

Molecular-Based Guide to Predict the pH of Eutectic Solvents: Promoting an Efficient Design Approach for New Green Solvents

Tarek Lemaoui,¹ Farah Abu Hatab,¹ Ahmad S. Darwish, Ayoub Attoui, Nour El Houda Hammoudi, Ghaiath Almustafa, Mohamed Benaicha, Yacine Benguerba, and Inas M. Alnashef*



Cite This: *ACS Sustainable Chem. Eng.* 2021, 9, 5783–5808



Read Online

ACCESS |

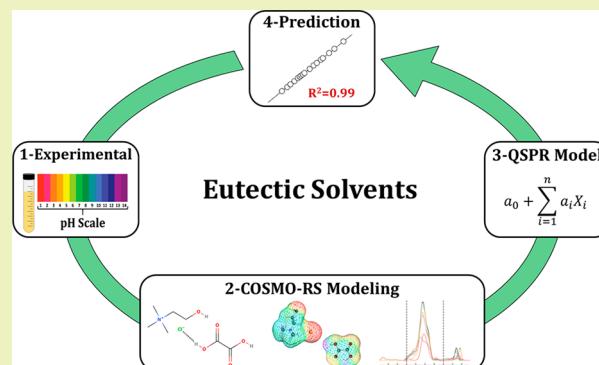
Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The case of sustainable solvents is of great interest both academically and industrially. With research communities becoming more aware of the negative impacts of conventional organic solvents, a range of greener and more sustainable solvents have been developed to counter the harmful drawbacks associated with conventional solvents. Among these, eutectic solvents (ESs) attracted considerable attention for their “green” properties and have proven their usefulness as environmentally benign alternatives to classical solvents. Among the various desirable characteristics of ESs, pH is a key property with significant implications for the design and control of industrial-scale applications. However, selecting an ES with the required pH for a particular application is a challenging task, especially with extensive experimentally determined data being time consuming and expensive. Therefore, in this work, the pH of various ESs have been predicted via novel quantitative structure–property relationships (QSPR) models using two machine learning algorithms, a multiple linear regression (MLR) and an artificial neural network (ANN), with a set of molecular descriptors generated by COSMO-RS. A total of 648 experimental points for 41 chemically unique ESs prepared from 9 HBAs and 21 HBDs at different temperatures were utilized for sufficient data set representation. On the basis of the statistical analysis of the models, it can be concluded that both approaches can be utilized as powerful predictive tools in estimating the pH of new ESs with the ANN model having better predictive capabilities and the MLR model being more interpretable. These models inspire and stimulate the development of robust models to predict the properties of designer solvents from the drawn molecular structures, which will save time and resources.

KEYWORDS: *Green solvents, pH, Eutectic solvents (ESs), Quantitative structure–property relationships (QSPR), Multiple linear regression (MLR), Artificial neural networks (ANN)*



1. INTRODUCTION

With the appearance of the green chemistry concept, and since new solvents for sustainable chemical processes are in constant demand nowadays, the selection and optimization of solvent systems are vital to many industrial applications including fuel purification, biochemical, metal refining, and water/wastewater treatment.^{1–4} The organic solvents traditionally used in such industries are often associated with a high toxicity profile.^{5–7} The volatility of these conventional solvents not only increases their exposure rate to the environment but also incurs significant economic losses due to solvent evaporation.⁸ Modern chemistry and chemical engineering practices reached considerable milestones in replacing conventional organic solvents with less hazardous ones.^{7,9} One of the major steps taken in this direction was the development made in the field of ionic liquids (ILs).^{10,11} Although ILs are chemically and thermally stable, highly nonvolatile, and tunable, they suffer from high production costs and potential toxic manifestations.^{12–14} Consequently, these challenges stunt their application as feasible alternatives.

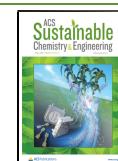
For this reason, green technology has actively sought to obtain new solvents to replace ILs.

A more recent milestone was achieved in 2003, when deep eutectic solvents (DES) emerged as potential analogs for ILs.¹⁵ In general, DESs are mixtures of two or more hydrogen-bond acceptors (HBA) and hydrogen-bond donors (HBD) that are associated with each other through hydrogen-bond interactions.¹⁶ The resulting eutectic mixture has a decreased melting point relative to the individual components. As the DESs are still considered in their infant stages, different arguments have been raised when defining DES, and several papers were arguing the objective of finding a distinct and

Received: October 8, 2020

Revised: March 25, 2021

Published: April 22, 2021



uniform definition.^{17,18} As stated by Martins et al.,¹⁹ a “deep eutectic solvent” is a mixture of two or more pure compounds for which the eutectic point temperature is below that of an ideal liquid mixture, presenting significant negative deviations from ideality. Additionally, the temperature depression should be such that the mixture is liquid at operating the temperature for a certain composition range. Otherwise, a simpler term “eutectic solvent” could be used to describe mixtures that do not fulfill these criteria. Therefore, since the definition of these solvents are still being reviewed,^{17–19} in this work, the term “eutectic solvent” (ES) was adopted.

This “new generation designer’s solvents” offer often advantageous characteristics such as simple synthesis, low production costs, and biodegradability.^{20,21} Moreover, their structures can be attuned by selecting both the type and the ratio of the hydrogen-bonding constituents, consequently allowing an additional degree of freedom.^{22,23} Therefore, each combination of ES, even those that are varied only by molar ratios, should be treated as distinct ESs as each combination can demonstrate different physicochemical, thermal, and even solvation properties. These versatile characteristics of ESs facilitated their use in different applications such as catalysis, separation, biochemistry, electrochemistry, and nanotechnology.^{24–32}

Characterization of this neoteric solvent is mandatory to identify its technical features and ecological impact. Several sets of physicochemical properties (e.g., density, viscosity, refractive index, and pH) of ESs were measured and can be found in the literature.^{17,33} Among these imperative properties, the solvent’s pH is an essential physical property that directly influences the design criteria of many industrial processes in terms of selection of construction materials, corrosion control, and operation conditions.^{34–37} The pH value of ESs affects the chemical and biochemical reactions in different treatment processes. For instance, in metallurgical processes, such as those dealing with material recovery from mining ores or end-of-life batteries, the acidity of ESs affects the dissolution of metal oxides; the varying solubility of metal oxides based on the pH value of the ES can be utilized for preferential separation of certain metals from their aqueous media.³⁸ Moreover, the pH value of ESs used in biomass treatment affects the product’s composition and properties resulting in extracts with different structures and yield.³⁹ Xu and co-workers⁴⁰ used an ES of choline chloride and glycerol (2:1) to consolidated a bioconversion process; the selection of the ES constituents, and accordingly the acidity of the solvent, were key factors to enable downstream bioconversion without additional treatments. Nevertheless, acidic ESs account for optimizing some fuel purification processes.^{41,42} For example, the use of ESs consisted of quaternary ammonium salts (i.e., choline chloride or tetrabutylammonium chloride) combined with *p*-toluenesulfonic acid could reach extraction efficiencies up to 97.25% and 95.90%, respectively, in deep oxidation/extraction desulfurization of fuels.⁴²

Owing to the importance of pH in various industries, several studies were concerned with the role of pH in optimizing processes. Such an example could be seen in the process of converting waste CO₂ into a carbon-based chemical where Varela⁴³ discussed the importance of the pHs of electrolytes in optimizing both the activity and selectivity of the CO₂ electrochemical reduction reaction (CO₂RR). Furthermore, measuring the solution’s pH is of great importance in determining the solubility of calcium and phosphorus, which are the major nutrients of an infant’s needs. Therefore, a study

conducted by Porcelli et al.⁴⁴ developed a model to predict the pH of the parenteral solution as a linear function of the individual parenteral component concentrations. Another study reported by Bi⁴⁵ predicted the pH values of acidic natural water based on a mathematical model where different factors were utilized to test their effect on the acidity of the water including the aluminum concentration, dissolved inorganic carbon, mineral phase, and temperature. Additionally, the effect of pH on the extraction of boron from wastewater and brines was studied by Almustafa et al.⁴⁶ using thymol, menthol, and 2-methyl-2,4-pentanediol-based ESs. The study showed the importance of optimizing the pH value of the solution to maximize the extraction efficiency of boron.

Nevertheless, studies reporting experimental values of the pHs of ESs are fairly scarce relative to other properties (i.e., density and viscosity). Thus, more investigations are required in this area to provide deeper insights on the ES’s acidity; however, considering the wide range of potential ESs, depending only on experimental measurements for solvent selection parts away from the sustainable approach and is not economically feasible. Therefore, mathematical and simulation models for the accurate prediction of ES properties are continuously being sought out. One of the furthermost usually applied methods to replace the experimental techniques is the use of computer-assisted quantitative structure–property relation (QSPR) models.⁴⁷ A QSPR model is a mathematical model that relates the component’s molecular-level structure to its macroscopic-level physicochemical properties.^{48,49} This modeling approach showed its reliability in predicting the physical, chemical, and biological properties of numerous solvents.^{50–52} In an attempt to develop a QSPR model, the following should be obtained: (1) an extensive data set which covers different chemicals for sufficient statistical analysis, (2) selection of molecular descriptors, (3) calculation and generating the molecular descriptors of these compounds, (4) selecting a proper algorithm to relate the molecular descriptors with the dependent variable, and (5) performing a statistical analysis to ensure the robustness and the applicability of the predictive model. The application of QSPR models in predicting the physicochemical properties of ionic liquids has been reported extensively in the literature. Such examples are density,^{53,54} viscosity,^{55,56} surface tension,⁵⁷ melting point,^{58,59} thermal decomposition,^{60–63} and heat capacity.⁶⁴

The conductor-like screening model for real solvents (COSMO-RS) is a tool that has been used previously to design molecular descriptors by representing the molecule’s surface with the surface polarity distribution (σ -profile) and their distribution areas (S_{σ} -profile).⁶⁵ The use of COSMO-RS-based molecular descriptors combined with the QSPR approach proved its applicability in predicting different properties. Koi et al.⁶⁶ predicted the viscosity of imidazolium-based ILs, and the obtained results showed the suitability of the proposed model in fitting the experimental data and predicting the viscosity of new ILs.⁶⁶ Zhao et al.⁶⁵ also predicted the viscosity of different ILs using the S_{σ} -profiles molecular descriptors coupled with linear and nonlinear QSPR models. A similar approach was adopted to predict the ecotoxicity of a wide range of ILs, and the suggested QSPR models were recommended for the preselection of ILs before application.⁶⁷ In terms of predicting the physicochemical properties of ESs, Lemaoui et al.⁶⁸ studied the application of QSPR models in predicting the densities and viscosities of 49 ESs reported in the literature. In continuation of the previous work, Lemaoui et al.⁶⁹ also proposed another two models for predicting the electrical conductivities of ESs, and the obtained

results exhibited robust models in describing the electrical conductivity properties of ESs.⁶⁹

Given the essential role of pH in identifying the suitability and compatibility of ESs in particular applications, the primary objective of this article is the development of the first QSPR model that can predict the pHs of ESs by simply correlating their molecular-level structure with the macroscopic-level properties. Here, we propose novel QSPR models, which can predict the pH values of different ESs based on COSMO-RS-based molecular descriptors (S_{σ} -profiles). The novelty of this work is shaded with filling the gaps observed in the literature, where studies reporting the experimental values of the pHs of ESs are fairly scarce. Subsequently, the model at hand can be utilized for studying and predicting the pHs of many ESs, which have not been tested experimentally, utilizing 10 basic molecular descriptors only enabling a straightforward, time-saving, and cheap method for screening new green and sustainable ESs with the required pH for a particular application.

According to the author's knowledge, this is the first article that discusses the use of predictive models in predicting the pH properties of ESs. Thus, this work presents an initiative toward developing robust models capable of predicting the pHs of designer solvents to promote a more efficient and "sound" design method for screening new sustainable solvents. To guarantee the generalizability of the developed model, the data set utilized covered all the experimentally determined data of ESs in the literature (to the best of our knowledge). Moreover, the S_{σ} -profiles molecular descriptors were developed using the COSMO-RS quantum chemical approach. Two machine learning algorithms were utilized to correlate descriptors and the pH property, namely, the multiple linear regression (MLR) model, which has been previously proven for its simplicity and high interpretability,⁷⁰ and the artificial neural network (ANN) model, which was also developed and applied extensively to a variety of engineering problems, and they both have been proven for their high accuracy and reliability in solving complex problems.⁷¹ The developed models were then assessed using various statistical approaches, and their "chemical space" has been computed using applicability domain analysis.

2. METHODOLOGY

The method used in this work is described in this section, starting by collecting the experimental data set, setting up the QSPR model, and evaluating the model based on several statistical parameters and tests. Figure 1 summarizes the methodology applied. The calculation and modeling approaches used are described in more detail in our previous works.^{68,69,72}

2.1. Experimental Data. To the best of our knowledge, the data set reported in the literature covered 41 ESs with 84 different compositions prepared from 9 HBAs and 21 HBDs resulting in a total of 648 experimental points covering a variety of cations, anions, and functional groups. The ESs utilized in the data set are considered to be a representative batch of a sufficiently extensive range of molecules that allow for a robust approach to predict the pH property of hydrophilic ESs. The utilized pH measurements also include a broad range of temperatures (358.15–293.15 K), water content (0.18–26.7), and molar ratios (9:1–1:16) all measured at 1.01 bar, which is also anticipated to improve the robustness of the model. Table 1 summarizes the ESs used in this work with their corresponding molar ratios and experimental pH measurements.⁷³ The complete experimental data set is available in Table S.1 in the Supporting Information. Also, since the presence of water is an

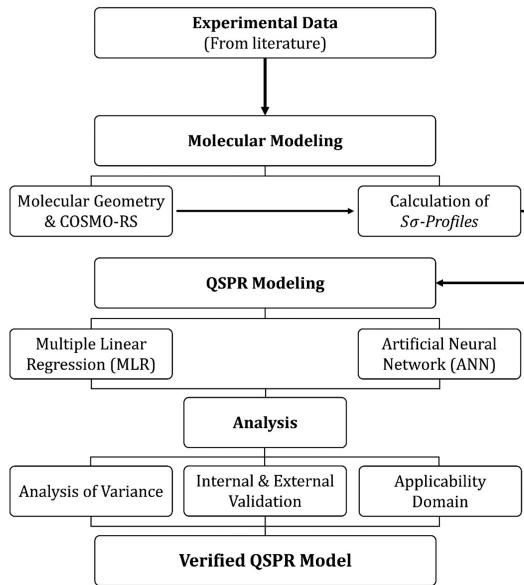


Figure 1. Summary of applied methodology.

important parameter that affects the pHs of ESs, the water contents of all the ESs have been accounted for, and their respective water molar ratios are shown in Table 1.

2.2. Generating the Molecular Descriptors. In order to generate the COSMO-RS-based molecular descriptors, first, the SMILES of each HBA and HBD molecule were imported to the Turbomole software (TmoleX version 4.5.1). Then, the 3D structures were geometrically optimized at the DFT level by combining the def-TZVP "triple- ζ valence polarized" functions with the BP86 "Becke–Perdew 86" generalized gradient approximation. The SCF margin for the calculations was set to 1×10^{-6} Hartree.⁸⁰ The files generated for each molecule were then exported as "COSMO" files and then imported into the COSMO-RS software "COSMOThermX". Using the COSMOThermX software, the 51 points of the σ -profile within the range of ± 0.025 e/Å were then extracted as "prf" data. The σ -profile data were then converted into molecular descriptors denoted as S_{σ} -profiles by entering their data into MATLAB to calculate the integral of the area under the σ -profile curves. In this sense, the discretization of the σ -profile curves into 4, 6, 8, 10, and 12 S_{σ} -profile descriptors has been determined. After calculating the S_{σ} -profile descriptor of each HBA and HBD, the S_{σ} -profile of the ESs were calculated as follows⁷²

$$S_i^{\text{ES}} = (x_{\text{HBA}})(S_i^{\text{HBA}}) + (x_{\text{HBD}})(S_i^{\text{HBD}}) + (x_{\text{H}_2\text{O}})(S_i^{\text{H}_2\text{O}}) \quad (1)$$

where S_i is the descriptor in region i (e/Å²), and x_{HBA} , x_{HBD} , and $x_{\text{H}_2\text{O}}$ are the mole fractions of the HBA, HBD, and water, respectively. Table S.2 in Supporting Information lists the calculated S_{σ} -profile descriptors for the ES constituents used in this study. This method of modeling is superior to the conventional method of defining the ES as a pseudo-pure component. Here, the ES is considered to be a mixture of three components: HBA, HBD, and water. The benefit of considering ES as a mixture is that the resulting model would be much more flexible with regard to modeling new combinations of ES as it enables changing the HBA, HBD, or their molar ratio with ease.^{81,82} This method also allows the developed model to account for the effect of water content on the pHs of ESs, which

Table 1. Constituents of Eutectic Solvents with Corresponding Molar Ratios, Water Contents, and Experimental pH Measurements^a

ES	Abbreviation	Mole ratio	pH	Temperature	ref
ES1	ATPPB:DEG:H ₂ O	1:4:0.17	1.49–0.50	293.15–343.15	75
ES1.1	ATPPB:DEG:H ₂ O	1:10:0.31	4.05–3.23	293.15–343.15	75
ES1.2	ATPPB:DEG:H ₂ O	1:16:0.39	4.21–3.34	293.15–343.15	75
ES2	ATPPB:TEG:H ₂ O	1:4:0.18	1.40–0.15	293.15–343.15	75
ES2.1	ATPPB:TEG:H ₂ O	1:10:0.35	3.15–1.90	293.15–343.15	75
ES2.2	ATPPB:TEG:H ₂ O	1:16:0.56	3.42–2.47	293.15–343.15	75
ES3	BTPC:EG	1:3	5.71–5.59	298.15–353.15	76
ES4	BTPC:Gly	1:5	6.90–7.02	298.15–353.15	76
ES5	ChCl:CA:H ₂ O	1:1:1.33	1.72–0.92	298.15–333.15	39
ES5.1	ChCl:CA:H ₂ O	2:1:1.44	1.33–0.98	298.15–333.15	39
ES6	ChCl:DEA	1:6	11.47–9.98	295.15–353.15	39
ES7	ChCl:EG:H ₂ O	1:2:0.33	4.38–4.00	298.15–333.15	39
ES8	ChCl:Fru	1:1	6.10–4.43	298.15–358.15	39
ES8.1	ChCl:Fru	1.5:1	6.91–6.32	298.15–358.15	39
ES8.2	ChCl:Fru	2:1	6.65–4.85	298.15–358.15	39
ES8.3	ChCl:Fru	2.5:1	7.10–6.41	298.15–358.15	39
ES9	ChCl:Glu	1:1	6.83–6.25	298.15–353.15	39
ES9.1	ChCl:Glu	1.5:1	7.10–5.99	298.15–353.15	39
ES9.2	ChCl:Glu	2:1	7.00–6.45	298.15–353.15	39
ES9.3	ChCl:Glu	2.5:1	7.11–6.47	298.15–353.15	39
ES10	ChCl:Gly:H ₂ O	1:2:0.33	4.47–4.12	298.15–333.15	39
ES11	ChCl:GlyA:H ₂ O	1:3:0.44	1.24–0.99	298.15–333.15	39
ES12	ChCl:LacA:H ₂ O	1:5:0.67	1.73–0.99	298.15–333.15	39
ES12.1	ChCl:LacA:H ₂ O	1:10:1.22	1.77–1.04	298.15–333.15	39
ES13	ChCl:MA:H ₂ O	1:1:0.22	1.61–0.94	298.15–333.15	39
ES13.1	ChCl:MA:H ₂ O	2:1:0.33	1.93–1.19	298.15–333.15	39
ES14	ChCl:MalA:H ₂ O	1:1:0.22	1.28–0.41	298.15–333.15	39
ES15	ChCl:MDEA	1:6	11.04–9.89	295.15–353.15	39
ES16	ChCl:MEA	1:6	12.81–11.12	295.15–353.15	39
ES17	ChCl:OxaA:H ₂ O	1:1:2.44	1.21–0.06	298.15–333.15	39
ES18	ChCl:TFA	1:2	3.97–3.86	298.15–353.15	39
ES19	DEEAC:MalA	1:1	2.41–2.29	298.15–353.15	77
ES20	EAC:Gly:H ₂ O	1:3:0.64	2.04–1.97	303.15–353.15	74
ES20.1	EAC:Gly:H ₂ O	1:4:0.95	2.42–2.33	303.15–353.15	74
ES20.2	EAC:Gly:H ₂ O	1:5:1.02	2.57–2.44	303.15–353.15	74
ES21	LacA:Ala:H ₂ O	9:1:1.11	2.15–1.42	298.15–333.15	39
ES22	LacA:Bet:H ₂ O	2:1:0.33	2.45–1.85	298.15–333.15	39
ES23	LacA:Glyi:H ₂ O	2:1:0.33	2.74–2.18	298.15–333.15	39
ES23.1	LacA:Glyi:H ₂ O	9:1:1.11	2.27–1.54	298.15–333.15	39
ES24	MA:Suc:H ₂ O	1:1:0.22	2.05–1.35	298.15–333.15	39
ES25	MTPB:EG	1:4	6.35–5.86	298.15–353.15	76
ES26	MTPB:Gly	1:1.75	6.97–6.70	298.15–353.15	76
ES27	MTPB: TFA	1:8	2.71–3.34	298.15–353.15	76
ES28	TBAC:EG	1:2	9.10–7.51	293.15–353.15	78
ES28.1	TBAC:EG	1:3	9.20–7.76	293.15–353.15	78
ES28.2	TBAC:EG	1:4	9.35–8.19	293.15–353.15	78
ES29	TBAC:Gly	1:3	6.51–6.11	293.15–353.15	78
ES29.1	TBAC:Gly	1:4	8.95–7.50	293.15–353.15	78
ES29.2	TBAC:Gly	1:5	6.81–6.42	293.15–353.15	78
ES30	TBAC:TEG	1:1	6.40–5.92	293.15–353.15	78
ES30.1	TBAC:TEG	2:1	6.97–6.21	293.15–353.15	78
ES30.2	TBAC:TEG	3:1	7.70–6.73	293.15–353.15	78
ES30.3	TBAC:TEG	4:1	8.06–7.03	293.15–353.15	78
ES31	TPAB:EG	1:3	6.41–5.97	293.15–353.15	34
ES31.1	TPAB:EG	1:4	6.53–6.14	298.15–353.15	34
ES31.2	TPAB:EG	1:5	7.23–6.57	298.15–353.15	34
ES32	TPAB:Gly	1:2	6.40–6.03	298.15–353.15	34
ES32.1	TPAB:Gly	1:3	5.96–5.85	298.15–353.15	34
ES32.2	TPAB:Gly	1:4	5.85–5.64	298.15–353.15	34
ES33	TPAB:TEG	1:2.5	5.09–4.80	298.15–353.15	34

