Thermophysical Property Prediction of Anion-Functionalized Ionic Liquids for CO_2 Capture

Austin N. Keller, Pratik Kelkar, Michael Baldea, Mark A. Stadtherr and Joan F. Brennecke*

John. J. McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712

*jfb@che.utexas.edu

Abstract

We develop a machine learning framework for predicting the density, viscosity, and heat capacity of a family of anion-functionalized ionic liquids for CO_2 capture, specifically those with tetraalkylphosphonium cations and aprotic *N*-heterocyclic anions (AHAs). We screen several feature sets using group contribution-based (GC) descriptors and descriptors extracted from COSMO-RS sigma profiles (SP) to build Support Vector Regression (SVR) and Gradient-Boosted Regression (GBR) machine learning models. Viscosities and densities were modeled based on data sets containing nearly 60 ILs each. The best fit for viscosity used GC-based descriptors and the SVR model, achieving a test set %AARD of 12.5% and R^2 of 0.989. Density was modeled using these same descriptors with the SVR model framework and was fitted with a test set %AARD of 1.0%. Heat capacity was fit as a function of molar volume and temperature, a general trend observed for all ILs in a family. Heat capacity predictions could then be made using the density SVR model with a test set accuracy of 3.0 %AARD. With these results, we have developed predictive models which can potentially be used in the design of new advanced ionic liquids for carbon capture.

I. Introduction

As the societies of the world develop mitigation strategies for rising atmospheric CO_2 concentrations while concurrently addressing rising energy demand, efficient carbon capture technologies will likely be necessary to reduce emissions from existing fossil fuel-based power generation. Aqueous amines are the currently preferred technology for post-combustion carbon capture [1], but they suffer from poor regeneration thermodynamics related to their strong enthalpy of reaction with CO_2, as well as water evaporation [2].

Ionic liquids with aprotic *N*-heterocyclic anions (AHA ILs) have been proposed [3,4] as an alternative solvent for CO_2 capture due to their favorable properties, such as negligible volatility, high thermal stability [5], and high tunability, especially with regard to their CO_2 binding energy [6]. The anion of the AHA IL is comprised of a cyclic secondary amine which reacts reversibly in an equimolar ratio with CO_2 at ambient pressure or below [6–8]. This is an advantage over physically absorbing ionic liquids which require high pressures to achieve moderate loading of CO_2 in the IL [9,10]. Additionally, there is no significant viscosity increase upon complexation

with CO_2 in the case of AHA ILs (as opposed to many other amine functionalized ionic liquids which experience large viscosity increases due to formation of hydrogen-bonding networks [9]). In addition to viscosity, density is also very weakly dependent on CO_2 loading, as was demonstrated, e.g., by the measurements of Makino et al. [11], who reported that at CO_2 pressures up to 0.7 bar the volume expansion of the AHA IL tributyl(octyl)phosphonium benzotriazolide is less than 1%.

Sensitivity and optimization studies on the properties of AHA ILs for post-combustion carbon capture were conducted by Seo et al. [12] and Hong et al. [4]. In each of these works, optimal values of AHA IL properties are proposed. However, in the absence of a model to map chemical structures to chemical properties, this approach has limitations because it is difficult to predict the tradeoff between different material properties. Moreover, a set of optimal physical properties may not correspond to any actual IL and therefore serve merely as a guideline for IL selection or *de novo* design. In this work, we aim to map the relevant thermophysical properties—particularly density, viscosity, and heat capacity—of AHA ILs to ab initio and structural predictors using machine learning models. Given the documented insensitivity of AHA IL density and viscosity to CO_2 loading, the predicted properties may be considered applicable to both CO_2 rich and lean process conditions and therefore be relevant to future process modeling studies.

Machine learning models and quantitative structure property relationship (QSPR) methods have been used in a variety of works for the prediction of IL properties. In particular, researchers have found that density as a function of temperature can be modeled well, often reporting errors of less than 2% AARD, using a variety of descriptors (e.g., group contribution [13–17]) and models (e.g., SVR [15,18], artificial neural networks (ANNs) [15,19], decision tree based methods [19–21], and multiple linear regression (MLR) [15,17,19,21–26]). Some models include other parameters such as pressure [13–15,22,24] and water content [15,21] as input variables. On the other hand, predicting viscosity is more challenging, often requiring authors to reduce the IL space to a fixed class of cations or anions [18,27–33] or to implement GC type descriptors to differentiate between IL families [17,34–39] to achieve accuracies better than 20% AARD. Models for IL heat capacity have found sufficient accuracy with MLR and other parametric regression approaches [17,40–45]. Decision-tree based methods [20] and ANNs [42] have also been investigated to this end. Nonetheless, these efforts have either focused on limited classes of ILs or exclusively used the NIST ILThermo database,[46,47] both of which do not include sufficient data to model ILs that chemically react with CO_2.

We investigate the efficacy of support vector machine regression (SVR) and gradient-boosted regression (GBR) for modeling and predicting thermophysical properties of AHA ILs. SVR has been used in a number of studies involving IL property prediction [42,48–50], including our prior work on dialkylimidazolium ILs [18]. GBR and similar tree-based methods have also been used in several studies on ILs. For example, Dhakal and Shah use extreme gradient boosting (XGBoost) to predict IL ionic conductivity [51] and Venkatraman et al. [52] found that tree-based methods offered slight advantages over SVR and *k*-nearest neighbor approaches in predicting ionic liquid melting points.

Descriptors and molecular representations which contain relevant information are key to accurate chemical and thermophysical property prediction. In a review on QSPR methods for material property prediction, Le et al. [53] assign chemical descriptors to five categories: constitutional, topological, physicochemical, structural, and quantum-chemical. Additionally, Philippi et al. [54] suggest that inputs to machine learning models must consider the factors which distinguish ionic liquid viscosities from those of molecular liquids: charge networks and coulombic compaction. One quantum-chemical descriptor set investigated in this work is based on the COSMO-RS sigma profile. COSMO-RS is a thermodynamic model [55,56] which uses surface screening charge distributions calculated with quantum chemical methods for individual species to model the thermodynamic behavior of mixtures. The surface screening charge representation used by COSMO-RS, called the sigma profile, is a histogram of surface screening charge density [57]. In this sense, it is a two-dimensional descriptor which contains by design information that describes the thermodynamic interactions of species in solution. However, this information can also be used to describe thermophysical properties; for example, we have used descriptors based on COSMO-RS sigma profiles to predict the density, viscosity, and ionic conductivity of imidazolium based ionic liquids [18]. Sigma profiles and other surface charge- and screening charge-based methods have also been used to predict other IL properties, such as infinite dilution activity coefficients [48], heat capacity [42], viscosity [31,58] and toxicity [59]. From the perspective of process optimization and molecular design, sigma-profile-based descriptors have the advantage that they are continuous variables that can be used in an optimization problem. However, it is non-trivial to map a set of sigma-profile values back to a physically realizable IL structure.

Group contribution-based (GC) descriptors are constitutional descriptors and have also been used to predict several of the IL properties of interest to this study. For example, Gharagheizi et al. [60] used GC descriptors with a multiple linear regression model to estimate the viscosity of a large general set of ILs. This type of descriptor has also been used to predict melting point [61,62] and CO_2 solubility [63]. GC descriptors are attractive because they correspond intuitively to structures; however, for the purpose of design optimization they may be less desirable because they are discrete (rather than continuous) variables.

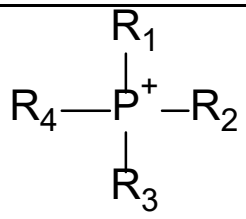
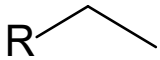
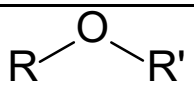
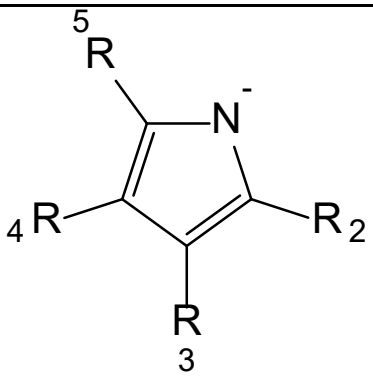
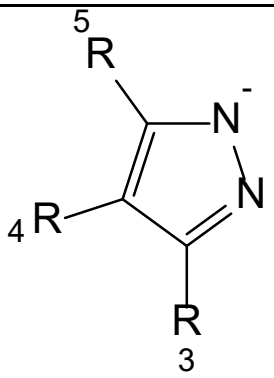
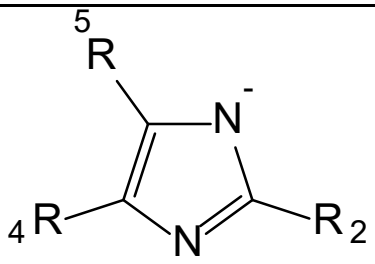
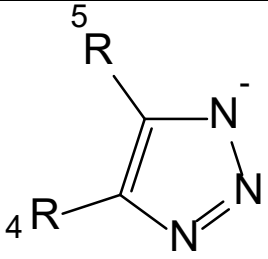
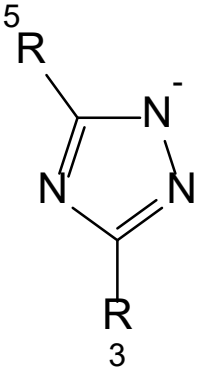
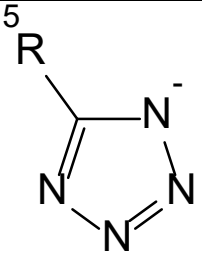
II. Methods

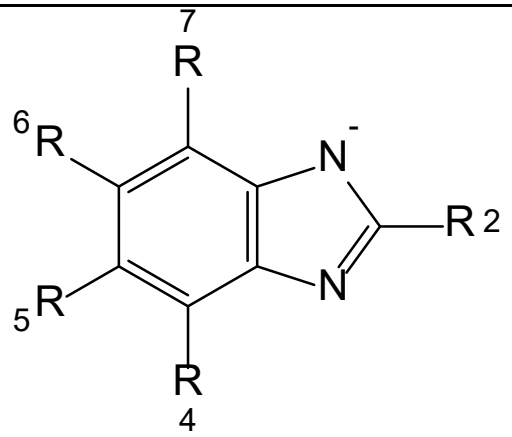
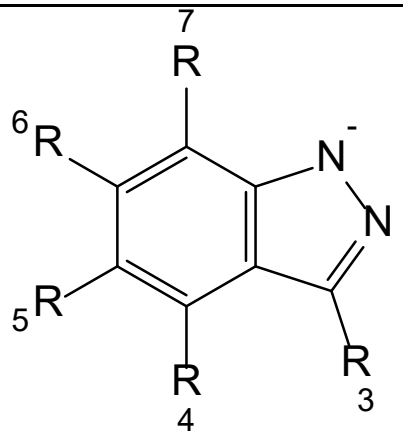
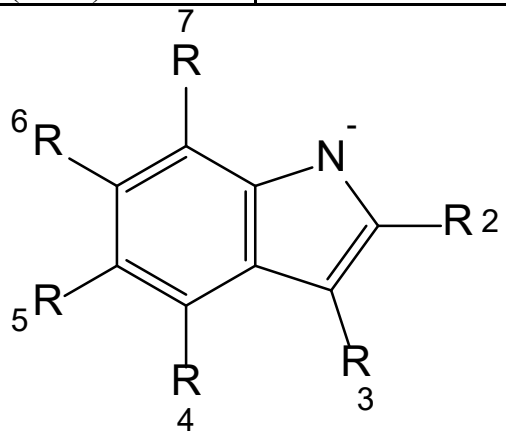
In this section we discuss how we assembled and processed the thermophysical property data, how we generated the group contribution and COSMO-RS features, the regression and evaluation process, and how we dealt with the heat capacity, since it depends on the density.

