(56) Das, R. N.; Roy, K.; Popelier, P. L. Interspecies quantitative structure-toxicity-toxicity (QSTTR) relationship modeling of ionic liquids. Toxicity of ionic liquids to V. fischeri, D. magna and S. vacuolatus. *Ecotoxicol. Environ. Saf.* **2015**, *122*, 497−520.
(57) Fatemi, M. H.; Izadiyan, P. Cytotoxicity estimation of ionic liquids based on their effective structural features. *Chemosphere* **2011**, *84* (5), 553−563.
(58) Sosnowska, A.; Barycki, M.; Zaborowska, M.; Rybinska, A.; Puzyn, T. Towards designing environmentally safe ionic liquids: the influence of the cation structure. *Green Chem.* **2014**, *16* (11), 4749−4757.
(59) Ahmadi, R.; Hemmateenejad, B.; Safavi, A.; Shojaeifard, Z.; Mohabbati, M.; Firuzi, O. Assessment of cytotoxicity of choline chloride-based natural deep eutectic solvents against human HEK-293 cells: A QSAR analysis. *Chemosphere* **2018**, *209*, 831−838.
(60) Kleandrova, V. V.; Ruso, J. M.; Speck-Planche, A.; Cordeiro, M. N. D. S. Enabling the discovery and virtual screening of potent and safe antimicrobial peptides. simultaneous prediction of antibacterial activity and cytotoxicity. *ACS Comb. Sci.* **2016**, *18* (8), 490−498.
(61) Speck-Planche, A.; Cordeiro, M. N. D. S. Fragment-based in silico modeling of multi-target inhibitors against breast cancer-related proteins. *Mol. Diversity* **2017**, *21* (3), 511−523.
(62) Speck-Planche, A.; Cordeiro, M. N. D. S. De novo computational design of compounds virtually displaying potent antibacterial activity and desirable in vitro ADMET profiles. *Med. Chem. Res.* **2017**, *26* (10), 2345−2356.
(63) Alonso, N.; Caamano, O.; Romero-Duran, F. J.; Luan, F.; Cordeiro, M. N. D. S.; Yanez, M.; Gonzalez-Diaz, H.; Garcia-Mera, X. Model for high-throughput screening of multitarget drugs in chemical neurosciences: synthesis, assay, and theoretic study of rasagiline carbamates. *ACS Chem. Neurosci.* **2013**, *4* (10), 1393−1403.
(64) Speck-Planche, A.; Kleandrova, V. V.; Ruso, J. M.; Cordeiro, M. N. D. S. First multitarget chemo-bioinformatic model to enable the discovery of antibacterial peptides against multiple gram-positive pathogens. *J. Chem. Inf. Model.* **2016**, *56* (3), 588−598.
(65) Tenorio-Borroto, E.; Penuelas-Rivas, C. G.; Vasquez-Chagoyan, J. C.; Castanedo, N.; Prado-Prado, F. J.; Garcia-Mera, X.; Gonzalez-Diaz, H. Model for high-throughput screening of drug immunotoxicity–study of the anti-microbial G1 over peritoneal macrophages using flow cytometry. *Eur. J. Med. Chem.* **2014**, *72*, 206−220.
(66) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Unified multi-target approach for the rational in silico design of anti-bladder cancer agents. *Anti-Cancer Agents Med. Chem.* **2013**, *13* (5), 791−800.
(67) Casanola-Martin, G. M.; Huong, L. T. T.; Perez-Gimenez, F.; Marrero-Ponce, Y.; Merino-Sanjuan, M.; Abad, C.; Gonzalez-Diaz, H. Multi-output model with Box-Jenkins operators of linear indices to predict multi-target inhibitors of ubiquitin-proteasome pathway. *Mol. Diversity* **2015**, *19* (2), 347−356.
(68) Speck-Planche, A.; Luan, F.; Cordeiro, M. N.D.S. Role of ligand-based drug design methodologies toward the discovery of new anti- Alzheimer agents: futures perspectives in Fragment-Based Ligand Design. *Curr. Med. Chem.* **2012**, *19* (11), 1635−1645.
(69) Speck-Planche, A.; Cordeiro, M. N. D. S. Advanced *in silico* approaches for drug discovery: mining information from multiple biological and chemical data through mtk-QSBER and pt-QSPR strategies. *Curr. Med. Chem.* **2017**, *24*, 1−18.
(70) Gonzalez-Diaz, H.; Arrasate, S.; Gomez-SanJuan, A.; Sotomayor, N.; Lete, E.; Besada-Porto, L.; Ruso, J. M. General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry. *Curr. Top. Med. Chem.* **2013**, *13* (14), 1713−41.