Table 1. continued

ES	Abbreviation	Mole ratio	pH	Temperature	ref
ES33.1	TPAB:TEG	1:3	5.22–4.94	298.15–353.15	34
ES33.2	TPAB:TEG	1:4	5.15–4.87	298.15–353.15	34
ES34	ChCl:LacA:H ₂ O	1:9:1.11	1.61–0.80	298.15–333.15	39
ES35	Bet:MA:H ₂ O	1:1:1.5	3.39–2.62	288.15–328.15	79
ES35.1	Bet:MA:H ₂ O	1:1:6	3.40–2.90	288.15–328.15	79
ES35.2	Bet:MA:H ₂ O	1:1:13.9	2.95–2.50	288.15–328.15	79
ES36	ChCl:CA:H ₂ O	2:1:3	0.63–0.67	288.15–328.15	79
ES36.1	ChCl:CA:H ₂ O	2:1:11.5	0.88–0.98	288.15–328.15	79
ES36.2	ChCl:CA:H ₂ O	2:1:26.7	1.11–1.18	288.15–328.15	79
ES37	ChCl:MA:H ₂ O	1:1:1.7	0.22–0.34	288.15–328.15	79
ES37.1	ChCl:MA:H ₂ O	1:1:6.5	0.55–0.78	288.15–328.15	79
ES37.2	ChCl:MA:H ₂ O	1:1:15.2	1.10–1.11	288.15–328.15	79
ES38	Bet:CA:H ₂ O	1:1:1.9	2.81–2.15	288.15–328.15	79
ES38.1	Bet:CA:H ₂ O	1:1:7.4	2.77–2.15	288.15–328.15	79
ES38.2	Bet:CA:H ₂ O	1:1:17.2	2.75–2.12	288.15–328.15	79
ES39	ChCl:Pro:MA:H ₂ O	1:1:1:2.4	3.63–3.58	288.15–328.15	79
ES39.1	ChCl:Pro:MA:H ₂ O	1:1:1:9.3	3.35–2.80	288.15–328.15	79
ES39.2	ChCl:Pro:MA::H ₂ O	1:1:1:21.6	2.95–3.03	288.15–328.15	79
ES40	Pro:MA:H ₂ O	1:1:1.5	2.17–2.19	288.15–328.15	79
ES40.1	Pro:MA:H ₂ O	1:1:5.9	2.87–2.29	288.15–328.15	79
ES40.2	Pro:MA:H ₂ O	1:1:13.8	2.86–2.28	288.15–328.15	79
ES41	MA:Glu:H ₂ O	1:1:1.9	0.37–0.46	288.15–328.15	79
ES41.1	MA:Glu:H ₂ O	1:1:7.5	0.45–0.67	288.15–328.15	79
ES41.2	MA:Glu:H ₂ O	1:1:17.4	0.76–0.81	288.15–328.15	79

^aAbbreviations are available in the Nomenclature section.

is crucial in screening new ESs with the required pH for a particular application.

2.3. Machine Learning Algorithms. **2.3.1. Multiple Linear Regression.** Multiple linear regression (MLR) is a machine learning technique reported extensively in the literature for its reliability and interpretability in expressing a simple linear relation between the dependent variable (i.e., physicochemical properties) and the independent variables (i.e., molecular descriptors).⁷⁰ The linear equation can be expressed as follows⁶⁸

$$y = a_0 + \sum_{i=1}^n a_i S_i + \sum_{i=1}^n \sum_{j=i+1}^n a_{i-j} (S_i - \bar{S}_i)(S_j - \bar{S}_j) \quad (2)$$

where a_0 represents the intercept of the linear equation, S_i represents descriptor i , a_i is coefficient of descriptor i , n is the total number of descriptors, and $a_{i-j}(S_i - \bar{S}_i)(S_j - \bar{S}_j)$ represents the binary interactions between a pair of descriptors. A list of all the descriptors (i.e., individual descriptors and their binary interactions) can be found in Table S.3 in the Supporting Information.

The MLR model was developed through the multilinear fitting toolbox of the JMP statistical software. The discretized S_{σ} -profile descriptors and the temperature were selected as the inputs, while the pHs of the ESs were selected as the output. The binary interactions between a pair of descriptors were added by selecting the second degree factorial option. The fitting method was set as “forward stepwise” with the parameter cost function (stopping rule) selected as minimum AIC_c, the corrected Akaike information criterion, which can be defined as follows

$$\text{AIC} = 2K - 2\ln(L) \quad (3)$$

$$\text{AIC}_c = \text{AIC} + \frac{2K^2 + 2K}{p - K - 1} \quad (4)$$

where K represents the number of estimated parameters in the model, L represents the maximum value of the likelihood function in the model, and p represents the number of training experimental data points. Note that as $p \rightarrow \infty$, the extra corrected penalty term in AIC_c converges to 0, and thus AIC_c converges to AIC. Using the stepwise AIC_c algorithm, only the significant descriptors that enhanced the model's information criterion were added to the model, while the ones that had an insignificant effect were eliminated.⁸³

2.3.2. Artificial Neural Network. In addition to the MLR machine learning algorithm, a feed-forward artificial neural network (ANN) has also been utilized to develop a robust nonlinear correlation between the descriptors and the pHs of ESs. The network consists of several processing elements denoted as “neuron nodes”. The neurons are associated with each other by direct communication activation functions that contain the information required to generate the output.^{84,85} The hyperbolic tangent sigmoid activation function of each hidden neuron (H_k) can be computed as follows⁸⁴

$$H_k = \tanh\left(\frac{1}{2}Y_k\right) \quad (5)$$

The tanh activation function transforms the Y_k values to be between -1 and 1. Y_k is a linear combination of the inputs linked to hidden neuron k , which can be calculated as follows

$$Y_k = \sum_{k=1}^M (W_{k,\text{input}})(S_i) + b_k \quad (6)$$

where $W_{k,\text{input}}$ represents the weight coefficient of the link between each input and hidden neuron k , and b_k represents the intercept bias of hidden neuron k .

The neural network was developed through the neural networks toolbox of the JMP statistical software. The discretized

S_σ -profile descriptors and the temperature were selected as the network's inputs, while the pHs of the ESs were selected as the output. The network's learning rate was fixed at 0.1 with a squared penalty method and an internal cross-validation holdback proportion of 25%.

2.4. Model Evaluation. The goodness-of-fit of the proposed QSPR models was evaluated depending on the regression coefficients (R^2) and Fisher's statistic (F_{Ratio}) values. Thereafter, to ensure that the descriptors selected are stable and not sensitive to changes in the input data set, five internal cross-validation tests based on different splitting techniques were applied: (1) pH at room temperature ordered response technique (Q^2_{pH}), (2) molecular weight ordered response technique (Q^2_{Mw}), (3) averaged molecular descriptor $S^{\text{ES}}_{1-10,\text{avg}}$ structural response technique (Q^2_{Savg}), (4) randomly leave-many-out (Q^2_{LMO}) technique, and (5) extensively used leave-one-out (Q^2_{LOO}) technique.⁸⁶ In the first three splitting techniques, the ESs were sorted in ascending order based on the selected parameter, and then, one out of every four ESs were excluded from the training set. In the fourth technique, 25% of the ESs were randomly excluded from the training set. In the fifth technique, the cross-validation is computed by excluding one ES from the training set and determining the model's internal fit. The process is then repeated multiple times until all the ESs have been removed once, and then, an average of the internal fits is computed as the Q^2_{LOO} cross-validation coefficient.

To confirm that the models are not unintentionally associated, the y -scrambling procedure was applied.⁸³ In this method, the models were developed on the basis of modifying the experimental data for 300 iterations to generate randomly reordered responses. The goodness of the developed models in fitting the new randomly reordered responses was evaluated based on the y -scrambling regression coefficient (R^2_{scramble}). Finally, the overall error for each model was calculated by the root-mean-square error (RMSE), the relative deviation (RD), the standard deviation (SD), and the absolute-average relative deviation (AARD).

2.5. Applicability Domain. Different methods have been used to define the applicability domain (AD), but the most commonly used is the *leverage* approach where the reliability of the model is evaluated based on the leverage value (h_i).⁸⁷ The leverage value is a quantitative measure to assess the presence of structural outliers by calculating h_i and comparing it with the critical value (h^*). The critical leverage value (h^*) is the edge of the applicability domain where the presence of any value beyond this critical value (i.e., structural outliers) indicates the unreliability in the prediction.⁸⁸ For instance, molecules with lower h_i values ($h_i < h^*$) indicate the similarity of the predicted molecule to that of the training set. In contrast, h_i values higher than the critical value ($h_i > h^*$) correspond to molecules that are "chemically different" to that of the molecules in the training set, and the prediction of these molecules could be considered somewhat unreliable due to the substantial extrapolation degree. The leverage value can be calculated as follows^{89,90}

$$h_i = v_i(V^T V)^{-1} \times v_i^T \quad (7)$$

where v_i represents a matrix consisting of the significant descriptors of ES i with dimensions of $1 \times d^*$ where d^* denotes the number of significant descriptors in the model. V is a $p \times d^*$ matrix where p denotes the number of experimental data points, and h_i is the leverage value of each experimental

data point.^{89,90} Likewise, the critical leverage value can be calculated by^{89,90}

$$h^* = \frac{3(d^* + 1)}{p} \quad (8)$$

To visualize the domain of applicability of a certain model, the William plot is used to plot the standardized residuals (SDR) versus the leverage values. In the William plot, the applicability domain boundaries range between $-3 < \text{SDR} < +3$ and $0 < h_i < h^*$. The standardized residuals can be calculated by⁹¹

$$\text{SDR} = \frac{y_{\text{pred}} - y_{\text{exp}}}{\sqrt{\frac{\sum_{m=1}^p (y_{\text{pred}} - y_{\text{exp}})^2}{p}}} \quad (9)$$

where y_{exp} and y_{pred} are the experimentally determined and the model predicted pH values, respectively. The coverage of the AD structural range could be evaluated based on the number of ESs within the AD compared to the outliers in a William plot by

$$\text{AD}_{\text{coverage}} = \frac{p_{\text{inside}}}{p_{\text{total}}} (100) \quad (10)$$

where p_{inside} represents the points contained inside the domain, while p_{total} represents the total number of data points.

3. RESULT AND DISCUSSIONS

3.1. Literature Review. **3.1.1. Concept of pH in ESs.** The knowledge of a solvent's pH is vital in designing many industrial processes and optimizing their operating conditions.^{3,36,92,93} In this section, the data sets collected from the literature (Table 1) were analyzed to study and identify the structural factors influencing the pH value of the ESs. As for the definition of pHs in ESs, it has been previously reported by Abbott et al.⁹⁴ that the concept of pH in ILs and ESs is poorly understood. According to their discussion, the pHs of ILs and ESs are based on the ability of the ES's cation, anion, and HBD to act as proton acceptors and proton donors in the same way that any other molecular liquid can. The protonation reaction of the ESs can be described as follows



where HA and Y represent the ES constituents. If water is present in the ES mixture, then an additional reaction would be observed as follows



For a more detailed description of the concept of pHs in ESs and ILs, the reader is referred to the discussion paper by Abbott et al.⁹⁴

3.1.2. Interpretation of Experimental Trends. In terms of the temperature effect, as expected, all the ESs in Table 1 show a linear acidity behavior; a decrease in the pH value when increasing the temperature, except for some ESs, e.g., MTPPB:TFA (1:8) and BTPC:Gly (1:5) where the increase in the temperature increases the pH values.⁷⁶ As for the effect of water content, it can be observed from Table 1 that the increase in the water content had a dual effect: either increasing the acidity of the ESs (e.g., ChCl:Pro:MA:H₂O or Bet:CA:H₂O ESs) or decreasing it (e.g., ChCl:CA:H₂O or Pro:MA:H₂O).

Regarding the effect of the ES's structure on pH values, the following was found: by keeping the HBD constant, i.e.,

triethylene glycol (TEG), the acidity of ATPPB:TEG:H₂O (1:4:0.18) was much higher than TPAB:TEG (1:4). This could be attributed to the aromaticity of the ATPPB and the addition of water.⁷⁵ Also, when comparing EAC:Gly (1:4)⁷⁴ and TBAC:Gly (1:4),⁷⁸ the pH values of TBAC:Gly (1:4) were higher than EAC:Gly (1:4). It can be seen that the increase in the number of alkyl chains increases the value of pH.

Considering the effect of HBD by keeping choline chloride (ChCl) constant, it can be seen that ESs with organic acid³⁹ HBDs (i.e., citric acid, glycolic acid, lactic acid, malic acid, malonic acid, and oxalic acid) had the highest acidity, followed by polyols³⁹ HBDs (e.g., ethylene glycol, glycerol) > sugars^{35,95} (e.g., fructose and glucose) > and finally amines⁸⁴ (e.g., ethanolamine and diethanolamine). Moreover, for the same HBD family, i.e., polyols,⁷⁵ it can be seen that the TEG, which contains three hydroxyl groups, showed higher acidity compared to DEG with two hydroxyl groups when comparing ATPPB:TEG:H₂O (1:10:0.35) to ATPPB:DEG:H₂O (1:10:0.31). Likewise, for constant HBA, i.e., TBAC, ethylene glycol (EG) was more basic compared to glycerol (Gly) when comparing TBAC:EG (1:3) to TBAC:Gly (1:3), which could be explained again based on the number of hydroxyl functional groups attached to each molecule.⁷⁸ Additionally, for amine HBDs, the primary amines (EA) showed higher basicity (high pH) compared to secondary (DEA) and tertiary (MDEA) amines.⁸⁴ From the data set available, it can be said that the nature of the HBD was significantly influencing the acidity of the ES relative to the effect of HBAs.^{76,77}

Another important feature is the molar ratio, which plays a significant role in altering the properties of ESs. It can be observed that the increase in the molar ratio had inconsistent trends. For example, the increase in the molar ratio of TPAB:Gly from 1:2 to 1:4 increases the acidity of the ES.³⁴ Conversely, increasing the molar ratio of ATPPB:DEG:H₂O⁷⁵ from 1:4:0.17 to 1:10:0.31 reduces the acidity of the ES. It should be noted that the further increase in the molar ratio of ATPPB:DEG:H₂O to 1:16:0.39 had an insignificant effect on the pH. This behavior could be attributed to the pH value of the ES converging toward the pH of the pure HBD (pH ≈ 7) in addition to the increase in the water content.⁷⁵ The same behavior was observed for ATPPB:TEG:H₂O, EAC:Gly:H₂O, and ChCl:Fru.³⁵ Other ESs showed an insignificant change in the pH when the molar ratio is increased as in the case of increasing the molar ratio of ChCl:LaA:H₂O from 1:5:0.67 to 1:10:1.22³⁹ and the increase in the molar ratio of TBAC:EG from 1:2 to 1:4.⁷⁸ Other ESs showed exceptional changes in their pH value when changing the HBA molar ratio. For instance, increasing the molar ratio of ChCl:Glu from 1:1 to 1.5:1 increased the basicity of the ES; however, the basicity is decreased with a further increase in the molar ratio to 2:1, yet increased again at a molar ratio of 2.5:1.⁹⁵ The same behavior was observed with TBAC:Gly.⁷⁸ Thus, it can be concluded that no clear trend can be deduced by changing the molar ratio. The observed behavior could be attributed to the fact that the pH property significantly depends on the molecular-level interactions between the HBA and HBD, and therefore, each case should be studied independently.

3.2. Physical Meaning of σ -Profiles. Figures 2 and 3 show the 2D chemical structures and the 3D geometrically optimized COSMO-RS molecular structures of the HBAs and HBDs. The green segments denote the nonpolar “neutral” parts of the molecules. The red segments denote the negatively

charged “hydrogen-accepting” parts of the molecule. Conversely, the blue segments denote the positively charged “hydrogen-donating” parts of the molecule.

Utilizing the chemical structures generated by COSMO-RS, the σ -profile of each HBA and HBD was determined. Figure 4 shows the developed σ -profile curves of the 9 HBAs and 21 HBDs categorized as follows: (a) salts, (b) amines and water, (c) fatty acids and amino acids, and d) polyols and sugars.

The importance of analyzing the σ -profile is that it describes the surface polarity of a molecule and gives the chemical information required in predicting the dispersal, electrostatic, and hydrogen-bonding interactions between the constituents of a mixture.⁹⁶ From Figure 4, depending on the density charge, the curves can be split into three primary regions; the positive polarity surface “HBD region” covering the σ range from -0.02500 up to -0.00625, the nonpolar region denoted by the σ range of -0.00625 < σ < +0.00625, and the negative polarity surface “HBA region” with a σ range of +0.00625 < σ < +0.02500.

