

## Article

# Predicting the Surface Tension of Deep Eutectic Solvents: A Step Forward in the Use of Greener Solvents

Amit Kumar Halder <sup>1,2,\*</sup> , Reza Haghbakhsh <sup>3,4</sup> , Iuliia V. Voroshylova <sup>1</sup> , Ana Rita C. Duarte <sup>4</sup> and Maria Natalia D. S. Cordeiro <sup>1,\*</sup> 

- <sup>1</sup> LAQV@REQUIMTE, Department of Chemistry and Biochemistry, Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal; up501129@g.uporto.pt
- <sup>2</sup> Dr B C, Roy College of Pharmacy and Allied Health Sciences, Dr. Meghnad Saha Sarani, Bidhannagar, Durgapur 713212, WB, India
- <sup>3</sup> Department of Chemical Engineering, Faculty of Engineering, University of Isfahan, Isfahan 81746-73441, Iran; haghbakhsh@gmail.com
- <sup>4</sup> LAQV@REQUIMTE, Department of Chemistry, NOVA School of Science and Technology, 2829-516 Caparica, Portugal; ard08968@fct.unl.pt
- \* Correspondence: amit.halder@fc.up.pt (A.K.H.); ncordeir@fc.up.pt (M.N.D.S.C.); Tel.: +351-2240-2502 (M.N.D.S.C.)

**Abstract:** Deep eutectic solvents (DES) are an important class of green solvents that have been developed as an alternative to toxic solvents. However, the large-scale industrial application of DESs requires fine-tuning their physicochemical properties. Among others, surface tension is one of such properties that have to be considered while designing novel DESs. In this work, we present the results of a detailed evaluation of Quantitative Structure-Property Relationships (QSPR) modeling efforts designed to predict the surface tension of DESs, following the Organization for Economic Co-operation and Development (OECD) guidelines. The data set used comprises a large number of structurally diverse binary DESs and the models were built systematically through rigorous validation methods, including ‘mixtures-out’- and ‘compounds-out’-based data splitting. The most predictive individual QSPR model found is shown to be statistically robust, besides providing valuable information about the structural and physicochemical features responsible for the surface tension of DESs. Furthermore, the intelligent consensus prediction strategy applied to multiple predictive models led to consensus models with similar statistical robustness to the individual QSPR model. The benefits of the present work stand out also from its reproducibility since it relies on fully specified computational procedures and on publicly available tools. Finally, our results not only guide the future design and screening of novel DESs with a desirable surface tension but also lays out strategies for efficiently setting up silico-based models for binary mixtures.

**Keywords:** DES; surface tension; in silico-based models; QSPR; validation; consensus modeling



**Citation:** Halder, A.K.; Haghbakhsh, R.; Voroshylova, I.V.; Duarte, A.R.C.; Cordeiro, M.N.D.S. Predicting the Surface Tension of Deep Eutectic Solvents: A Step Forward in the Use of Greener Solvents. *Molecules* **2022**, *27*, 4896. <https://doi.org/10.3390/molecules27154896>

Academic Editors: Mónia A. R. Martins and Nicolas Schaeffer

Received: 11 July 2022

Accepted: 28 July 2022

Published: 31 July 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The last two decades have witnessed a significant shift in the design and development of new chemicals for large-scale industrial applications. One of such efforts has been driven towards the replacement of flammable and environmentally hazardous substances with green and sustainable solvents. More benign solvents, even if dispensed in a large amount into the environment, are known to produce less harmful effects [1–3]. Deep eutectic solvents (DES) represent a class of “green solvents” with tremendous potential to replace conventional toxic chemicals. Indeed, apart from having a wide range of applications, DESs exhibit much less environmental toxicity, even when compared to their predecessor, ionic liquids (ILs) [4,5]. Thus, it is not surprising that the emergence of DESs at the beginning of this century drew considerable attention from the scientific community, as is confirmed by the growing number of publications related to DESs in the last two decades [4,5]. DESs may

simply be defined as the low melting point mixture of at least two compounds, one acting as a hydrogen bond acceptor (HBA) and another as a hydrogen bond donor (HBD) in a specific molar ratio [6,7]. Their low melting point, which is a result of complex hydrogen bonding interactions between the components of DESs, allows them to remain in the liquid phase at room temperature [8]. Besides being less eco-toxic in nature, DESs are generally easy to prepare, cost-effective and biocompatible [7–9]. However, even with many advantages, the suitability of any chemical for long-term industrial applications often depends on its fundamental physical properties, such as density, viscosity, surface tension, vapor pressure, the speed of sound etc. [4,5], and DESs are no exception. The physicochemical profile of a DES can readily be tailored by choosing different combinations of their starting components or by modifying their chemical structures [10,11]. Still, the number of possible combinations that can be envisaged to form DESs is extremely high. As such, without detailed knowledge of the relation between structure and properties, their fine-tuning is barely applicable in practice and often limited to a trial-and-error procedure.

Surface tension is one of the most crucial physical properties that must be considered, as it is required for the set-up of industrial processes, such as the design of heating systems, distillation columns and heat exchangers [12]. Therefore, the measurement of their surface tension is essential to assess the suitability of DESs for industrial applications. Normally, with increasing temperature, the surface tension of a DES decreases, but it is well-known that its components and their molar ratio are also responsible for their resulting surface tension [13].

Previously, we have reported a general thermodynamic model to estimate the surface tension for DESs of different nature [14]. The model has been developed with an up-to-date data bank containing surface tension values for a large number of structurally diverse DESs. The question that yet remains is whether a more predictive model for the surface tension of DESs can be achieved by alternative in silico-based modeling approaches such as the one proposed here, Quantitative Structure-Property Relationships (QSPR). In fact, the application of QSPR modeling techniques has long stood as particularly useful to estimate a wide range of properties of different materials [15–21]. Thus, many QSPR modeling studies have been devoted to different physicochemical properties of DESs. Very recently, for example, Wang et al. developed QSPR models based on Conductor-like Screening Model for Real Solvents (COSMO-RS) descriptors for characterizing the CO<sub>2</sub> solubility in DESs [22]. The authors found that the linear model was unable to successfully fit the whole dataset but a random forest non-linear model showed greater reliability, judging from the Absolute-Average-Relative-Deviation (AARD) value of 7.8% attained for that data (59 DESs). Balali and co-workers developed also QSPR models for probing the thiophene distribution between choline chloride (ChCl)-based DES and hydrocarbon phases in ternary systems [23]. The proposed linear models displayed good accuracy and included topological descriptors, which indicate the influence of the degree of the structure of HBDs on the thiophene distribution. In another study, Khajeh et al. employed QSPR modeling for predicting the melting and freezing points of DESs [24]. Their results showed that both properties of 181 DESs could be predicted with good accuracy ( $R^2 \sim 0.80$ ) by the derived linear models. Multiple attempts have been undertaken as well to set up linear and non-linear QSPR-based models for estimating the density and viscosity of DESs [25–27]. All the later models resorted to COSMO-RS descriptors and showed great predictivity performance ( $R^2 > 0.95$ ). However, these models have been based on a small number of data points pertaining to just certain families of DESs (e.g.: 49 hydrophilic [26] or 54 hydrophobic [27] DESs), which thus limits their general applicability.

The present work is encouraged by our very recent QSPR modeling efforts on the density of DESs [28], which yielded statistically more robust models than a thermodynamic model developed with the same dataset [29]. On one hand, such QSPR efforts demonstrated to offer more options and versatility for setting up predictive models as compared to thermodynamic modeling. On the other hand, QSPR models demand several statistical conditions be satisfied just as inspected here as per the guidelines of the Organization for

Economic Co-operation and Development (OECD) [30] to expand their overall applicability as well as statistical reliability [31]. Moreover, in order to address the requirement of robust validation strategies applicable for the binary mixtures, we have recently designed an open-source standalone Python-based tool “QSAR-Mx” (freely available to download at <https://github.com/ncordeirfcup/QSAR-Mx>, last accessed on 28 April 2022) [28]. This work extensively utilizes such a tool to set up the predictive QSPR models for probing the surface tensions of DESs. Therefore, the scope of the current work goes beyond the development of QSPR models by proposing and comparatively testing novel methodologies that can be utilized in the future towards reliably predicting the properties of binary mixtures.

## 2. Materials and Methods

### 2.1. Dataset Collection and Splitting

The dataset employed for the development of the QSPR models was adopted from our previously published work on DES surface tension estimation [14]. It contains 553 data points, compiled from 112 different binary and ternary DESs of diverse compositions. However, here we solely focused on binary DESs and therefore the current dataset was reduced to 530 data points pertaining to 99 different binary DESs. This dataset was combined with an additional set comprising 89 data points that were collected from measurements reported in the literature since 2020 [32–35] and that, thus, were not included in our previously thermodynamic model. The final dataset comprises 619 unique data points coming from 113 different types of binary DESs. It is worth noting that the experimental surface tension of each DES in the dataset was measured at atmospheric pressure in a large temperature range (278.15–358.15 K), rendering the temperature an important independent variable to consider in the QSPR modeling for understanding how it influences such physical property.

Predictive validation is a required but delicate task in any QSPR modeling—i.e., to assess model adequacy to new mixtures, and it is related directly to the dataset division scheme adopted. In fact, as shown by Muratov et al. [36], the random division of the original dataset for validation purposes is unacceptable since it can lead to unreliable QSPR models and to an over-optimistic estimation of their predictive performance. The authors have thus proposed and described in detail different validation strategies for the QSPR modeling of mixtures [36]. In this work, two such validation strategies were utilized to search for the most predictive QSPR models, namely: the mixtures-out (MO) and compounds-out (CO) schemes. Briefly, in the MO scheme, mixtures of the modeling set are distributed among the training and the test sets without repetition. By contrast, in the CO scheme, at least one chemical of the dataset is never placed in the training set. Naturally, these validation strategies are only applicable to binary mixtures and require some guidance to follow. Due to the complexity of the data matrices, any random MO- or CO-based division scheme may not yield the most predictive model since variables selection depends largely on the training set [28]. Even though the CO-based validation is considered to be the most robust strategy [36], it may give rise to underfitted models with poor statistical quality. At the same time, while the MO-based validation is less robust, this strategy definitely provides more meaningful solutions than any random data distributions or other validation division schemes such as the points-out one proposed by the same authors [28,36–38].