(71) Radosevic, K.; Canak, I.; Panic, M.; Markov, K.; Bubalo, M. C.; Frece, J.; Srcek, V. G.; Redovnikovic, I. R. Antimicrobial, cytotoxic and antioxidative evaluation of natural deep eutectic solvents. *Environ. Sci. Pollut. Res.* **2018**, *25* (14), 14188−14196.
(72) Passino, D. R. M.; Smith, S. B. Acute bioassays and hazard evaluation of representative contaminants detected in great lakes fish. *Environ. Toxicol. Chem.* **1987**, *6* (11), 901−907.
(73) García, G.; Aparicio, S.; Ullah, R.; Atilhan, M. Deep eutectic solvents: physicochemical properties and gas separation applications. *Energy Fuels* **2015**, *29* (4), 2616−2644.
(74) Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Clarendon Press: Oxford; New York, 1990.
(75) García, G.; Atilhan, M.; Aparicio, S. An approach for the rationalization of melting temperature for deep eutectic solvents from DFT. *Chem. Phys. Lett.* **2015**, *634* (Supplement C), 151−155.
(76) Oprisiu, I.; Novotarskyi, S.; Tetko, I. V. Modeling of non-additive mixture properties using the Online CHEmical database and Modeling environment (OCHEM). *J. Cheminf.* **2013**, *5* (1), 4.
(77) Ajmani, S.; Rogers, S. C.; Barley, M. H.; Livingstone, D. J. Application of QSPR to mixtures. *J. Chem. Inf. Model.* **2006**, *46* (5), 2043−2055.
(78) Oprisiu, I.; Varlamova, E.; Muratov, E.; Artemenko, A.; Marcou, G.; Polishchuk, P.; Kuz'min, V.; Varnek, A. QSPR approach to predict nonadditive properties of mixtures. application to bubble point temperatures of binary mixtures of liquids. *Mol. Inf.* **2012**, *31* (6−7), 491−502.
(79) Luan, F.; Xu, X.; Liu, H. T.; Cordeiro, M. N. D. S. Prediction of the baseline toxicity of non-polar narcotic chemical mixtures by QSAR approach. *Chemosphere* **2013**, *90* (6), 1980−1986.
(80) Wang, T.; Tang, L. L.; Luan, F.; Cordeiro, M. N. D. S. Prediction of the toxicity of binary mixtures by qsar approach using the hypothetical descriptors. *Int. J. Mol. Sci.* **2018**, *19* (11), 3423.
(81) Hechinger, M.; Leonhard, K.; Marquardt, W. What is wrong with quantitative structure-property relations models based on three-

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660

dimensional descriptors? *J. Chem. Inf. Model.* **2012**, *52* (8), 1984−1993.
(82) Concu, R.; Kleandrova, V. V.; Speck-Planche, A.; Cordeiro, M. N. D. S. Probing the toxicity of nanoparticles: a unified in silico machine learning model based on perturbation theory. *Nanotoxicology* **2017**, *11* (7), 891−906.
(83) Helguera, A. M.; Cordeiro, M. N. D. S.; Perez, M. A.; Combes, R. D.; Gonzalez, M. P. Quantitative structure carcinogenicity relationship for detecting structural alerts in nitroso-compounds: species: rat; sex: male; route of administration: water. *Toxicol. Appl. Pharmacol.* **2008**, *231* (2), 197−207.
(84) Speck-Planche, A.; Cordeiro, M. N. D. S. Chemoinformatics for medicinal chemistry: in silico model to enable the discovery of potent and safer anti-cocci agents. *Future Med. Chem.* **2014**, *6* (18), 2013−2028.
(85) Speck-Planche, A.; Kleandrova, V. V.; Cordeiro, M. N. D. S. Chemoinformatics for rational discovery of safe antibacterial drugs: simultaneous predictions of biological activity against streptococci and toxicological profiles in laboratory animals. *Bioorg. Med. Chem.* **2013**, *21* (10), 2727−2732.
(86) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Predicting multiple ecotoxicological profiles in agrochemical fungicides: a multi-species chemoinformatic approach. *Ecotoxicol. Environ. Saf.* **2012**, *80*, 308−313.
(87) CambridgeSoft. *Chem Draw Ultra v8.0*; Cambridge, 2003; http://www.cambridgesoft.com/.
(88) Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. Dragon software: An easy approach to molecular descriptor calculations. *MATCH Commun. Math. Comput. Chem.* **2006**, *56*, 237−248.
(89) Romero-Duran, F. J.; Alonso, N.; Yanez, M.; Caamano, O.; Garcia-Mera, X.; Gonzalez-Diaz, H. Brain-inspired cheminformatics of drug-target brain interactome, synthesis, and assay of TVP1022 derivatives. *Neuropharmacology* **2016**, *103*, 270−278.