The added value of analyzing the σ -profile is that the nature of atoms and their concentration in each molecule can be detected. In Figure 4, the σ -profile curves were analyzed, and the position of each atom was observed as the following. In the HBD region, the positively charged hydrogen atoms (H⁺) were found at the -0.02500 < σ < -0.01250 region. The atoms found in the weak donor region (nearby the vertical dashed lines) were the ammonium (N⁺) and phosphonium (P⁺) cations located in the -0.01250 < σ < -0.00625 region. The alkyl groups (-CH₃, -CH₂, and -CH) were found in the nonpolar region between -0.00625 < σ < +0.00625. The double bond atoms (C=C) in the aromatic group and the carbons in the carbonyl group (C=O) were positioned in a weak acceptor region of the +0.00625 < σ < +0.01250 range. The HBA region covered oxygen (O⁻) and nitrogen (N⁻) atoms that belong to the O-H and N-H groups located in the +0.01250 < σ < +0.01875 area, and fluoride (F⁻), chloride (Cl⁻), and bromide (Br⁻) anions were found in the +0.01875 < σ < +0.02500 region. Table 2 summarizes the observed atomic contributions in the σ range.⁹⁷

3.3. MLR Model. **3.3.1. Model Development.** In the early stages of developing this work, the authors have attempted to create one general MLR model capable of predicting the pH of all 41 ESs. However, in the model evaluation step, the results were not promising (further discussion is available in Supporting Information). This behavior is presumably due to three main reasons: (1) The number of experimental points available in the literature may not be sufficient enough to create one general linear model. (2) The sensitivity of the pH property, where it is generally governed from values of -2 up to +14, and (3) the pH property between various types of different ES families could be exhibiting a nonlinear relationship. Therefore, the data set was split into two subsets that are “more specific” to certain families of ESs. The two categories were created based on the families of the HBDs as it was found that the nature of the HBD was much more pronounced relative to the effect of HBAs (see Section 3.1). The first category denoted as “Family A” consists of the ESs that are more acidic in nature such as fatty acids and amino acids. Sugars were also added to the Family A group as many ESs in Table 1 are combinations of acids and sugars (such as MA:Glu and MA:Suc). On the other hand, the other relatively less acidic ESs with higher pH values were grouped in Family B, which consists of amines and polyols. Subsequently, the Family

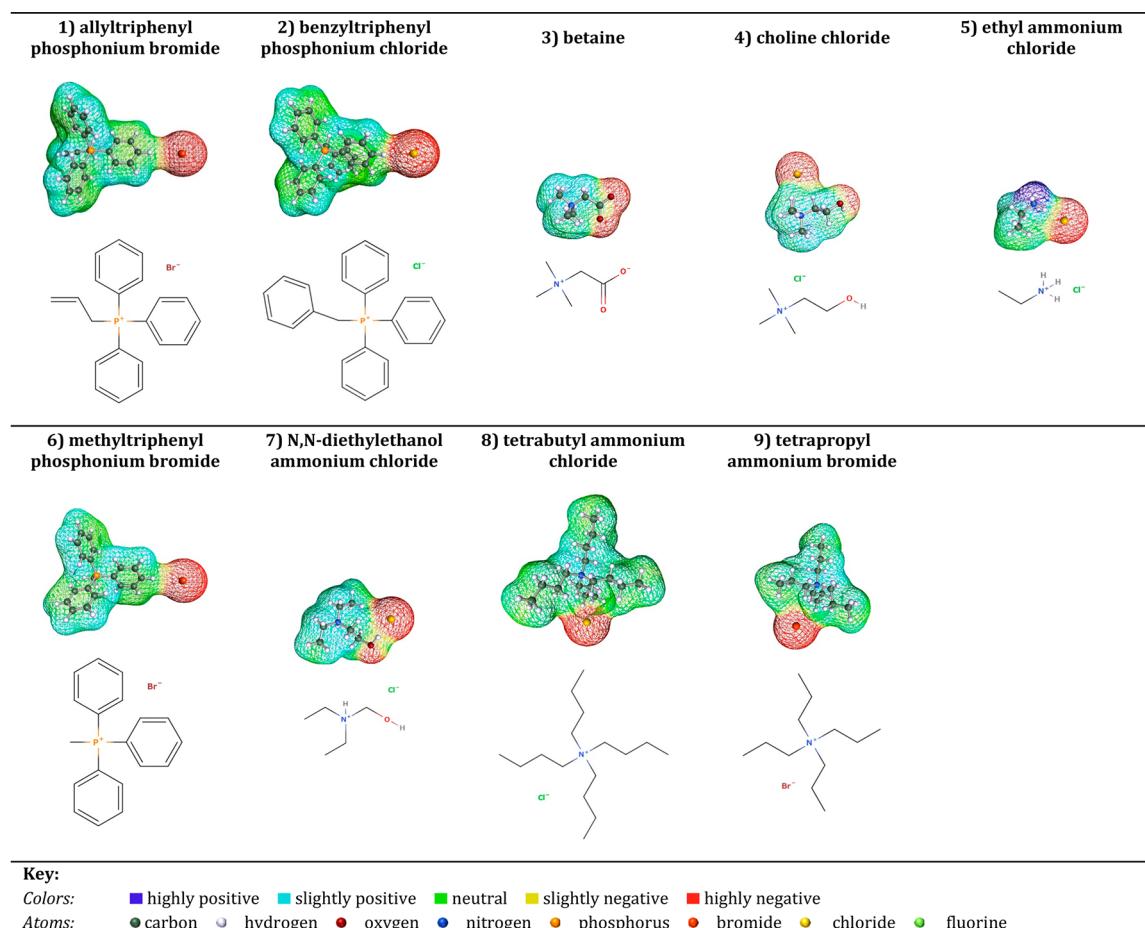


Figure 2. 3D and 2D molecular structures of the nine modeled hydrogen bond acceptors (HBAs).

A set contained 21 ESs with 24 compositions, and the Family B set contained 20 ESs with 19 compositions. These divisions were expected to improve the model's predictive power in estimating the pH of the ESs.⁹⁸ Accordingly, the Family A set consisted of 313 experimental points, while the Family B set consisted of 335 points presenting an extensive range of pH measurements. The divisions of the data set for both Family A and Family B models are shown in Table 3.

Consequently, the prediction of new ESs should be made exclusively based on the nature of the HBD, where if the HBD constituent of the ES is a fatty acid, amino acid, hydroxy acid, or sugar then the Family A model must be utilized. On the other hand, if the HBD constituent of the ES is an amine or a polyol, then the Family B model must be utilized instead. Otherwise, if the HBD constituent at hand belongs to neither family, then the developed models would not be applicable to predict their pH behavior.

To test the predictivity of the proposed models, the data set in each model was split into a training set and a testing set. The data in the training set were utilized in the development of the model, whereas the data in the testing set were only used to externally evaluate the predictivity of the model. The testing set was fixed to 75% of the ESs, and the remaining 25% were used as the testing set. The selection of the ESs in the testing set was done by first choosing representative external molecules for each family. For instance, in the Family A set, DEEAC was selected as an HBA representative, MalA and OxaA as acid representatives, and Suc as a sugar representative. Conversely,

for Family B, BTPC was selected as an HBA representative, DEA as an amine representative, and DEG as a polyol representative. Accordingly, all ESs composed of DEEAC, BTPC, MalA, OxaA, Suc, DEA, and DEG were taken for external validation. The selection of these molecules was done based on the number of their data points, where, for instance, only two ESs were composed of BTPC. If, for instance, ChCl would have been chosen as an external representative, then 18 ESs would have to be excluded from the training set, which would not be practical for the development of a machine learning-based model. To reach the 25% testing set threshold, the "ordered response" technique was utilized.⁹⁹ In this technique, first, the pH values of all ESs at room temperature were organized from lowest to highest. Then, one ES out of each seven ESs of the Family A set and one ES out of each nine ESs of the Family B set were added into the testing set to cover the 25% ES threshold. These 25% ESs in the testing set were not used to develop the models and were only utilized to test the final models. Accordingly, the Family A test set consisted of 11 out of 45 ES compositions, while the Family B test set consisted of 10 out of 39 ES compositions. The performance of the models in predicting the testing set was evaluated based on the external regression coefficient (R^2_{external}) and various other statistical parameters.

3.3.2. S_{σ} -profiles Discretization. The σ -profile of most molecules has been reported to contain 51 points within the range of $\pm 0.025 \text{ e}/\text{\AA}$.¹⁰⁰ In this sense, the discretization of the σ -profile curves into 4, 6, 8, 10, and 12 S_{σ} -profile descriptors has

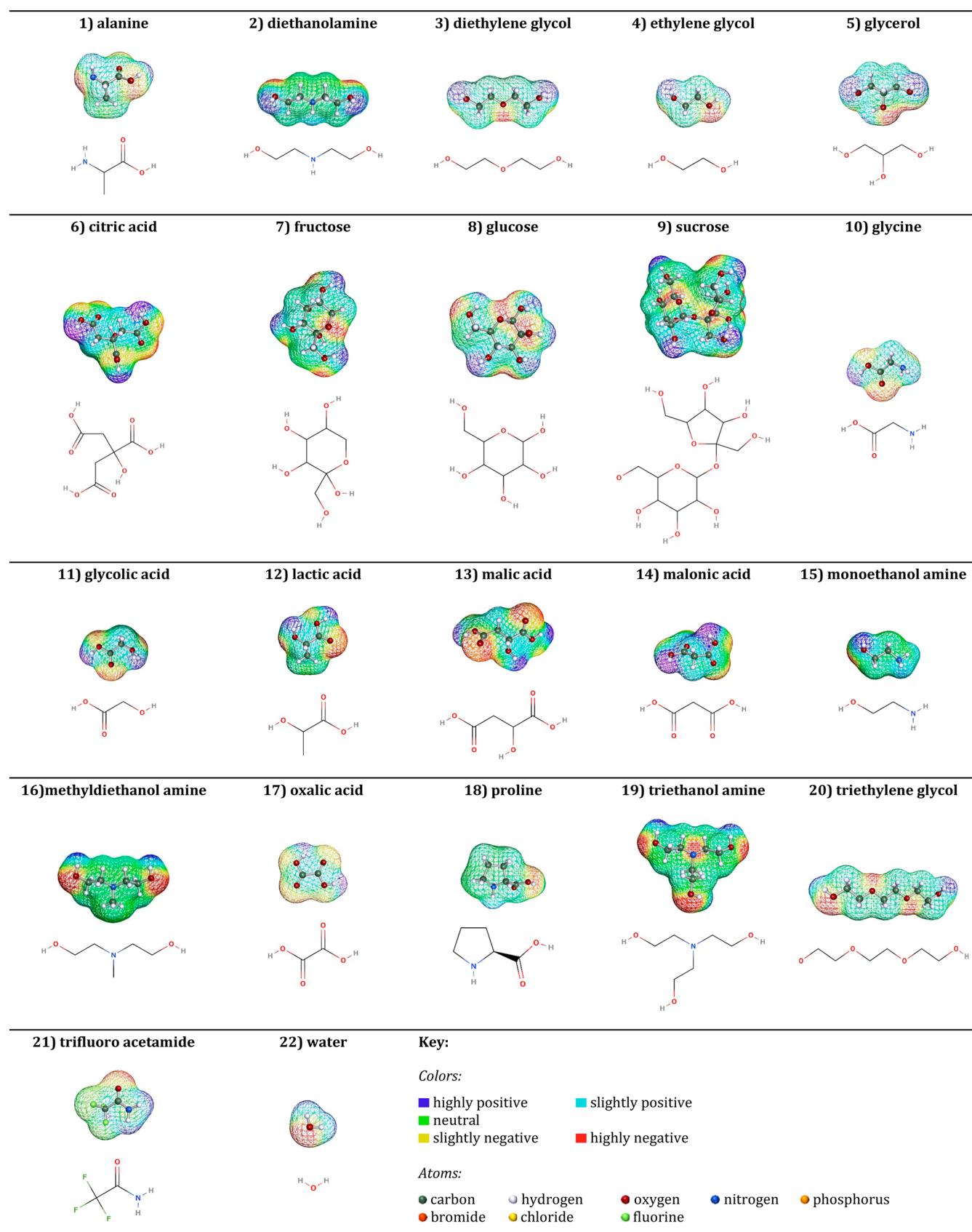


Figure 3. 3D and 2D molecular structures of the 21 modeled hydrogen bond donors (HBDs) and water.

been systematically investigated in this work. Figure 5 represents an example of the discretized S_{σ} -profile descriptors for ethylene glycol.

As the discretization of the S_{σ} -profile increases, the accuracy of the developed models is also significantly increased as a result of more fitting parameters. Nevertheless, that would

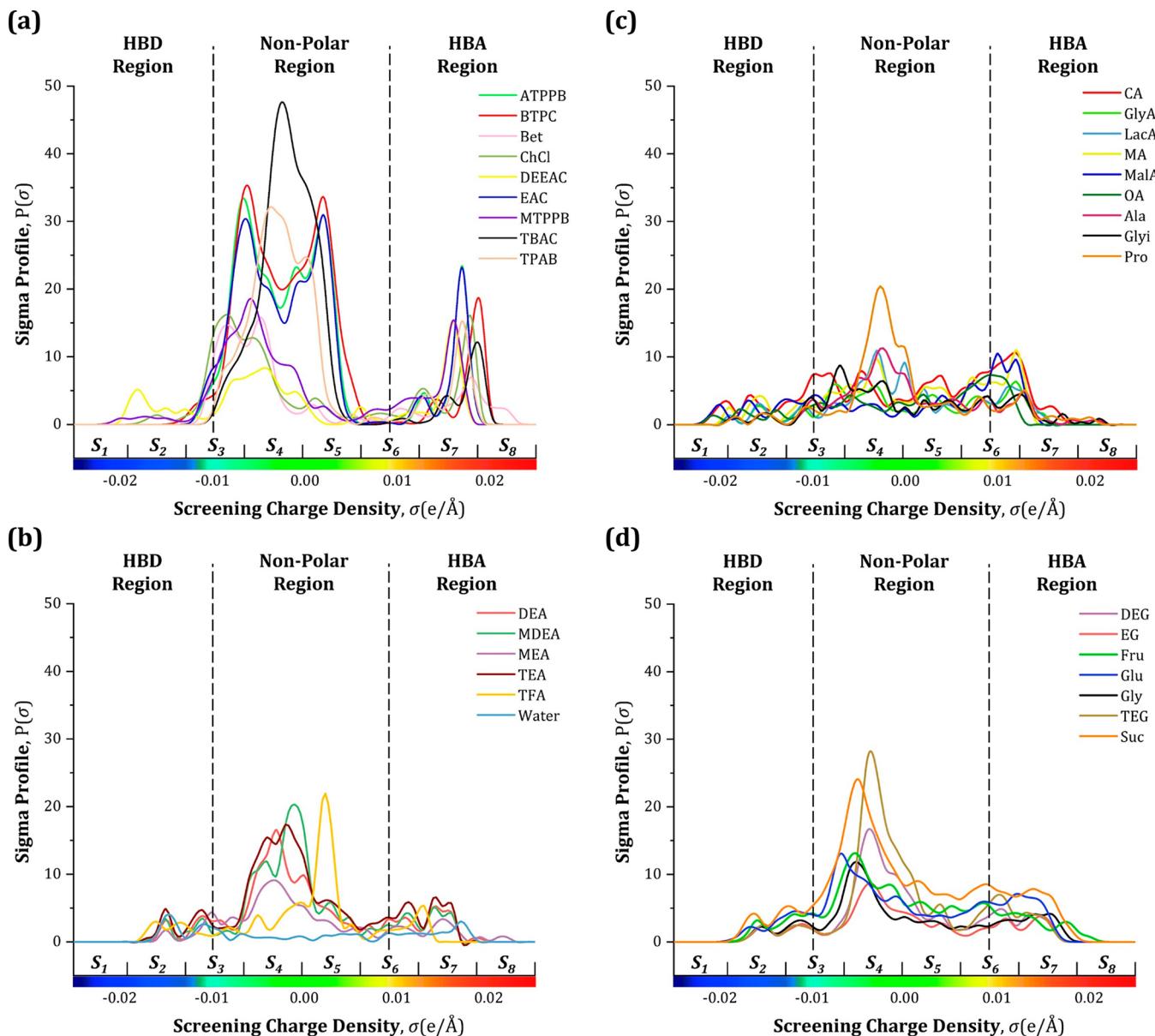


Figure 4. Calculated σ -profile of the 31 constituents modeled as (a) salts, (b) amines and water, (c) fatty acids and amino acids, and d) polyols and sugars. The abbreviations are available in the [Nomenclature](#) section.

Table 2. Partitioning of Atomic Contributions Contained within the σ Range

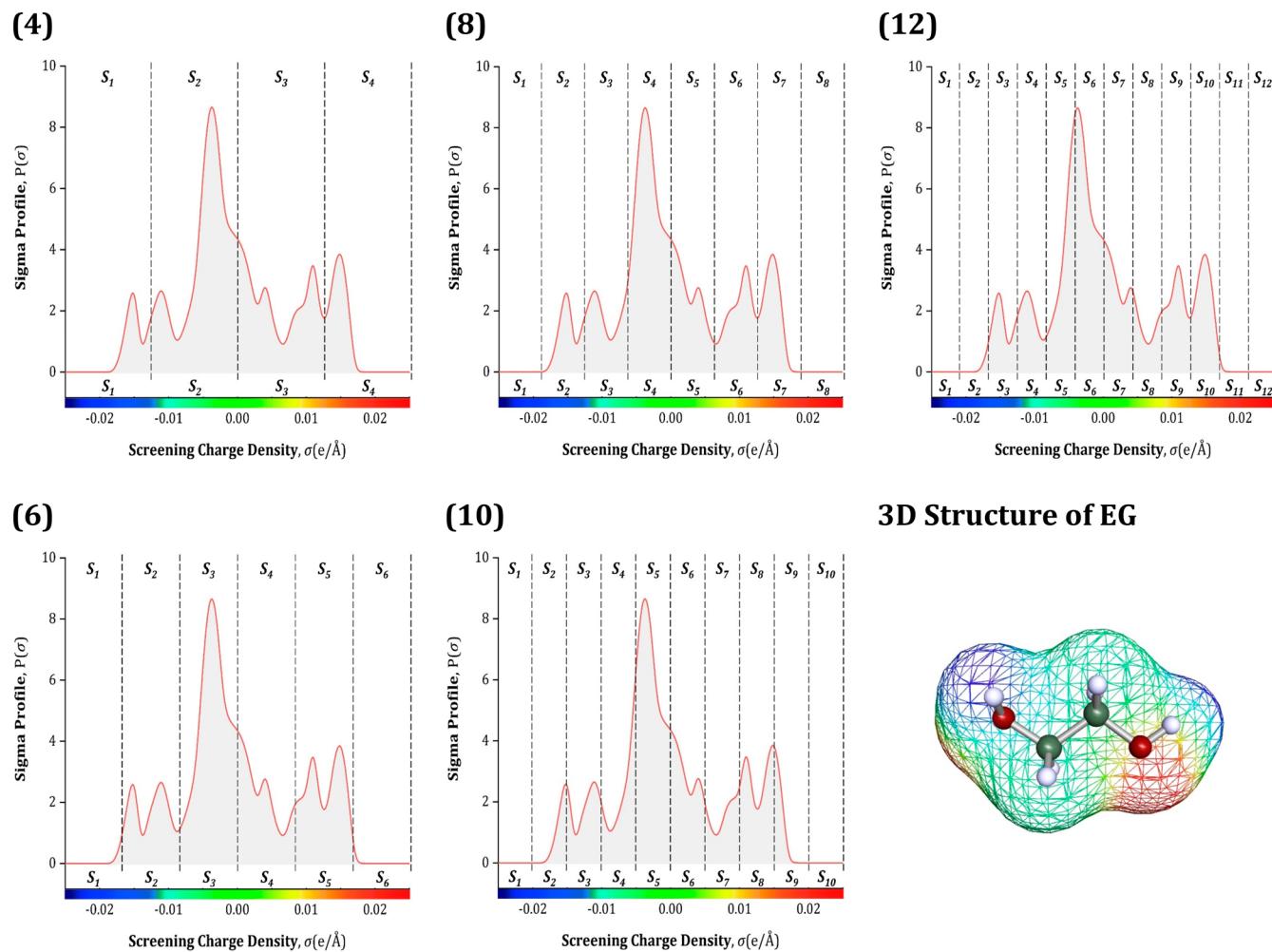
Descriptors	σ Values ($e/\text{\AA}$)	Atoms contained
Hydrogen bond donor region		
S_1	$-0.02500 < \sigma < -0.01875$	H^+ of COOH
S_2	$-0.01875 < \sigma < -0.01250$	H^+ of O–H and N–H
Weak donor region		
S_3	$-0.01250 < \sigma < -0.00625$	$\text{C}^{\delta+}, \text{H}^{\delta+}, \text{N}^+, \text{P}^+$
Nonpolar region		
S_4	$-0.00625 < \sigma < 0.00000$	C and H of $-\text{CH}_3, -\text{CH}_2$, and $-\text{CH}_3$
S_5	$0.00000 < \sigma < +0.00625$	
Weak acceptor region		
S_6	$+0.00625 < \sigma < +0.01250$	$\text{C}=\text{C}^{(\delta-)}, \text{C}=\text{O}^{(\delta-)}$
Hydrogen bond acceptor region		
S_7	$+0.01250 < \sigma < +0.01875$	$\text{O}^{\delta-}$ and $\text{N}^{\delta-}$ of O–H and N–H
S_8	$+0.01875 < \sigma < +0.02500$	$\text{Cl}^-, \text{Br}^-, \text{F}^-$

also lead to an increase in the complexity of the resulting model. Thus, a compromise should be made between the

number of fitting parameters and the accuracy of the developed model. [Figure 6](#) presents a heatmap of the regression coefficient

Table 3. Partitioning of Family A and Family B Data Sets

	Family A	Family B
Total number of HBA	6	7
Total number of HBD	8	7
Total compositions	45	39
Training ESs	33	29
External ESs	12	10
Data points	313	335
Training points	222	251
External points	91	84
Training ESs	ESs 8.2, 8.3, 9, 9.2, 9.3, 11, 12, 12.1, 13, 13.1, 22, 23, 23.1, 34, 35, 35.1, 36, 36.1, 37, 37.1, 37.2, 38, 38.1, 38.2, 39, 39.1, 39.2, 40, 40.1, 40.2, 41.1	ESs 2, 7, 10, 15, 16, 18, 20, 20.1, 20.2, 25, 26, 27, 28, 28.1, 28.2, 29, 29.1, 29.2, 30, 30.1, 30.2, 31, 31.1, 32, 32.1, 32.2, 33, 33.2
External ESs	ESs 8.1, 9.1, 14, 17, 19, 21, 24, 35.2, 36.2, 41, 41.2	ESs 1, 1.1, 1.2, 2.1, 2.2, 3, 4, 6, 30.3, 31.2, 33.1
External molecules	DEEAC, MalA, OxaA, Suc	BTPC, DEA, DEG

**Figure 5.** Representation of the discretized S_{σ} -profile descriptors in 4, 6, 8, 10, and 12 segments for ethylene glycol as an example.