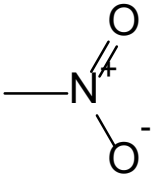
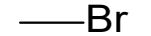
II. A. Data Acquisition and Processing

Data for viscosity, density, and heat capacity of ionic liquids with tetraalkylphosphonium cations and AHA anions were obtained from the literature [5–9,11,64–74]. All tetraalkylphosphonium cations contain *n*-alkane and/or ether groups of varying lengths. All cation and anion core groups and functional groups that were included in the data set are summarized in **Table 1**.

Table 1: Cation and anion core and functional groups involved in this study.

Cation Core		
		
Cation Groups		
		
Alkane	Ether	
Anion Cores		
		
Pyrrolide (Pyr)	Pyrazolide (Pyra)	Imidazolide (Im)
		
1, 2, 3-triazolide (3-Triz)	1, 2, 4-triazolide (4-Triz)	Tetrazolide (Tetz)

	
Benzimidazolide (BnIm)	Indazolide (Inda)
	
Indolide (Indo)	
Anion Groups	

		
Methyl (CH_3)	Trifluoromethyl (CF_3)	Nitro (NO_2)
		
Chloro (Cl)	Bromo (Br)	Cyano (CN)

The viscosity data set consists of ILs containing 15 unique cations, 26 unique anions and a total of 59 unique combinations. Low temperature data (below 290 K) were excluded because measuring such data is typically less reliable; there is increased risk of water condensation at these temperatures and experimental uncertainties may be higher at extremely high viscosities. Therefore, only data between 290 K and 373 K were included. The distribution of the 527 viscosity observations in this set is shown in **Figure 1** and the data are given in **Table S1.A** of the **Supporting Data** spreadsheet. The density data set has 58 unique ILs consisting of 13 unique cations and 25 unique anions. The temperature ranges from 293 K to 363 K. The distribution of

the 549 density observations also appears in **Figure 1** and the data are provided in **Table S2** of the **Supporting Data** spreadsheet. The available heat capacity data were constrained to ILs with melting points below 283 K. There were significantly fewer C_p data available for AHA ILs than for viscosities and densities. In fact, there are just 15 unique ILs made from 12 anions and 3 cations in the heat capacity data set. Nonetheless, heat capacity is an important thermophysical property for CO_2 capture solvents since the absorption and regeneration processes generally involve a temperature swing. The 37 heat capacity observations are shown on the right most plot on **Figure 1** and the data are given in **Table S3** of the **Supporting Data** spreadsheet.

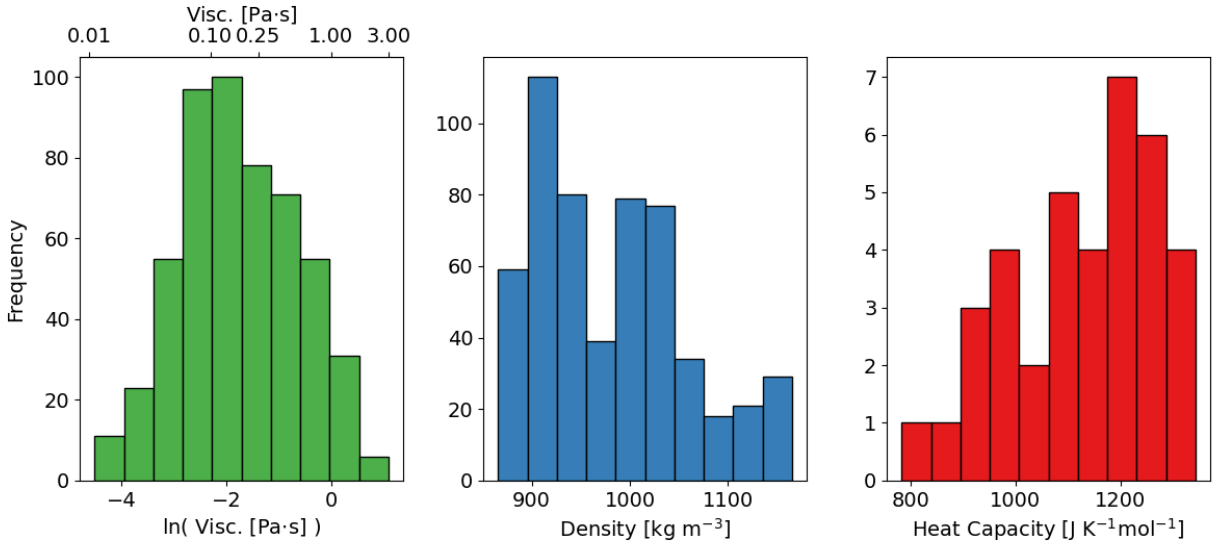


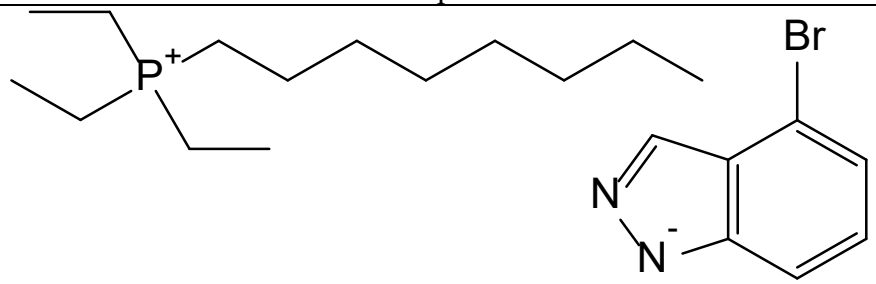
Figure 1. Histograms of literature experimental data for the properties of interest [5–8,64–70]. The natural logarithm of viscosity (left) is shown in addition to the viscosity values on the secondary abscissa. Density (middle) and heat capacity (right) are plotted on linear scales.

We hypothesized that only AHA-based ionic liquids would be relevant to prediction of AHA IL properties due to the highly polar and chemically basic nature of the AHA anions. Nonetheless, we also compiled additional phosphonium IL viscosity data from the ILThermo database [46,47] as well as from the literature [75–78] to fit to a larger dataset and validate this hypothesis. In total, the non-AHA viscosity dataset consists of an additional 55 ILs, 937 viscosity observations, and 45 unique anions. These data are given in **Table S1.B** of the **Supporting Data** spreadsheet.

II. B. Group Contribution Feature Generation

The functional groups employed in developing the corresponding group contribution (GC) descriptors are shown in **Table 2**. There are six anion core groups corresponding to all the monocyclic AHA anions and a bicyclic descriptor to allow for bicyclic anions. For instance, an indazolide anion has one pyrazolide core and one bicyclic functional group. All cations have a phosphonium core and terminal -CH_3 groups so the features on the cation are the number of interior -CH_2- and -O- groups. Group contribution representations for all ions in this study are available in the **Supporting Data** spreadsheet, **Table S4**.

Table 2. GC descriptors for cations and anions used in this work with triethyl(octyl)phosphonium 4-bromoindazolide as an example.

Example IL	
	
Cation Groups	
Group	No. in Example IL
Alkyl [-CH_2-]	10
Ether [-O-]	0
Anion Cores	
Pyrrolide [Pyr]	0
Pyrazolide [Pyra]	1
Imidazolide [Im]	0
1, 2, 3-triazolide [3-Triz]	0
1, 2, 4-triazolide [4-Triz]	0
Tetrazolide [Tetz]	0
Anion Groups	
Bicyclic	1
Methyl [-CH_3]	0
Trifluoromethyl [-CF_3]	0
Cyano [-CN]	0
Nitro [-NO_2]	0
Chloro [-Cl]	0
Bromo [-Br]	1

II. C. COSMO-RS Feature Generation

All COSMO-RS based descriptors used in this work are developed from each ion's sigma profile. Sigma profiles were generated using a method similar to that used by Nordness et al. [18] where the geometry of the ions is optimized individually in the gas phase using Gaussian 16 [79] with the BVP86 method, triple-ζ valence potential (TZVP) basis set, and the DGA1 density fitting set. After geometry optimization, a frequency calculation is performed to confirm an optimal geometry is attained (i.e. not a transition state). A conductor-like polarizable continuum model calculation was performed with the same method, basis set, and fitting set, using Gaussian's COSMO-RS calculation option and a *.cosmo file is obtained. A sample Gaussian input *.com file for 1, 2, 4,-triazolide is given in **Section 1.1** of the **Supplementary Materials** document. The individual ion sigma profiles were then calculated from each *.cosmo file using COSMOthermX19 [80–82] and the BP_TZVP_19 parameterization. Before further calculations were performed, the

cation and anion sigma profiles for each species of interest were additively combined to obtain a sigma profile for the IL; this is referred to as the meta-file approach in the literature [83]. The cation and anion sigma profiles are available in **Tables S5.A** and **S5.B** of the **Supporting Data** spreadsheet.