As referred to above, we have recently developed a Python-based tool named QSAR-Mx [28], specifically devised to address and automate some crucial steps involved in the QSPR modeling of binary mixtures. A detailed description of the functionalities of this tool can be found in its instruction manual (accessible from <https://github.com/ncordeirfcup/QSAR-Mx>, last accessed on 28 April 2022). Essentially, QSAR-Mx lets the user generate multiple MO- and CO-based data distributions and then develop models with the latter and select the most predictive ones based on their statistical metrics. Firstly, the user should choose two parameters—i.e., seed and interval, for generating the MO- and CO-based data distributions. In the MO division scheme, QSAR-Mx detects unique mixtures present in the dataset and then sorts them by the number of occurrences in descending order. From the

sorted list, the mixtures are grouped according to the seed (the starting point for selection) and interval values given. The unique mixtures selected are then incorporated into the test set. Likewise, in the CO scheme, the QSAR-Mx tool begins by sorting the unique chemicals that belong to component-1 of the mixtures, followed by sorting them in descending order and lastly, by choosing some chemicals based upon the maximum values of the seed and interval chosen. This procedure is then replicated for the unique chemicals belonging to component-2 of the mixtures. The unique chemicals selected are then placed in the test set. One should notice however here that the following QSPR models were always derived with generated data distributions in which the training set size was always greater than the test set size and, simultaneously, the size of the latter was at least 15% of the former. It should be also mentioned that the QSAR-Mx tool has been slightly modified since our previous work [28] because we found that the MO-based data distributions vary from one run to another. The new version of QSAR-Mx is now able to generate the same data distributions (i.e., MO-based training and test sets) every time, independently of the seed and interval given, leading thus to more reproducible modeling results.

In this work, to begin with, we divided the whole dataset into a modeling set and an external validation set (535 and 84 data points, respectively), using the CO-based division scheme with values for the seed and interval of 3 and 4, respectively. The modeling dataset was subsequently divided into training and test sets by MO- and CO-based schemes, setting both the maximum seed and maximum interval as 6. The DESs in the training set coming from the two schemes were employed separately for the development of the QSPR models, and those from the test sets were only used to test such models. The DESs in the external validation set were utilized for extra validation of the final most predictive QSPR models found. Details about the investigated DESs along with their experimental surface tension values, and corresponding references are given in Table S1 of the Supplementary Materials.

## 2.2. Mixture Descriptors

Due to the unique nature of binary DESs, the calculation of their descriptors requires additional steps to take into account the specificity of each component as well as the molar fractions. Here, we resorted to the strategy previously suggested by Oprisiu et al. [37], in which the descriptors are initially calculated for each component and then modified on the basis of ‘mixture descriptors weighted by molar fraction’ formulas. As such, two types of modified descriptors (from now on, referred to as WM descriptors), i.e.,  $D_{\text{pmix}}$  and  $D_{\text{nmix}}$ , were computed by the following formulas:

$$D_{\text{pmix}} = x_1 D_1 + x_2 D_2 \quad (1)$$

$$D_{\text{nmix}} = |x_1 D_1 - x_2 D_2| \quad (2)$$

where  $D_i$  stands for the descriptor of each component  $i$  ( $i = 1, 2$ ) and  $x_i$  for the respective molar fraction in the mixture.

Both formulas have already been successfully applied to generate predictive models for various properties of DESs [9,28,38] and implemented in the QSAR-Mx tool. Basically, QSAR-Mx includes two methods for calculating such mixture descriptors. Starting from the descriptors previously obtained for each component ( $D_1$  and  $D_2$ ), ‘Method-1’ calculates only the  $D_{\text{pmix}}$  descriptors whereas ‘Method-2’ provides both the  $D_{\text{pmix}}$  and  $D_{\text{nmix}}$  descriptors. In the present work, we used both of the aforementioned methods separately and then performed a comparative analysis to elucidate the method that provides the best solution as far as the predictivity of the overall model is concerned.

To start with, the 3D structures of each DES component were obtained by inputting the SMILES (Simplified Molecular Input Line Entry Specification) strings into the application MarvinView (<https://docs.chemaxon.com/display/docs/marvinview.md>, last accessed on 15 March 2022) and subsequently standardized by the ChemAxon Standardizer tool with the following options: strip salts, aromatize, neutralize and add explicit hydrogen atoms [39]. Here, we have resorted to the 0D-2D descriptors available in the Dragon software [40] for



describing each DES component. Actually, 3D descriptors were excluded due to the high computational effort required for structure optimization of each component, especially for large datasets, and the fact that those may also give rise to misleading information not ensuring reliable property prediction by 3D QSPR [28,41]. Finally, along with the WM descriptors, three independent variables were also included, namely: the measuring temperature,  $T$  (in K), the presence/absence of chlorine ions, and the presence/absence of bromine ions. The importance of temperature for the modeling was referred to before. Note in addition that only the cationic part was considered during the calculation of WM descriptors for the DES HBA component. Hence, two binary indicator variables (i.e., the presence (1) or absence (0) of halide ions) were required to be included to account for the anionic part of the HBA components.

### 2.3. Modeling Techniques and Evaluation

As to the modeling techniques, we started by opting for a regression-based approach like in our previous work [28] thanks to its easy interpretation but also high reliability. Specifically, the regression coefficients were obtained by the multiple linear regression analysis (MLR) implemented in our in-house QSAR-Mx tool and by selecting the variables through the sequential forward selection (SFS) algorithm using the Sequential Feature Selector module of Mlxtend (<http://rasbt.github.io/mlxtend/>) [42]. The following different conditions were applied for scoring the SFS selection:

- (a) determination coefficient ( $R^2$ ), no cross-validation;
- (b) negative mean absolute error (NMAE), no cross-validation;
- (c) negative mean Poisson deviance (NMPD), no cross-validation;
- (d) determination coefficient ( $R^2$ ), five-fold cross-validation (CV) or ten-fold CV.

Yet, a correlation cutoff of 0.95 and variance cutoff of 0.001 were always set to discard highly intercorrelated and near-constant descriptors. Additionally, the selection of the optimal number of descriptors for the MLR models was controlled by the %MAE<sub>LOO</sub> reduction policy, also implemented in QSAR-Mx. The %MAE<sub>LOO</sub> reduction scheme guarantees that no new descriptor is added to the model during feature selection if its inclusion does not reduce the leave-one-out (LOO) cross-validation and mean absolute error (MAE<sub>LOO</sub>) by at least 5% of the previous model. As such, this policy guarantees that the optimal number of descriptors is present in the model and, at the same time, that models generated with multiple model development strategies may be compared from a neutral ground [28].

To check if higher accuracy could be achieved when estimating the surface tension of DESs, non-linear models were also developed using five different machine learning (ML) techniques, i.e.: (i)  $k$ -Nearest Neighbors ( $k$ -NN), (ii) Random Forests (RF), (iii) Support Vector Machines (SVM), (iv) Neural Network Multilayer Perceptron (NN-MLP), and (v) Gradient boosting (GB) [43–47]. Such ML-based models were set up by resorting to the tools available in the Scikit-learn programs (<https://scikit-learn.org/stable/>) with QSAR-Mx (last accessed on 28 April 2022) and for each of them, hyperparameter tuning was performed by varying their crucial parameters (see the list in Table S2). The best parameters for a given ML estimator were determined by a 5-fold cross-validation scheme using the same training sets as before. In the same manner, the external predictivity of the promising non-linear models found was firstly accessed and further validated using the same test and external sets.

As described above, QSAR-Mx generates multiple models based on three types of inputs provided by the user, namely: (i) descriptor calculation strategy (Method-1 or -2), (ii) dataset division schemes (MO- or CO-based data-division), and (iii) scoring conditions. No matter the model generation strategy followed, any QSPR regression model requires an evaluation with robust diagnostic tools to assess and compare its acceptability as well as quality over other models.

In this work, the internal predictivity of the developed QSPR regression models was primarily checked by statistical parameters, such as the MAE<sub>LOO</sub> and  $Q^2_{LOO}$  (LOO cross-validation  $R^2$ ) [48,49]. Keeping in mind the importance of the compounds-out validation,

we have recently introduced two new statistical parameters based on the so-called leave-chemical-out (LCO) cross-validation, which is conceptually similar to the well-known leave-many-out CV but more effective whenever dealing with mixtures [28]. These new parameters, i.e.,  $Q^2_{\text{LCO}}$  and  $\text{MAE}_{\text{LCO}}$ , are mostly important for the MO-based data distributions. Indeed, even though the latter often produce more predictive models than the CO-based data distributions, their predictivity remains questionable due to the lack of CO-based validation. Both  $Q^2_{\text{LCO}}$  and  $\text{MAE}_{\text{LCO}}$  have actually helped us monitor the model performance upon the removal of each component (belonging either to HBA or to HBD) one by one from the training set with further model redevelopment using the remaining components. A detailed description of these two parameters can be found in our previous work [28]. Importantly, while assessing the quality of the models, the difference between  $Q^2_{\text{LOO}}$  and  $Q^2_{\text{LCO}}$  should also be evaluated. In fact, a large discrepancy between the values of the latter suggests that the mixtures based on one or more components are not predicted well enough by the QSPR model. Besides the above-mentioned statistics, the internal predictivity of the final regression models was also evaluated by using scaled  $r_m^2$  validation metrics, such as  $r_m^2(\text{LOO})$  and  $\Delta r_m^2(\text{LOO})$  [50]. Basically,  $r_m^2$  metrics are based on the correlation between the observed and predicted values, with and without setting to zero the intercept of the least square regression lines. In addition, the AARD calculated for each data distribution was also used for checking the overall errors of the derived models. Although not quite common in QSPR modeling, the latter allowed us to compare the quality of our QSPR models with that of previously reported thermodynamic models [29]. To access the external predictive ability of the models, similar statistical validation metrics were also employed, i.e., the mean absolute error for the test set or external validation set ( $\text{MAE}_{\text{test}}$  and  $\text{MAE}_{\text{ext}}$ ) and the variance explained for external prediction ( $R^2_{\text{Pred}}$ ) [44] along with the scaled  $r_m^2(\text{test})$ ,  $\Delta r_m^2(\text{test})$ ,  $r_m^2(\text{ext})$ , and  $\Delta r_m^2(\text{ext})$  metrics [50].

