ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Development of an Infinite Dilution Activity Coefficient Prediction Model for Organic Solutes in Ionic Liquids with Modified Partial Equalization Orbital Electronegativity Method Derived Descriptors

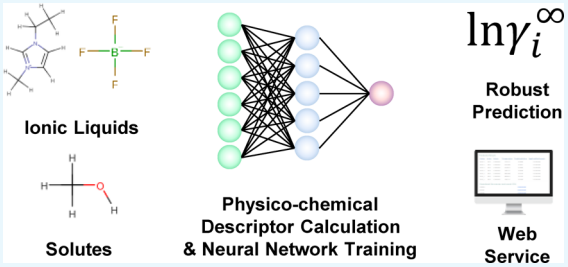
Hyeon-Nae Jeon, Hyun Kil Shin, Sungbo Hwang, and Kyoung Tai No*

Cite This: *ACS Omega* 2021, 6, 15361−15373

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The objective of this study was to develop a robust prediction model for the infinite dilution activity coefficients (γ^{∞}) of organic molecules in diverse ionic liquid (IL) solvents. Electrostatic, hydrogen bond, polarizability, molecular structure, and temperature terms were used in model development. A feed-forward model based on artificial neural networks was developed with 34,754 experimental activity coefficients, a combination of 195 IL solvents (88 cations and 38 anions), and 147 organic solutes at a temperature range of 298 to 408 K. The root mean squared error (RMSE) of the training set and test set was 0.219 and 0.235, respectively. The R^2 of the training set and the test set was 0.984 and 0.981, respectively. The applicability domain was determined through a Williams plot, which implied that water and halogenated compounds were outside of the applicability domain. The robustness test shows that the developed model is robust. The web server supports using the developed prediction model and is freely available at https://preadmet.bmdrc.kr/activitycoefficient_mainpage/prediction/.



1. INTRODUCTION

Ionic liquids (ILs) are liquid salts that exist as liquids even below room temperature, i.e., 298 K. Due to their wide range of innovative properties, such as negligible vapor pressure, high thermal stability, high chemical stability, and inflammability, ILs are considered promising materials in the industrial sector.

In particular, ILs are expected to be substituents of organic solvents owing to their nonvolatility.[1] Since the physicochemical properties of ILs depend on the combination of their constituent cations and anions, to select an optimal IL as the solvent of a solute, it is crucial to understand the interactions among the cation, anion, and solute.

The activity coefficient γ is a factor that describes the deviation from the ideal behavior of a mixture of substances. Thus, γ is crucial for estimating the solute mole fraction in a solution at equilibrium, at any temperature and for any mole fraction. The activity coefficient at infinite dilution γ^{∞} contains information on the interaction between the solute and solvent at a negligible solute concentration. Since γ^{∞} provides information about pure solute−solvent interactions, γ^{∞} is frequently used to estimate separation factors such as the selectivity S_{ij}^{∞} or the capacity k_i^{∞}, which are required information for designing separation processes.[2] Experimentally, γ^{∞} has been measured using gas−liquid chromatography,[3−8] a dilutor technique,[5,9,10] and vapor−liquid equilibria methods.[11,12] The temperature dependency of γ^{∞} is important for practical applications of the solute−solvent pair of a solution. Revelli et al. derived the temperature dependency of γ^{∞} from their experimental results as follows:[13]

$$\ln \gamma_i^{\infty} = a_i + \frac{b_i}{T} \tag{1}$$

where a_i and b_i are the coefficients of the ith solute. The coefficients of the ith solute differ for each solvent.

It is practically unfeasible to measure γ^{∞} for all possible combinations of IL-anions and IL-cations and solutes. Therefore, prediction methods for γ^{∞} facilitate the maximization of the utility of IL as a solvent. Two categories of methods have principally been used to develop γ^{∞} prediction models: one is a thermodynamic method, and the other is a Quantitative Structure−Property Relationship (QSPR) method. The first thermodynamic model for predicting the γ^{∞} of organic solutes in ILs was proposed by Diedenhofen et al. in 2003.[14] They used the COnductor-like Screening Model for Real Solvents (COSMO-RS) method,[15] which uses both quantum mechanical and statistical mechanical calculation to compute the thermodynamic properties of liquid mixtures. The

Received: March 29, 2021
Accepted: May 25, 2021
Published: June 3, 2021

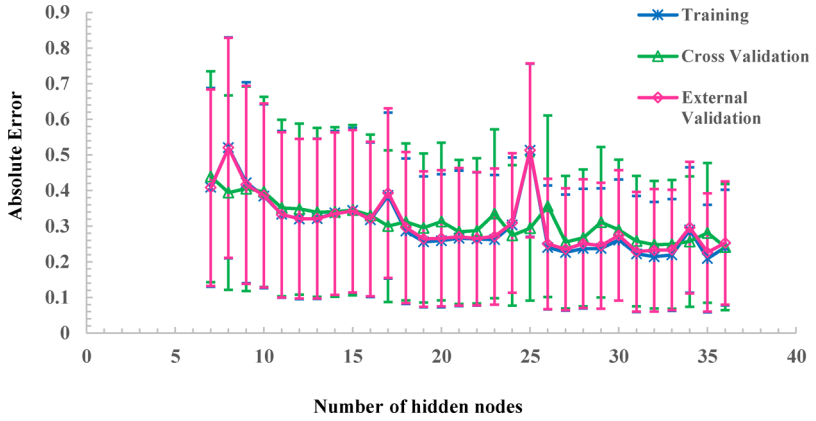


Figure 1. The parity plot of absolute error (AE), with the deviation from the training set, cross-validation, and external validation set by a variant of the hidden node number in FFANN. The error bars were plotted based on standard deviations of each data set. By considering the AE and the deviation for each hidden node number, the optimal hidden node number was selected as 31.

prediction was applied to 330 γ^{∞} data points containing three IL pairs: 1-ethyl-3-methylimidazolium with NTf2, 1-ethyl-2,3-dimethylimidazolium with NTf2, and 4-methyl-*N*-butylpyridinium with tetrafluoroborate. The RMSE of the prediction was 0.393, a reliable value for a prediction model. Wang et al. predicted γ^{∞}values of solutes in ILs using a UNIversial quasichemical Functional-group Activity Coefficients (UNIFAC) model.[16] The UNIFAC model predicts non-electrolyte activity coefficients in non-ideal mixtures by calculating the interactions of each functional group in molecules based on semi-empirical methods.[17] This model developed by Wang et al. achieved decent performance on two ILs: imidazoliums-NTf2 and imidazoliums-dimethyl phosphate. Mutelet et al.[18] calculated the γ^{∞} of organic solutes in 40 ILs using Linear Solvation free Energy Relationship (LSER) parameters.[19] The data set was analyzed using an Abraham solvation parameter model to determine the contributions of cations and anions. The model predicted a partition coefficient K_L, which is closely related to γ^{∞}, of the organic solutes in ILs that contain alkyl-based cations to within 0.13 log units. Recently, Paduszyński has developed a prediction model based on a COSMO-RS with 43,820 experimental activity coefficients as constraints for IL solvent data at infinite dilution.[20] This model is the first comprehensive evaluation in γ^{∞} prediction and works well particularly at low γ^{∞} values.

A QSPR method, based on various types of descriptors, was applied to simplify the approach. Multiple linear regression (MLR) and artificial neural networks (ANNs) have been widely used in knowledge-based prediction model development. Eike et al. developed three MLR γ^{∞} prediction models for organic solutes in three types of ILs.[21] For the MLR models, four solute properties were introduced as descriptors: an octanol/water partition coefficient, the number of hydrogen bonds, the surface-weighted partial negative surface area, and the sum of the E-state values for carbon atoms. However, the properties of the ILs were not introduced as descriptors of the MLR. The model predicted the experimental γ^{∞} within a 30% error without using a first-principles-based approach. Xi et al.[22] developed a QSPR model using an ordinary least-square regression algorithm to describe the temperature dependency of γ^{∞} of solutes in one IL, i.e., trihexyl(tetradecyl)-phosphonium bis(trifluorosulfonyl)imide. The descriptors were selected through evolution using a genetic algorithm, and the descriptors selected were the reciprocal temperature, relative positive charge, energy gap between the highest occupied molecular orbital and the lowest unoccupied molecular orbital, total charge weighted partial positive surface area, minimum valency of a carbon atom, and the maximum valency of a carbon atom. The squared correlation coefficient value of the predicted γ^{∞} of the external validation set was high (0.938). This study concluded that polar interaction is the preferential determinant of γ^{∞}. Paduszyński developed three models using StepWise MLR (SWMLR), Least Square Support Vector Machine (LSSVM), and Feed-Forward Artificial Neural Network (FFANN).[23] The models proposed by Paduszyński have a wider applicability domain than the previously developed models. For the development of γ^{∞} prediction covering a wide range of ILs, the author surveyed experimental data from the literature published since 2001. The SWMLR showed less than a 40% average absolute relative deviation (AARD), whereas the LSSVM and FFANN models gave an AARD of less than 20%.

The purpose of this work is to provide a computational tool to select proper ILs as solvents for any target organic solute, by providing the γ^{∞} value for IL solutions. Since the number of possible combinations of cations and anions of ILs is huge, one cannot perform mechanics-based simulation, such as molecular dynamics or Monte Carlo simulations. Instead, it is necessary to develop a knowledge-based prediction model.