The sigma profile generated by the COSMOtherm software was binned in equal size bins separated by 0.001 e/Å^2, where 'e' represents the elementary charge, along the screening charge axis. Following the work of Palomar et al. [84] we further discretized the sigma profile into larger bins to reduce the dimensionality of the descriptors. Since Nordness et al. [18] and others chose the number of equal sized bins somewhat arbitrarily, we screened multiple bin sizes. We focused on the nonzero sigma region of the sigma profile for ILs in this study between [-0.02 e/Å^2, 0.03 e/Å^2] and discretized this range into 9, 15, and 19 equal size bins, with the first and last bin being for overflow (referred to as SP9, SP15, and SP19, respectively). The boundaries of the bins are summarized in **Figure 2**, using triethyl(octyl)phosphonium 4-triazolide as an example. In addition, we performed regression with all 51 bins from the COSMOtherm output, having 0.001 e/Å^2 width each (SP51). These feature sets are available for all AHA ionic liquids in the viscosity set in **Tables S6-S9** of the **Supporting Data** spreadsheet. **Table S6.B** contains the SP9 feature set for the larger dataset including AHA and non-AHA phosphonium ILs (SP9+NA).

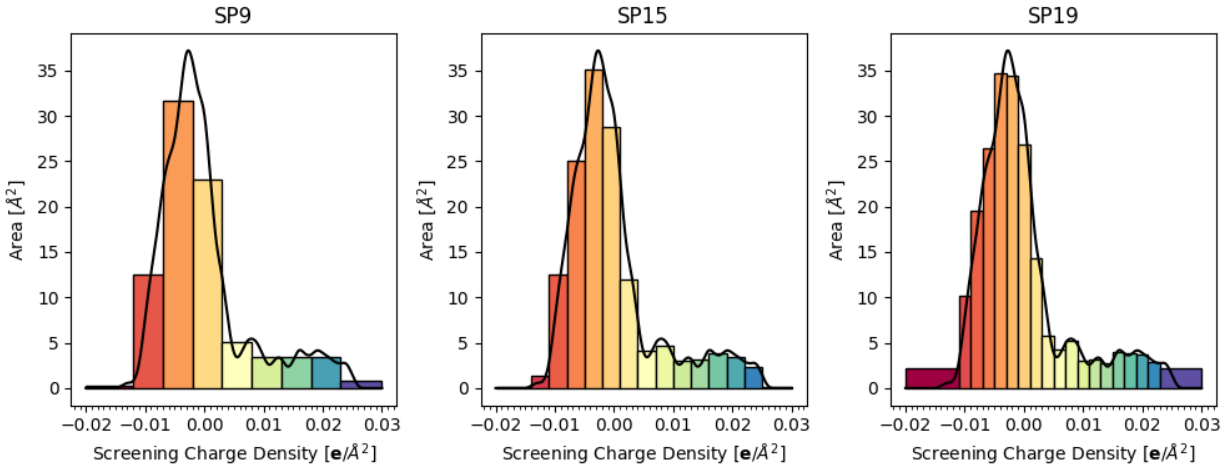


Figure 2: Sigma-profile bin sets for triethyl(octyl)phosphonium 4-triazolide. **e** is the elementary charge (1.6×10^{-19} coulomb) and Å is angstrom (1×10^{-10} meter); these are the standard COSMO-RS units.

When performing analysis with the SP9, SP15, SP19, and SP51 feature sets, it became apparent that many of the bins exhibited significant multicollinearity, i.e. the bin heights were correlated with each other; this was especially true for adjacent bins. To remediate this problem, additional feature sets were prepared in which one bin of any pair of bins with a correlation, $R > 0.9$, was dropped based on which bin had the lower standardized residual in a multilinear regression model with all other bins included. These new feature sets are referred to as CovX-SP9, CovX-SP15, and CovX-SP19. It was also hypothesized that the reason some bins were collinear was because they were canonically part of the same feature of the sigma profile. To further analyze this, we prepared a feature set from all SP51 where bins with correlation, $R > 0.9$, were iteratively

combined. The resulting feature set, SP-Cov, and the ultimate bin boundaries are discussed in section III. B. The covariance adjusted feature sets (CovX-SP9, CovX-SP15, and CovX-SP19) and the SP-Cov set are available in **Tables S10-13** of the **Supplementary Data** spreadsheet for modeling of the AHA IL viscosity data set.

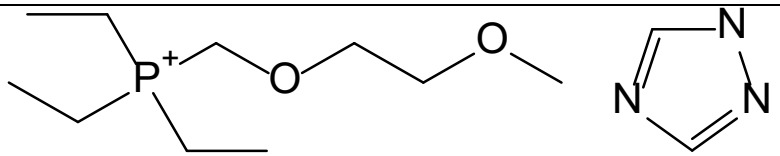
In addition to the sigma-profile-based feature sets, we also calculated COSMO-RS interaction energies as a feature set, which was the method used by Koi et al. [50]. However, these features performed poorly when applied to both a linear regression model and an SVR model, so they are not discussed further. We note that a key difference between the ionic liquids in the study by Koi et al. and our work are the anion moieties. In particular, the AHA anions have much greater charge polarization.

II. D. Regression and Evaluation

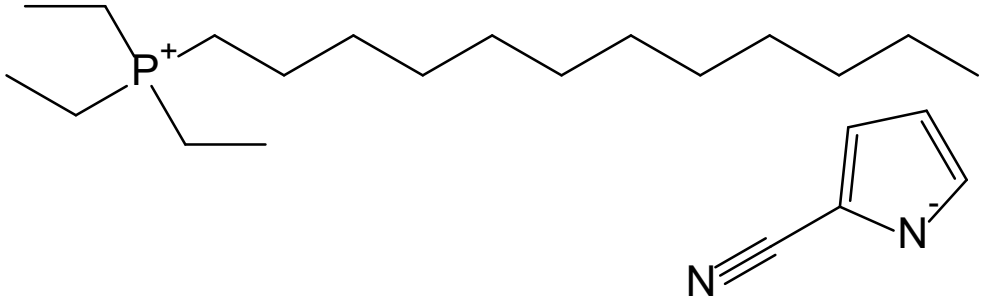
The density and viscosity data sets were randomly split into 5-folds of roughly equal size to test the predictive capability of each model that is developed through cross-validation. This is a procedure in which a portion of data, or a fold, is held out from the training set and the performance of the model at predicting that particular fold is tested. The folds are created such that no observations of one IL appears in another fold, similar to the work of Nordness et al.[18] and Baskin et al. [19]. This is done because, as observed by Faundez et al. [85], the *predictive* ability of machine learning models cannot be accurately quantified by the model's performance on species which appear in the training set. Rather they must predict properties for chemical species which are not included in the development of the model. However, since ILs may be considered a simple mixture of cation and anion species, models which perform more strict segregation by not only IL species but also by ion species, as demonstrated by the works of Makarov et al. [86,87], have even better predictive power than those without. With such a limited dataset containing a large variety of ion species, we do not adhere to the stricter approach of Makarov et al. and therefore acknowledge this limitation that our model will perform best for the 390 combinations of 15 cations and 26 anions in the viscosity set and the 325 combinations of 13 cations and 25 anions in the density set. Therefore, predictions for ILs outside of the support of the unique cations and anions present in the model are extrapolations and must be viewed critically until their properties are experimentally verified.

The ILs contained in the 5-folds for both the density and viscosity sets are available in the **Supporting Data** spreadsheet, **Tables S14-15**. Additionally, a set of representative ILs, shown in **Table 2**, is held out from training for use as a test set and the final model developed is based on best performance for this set. These ILs were selected because they include 4 of the 6 anion base groups, 5 of the 7 anion functional groups, have various cation sizes, and include oxygenated cations. Comparatively, the training set has the full support of anion base groups and functional groups and a similar range of cation sizes. In the case of the larger dataset containing AHAs and non-AHAs, the validation sets (5-fold and test) were made to only contain AHAs because the goal of this work was to predict the properties of tetraalkylphosphonium AHAs. For simplicity, the AHA/non-AHA combined model uses the same 5-folds and test set as the AHA only models.

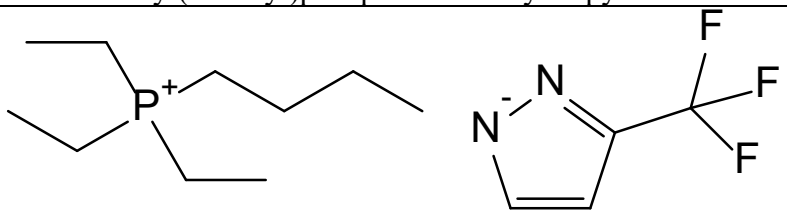
Table 2. Test set ILs for density and viscosity.



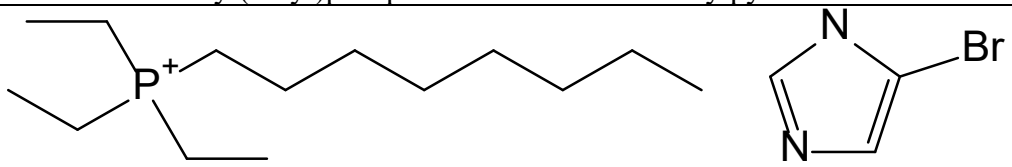
triethyl((2-methoxyethoxy)methyl)phosphonium 4-triazolide



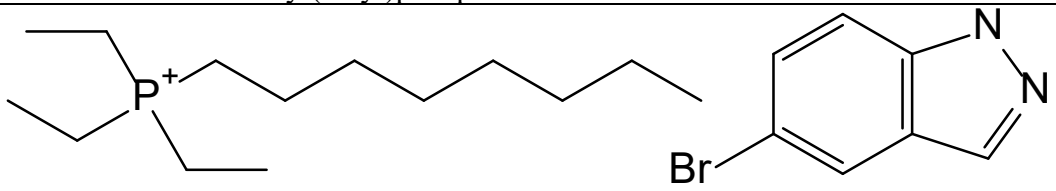
triethyl(dodecyl)phosphonium 2-cyanopyrrolide