(R^2) and the number of fitting parameters based on the partitioning of the S_{σ} -profile for the Family A and Family B models. It can be observed that as the partitioning increases to 12 segments, the R^2 values are also improved up to 0.8970 and 0.8580 for Family A and Family B, respectively. However, these R^2 values may not be considered good enough to facilitate solvent screening studies. These R^2 values also indicate that the machine learning algorithm requires more statistically significant descriptors in order to represent the data more accurately.

Therefore, in order to improve the performance of the models, the addition of more descriptors (other than S_{σ} -profiles) or complexity terms between S_{σ} -profile descriptors could be utilized. However, if more descriptors (such as ES critical properties for instance) were utilized, then the “coherence” of the regression would be damaged as a result of the addition of these macroscopic descriptors with the microscopic S_{σ} -profile descriptors. Other than that, the S_{σ} -profile molecular descriptors were selected as they have a sound physical basis and can easily be obtained through basic COSMO-RS modeling. Other types

of descriptors may not have the same physical basis that S_{σ} -profiles have and may not be as easy to calculate and obtain. Additionally, most ES descriptors assume the ES to be a pseudo-pure component, meaning that the descriptors need to be calculated for each ES individually, unlike the methodology used in this work, where the S_{σ} -profile descriptors are a special type of parameter calculated from an atomic basis.¹⁰¹ Thus, it takes into account the charged contribution and the relative concentration of each atom that constitutes the “ES mixture”, which is very useful in the field of ESs as it removes the need to define the ES as a pseudo-pure component, and it easily allows for defining mixtures of ESs with their water content, which are critical in pH studies. Thus, to avoid these issues, the addition of complexity terms between S_{σ} -profile descriptors has been utilized in this work.

The complexity terms selected in this work are based on second degree factorials, which can be utilized to investigate the interactions between a pair of descriptors. The binary interaction terms can be categorized into two main groups: (1) the binary interactions between temperature and each molecular surface descriptor denoted as $\{S_i, T\}$ and (2) the binary interactions between two molecular surface descriptors denoted as $\{S_i, S_j\}$. The two types of binary interactions can be calculated through basic multiplications of a pair of descriptors expressing their binary interaction (shown in eq 2).

The R^2 values, after including the binary interaction terms, are shown in the heatmap in Figure 6. It can be observed that the addition of the interaction terms increased the performance of the developed models. Nevertheless, the complexity of the models also increased as a result. On the basis of the obtained heatmap, the 8-partition model including the binary interactions between the descriptors has been selected as the optimal compromise between accuracy and fitting parameters as it is the simplest model with a regression coefficient above $R^2 > 0.99$. Presumably, the discretization of the σ -profile into eight segments is found to be sufficient enough to reasonably characterize the effect of all the functional groups that exist within the ESs. Using the 8-partition profile, the σ -profile curve can be classified into five main regions based on their charges: HBD region, weak donor region, nonpolar region, weak acceptor

region, and HBA region with the representative descriptors being $[S_1, S_2]$, $[S_3]$, $[S_4, S_5]$, $[S_6]$, and $[S_7, S_8]$, respectively. The chemical information of strong-, regular-, and weak-donating functional groups is stored in S_1 , S_2 , and S_3 , respectively, while the chemical information of strong-, regular-, and weak-accepting functional groups is stored within S_8 , S_7 , and S_6 , respectively.

3.3.3. MLR Model Evaluation. To develop the Family A MLR model, 33 ES compositions with their mixtures were utilized in the development of the model (i.e., training), and the remaining 12 ES compositions with their mixtures were used to test the predictivity of the model (i.e., testing set). The initial step in evaluating the model is to check the impact of each descriptor and its binary interactions on the model. For that reason, an analysis of variance study was conducted, and the results are shown in Table 4.

In Table 4, it can be seen that the intercept and all the eight molecular descriptors with the temperature descriptor had a significant influence on the model development as their P_{values} were less than 5%. The effect of the binary interactions on the model was also analyzed, and it was found that the interactions between $\{S_{6,+}, T\}$ and $\{S_{7,+}, T\}$ molecular descriptors and the temperature descriptor were insignificant on the Family A model. The same findings were obtained for the binary interactions between some pairs of molecular descriptors where the $\{S_{2,-}, S_{3,-}\}$, $\{S_{2,-}, S_{4,-}\}$, $\{S_{2,-}, S_{5,+}\}$, $\{S_{4,-}, S_{5,+}\}$, and $\{S_{5,+}, S_{6,+}\}$ pairs did not affect the model. The remaining binary interactions showed a significant effect on the model with P_{values} less than 5% except for the $\{S_{1,+}, T\}$ and $\{S_{5,+}, S_{8,+}\}$ descriptors where their impact was considered to be less pronounced as their P_{values} were greater than 5% at 11.25% and 6.71%, respectively. Nonetheless, these descriptors were still considered influential by minimizing the corrected Akaike information criterion (AIC_c), indicating that the descriptors contain valuable information from a statistical viewpoint and that they improve the model more than they would be expected to by chance. Therefore, on the basis of the stepwise fitting algorithm and the analysis of variance study, the resultant Family A model can be expressed as follows

$$\begin{aligned} \text{pH}_{\text{Family A}} = & 39.31 + 9401.19(S_1) - 13070.83(S_2) - 3324.20(S_3) - 1346.91(S_4) + 435.42(S_5) + 4992.10(S_6) + 8326.25(S_7) \\ & + 37435.97(S_8) - 1.29 \times 10^{-2}(T) - 35370219.00(S_1 - \bar{S}_1)(S_2 - \bar{S}_2) - 7380316.00(S_1 - \bar{S}_1)(S_3 - \bar{S}_3) \\ & - 105501.00(S_1 - \bar{S}_1)(S_4 - \bar{S}_4) + 9001733.20(S_1 - \bar{S}_1)(S_5 - \bar{S}_5) + 1853102.80(S_1 - \bar{S}_1)(S_6 - \bar{S}_6) + 12118890.00(S_1 - \bar{S}_1)(S_7 - \bar{S}_7) \\ & - 56305137.00(S_1 - \bar{S}_1)(S_8 - \bar{S}_8) + 10.29 \times 10^{-1}(S_1 - \bar{S}_1)(T - \bar{T}) + 1272799.70(S_2 - \bar{S}_2)(S_6 - \bar{S}_6) - 1329440.00(S_2 - \bar{S}_2)(S_7 - \bar{S}_7) \\ & - 32845561.00(S_2 - \bar{S}_2)(S_8 - \bar{S}_8) + 34.44 \times 10^{-1}(S_2 - \bar{S}_2)(T - \bar{T}) + 72611.93(S_3 - \bar{S}_3)(S_4 - \bar{S}_4) - 2685473.00(S_3 - \bar{S}_3)(S_5 - \bar{S}_5) \\ & + 2047090.80(S_3 - \bar{S}_3)(S_6 - \bar{S}_6) - 84229.96(S_3 - \bar{S}_3)(S_7 - \bar{S}_7) - 9971815.00(S_3 - \bar{S}_3)(S_8 - \bar{S}_8) - 2.95 \times 10^{-2}(S_3 - \bar{S}_3)(T - \bar{T}) \\ & - 268050.70(S_4 - \bar{S}_4)(S_6 - \bar{S}_6) - 247666.60(S_4 - \bar{S}_4)(S_7 - \bar{S}_7) - 5042131.00(S_4 - \bar{S}_4)(S_8 - \bar{S}_8) + 5.69 \times 10^{-1}(S_4 - \bar{S}_4)(T - \bar{T}) \\ & + 7438614.10(S_5 - \bar{S}_5)(S_7 - \bar{S}_7) - 3346227.00(S_5 - \bar{S}_5)(S_8 - \bar{S}_8) - 28.99 \times 10^{-1}(S_5 - \bar{S}_5)(T - \bar{T}) - 4790982.00(S_6 - \bar{S}_6)(S_7 - \bar{S}_7) \\ & + 26080014.00(S_6 - \bar{S}_6)(S_8 - \bar{S}_8) + 11958774.00(S_7 - \bar{S}_7)(S_8 - \bar{S}_8) - 20.20 \times 10^{-1}(S_8 - \bar{S}_8)(T - \bar{T}) \end{aligned} \quad (13)$$

where S_i and \bar{S}_i are the molecular descriptors and their mean values ($\text{e}/\text{\AA}^2$), respectively, and T is the temperature (K).

After generating the Family A model, its performance was analyzed, and the results are summarized in Table S.5. The model showed excellent performance in terms of fitting the experimental data for the pH property with high regression

coefficient ($R^2 = 0.9947$). Moreover, the internal robustness of the model was evaluated based on the values of the cross-validation coefficients. The high cross-validation coefficients (Q^2) reflects the stability of the model. Furthermore, several chance correlation tests have been conducted. The low y-scrambling regression coefficient ($R^2_{\text{scramble}} = 0.0072$) indicates that the

	4 Partitions (S ₁ -S ₄)	6 Partitions (S ₁ -S ₆)	8 Partitions (S ₁ -S ₈)	10 Partitions (S ₁ -S ₁₀)	12 Partitions (S ₁ -S ₁₂)
Including Interactions	R ² =0.9191	R ² =0.9644	R²=0.9947	R ² =0.9974	R ² =0.9975
	15 coefficients	26 coefficients	38 coefficients	53 coefficients	51 coefficients
Excluding Interactions	R ² =0.7664	R ² =0.8705	R ² =0.8548	R ² =0.8851	R ² =0.8970
	5 coefficients	7 coefficients	9 coefficients	11 coefficients	13 coefficients
(b)	4 Partitions (S ₁ -S ₄)	6 Partitions (S ₁ -S ₆)	8 Partitions (S₁-S₈)	10 Partitions (S ₁ -S ₁₀)	12 Partitions (S ₁ -S ₁₂)
	R ² =0.6767	R ² =0.9429	R²=0.9969	R ² =0.9964	R ² =0.9981
Excluding Interactions	R ² =0.4912	R ² =0.6138	R ² =0.6390	R ² =0.8036	R ² =0.8580
	5 coefficients	7 coefficients	9 coefficients	10 coefficients	12 coefficients

Figure 6. Heatmap of the regression coefficient (R^2) in several MLR models based on descriptor interactions and partitioning of S_{σ} -profile for the (a) Family A model and the (b) Family B model.

Table 4. Coefficients of 38 Significant Descriptors and Intercept of the Family A Model

Category	Term	Estimate	Standard error	t _{Ratio}	P _{value}
Intercept	a ₀	39.31	14.78	2.66	0.0083
HBD	S ₁	9401.19	4821.12	1.95	0.0465
	S ₂	-13070.83	3251.45	4.02	<0.0001
Weak donor	S ₃	-3324.20	972.23	3.42	0.0007
Nonpolar	S ₄	-1346.91	496.74	2.71	0.0071
	S ₅	435.42	219.90	1.98	0.0455
Weak acceptor	S ₆	4992.10	1509.15	3.31	0.0011
HBA	S ₇	8326.25	2334.65	3.57	0.0004
	S ₈	37435.97	9315.42	4.02	<0.0001
Temperature	T	-1.29 × 10 ⁻²	8.51 × 10 ⁻⁴	15.15	<0.0001
{HBD, Temperature}	(S ₁ -S̄ ₁)(T-T̄)	10.29 × 10 ⁻¹	6.86 × 10 ⁻¹	1.50	0.1125
	(S ₂ -S̄ ₂)(T-T̄)	34.44 × 10 ⁻¹	9.69 × 10 ⁻¹	3.55	0.0004
{Weak donor, Temperature}	(S ₃ -S̄ ₃)(T-T̄)	-2.95 × 10 ⁻²	1.536 × 10 ⁻²	1.92	0.0482
{Nonpolar, Temperature}	(S ₄ -S̄ ₄)(T-T̄)	5.69 × 10 ⁻¹	2.50 × 10 ⁻¹	2.28	0.0236
	(S ₅ -S̄ ₅)(T-T̄)	-28.99 × 10 ⁻¹	6.67 × 10 ⁻¹	4.35	<0.0001
{HBA, Temperature}	(S ₈ -S̄ ₈)(T-T̄)	-1.29 × 10 ⁻²	9.77 × 10 ⁻¹	2.25	0.0271
{HBD, HBD}	(S ₁ -S̄ ₁)(S ₂ -S̄ ₂)	-35370219.00	12580221.00	2.81	0.0053
{HBD, Weak donor}	(S ₁ -S̄ ₁)(S ₃ -S̄ ₃)	-7380316.00	2627407.00	2.81	0.0053
{HBD, Nonpolar}	(S ₁ -S̄ ₁)(S ₄ -S̄ ₄)	-105501.00	32866.36	3.21	0.0015
	(S ₁ -S̄ ₁)(S ₅ -S̄ ₅)	9001733.20	3048929.00	2.95	0.0034
{HBD, Weak acceptor}	(S ₁ -S̄ ₁)(S ₆ -S̄ ₆)	1853102.80	661646.20	2.80	0.0055
	(S ₂ -S̄ ₂)(S ₆ -S̄ ₆)	1272799.70	434961.90	2.93	0.0037
{HBD, HBA}	(S ₁ -S̄ ₁)(S ₇ -S̄ ₇)	12118890.00	4277145.00	2.83	0.0049
	(S ₁ -S̄ ₁)(S ₈ -S̄ ₈)	-56305137.00	17131621.00	3.29	0.0011
	(S ₂ -S̄ ₂)(S ₇ -S̄ ₇)	-1329440.00	500806.20	2.65	0.0084
	(S ₂ -S̄ ₂)(S ₈ -S̄ ₈)	-32845561.00	11364365.00	2.89	0.0042
{Weak donor, Nonpolar}	(S ₃ -S̄ ₃)(S ₄ -S̄ ₄)	72611.93	33275.65	2.18	0.0299
	(S ₃ -S̄ ₃)(S ₅ -S̄ ₅)	-2685473.00	989645.80	2.71	0.0071
{Weak donor, Weak acceptor}	(S ₃ -S̄ ₃)(S ₆ -S̄ ₆)	2047090.80	731873.90	2.80	0.0055
{Weak donor, Nonpolar}	(S ₃ -S̄ ₃)(S ₇ -S̄ ₇)	-84229.96	32120.71	2.62	0.0092
	(S ₃ -S̄ ₃)(S ₈ -S̄ ₈)	-9971815.00	3085243.00	3.23	0.0014
{Nonpolar, Weak acceptor}	(S ₄ -S̄ ₄)(S ₆ -S̄ ₆)	-268050.70	94521.37	2.84	0.0049
{Nonpolar, HBA}	(S ₄ -S̄ ₄)(S ₇ -S̄ ₇)	-247666.60	100270.30	2.47	0.0141
	(S ₄ -S̄ ₄)(S ₈ -S̄ ₈)	-5042131.00	1805414.00	2.79	0.0056
	(S ₅ -S̄ ₅)(S ₇ -S̄ ₇)	7438614.10	2752236.00	2.70	0.0073
	(S ₅ -S̄ ₅)(S ₈ -S̄ ₈)	-3346227.00	1820323.00	1.84	0.0671
{Weak acceptor, HBA}	(S ₆ -S̄ ₆)(S ₇ -S̄ ₇)	-4790982.00	1723914.00	2.78	0.0058
	(S ₆ -S̄ ₆)(S ₈ -S̄ ₈)	26080014.00	7454203.00	3.50	0.0005
{HBA, HBA}	(S ₇ -S̄ ₇)(S ₈ -S̄ ₈)	24500561.00	7327952.00	3.34	0.0009

^aMean values: S̄₁ = 1.10 × 10⁻³; S̄₂ = 8.71 × 10⁻³; S̄₃ = 3.01 × 10⁻²; S̄₄ = 2.34 × 10⁻²; S̄₅ = 1.19 × 10⁻²; S̄₆ = 1.49 × 10⁻²; S̄₇ = 1.47 × 10⁻²; S̄₈ = 3.20 × 10⁻⁴; T̄ = 316.169.

model parameters were not correlated by chance. Similarly, the high Fisher statistics ($F_{\text{Ratio}} = 905.51$; $P_{\text{value, Fisher}} < 0.0001$) suggest that large variations due to systematic variances in the descriptors are exhibited by the model rather than differences caused by chance. This also can be supported by the high adjusted regression coefficient ($R^2_{\text{adjusted}} = 0.9936$), indicating that the descriptors enhanced the model more than it would be expected by chance.

Figure 7(a) shows a parity plot of the experimental and the predicted data in the training set. It can be seen that the model can be considered reliable in fitting the experimental data as all the points in training were lying close to the diagonal line showing minor dispersion ($R^2 = 0.9947$) and the root-mean-square error (RMSE) was calculated to be 0.1589. Figure 7(b) shows the parity graph plotted for the testing set. It can be seen

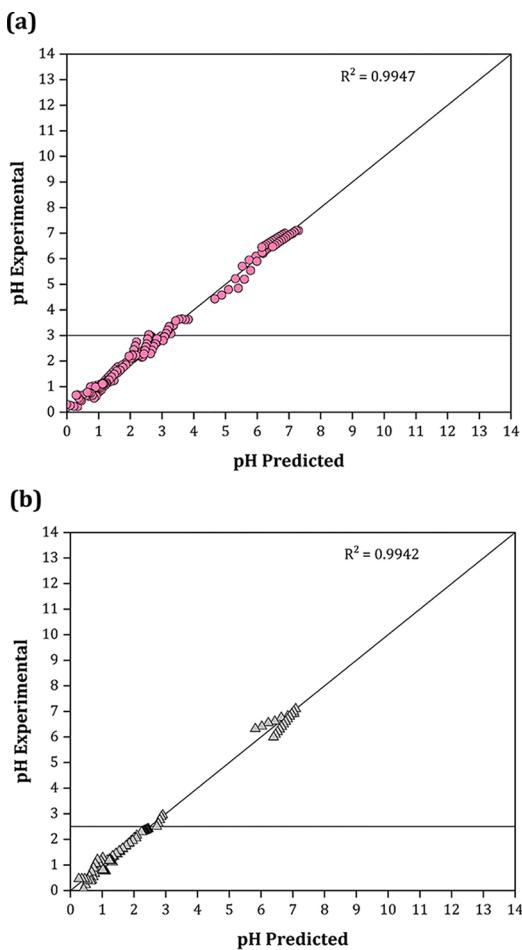


Figure 7. Parity graph of the experimental and predicted pH values of Family A's MLR model in (a) training and (b) external testings.

that the model was able to show high predictive capabilities of the ESs in the testing set ($R^2_{\text{external}} = 0.9942$ and RMSE = 0.1722).

For further model evaluation, the residual plot was utilized to examine the accuracy of the model. Figure 8 shows the

excellence of the proposed model in predicting the pH property of ESs where the residuals were at a range of ± 0.5 with an absolute-average relative deviation (AARD) of $\pm 7.54\%$ in training and an AARD of $\pm 17.38\%$ in testing. It should be mentioned that the reason the model shows a systematic linear deviation in both Figures 7 and 8 is that the pH model was developed for a variety of ES structures and their temperature dependence, which exhibits a linear behavior (as discussed in Section 3.1). This allows the model's field of investigation to include both the ES structure and the temperature, which is very important for screening new green and sustainable ESs with the required pH for a specific application.

Moreover, Figure 9 shows the performance of the Family A MLR model in predicting the pH values of the external molecule ES set as a function of temperature. It can be seen that the model predicts the external molecules of DEEAC: MalA (ES19) and MA:Suc:H₂O (ES24) quite accurately. However, the external performances of ChCl:MalA:H₂O (ES14) and ChCl:OxaA:H₂O (ES17) are a lot lower with standard deviations (SD_{avg}) of ± 0.18 and ± 0.25 , respectively.

In the second model, 39 amine- and polyol-based ES compositions were selected to develop the model, where 29 ES compositions were utilized for training, and the remaining 10 ES compositions were utilized in external validation. The model results are listed in Table 5.