To develop a knowledge-based prediction model, the choice of appropriate descriptors is important. In this case, the descriptor should reflect prior knowledge of the underlying physics of the interactions among the ion pairs in ILs and solutes, irrespective of the type of IL. Electrostatic interactions are the principal contributors to the total free energy of the solvation of the solute to IL. Such interactions include dipoles, induced dipoles, and higher perturbation interactions among the cation, anion, and solute in an IL solution. Therefore, it is important to describe electrostatic interactions accurately. There are a large number of possible IL solutions, which are combinations of cations, anions, and solutes. Therefore, using a practical approach, this study aims to describe electrostatic interactions with monopole−monopole (coulombic) and monopole-induced dipole interactions (polarization). The m-PEOE method[24−28] and CDEAP method[29] were used to describe monopole−monopole interactions and monopole-induced dipole interactions. For more precise descriptions of

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

the interactions, hydrogen bond descriptors and basic molecular properties were also used.

Since interactions in IL solutions are too complex to describe using linear functions, a prediction model was developed based on FFANN and a descriptor pool containing cross-term descriptors. The model developed herein can be used by scientists unfamiliar with cheminformatics or programming, and the model does not require any installation on users' desktops, as the model can be run in a web browser.

2. RESULTS AND DISCUSSION

2.1. Determining the Optimal ANN Architecture. In the model-building process, the performance values were explored for different hidden node numbers after the hidden layer number was fixed to 1 because an overfitting problem occurred when the hidden layer number was larger than 2. Equivalently, the model was overfitted to the training set when the hidden layer number was larger than 1. The hidden node number was changed from 7 to 36, as determined based on the absolute errors (AEs) and their deviations. The performance values for different hidden node numbers are presented in Figure 1, in which the AEs are illustrated for training, cross-validation, and external validation. Despite fluctuation in the AEs, a decreasing tendency was observed with every increment of the hidden node number. However, when the hidden node number exceeded 31, there was no significant decreasing tendency in the AEs. Thus, it is concluded that more weight parameters do not make a significant difference to the model. Finally, the optimal hidden node number was set to 31, accounting for the AEs, deviation in the AEs, and the number of neural network weight parameters. The weight value matrix of the final FFANN topology is presented in Table S1a,b.

2.2. Performance of the Model. The statistics for the performance of the final model are shown in Table 1, and the prediction results for the training set and the external validation set are depicted in Figure 2a,b.

Table 1. Statistics for the Model Results

criteria	training set	cross-validation	test set
RMSE	0.219	0.262	0.235
AE	0.151	0.185	0.158
R^2	0.984	0.980	0.981

In Table 1, the RMSE values for the training set, cross-validation, and external validation set are 0.219, 0.262, and 0.235, respectively. The R^2 values are 0.984, 0.980, and 0.981, respectively. Since these values showed no significant differences for any of the data sets regarding the RMSE and R^2, it was concluded that the FFANN model was not overfitted. All data sets show high squared correlation coefficient (R^2) values, specifically larger than 0.9. Thus, the results may be interpreted as showing that the performance of the model is acceptable and that the model may make accurate predictions for γ^{∞} in various combinations of cations, anions, and solutes. The experimental and predicted γ^{∞} values are included in Table S2.

2.3. Model Description. Since each subgroup contains at least one descriptor for calculating the activity coefficient, it is possible to reflect the overall interaction. To explain the polarization, descriptors of the local dipole, charge polarization, and CDEAP were used. The LocalDipole and ChargePolarization descriptors simply describe the distribution of the polarities in a molecule. These polarization terms were not used in the cross-term because the complexity of the FFANN's structure was sufficient to describe a first-order polarization perturbation term for the system. Electrostatic potential is a key factor for electrostatic interaction. Thus, AvgNegESP and AvgPosESP for each molecule were chosen for the model due to representing their molecule's electrostatic property. Hence, the AvgNegESP of cations and AvgPosESP of anions are not dominant in each of their molecules, and they were not used. WNSA3 and WPSA3 were used to reflect the effective charge−charge interactions. Like the former descriptors, WNSA3 for cations and WPSA3 for anions were not used. To describe more complex interactions between solutes and cations and between solutes and anions, ESPD descriptors were chosen in the model-building process.

Next, NegEonHBA, PosEonHBD, HRNCG, HBMix_RI/T, and CSWHBMix3/T were selected as the hydrogen bond descriptors. NegEonHBA, PosEonHBD, and HRNCG describe how the molecule easily makes a hydrogen bond. HBMix_RI/T and CSWHBMix3/T were used as temperature-dependent forms to explain the decrease of a hydrogen bond.

Finally, SAVR, HPhobSA, RBF, MVR, MWR, and AAF were used to describe the molecule's basic properties. As rotatable bonds and aromaticity significantly affect the properties of cations, RBF and AAF were used for cations.

2.4. Importance of Descriptors with Examples. The RMSE increments by shuffling descriptor values for each

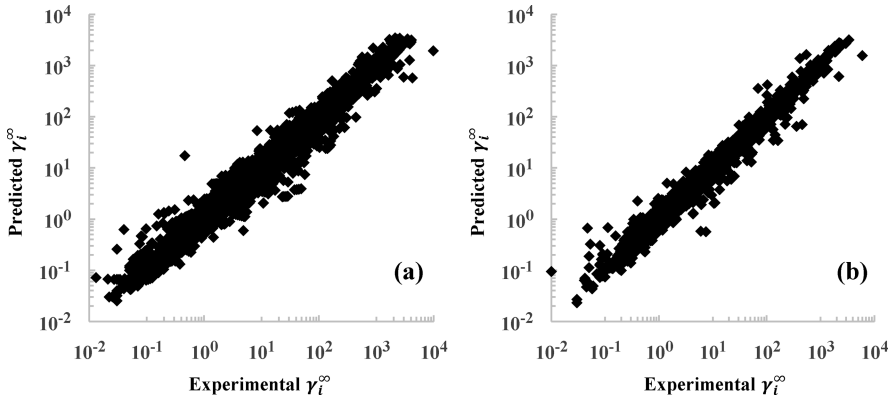


Figure 2. The prediction results of the developed model. (a) Experimental IDACs versus predicted IDACs for the training data set. (b) Experimental IDACs versus predicted IDACs for the external validation data set.

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

descriptor are listed in Table S3. The most distinctive increment conveying descriptor was considered to be crucial for the model performance. HRNCG_A shows the highest RMSE increment value, 4.885, which is followed by ChargePolarization_A of which the value is 4.380. To explain their role in the prediction, we selected example solute−IL combinations as follows: 1-ethyl-3-methyl-1H-imidazol-3-ium (EMIM) for cation; bis(trifluoromethylsulfonyl)azanide (NTF2), dicyanoazanide (DCA), and tricyanomethide (CCN3) for anions; and ethanol and n-hexane for solutes. The HRNCG_A and ChargePolarization values of the selected anions are listed in Table 2, and the experimental γ^{∞} values of

Table 2. Descriptor Values of Selected Anions

descriptors\anions	[NTF2]−	[DCA]−	[CCN3]−
HRNCG_A	0.1547	0.3599	0.2872
ChargePolarization_A	6.320	1.659	4.046

Table 3. Experimental Infinite Dilution Activity Coefficients of the Selected Solute-IL Combinations

cation	anions	solutes	temperatures (K)	γ^{∞}
EMIM	NTF2	ethanol	328.15	0.3016
	NTF2	n-hexane	328.15	3.127
	DCA	ethanol	328.15	−0.2769
	DCA	n-hexane	328.15	4.653
	CCN3	ethanol	328.15	0.1570
	CCN3	n-hexane	328.15	3.953

selected solute−IL combinations are listed in Table 3. The HRNCG_Anion represents the relative electron enrichment in a hydrogen bond acceptor atom to others. It has a large positive value when a hydrogen bond acceptor atom in an anion has a large negative atomic point charge. In this example, DCA has the highest HRNCG values and makes the highest attractive interaction (the lowest γ^{∞}) with ethanol compared to CCN3 and NTF2. The ChargePolarization descriptor is related to dipole strength. It has a large positive value when atomic point charges in an anion are disproportionally distributed. In this case, DCA has the lowest ChargePolarization values and makes the lowest attractive interaction (the highest γ^{∞}) with n-hexane compared to CCN3 and NTF2.

2.5. Y-Randomization Test. The results for the Y-randomization test are depicted in Figure 3 and Table 4. Specifically, the RMSE values from the training set and cross-validation set of the random models and the model in this study are presented in Figure 3, with the RMSE values for the randomized models being approximately 6 times higher than those for the QSPR model. The average R^2, $R_{CV}{}^2$, RMSE, and $RMSE_{CV}$ of randomized models are listed in Table 4. And ${}^{c}R_{p}{}^{2}$ value, which should be more than 0.5 to pass the test, is 0.979. Hence, the correlation of the chosen model has not occurred by chance. The RMSE and squared correlation coefficient values of randomized models are listed in Table S4.