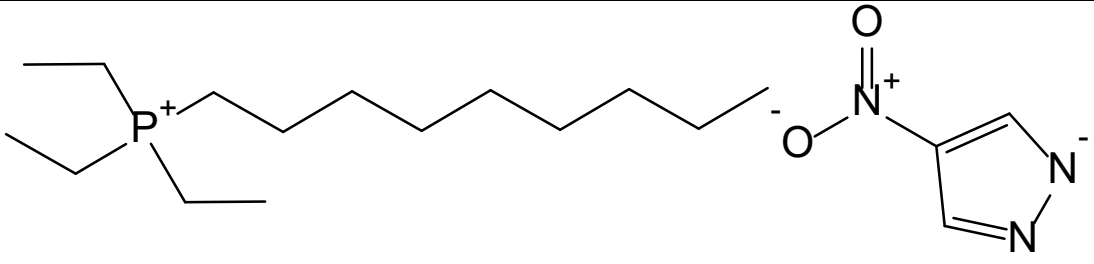
triethyl(butyl)phosphonium 3-trifluoromethylpyrazolide



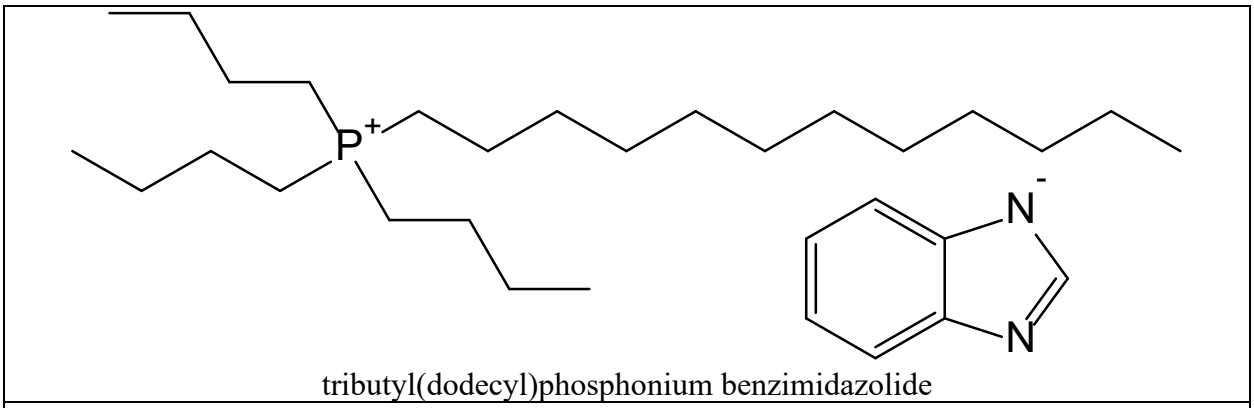
triethyl(octyl)phosphonium 4-bromoimidazolide



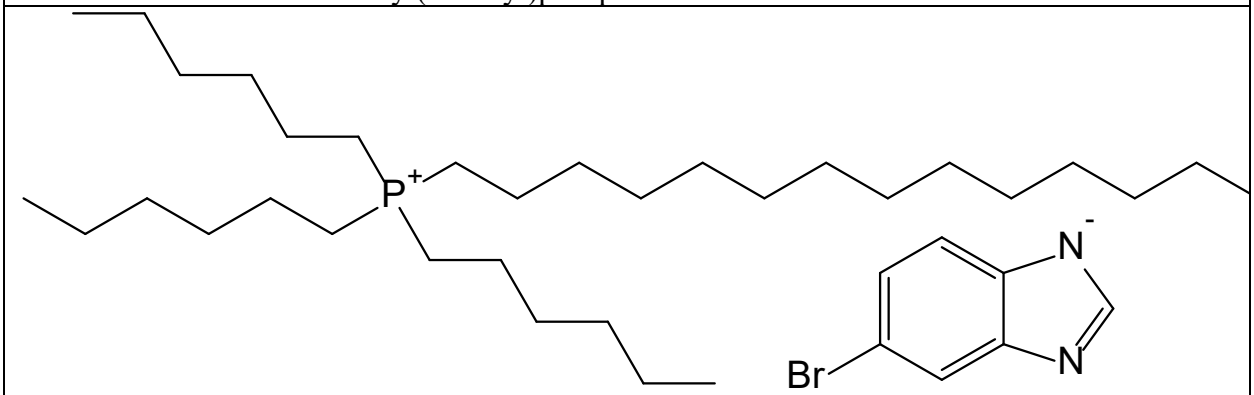
triethyl(octyl)phosphonium 5-bromoindazolide



triethyl(nonyl)phosphonium 4-nitropyrazolide



tributyl(dodecyl)phosphonium benzimidazolide



trihexyl(tetradecyl)phosphonium 6-bromo-benzimidazolide

Both machine learning methods, SVR and GBR, employ functions from the Scikit-learn package [88] in Python. We also considered artificial neural network (ANN) and Gaussian process regression (GPR) models but they did not perform well in initial work so they were not pursued further. The SVR method employed uses the radial basis function (RBF) kernel as it has lower bias and is better able to represent nonlinearity than other common kernels. Often transformations of the dependent and/or independent variables in a regression problem can be helpful for fitting. For the density data set, mean centering and standard scaling of neither the dependent nor independent variable transformations appeared to produce significant differences in accuracy of the resulting fit so these values are used as is. In the case of viscosity, the best results were observed when the independent variables are not transformed but a natural logarithm transform is used for viscosity; i.e., the fitted property is,

$$\eta' = \ln\left(\eta\ [\mathrm{Pa \cdot s}]\right) \tag{1}$$

where η' is the natural logarithm of viscosity which is predicted and η is the dynamic viscosity in Pascal seconds. The SVR and GBR models perform differently based on the values of hyperparameters which are assigned to them. A grid search is used to find near optimal hyperparameters for each descriptor set, model, and assessment method (test or 5-fold) used. Transforming viscosity by dividing by density and molecular weight and then taking the logarithm

is done in the Orrick-Erbar type approach adopted by Gardas and Coutinho [89] for ionic liquid viscosity. That is, fitting the property,

$$\eta^{\star} = \ln\left(\frac{\eta}{\rho * MW}\right) \tag{2}$$

where $\eta^{\star}$ is the predicted viscosity variable, η is the viscosity in Pascal seconds, ρ is the density in kilograms per cubic meter, and MW is the molecular weight in grams per mole. We implemented this equation in our GC/GBR model but saw negligible improvement and therefore opted to fit η' from **Equation 1** only in this study.

Each fit is evaluated using percentage average absolute relative deviation (%AARD) and coefficient of determination (R^2). The %AARD is calculated as,

$$\%AARD = \frac{1}{N}\sum_{i}^{N}\left|\frac{\hat{y}_i - y_i}{y_i}\right| * 100\% \tag{3}$$

For N observations of the dependent variable y where $\hat{y}_i$ is the model's estimate of the i^{th} observation of y. The R^2 is calculated as,

$$R^2 = \frac{SSR}{SSTO}\ , \qquad SSR = \sum_{i}^{N}(\hat{y}_i - \bar{y})^2\ , \qquad SSTO = \sum_{i}^{N}(y_i - \bar{y})^2$$

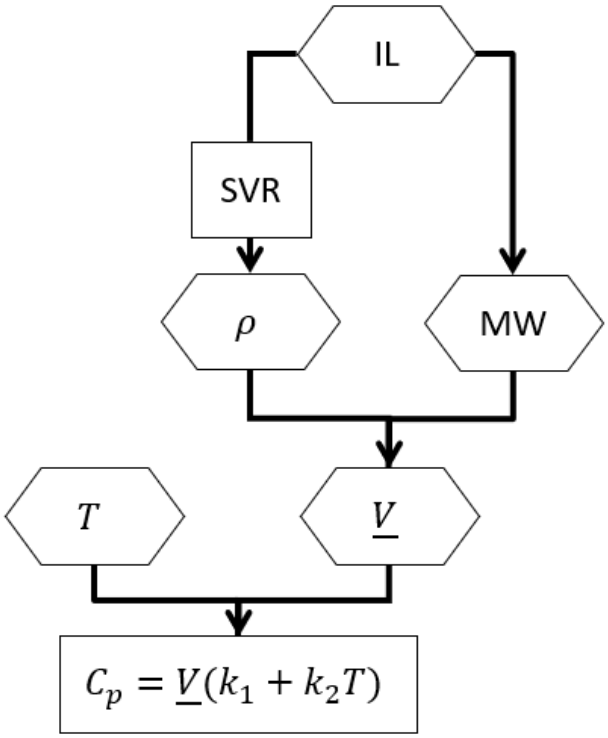
where SSR is the regression sum of squares or the square variation of the estimates around the mean of the dependent variable observations, $SSTO$ is the total sum of squares or the square variation of the data around the mean, $\hat{y}_i$ is the model's estimate of the i^{th} observation of y, and $\bar{y}$ is the arithmetic mean of y observations in the data set. A sample script which performs the model fitting and analysis for the density and viscosity sets is available in **Sections 2.1** and **2.2** of the **Supplementary Materials** document.

II. E. Heat Capacity Fitting

In order to fit the constant pressure heat capacity of these ILs, we first consider an observation made by Strechan et al. [90] and Gardas and Coutinho [40], that heat capacity on a volumetric basis at room temperature, i.e., the heat capacity in *energy per temperature unit per volume*, is roughly constant for most ILs. This quantity has been referred to as the volumic heat capacity [91]. We use the later observation by Paulechka et al. [91] that this relationship, when expanded to be a linear function of temperature, gives a good approximation of volumic heat capacity in a moderate temperature range:

$$\frac{C_p}{\rho_L^{-1} \cdot MW} \approx k_1 + k_2 \cdot T \quad (4)$$

where C_p is the molar heat capacity in J/mol K^{-1}, ρ_L is liquid density in g/cm^3, MW is molecular weight in g/mol, T is temperature in Kelvin, and k_1 and k_2 are fitted parameters that are appropriate for an entire set of ILs. Once k_1 and k_2 have been evaluated for a few ILs in the family, the heat capacity (C_p) of all the other AHA ILs can be predicted as a function of temperature using the density from the ML model and the molecular weight, which is known from the structure. This routine is shown graphically in **Scheme 1**.



Scheme 1: Flow chart representation of heat capacity fitting. ρ is IL density, MW is molecular weight, $\underline{V}$ is IL molar volume, T is temperature in Kelvin, and k_1 and k_2 are constants from **Eq. 4**.