In this model, 40 descriptors were observed to have a strong influence on the pH of ESs. The eight molecular descriptors had an impact on the model, as their estimated coefficients had $P_{\text{values}} < 5\%$. As for the binary interaction between each molecular descriptor and the temperature, it was observed that all the interactions pairs $\{S_i, T\}$ were found to be statistically significant and affecting the pH of ESs, except for $\{S_4, T\}$ where its effect was less pronounced and therefore excluded from the model. With regard to the interactions among two surface descriptors, it was observed that the pairs $\{S_1, S_8\}$ $\{S_2, S_6\}$, $\{S_4, S_6\}$, and $\{S_6, S_8\}$ did not affect the pH of ESs. The effect of all the molecular descriptors and their binary interactions on the developed model was confirmed by the low P_{values} , the high values of t_{Ratio} , and improving the AIC_c information criterion. Consequently, the resulting model can be expressed as follows

$$\begin{aligned}
 \text{pH}_{\text{Family B}} = & 2712.99 + 12195344.00(S_1) + 2007820.40(S_2) + 1625622.70(S_3) + 299995.39(S_4) - 1219162.00(S_5) \\
 & - 242091.30(S_6) - 2446456.00(S_7) - 3275372.00(S_8) - 9.91 \times 10^{-3}(T) + 26370000000.00(S_1 - \bar{S}_1)(S_2 - \bar{S}_2) \\
 & + 7141600000.00(S_1 - \bar{S}_1)(S_3 - \bar{S}_3) + 2172600000.00(S_1 - \bar{S}_1)(S_4 - \bar{S}_4) - 325700000.00(S_1 - \bar{S}_1)(S_5 - \bar{S}_5) \\
 & - 509500000.00(S_1 - \bar{S}_1)(S_6 - \bar{S}_6) - 7338000000.00(S_1 - \bar{S}_1)(S_7 - \bar{S}_7) - 40.15 \times 10^{-1}(S_1 - \bar{S}_1)(T - \bar{T}) \\
 & - 741500000.00(S_2 - \bar{S}_2)(S_3 - \bar{S}_3) - 46429937.00(S_2 - \bar{S}_2)(S_4 - \bar{S}_4) + 93297386.00(S_2 - \bar{S}_2)(S_5 - \bar{S}_5) \\
 & + 1716100000.00(S_2 - \bar{S}_2)(S_7 - \bar{S}_7) + 924971536.00(S_2 - \bar{S}_2)(S_8 - \bar{S}_8) + 51.57 \times 10^{-1}(S_2 - \bar{S}_2)(T - \bar{T}) \\
 & - 33295616.00(S_3 - \bar{S}_3)(S_4 - \bar{S}_4) + 38422954.00(S_3 - \bar{S}_3)(S_5 - \bar{S}_5) + 151526609.00(S_3 - \bar{S}_3)(S_6 - \bar{S}_6) \\
 & + 4999312.60(S_3 - \bar{S}_3)(S_7 - \bar{S}_7) + 1287300000.00(S_3 - \bar{S}_3)(S_8 - \bar{S}_8) - 8.32 \times 10^{-1}(S_3 - \bar{S}_3)(T - \bar{T}) \\
 & - 5079827.00(S_4 - \bar{S}_4)(S_5 - \bar{S}_5) + 77090308.00(S_4 - \bar{S}_4)(S_7 - \bar{S}_7) + 187576063.00(S_4 - \bar{S}_4)(S_8 - \bar{S}_8) \\
 & + 26647478.00(S_5 - \bar{S}_5)(S_6 - \bar{S}_6) - 66022050.00(S_5 - \bar{S}_5)(S_7 - \bar{S}_7) - 1191000000.00(S_5 - \bar{S}_5)(S_8 - \bar{S}_8) + 5.77 \\
 & \times 10^{-1}(S_5 - \bar{S}_5)(T - \bar{T}) - 382400000.00(S_6 - \bar{S}_6)(S_7 - \bar{S}_7) - 7.69 \times 10^{-1}(S_6 - \bar{S}_6)(T - \bar{T}) - 2196000000.00(S_7 - \bar{S}_7)(S_8 - \bar{S}_8) \\
 & + 25.12 \times 10^{-1}(S_7 - \bar{S}_7)(T - \bar{T}) + 26.19 \times 10^{-1}(S_8 - \bar{S}_8)(T - \bar{T})
 \end{aligned} \quad (14)$$

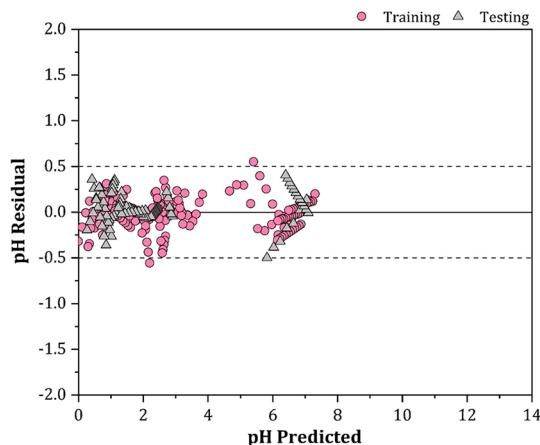


Figure 8. Residual deviation between the experimentally determined and model predicted pH values in the Family A MLR model.

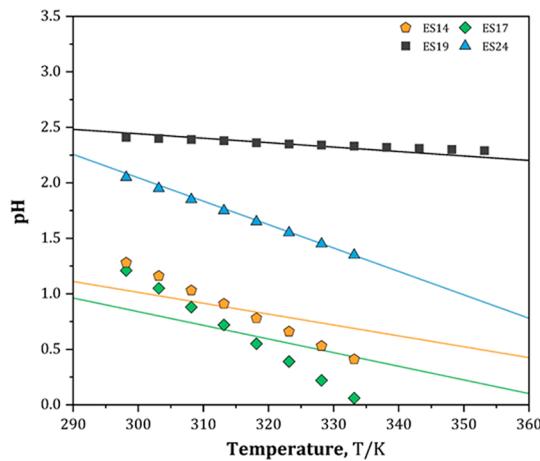


Figure 9. Experimental and predicted pH values as a function of temperature in external molecule validation for Family A's MLR model. The points represent the experimental data, and the solid lines represent the model predictions.

Table 5. Coefficients of 40 Significant Descriptors and Intercept of the Family B Model

Category	Term	Estimate	Standard error	t _{Ratio}	P _{value}
Intercept	a_0	2712.99	998.45	2.72	0.0070
HBD	S_1	12195344.00	919519.90	13.26	<0.0001
	S_2	2007820.40	108337.40	18.53	<0.0001
Weak donor	S_3	1625622.70	64483.41	25.21	<0.0001
Nonpolar	S_4	299995.39	11001.01	27.27	<0.0001
	S_5	-1219162.00	53413.79	22.82	<0.0001
Weak acceptor	S_6	-242091.30	17662.87	13.71	<0.0001
HBA	S_7	-2446456.00	94746.93	25.82	<0.0001
	S_8	-3275372.00	125348.20	26.13	<0.0001
Temperature	T	-9.91×10^{-3}	4.38×10^{-4}	22.63	<0.0001
{HBD, Temperature}	$(S_1 - \bar{S}_1)(T - \bar{T})$	-40.15×10^{-1}	24.02×10^{-1}	1.67	0.0957
	$(S_2 - \bar{S}_2)(T - \bar{T})$	51.57×10^{-1}	6.41×10^{-1}	8.05	<0.0001
{Weak donor, Temperature}	$(S_3 - \bar{S}_3)(T - \bar{T})$	-8.32×10^{-1}	1.74×10^{-1}	4.78	<0.0001
{Nonpolar, Temperature}	$(S_5 - \bar{S}_5)(T - \bar{T})$	5.77×10^{-1}	7.60×10^{-2}	7.59	<0.0001
	$(S_6 - \bar{S}_6)(T - \bar{T})$	-7.69×10^{-1}	1.38×10^{-1}	5.56	<0.0001
{HBA, Temperature}	$(S_7 - \bar{S}_7)(T - \bar{T})$	25.12×10^{-1}	3.53×10^{-1}	7.11	<0.0001
	$(S_8 - \bar{S}_8)(T - \bar{T})$	26.19×10^{-1}	7.76×10^{-1}	3.37	0.0008
{HBD, HBD}	$(S_1 - \bar{S}_1)(S_2 - \bar{S}_2)$	2637000000.00	1390000000.00	18.97	<0.0001
{HBD, Weak donor}	$(S_1 - \bar{S}_1)(S_3 - \bar{S}_3)$	7141600000.00	294900000.00	24.21	<0.0001
	$(S_2 - \bar{S}_2)(S_3 - \bar{S}_3)$	-741500000.00	34900324.00	21.25	<0.0001
{HBD, Nonpolar}	$(S_1 - \bar{S}_1)(S_4 - \bar{S}_4)$	2172600000.00	77953697.00	27.87	<0.0001

where S_i and \bar{S}_i are the molecular descriptors, and their mean values ($e/\text{\AA}^2$), respectively, and T is the temperature (K).

Statistical analysis was then performed on the developed model, and the results are summarized in Table S.6. On the basis of the obtained results, it was concluded that the Family B model also established a strong performance statistically. The regression coefficient (R^2) and the Fisher's statistic (F_{Ratio}) values were high; $R^2 = 0.9969$ and $F_{\text{Ratio}} = 1991.87$ with a $P_{\text{value, Fisher}} < 0.0001$ confirming the suitability of fitting the experimental pH values. The cross-validation coefficients were high suggesting the robustness of the model, and the low value of the y -scrambling regression coefficient ($R^2_{\text{scramble}} = 0.0056$) and high value of adjusted regression coefficient ($R^2_{\text{adjusted}} = 0.9963$) indicate the absence of chance regression correlation.

Figure 10(a) shows the experimental versus the predicted data of the training set. The high regression coefficient indicated the linearity of the model as all the experimental and predicted data were lying on the diagonal line with narrow dispersion ($R^2 = 0.9969$), and the RMSE = 0.1205. The external predictivity of the model was then evaluated with 10 ESSs used as a testing set. In Figure 10(b), it can be seen that the model showed high predictive capabilities as the regression coefficient (R^2_{external}) obtained was 0.9960, and the RMSE was 0.1433. Figure 11 further confirms the quality of the model in predicting the pH of Family B's ESSs where the residuals was at a range of $\pm 0.50\%$ with an AARD of $\pm 3.68\%$ and an SD_{avg} of ± 0.08 .

Furthermore, Figure 12 shows the performance of the Family B MLR model in predicting the pH values of the external molecule ESS set as a function of temperature. It can be observed that the experimental and the predicted points were fairly overlying indicating high reliability of the proposed model except for ATPPB:DEG:H₂O (ES1), BTPC:EG (ES3), and ChCl:DEA (ES 6) where their external performance was deviating at lower temperatures. Hence, their average standard deviations were ± 0.27 , ± 0.24 , and ± 0.30 , respectively. Nevertheless, the model still computes the correct qualitative trend with the pHs of ES6 > ES3 > ES1, which is consistent with the trend in the experimental data.

Table 5. continued

Category	Term	Estimate	Standard error	t _{Ratio}	P _{value}
	$(S_1 - \bar{S}_1)(S_5 - \bar{S}_5)$	-3257000000.00	228200000.00	14.27	<0.0001
	$(S_2 - \bar{S}_2)(S_4 - \bar{S}_4)$	-46429937.00	2273852.00	20.42	<0.0001
	$(S_2 - \bar{S}_2)(S_5 - \bar{S}_5)$	93297386.00	4099702.00	22.76	<0.0001
{HBD, Weak acceptor}	$(S_1 - \bar{S}_1)(S_6 - \bar{S}_6)$	-5095000000.00	241800000.00	21.07	<0.0001
{HBD, HBA}	$(S_1 - \bar{S}_1)(S_7 - \bar{S}_7)$	-7338000000.00	281300000.00	26.09	<0.0001
	$(S_2 - \bar{S}_2)(S_7 - \bar{S}_7)$	1716100000.00	80438510.00	21.33	<0.0001
	$(S_2 - \bar{S}_2)(S_8 - \bar{S}_8)$	924971536.00	34174322.00	27.07	<0.0001
{Weak donor, Nonpolar}	$(S_3 - \bar{S}_3)(S_4 - \bar{S}_4)$	-33295616.00	1559555.00	21.35	<0.0001
	$(S_3 - \bar{S}_3)(S_5 - \bar{S}_5)$	38422954.00	2005305.00	19.16	<0.0001
{Weak donor, Weak acceptor}	$(S_3 - \bar{S}_3)(S_6 - \bar{S}_6)$	151526609.00	6987297.00	21.69	<0.0001
{Weak donor, HBA}	$(S_3 - \bar{S}_3)(S_7 - \bar{S}_7)$	4999312.60	274929.90	18.18	<0.0001
	$(S_3 - \bar{S}_3)(S_8 - \bar{S}_8)$	1287300000.00	50105026.00	25.69	<0.0001
{Nonpolar, Nonpolar}	$(S_4 - \bar{S}_4)(S_5 - \bar{S}_5)$	-5079827.00	424574.70	11.96	<0.0001
{Nonpolar, HBA}	$(S_4 - \bar{S}_4)(S_7 - \bar{S}_7)$	77090308.00	3600634.00	21.41	<0.0001
	$(S_4 - \bar{S}_4)(S_8 - \bar{S}_8)$	187576063.00	7419743.00	25.28	<0.0001
	$(S_5 - \bar{S}_5)(S_7 - \bar{S}_7)$	-66022050.00	3544836.00	18.62	<0.0001
	$(S_5 - \bar{S}_5)(S_8 - \bar{S}_8)$	-1191000000.00	47194228.00	25.24	<0.0001
{Nonpolar, Weak acceptor}	$(S_5 - \bar{S}_5)(S_6 - \bar{S}_6)$	26647478.00	2158303.00	12.35	<0.0001
{Weak acceptor, HBA}	$(S_6 - \bar{S}_6)(S_7 - \bar{S}_7)$	-382400000.00	17597237.00	21.73	<0.0001
{HBA, HBA}	$(S_7 - \bar{S}_7)(S_8 - \bar{S}_8)$	-2196000000.00	83528785.00	26.29	<0.0001

^aMean values: $\bar{S}_1 = 7.32 \times 10^{-5}$; $\bar{S}_2 = 6.89 \times 10^{-3}$; $\bar{S}_3 = 2.15 \times 10^{-2}$; $\bar{S}_4 = 6.91 \times 10^{-2}$; $\bar{S}_5 = 2.30 \times 10^{-2}$; $\bar{S}_6 = 1.23 \times 10^{-2}$; $\bar{S}_7 = 1.68 \times 10^{-2}$; $\bar{S}_8 = 8.40 \times 10^{-4}$; $T = 324.019$.

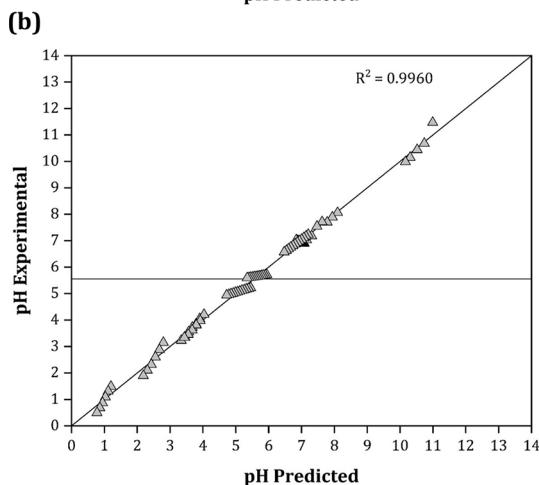
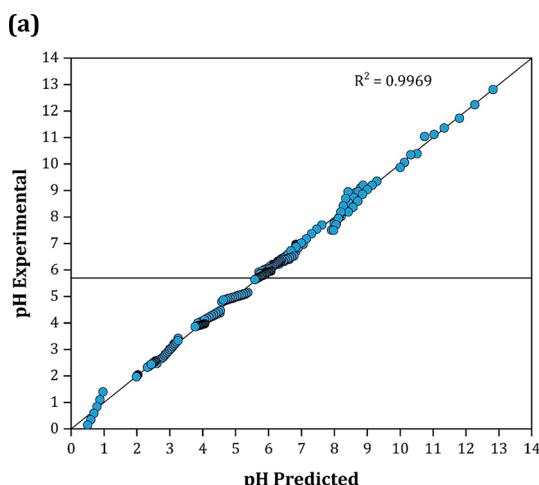


Figure 10. Parity graph of experimental and predicted pH values of Family B's MLR model in (a) training and (b) external testings.

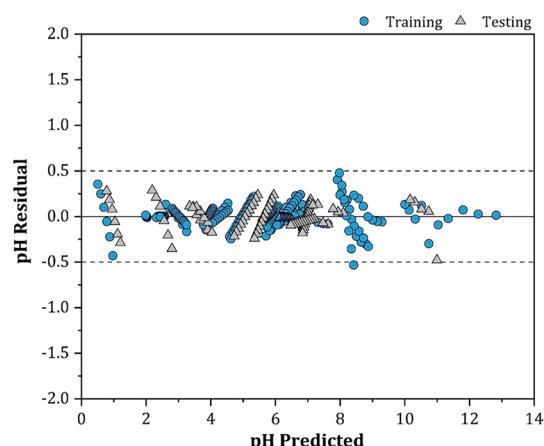


Figure 11. Residual deviation between the experimentally determined and model-predicted pH values in the Family B's MLR model.

It should be mentioned that although the developed models consist of multidescriptors, the parameters are all essentially constructed based on eight basic molecular descriptors only ($S_1 - S_8$) and the temperature descriptor (T). The remaining parameters are just basic multiplications of a pair of descriptors expressing the binary interactions between the descriptors. Subsequently, the model at hand can be utilized for studying and predicting the pH of a large amount of ESs, which have not been tested experimentally, utilizing eight basic molecular descriptors only that can be obtained through simple σ -profiles of the desired HBAs and the HBDs, enabling a straightforward and cheap method for screening new green and sustainable ESs with the required pH for a particular application. Nevertheless, a consistent physical interpretation cannot be made for the two models due to the complexity in the binary interaction effect and the difficulty in generalizing the results for both models. Moreover, splitting the data set can violate the QSPR modeling guidelines of utilizing a sufficient data set to generate QSPR

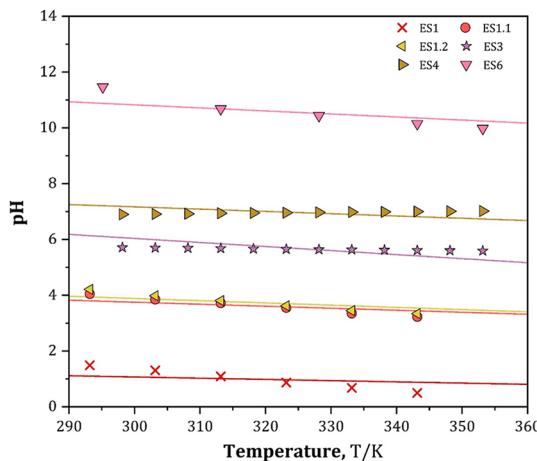


Figure 12. Experimental and predicted pH values as a function of temperature in external molecule validation for Family B's MLR model. The points represent the experimental data, and the solid lines represent the model predictions.

models. Therefore, another machine learning method was utilized to predict the pH of ESSs which is the artificial neural network. It should be mentioned that the authors presented some preliminary approaches for the physical interpretation of the two models in the [Supporting Information](#).

3.4. Artificial Neural Network. **3.4.1. S_{σ} -profiles Discretization and Architecture Optimization.** As a comparison to the performance of the linear MLR machine learning algorithm, another nonlinear model of the experimental pH data of the ESSs has also been developed utilizing a feed-forward artificial neural network (ANN). The 4, 6, 8, 10, and 12 discretized S_{σ} -profile descriptors and the temperature were selected as the network's inputs, while the pHs of the ESSs were selected as the output. During model development, it was found that an accurate fit for the entire experimental data can be determined using the ANN model (unlike the MLR model) eliminating the need to split the data into two subsets that are more specific to certain families of ESSs. Out of the 84 ESS compositions, 32 ESS

compositions were selected for network learning, and the remaining 19 ESS compositions were utilized in external validation.

The number of neurons in the hidden layer is a crucial parameter that has a substantial influence on the accuracy and the complexity of the developed model.^{71,85} Thus, to avoid developing a complicated or an overfitted model, several network architectures with 2, 4, 6, 8, 10, and 12 hidden neurons were studied. A heatmap of the regression coefficient (R^2) in several neural network architectures based on the number of hidden neurons and partitioning of S_{σ} -profile is shown in **Figure 13**. On the basis of the obtained results, the 8-partition network with a neuron architecture of 9–6–1 (i.e., 60 weight coefficients and 7 bias intercepts) has been selected as the optimal compromise between accuracy and fitting parameters as it is the simplest model with a regression coefficient above $R^2 > 0.99$. The selected neural configuration is visually depicted in **Figure 14**.