2.6. Applicability Domain of the Model. Figure 4 shows a plot of the leverages and standard residuals of the descriptor set used in the training process. Where h_i is greater than the warning leverage h^*, this implies that the descriptors for molecule i differ considerably from the descriptors for other

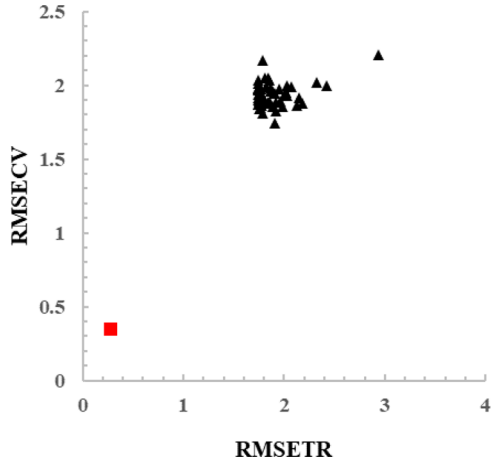


Figure 3. RMSE value comparison between the QSPR model and Y-randomized models. RMSETR is the RMSE value for the training set, and RMSECV is the RMSE value for the cross-validation set.

Table 4. Performance Comparison of Y-Randomized Models with Developed Model

	R^2	$R_{CV}{}^2$	RMSE	$RMSE_{CV}$
developed model	0.984	0.980	0.219	0.262
randomized models (averages)	0.0095	0.0001	1.809	1.865

molecules. Thus, the prediction for molecule i may be regarded as extrapolated and thus unreliable.

In Figure 4, there are no outliers for the standardized residual axis, but 493 outliers were detected for the leverage axis. The major component of the outliers was analyzed from three perspectives: cations, anions, and solutes. From the perspective of cations, various types of cations were included, reducing the propensity to specify whether the cations dominate the outlier proportion. The anions showed similar tendencies to the cations. There were no dominant anion types. Yet solutes with halogens (fluorine and chlorine) or water account for nearly 55% of the outliers. In particular, 43% of the total data for water are outliers. Additionally, the predicted results conveyed that the combination containing water (IL with solute) produced a higher error rate than other types of combinations. Ergo, outlier analysis demonstrated that the specific type of solute greatly affects the proportion of outliers.

Thus, it is concluded that the descriptions for water solutes and solutes with halogen atoms were extrapolated, meaning that the descriptors used in the model-building process were insufficient to describe the hydrogen bond system affected by water or the interactions of halogen atoms. Therefore, the model developed in this study covers the systems containing imidazolium; piperidinium; pyrrolidinium; pyridinium; morpholinium; ammonium; phosphonium; sulfonium-based cations; non-transition metal-based anions; and various types of organic compounds including alcohols, amines, carboxylic acids, esters, ethers, nitriles, ketones, and aldehydes, except for water and halogenated hydrocarbons.

2.7. Robustness Test. The experimental and predicted γ^{∞} values are included in Table S5. The plot in Figure 5 represents the experimental γ^{∞} and predicted γ^{∞}. The RMSE value is 0.397, and the R^2 value is 0.954. In Figure 5, the data points are distributed with a positive correlation. The performance

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

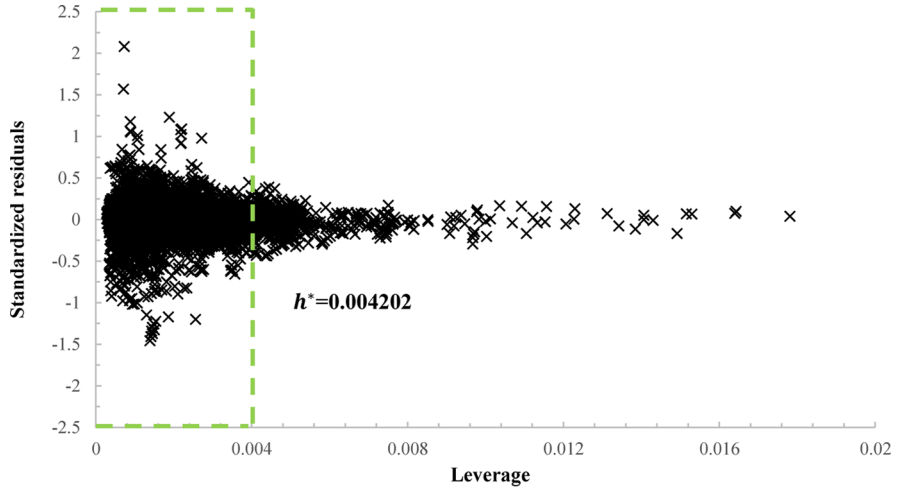


Figure 4. Williams plot for the descriptor set used in the training process. The green broken lines were drawn based on standardized residuals and warning leverage threshold. A total of 493 outliers exceeded the warning leverage threshold.

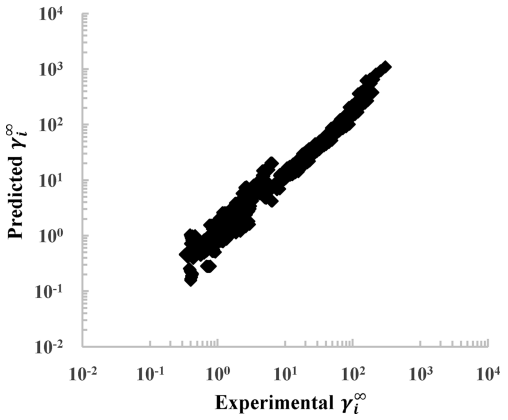


Figure 5. Robustness test result. The experimental data that were included in neither the training set nor the test set were predicted by using the developed model. The predicted values and experimental values were plotted.

metrics and point distribution in Figure 5 imply that the developed model is robust to new data.

2.8. Performance Comparison with Previously Reported Models. The performance comparison is represented in Table 5. Specifically, the RMSE values of our model on

Table 5. RMSE Comparison between the Developed Model and Reported Models

	developed model	SWMLR (ref 23)	FFANN (ref 23)	LSSVM (ref 23)
training	0.219	0.453	0.181	0.174
test	0.235	0.465	0.230	0.222
robustness test	0.397	0.586	2.404	1.499

training and test data sets were lower than those calculated from the SWMLR model but were marginally higher than those from either FFANN or LSSVM models. Nevertheless, our model overperformed all other benchmark models when it was applied to the robustness test data set. Because the robustness test data set contains no duplicate cation and anion combinations in the original data set, we concluded that our model is more robust and generalized than models in reference. The compared prediction results are listed in Table S6.

2.9. Web Server. The developed prediction model was reimplemented in Java and used to construct a web server. The user can predict and download γ^∞ values by uploading cation, anion, and solute files and confirming whether the predicted value is in or out of the domain. The example molecule set is also provided. An example output table returned from the web server is shown in Figure 6. The web server is freely accessible in the public domain at https://preadmet.bmdrc.kr/activitycoefficient_mainpage/prediction/.

3. CONCLUSIONS

In this study, a FFANN model was built based on physically meaningful descriptors to predict activity coefficients for organic solutes in ILs. The high R^2 and low RMSE values for the developed model indicate that the m-PEOE-IL- and CDEAP-derived descriptors are sufficient for describing the intermolecular interactions between a solute and ILs or between cations and anions. A Y-randomization test showed that the model developed was not built by chance but was solely built on the correlations of physically meaningful descriptors for γ^∞. Applicability domain analysis further validated the developed model's applicability for addressing various types of IL solvents and organic molecule systems, except in the case of solute organic molecules containing water or halogens. The robustness test data set was applied to the developed model, and on this data set, the model showed good performance. Based on these results, it is concluded that the model is robust for predicting γ^∞ for various organic solutes in ILs with great accuracy. And the performance comparison with three previously developed models shows that our model is more robust and generalized. Equivalently, the employed descriptors sufficiently predict molecular interactions and so can be applied to predict γ^∞ values of a new combination of ILs and solutes. Since water is not included in the applicability domain of the developed model, future research will investigate the description of the replenishment of hydrogen bond interactions. So too, an extension of the applicability domain of m-PEOE-IL and CDEAP will be challenged in future research to produce a quick approach for γ^∞ prediction. The developed model is freely available as a web server, and