Since there was a limited amount of liquid heat capacity data, prediction is assessed based on only 4 ILs from the 17 IL set. The 4 ILs held in the test set are tetraoctylphosphonium benzimidazolide, trihexyl(tetradecyl)phosphonium pyrazolide, tributyl(dodecyl)phosphonium benzimidazolide, and tributyl(dodecyl)phosphonium 3-cyanopyrrolide. These 4 ILs represent 3 of the 4 unique cations, and 3 of the 13 unique anions in the set, and contain both bicyclic and monocyclic anions. A sample script which fits and evaluates this model is available in **Section 2.3** of the **Supplementary Materials** document.

III. Results and Discussion

Below we discuss the results for modeling of viscosity, density and heat capacity of AHA ILs. Since viscosity is the most challenging property to model, we discuss it first, comparing the various group contribution (GC) and sigma-profile (SP) descriptors that we explored. There is a separate section on our attempts to eliminate correlation between adjacent sigma-profile bins in the modeling of viscosity.

III. A. Viscosity

The decision of which feature set to use for the final model was made based upon the features that are best able to predict viscosity, because IL density was fit to similar accuracy with most of the feature sets that were considered. We evaluated performance of each of the GBR and SVR models with each feature set using both the test set and 5-fold cross validation. The SVR model constructed using the GC descriptor set had both the best %AARD and R^2 for the test set and the best average %AARD for the 5-fold cross validation. The performance at predicting the viscosities of the test set ILs for the SVR and GBR models using the GC, all four sigma-profile (SP51, SP9, SP15 and SP19) based descriptor sets, and the augmented dataset using SP9 and additional non-AHA data (SP9+NA) is shown in **Figure 3** and **Table 3**, and the performance for the aforementioned models at predicting the viscosities of the 5-fold set ILs is also given in **Table 3**.

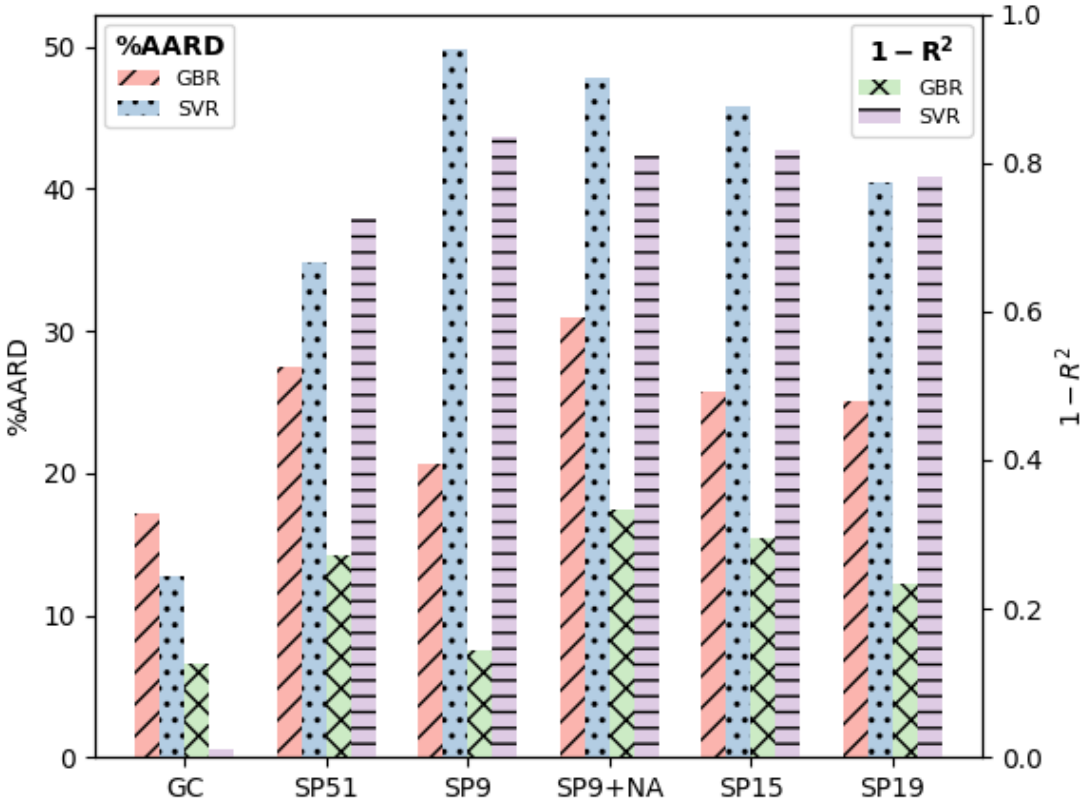


Figure 3. Bar graph comparing test set performance of SVR and GBR viscosity models based on

GC and SP. The performance metrics shown are %AARD (red and blue bars with left axis) and $1 - R^2$ (green and violet bars on right axis).

Table 3. Performance summary for the GC/SP-based SVR/GBR viscosity models.

	Test				5-fold CV Average			
	GBR		SVR		GBR		SVR	
Feature Set	%AARD	R^2	%AARD	R^2	%AARD	R^2	%AARD	R^2
GC	17.2	0.873	12.7	0.988	22.1	0.844	21.6	0.810
SP51	27.5	0.727	34.8	0.275	33.5	0.673	45.4	0.432
SP9	20.6	0.855	49.8	0.164	26.8	0.780	51.7	0.388
SP9+NA	31.0	0.845	35.4	0.700	47.8	0.19	53.7	0.320
SP15	25.7	0.706	45.8	0.184	31.2	0.724	48.9	0.376
SP19	25.1	0.767	40.4	0.220	30.4	0.655	47.6	0.397

Both the GC SVR and GC GBR models had superior performance to all SP-based models. The GBR model worked best for the sigma-profile-based sets. Additionally, within the set of GBR models the number of sigma-profile bins did not appear to have a definite trend, but for the SVR set more sigma-profile bins improved model performance. This is a reasonable result as decision tree-based models are often considered to be robust to overfitting of high dimensional data. The larger SP9+NA dataset containing non-AHA phosphonium ILs performed more poorly when used with a GBR model than the SP9, and showed only slight improvement over the SP9 SVR model, but still with a poor fit. Therefore, we only consider models constructed with the AHA only datasets herein. Ultimately, the GC SVR model was selected due to its superior performance in both testing and cross-validation and a parity plot of these models for the test set and the 5-fold CV set are shown in **Figure 4**.

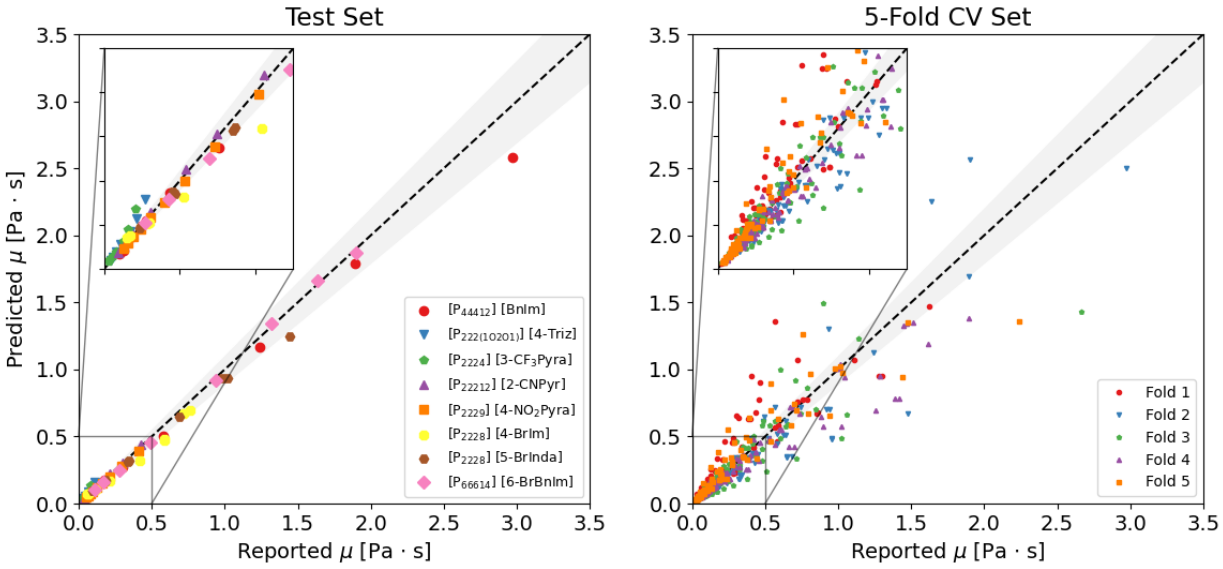


Figure 4. Parity plots for GC SVR viscosity model for test set ILs and 5-fold cross validation sets. Gray shaded region is ±10% of the reported viscosity.

The test set performance is significantly better than the 5-fold cross validation set. We believe this to be a result of how we chose the test set. The test set contains ILs whose cations and anions appear in multiple different pairings in the support. The 5-fold cross validation set was randomly selected, however, and as a result the poorest performing ILs either have few or no similar ILs in the support or are the product of a broken trend such as the increase in viscosity that occurs when exchanging the smaller $[P_{2228}]^+$ cation with the $[P_{66614}]^+$ cation in the case of $[6BrBnIm]^-$.

III. B. Sigma Profile Covariance

In the modeling of viscosity with sigma-profile descriptors, we observed that several adjacent sigma-profile bins were significantly correlated across the data set. In an effort to improve the SP-based models, we attempted to remove the redundant information by removing bins which were highly correlated with another bin by the procedure described in **Section II. C**. Additionally, it was hypothesized that high correlation between adjacent bins may suggest that the bins represent the same canonical feature. To investigate, we developed another SP-based feature set referred to as SP-Cov with bins of varying sizes, the boundaries of which are constructed according to the procedure described in **Section II. C.** The resulting bin boundaries are shown graphically in **Figure 5**, using triethyl(octyl)phosphonium 4-triazolide as an example. The performance of models for viscosity constructed using these covariance remediation methods is compared to those models without covariance remediation in **Figure 6** and **Table 4**.