3.4.2. ANN Model Evaluation. The weight coefficient and the bias intercept estimates of the 9–6–1 network are listed in **Table 6**. Consequently, the resulting ANN model can be expressed as follows

$$\begin{aligned} \text{pH} = & 2.46(H_1) - 6.87(H_2) + 5.37(H_3) + 4.46(H_4) \\ & - 1.14(H_5) - 2.04(H_6) + 2.16 \end{aligned} \quad (15)$$

where the hidden neurons H_1, H_2, H_3, H_4, H_5 , and H_6 are expressed as follows

$$\begin{aligned} H_1 = & \tanh\left(\frac{1}{2}(4159.25(S_1) - 6129.54(S_2) + 1125.83(S_3) \right. \\ & - 572.51(S_4) - 238.14(S_5) + 2373.34(S_6) - 2396.48(S_7) \\ & \left. + 1882.06(S_8) - 5.67 \times 10^{-4}(T) + 72.82)\right) \end{aligned} \quad (16)$$

$$\begin{aligned} H_2 = & \tanh\left(\frac{1}{2}(-2166.37(S_1) + 1316.91(S_2) - 929.68(S_3) \right. \\ & - 16.38(S_4) + 805.43(S_5) - 80.86(S_6) + 1073.54(S_7) \\ & \left. + 205.36(S_8) + 1.21 \times 10^{-5}(T) - 23.81)\right) \end{aligned} \quad (17)$$

	2 Neurons (3 intercepts)	4 Neurons (5 intercepts)	6 Neurons (7 intercepts)	8 Neurons (9 intercepts)	10 Neurons (11 intercepts)	12 Neurons (13 intercepts)
12 Partitions (S_1-S_{12})	$R^2=0.8952$	$R^2=0.9814$	$R^2=0.9971$	$R^2=0.9991$	$R^2=0.9997$	$R^2=0.9999$
	28 coefficients	56 coefficients	84 coefficients	112 coefficients	140 coefficients	168 coefficients
10 Partitions (S_1-S_{10})	$R^2=0.8763$	$R^2=0.9782$	$R^2=0.9964$	$R^2=0.9983$	$R^2=0.9993$	$R^2=0.9996$
	24 coefficients	48 coefficients	72 coefficients	96 coefficients	120 coefficients	144 coefficients
8 Partitions (S_1-S_8)	$R^2=0.8557$	$R^2=0.9679$	$R^2=0.9951$	$R^2=0.9978$	$R^2=0.9986$	$R^2=0.9992$
	20 coefficients	40 coefficients	60 coefficients	80 coefficients	100 coefficients	120 coefficients
6 Partitions (S_1-S_6)	$R^2=0.8432$	$R^2=0.9408$	$R^2=0.9716$	$R^2=0.9943$	$R^2=0.9968$	$R^2=0.9981$
	16 coefficients	32 coefficients	48 coefficients	64 coefficients	80 coefficients	96 coefficients
4 Partitions (S_1-S_4)	$R^2=0.7645$	$R^2=0.8862$	$R^2=0.9209$	$R^2=0.9635$	$R^2=0.9793$	$R^2=0.9888$
	12 coefficients	24 coefficients	36 coefficients	48 coefficients	60 coefficients	72 coefficients

Figure 13. Heatmap of the regression coefficient (R^2) in several neural network architectures based on the number of hidden neurons and partitioning of S_{σ} -profile.

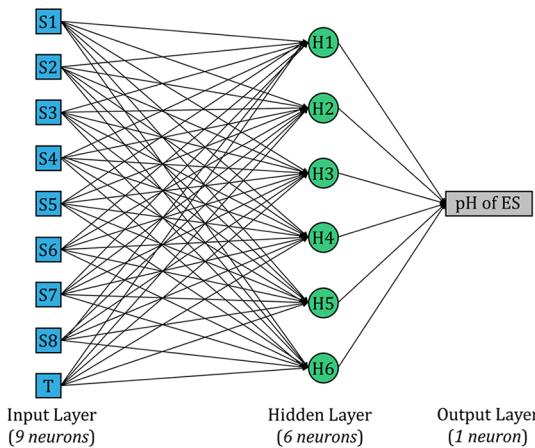


Figure 14. 9–6–1 architecture configuration of the artificial neural network for predicting the pH of ESs.

$$\begin{aligned} H_3 = \tanh\left(\frac{1}{2}\left(-769.41(S_1) - 492.08(S_2) - 136.95(S_3) \right.\right. \\ \left.\left. - 5.09(S_4) + 305.81(S_5) + 143.56(S_6) + 600.8(S_7) \right.\right. \\ \left.\left. - 62.57(S_8) - 4.43 \times 10^{-3}(T) - 7.48\right)\right) \quad (18) \end{aligned}$$

$$\begin{aligned} H_4 = \tanh\left(\frac{1}{2}\left(-28529.31(S_1) - 1669.83(S_2) + 386.62(S_3) \right.\right. \\ \left.\left. + 99.2(S_4) + 34.18(S_5) + 41.16(S_6) - 1725.9(S_7) \right.\right. \\ \left.\left. - 3654.55(S_8) - 3.76 \times 10^{-3}(T) + 26.5\right)\right) \quad (19) \end{aligned}$$

$$\begin{aligned} H_5 = \tanh\left(\frac{1}{2}\left(-9600.19(S_1) - 4736.25(S_2) - 812.63(S_3) \right.\right. \\ \left.\left. - 741.55(S_4) + 310.38(S_5) + 2614.84(S_6) + 1154.73(S_7) \right.\right. \\ \left.\left. + 13877.77(S_8) + 4.49 \times 10^{-2}(T) + 12.65\right)\right) \quad (20) \end{aligned}$$

$$\begin{aligned} H_6 = \tanh\left(\frac{1}{2}\left(10169.21(S_1) + 4741.2(S_2) - 633.29(S_3) \right.\right. \\ \left.\left. - 340.01(S_4) + 2496.14(S_5) - 746.18(S_6) + 2610.79(S_7) \right.\right. \\ \left.\left. - 11328.32(S_8) - 1.34 \times 10^{-4}(T) - 70.73\right)\right) \quad (21) \end{aligned}$$

where S_i is the molecular descriptor of region i ($\text{e}/\text{\AA}^2$), and T is the temperature (K).

A statistical summary of the performance of the ANN model is presented in Table 7. When compared to the MLR model, the prediction performance of the ANN model was found to be

Table 6. Estimates of 60 Weight Coefficients and Seven Bias Intercepts of the 9–6–1 ANN Model

Neuron	Term	Coefficient estimate	P _{value}	Neuron	Term	Coefficient estimate	P _{value}
H_1	S_1	4159.25	<0.0001	H_5	S_5	34.18	<0.0001
	S_2	-6129.54	<0.0001		S_6	41.16	<0.0001
	S_3	1125.83	<0.0001		S_7	-1725.90	<0.0001
	S_4	-572.51	<0.0001		S_8	-3654.55	<0.0001
	S_5	-238.14	<0.0001		T	-3.76×10^{-3}	<0.0001
	S_6	2373.34	<0.0001		bias b_4	26.50	<0.0001
	S_7	-2396.48	<0.0001		S_1	-9600.19	<0.0001
	S_8	1882.06	<0.0001		S_2	-4736.25	<0.0001
	T	-5.67×10^{-4}	<0.0001		S_3	-812.63	<0.0001
	bias b_1	72.82	<0.0001		S_4	-741.55	<0.0001
H_2	S_1	-2166.37	<0.0001		S_5	310.38	<0.0001
	S_2	1316.91	<0.0001		S_6	2614.84	<0.0001
	S_3	-929.68	<0.0001		S_7	1154.73	<0.0001
	S_4	-16.38	<0.0001		S_8	13877.77	<0.0001
	S_5	805.43	<0.0001		T	4.49×10^{-2}	<0.0001
	S_6	-80.86	<0.0001		bias b_5	12.65	<0.0001
	S_7	1073.54	<0.0001		S_1	10169.21	<0.0001
	S_8	205.36	<0.0001		S_2	4741.20	<0.0001
	T	1.21×10^{-5}	<0.0001		S_3	-633.29	<0.0001
	bias b_2	-23.81	<0.0001		S_4	-340.01	<0.0001
H_3	S_1	-769.41	<0.0001		S_5	2496.14	<0.0001
	S_2	-492.08	<0.0001		S_6	-746.18	<0.0001
	S_3	-136.95	<0.0001		S_7	2610.79	<0.0001
	S_4	-5.09	<0.0001		S_8	-11328.32	<0.0001
	S_5	305.81	<0.0001		T	-1.34×10^{-4}	<0.0001
	S_6	143.56	<0.0001		bias b_6	-70.73	<0.0001
	S_7	600.80	<0.0001		pH	H_1	2.46
	S_8	-62.57	<0.0001		H_2	-6.87	<0.0001
	T	-4.43×10^{-3}	<0.0001		H_3	5.37	<0.0001
	bias b_3	-7.48	<0.0001		H_4	4.46	<0.0001
H_4	S_1	-28529.31	<0.0001		H_5	-1.14	<0.0001
	S_2	-1669.83	<0.0001		H_6	-2.04	<0.0001
	S_3	386.62	<0.0001		bias b	2.16	<0.0001
	S_4	99.20	<0.0001				

Table 7. ANN Model's Statistical Performance

Parameter	Value
Training set	
Total number of HBA	8
Total number of HBD	14
Number of compositions	62
ESs	ESs 2, 2.2, 5, 7, 8, 8.2, 8.3, 9, 9.2, 9.3, 10, 11, 12, 12.1, 13, 13.1, 15, 16, 18, 20, 20.1, 20.2, 22, 23, 25, 26, 27, 28, 28.1, 28.2, 29, 29.1, 29.2, 30, 30.1, 30.1, 31, 31.1, 32, 32.1, 32.2, 33, 33.4, 34, 35, 35.1, 36, 36.1, 37, 37.1, 37.2, 38, 38.1, 38.2, 39, 39.1, 39.2, 40, 40.1, 40.2, 41.1
Data points	473
R^2	0.9951
AIC _c	-102.99
$P_{\text{value, Fisher}}$	<0.0001
$Q^2_{\text{LMO, 25\%}}$	0.9938
R^2_{scramble}	0.1216
RMSE	0.2241
SD _{avg}	±0.11
AARD	±9.48%
AD _{coverage}	98.1%
Testing set	
Total number of HBA	22
Total number of HBD	7
Number of ESs	15
ESs	ESs 1, 1.1, 1.2, 2.1, 3, 4, 5.1, 6, 8.1, 9.1, 14, 17, 19, 21, 24, 30.3, 31.2, 33.1, 35.2, 36.2, 41, 41.2
External molecules	DEEAC, BTPC, MalA, OxaA, Suc, DEA, DEG
Data points	175
R^2_{external}	0.9923
RMSE	0.1890
SD _{avg}	±0.11
AARD	±11.42%
AD _{coverage}	96.0%

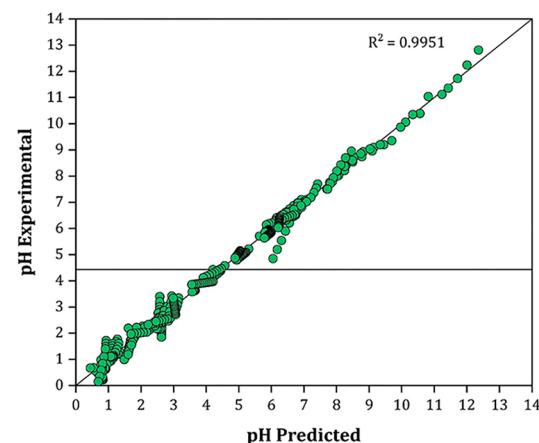
better as the model can fit the pH property of acids, sugars, polyols, and amines in a single correlation with an internal R^2 fit of 0.9951. The high regression coefficient can be observed visually in Figure 15(a) as the experimental and predicted data were lying on the parity graph's diagonal line with a narrow dispersion (RMSE = 0.2241).

The external predictivity of the ANN model was then assessed with a testing set composed of 22 ESs. It can be seen from Figure 15(b) that the model showed high predictive capabilities as the R^2_{external} was obtained to be 0.9923. This result was also confirmed by the prediction residuals shown in Figure 16 where the majority of the residuals were at a range of ±0.5 with an AARD of ±11.42% and an SD_{avg} of ±0.11.

Figure 17 shows the performance of the ANN model in predicting the pH values of the external molecule ES set as a function of temperature. Overall, it can be seen that the model predicts the external molecules quite accurately. However, the performance of ChCl:MalA:H₂O (ES14) is a lot lower than the rest with an SD_{avg} of ±0.30.

3.5. Applicability Domain. To strengthen the validity of a QSPR model, the Organization for Economic Cooperation and Development (OECD) suggested the use of the applicability domain (AD) as a validation tool to estimate the uncertainties in predicting new compounds by a proposed QSPR model.^{91,102} The AD is a theoretical region in the chemical space, which represents the limitation of a model.¹⁰³ The model is considered

(a)



(b)

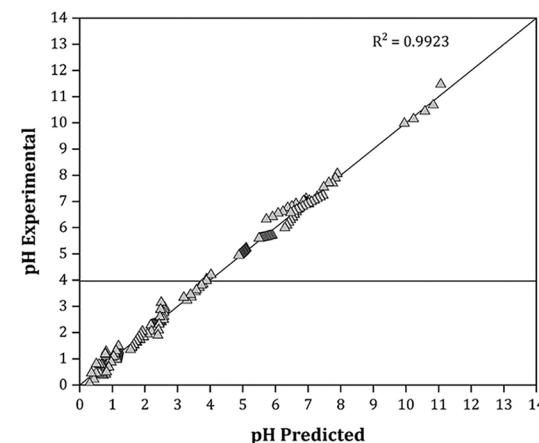


Figure 15. Parity graph of experimental and predicted pH values of the 9–6–1 ANN model in (a) training and (b) external testings.

to be valid and able to predict new molecules if the training and testing data sets are within the domain of applicability of that model.⁸³ The points located outside the domain are treated as outliers. These outliers are considered of lower reliability as a result of their extrapolation and consequently must be regarded with more attention. The outliers are a result of the differences in the chemical structure of the selected experimental points for training and testing sets. Hence, it is important to carefully select the experimental points that represent both sets. In other words, the sets should be selected wisely and should cover a variety of different data set points to maximize the domain and to shield the entire chemical space for all the data set molecules with proper handling of outliers.⁸⁷

The AD margins are defined as follows: (i) The x -axis represents the leverage values where $0 < h_i < h^*$. (ii) The y -axis represents the standardized residuals where $-3 < SDR < +3$.⁸³ Figure 18 shows the William Plots for the developed models.

In the first step of obtaining the AD, the critical leverage values (h^*) were calculated for the models to be 0.53, 0.49, and 0.48 for the Family A MLR model, the Family B MLR model, and the ANN model, respectively. After calculating h^* , the domains were obtained to test the applicability of each model in covering and predicting a wide range of new ES combinations. From Figure 18(a) and (b), it can be seen that all the ESs in the training set and testing set in both Family A

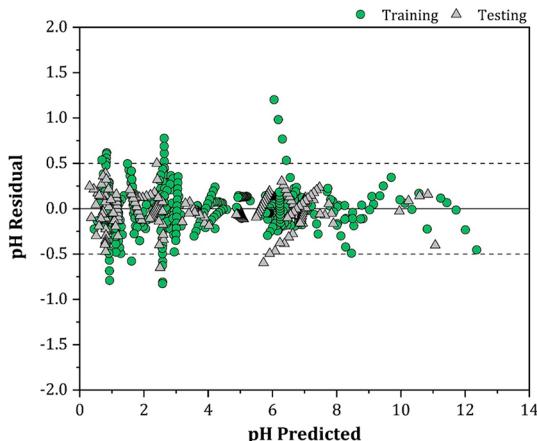


Figure 16. Residual deviation between the experimentally determined and model predicted pH values in the 9–6–1 ANN model.

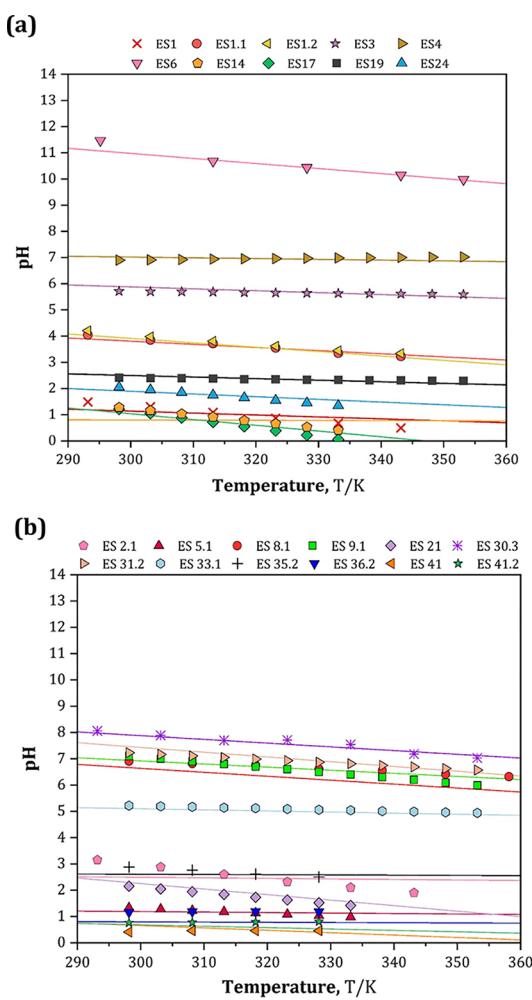


Figure 17. Experimental and predicted pH values as a function of temperature in (a) external molecule validation and (b) regular external validation for the ANN model. The points represent the experimental data, and the solid lines represent the model predictions.

and Family B MLR models were within the AD margins and present no outliers with an $AD_{coverage} = 100.0\%$, indicating that the families are homogeneous and can be well described by the proposed machine learning algorithm. As for the ANN model

in Figure 18(c), it can be observed that the model presents no structural outliers as all the points present a leverage value significantly lower than the critical leverage value ($h_i < h^*$). Nonetheless, the estimates of a few ESs were considered as “response outliers” as they exhibit standardized residuals values above the ± 3 boundaries, which bring down the $AD_{coverage}$ to 97.5%. The response outliers in the general ANN model include ATPPB:DEG:H₂O (ES1), ATPPB:TEG:H₂O (ES2), ChCl:DEA (ES6), and ChCl:MDEA (ES15). On the basis of the obtained AD analysis, it can be concluded that the prediction of a new combination of hydrophilic ESs that (i) are within the model’s applicability domain and (ii) contain similar constituents to the ones utilized in the training set could be considered reliable. However, the prediction of new hydrophilic ESs that are not within the model’s applicability domain should be treated with more attention due to their high extrapolation degree.

3.6. Models Comparison. The performance of the general ANN model has been compared to the performance of the Family A and Family B MLR models in Table 8. On the basis of the obtained results, it can be concluded that ANN models based on $S_{\sigma\text{-profile}}$ descriptors are excellent at predicting the pH of ESs, especially when compared to the developed MLR models that required the data to be split into two family specific subsets. Nonetheless, the ANN model still requires a large amount of estimated fitting parameters, and even though ANNs provide highly accurate and reliable predictive models, one of their well-known shortcomings is the lack of physical interpretation between the inputs and the predicted property.⁷¹ However, all in all, the utilization of the ANN model can provide a robust method capable of predicting the pH of ESs to promote a more efficient and “sound” design method for prescreening new sustainable solvents.

4. CONCLUSION

The need for facilitating studies concerned with screening green and sustainable solvents encouraged the use of computer-assisted methods in predicting the properties and behaviors of designer solvents. In this article, novel machine learning-based QSPR models were developed to predict the pHs of 41 ESs composed of various combinations of 9 HBAs and 21 HBDs resulting in a total of 648 experimental data points. For the sake of comparing, the performance of the linear machine learning algorithm and another nonlinear algorithm were utilized. To construct these models, the 3D molecular structure of the ES constituents was first successfully modeled. Then, using COSMO-RS, the $S_{\sigma\text{-profile}}$ of these molecules were calculated and assigned as molecular descriptor inputs to both machine learning algorithms. Statistical analysis was then applied to evaluate the robustness, reliability, and predictivity of the pH models. The results confirmed the validity of the models as they can capture the behavior of the training data set with high R^2 values and predict the external testing data set successfully with low standard deviations. Moreover, the models showed a domain of applicability that covered a wide space of molecular structures.