Other aspects that deserve special attention are the absence of highly collinear descriptors and the lack of chance correlations in the final derived models. Highly collinear variables were simply checked by inspecting the cross-correlation matrix of the models' descriptors. On the other hand, the  $Y$ -randomization technique identifies models with chance correlations, using the  $cR_p^2$  parameter [51], after the sequence of the response vector has been randomly modified. Here, the procedure was repeated 1000 times, and new models were developed with the randomly reordered responses employing the same set of variables. The uniqueness of the final regression model and its lack of chance correlations is confirmed by the value obtained for  $cR_p^2$ , which should be closer to one [51].

Finally, apart from inspecting the models' robustness and predictivity, one should also define their applicability domain (AD), that is, the response and chemical structure space for which the models form reliable predictions without extrapolating. In this work, the AD of the developed models was determined by the leverage approach [52], which renders a measure of the similarity of a particular substance from all other substances (distance between its descriptor values and the average for all descriptor values). So, one can plot the standardized residuals against the leverage values for each DES of the several sets. From such a plot, the so-called William's plot [52], we were able to identify the response and structural DES outliers. All plots shown in the present work were conceived with Matplotlib [53].

#### 2.4. Consensus Modeling

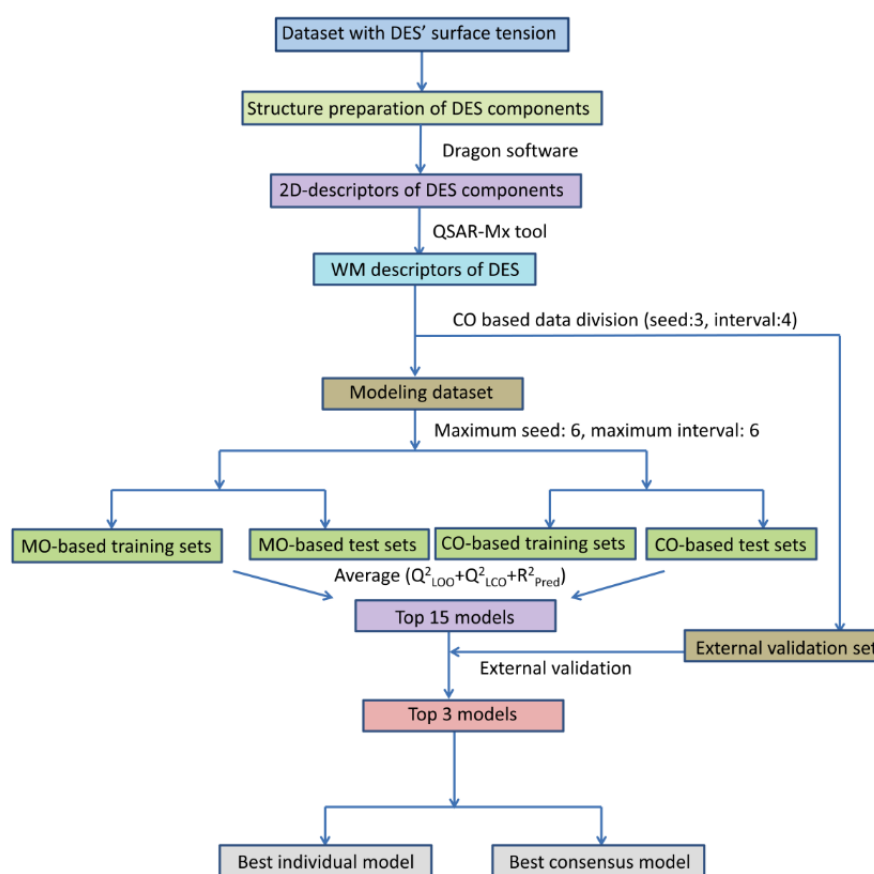
The task here is to explore whether the overall quality of predictions for new substances might be improved by an "Intelligent" selection of multiple models. Towards that end, the most predictive QSPR models derived were subjected to consensus modeling, using the software tool Intelligent Consensus Predictor (freely available through the web <https://dtclab.webs.com/software-tools>, last accessed on 23 March 2022) developed by Roy et al. [54]. The four strategic techniques of this tool were applied, namely: Consensus Models (CM) 0–3, just as in our previous work [28], and as fully described in the work by the authors [54]. In short, CM0 is the simplest strategy and consists in computing

the arithmetic average of predicted response values from all input individual models. In contrast, CM1 is based on the simple arithmetic average of predictions from all qualified individual models. CM2 corresponds to weighted average predictions from all qualified models, formerly giving appropriate weightage to those models. Finally, CM3 applies compound-wise predictions based on the best selection coming from the qualified models. Independently of the consensus modeling methodology, our main purpose was thus to combine multiple statistically robust models to improve the predictivity over the external validation set.

### 3. Results

#### 3.1. Model Calibration and Evaluation

Figure 1 depicts the workflow chart followed in the present QSPR modeling, which was mostly carried out using our recently developed tool, QSAR-Mx. As can be seen, all the involved steps and methodology employed to cope with the major goal of this work are shown, i.e., to build reliable predictive QSPR regression models from the compiled data that could be used to estimate the surface tension of DESs.



**Figure 1.** Basic workflow chart for the QSPR regression modeling, followed in this study.

In total, 258 models were set up by varying data splitting schemes, descriptor calculation methods (Method-1 or Method-2) and SFS-MLR modeling. Among these, 136 models pertained to the MO-based division scheme, whereas the remaining 112 models were generated with the CO-based division scheme. The overall predictive quality of each of these regression models was judged by means of the average value computed for the statistical parameters  $Q^2_{LOO}$ ,  $Q^2_{LCO}$  and  $R^2_{Pred}$ . Essentially, the two parameters— $Q^2_{LOO}$  and  $R^2_{Pred}$ —account for the internal and external predictivity of the QSPR models, respectively. Nevertheless, the parameter  $Q^2_{LCO}$  was also included to ensure that the most predictive models do not suffer overfitting due to bias towards some specific components of the binary

DES mixtures. Naturally, the higher the average value obtained from these three parameters is, the more predictive the model is. Considering this, we selected the top 15 unique models for further processing. A summary of the statistical results of these models is given in Table 1. Interestingly, out of these 15 models, 14 were derived from MO-based data distributions, and only one model arose from CO-based distributions. Undoubtedly, that clearly shows that MO-based data distributions are more likely to produce more predictive models in comparison to the CO-based data distributions since the latter provides a more rigorous validation strategy.

As referred to before, in the entire model-building process, the data distributions were varied for the selection of the most predictive models. Therefore, the generated test sets serve as a validation set to estimate the external predictivity of the models but, at the same time, as calibration sets for the selection of the best models. In contrast, the external validation sets (containing 84 data points) were treated as the ‘true validation set’ for assessing the external predictivity of the models. The latter was built with the CO-based data-distribution scheme, thus holding a significant challenge to the generated models as far as their external predictivity is concerned. A comparison of the predictivity of the top 15 models is shown in Table 2.

As can be clearly observed from Table 2, only a few models show satisfactory predictions against the external validation set. Nevertheless, six of these models had  $R^2_{\text{Pred}}$  values greater than 0.50, as well as average %AARD values lower than 12. Moreover, three models, namely M09, M10 and M12, supplied the most satisfactory predictivity towards such external validation set with  $R^2_{\text{Pred}} > 0.65$  and %AARD < 10. Therefore, these three models were considered the best models obtained for predicting the surface tension of DESs. Remarkably, M10, the only CO-based model included in the top 15, emerged as one of the most predictive models. Still, on the basis of overall predictivity, M12 was selected as the best individual QSPR model, even taking into account its slightly lower internal predictivity, compared to M10, and its slightly lower external predictivity towards the test set, as compared to M09. Even so, M12 afforded a balanced prediction against all three sets with an average %AARD value of 7.126, which is lower than that obtained for the other two models. At the same time, model M12 provides the best solution if the average value of  $Q^2_{\text{LOO}}$ ,  $Q^2_{\text{LCO}}$  and  $R^2_{\text{Pred}}$  (against the two validation sets) is considered. In fact, for M12, this average value was found to be 0.859, while for M09 and M10, the average values were estimated as 0.820 and 0.831, respectively.