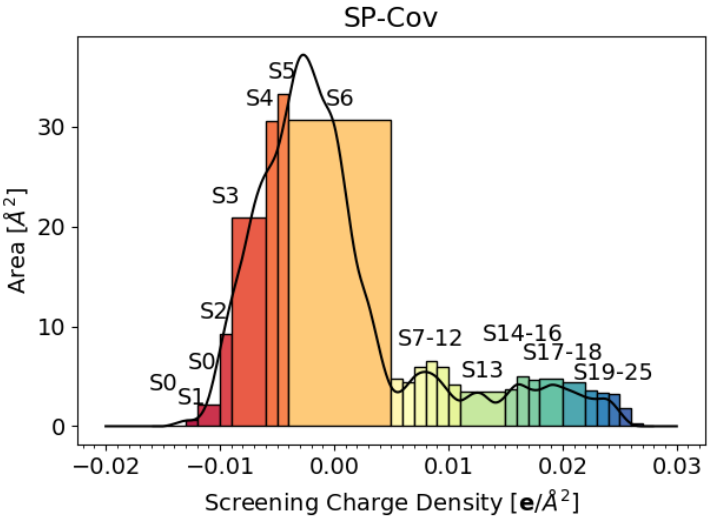


Figure 5. Sigma profile for triethyl(octyl)phosphonium 4-triazolide (black curve) with SP-Cov bins. Bin S1 bisects bin S0, as it contains additional variance uncorrelated to either side of S0.

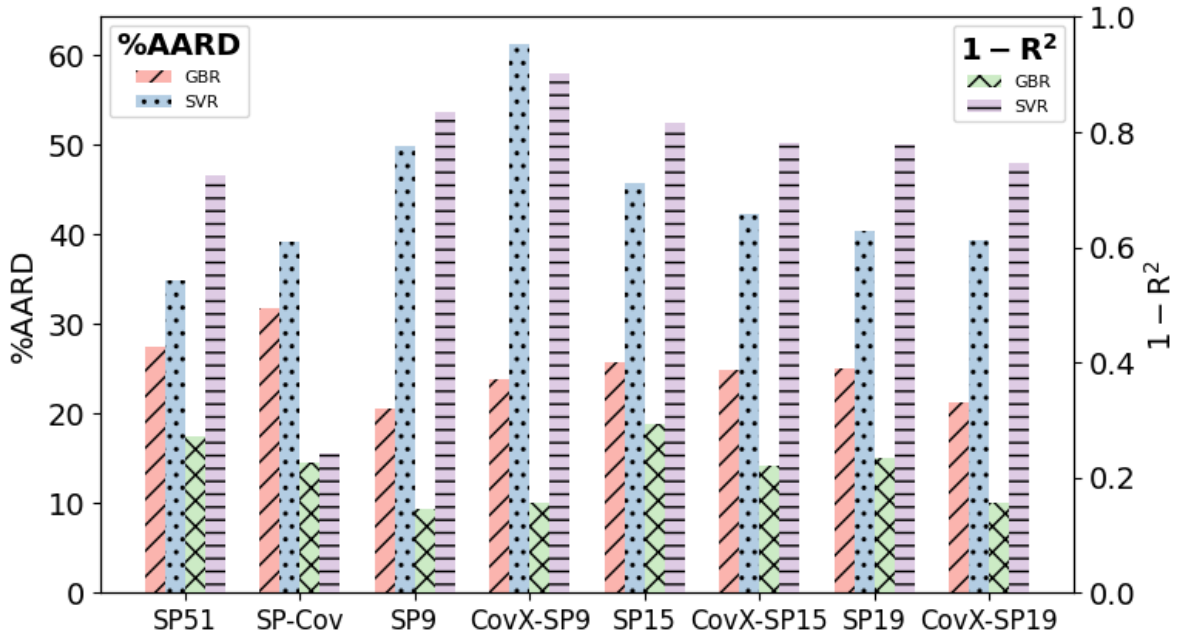


Figure 6. Bar graph comparing performance of models with and without highly covariate features removed or reconstructed. Performance metrics are %AARD (red and blue bars with left axis) and $1 - R^2$ (green and violet bars on right axis) for the test set.

Table 4. Performance summary/comparison for covariance-based feature sets.

	Test				5-fold CV Average			
	GBR		SVR		GBR		SVR	
Feature Set	%AARD	R^2	%AARD	R^2	%AARD	R^2	%AARD	R^2

SP51	27.5	0.727	34.8	0.275	33.5	0.673	45.4	0.432
SP-Cov	27.6	0.710	58.8	0.131	34.0	0.655	54.0	0.308
SP9	31.8	0.773	39.2	0.757	29.7	0.709	44.6	0.504
CovX-SP9	20.6	0.855	49.8	0.164	26.8	0.780	51.7	0.388
SP15	23.9	0.845	61.2	0.097	28.8	0.783	60.6	0.291
CovX-SP15	25.7	0.706	45.8	0.184	31.2	0.724	48.9	0.376
SP19	24.9	0.780	42.3	0.219	29.1	0.755	47.4	0.413
CovX-SP19	25.1	0.767	40.4	0.220	30.4	0.655	47.6	0.397

Reducing the number of bins from the SP51 case to the SP-Cov case does appear to provide a benefit in significantly increasing the R^2 for both the SVR and GBR models but has a slightly negative effect on the %AARD. Since R^2 as a performance metric preferentially penalizes poor estimates away from the mean and %AARD preferentially penalizes poor estimates near the origin this may suggest that the SP-Cov model is improving the performance of the model at high viscosity values at the cost of low viscosity accuracy. A similar trend appears between the SP19 and CovX-SP19 feature sets as well as the SP9 and CovX-SP9 sets. Overall, the GBR models still performed much better than the SVR models for all SP-based feature sets. We conclude that while removing some covariant features using the methodologies described did increase the R^2 in most cases, the ultimate performance of the SP models was still dwarfed by that of the GC-based model for this narrowly defined data set of phosphonium AHA ILs.

Finally, to rationalize the results of the SP-Cov models we compare the SP-Cov features to the GC features by analyzing the correlation between these two sets of features. A heat map of this analysis is shown in **Figure 7**. Notably, in the case of the cation features, S1 is highly correlated with presence of ether groups and S0, S3, and S4 are correlated with the length of the alkyl chains on the phosphonium. For the anion base groups, the pyrrolide (Pyr) group and bicyclic core are described with S17-S19 and S10/S15, respectively. The other core groups are not clearly correlated to any single SP-Cov bin. Additionally, while many of the anion functional groups are correlated with the SP-Cov bins, the methyl group and chloro groups are not. In other words, there doesn't appear to be a distinghishable and characteristic change in the sigma profiles in this dataset for functionalization with methyl and chloride groups and for base groups other than pyrrolide. In this sense, while much of the information about which functional groups are present is directly translated to the SP-Cov feature set, there is a significant amount of information which is represented in the GC set but not the SP-Cov set, providing a possible explanation for why the two perform so differently. Given that GC features cannot be reproduced by the SP-Cov set we hypothesize that either there is some information in GC that is absent in the sigma profile, such as structural insights, or there is insufficient precision in the sigma profile and/or variety in the sigma-profile dataset for the GC features to be deconvoluted.

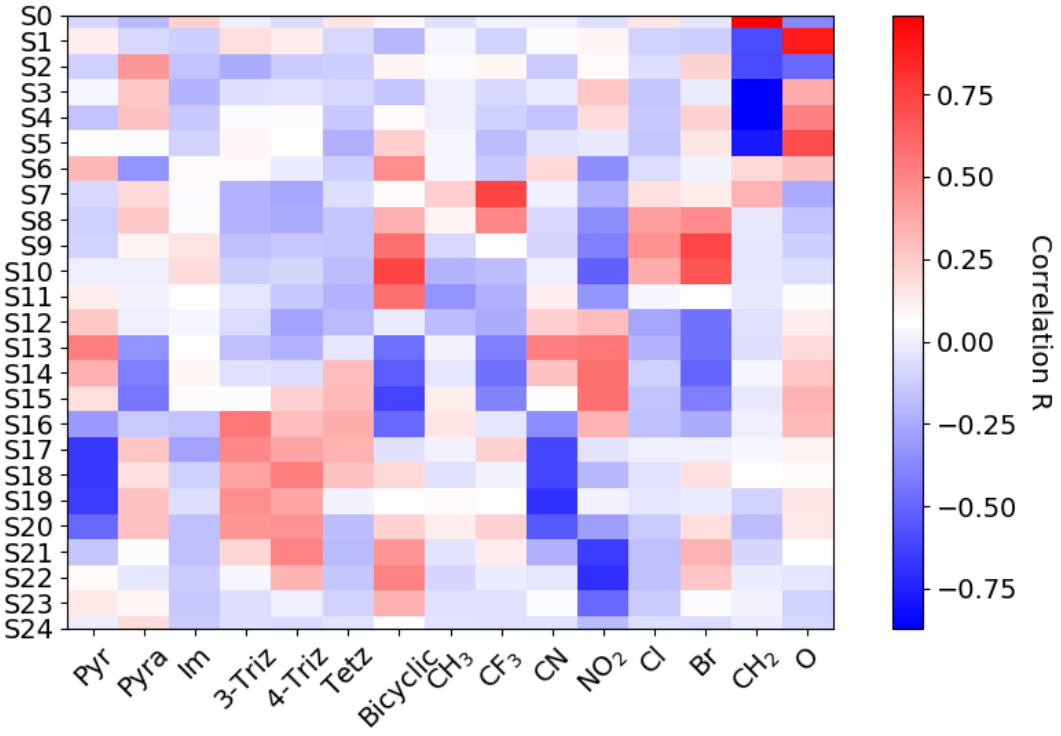


Figure 7. Heat map of correlation, *R*, between SP-Cov bins and GC groups for ILs studied in this work. CH_2 and O are cation groups and all others are anion groups.

III. C. Density

Since the GC feature set was the superior set for predicting viscosity, the same feature set is used for density with the goal of creating a set of model with one set of inputs. Models using the SVR and GBR methods were developed and performed similarly in both the test set case and the 5-fold cross validation case, having %AARD of around 1% each and R^2 of 0.961 and 0.968, respectively. For the purpose of consistency with the viscosity model, the GC/SVR model is ultimately selected and the results of that fit are shown in **Figure 8**.