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

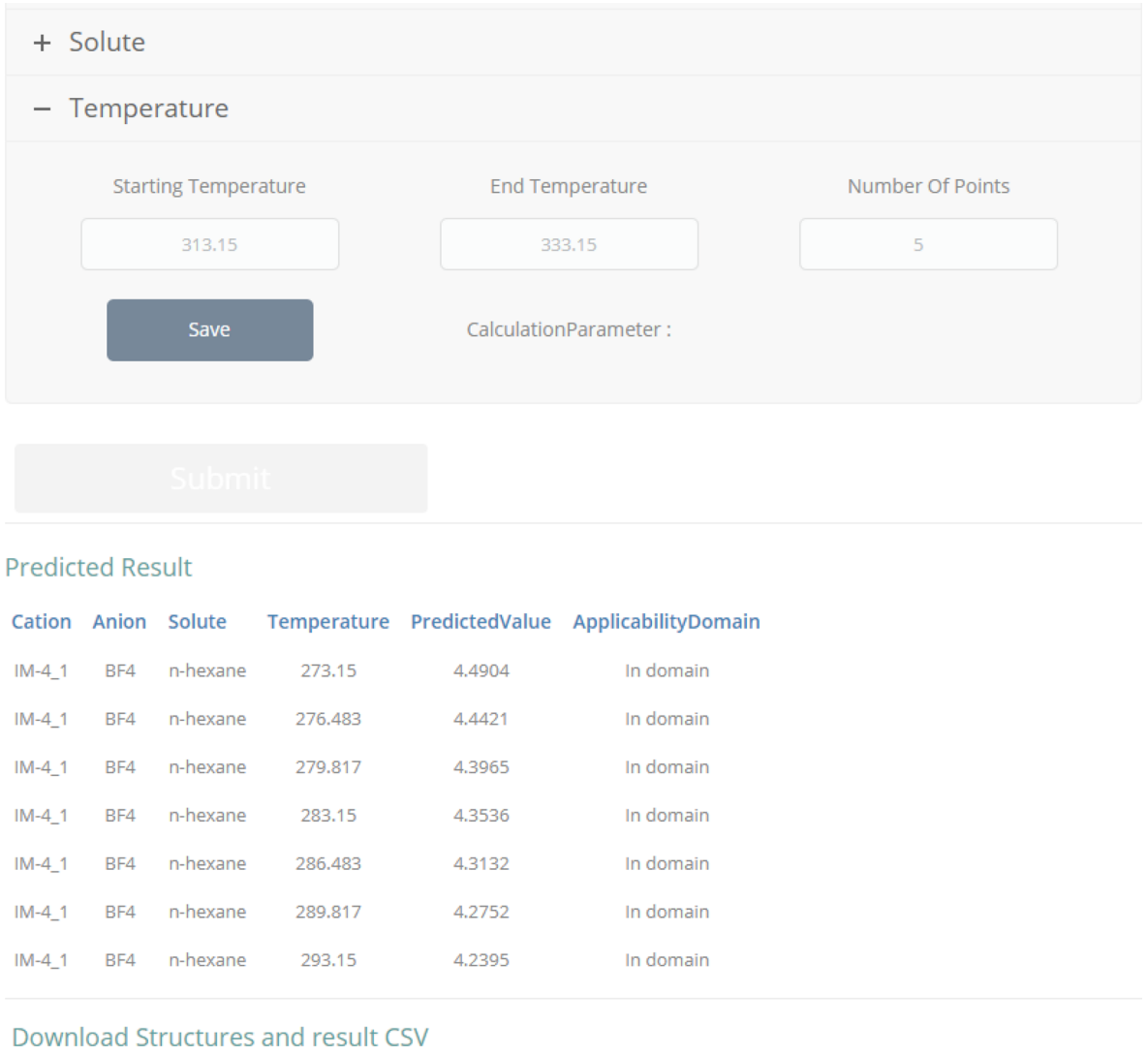


Figure 6. Sample result table from the developed web server. In this example, a test calculation was performed. The ″Download″ section exists below the prediction table.

accurate and convenient predictions from this model may be useful to researchers interested in ILs.

4. COMPUTATIONAL DETAILS

4.1. Database Construction for Prediction Model Training. The γ^{∞} data set in this study was mainly constructed using Paduzynski's database,[20] which contains 41,868 experimental γ^{∞} measurements. When constructing the data set, the measurement with the smallest experimental error was selected if there were more than two measurements for the same solute and IL. Then, experimentally measured γ^{∞} values for a 1,4-butanediol solute,[12] which were not included in Paduzynski's database, were included in data set. Finally, data points with ambiguous temperature dependency and data points measured at a single temperature were removed. After the above procedure, 34,754 data points remained for the γ^{∞}prediction model development. Then, the data set was randomly divided into a training set (27,804 data points, 80%) and a test set (6950 data points, 20%) by using an in-house code implemented in JAVA. The basic chemical information (the abbreviation, molecule name, molecular weight, charge, and SMILES) of the cations, anions, solutes, and IL combinations is summarized in Table S7−S10. The references for the database are included in the Electronic Supporting Information.

4.2. Description of the Energy of the IL Solution System. Since the activity coefficient of the IL solution depends on the intermolecular interactions among the cation, anion, and solute in an IL solution, the total energy of interaction of the system may be described as follows:

$$V_{\text{total}} = V_{\text{polar}} + V_{\text{shape}} \quad (2)$$

$$V_{\text{polar}} = V_{\text{ES}} + V_{\text{HB}} + V_{\text{POL}} \quad (3)$$

where V_{ES}, V_{HB}, and V_{POL} are the contribution from electrostatic interaction, hydrogen bonding, and polarization interactions, especially in high dielectric environments. V_{shape} represents the stabilization energy contributed by the degree of contact through the surface among the anion, cation, and solute.

Although there may be higher perturbation terms apart from the terms in eq 3, only highly contributing terms were introduced: terms for electrostatic, polarization, and hydrogen bond interaction. The shape of the components of the IL

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

Table 6. The New m-PEOE Parameters for the Cations

parameter	description	atom	A	B	initial charge
142	S3+	S	2.131	24.923	0.70
211	Csp3-S3+	C	11.705	2.001	0.10
152	P4-	P	15.672	15.309	0.00
212	Csp3-P4-	C	12.243	2.007	0.25
213	Nimi-Car-Nar+in pyridine	C	2.029	43.209	0.35
232	Nimi	N	35.438	38.366	0.00
214	Car-Nimi	C	37.518	44.882	0.00
215	Car-Nar+in pyridine	C	2.285	43.400	0.35
231	Ntriaz	N	17.593	21.033	0.00
216	Csp3-Npip	C	10.807	6.071	0.35
217	Csp3-Csp3-Npip	C	2.715	44.695	0.00
218	Csp3-Nmor	C	2.005	29.798	0.35
219	Csp3-Csp3-Nmor	C	10.853	38.618	0.00
311	Csp3-Npyrol	C	6.099	22.749	0.35
135	Npyrid	N	27.892	44.958	−0.05
312	Csp3-ortho-pyrid	C	8.238	13.658	0.35
313	Csp3-orthoSub-pyrid	C	8.637	1.002	0.35
314	Csp3-meta-pyrid	C	11.669	43.766	0.0
315	Csp3-metaSub-pyrid	C	18.925	21.993	0.0
316	Csp3-para-pyrid	C	12.682	15.943	0.0
317	Csp3-paraSub-pyrid	C	14.559	1.171	0.0

Table 7. The New m-PEOE Parameters for the Anions

parameter	description	atom	A	B	initial charge
82	B in borate ring	B	3.431	1.556	1.40
84	B bounded to 3F	B	7.329	1.340	1.00
523	O in borate ring	O	17.323	12.860	−0.60
611	C in sulfonate	C	5.166	36.128	0.10
651	P in octahedral phosphate3	P	6.281	2.940	0.80
711	dionate	C	1.722	6.839	0.60
714	dionate	C	6.977	2.122	−0.10
717	C in cyanamide	C	19.680	10.660	−0.30
722	dionate	O	11.594	15.133	−0.70
736	amide	N	21.774	25.520	−1.40
737	N in cyanamide	N	26.099	12.517	0.80
751	phosphateOct4	P	5.379	2.940	1.00
837	cyanamide	N	27.257	12.164	−0.60
841	S in sulfonate	S	3.272	3.919	0.40
871	borate3F	F	21.553	25.163	−0.50
914	sulfonamide	C	9.293	2.004	0.20
919	borate3F	C	18.743	16.870	−0.50
927	sulfinate	O	13.839	13.260	−1.30
941	sulfonamide	S	1.210	6.821	0.80
945	sulfinate	S	7.900	3.919	1.60
8725	sulfonate	O	13.260	15.354	−0.50
8819	PFA2	C	6.977	3.603	1.00
8871	PFA2	F	24.226	25.741	−0.50
8919	PFA3	C	5.166	4.967	1.05
8925	sulfate1	O	14.997	13.839	−0.65
8971	PFA3	F	13.260	6.483	−0.35
9671	phosphateOct3	F	24.447	25.741	−0.60
9725	phosphate3	O	21.852	8.582	−1.00
9771	phosphateOct4	F	17.449	25.741	−0.40
9925	phosphate1	O	21.773	18.365	−0.60

Table 8. CDEAP Parameters

parameter	description	atom	initial polarizability	displacement
C0	C	C	1.490	1.110
C1	Csp2(ethylene)	C	1.517	0.568
C2	Csp2(aromatic)	C	1.450	0.763
C3	Csp2(carbonyl)	C	1.253	0.862
C4	Csp3	C	1.031	0.590
H1	Hsp3	H	0.396	0.219
H2	Hsp2(aromatic)	H	0.298	0.404
O1	Osp2	O	0.720	0.347
O2	Osp3	O	0.623	0.281
N0	N	N	0.980	0.310
N1	Nsp2(aromatic, pyrrole)	N	0.871	0.424
N2	Nsp2(aromatic, pyridine)	N	0.656	0.436
N3	Nsp2(amide)	N	0.821	0.422
N4	Nsp3	N	0.966	0.437
B4[a]	Bsp3	B	1.096	0.743
S1	S	S	2.688	1.319
S6	S6	S	5.152	−1.730
F1	F	F	0.226	0.144
Cl1	Cl	Cl	2.180	1.089
Br1	Br	Br	3.114	1.402
I1	I	I	5.166	2.573
P5	P5	P	11.101	−7.006

[a]Newly added parameter.

solution also plays an important role in stabilizing the system. For example, a high-density IL solution means that the components of the IL solution are well packed and contribute to the stabilization of the system. In addition, since information on the temperature dependency of γ^{∞} is crucial for handling IL solutions, the model was developed to express the temperature dependence of target IL solutions.