In summary, the best predictive model found for the DESs’ surface tension (a six-variable equation, model M12) can be expressed as detailed below, while the meaning of the selected WM descriptors is given in Table 3.

$$\begin{aligned} \sigma = & +89.611 (\pm 3.452) \\ & +0.405 (\pm 0.026) P\_VSA\_MR\_6\text{pmix} \\ & -5.034 (\pm 0.874) Eig02\_EA(\text{dm})\text{pmix} \\ & -23.145 (\pm 3.320) CATS2D\_02\_AN\text{pmix} \\ & +8.835 (\pm 0.174) BLTF96\text{pmix} \\ & -25.191 (\pm 2.352) MATS5sn\text{mix} \\ & -0.104 (\pm 0.011) T \end{aligned} \quad (1)$$



**Table 1.** Statistical results of the top 15 unique QSPR regression models generated.

Model	Seed; Interval	Descriptor <sup>a</sup>	Split <sup>b</sup>	Scoring <sup>c</sup>	$N_{tr}$ <sup>d</sup>	$Q^2_{LOO}$ <sup>e</sup>	$Q^2_{LCO}$ <sup>f</sup>	$MAE_{LOO}$ <sup>g</sup>	$N_{ts}$ <sup>h</sup>	$R^2_{Pred}$ <sup>i</sup>	$MAE_{test}$ <sup>j</sup>	Avg <sup>k</sup>
M01	2; 5	Method-1	MO	NMAE	408	0.884	0.854	2.586	127	0.907	3.863	0.882
M02	4; 5	Method-2	MO	NMAE	435	0.873	0.849	2.658	100	0.899	3.966	0.874
M03	1; 4	Method-2	MO	NMAE	408	0.898	0.854	2.775	127	0.865	2.569	0.872
M04	5; 4	Method-2	MO	NMAE	409	0.898	0.855	2.771	126	0.862	2.584	0.872
M05	4; 5	Method-1	MO	NMAE	435	0.871	0.839	2.635	100	0.898	4.039	0.869
M06	3; 5	Method-1	MO	NMAE	443	0.881	0.849	2.671	92	0.871	4.073	0.867
M07	1; 3	Method-1	MO	NMAE	359	0.901	0.862	2.369	176	0.836	4.223	0.866
M08	4; 5	Method-1	MO	$R^2$	435	0.883	0.858	2.854	100	0.855	4.695	0.865
M09	4; 3	Method-1	MO	NMAE	360	0.906	0.854	2.322	175	0.83	4.389	0.864
M10	1; 2	Method-2	CO	$R^2$	301	0.931	0.903	1.660	234	0.754	6.030	0.862
M11	4; 5	Method-2	MO	5-fold	435	0.865	0.841	3.000	100	0.876	4.282	0.861
M12	4; 3	Method-2	MO	$R^2$	360	0.908	0.882	2.608	175	0.783	5.134	0.858
M13	4; 5	Method-1	MO	5-fold	435	0.871	0.845	2.706	100	0.857	4.626	0.858
M14	1; 4	Method-1	MO	10-fold	408	0.869	0.849	3.340	127	0.847	2.050	0.855
M15	5; 4	Method-1	MO	10-fold	409	0.869	0.85	3.336	126	0.844	2.052	0.855

<sup>a</sup> Descriptor calculation method used. <sup>b</sup> Data splitting scheme utilized. <sup>c</sup> Scoring condition applied. <sup>d</sup> Number of data points in the training set. <sup>e</sup> Leave-one-out cross-validation determination coefficient. <sup>f</sup> Leave-chemical-out cross-validation determination coefficient. <sup>g</sup> LOO cross-validation mean absolute error. <sup>h</sup> Number of data points in the test set. <sup>i</sup> Variance explained for external prediction. <sup>j</sup> Mean absolute error of the test set. <sup>k</sup> Average value of  $Q^2_{LOO}$ ,  $Q^2_{LCO}$  and  $R^2_{Pred}$ .

**Table 2.** Internal and external predictivity for the top 15 regression models against the training, test and external validation sets <sup>a</sup>.

Model <sup>b</sup>	Training Set				Test Set			External Validation Set		
	$N_{tr}$	$Q^2_{LOO}$	$Q^2_{LCO}$	%AARD	$N_{ts}$	$R^2_{Pred}$	%AARD	$N_{ex}$ <sup>c</sup>	$R^2_{Pred}$	%AARD
M01	408	0.884	0.854	5.541	127	0.907	11.843	84	−0.335	15.931
M02	435	0.873	0.849	6.11	100	0.899	7.517	84	0.464	12.176
M03	408	0.898	0.854	6.063	127	0.865	5.418	84	−0.196	15.833
M04	409	0.898	0.855	6.057	126	0.862	5.446	84	−0.19	15.818
M05	435	0.871	0.839	5.965	100	0.898	7.538	84	0.392	11.838
M06	443	0.881	0.849	6.052	92	0.871	7.65	84	0.516	11.021
M07	359	0.901	0.862	5.222	176	0.836	9.331	84	−7.225	27.7
M08	435	0.883	0.858	6.6	100	0.855	8.456	84	0.466	11.45
<b>M09</b>	<b>360</b>	<b>0.906</b>	<b>0.854</b>	<b>5.202</b>	<b>175</b>	<b>0.83</b>	<b>9.872</b>	<b>84</b>	<b>0.688</b>	<b>8.527</b>
<b>M10</b>	<b>301</b>	<b>0.931</b>	<b>0.903</b>	<b>4.208</b>	<b>234</b>	<b>0.754</b>	<b>12.754</b>	<b>84</b>	<b>0.734</b>	<b>7.777</b>
M11	435	0.865	0.841	6.875	100	0.876	7.78	84	0.568	10.047
<b>M12</b>	<b>360</b>	<b>0.908</b>	<b>0.882</b>	<b>5.805</b>	<b>175</b>	<b>0.783</b>	<b>11.155</b>	<b>84</b>	<b>0.862</b>	<b>4.418</b>
M13	435	0.871	0.845	6.204	100	0.857	8.442	84	0.544	11.093
M14	408	0.869	0.849	7.344	127	0.847	3.943	84	0.352	11.642
M15	409	0.869	0.85	7.339	126	0.844	3.929	84	0.353	11.638

<sup>a</sup> For the meaning of  $N_{tr}$ ,  $N_{ts}$ ,  $Q^2_{LOO}$ ,  $Q^2_{LCO}$ ,  $R^2_{Pred}$  and %AARD, please check the footnotes of Table 1. <sup>b</sup> The more predictive models are marked in bold. <sup>c</sup> Number of data points in the external validation set.

**Table 3.** The five WM molecular descriptors selected for model M12—Equation (3).

Symbol	Definition [55–58]	Class
P_VSA_MR_6 <sub>pmix</sub>	P_VSA-like on Molar Refractivity, at bin size 6	P_VSA-like descriptor <sup>a</sup> ( $D_{pmix}$ type)
Eig02_EA(dm) <sub>pmix</sub>	eigenvalue n. 2 from edge adjacency matrix, weighted by dipole moment	Edge adjacency indices ( $D_{pmix}$ type)
CATS2D_02_AN <sub>pmix</sub>	CATS2D Acceptor-Negative at lag 2	2D CATS <sup>b</sup> ( $D_{pmix}$ type)
BLTF96 <sub>pmix</sub>	Verhaar Fish base-line toxicity from MLOGP (mmol/L)	Molecular properties ( $D_{pmix}$ type)
MATS5S <sub>nmix</sub>	Moran autocorrelation of lag 5, weighted by I-state <sup>c</sup>	2D autocorrelations ( $D_{nmix}$ type)

<sup>a</sup> P\_VSA-like descriptors stand for the van der Waals surface area (VSA) with a particular property (P), in this case, the molar refractivity (MR) [57]. <sup>b</sup> Chemically Advanced Template Search (CATS) descriptors expressly designed to identify scaffold hops [58]. <sup>c</sup> I-states are based on the Kier-Hall atomic electronegativity modified by the number of  $\sigma$  bonds, number of hydrogen atoms, number of electrons in  $\pi$  orbitals, and number of lone pair electrons [55,56].

In this equation,  $X_{pmix}$  and  $X_{nmix}$  stand for WM descriptors of the type  $D_{pmix}$  in line with Equation (1) and  $D_{nmix}$  following Equation (2), respectively,  $T$  is the temperature (in K) under which the surface tension has been measured, and  $\sigma$  is the surface tension (in mN/m).

A summary of the extended statistical results for model M12 is given in Table 4. The determination coefficient values ( $R^2 = 0.916$  and  $R^2_{Adj} = 0.915$ ), the sample size ( $N_{tr} = 360$ ), the Fisher ratio ( $F = 642.4$ ), but especially the high ratio between the number of data points to adjustable variables ( $\rho = 60$ ) [59] are indicative of the model's statistical significance and fitness. Model M12 also provides a satisfactory internal and external predictivity as follows from the cross-validation,  $r_m^2$  and  $R^2_{Pred}$  metrics values (see Table 4). Moreover, built with only six descriptors, it led to %AARD values of 5.805, 11.155 and 4.418 against the training, test and validation sets, respectively. The model prediction ability was further checked by analyzing the relative deviations ( $\%RD = 100 \cdot (\sigma_{Pred} - \sigma_{exp}) / \sigma_{exp}$ ) between the predicted and experimental DES surface tension values for all three sets. As Figure 2 shows, model M12 performs more accurately regarding the training and external validation sets than the test set. Yet, the latter also demonstrates a normal behavior considering the shape of the RD distribution according to the proposed model, also displayed in Figure 2. This histogram plot clearly depicts that most of the RD error values are within  $\pm 20\%$  and that those are normally distributed, suggesting that the model estimations are not biased.