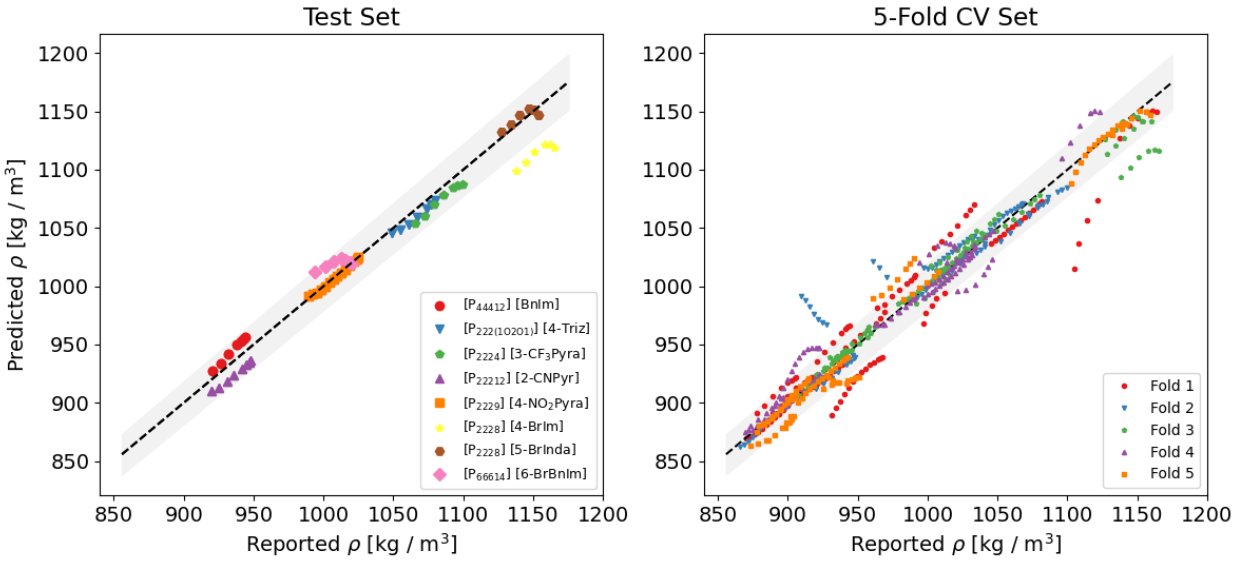


Figure 8. Parity plot of GC/SVR density model for test set and 5-fold cross validation sets. Gray shaded region is ±2% of the measured density. Both the test set and the 5-fold CV set have %AARD of 1.0%.

While the test set data are very well reproduced by the model, there are two ILs in Fold 2 of the 5-fold CV set for which the model predictions versus reported values have a negative slope; i.e., the model predicts that the density increases with increasing temperature. These two species were identified as [P_{4444}][BnIm] which has a melting temperature of 71 °C [64] and [P_{4444}][Im] which is generally a solid at room temperature but may liquify upon mechanical disturbance [69]. Therefore, it is hypothesized that the error in the trend predicted by the model could be due to these ILs, and the other [P_{4444}] based ILs in the data set, being near a phase transition at the measurement temperatures. Additionally, as a result of being solids at room temperature, the measured temperatures in the data available for each of these ILs are among the highest present in the data set, ranging from 353-363 K and 333-363 K for [P_{4444}][BnIm] and [P_{4444}][Im], respectively. It is also possible that the model performance could be hindered at these higher temperatures. Similarly, [P_{4444}][6BrBnIm] which is an IL present in Fold 1 with reported densities between 1100 and 1150 kg/m^3, is also very poorly described by the model and has a melting temperature of 69 °C [64].

III. D. Heat Capacity

To fit and predict the heat capacity data, first we fit the volumic heat capacity data for the 13 training ILs to **Eq. 4** This resulted in a reasonably good fit having a %AARD of 2.9%. The resulting model is given below and plotted in **Figure 9**:

$$\frac{\widehat{C_p}}{\rho_L^{-1} \cdot MW}\left[\frac{\mathrm{J}}{\mathrm{cm}^3 \cdot \mathrm{K}}\right] \approx 1.181 + 2.01 \cdot 10^{-3}(T\ [\mathrm{K}]) \qquad (5)$$

Where $\widehat{C_p}$ is the model predicted heat capacity, with the hat indicating it is a regression model estimate. The temperature dependence term of $2.01 \cdot 10^{-3} \frac{\mathrm{J}}{\mathrm{cm^3 \cdot K^2}}$ is more than twice as large as that of $8.33 \cdot 10^{-4} \frac{\mathrm{J}}{\mathrm{cm^3 \cdot K^2}}$ reported by Paulechka et al. [91] who, however, considered no phosphonium cations and the anions considered were less basic (in terms of the ability to accept a proton) than those investigated here. In Figure 9, the gray shaded region on either side of the dashed line fit represents ±5% of the predicted volumic heat capacity. Also shown in the graph as a red dotted line is the fit from Paulechka et al. [91], along with red shaded ±5% values on either side of that line.

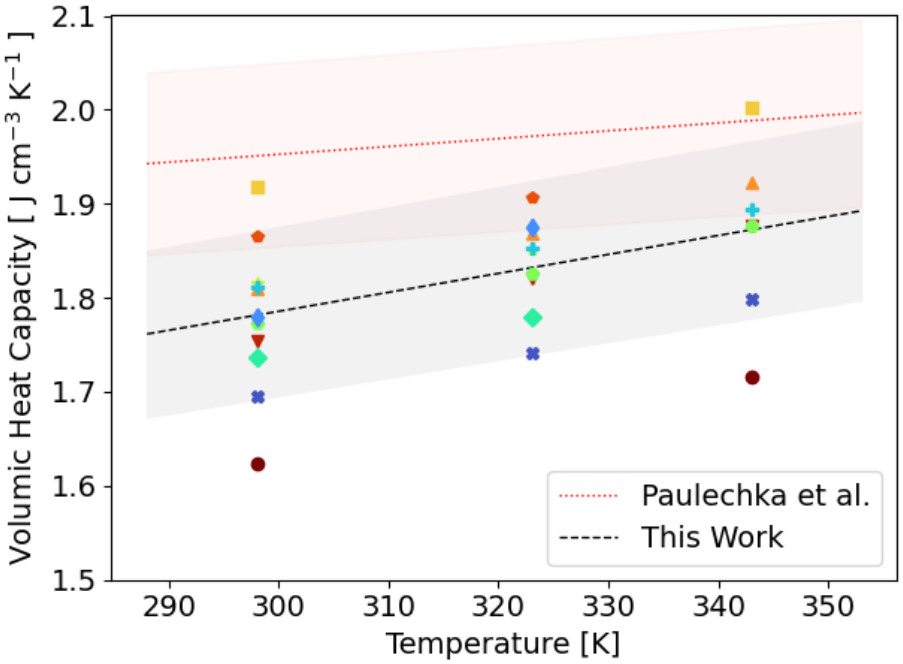


Figure 9. Fitted line of volumic heat capacity versus temperature of 13 AHA ILs. Gray shaded region is ±5% of the line. The red dotted line and ±5% is the fit by Paulechka et al. [91] for a different set of ILs.

With this relationship established between the volumic heat capacity and temperature, a GC/SVR density model was trained on all of the data besides the four test set ILs and was used to predict the densities of these ILs at the temperature of each heat capacity observation. Using these densities, the volumic heat capacity estimated with **Eq. 5**, and the molecular weights of each IL, the molar heat capacity was predicted and has a %AARD of 3.0% relative to the experimental data. The results of this prediction are shown in a parity plot in **Figure 10**.

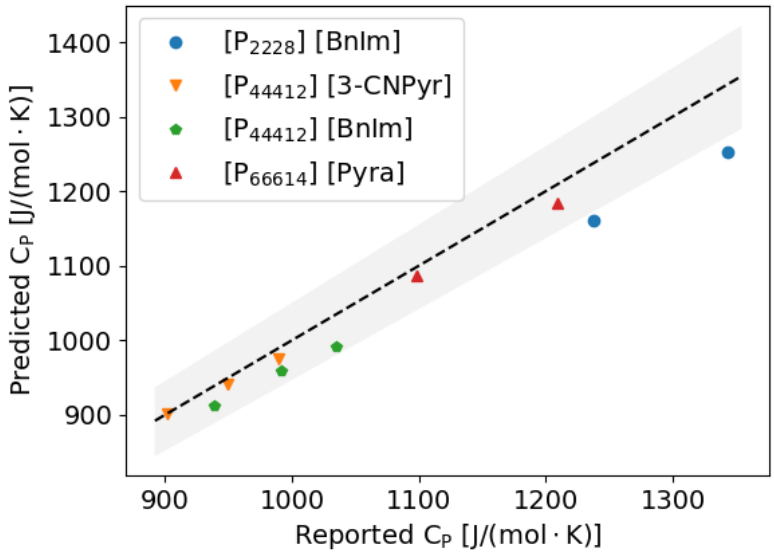


Figure 10. Parity plot of heat capacity model. Gray shaded region is ±5%.

IV. Conclusions and Recommendations

We have expanded upon the framework developed by Nordness et al. [18] to predict the viscosity, density, and heat capacity of phosphonium AHA ionic liquids. This is an important class of ILs that can reversibly react with CO_2 and have potential for CO_2 capture applications. Several bin sizes for feature extraction from COSMO-RS sigma profiles were considered as was covariance between the extracted features and its impacts on regression. We screened through multiple SVR and GBR models and analyzed both group contribution (GC) and several COSMO-RS sigma-profile-based descriptor sets and determined that SVR models with GC-based descriptors offered the best predictions for this data set. The ultimate goal is incorporation of these models into a combined chemical process and material design framework to aid in the discovery of even better AHA ILs for CO_2 capture.

CRediT authorship contribution statement

Austin N. Keller: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – original draft. **Pratik Kelkar:** Methodology, Software, Formal analysis. **Michael Baldea:** Supervision, Writing – review & editing, Funding acquisition. **Mark A. Stadtherr:** Supervision, Writing – review & editing, Funding acquisition. **Joan F. Brennecke:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the University of Texas Energy Institute under the Fueling a Sustainable Energy Transition program. It was also partially funded by the Robert A. Welch Foundation (grant no. F-1945-20180324) as well as by the National Science Foundation (NSF) under the Environmental Convergence Opportunities in Chemical, Bioengineering, Environmental, and Transport Systems program (ECO-CBET). We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC and storage resources that have contributed to the research results reported within this paper.