The temperature dependency of γ^{∞} is described by eq 1, which states that the γ^{∞} is proportional to the reciprocal values of the experimental temperatures. Hence, the descriptors calculated in Sections 4.2.1−4.2.4 do not reflect temperature dependency, and temperature-dependent descriptors were calculated by dividing the descriptor vectors by constants, the experimental condition temperatures. Thus, five types of descriptors were calculated (electrostatic, hydrogen bonding, polarization, molecular structure, and temperature-dependent).

4.2.1. Description of Electrostatic Interactions in the IL Solution. To describe the electrostatic interactions among cations, anions, and polar solutes, an atom-centered point charge model was used. For the net atomic charge calculation, a modified-Partial Equalization of Orbital Electronegativity (m-PEOE) method[24−28] was used. This method is a modified version of the PEOE method developed by Gasteiger,[30] which is based on the partial equalization of orbital electronegativity between covalently bonded atoms in a molecule. The net atomic charge on atom A in the molecule is obtained using an iterative procedure in which charge transfer takes place between A-connected atoms and atom A. The iteration steps are continued until a fractional charge transfer reaches its threshold, i.e., 0.0001.

The electronegativity of the *i*th atom is expressed in the following equation:

$$\chi_i^{<n>} = a_i + b_i Q_i^{<n>} \quad (4)$$

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

Table 9. Descriptors Used in the FFANN Model

descriptor	examination
1000/T	1000/temperature (K)
LnT	natural logarithm of temperature (K)
AAF_CS	aromatic atom fraction ratio between cations and solutes
AvgNegESP_A	average negative ESP of anions
AvgNegESP_S	average negative ESP of solutes
AvgPosESP_C	average positive ESP of cations
AvgPosESP_S	average positive ESP of solutes
CSWHBMix3_ILS/T	charge and surface-weighted hydrogen bond effect between ILs and solutes
ChargePolarization_A	average absolute atomic charge in anions
ChargePolarization_C	average absolute atomic charge in cations
ChargePolarization_S	average absolute atomic charge in solutes
ESPD_SA	ESP difference between solutes and anions
ESPD_SC	ESP difference between solutes and cations
RBF_C	rotatable bond fraction in cations
HBMix_RI/T	relative hydrogen bond effect in a mixture, divided by temperature
HPhobSA_A	hydrophobic surface area of anions
HPhobSA_C	hydrophobic surface area of cations
HPhobSA_S	hydrophobic surface area of solutes
HRNCG_A	relative negative charge of HBA atoms in anions
HRNCG_C	relative negative charge of HBA atoms in cations
LocalDipole_A	average charge difference between two covalently bonded atoms in anions
LocalDipole_C	average charge difference between two covalently bonded atoms in cations
LocalDipole_S	average charge difference between two covalently bonded atoms in solutes
Polarizability_A	polarizability of anions, calculated using the CDEAP method
Polarizability_C	polarizability of cations, calculated using the CDEAP method
Polarizability_S	polarizability of solutes, calculated using the CDEAP method
MVR_ILS	molar volume ratio between ILs and solutes
MWR_SIL	molecular weight ratio between solutes and ILs
NegEonHBAAtom_A	negative electrostatic potential of HBA atoms in anions
NegEonHBAAtom_C	negative electrostatic potential of HBA atoms in cations
PosEonHBDAtom_S	positive electrostatic potential of HBD atoms in solutes
SAVR_C	surface area-to-volume ratio of cations
SAVR_S	surface area-to-volume ratio of solutes
WNSA3_A	negative charge weighted surface area type 3 for anions
WNSA3_S	negative charge weighted surface area type 3 for solutes
WPSA3_C	positive charge weighted surface area type 3 for cations
WPSA3_S	positive charge weighted surface area type 3 for solutes

where $\chi_i^{<n>}$ is the atomic electronegativity in the nth iteration, a_i is the m-PEOE parameter for the initial charge, b_i is another m-PEOE parameter for the charge transfer, and $Q_i^{<n>}$ is the charge transferred in the nth iteration. The iterative procedure of PEOE begins with the assignment of the initial charges to each atom in the molecule, as follows:

$$\chi_i^{<0>} = a_i + b_i Q_i^{<0>} \tag{5}$$

For this approach to ILs, the initial charge $Q_i^{<0>}$ value was set to +1 for a cation and −1 for an anion. The procedure for obtaining the net atomic charge of a molecule is well explained elsewhere in our previous works.[24−28]

In the iterative procedure for charge transfer, the transfer of the fractional charges between all the covalent bonding pairs of atoms becomes zero. The fractional charge transfer between atoms A and B, which are covalently bonded, can be expressed as follows:

$$dq^{<n>} = \frac{\chi_B^{<n-1>} - \chi_A^{<n-1>}}{\chi_A^{+}} (f_{AB})^n \text{ if } \chi_B^{<n-1>} > \chi_A^{<n-1>} \tag{6}$$

where χ_A, χ_B, and f_{AB} are the electronegativity of atom A, the electronegativity of atom B, and a damping factor between atom A and B, respectively. The damping factor f_{AB} controls the amount of fractional charge transfer through the covalent bond by attenuating the magnitude of the charge transfer during the iteration process. The value of the damping factor was originally fixed to 1/2 in the PEOE model;[30] however, the m-PEOE method adopted different damping factors for different types of bonds.

In the nth iteration, the net atomic charge of the atom A is given by $Q_A^{<n>}$, as follows:

$$Q_A^{<n>} = Q_A^{<0>} + \sum_n \sum_B dq_{AB}^{<n>} \tag{7}$$

where $Q_A^{<0>}$ is the initial net atomic charge on atom A.

From the point charge of all atoms in a molecule, the electrostatic potentials at each grid point can be calculated as follows:

$$V_{i,PD}^k(r_i) = \sum_j^{N_A^k} \frac{eQ_{j,k}}{|r_i - R_{j,k}|} \tag{8}$$

In this equation, $N_A^k$is the number of atoms in the kth molecule, $Q_{j,k}$ is the net atomic charge of the jth atom in the kth molecule, e is the electronic charge of the jth atom in the kth molecule, and $R_{j,k}$ is the position of the jth atom in the kth molecule.

Based on the m-PEOE charge, where $\vec{r}_i$ is the vector from the center of mass of a molecule to the ith atom, the molecular dipole moment may be calculated as:

$$\vec{\mu} = \sum_i \vec{r}_i Q_i \tag{9}$$

In summary, the overall m-PEOE procedure is as follows:

1. Calculate the connectivity information and determine the initial charges for the given molecule data.
2. Load the m-PEOE parameter presets.
3. Compute the electronegativity ($\chi_{AB}^{<n>}$) for all the bond pairs of the input molecule.
4. Compute the charge transfer ($dq_{AB}^{<n>}$) for all the bond pairs of the input molecule.
5. Calculate the net atomic charges for each atom A ($Q_A^{<n+1>}$).
6. Check the charge transfer quantity.
7. Iterate steps 3−6 until the charge transfer quantity reaches a threshold (threshold = 0.0001).

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

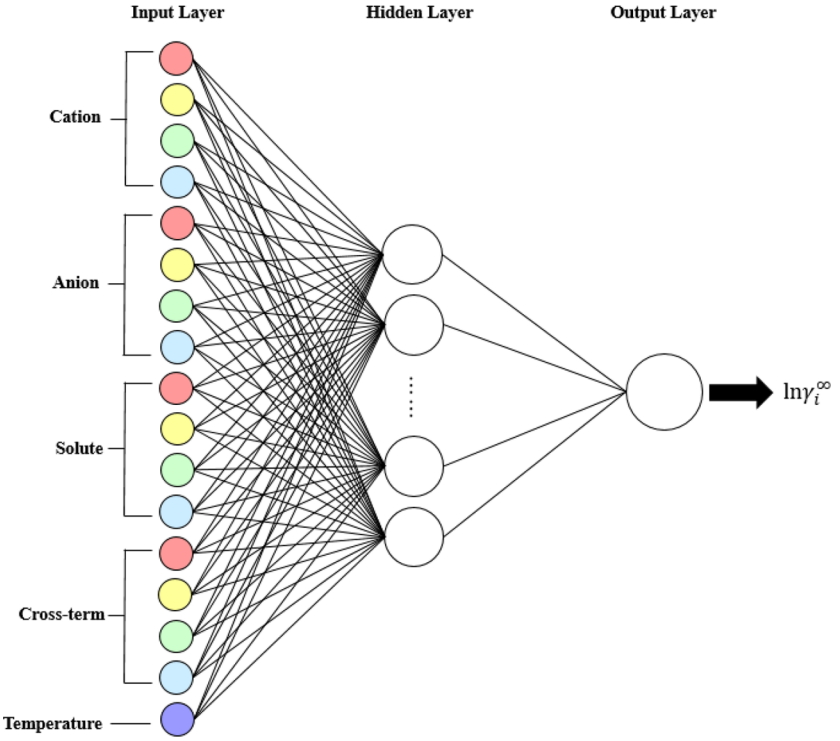


Figure 7. Schematic of the feed-forward neural network architecture. The input vector of FFANN contains the basic molecular properties (red), electrostatic interaction terms (yellow), polarization terms (green), and hydrogen bond terms (blue) of the cation, anion, solute, and cross-term. The temperature terms are shown in purple.