**Table 4.** MLR statistical results for model M12—Equation (3) <sup>a</sup>.

Training Set	Test Set	External Set
$N_{tr} = 360$ ; $R^2 = 0.916$ ; $R^2_{Adj} = 0.915$ ; $F(6353) = 642.4$ ; $Q^2_{LOO} = 0.908$ ; $MAE_{LOO} = 2.608$ ; $Q^2_{LCO} = 0.882$ ; $MAE_{LCO} = 3.122$ ; $r_m^2(LOO) = 0.869$ , $\Delta r_m^2(LOO) = 0.066$ ; $\%AARD = 5.805$ ; ${}^cR_p^2(1000 \text{ runs}) = 0.908$	$N_{ts} = 175$ ; $R^2_{Pred} = 0.783$ ; $MAE = 5.134$ ; $r_m^2(test) = 0.573$ , $\Delta r_m^2(test) = 0.197$ ; $\%AARD = 11.155$	$N_{ex} = 84$ ; $R^2_{Pred} = 0.862$ ; $MAE = 1.777$ ; $r_m^2(ext) = 0.767$ , $\Delta r_m^2(ext) = 0.097$ ; $\%AARD = 4.418$

<sup>a</sup>  $R^2$ : Determination coefficient;  $R^2_{Adj}$ : Adjusted  $R^2$ ;  $F(6,353)$ : Fisher's statistic;  $MAE_{LOO}$  and  $MAE_{LCO}$ : Leave-one-out and leave-chemicals-out cross-validation mean absolute error, respectively;  $r_m^2(LOO)$  and  $\Delta r_m^2(LOO)$ : LOO cross-validation  $r_m^2$  and its associated deviation, respectively;  $r_m^2(test)$  and  $\Delta r_m^2(test)$ :  $r_m^2$  of the test set and its associated deviation, respectively;  $r_m^2(ext)$  and  $\Delta r_m^2(ext)$ :  $r_m^2$  of the external test set and its associated deviation, respectively. For the meaning of  $N_{tr}$ ,  $N_{ts}$ ,  $N_{ex}$ ,  $Q^2_{LOO}$ ,  $Q^2_{LCO}$ ,  $R^2_{Pred}$ , and  $\%AARD$ , check the footnotes of Tables 1 and 2.

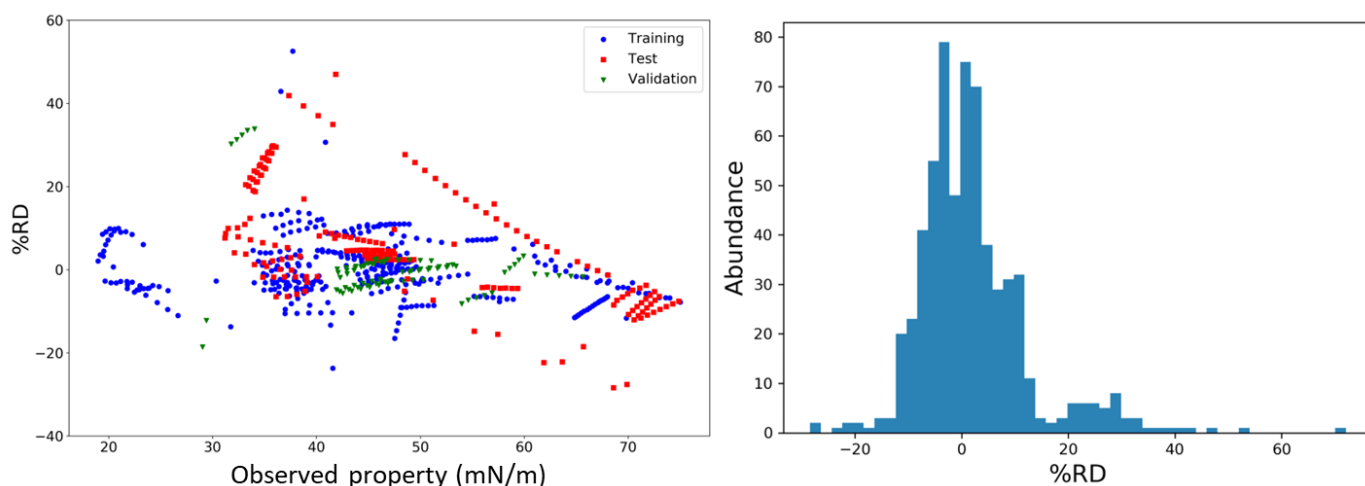
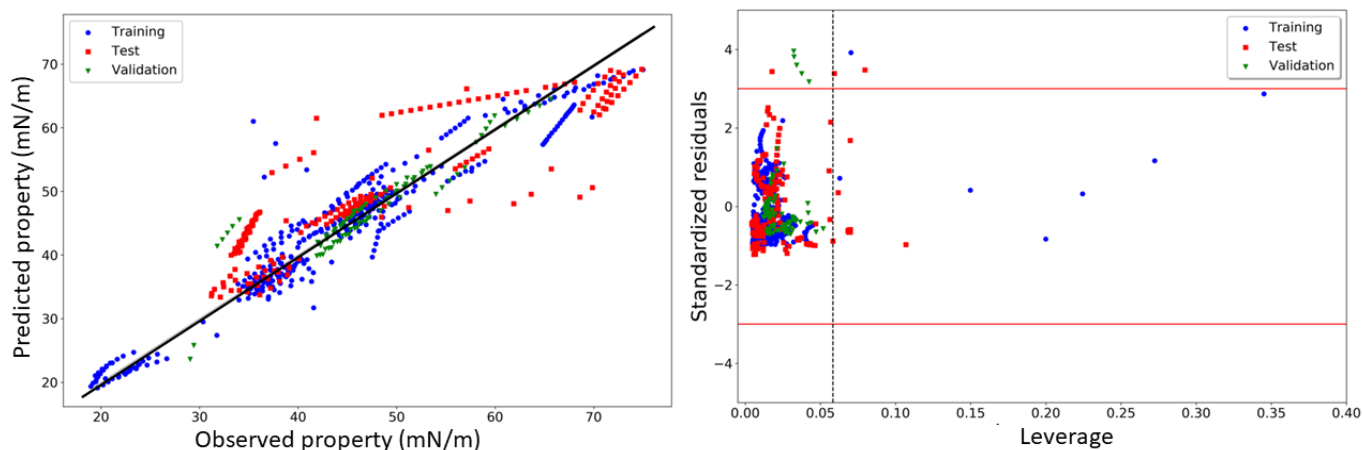
**Figure 2.** Relative deviations (%RD) between the predicted and observed DES surface tensions (**left**) and histogram plot of the distribution of %RD values (**right**).

Figure 3 shows the plot of the predicted surface tensions obtained from model M12 vs. the observed experimental ones. As seen, the majority of the data points are sufficiently close to the diagonal line, denoting the model's reliability and soundness of its predictions. Indeed, the model's performance is even better than that of the previously developed thermodynamic model for the DES surface tension [14], which, despite having fewer data points (a total of 530 data points, considering only the 99 unique binary DES), led to  $\%AARD$  values of 8.87 and 14.81 for the training and test sets, respectively. However, the purpose and outcomes of the current QSPR modeling are different from that of any thermodynamic model, as the former demands several different conditions to be satisfied, apart from validation, to establish the statistical robustness of the model. For example, so far, we have demonstrated the acceptable results on the reliability of the QSPR model M12, but it is also important to inspect the non-intercollinearity among any two of its descriptors. The latter was found to be 0.238, indicating that the variables included in the model are indeed not interrelated to each other. Furthermore, the model was itself checked for its uniqueness by the  $Y$ -based randomization technique, which was performed by scrambling the endpoint responses for the training set. The high value obtained for  $cR_p^2$  (=0.908) implies that the model is not correlated by chance. Another crucial aspect is related to the applicability domain of the model that here was assessed by analyzing the Williams plot (plot of standardized residual vs. leverages). As seen in Figure 3, eight data points from the training set and thirteen from the test set can be considered structural outliers of the model, but no structural outliers were found in the external validation set. Interestingly,

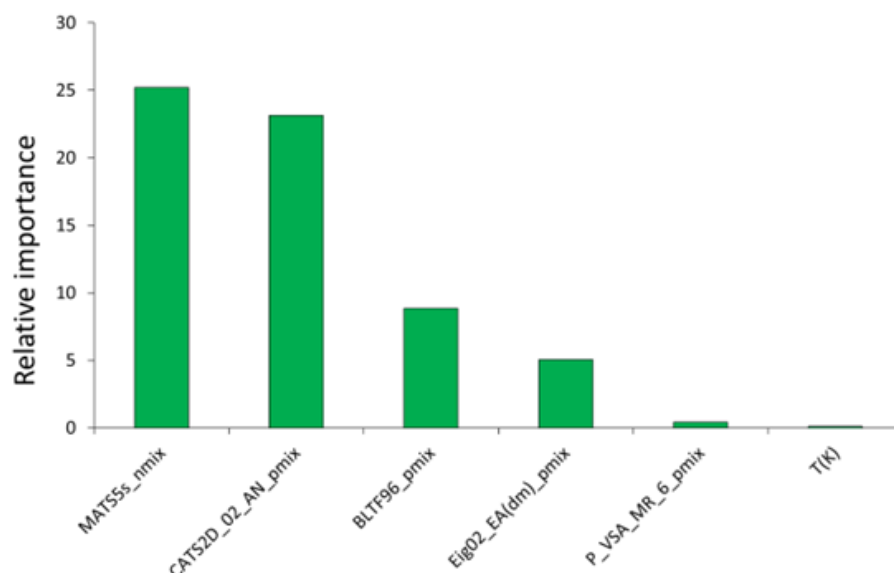
most of these structural outliers were well predicted by the model and were thus retained, as previously suggested by Gramatica et al. [49]. In addition, only twelve data points of the entire dataset were found to be response outliers, which also proves the high predictive accuracy of the model [60].



**Figure 3.** Predicted surface tension values vs. observed experimental ones (left) and Williams plot (right) obtained for model M12.