In summary, on the basis of the statistical analysis of the models, it can be concluded that both the MLR and ANN approaches could be considered reliable and can be utilized for screening purposes in the absence of experimental data for the determination of the pH properties of ESs, particularly for screening new green and sustainable ESs for process design and industrial scale-up. The ANN model presents better

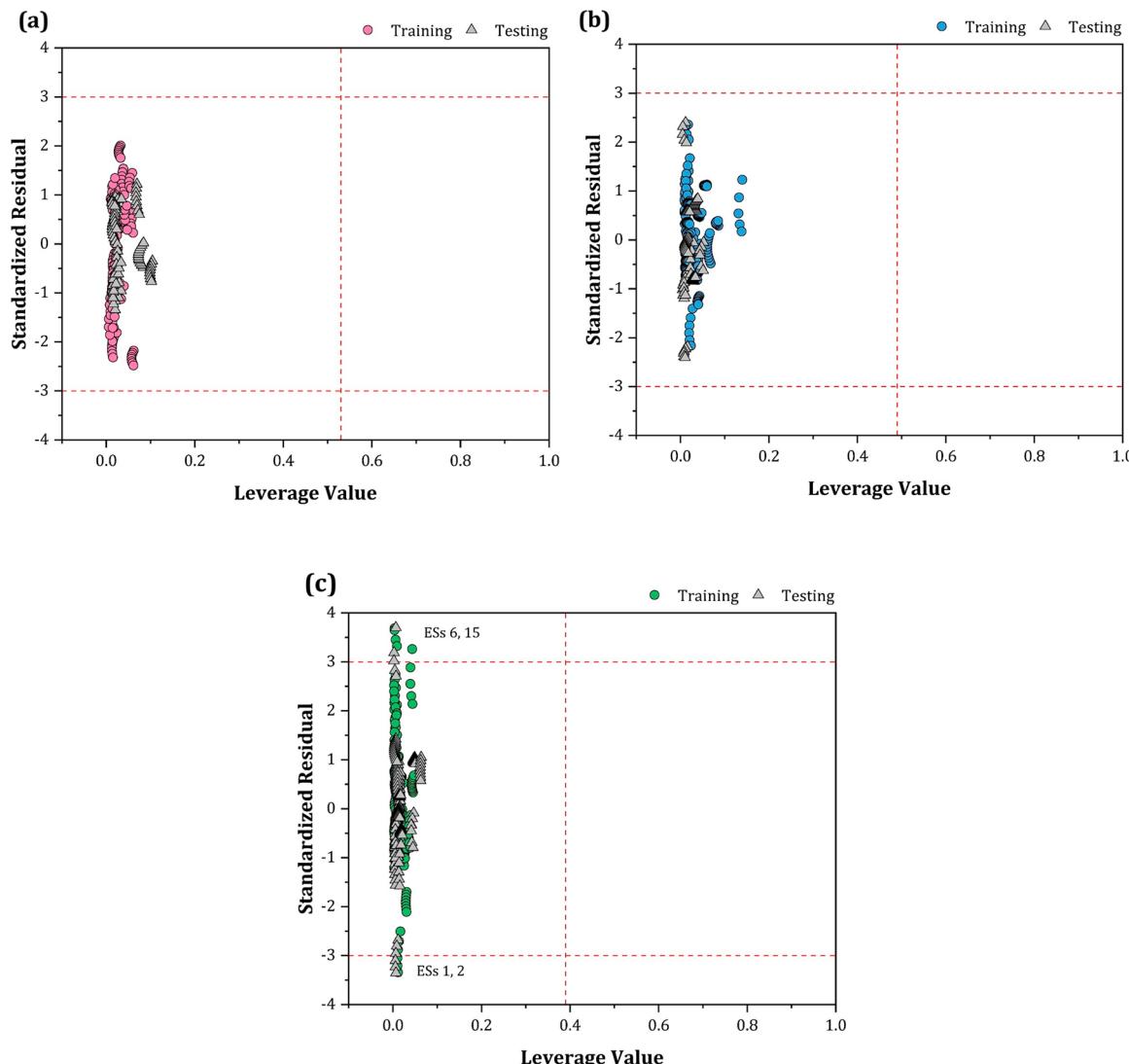


Figure 18. William plots for the (a) Family A MLR model, the (b) Family B MLR model, and the (c) general ANN model. The dashed lines represent the boundaries of the applicability domain.

Table 8. Comparison of Family-Specific MLR Models and General 9-6-1 ANN Model in Predicting the External Molecule ES set^a

ES	Temperature	Experimental	MLR prediction	ANN prediction
Family A's external molecule ES set				
ES14, ChCl:MalA:H ₂ O (1:1:0.22)	298.15–333.15	1.28–0.41	1.02–0.67 ($SD_{avg} = \pm 0.18$)	0.81–0.79 ($SD_{avg} = \pm 0.30$)
ES17, ChCl:OxaA:H ₂ O (1:1:2.44)	298.15–333.15	1.21–0.06	0.85–0.42 ($SD_{avg} = \pm 0.25$)	1.08–0.31 ($SD_{avg} = \pm 0.13$)
ES19, DEEAC:MalA (1:1)	298.15–353.15	2.41–2.29	2.46–2.24 ($SD_{avg} = \pm 0.04$)	2.51–2.18 ($SD_{avg} = \pm 0.07$)
ES 24, MA:Suc:H ₂ O, (1:1:0.22)	298.15–333.15	2.05–1.35	2.07–1.33 ($SD_{avg} = \pm 0.01$)	1.92–1.56 ($SD_{avg} = \pm 0.12$)
Family B's external molecule ES set				
ES1, ATPPB:DEG:H ₂ O (1:4:0.17)	293.15–343.15	1.49–0.50	1.10–0.88 ($SD_{avg} = \pm 0.27$)	1.19–0.82 ($SD_{avg} = \pm 0.22$)
ES1.1, ATPPB:DEG:H ₂ O (1:10:0.31)	293.15–343.15	4.05–3.23	3.80–3.44 ($SD_{avg} = \pm 0.16$)	3.89–3.29 ($SD_{avg} = \pm 0.08$)
ES1.2, ATPPB:DEG:H ₂ O (1:16:0.39)	293.15–343.15	4.21–3.34	3.94–3.54 ($SD_{avg} = \pm 0.17$)	4.03–3.19 ($SD_{avg} = \pm 0.12$)
ES3, BTPC:EG (1:3)	298.15–353.15	5.71–5.59	6.05–5.25 ($SD_{avg} = \pm 0.24$)	5.90–5.50 ($SD_{avg} = \pm 0.10$)
ES4, BTPC:Gly (1:5)	298.15–353.15	6.90–7.02	7.19–6.74 ($SD_{avg} = \pm 0.20$)	7.02–6.86 ($SD_{avg} = \pm 0.10$)
ES6, ChCl:DEA (1:6)	295.15–353.15	11.47–9.98	10.89–10.26 ($SD_{avg} = \pm 0.30$)	11.07–9.95 ($SD_{avg} = \pm 0.15$)

^aAverage standard deviation between the two points (SD_{avg}).

predictive capabilities and a more robust method overall, while the MLR model can be considered as more interpretable. These models inspire and stimulate the development of robust models to predict the properties of designer solvents from the drawn molecular structures, which will save time and resources.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acssuschemeng.0c07367>.

List of all pH experimental data used in this work, calculated S_{σ} -profile values of HBAs and HBDs, list of all 45 descriptors, heatmap of regression coefficient in the general MLR model, applicability domain of the general MLR model, general MLR model's statistical performance, Family A MLR model's statistical performance, summary of signs of the interaction descriptors in the Family A model, Family B MLR model's statistical performance, and summary of signs of interaction descriptors in the Family B model ([PDF](#))

CIRA-2018-023 is gratefully acknowledged. The authors are also grateful for the support received from the Center for Membrane and Advanced Water Technology (CMAT, RC2-2018-009) at Khalifa University in UAE, the generous support of Laboratoire des Matériaux Polymères Multiphasiques (LMPMP), Energetics and Solid-State Electrochemistry Laboratory (LEES) at Université Ferhat ABBAS in Algeria to perform this research, and Directorate General for Scientific Research and Technological Development (DGRSDT).

AUTHOR INFORMATION

Corresponding Author

Inas M. Alnashef – Chemical Engineering Department and Center for Membrane and Advanced Water Technology (CMAT), Khalifa University, Abu Dhabi, United Arab Emirates; [orcid.org/0000-0003-4654-2932](#); Email: enas.nashef@ku.ac.ae

Authors

Tarek Lemaoui – Laboratoire des Matériaux Polymères Multiphasiques (LMPMP) and Energetics and Solid-State Electrochemistry Laboratory (LEES), Université Ferhat ABBAS Sétif-1, 19000 Sétif, Algeria; [orcid.org/0000-0002-9634-7134](#)

Farah Abu Hatab – Chemical Engineering Department and Center for Membrane and Advanced Water Technology (CMAT), Khalifa University, Abu Dhabi, United Arab Emirates

Ahmad S. Darwish – Chemical Engineering Department and Center for Membrane and Advanced Water Technology (CMAT), Khalifa University, Abu Dhabi, United Arab Emirates

Ayoub Attoui – Laboratoire des Matériaux Polymères Multiphasiques (LMPMP), Université Ferhat ABBAS Sétif-1, 19000 Sétif, Algeria

Nour El Houda Hammoudi – Laboratoire des Matériaux Polymères Multiphasiques (LMPMP) and Energetics and Solid-State Electrochemistry Laboratory (LEES), Université Ferhat ABBAS Sétif-1, 19000 Sétif, Algeria

Ghaiath Almustafa – Chemical Engineering Department and Center for Membrane and Advanced Water Technology (CMAT), Khalifa University, Abu Dhabi, United Arab Emirates

Mohamed Benacha – Energetics and Solid-State Electrochemistry Laboratory (LEES), Université Ferhat ABBAS Sétif-1, 19000 Sétif, Algeria

Yacine Benguerba – Laboratoire des Matériaux Polymères Multiphasiques (LMPMP), Université Ferhat ABBAS Sétif-1, 19000 Sétif, Algeria; [orcid.org/0000-0002-8251-9724](#)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acssuschemeng.0c07367>

Author Contributions

¹Tarek Lemaoui and Farah Abu Hatab have contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The financial support provided to this work by Khalifa University in UAE through Awards CIRA-2018-069 and

NOMENCLATURE

Symbols

- a_0 = intercept
 a_i = coefficient of descriptor i
 a_{i-j} = interaction coefficient
 $AD_{coverage}$ = applicability domain coverage
AIC = Akaike information criterion
AIC_c = corrected Akaike information criterion
 b_k = intercept bias of hidden neuron k
 d^* = number of significant model descriptors
 F_{Ratio} = Fisher's statistic
 h_i = leverage value of each data point
 h^* = critical leverage value
 H_k = activation function of hidden neuron k
 K = number of estimated parameters
 L = maximum value of the likelihood function
 P_{value} = probability value
 Q^2 = internal cross-validation coefficient
 R^2 = regression coefficient
 $R^2_{adjusted}$ = adjusted regression coefficient
 $R^2_{external}$ = testing regression coefficient
 $R^2_{scramble}$ = y -scrambling regression coefficient
RMSE = root-mean-square-error
 S_i = descriptor i
 \bar{S}_i = mean value of the descriptor i
 SD_{avg} = average standard deviation
SDR = standardized residual
 T = temperature, K
 T (superscript) = matrix transpose operator
 t_{Ratio} = t test ratio
 V = matrix of data and ES descriptors with dimensions of $p \times d^*$
 V^T = transpose of matrix V with dimensions of $d^* \times p$
 v_i = matrix of ES descriptors with dimensions of $1 \times d^*$
 v_i^T = transpose of row-vector v_i with dimensions of $d^* \times 1$
 $W_{k,input}$ = weight coefficient of each input with hidden neuron k
 x_j = mole fraction of component j , mol %
 y = generic property of the ES
 Y_k = linear combination of the inputs linked to hidden neuron k

Greek Symbols

- σ = surface charge density, e/Å
 S_{σ} -profile = area under the surface charge density distribution, e/Å²
 $P(\sigma)$ = σ -profile

Abbreviations

- AD = applicability domain
Ala = alanine
ANN = artificial neural network
ATPPB = allyltriphenyl phosphonium bromide
Bet = betaine

BTPC = benzyltriphenyl phosphonium chloride
CA = citric acid
ChCl = choline chloride
COSMO-RS = conductor-like screening model for real solvents
DEA = diethanolamine
DEEAC = *N,N*-diethylethanol ammonium chloride
DEG = diethylene glycol
DES = deep eutectic solvent
DFT = density functional theory
EAC = ethylammonium chloride
EG = ethylene glycol
ES = Eutectic solvent
Fru = fructose
BP86 = Becke–Perdew 86
Glu = glucose
Gly = glycerol
Glyi = glycine
GlyA = glycolic acid
H₂O = water
ILs = ionic liquids
LacA = lactic acid
MA = malic acid
MalA = malonic acid
MDEA = methyltriethanolamine
MEA = monoethanolamine
MLR = multiple linear regression
MTPPB = methyltriphenyl phosphonium bromide
OA = oxalic acid
OECD = organization for economic cooperation and development
Pro = proline
QSPR = quantitative structure–property relationship
Suc = sucrose
TBAC = tetrabutyl ammonium chloride
TEA = triethanolamine
TEG = triethylene glycol
TFA = 2,2,2-trifluoracetamine
TPAB = tetrapropyl ammonium bromide
TZVP = triple- ζ valence polarized

■ REFERENCES

- (1) Anastas, P. T.; Kirchhoff, M. M. Origins, Current Status, and Future Challenges of Green Chemistry. *Acc. Chem. Res.* **2002**, *35* (9), 686–694.
- (2) Clarke, C. J.; Tu, W. C.; Levers, O.; Bröhl, A.; Hallett, J. P. Green and Sustainable Solvents in Chemical Processes. *Chem. Rev.* **2018**, *118* (2), 747–800.
- (3) Carvalheda, C. A.; Campos, S. R. R.; Machuqueiro, M.; Baptista, A. M. Structural Effects of Ph and Deacylation on Surfactant Protein c in an Organic Solvent Mixture: A Constant-PH MD Study. *J. Chem. Inf. Model.* **2013**, *53* (11), 2979–2989.
- (4) Li, C.; Li, D.; Zou, S.; Li, Z.; Yin, J.; Wang, A.; Cui, Y.; Yao, Z.; Zhao, Q. Extraction Desulfurization Process of Fuels with Ammonium-Based Deep Eutectic Solvents. *Green Chem.* **2013**, *15*, 2793–2799.
- (5) Gani, R.; Jiménez-González, C.; Constable, D. J. C. Method for Selection of Solvents for Promotion of Organic Reactions. *Comput. Chem. Eng.* **2005**, *29* (7), 1661–1676.
- (6) Warrag, S. E. E.; Darwish, A. S.; Abuhatab, F. O. S.; Adeyemi, I. A.; Kroon, M. C.; Alnashef, I. M. Combined Extractive Dearomatization, Desulfurization, and Denitrogenation of Oil Fuels Using Deep Eutectic Solvents: A Parametric Study. *Ind. Eng. Chem. Res.* **2020**, *59* (25), 11723–11733.
- (7) Hatab, F. A.; Darwish, A. S.; Lemaoui, T.; Warrag, S. E. E.; Benguerba, Y.; Kroon, M. C.; Alnashef, I. M. *J. Chem. Eng. Data* **2020**, *65*, 5443.
- (8) Gao, J.; Zhang, J.; Li, H.; Li, L.; Xu, L.; Zhang, Y.; Wang, Z.; Wang, X.; Zhang, W.; Chen, Y.; Cheng, X.; Zhang, H.; Peng, L.; Chai, F.; Wei, Y. Comparative Study of Volatile Organic Compounds in Ambient Air Using Observed Mixing Ratios and Initial Mixing Ratios Taking Chemical Loss into Account – A Case Study in a Typical Urban Area in Beijing. *Sci. Total Environ.* **2018**, *628*–629 (8), 791–804.
- (9) Ranu, B. C.; Saha, A.; Dey, R. Using More Environmentally Friendly Solvents and Benign Catalysts in Performing Conventional Organic Reactions. *Curr. Opin. Drug Discovery Dev.* **2010**, *13* (6), 658–668.
- (10) Plechkova, N. V.; Seddon, K. R. Applications of Ionic Liquids in the Chemical Industry. *Chem. Soc. Rev.* **2008**, *37* (1), 123–150.
- (11) Fegade, S. L.; Trembley, J. P. Dispute over Designating an Ionic Liquid as a Green Solvent. *Chem. Eng. J.* **2015**, *273* (April), 668.
- (12) Deetlefs, M.; Seddon, K. R. Assessing the Greenness of Some Typical Laboratory Ionic Liquid Preparations. *Green Chem.* **2010**, *12* (1), 17–30.
- (13) Ghandi, K. A Review of Ionic Liquids, Their Limits and Applications. *Green Sustainable Chem.* **2014**, *04* (01), 44–53.
- (14) Romero, A.; Santos, A.; Tojo, J.; Rodríguez, A. Toxicity and Biodegradability of Imidazolium Ionic Liquids. *J. Hazard. Mater.* **2008**, *151* (1), 268–273.
- (15) Abbott, A. P.; Capper, G.; Davies, D. L.; Rasheed, R. K.; Tambyrajah, V. Novel Solvent Properties of Choline Chloride/Urea Mixtures. *Chem. Commun.* **2003**, *9* (1), 70–71.
- (16) 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.
- (17) Zhang, Q.; De Oliveira Vigier, K.; Royer, S.; Jérôme, F. Deep Eutectic Solvents: Syntheses, Properties and Applications. *Chem. Soc. Rev.* **2012**, *41* (21), 7108–7146.
- (18) Smith, E. L.; Abbott, A. P.; Ryder, K. S. Deep Eutectic Solvents (DESs) and Their Applications. *Chem. Rev.* **2014**, *114* (21), 11060–11082.
- (19) Martins, M. A. R.; Pinho, S. P.; Coutinho, J. A. P. Insights into the Nature of Eutectic and Deep Eutectic Mixtures. *J. Solution Chem.* **2019**, *48* (7), 962–982.
- (20) Abbott, A. P.; Harris, R. C.; Ryder, K. S.; D'Agostino, C.; Gladden, L. F.; Mantle, M. D. Glycerol Eutectics as Sustainable Solvent Systems. *Green Chem.* **2011**, *13* (1), 82–90.
- (21) Khandelwal, S.; Tailor, Y. K.; Kumar, M. Deep Eutectic Solvents (DESs) as Eco-Friendly and Sustainable Solvent/Catalyst Systems in Organic Transformations. *J. Mol. Liq.* **2016**, *215*, 345–386.
- (22) Florindo, C.; Oliveira, F. S.; Rebelo, L. P. N.; Fernandes, A. M.; Marrucho, I. M. Insights into the Synthesis and Properties of Deep Eutectic Solvents Based on Cholinium Chloride and Carboxylic Acids. *ACS Sustainable Chem. Eng.* **2014**, *2* (10), 2416–2425.
- (23) 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.
- (24) Chen, J.; Wang, Y.; Wei, X.; Xu, W.; Xu, P.; Ni, R.; Meng, J. First Investigation of the Micelles Forming in a Novel Deep Eutectic Solvents-Based Aqueous Micellar Two-Phase System: Partitioning of Cationic/Neutral/Anionic Pigments. *ACS Sustainable Chem. Eng.* **2019**, *7* (6), 6078–6092.
- (25) Tomé, L. C.; Mecerreyes, D. Emerging Ionic Soft Materials Based on Deep Eutectic Solvents. *J. Phys. Chem. B* **2020**, *124*, 8465–8478.
- (26) Hu, Y.; Liu, L.; Yu, J.; Wang, Z.; Fan, Y. Preparation of Natural Multicompatible Silk Nanofibers by Green Deep Eutectic Solvent Treatment. *ACS Sustainable Chem. Eng.* **2020**, *8* (11), 4499–4510.
- (27) Cunha, S. C.; Fernandes, J. O. Extraction Techniques with Deep Eutectic Solvents. *TrAC, Trends Anal. Chem.* **2018**, *105*, 225–239.