Table 10. Key Information of the FFANN Model

name	value/type
number of hidden layers	1
number of hidden nodes	7−36
number of input node	37
number of parameters	1180
number of validations	10
sampling type	shuffled sampling

The m-PEOE procedure for the atomic monopole calculation was implemented in Java, and the parameters for the original m-PEOE are listed in Table S11.

Since the atom types in the existing m-PEOE method had limited coverage in ionic molecules, new m-PEOE parameters were defined and their values were optimized to develop an m-PEOE method that can be applied to IL (m-PEOE-IL).

4.2.1.1. Ab Initio Calculation of Properties of Cations and Anions. Sixty-eight cation and 134 anion structures encoded in the Molecular Design Limited SDF file format were collected from a PubChem database or manually drawn in Discovery Studio 2016 Client[31] when the SDF file was not available in PubChem. Due to the lack of experimental dipole moments of ILs, ab initio dipole moments were used as constraints in the m-PEOE parameter optimization process. Since ionic bonding is a dominant force in the intermolecular interaction of ILs and the ambient environment of solutes lain in IL solvents, we selected the Quantum Mechanical (QM) calculation method and basis set according to the work of Izgorodina et al.[32] The M06-2X functional with a 6-311++G(3df,2p) basis set was selected. The QM calculation was performed using the Gaussian 03 program.[33] The list of molecules used in ab initio calculation is presented in Table S12.

4.2.1.2. Parameter Supplement and Optimization. Fifty new atom types for cations and anions were newly defined for the development of m-PEOE-IL. After the atom type was extended, the existing m-PEOE parameters and m-PEOE-IL parameters were optimized to predict the ab initio electrostatic potential and dipole moment of the ionic molecules. The following cost function was minimized in an optimization process:

$$F = \sum_{i}^{N} w_D\{\mu_{i,ab} - \mu_{i,calc}\}^2 + \sum_{i}^{N}\sum_{j}^{N_p} w_E V_{i,ab}(r_j) - V_{i,calc}(r_j)^2 \quad (10)$$

where w_D and w_E are the weighting factors of the dipole moment term and the electrostatic potential term, respectively. Here, $\mu_{i,ab}$, $\mu_{i,calc}$, $V_{i,ab}(r_j)$, and $V_{i,calc}(r_j)$ are the ab initio dipole moment of the ith molecule, the m-PEOE-IL-derived dipole moment of the ith molecule, the ab initio electrostatic potential at point r_j, and the m-PEOE-IL-derived electrostatic potential at point r_j, respectively. The golden-section search algorithm was applied in a minimization step to find the optimal parameter value by dividing the given search range by the golden ratio. The m-PEOE-IL parameters and their values are listed in Table 6 for the cations and in Table 7 for the anions.

4.2.2. Description of Polarization Interactions in IL Solution. The Charge Dependence of Effective Atomic Polarizabilities (CDEAP) method uses the dipole polarizability proposed by Shevelko and Vinogradov,[34] expressed using the following equation:

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

$$\alpha \cong \frac{1}{3}\int_0^{r_0} \frac{4\sqrt{2}}{\pi}\left(\frac{Z}{r} - r_0\right)^{1/2} r^4 dr = \frac{N^3}{Z^4}V \quad (11)$$

Here, Z, N, r_0, and V are the nuclear charge, the number of electrons, the radius of the atom or ion, and a constant that has the volume dimensions. If it is assumed that the motions of the electrons in a molecule are strongly restricted by the nuclei in a molecule, the atoms in the molecule are assumed to be perturbed by their environments. This perturbation can be expressed by a change in the electron population (dq), as in the following equation:

$$\alpha_{ij}^* = \alpha_{ij,0}^* - a_{ij}(dq_{ij}) + b_{ij}(dq_{ij})^2 \quad (12)$$

If dq_{ij} is small, the CDEAP for atoms heavier than hydrogen can be expressed as a first-order function of dq_{ij} by neglecting the $(dq_{ij})^2$ term. Thus, eq 12 can be approximated by the following:

$$\alpha_{ij}^* = \alpha_{ij,0}^* - a_{ij}(dq_{ij}) \quad (13)$$

The α_{ij}^* and a_{ij} values for CDEAP are listed in Table 8.

The original CDEAP parameters can explain the polarizabilities of the ILs, except for boron-containing anions because of the lack of boron parameters. Thereby, we derived parameter values (α_{ij}^*, and a_{ij}) for boron by interpolating from a trend in the periodic table. As boron is placed between carbon and nitrogen in the periodic table, the parameter values for boron were interpolated from those of carbon and nitrogen. The new parameter values for boron are marked using footnote *a* in Table 8.

4.2.3. Description of the Hydrogen Bond Interactions in the IL Solution. Simple hydrogen bond descriptors (the number of hydrogen bond acceptors and the number of hydrogen bond donors) were obtained using the TopoMol package in pre-ADMET v.3.0.[35] The cross-term hydrogen bond descriptors, called HB_Mix, were obtained using the proposal of Ajamani et al.[36,37] This equation can be expressed as follows:

$$\mathrm{HB_{Mix}} = 2|A_1 - A_2| + 2|D_1 - D_2| - A_1D_2 - D_1A_2 \quad (14)$$

where A_m and D_m are the number of hydrogen acceptors for the molecule m and the number of hydrogen bond donors for molecule m, respectively. The monopole charge or electrostatic potential-based hydrogen bond descriptors were obtained by combining the monopole charges and electrostatic potentials derived from the m-PEOE-IL.

4.2.4. Description of the Molecular Structure of the Components of the IL Solution. Basic molecular properties (the molecular weight, fraction of rotatable bonds, fraction of aromatic atoms, van der Waals surface area, molar volume, and surface area-to-volume ratio) of the IL-cation, IL-anion, and solute were obtained using the TopoMol package in pre-ADMET v.3.0.[35]

4.3. Descriptor Selection. To select the descriptors for the model, a pool of physically meaningful molecular descriptors was generated from five types of calculated descriptors (electrostatic, hydrogen bonding, polarization, molecular structure, and temperature-dependent). These were divided into 17 subgroups, 16 of which were composed of polarization descriptors; electrostatic descriptors; hydrogen bond descriptors; and molecular structure descriptors of the cation, anion, solute, and a cross-term. The final subgroup included descriptors only containing a temperature term and thus was labeled as the "temperature group". Hence, the number of available combinations of descriptors was large, so descriptor selection was performed using following three criteria:

a. To avoid the omission of key interactions, the selected descriptor set must contain at least one descriptor for each of the 17 subgroups.
b. Since a single temperature-dependent descriptor does not reflect the temperature dependency precisely, the selected descriptor set must contain at least two temperature-dependent descriptors.
c. Since electrostatic interaction is a key interaction in high-dielectric environments, the number of electrostatic descriptors should be greater than the number of other types of descriptors.

Non-identical subsets of descriptors were explored for the following construction of the model. Based on the model performance, a final descriptor set was selected. Each descriptor used in the model development is described in Table 9. Extended explanation and calculation methods for the ultimately selected descriptor candidates are listed in Table S13. Then, a correlation matrix of used descriptors was generated. The correlation matrix is represented in Table S14.

4.4. Development of a FFANN-Based γ^∞ Prediction Model. Since the variables and physical properties of the cation, anion, and solute influence the γ^∞ values of ILs and the variables are strongly dependent on each other, it is impossible to predict γ^∞ using a linear regression model. As the FFANN algorithm can represent the complexity of the molecular interaction between an IL solvent and an organic solute due to their ability to cover all possible interaction combinations through nonlinearity and complexity, FFANN was introduced to describe the dependencies among the physical properties of anions, cations, and solutes in an IL solution. The selected descriptors were used to construct an input layer vector. The scheme for the final FFANN architecture is portrayed in Figure 7. The FFANN model was trained using RapidMiner Studio 5.3.008.[38] Ten-fold cross-validation with shuffled sampling was used for training.

The performance of the QSPR model was evaluated using the RMSE, AE, and squared correlation coefficient (R^2) for the training set, cross-validation set, and external validation set. The hyperparameters for FFANN and the training process are presented in Table 10. The topologies of the model (the hidden layer number and the hidden node number) were optimized by comparing the AE values of various combinations of the hidden layer number and the hidden node number. Further validations were carried out to one of the developed models.

4.5. Importance of Descriptors with Examples. To obtain the importance of descriptors, we performed a process based on descriptor value scrambling.[39] We denoted the RMSE of the developed model as the "control RMSE". Then, we executed the following processes:

1. Generate a new data by randomly scrambling the first input descriptor.
2. Apply the generated data into the trained model.
3. Calculate RMSE.
4. Do procedures 1−3 for each of the input descriptors.
5. Iterate procedures 1−4 20 times.
6. Average the RMSE values of each descriptor.

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

The descriptor scrambling process was implemented in JAVA. The importance of each descriptor was measured by calculating the difference between the averaged RMSE value and control RMSE. An interpretation was done by comparing values of descriptors and experimental γ^{∞} on the selected example solute-IL combinations.