### 3.2. Model Interpretation

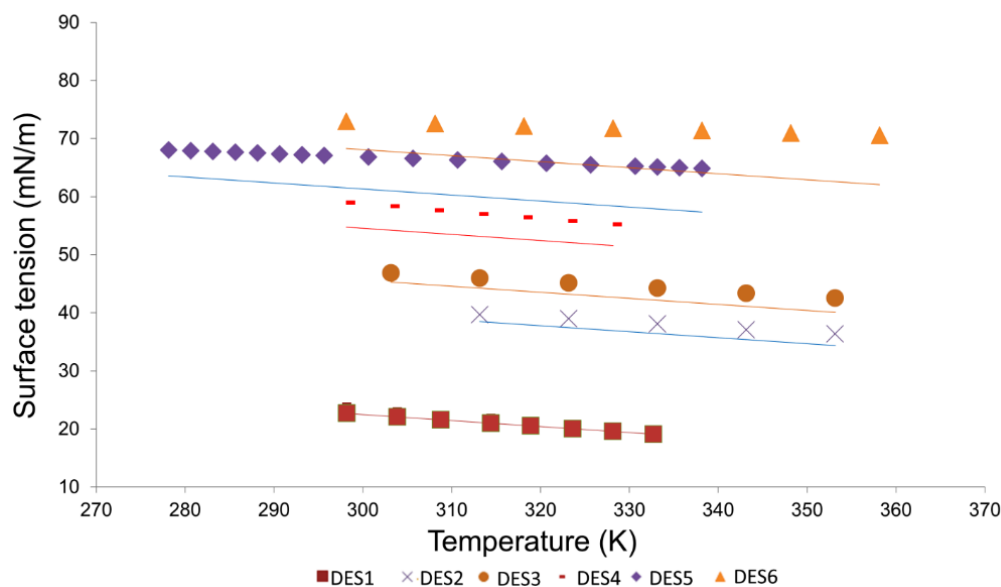
In our previous investigation on density [28] we observed that, in spite of providing less mechanistic interpretability, graph-based topological descriptors often help in characterizing the physicochemical properties. In the present work, a number of topological descriptors were also proven to be significant for describing the surface tension of DESs. Figure 4 shows the relative importance of each descriptor of model M12, estimated on the basis of the absolute value of its regression coefficients.



**Figure 4.** Relative importance of the descriptors found in the best individual model M12.

As can be observed, the WM descriptor MATS5s<sub>nmix</sub> was found to have the highest importance and besides, it is the only  $D_{nmix}$  type descriptor in the model. Being derived from graph-based topological descriptors, MATS5d<sub>nmix</sub> points out that the differences in topological geometry of the DESs' components may play a significant role in the surface tension of these solvents. The  $D_{pmix}$  type WM descriptor CATS2D\_02\_AN<sub>pmix</sub> is the second

most influencing descriptor of the model. Chemically Advanced Template Search (CATS) descriptors are a useful group of descriptors that account for the topological distance among scaffold features in the molecules [58]. CATS2D\_02\_AN, in particular, means that the acceptor and negatively charged groups are separated by a small topological distance (=2). In this case, higher values of this descriptor are found to be negatively correlated to the surface tension. Descriptor BLTF96<sub>pmix</sub> appears as the third most important descriptor in the model. Unlike the first two descriptors of topological nature, this descriptor is based on an important molecular property—lipophilicity [55,56]. Since this descriptor belongs to the  $D_{pmix}$  type, it may be inferred that higher lipophilicity of the components would trigger higher surface tension for the DESs. Apart from lipophilicity, another well-known physicochemical property—dipole moment—was also found to have important contributions in ascertaining the DES surface tension. The importance of the dipole moment is derived from the presence of descriptor Eig02\_EA(dm)<sub>pmix</sub>. The fifth most important descriptor belongs to the class of P\_VSA descriptors, which represent the amount of van der Waals surface area (VSA) having a property (P) in a certain range [56]. In the case of the descriptor P\_VSA\_MR\_6<sub>pmix</sub>, the property is the molar refractivity (MR) at a larger range (bin size 6). The positive relation of P\_VSA\_MR\_6<sub>pmix</sub> with the dependent property is highly significant as it suggests that increased MR (i.e., polarizability) within the van der Waals surface of each component contributes towards a higher surface tension for the respective DESs. Finally, the last descriptor of the model is the temperature of surface tension measurements,  $T$ . As expected, with increasing temperature the surface tension is found to decrease, which fits well with the experimental findings. Still, to further check how model M12 actually addresses the influence of temperature, we randomly selected six DESs with a range of surface tension values. From Figure 5, it can be clearly seen that both experimental and predicted properties followed the same trend, i.e., the surface tension gradually decreases as the temperature is increased.



**Figure 5.** Comparison of surface tension calculated by the M12 model with literature data in the temperature range from 278.15–358.15 K for six DESs at atmospheric pressure. DES1: DL-menthol and octanoic acid (3:1); DES2: tetrabutylammonium chloride and arginine (8:1); DES3: tetraprpylammonium bromide and ethylene glycol (1:6); DES4: N,N-diethylethanolammonium chloride and glycerol (1:5); DES5: choline chloride and glycerol (1:5); DES6: choline chloride and D-glucose (1:1).



### 3.3. Non-Linear Models

Albeit M12 emerged as the most accurate linear model in estimating the DES surface tension, the question that still remains is whether a non-linear model based on its descriptors might have better performance. Table 5 shows a statistical summary of the performance of the non-linear models resulting from applying the five different machine learning techniques, i.e.: *k*-NN, RF, SVM, NN-MLP and GB [49–53]. It can be observed that most of these ML techniques failed to produce predictive models and their results are not accurate either. Still, the SVM technique yields a predictive model and thus has the highest performance among the other ML techniques, though both the internal and the external predictivity of the latter remain inferior to the linear model. These results indicate that for the selected set of descriptors, the multiple linear regression-based model has the best accuracy in estimating the surface tension of DESs as well as sufficient predictivity that cannot be achieved with other model development techniques.

**Table 5.** Summary of the statistical parameters obtained from non-linear models based on different machine learning methods.

Method <sup>a</sup>	Training Set ( $Q^2_{5\text{-fold}}$ )	Test Set ( $R^2_{\text{Pred}}$ )	External Set ( $R^2_{\text{Pred}}$ )
<i>k</i> -NN	0.176	0.597	not determined
RF	0.473	0.746	not determined
SVM	0.874	0.774	0.767
MLP	0.541	0.269	not determined
GB	0.453	0.471	not determined

<sup>a</sup> *k*-NN: *k*-Nearest Neighbors; RF: Random Forests; SVM: Support Vector Machines; NN-MLP: Neural Network Multilayer Perceptron; GB: Gradient boosting.

### 3.4. Consensus Modeling

Finally, we applied the intelligent consensus modeling [54] to see whether the surface tension predictions for the external validation set could be improved. To do so, sets of the three most predictive linear models—M09, M10 and M12—were subjected to consensus predictions in different combinations, namely: (a) C1-based using models M09, M10 and M12; (b) C2-based using models M10 and M12; (c) C3-based using models M09 and M10; and (d) C4-based with models M09 and M12. In each case, the modeling dataset containing 535 data points was treated as the training set whereas the external validation set was used to check the external predictivity of the consensus model. The results of all consensus modeling attempts are presented in Table 6.

**Table 6.** External predictivity of the best individual model M12 and consensus models (C1–C4) built with different combinations of the top three models (M09, M10 and M12).

Consensus Models	Models	CM <sup>a</sup>	$R^2_{\text{Pred}}$ <sup>b</sup>	$r_m^2(\text{test})$ <sup>c</sup>	MAE <sub>test</sub> <sup>d</sup>	%AARD <sup>e</sup>
C1	M09, M10, M12	0	0.864	0.801	1.869	4.459
C2	M10 and M12	2	0.823	0.812	2.089	4.732
C3	M09 and M10	2	0.853	0.787	1.979	4.393
C4	M09 and M12	None	—	—	—	—
M12	—	—	0.862	0.767	1.777	4.418

<sup>a</sup> Method of Intelligent consensus prediction that yielded the best external validation result. <sup>b</sup> Variance explained for the external prediction. <sup>c</sup> Metric  $r_m^2$  for the test set. <sup>d</sup> Mean absolute error for the test set. <sup>e</sup> Absolute average relative deviation.

Interestingly, the resulting models C1 and C4 lead to similar predictivities. Yet, none of the later consensus models display an external predictivity considerably better than that of the best individual model, M12. The  $R^2_{\text{Pred}}$  and %AARD values obtained for consensus model C1 are 0.864 and 4.459, respectively, and similarly for C3 (i.e., 0.854 and 4.393). As can be seen, both C1 and C3 may therefore be projected as alternative models to M12.

However, let us mainly focus on C3, since it reveals that M09 and M10 may indeed work as complementary models for each other towards improving the external predictivity. Details about the M09 and M10 models are provided in Table S3 of the SI.

Since M09 was developed with the same data distribution as M12, these two models have four descriptors in common, namely: CATS2D\_02\_AN<sub>pmix</sub>, P\_VSA\_MR\_6<sub>pmix</sub>, BLTF96<sub>pmix</sub>, and *T*. Obviously, these four descriptors have a high significance in describing the surface tension of DESs. Significantly, CATS2D\_02\_AN<sub>pmix</sub>, which was found to be the second most important descriptor of M12, appears to be the most influential descriptor of M09. It undoubtedly indicates that this descriptor may be considered the most crucial descriptor in predicting the surface tension of DESs. Presumably, due to the similarity between M12 and M09, consensus modeling with these two models failed to provide any better solution. Model M10, in contrast, is established as a unique model because, save for *T*, none of its descriptors is found either in M09 or in M12. Most likely due to this reason, its combination with the other two models produces good consensus models. Unlike models M09 and M12, model M10 yields are slightly higher but still have acceptable intercollinearity between descriptors, with a maximum  $R^2$  value of 0.713. The selected descriptors for models M09 and M10 are described in detail in Table S4 of the SI.