- (28) Florindo, C.; Romero, L.; Rintoul, I.; Branco, L. C.; Marrucho, I. M. From Phase Change Materials to Green Solvents: Hydrophobic Low Viscous Fatty Acid-Based Deep Eutectic Solvents. *ACS Sustainable Chem. Eng.* **2018**, *6* (3), 3888–3895.
- (29) Warrag, S. E. E.; Darwish, A. S.; Adeyemi, I. A.; Hadj-Kali, M. K.; Kroon, M. C.; Alnashef, I. M. Extraction of Pyridine from N-Alkane Mixtures Using Methyltriphenylphosphonium Bromide-Based Deep Eutectic Solvents as Extractive Denitrogenation Agents. *Fluid Phase Equilib.* **2020**, *517*, 112622.
- (30) Cruz, H.; Jordao, N.; Pinto, A. L.; Dionisio, M.; Neves, L. A.; Branco, L. C. Alkaline Iodide-Based Deep Eutectic Solvents for Electrochemical Applications. *ACS Sustainable Chem. Eng.* **2020**, DOI: [10.1021/acssuschemeng.9b06733](https://doi.org/10.1021/acssuschemeng.9b06733).
- (31) Chandran, D.; Khalid, M.; Walvekar, R.; Mubarak, N. M.; Dharaskar, S.; Wong, W. Y.; Gupta, T. C. S. M. Deep Eutectic Solvents for Extraction-Desulphurization: A Review. *J. Mol. Liq.* **2019**, *275*, 312–322.
- (32) Li, X.; Row, K. H. Development of Deep Eutectic Solvents Applied in Extraction and Separation. *J. Sep. Sci.* **2016**, *39* (18), 3505–3520.
- (33) Ruß, C.; König, B. Low Melting Mixtures in Organic Synthesis - An Alternative to Ionic Liquids? *Green Chem.* **2012**, *14* (11), 2969–2982.
- (34) Jibril, B.; Mjalli, F.; Naser, J.; Gano, Z. New Tetrapropylammonium Bromide-Based Deep Eutectic Solvents: Synthesis and Characterizations. *J. Mol. Liq.* **2014**, *199*, 462–469.
- (35) Hayyan, A.; Mjalli, F. S.; Alnashef, I. M.; Al-Wahaibi, T.; Al-Wahaibi, Y. M.; Hashim, M. A. Fruit Sugar-Based Deep Eutectic Solvents and Their Physical Properties. *Thermochim. Acta* **2012**, *541*, 70–75.
- (36) Uslu, H.; Bamufleh, H. S. Effect of Solvent and PH on the Extraction of Carabolic Acid from Aqueous Solution by TOMAC. *J. Chem. Eng. Data* **2016**, *61* (4), 1676–1680.
- (37) Brimecombe, R. D.; Limson, J. L. Electrochemical Investigation of the Effect of PH and Solvent on Amitraz Stability. *J. Agric. Food Chem.* **2006**, *54* (21), 8139–8143.
- (38) Pateli, I. M.; Thompson, D.; Alabdullah, S. S. M.; Abbott, A. P.; Jenkin, G. R. T.; Hartley, J. M. The Effect of PH and Hydrogen Bond Donor on the Dissolution of Metal Oxides in Deep Eutectic Solvents. *Green Chem.* **2020**, *22* (16), 5476–5486.
- (39) Skulcova, A.; Russ, A.; Jablonsky, M.; Sima, J. The PH Behavior of Seventeen Deep Eutectic Solvents. *BioResources* **2018**, *3*, 5042–5051.
- (40) Xu, F.; Sun, J.; Wehrs, M.; Kim, K. H.; Rau, S. S.; Chan, A. M.; Simmons, B. A.; Mukhopadhyay, A.; Singh, S. Biocompatible Choline-Based Deep Eutectic Solvents Enable One-Pot Production of Cellulosic Ethanol. *ACS Sustainable Chem. Eng.* **2018**, *6* (7), 8914–8919.
- (41) Lemaoui, T.; Benguerba, Y.; Darwish, A. S.; Hatab, F. A.; Warrag, S. E. E.; Kroon, M. C.; Alnashef, I. M. Simultaneous Dearomatization, Desulfurization, and Denitrogenation of Diesel Fuels Using Acidic Deep Eutectic Solvents as Extractive Agents: A Parametric Study. *Sep. Purif. Technol.* **2021**, *256* (3), 117861.
- (42) Yin, J.; Wang, J.; Li, Z.; Li, D.; Yang, G.; Cui, Y.; Wang, A.; Li, C. Deep Desulfurization of Fuels Based on an Oxidation/Extraction Process with Acidic Deep Eutectic Solvents. *Green Chem.* **2015**, *17* (9), 4552–4559.
- (43) Varela, A. S. The Importance of PH in Controlling the Selectivity of the Electrochemical CO₂ Reduction. *Curr. Opin. Green Sustain. Chem.* **2020**, *26*, 100371.
- (44) Porcelli, P. J.; O'Shea, T. M.; Dillard, R. G. A Linear Regression Model to Predict the PH of Neonatal Parenteral Nutrition Solution. *J. Clin. Pharm. Ther.* **2000**, *25* (1), 55–59.
- (45) Bi, S. A Simple Model for Predicting the PH Values of Acidic Natural Waters. *Anal. Chim. Acta* **1995**, *314* (1–2), 111–119.
- (46) Almustafa, G.; Sulaiman, R.; Kumar, M.; Adeyemi, I.; Arafat, H. A.; AlNashef, I. Boron Extraction from Aqueous Medium Using Novel Hydrophobic Deep Eutectic Solvents. *Chem. Eng. J.* **2020**, *395*, 125173.
- (47) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'Min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57* (12), 4977–5010.
- (48) Yousefinejad, S.; Hemmateenejad, B. Chemometrics Tools in QSAR/QSPR Studies: A Historical Perspective. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 177–204.
- (49) Hartman, G. H. A Short History Of. *Chicago Rev.* **1962**, *15* (4), 122.
- (50) Liu, X.; Wang, C.; Lan, T.; He, M.; Zhang, Y. Prediction of Thermal Conductivity for Guiding Molecular Design of Liquids. *ACS Sustainable Chem. Eng.* **2020**, *8* (15), 6022–6032.
- (51) Blay, V.; Gullón-Soleto, J.; Gálvez-Llompart, M.; Gálvez, J.; García-Domenech, R. Biodegradability Prediction of Fragrant Molecules by Molecular Topology. *ACS Sustainable Chem. Eng.* **2016**, *4* (8), 4224–4231.
- (52) Calvo-Serrano, R.; González-Miquel, M.; Guillén-Gosálbez, G. Integrating COSMO-Based σ -Profiles with Molecular and Thermodynamic Attributes to Predict the Life Cycle Environmental Impact of Chemicals. *ACS Sustainable Chem. Eng.* **2019**, *7* (3), 3575–3583.
- (53) Paduszyński, K. Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 1. Density. *Ind. Eng. Chem. Res.* **2019**, *58* (13), 5322–5338.
- (54) El-Harawi, M.; Samir, B. B.; Babaa, M. R.; Mutualib, M. I. A. A New QSPR Model for Predicting the Densities of Ionic Liquids. *Arab. J. Sci. Eng.* **2014**, *39* (9), 6767–6775.
- (55) Chen, B. K.; Liang, M. J.; Wu, T. Y.; Wang, H. P. A High Correlate and Simplified QSPR for Viscosity of Imidazolium-Based Ionic Liquids. *Fluid Phase Equilib.* **2013**, *350*, 37–42.
- (56) Paduszyński, K. Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity. *Ind. Eng. Chem. Res.* **2019**, *58* (36), 17049–17066.
- (57) Gardas, R. L.; Coutinho, J. A. P. Applying a QSPR Correlation to the Prediction of Surface Tensions of Ionic Liquids. *Fluid Phase Equilib.* **2008**, *265* (1–2), 57–65.
- (58) Bini, R.; Chiappe, C.; Duce, C.; Micheli, A.; Solaro, R.; Starita, A.; Tiné, M. R. Ionic Liquids: Prediction of Their Melting Points by a Recursive Neural Network Model. *Green Chem.* **2008**, *10* (3), 306–330.
- (59) Venkatraman, V.; Evjen, S.; Knuutila, H. K.; Fiksdahl, A.; Alsberg, B. K. Predicting Ionic Liquid Melting Points Using Machine Learning. *J. Mol. Liq.* **2018**, *264*, 318–326.
- (60) Zhao, X.; Pan, Y.; Jiang, J.; Xu, S.; Jiang, J.; Ding, L. Thermal Hazard of Ionic Liquids: Modeling Thermal Decomposition Temperatures of Imidazolium Ionic Liquids via QSPR Method. *Ind. Eng. Chem. Res.* **2017**, *56* (14), 4185–4195.
- (61) Venkatraman, V.; Alsberg, B. K. Quantitative Structure-Property Relationship Modelling of Thermal Decomposition Temperatures of Ionic Liquids. *J. Mol. Liq.* **2016**, *223*, 60–67.
- (62) Gharagheizi, F.; Sattari, M.; Ilani-Kashkouli, P.; Mohammadi, A. H.; Ramjugernath, D.; Richon, D. Quantitative Structure-Property Relationship for Thermal Decomposition Temperature of Ionic Liquids. *Chem. Eng. Sci.* **2012**, *84*, 557–563.
- (63) Jiang, J.; Duan, W.; Wei, Q.; Zhao, X.; Ni, L.; Pan, Y.; Shu, C. M. Development of Quantitative Structure-Property Relationship (QSPR) Models for Predicting the Thermal Hazard of Ionic Liquids: A Review of Methods and Models. *J. Mol. Liq.* **2020**, *301*, 112471.
- (64) Zhao, Y.; Huang, Y.; Zhang, X.; Zhang, S. Prediction of Heat Capacity of Ionic Liquids Based on COSMO-RS σ -Profile. *Comput.-Aided Chem. Eng.* **2015**, *37*, 251.
- (65) Zhao, Y.; Huang, Y.; Zhang, X.; Zhang, S. A Quantitative Prediction of the Viscosity of Ionic Liquids Using σ -Profile Molecular Descriptors. *Phys. Chem. Chem. Phys.* **2015**, *17* (5), 3761–3767.
- (66) Koi, Z. K.; Yahya, W. Z. N.; Abu Talip, R. A.; Kurnia, K. A. Prediction of the Viscosity of Imidazolium-Based Ionic Liquids at

- Different Temperatures Using the Quantitative Structure Property Relationship Approach. *New J. Chem.* **2019**, *43* (41), 16207–16217.
- (67) Ghanem, O. B.; Mutalib, M. I. A.; Lévéque, J. M.; El-Harbowi, M. Development of QSAR Model to Predict the Ecotoxicity of Vibrio Fischeri Using COSMO-RS Descriptors. *Chemosphere* **2017**, *170*, 242–250.
- (68) 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 σ -Profile as Molecular Descriptors. *J. Mol. Liq.* **2020**, *309*, 113165.
- (69) Lemaoui, T.; Darwish, A. S.; Hammoudi, N. E. H.; Abu Hatab, F.; Attou, A.; Alnashef, I. M.; Benguerba, Y. Prediction of Electrical Conductivity of Deep Eutectic Solvents Using COSMO-RS Sigma Profiles as Molecular Descriptors: A Quantitative Structure–Property Relationship Study. *Ind. Eng. Chem. Res.* **2020**, *59* (29), 13343–13354.
- (70) Gu, G. H.; Noh, J.; Kim, I.; Jung, Y. Machine Learning for Renewable Energy Materials. *J. Mater. Chem. A* **2019**, *7* (29), 17096–17117.
- (71) Tu, J. V. Advantages and Disadvantages of Using Artificial Neural Networks versus Logistic Regression for Predicting Medical Outcomes. *J. Clin. Epidemiol.* **1996**, *49* (11), 1225–1231.
- (72) Benguerba, Y.; Alnashef, I. M.; Erto, A.; Balsamo, M.; Ernst, B. A Quantitative Prediction of the Viscosity of Amine Based DESSs Using σ -Profile Molecular Descriptors. *J. Mol. Struct.* **2019**, *1184*, 357–363.
- (73) Mjalli, F. S.; Naser, J.; Jibril, B.; Alizadeh, V.; Gano, Z. Tetrabutylammonium Chloride Based Ionic Liquid Analogues and Their Physical Properties. *J. Chem. Eng. Data* **2014**, *59* (7), 2242–2251.
- (74) Saputra, R.; Walvekar, R.; Khalid, M.; Mubarak, N. M. Synthesis and Thermophysical Properties of Ethylammonium Chloride-Glycerol-ZnCl₂ Ternary Deep Eutectic Solvent. *J. Mol. Liq.* **2020**, *310*, 113232.
- (75) Ghaedi, H.; Ayoub, M.; Sufian, S.; Hailegiorgis, S. M.; Murshid, G.; Khan, S. N. Thermal Stability Analysis, Experimental Conductivity and PH of Phosphonium-Based Deep Eutectic Solvents and Their Prediction by a New Empirical Equation. *J. Chem. Thermodyn.* **2018**, *116*, 50–60.
- (76) Kareem, M. A.; Mjalli, F. S.; Hashim, M. A.; Alnashef, I. M. Phosphonium-Based Ionic Liquids Analogues and Their Physical Properties. *J. Chem. Eng. Data* **2010**, *55*, 4632.
- (77) Bahadori, L.; Chakrabarti, M. H.; Mjalli, F. S.; Alnashef, I. M.; Manan, N. S. A.; Hashim, M. A. Physicochemical Properties of Ammonium-Based Deep Eutectic Solvents and Their Electrochemical Evaluation Using Organometallic Reference Redox Systems. *Electrochim. Acta* **2013**, *113*, 205–211.
- (78) Mjalli, F. S.; Naser, J.; Jibril, B.; Alizadeh, V.; Gano, Z. Tetrabutylammonium Chloride Based Ionic Liquid Analogues and Their Physical Properties. *J. Chem. Eng. Data* **2014**, *59*, 2242–2251.
- (79) Mitar, A.; Panić, M.; Prlić Kardum, J.; Halambek, J.; Sander, A.; Zagajski Kučan, K.; Radojčić Redovniković, I.; Radošević, K. Physicochemical Properties, Cytotoxicity, and Antioxidative Activity of Natural Deep Eutectic Solvents Containing Organic Acid. *Chem. Biochem. Eng. Q.* **2019**, *33* (1), 1–18.
- (80) Alareeqi, S.; Bahamon, D.; Nogueira, R. P.; Vega, L. F. Understanding the Relationship between the Structural Properties of Three Corrosion Inhibitors and Their Surface Protectiveness Ability in Different Environments. *Appl. Surf. Sci.* **2021**, *542*, 148600.
- (81) Lloret, J. O.; Vega, L. F.; Llorell, F. Accurate Description of Thermophysical Properties of Tetraalkylammonium Chloride Deep Eutectic Solvents with the Soft-SAFT Equation of State. *Fluid Phase Equilib.* **2017**, *448*, 81–93.
- (82) Zubeir, L. F.; Held, C.; Sadowski, G.; Kroon, M. C. PC-SAFT Modeling of CO₂ Solubilities in Deep Eutectic Solvents. *J. Phys. Chem. B* **2016**, *120* (9), 2300–2310.
- (83) Gramatica, P. Principles of QSAR Models Validation: Internal and External. *QSAR Comb. Sci.* **2007**, *26* (5), 694–701.
- (84) Adeyemi, I.; Abu-Zahra, M. R. M.; AlNashef, I. M. Physicochemical Properties of Alkanolamine-Choline Chloride Deep Eutectic Solvents: Measurements, Group Contribution and Artificial Intelligence Prediction Techniques. *J. Mol. Liq.* **2018**, *256*, 581–590.
- (85) Shahbaz, K.; Baroutian, S.; Mjalli, F. S.; Hashim, M. A.; Alnashef, I. M. Densities of Ammonium and Phosphonium Based Deep Eutectic Solvents: Prediction Using Artificial Intelligence and Group Contribution Techniques. *Thermochim. Acta* **2012**, *527*, 59–66.
- (86) Gramatica, P.; Papa, E.; Sangion, A. QSAR Modeling of Cumulative Environmental End-Points for the Prioritization of Hazardous Chemicals. *Environ. Sci. Process. Impacts* **2018**, *20* (1), 38–47.
- (87) Roy, K.; Kar, S.; Ambure, P. On a Simple Approach for Determining Applicability Domain of QSAR Models. *Chemom. Intell. Lab. Syst.* **2015**, *145*, 22–29.
- (88) Melagraki, G.; Afantitis, A. Enalos KNIME Nodes: Exploring Corrosion Inhibition of Steel in Acidic Medium. *Chemom. Intell. Lab. Syst.* **2013**, *123*, 9–14.
- (89) Almi, I.; Belaidi, S.; Zerroug, E.; Alloui, M.; Ben Said, R.; Linguerri, R.; Hochlaf, M. QSAR Investigations and Structure-Based Virtual Screening on a Series of Nitrobenzoxadiazole Derivatives Targeting Human Glutathione-S-Transferases. *J. Mol. Struct.* **2020**, *1211*, 128015.
- (90) Oluwaseye, A.; Uzairu, A.; Shallangwa, G. A.; Abechi, S. E. Quantum Chemical Descriptors in the QSAR Studies of Compounds Active in Maxima Electroschock Seizure Test. *J. King Saud Univ., Sci.* **2020**, *32* (1), 75–83.
- (91) Hammoudi, N. E. H.; Benguerba, Y.; Attou, A.; Hognon, C.; Lemaoui, T.; Sobhi, W.; Benaicha, M.; Badawi, M.; Monari, A. In Silico Drug Discovery of IKK- β Inhibitors from 2-Amino-3-Cyano-4-Alkyl-6-(2-Hydroxyphenyl) Pyridine Derivatives Based on QSAR, Docking, Molecular Dynamics and Drug-Likeness Evaluation Studies. *J. Biomol. Struct. Dyn.* **2020**, *0* (0), 1–17.
- (92) Paris, J.; Telzerow, A.; Ríos-Lombardía, N.; Steiner, K.; Schwab, H.; Morís, F.; Gröger, H.; González-Sabín, J. Enantioselective One-Pot Synthesis of Biaryl-Substituted Amines by Combining Palladium and Enzyme Catalysis in Deep Eutectic Solvents. *ACS Sustainable Chem. Eng.* **2019**, *7* (5), 5486–5493.
- (93) Farias, F. O.; Passos, H.; Coutinho, J. A. P.; Mafra, M. R. PH Effect on the Formation of Deep-Eutectic-Solvent-Based Aqueous Two-Phase Systems. *Ind. Eng. Chem. Res.* **2018**, *57* (49), 16917–16924.
- (94) Abbott, A. P.; Alabdullah, S. S. M.; Al-Murshedi, A. Y. M.; Ryder, K. S. Brønsted Acidity in Deep Eutectic Solvents and Ionic Liquids. *Faraday Discuss.* **2018**, *206*, 365–377.
- (95) Hayyan, A.; Mjalli, F. S.; Alnashef, I. M.; Al-Wahaibi, Y. M.; Al-Wahaibi, T.; Hashim, M. A. Glucose-Based Deep Eutectic Solvents: Physical Properties. *J. Mol. Liq.* **2013**, *178*, 137–141.
- (96) Palomar, J.; Torrecilla, J.; Lemus, J.; Ferro, V. R.; Rodriguez, F. A COSMO-RS Based Guide to Analyze/Quantify the Polarity of Ionic Liquids and Their Mixtures with Organic Cosolvents. *Phys. Chem. Chem. Phys.* **2010**, *12* (8), 1991–2000.
- (97) Aissaoui, T.; Benguerba, Y.; AlOmar, M. K.; AlNashef, I. M. Computational Investigation of the Microstructural Characteristics and Physical Properties of Glycerol-Based Deep Eutectic Solvents. *J. Mol. Model.* **2017**, *23* (10), 277 DOI: [10.1007/s00894-017-3450-5](https://doi.org/10.1007/s00894-017-3450-5).
- (98) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify: On the Importance of Chemical Structure Curation in Cheminformatics and QSAR Modeling Research. *J. Chem. Inf. Model.* **2010**, *50* (7), 1189–1204.
- (99) Gramatica, P.; Cassani, S.; Sangion, A. Aquatic Ecotoxicity of Personal Care Products: QSAR Models and Ranking for Prioritization and Safer Alternatives' Design. *Green Chem.* **2016**, *18* (16), 4393–4406.
- (100) Lin, S. T.; Sandler, S. I. A Priori Phase Equilibrium Prediction from a Segment Contribution Solvation Model. *Ind. Eng. Chem. Res.* **2002**, *41* (5), 899–913.

- (101) Torrecilla, J. S.; Palomar, J.; Lemus, J.; Rodríguez, F. A Quantum-Chemical-Based Guide to Analyze/Quantify the Cytotoxicity of Ionic Liquids. *Green Chem.* **2010**, *12* (1), 123–13.
- (102) OECD Principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure–Activity Relationship Model. In *37th Joint Meeting of the Chemicals Committee and Working Party on Chemicals, Pesticides and Biotechnology*, November 2004.
- (103) Netzeva, T. I.; Worth, A. P.; Aldenberg, T.; Benigni, R.; Cronin, M. T. D.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D. W.; Schultz, T. W.; Stanton, D. T.; Van De Sandt, J. J. M.; Tong, W.; Veith, G.; Yang, C. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships. *ATLA, Altern. Lab. Anim.* **2005**, *33* (2), 155–173.