4.6. Model Validation. *4.6.1. Y-Randomization Test.* For the purposes of confirming whether the correlation was established by chance or not, a Y-randomization test was performed 60 times by shuffling experimental γ^{∞} values of the original data set. Each shuffled experimental γ^{∞} value set was generated by using 60 different random seeds. Then, 60 models were built from each of the shuffled data sets. The model topology and model-building descriptors were the same as for the final model. The RMSE values for the training sets and cross-validation sets were the criteria for the coincidence of the models. Another metric, ${}^{c}R_p^2$, is also calculated as follows:

$$ {}^{c}R_p^2 = R \times \sqrt{R^2 - R_r^2} \tag{15} $$

where R is the correlation coefficient of the developed model and R_r^2 is the squared mean correlation coefficient of the randomized model.[40]

4.6.2. Applicability Domain. In addition to the predictive ability validation, the structure space of the ILs and organic solutes was also verified, wherein the model made predictions with the most optimal reliability. Defining the borders of the space, the so-called ″optimum prediction space″ or ″applicability domain″ is especially important for compounds without experimental data to verify the quality of the predictions. To visualize the applicability domain of a QSPR model, a Williams plot was adopted, which visualizes outliers with leverage values and standardized residuals for each molecule.[41] This plot can provide the immediate and simple graphical detection of outliers. The applicability domain is defined within ±3 standard residuals and a leverage threshold. The leverage value h_i for each compound i may be calculated from the descriptor matrix as follows:

$$ h_i = x_i^T (X^T X)^{-1} x_i \tag{16} $$

where x_i is the molecular descriptor vector for each compound i. The warning leverage h^* value may be calculated as follows:

$$ h^* = \frac{3(p+1)}{n} \tag{17} $$

where n is the total compound number and p is the number of descriptors used in the model-building process.[42]

4.6.3. Robustness Test. To verify the robustness of the model, the experimental data set was included in neither the training set nor the test set of any other research.[43−45] The collected data set contained 1269 data points that were composed of 4 cations, 3 anions, and 77 solutes. The cation and anion combinations in the data set had no duplicate combinations in the original data set. The descriptors for the molecules, cations, anions, and solutes in the experimental data were calculated and applied to the neurons in the input layer.

4.6.4. Performance Comparison with Previously Reported Models. To test our model in control with previously reported models, we compared the performance of our model with the three models (SWMLR, FFANN, and LSSVM) developed by Paduszyński.[23] The prediction was performed on a training data set, a test data set, and a robustness test data set. RMSE values of each model were used for performance comparison.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c01690.

Weight matrix of FFANN model; predicted infinite dilution activity coefficients versus experimental infinite dilution activity coefficients in model development; performance changes by scrambled descriptors; performances of randomized models; predicted infinite dilution activity coefficients versus experimental infinite dilution activity coefficients of the robustness test data set; performance comparison between our model and reference models; basic chemical information of the cations included in the training and test data set; basic chemical information of the anions included in the training and test data set; basic chemical information of the solutes included in the training and test data set; data points of each IL combination; existing m-PEOE parameters and their values; basic chemical information of the cations and anions used in m-PEOE optimization; descriptors used in model building process; correlation matrix of descriptors; and references for experimental data (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Kyoung Tai No − *Department of Biotechnology, Yonsei University, Seoul 03722, Republic of Korea;* Present Address: Institute of Convergence Science and Technology, Yonsei University, Incheon 21983, Republic of Korea; Present Address: Bioinformatics & Molecular Design Research Center, Incheon 21983, Republic of Korea (K.T.N).; orcid.org/0000-0003-3187-8193; Email: ktno@yonsei.ac.kr

Authors

Hyeon-Nae Jeon − *Department of Biotechnology, Yonsei University, Seoul 03722, Republic of Korea;* orcid.org/0000-0002-6332-1990

Hyun Kil Shin − *Department of Biotechnology, Yonsei University, Seoul 03722, Republic of Korea;* Present Address: Department of Predictive Toxicology, Korea Institute of Toxicology, Gajeong-ro 141, Daejeon 34114, Republic of Korea (H.K.S. and S.H.).; orcid.org/0000-0003-3665-0841

Sungbo Hwang − *Department of Biotechnology, Yonsei University, Seoul 03722, Republic of Korea;* Present Address: Department of Predictive Toxicology, Korea Institute of Toxicology, Gajeong-ro 141, Daejeon 34114, Republic of Korea (H.K.S. and S.H.).; orcid.org/0000-0002-1610-5259

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c01690

Author Contributions

H.N.J. and K.T.N. designed the study. H.N.J performed the quantum mechanical calculations, model training, and web-server construction. K.T.N. helped supervise the project. H.K.S. helped carry out the machine learning model

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

construction. S.H. helped carry out the web-server construction. H.N.J took the lead in writing the manuscripts. All authors provided critical feedback and analysis.

Funding

This research was financially supported by the Ministry of Trade, Industry, and Energy (MOTIE), Korea, under the "Infrastructure Support Program for Industry Innovation"(reference number P0014714) supervised by the Korea Institute for Advancement of Technology (KIAT).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the BK21 PLUS program (Brain Korea 21 for Leading Universities & Students) [grant number 2018-11-0004].

■ ABBREVIATIONS

IL, ionic liquid; QSPR, Quantitative Structure−Property Relationship; COSMO-RS, Conductor-like Screening Model for Real Solvents; UNIFAC, UNIversial quasi-chemical Functional-group Activity Coefficients; LSER, Linear Solvation free Energy Relationship; MLR, multiple linear regression; ANN, artificial neural network; SWMLR, StepWise Multiple Linear Regression; LSSVM, Least Square Support Vector Machine; FFANN, Feed-Forward Artificial Neural Network; AARD, average absolute relative deviation; m-PEOE, modified Partial Equalization of Orbital Electronegativity; m-PEOE-IL, m-PEOE method for IL; QM, quantum mechanical; CDEAP, Charge Dependence of Effective Atomic Polarizability; AE, absolute error; EMIM, 1-ethyl-3-methyl-1H-imidazol-3-ium; NTF2, bis(trifluoromethylsulfonyl)azanide; DCA, dicyanoazanide; CCN3, tricyanomethide

■ REFERENCES

(1) Grodowska, K.; Parczewski, A. Organic solvents in the pharmaceutical industry. *Acta Pol. Pharm.* **2010**, *67*, 3−12.

(2) Wlazło, M.; Gawkowska, J.; Domańska, U. Separation Based on Limiting Activity Coefficients of Various Solutes in 1-Allyl-3-methylimidazolium Dicyanamide Ionic Liquid. *Ind. Eng. Chem. Res.* **2016**, *55*, 5054−5062.

(3) Domańska, U.; Redhi, G. G.; Marciniak, A. Activity coefficients at infinite dilution measurements for organic solutes and water in the ionic liquid 1-butyl-1-methylpyrrolidinium trifluoromethanesulfonate using GLC. *Fluid Phase Equilib.* **2009**, *278*, 97−102.

(4) Letcher, T. M.; Marciniak, A.; Marciniak, M.; Domańska, U. Activity coefficients at infinite dilution measurements for organic solutes in the ionic liquid 1-hexyl-3-methyl-imidazolium bis-(trifluoromethylsulfonyl)-imide using g.l.c. at T=(298.15, 313.15, and 333.15) K. *J. Chem. Thermodyn.* **2005**, *37*, 1327−1331.

(5) Kato, R.; Gmehling, J. Activity coefficients at infinite dilution of various solutes in the ionic liquids [MMIM]+[CH_3SO_4]−, [MMIM]+[$CH_3OC_2H_4SO_4$]−, [MMIM]+[$(CH_3)_2PO_4$]−, [$C_5H_5NC_2H_5$]+[$(CF_3SO_2)_2N$]− and [C_5H_5NH]+[$C_2H_5OC_2H_4OSO_3$]−. *Fluid Phase Equilib.* **2004**, *226*, 37−44.

(6) Domańska, U.; Laskowska, M. Measurements of activity coefficients at infinite dilution of aliphatic and aromatic hydrocarbons, alcohols, thiophene, tetrahydrofuran, MTBE, and water in ionic liquid [BMIM][SCN] using GLC. *J. Chem. Thermodyn.* **2009**, *41*, 645−650.

(7) Domańska, U.; Marciniak, A. Measurements of activity coefficients at infinite dilution of aromatic and aliphatic hydrocarbons, alcohols, and water in the new ionic liquid [EMIM][SCN] using GLC. *J. Chem. Thermodyn.* **2008**, *40*, 860−866.

(8) Heintz, A.; Kulikov, D. V.; Verevkin, S. P. Thermodynamic Properties of Mixtures Containing Ionic Liquids. 1. Activity Coefficients at Infinite Dilution of Alkanes, Alkenes, and Alkylbenzenes in 4-Methyl-*n*-butylpyridinium Tetrafluoroborate Using Gas−Liquid Chromatography. *J. Chem. Eng. Data* **2001**, *46*, 1526−1529.