#### 4. Conclusions

The present work aimed to establish a systematic and well-designed QSPR modeling for predicting the surface tension of a wide range of DESs, following the OECD guidelines. Towards such aim, the largest surface tension data bank of binary DESs known to date, comprising 619 data points from 113 unique DESs of various families was gathered from the literature. In addition, special emphasis was put on employing robust validation strategies for setting up the QSPR models. In so doing, the best QSPR models were set up with multiple data distributions resulting from MO- and CO-data splitting schemes along with a weighted mixture type of descriptors, using our in-house open access tool QSAR-Mx. After considering several statistical parameters, the top three individual linear regression models stand out for their accuracy and robustness. The most predictive individual linear model was however selected based on the predictivity towards the external validation set.

Similar to our previous study on the development of a thermodynamic model [14], the surface tension of DESs was found to be a particularly difficult property to predict. This may be related to the challenging nature of the accurate experimental estimation of surface tensions, associated with the presence of surface-active impurities and differences in the measuring protocols [33]. Nevertheless, our most predictive individual QSPR regression model (i.e., model M12) yielded a satisfactory overall %AARD value (=7.126), especially when compared to the aforementioned thermodynamic model (%AARD = 10.31, considering only the binary DESs therein) [14]. This model depicted also the structural and physicochemical features related to the surface tension of DESs. Just as in our previously developed model for the density of DESs [28], graph-based topological descriptors were found to be highly useful in this respect. Some physicochemical factors, such as the lipophilicity, polarizability and dipole moment of the DESs' components, were found to be responsible for ruling their surface tensions. We also attempted to generate consensus models based on the top three individual linear models. Interestingly, consensus models based on the two other best individual models—M09 and M10—were found to be equally predictive towards the external validation set.

Overall, this work definitely provides valuable information about the structural and physicochemical features required for predicting the surface tension of binary DESs. At the same time, it also lends important guidelines to set up predictive and validated linear interpretable QSPR models for the various properties of binary mixtures. The high predictivity of the models ensures that these models may be used on the industrial scale to at least predict the surface tension of the DESs that are newly developed or under-developed to assess their suitability as an industrial solvent. The models may also be used to screen a large number of DESs (obtained from databases) to predict the DESs with desirable surface

tension properties. What is more, all the proposed models are easily reproducible since they rely on fully specified computational procedures and were built with non-commercial software tools. Finally, both the individual and consensus models developed in this work shall help the future screening as well as the design of new sustainable DES, with major time and cost savings.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules27154896/s1>, Table S1. List of all investigated DESs and experimental surface tension data. (XLSX); Table S2. Hyperparameter tuning of the different machine learning techniques. (PDF); Table S3. Detailed description of the M09 and M10 models. (PDF); Table S4. Descriptors used in the M09 and M10 models. (PDF)

**Author Contributions:** Conceptualization, A.K.H., R.H., A.R.C.D. and M.N.D.S.C.; methodology, A.K.H., R.H. and M.N.D.S.C.; software, A.K.H.; formal analysis, A.K.H. and R.H.; investigation, A.K.H., R.H. and I.V.V.; writing—original draft preparation, A.K.H. and R.H.; writing—review and editing, I.V.V. and M.N.D.S.C.; supervision, A.R.C.D. and M.N.D.S.C.; project administration, A.R.C.D. and M.N.D.S.C.; funding acquisition, A.R.C.D. and M.N.D.S.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by UIDB/50006/2020 with funding from FCT/MCTES through national funds.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All the data files pertaining to the QSPR modeling are available from the authors. The training, test and external datasets were taken from cited publications, and the DES chemical structures along with the collected surface tension data are provided in the Supporting Information (Table S1). Dragon 7.0, MarvinView, and Standardizer were used in this study under academic license (see Material and methods section). Three other open source software tools were also used in this study, namely: QSAR-Mx, a Python-based tool developed by the authors that is available to download at <https://github.com/ncordeirfcup/QSAR-Mx> (last accessed on 28 April 2022); Mlxtend, a Python library of useful tools that is accessible from <https://rasbt.github.io/mlxtend/>; scikit-learn, a Python library of useful machine learning tools that is accessible from <https://scikit-learn.org/stable/>; and Intelligent Consensus Predictor, a Java-based tool available through the web <https://dtclab.webs.com/software-tools> (last accessed on 23 March 2022)(see Material and methods section).

**Acknowledgments:** The authors are thankful to ChemAxon for providing the academic licenses of MarvinView and Standardizer to AKH.

**Conflicts of Interest:** The authors declare no conflict of interest.

**Sample Availability:** Samples of the compounds (full dataset used for modeling) are available from the authors.

## References

1. Clarke, C.J.; Tu, W.C.; Levers, O.; Brohl, A.; Hallett, J.P. Green and Sustainable Solvents in Chemical Processes. *Chem. Rev.* **2018**, *118*, 747–800. [[CrossRef](#)]
2. Sheldon, R.A. Green Solvents for Sustainable Organic Synthesis: State of the Art. *Green Chem.* **2005**, *7*, 267–278. [[CrossRef](#)]
3. Sheldon, R.A. Fundamentals of Green Chemistry: Efficiency in Reaction Design. *Chem. Soc. Rev.* **2012**, *41*, 1437–1451. [[CrossRef](#)] [[PubMed](#)]
4. García, G.; Aparicio, S.; Ullah, R.; Atilhan, M. Deep Eutectic Solvents: Physicochemical Properties and Gas Separation Applications. *Energy Fuels* **2015**, *29*, 2616–2644. [[CrossRef](#)]
5. Hansen, B.B.; Spittle, S.; Chen, B.; Poe, D.; Zhang, Y.; Klein, J.M.; Horton, A.; Adhikari, L.; Zelovich, T.; Doherty, B.W.; et al. Deep Eutectic Solvents: A Review of Fundamentals and Applications. *Chem. Rev.* **2021**, *121*, 1232–1285. [[CrossRef](#)] [[PubMed](#)]
6. El Achkar, T.; Greige-Gerges, H.; Fourmentin, S. Basics and Properties of Deep Eutectic Solvents: A Review. *Environ. Chem. Lett.* **2021**, *19*, 3397–3408. [[CrossRef](#)]
7. Abbott, A.P.; Capper, G.; Davies, D.L.; Rasheed, R.K.; Tambyrajah, V. Novel Solvent Properties of Choline Chloride/Urea Mixtures. *Chem. Commun.* **2003**, 70–71. [[CrossRef](#)]