(90) Speck-Planche, A.; Cordeiro, M. N. D. S. Simultaneous modeling of antimycobacterial activities and ADMET profiles: a chemoinformatic approach to medicinal chemistry. *Curr. Top. Med. Chem.* **2013**, *13* (14), 1656−1665.
(91) Hill, T.; Lewicki, P. *STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining*; StatSoft: Tulsa, 2006.
(92) Gore, P. A. J. Cluster Analysis. *Handbook of Applied Multivariate Statistics and Mathematical Modeling*; Academic Press: New York, 2000; pp 298−318.
(93) Statsoft-Team. *STATISTICA. Data Analysis Software System, v6.0*; Tulsa, 2001.
(94) Ojha, P. K.; Roy, K. Chemometric modeling, docking and in silico design of triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors as antimalarials. *Eur. J. Med. Chem.* **2010**, *45* (10), 4645−4656.
(95) Pearson, K. Note on regression and inheritance in the case of two parents. *Proc. R. Soc. London* **1895**, *58*, 240−242.
(96) Rucker, C.; Rucker, G.; Meringer, M. y-Randomization and its variants in QSPR/QSAR. *J. Chem. Inf. Model.* **2007**, *47* (6), 2345−2357.
(97) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Rational drug design for anti-cancer chemotherapy: Multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. *Bioorg. Med. Chem.* **2012**, *20* (15), 4848−4855.
(98) Hanczar, B.; Hua, J.; Sima, C.; Weinstein, J.; Bittner, M.; Dougherty, E. R. Small-sample precision of ROC-related estimates. *Bioinformatics* **2010**, *26* (6), 822−830.
(99) Perez-Garrido, A.; Helguera, A. M.; Borges, F.; Cordeiro, M. N.D.S.; Rivero, V.; Escudero, A. G. Two new parameters based on distances in a receiver operating characteristic chart for the selection of classification models. *J. Chem. Inf. Model.* **2011**, *51* (10), 2746−2759.
(100) Ambure, P.; Bhat, J.; Puzyn, T.; Roy, K. Identifying natural compounds as multi-target-directed ligands against Alzheimer's disease: an in silico approach. *J. Biomol. Struct. Dyn.* **2019**, *37*, 1−25.
(101) Zheng, S.; Luo, X.; Chen, G.; Zhu, W.; Shen, J.; Chen, K.; Jiang, H. A new rapid and effective chemistry space filter in recognizing a druglike database. *J. Chem. Inf. Model.* **2005**, *45* (4), 856−862.
(102) Oprisiu, I.; Novotarskyi, S.; Tetko, I. V. Modeling of non-additive mixture properties using the Online CHEmical database and Modeling environment (OCHEM). *J. Cheminf.* **2013**, *5* (1), 4.
(103) Muratov, E. N.; Varlamova, E. V.; Artemenko, A. G.; Polishchuk, P. G.; Kuz'min, V. E. Existing and developing approaches for qsar analysis of mixtures. *Mol. Inf.* **2012**, *31* (3−4), 202−21.
(104) Macario, I. P. E.; Ventura, S. P. M.; Pereira, J. L.; Goncalves, A. M. M.; Coutinho, J. A. P.; Goncalves, F. J. M. The antagonist and synergist potential of cholinium-based deep eutectic solvents. *Ecotoxicol. Environ. Saf.* **2018**, *165*, 597−602.
(105) Macario, I. P. E.; Jesus, F.; Pereira, J. L.; Ventura, S. P. M.; Goncalves, A. M. M.; Coutinho, J. A. P.; Goncalves, F. J. M. Unraveling the ecotoxicity of deep eutectic solvents using the mixture toxicity theory. *Chemosphere* **2018**, *212*, 890−897.
(106) Luan, F.; Kleandrova, V. V.; Gonzalez-Diaz, H.; Ruso, J. M.; Melo, A.; Speck-Planche, A.; Cordeiro, M. N. D. S. Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach. *Nanoscale* **2014**, *6* (18), 10623−10630.
(107) Ambure, P.; Halder, A. K.; González-Díaz, H.; Cordeiro, M. N. D. S. QSAR-Co: An Open Source Software for Developing Robust Multi-tasking or Multi-target Classification-Based QSAR Models. *J. Chem. Inf. Model.* **2019**, 1 DOI: 10.1021/acs.jcim.9b00295.

DOI: 10.1021/acssuschemeng.9b01306
ACS Sustainable Chem. Eng. 2019, 7, 10649−10660