(9) Dobryakov, Y. G.; Tuma, D.; Maurer, G. Activity Coefficients at Infinite Dilution of Alkanols in the Ionic Liquids 1-Butyl-3-methylimidazolium Hexafluorophosphate, 1-Butyl-3-methylimidazolium Methyl Sulfate, and 1-Hexyl-3-methylimidazolium Bis-(trifluoromethylsulfonyl) Amide Using the Dilutor Technique. *J. Chem. Eng. Data* **2008**, *53*, 2154−2162.

(10) Krummen, M.; Wasserscheid, P.; Gmehling, J. Measurement of Activity Coefficients at Infinite Dilution in Ionic Liquids Using the Dilutor Technique. *J. Chem. Eng. Data* **2002**, *47*, 1411−1417.

(11) Kato, R.; Krummen, M.; Gmehling, J. Measurement and correlation of vapor−liquid equilibria and excess enthalpies of binary systems containing ionic liquids and hydrocarbons. *Fluid Phase Equilib.* **2004**, *224*, 47−54.

(12) Vasiltsova, T. V.; Verevkin, S. P.; Bich, E.; Heintz, A.; Bogel-Lukasik, R.; Domanska, U. Thermodynamic Properties of Mixtures Containing Ionic Liquids. Activity Coefficients of Ethers and Alcohols in 1-Methyl-3-Ethyl-Imidazolium Bis(Trifluoromethyl-sulfonyl) Imide Using the Transpiration Method. *J. Chem. Eng. Data* **2005**, *50*, 142−148.

(13) Revelli, A.-L.; Sprunger, L. M.; Gibbs, J.; Acree, W. E., Jr.; Baker, G. A.; Mutelet, F. Activity Coefficients at Infinite Dilution of Organic Compounds in Trihexyl(tetradecyl)phosphonium Bis-(trifluoromethylsulfonyl)imide Using Inverse Gas Chromatography. *J. Chem. Eng. Data* **2009**, *54*, 977−985.

(14) Diedenhofen, M.; Eckert, F.; Klamt, A. Prediction of Infinite Dilution Activity Coefficients of Organic Compounds in Ionic Liquids Using COSMO-RS†. *J. Chem. Eng. Data* **2003**, *48*, 475−479.

(15) Klamt, A. Conductor-Like Screening Model for Real Solvents: a New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(16) Wang, J.; Sun, W.; Li, C.; Wang, Z. Correlation of infinite dilution activity coefficient of solute in ionic liquid using UNIFAC model. *Fluid Phase Equilib.* **2008**, *264*, 235−241.

(17) Fredenslund, A.; Jones, R. L.; Prausnitz, J. M. Group-contribution estimation of activity coefficients in nonideal liquid mixtures. *AIChE J.* **1975**, *21*, 1086−1099.

(18) Mutelet, F.; Ortega-Villa, V.; Moïse, J.-C.; Jaubert, J.-N.; Acree, W. E., Jr. Prediction of Partition Coefficients of Organic Compounds in Ionic Liquids Using a Temperature-Dependent Linear Solvation Energy Relationship with Parameters Calculated through a Group Contribution Method. *J. Chem. Eng. Data* **2011**, *56*, 3598−3606.

(19) Abraham, M. H.; Whiting, G. S.; Doherty, R. M.; Shuely, W. J. Hydrogen Bonding. Part 13. A New Method for the Characterisation of Glc Stationary Phases—the Laffort Data Set. *J. Chem. Soc., Perkin Trans. 2* **1990**, 1451−1460.

(20) Paduszyński, K. An overview of the performance of the COSMO-RS approach in predicting the activity coefficients of molecular solutes in ionic liquids and derived properties at infinite dilution. *Phys. Chem. Chem. Phys.* **2017**, *19*, 11835−11850.

(21) Eike, D. M.; Brennecke, J. F.; Maginn, E. J. Predicting Infinite-Dilution Activity Coefficients of Organic Solutes in Ionic Liquids. *Ind. Eng. Chem. Res.* **2004**, *43*, 1039−1048.

(22) Xi, L.; Sun, H.; Li, J.; Liu, H.; Yao, X.; Gramatica, P. Prediction of infinite-dilution activity coefficients of organic solutes in ionic liquids using temperature-dependent quantitative structure−property relationship method. *Chem. Eng. J.* **2010**, *163*, 195−201.

(23) Paduszyński, K. In Silico Calculation of Infinite Dilution Activity Coefficients of Molecular Solutes in Ionic Liquids: Critical Review of Current Methods and New Models Based on Three Machine Learning Algorithms. *J. Chem. Inf. Model.* **2016**, *56*, 1420−1437.

(24) Park, J. M.; No, K. T.; Jhon, M. S.; Scheraga, H. A. Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method. III. Application to

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373

Halogenated and Aromatic Molecules. *J. Comput. Chem.* **1993**, *14*, 1482−1490.
(25) Park, J. M.; Kwon, O. Y.; No, K. T.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. IV. Application to hypervalent sulfur- and phosphorus-containing molecules. *J. Comput. Chem.* **1995**, *16*, 1011−1026.
(26) Suk, J. E.; No, K. T. Determination of Net Atomic Charges Using a Modified Partial Equalization of Orbital Electronegativity Method V. Application to Silicon-Containing Organic Molecules and Zeolites. *Bull. Korean Chem. Soc.* **1995**, *16*, 915−923.
(27) No, K. T.; Grant, J. A.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4732−4739.
(28) No, K. T.; Grant, J. A.; Jhon, M. S.; Scheraga, H. A. Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 2. Application to ionic and aromatic molecules as models for polypeptides. *J. Phys. Chem.* **1990**, *94*, 4740−4746.
(29) No, K. T.; Cho, K. H.; Jhon, M. S.; Scheraga, H. A. An empirical method to calculate average molecular polarizabilities from the dependence of effective atomic polarizabilities on net atomic charge. *J. Am. Chem. Soc.* **1993**, *115*, 2005−2014.
(30) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(31) *Discovery Studio Visualizer v16.1.0.15340*; Dassault Systèmes BIOVIA: 2015.
(32) Izgorodina, E. I.; Bernard, U. L.; MacFarlane, D. R. Ion-Pair Binding Energies of Ionic Liquids: Can DFT Compete with Ab Initio-Based Methods? *J. Phys. Chem. A* **2009**, *113*, 7064−7072.
(33) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C. et al. *Gaussian 03; Revision C.02*; Gaussian, Inc.: Wallingford CT, 2004.
(34) Shevelko, V. P.; Vinogradov, A. V. Static dipole polarizability of atoms and ions in the Thomas-Fermi Model. *Phys. Scr.* **1979**, *19*, 275−282.
(35) Lee, S.; Lee, I. H.; Kim, H. J.; Chang, G. S.; Chung, J. E.; No, K. T. The PreADME Approach: Web-based program for rapid prediction of physico-chemical, drug absorption and drug-like properties. *EuroQSAR designing drugs and crop protectants: processes, problems and solutions* 2002; p 418−420.
(36) Ajmani, S.; Rogers, S. C.; Barley, M. H.; Burgess, A. N.; Livingstone, D. J. Characterization of Mixtures Part 1: Prediction of Infinite-Dilution Activity Coefficients Using Neural Network-Based QSPR Models. *QSAR Comb. Sci.* **2008**, *27*, 1346−1361.
(37) Muratov, E. N.; Varlamova, E. V.; Artemenko, A. G.; Polishchuk, P. G.; Kuz'min, V. E. Existing and Developing Approaches for QSAR Analysis of Mixtures. *Mol. Inf.* **2012**, *31*, 202−221.
(38) Mierswa, I.; Wurst, M.; Klinkenberg, R.; Scholz, M.; Euler, T. In YALE: rapid prototyping for complex data mining tasks, the *12th ACM SIGKDD international conference, 2006*; ACM Press: 2006; p 935, DOI: 10.1145/1150402.1150531.
(39) Guha, R.; Jurs, P. C. Interpreting computational neural network QSAR models: a measure of descriptor importance. *J. Chem. Inf. Model.* **2005**, *45*, 800−806.
(40) Gajewicz, A.; Puzyn, T. *Computational Nanotoxicology: Challenges and Perspectives*; CRC Press, 2019.
(41) Jaworska, J.; Nikolova-Jeliazkova, N.; Aldenberg, T. QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *Altern. Lab. Anim.* **2005**, *33*, 445−459.
(42) Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694−701.
(43) Marciniak, A.; Wlazło, M. Activity coefficients at infinite dilution and physicochemical properties for organic solutes and water in the ionic liquid trihexyl-tetradecyl-phosphonium tricyanomethanide. *J. Chem. Thermodyn.* **2018**, *120*, 72−78.
(44) Mutelet, F.; Baker, G. A.; Ravula, S.; Qian, E.; Wang, L.; Acree, W. E., Jr. Infinite dilution activity coefficients and gas-to-liquid partition coefficients of organic solutes dissolved in 1-*sec*-butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide and in 1-*tert*-butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide. *Phys. Chem. Liq.* **2019**, *57*, 453−472.
(45) Wlazło, M.; Zawadzki, M.; Domanska, U. Separation of water/butan-1-ol based on activity coefficients at infinite dilution in 1,3-didecyl-2-methylimidazolium dicyanamide ionic liquid. *J. Chem. Thermodyn.* **2018**, *116*, 316−322.

https://doi.org/10.1021/acsomega.1c01690
ACS Omega 2021, 6, 15361−15373