8. Zhang, Q.; De Oliveira Vigier, K.; Royer, S.; Jérôme, F. Deep Eutectic Solvents: Syntheses, Properties and Applications. *Chem. Soc. Rev.* **2012**, *41*, 7108–7146. [[CrossRef](#)]
9. Halder, A.K.; Cordeiro, M.N.D.S. Probing the Environmental Toxicity of Deep Eutectic Solvents and Their Components: An In Silico Modeling Approach. *ACS Sustain. Chem. Eng.* **2019**, *7*, 10649–10660. [[CrossRef](#)]
10. Palmelund, H.; Andersson, M.P.; Asgreen, C.J.; Boyd, B.J.; Rantanen, J.; Löbmann, K. Tailor-Made Solvents for Pharmaceutical Use? Experimental and Computational Approach for Determining Solubility in Deep Eutectic Solvents (DES). *Int. J. Pharm. X* **2019**, *1*, 100034. [[CrossRef](#)]
11. Nam, M.W.; Zhao, J.; Lee, M.S.; Jeong, J.H.; Lee, J. Enhanced Extraction of Bioactive Natural Products Using Tailor-Made Deep Eutectic Solvents: Application to Flavonoid Extraction from Flos Sophorae. *Green Chem.* **2015**, *17*, 1718–1727. [[CrossRef](#)]
12. Chen, Y.; Chen, W.; Fu, L.; Yang, Y.; Wang, Y.; Hu, X.; Wang, F.; Mu, T. Surface Tension of 50 Deep Eutectic Solvents: Effect of Hydrogen-Bonding Donors, Hydrogen-Bonding Acceptors, Other Solvents, and Temperature. *Ind. Eng. Chem. Res.* **2019**, *58*, 12741–12750. [[CrossRef](#)]
13. Ghaedi, H.; Ayoub, M.; Sufian, S.; Shariff, A.M.; Lal, B. The Study on Temperature Dependence of Viscosity and Surface Tension of Several Phosphonium-Based Deep Eutectic Solvents. *J. Mol. Liq.* **2017**, *241*, 500–510. [[CrossRef](#)]
14. Haghbakhsh, R.; Taherzadeh, M.; Duarte, A.R.C.; Raeissi, S. A General Model for the Surface Tensions of Deep Eutectic Solvents. *J. Mol. Liq.* **2020**, *307*, 112972. [[CrossRef](#)]
15. Le, T.; Epa, V.C.; Burden, F.R.; Winkler, D.A. Quantitative Structure-Property Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112*, 2889–2919. [[CrossRef](#)]
16. Mikolajczyk, A.; Gajewicz, A.; Rasulev, B.; Schaeublin, N.; Maurer-Gardner, E.; Hussain, S.; Leszczynski, J.; Puzyn, T. Zeta Potential for Metal Oxide Nanoparticles: A Predictive Model Developed by a Nano-Quantitative Structure-Property Relationship Approach. *Chem. Mater.* **2015**, *27*, 2400–2407. [[CrossRef](#)]
17. Kim, M.; Li, L.Y.; Grace, J.R. Predictability of Physicochemical Properties of Polychlorinated Dibenzo-p-Dioxins (PCDDs) Based on Single-Molecular Descriptor Models. *Environ. Pollut.* **2016**, *213*, 99–111. [[CrossRef](#)] [[PubMed](#)]
18. Moura, A.S.; Halder, A.K.; Cordeiro, M.N.D.S. From Biomedicinal to In Silico Models and Back to Therapeutics: A Review on the Advancement of Peptidic Modeling. *Future Med. Chem.* **2019**, *11*, 2313–2331. [[CrossRef](#)] [[PubMed](#)]
19. Sepehri, B. A Review on Created QSPR Models for Predicting Ionic Liquids Properties and Their Reliability from Chemometric Point of View. *J. Mol. Liq.* **2020**, *297*, 112013. [[CrossRef](#)]
20. Muratov, E.N.; Bajorath, J.; Sheridan, R.P.; Tetko, I.V.; Filimonov, D.; Poroikov, V.; Oprea, T.I.; Baskin, I.I.; Varnek, A.; Roitberg, A.; et al. QSAR Without Borders. *Chem. Soc. Rev.* **2020**, *49*, 3525–3564. [[CrossRef](#)]
21. Awfa, D.; Ateia, M.; Mendoza, D.; Yoshimura, C. Application of Quantitative Structure-Property Relationship Predictive Models to Water Treatment: A Critical Review. *ACS EST Water* **2021**, *1*, 498–517. [[CrossRef](#)]
22. Wang, J.; Song, Z.; Chen, L.; Xu, T.; Deng, L.; Qi, Z. Prediction of CO<sub>2</sub> Solubility in Deep Eutectic Solvents using Random Forest Model Based on COSMO-RS-Derived Descriptors. *Green Chem. Eng.* **2021**, *2*, 431–440. [[CrossRef](#)]
23. Balali, M.; Sobati, M.A.; Gorji, A.E. QSPR Modeling of Thiophene Distribution Between Deep Eutectic Solvent (DES) and Hydrocarbon Phases: Effect of Hydrogen Bond Donor (HBD) Structure. *J. Mol. Liq.* **2021**, *342*, 117496. [[CrossRef](#)]
24. Khajeh, A.; Shakourian-Fard, M.; Parvaneh, K. Quantitative Structure-Property Relationship for Melting and Freezing Points of Deep Eutectic Solvents. *J. Mol. Liq.* **2021**, *321*, 114744. [[CrossRef](#)]
25. Benguerba, Y.; Alnashef, I.M.; Erto, A.; Balsamo, M.; Ernst, B. A Quantitative Prediction of the Viscosity of Amine Based DESs Using S<sub>s-profile</sub> Molecular Descriptors. *J. Mol. Struct.* **2019**, *1184*, 357–363. [[CrossRef](#)]
26. Lemaoui, T.; Hammoudi, N.E.H.; Alnashef, I.M.; Balsamo, M.; Erto, A.; Ernst, B.; Benguerba, Y. Quantitative Structure Properties Relationship for Deep Eutectic Solvents Using S<sub>G-profile</sub> as Molecular Descriptors. *J. Mol. Liq.* **2020**, *309*, 113165. [[CrossRef](#)]
27. Lemaoui, T.; Darwish, A.S.; Attoui, A.; Hatab, F.A.; Hammoudi, N.E.H.; Benguerba, Y.; Vega, L.F.; Alnashef, I.M. Predicting the Density and Viscosity of Hydrophobic Eutectic Solvents: Towards the Development of Sustainable Solvents. *Green Chem.* **2020**, *22*, 8511–8530. [[CrossRef](#)]
28. Halder, A.K.; Haghbakhsh, R.; Voroshylova, I.V.; Duarte, A.R.C.; Cordeiro, M.N.D.S. Density of Deep Eutectic Solvents: The Path Forward Cheminformatics-Driven Reliable Predictions for Mixtures. *Molecules* **2021**, *26*, 5779. [[CrossRef](#)]
29. Haghbakhsh, R.; Bardool, R.; Bakhtyari, A.; Duarte, A.R.C.; Raeissi, S. Simple and Global Correlation for the Densities of Deep Eutectic Solvents. *J. Mol. Liq.* **2019**, *296*, 111830. [[CrossRef](#)]
30. Organization for Economic Co-Operation and Development (OECD). *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship ((Q)SAR) Models*; OECD Series on Testing and Assessment 69; OECD Document ENV/JM/MONO2007; OECD Publishing: Paris, France, 2014; pp. 55–65.
31. Toropov, A.A.; Toropova, A.P. QSPR/QSAR: State-of-Art, Weirdness, the Future. *Molecules* **2020**, *25*, 1292. [[CrossRef](#)]
32. Omar, K.A.; Sadeghi, R. Novel Deep Eutectic Solvents Based on Pyrogallol: Synthesis and Characterizations. *J. Chem. Eng. Data* **2021**, *66*, 2088–2095. [[CrossRef](#)]
33. Nunes, R.J.; Saramago, B.; Marrucho, I.M. Surface Tension of dl-Menthol:Octanoic Acid Eutectic Mixtures. *J. Chem. Eng. Data* **2019**, *64*, 4915–4923. [[CrossRef](#)]
34. Lapeña, D.; Bergua, F.; Lomba, L.; Giner, B.; Lafuente, C. A Comprehensive Study of the Thermophysical Properties of Reline and Hydrated Reline. *J. Mol. Liq.* **2020**, *303*, 112679. [[CrossRef](#)]



35. Abdallah, M.M.; Müller, S.; González de Castilla, A.; Gurikov, P.; Matias, A.A.; Bronze, M.d.R.; Fernández, N. Physicochemical Characterization and Simulation of the Solid–Liquid Equilibrium Phase Diagram of Terpene-Based Eutectic Solvent Systems. *Molecules* **2021**, *26*, 1801. [[CrossRef](#)] [[PubMed](#)]
36. 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. Inform.* **2012**, *31*, 202–221. [[CrossRef](#)] [[PubMed](#)]
37. Oprisiu, I.; Novotarskyi, S.; Tetko, I.V. Modeling of Non-Additive Mixture Properties Using the Online CHEMical Database and Modeling Environment (OCHEM). *J. Cheminformatics* **2013**, *5*, 4. [[CrossRef](#)]
38. Halder, A.K.; Cordeiro, M.N.D.S. Development of Predictive Linear and Non-linear QSTR Models for *Aliivibrio Fischeri* Toxicity of Deep Eutectic Solvents. *IJQSPR* **2019**, *4*, 50–69. [[CrossRef](#)]
39. ChemAxon. *Standardizer*; Version 15.9.14.0 Software; ChemAxon: Budapest, Hungary, 2010.
40. Mauri, A.C.V.; Pavan, M.; Todeschini, R. Dragon Software: An Easy Approach to Molecular Descriptor calculations. *MATCH Commun. Math. Comput. Chem.* **2006**, *56*, 237–248.
41. Hechinger, M.; Leonhard, K.; Marquardt, W. What is Wrong with Quantitative Structure–Property Relations Models Based on Three-Dimensional Descriptors? *J. Chem. Inf. Model.* **2012**, *52*, 1984–1993. [[CrossRef](#)]
42. Raschka, S. MLxtend: Providing Machine Learning and Data Science Utilities and Extensions to Python's Scientific Computing Stack. *J. Open Source Softw.* **2018**, *3*, 638. [[CrossRef](#)]
43. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inform. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
44. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
45. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory ACM, Pittsburgh, PA, USA, 27–29 July 1992; pp. 144–152.
46. Guang-Bin, H.; Babri, H.A. Upper Bounds on the Number of Hidden Neurons in Feedforward Networks with Arbitrary Bounded Nonlinear Activation Functions. *IEEE Trans. Neural Netw.* **1998**, *9*, 224–229. [[CrossRef](#)] [[PubMed](#)]
47. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
48. Golbraikh, A.; Tropsha, A. Beware of Q<sup>2</sup>! *J. Mol. Graph. Model.* **2002**, *20*, 269–276. [[CrossRef](#)]
49. Gramatica, P. On the Development and Validation of QSAR Models. *Methods Mol. Biol.* **2013**, *930*, 499–526. [[PubMed](#)]
50. Roy, P.P.; Paul, S.; Mitra, I.; Roy, K. On Two Novel Parameters for Validation of Predictive QSAR Models. *Molecules* **2009**, *14*, 1660–1701.
51. Ojha, P.K.; Roy, K. Comparative QSARs for Antimalarial Endochins: Importance of Descriptor-Thinning and Noise Reduction Prior to Feature Selection. *Chemom. Intell. Lab. Syst.* **2011**, *109*, 146–161. [[CrossRef](#)]
52. Gramatica, P. Principles of QSAR Models Validation: Internal and External. *QSAR Comb. Sci.* **2007**, *26*, 694–701. [[CrossRef](#)]
53. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
54. Roy, K.; Ambure, P.; Kar, S.; Ojha, P.K. Is It Possible to Improve the Quality of Predictions from an “Intelligent” Use of Multiple QSAR/QSPR/QSTR Models? *J. Chemom.* **2018**, *32*, e2992. [[CrossRef](#)]
55. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
56. Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*, 2nd ed.; Wiley-VCH: Weinheim, Germany, 2009.
57. Labute, P. A Widely Applicable Set of Descriptors. *J. Mol. Graph. Model.* **2000**, *18*, 464–477. [[CrossRef](#)]
58. Reutlinger, M.; Koch, C.P.; Reker, D.; Todoroff, N.; Schneider, P.; Rodrigues, T.; Schneider, G. Chemically Advanced Template Search (CATS) for Scaffold-Hopping and Prospective Target Prediction for ‘Orphan’ Molecules. *Mol. Inform.* **2013**, *32*, 133–138. [[CrossRef](#)] [[PubMed](#)]
59. García-Domenech, R.; Julián-Ortiz, J.V. Antimicrobial Activity Characterization in a Heterogeneous Group of Compounds. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 445–449. [[CrossRef](#)] [[PubMed](#)]
60. Khan, K.; Khan, P.M.; Lavado, G.; Valsecchi, C.; Pasqualini, J.; Baderna, D.; Marzo, M.; Lombardo, A.; Roy, K.; Benfenati, E. QSAR Modeling of *Daphnia magna* and Fish Toxicities of Biocides Using 2D Descriptors. *Chemosphere* **2019**, *229*, 8–17. [[CrossRef](#)] [[PubMed](#)]