Appendix A. Supplementary Material

1) Supplementary Materials document containing sample DFT command file for generating sigma profiles and code snippets of sample codes for density, viscosity and heat capacity fitting
2) SI Zip file containing
    a. Supplementary Data Spreadsheet – all sigma profiles, gc parametrizations and data used in this work
    b. Python files of sample code
    c. Binary file with cross validation folds needed to run sample code ("ViscFolds.pkl" and "DensFolds.pkl")

References

[1] D.Y.C. Leung, G. Caramanna, M.M. Maroto-Valer, An overview of current status of carbon dioxide capture and storage technologies, Renewable and Sustainable Energy Reviews. 39 (2014) 426–443. https://doi.org/10.1016/j.rser.2014.07.093.

[2] G.T. Rochelle, Conventional amine scrubbing for CO_2 capture, in: P.H.M. Feron (Ed.), Absorption-Based Post-Combustion Capture of Carbon Dioxide, Woodhead Publishing, 2016: pp. 35–67. https://doi.org/10.1016/B978-0-08-100514-9.00003-2.

[3] B.E. Gurkan, T.R. Gohndrone, M.J. McCready, J.F. Brennecke, Reaction kinetics of CO_2 absorption in to phosphonium based anion-functionalized ionic liquids, Phys. Chem. Chem. Phys. 15 (2013) 7796–7811. https://doi.org/10.1039/C3CP51289D.

[4] B. Hong, L.D. Simoni, J.E. Bennett, J.F. Brennecke, M.A. Stadtherr, Simultaneous Process and Material Design for Aprotic *N*-Heterocyclic Anion Ionic Liquids in Postcombustion CO_2 Capture, Ind. Eng. Chem. Res. 55 (2016) 8432–8449. https://doi.org/10.1021/acs.iecr.6b01919.

[5] J.J. Fillion, H. Xia, M.A. Desilva, M. Quiroz-Guzman, J.F. Brennecke, Phase Transitions, Decomposition Temperatures, Viscosities, and Densities of Phosphonium, Ammonium, and Imidazolium Ionic Liquids with Aprotic Heterocyclic Anions, J. Chem. Eng. Data. 61 (2016) 2897–2914. https://doi.org/10.1021/acs.jced.6b00269.

[6] A.N. Keller, C.L. Bentley, O. Morales-Collazo, J.F. Brennecke, Design and Characterization of Aprotic *N*-Heterocyclic Anion Ionic Liquids for Carbon Capture, J. Chem. Eng. Data. (2022). https://doi.org/10.1021/acs.jced.1c00827.

[7] T. Song, G.M. Avelar Bonilla, O. Morales-Collazo, M.J. Lubben, J.F. Brennecke, Recyclability of Encapsulated Ionic Liquids for Post-Combustion CO_2 Capture, Ind. Eng. Chem. Res. 58 (2019) 4997–5007. https://doi.org/10.1021/acs.iecr.9b00251.
[8] S. Seo, M. Quiroz-Guzman, M.A. DeSilva, T.B. Lee, Y. Huang, B.F. Goodrich, W.F. Schneider, J.F. Brennecke, Chemically Tunable Ionic Liquids with Aprotic Heterocyclic Anion (AHA) for CO_2 Capture, J. Phys. Chem. B. 118 (2014) 5740–5751. https://doi.org/10.1021/jp502279w.
[9] B. Gurkan, B.F. Goodrich, E.M. Mindrup, L.E. Ficke, M. Massel, S. Seo, T.P. Senftle, H. Wu, M.F. Glaser, J.K. Shah, E.J. Maginn, J.F. Brennecke, W.F. Schneider, Molecular Design of High Capacity, Low Viscosity, Chemically Tunable Ionic Liquids for CO_2 Capture, J. Phys. Chem. Lett. 1 (2010) 3494–3499. https://doi.org/10.1021/jz101533k.
[10] M.B. Shiflett, D.W. Drew, R.A. Cantini, A. Yokozeki, Carbon Dioxide Capture Using Ionic Liquid 1-Butyl-3-methylimidazolium Acetate, Energy Fuels. 24 (2010) 5781–5789. https://doi.org/10.1021/ef100868a.
[11] T. Makino, K. Tsunashima, M. Kanakubo, CO_2 absorption and physical properties of tributyloctylphosphonium benzotriazolate, Fluid Phase Equilibria. 494 (2019) 1–7. https://doi.org/10.1016/j.fluid.2019.04.025.
[12] K. Seo, C. Tsay, B. Hong, T.F. Edgar, M.A. Stadtherr, M. Baldea, Rate-Based Process Optimization and Sensitivity Analysis for Ionic-Liquid-Based Post-Combustion Carbon Capture, ACS Sustainable Chem. Eng. 8 (2020) 10242–10258. https://doi.org/10.1021/acssuschemeng.0c03061.
[13] N.S. Evangelista, F.R. do Carmo, R.S. de Santiago-Aguiar, H.B. de Sant'Ana, Development of a New Group Contribution Method Based on GCVOL Model for the Estimation of Pure Ionic Liquid Density over a Wide Range of Temperature and Pressure, Ind. Eng. Chem. Res. 53 (2014) 9506–9512. https://doi.org/10.1021/ie501031r.
[14] R.L. Gardas, J.A.P. Coutinho, Extension of the Ye and Shreeve group contribution method for density estimation of ionic liquids in a wide range of temperatures and pressures, Fluid Phase Equilibria. 263 (2008) 26–32. https://doi.org/10.1016/j.fluid.2007.09.016.
[15] K. Paduszyński, Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 1. Density, Ind. Eng. Chem. Res. 58 (2019) 5322–5338. https://doi.org/10.1021/acs.iecr.9b00130.
[16] J. Jacquemin, P. Nancarrow, D.W. Rooney, M.F. Costa Gomes, P. Husson, V. Majer, A.A.H. Pádua, C. Hardacre, Prediction of Ionic Liquid Properties. II. Volumetric Properties as a Function of Temperature and Pressure, J. Chem. Eng. Data. 53 (2008) 2133–2143. https://doi.org/10.1021/je8002817.
[17] Y. Chen, G.M. Kontogeorgis, J.M. Woodley, Group Contribution Based Estimation Method for Properties of Ionic Liquids, Ind. Eng. Chem. Res. 58 (2019) 4277–4292. https://doi.org/10.1021/acs.iecr.8b05040.
[18] O. Nordness, P. Kelkar, Y. Lyu, M. Baldea, M.A. Stadtherr, J.F. Brennecke, Predicting thermophysical properties of dialkylimidazolium ionic liquids from sigma profiles, Journal of Molecular Liquids. 334 (2021) 116019. https://doi.org/10.1016/j.molliq.2021.116019.
[19] I. Baskin, A. Epshtein, Y. Ein-Eli, Benchmarking machine learning methods for modeling physical properties of ionic liquids, Journal of Molecular Liquids. 351 (2022) 118616. https://doi.org/10.1016/j.molliq.2022.118616.

[20] V. Venkatraman, S. Evjen, K.C. Lethesh, J.J. Raj, H.K. Knuutila, A. Fiksdahl, Rapid, comprehensive screening of ionic liquids towards sustainable applications, Sustainable Energy Fuels. 3 (2019) 2798–2808. https://doi.org/10.1039/C9SE00472F.
[21] S.J. Brown, D. Yalcin, S. Pandiancherri, T.C. Le, I. Orhan, K. Hearn, Q. Han, C.J. Drummond, T.L. Greaves, Characterising a protic ionic liquid library with applied machine learning algorithms, Journal of Molecular Liquids. 367 (2022) 120453. https://doi.org/10.1016/j.molliq.2022.120453.
[22] S. Zhang, Q. Jia, F. Yan, S. Xia, Q. Wang, Evaluating the properties of ionic liquid at variable temperatures and pressures by quantitative structure–property relationship (QSPR), Chemical Engineering Science. 231 (2021) 116326. https://doi.org/10.1016/j.ces.2020.116326.
[23] S. Trohalaki, R. Pachter, G.W. Drake, T. Hawkins, Quantitative Structure−Property Relationships for Melting Points and Densities of Ionic Liquids, Energy Fuels. 19 (2005) 279–284. https://doi.org/10.1021/ef049858q.
[24] J.A. Lazzús, ρ(T, p) model for ionic liquids based on quantitative structure–property relationship calculations, Journal of Physical Organic Chemistry. 22 (2009) 1193–1197. https://doi.org/10.1002/poc.1576.
[25] F. Yan, Q. Shang, S. Xia, Q. Wang, P. Ma, Application of Topological Index in Predicting Ionic Liquids Densities by the Quantitative Structure Property Relationship Method, J. Chem. Eng. Data. 60 (2015) 734–739. https://doi.org/10.1021/je5008668.
[26] S. Hada, R.H. Herring, S.E. Davis, M.R. Eden, Multivariate characterization, modeling, and design of ionic liquid molecules, Computers & Chemical Engineering. 81 (2015) 310–322. https://doi.org/10.1016/j.compchemeng.2015.04.009.
[27] R. Bini, M. Malvaldi, W.R. Pitner, C. Chiappe, QSPR correlation for conductivities and viscosities of low-temperature melting ionic liquids, Journal of Physical Organic Chemistry. 21 (2008) 622–629. https://doi.org/10.1002/poc.1337.
[28] C. Han, G. Yu, L. Wen, D. Zhao, C. Asumana, X. Chen, Data and QSPR study for viscosity of imidazolium-based ionic liquids, Fluid Phase Equilibria. 300 (2011) 95–104. https://doi.org/10.1016/j.fluid.2010.10.021.
[29] B.-K. Chen, M.-J. Liang, T.-Y. Wu, H.P. Wang, A high correlate and simplified QSPR for viscosity of imidazolium-based ionic liquids, Fluid Phase Equilibria. 350 (2013) 37–42. https://doi.org/10.1016/j.fluid.2013.04.009.
[30] G. Yu, L. Wen, D. Zhao, C. Asumana, X. Chen, QSPR study on the viscosity of bis(trifluoromethylsulfonyl)imide-based ionic liquids, Journal of Molecular Liquids. 184 (2013) 51–59. https://doi.org/10.1016/j.molliq.2013.04.021.
[31] R. Alcalde, G. García, M. Atilhan, S. Aparicio, Systematic Study on the Viscosity of Ionic Liquids: Measurement and Prediction, Ind. Eng. Chem. Res. 54 (2015) 10918–10924. https://doi.org/10.1021/acs.iecr.5b02713.
[32] W. Beckner, C.M. Mao, J. Pfaendtner, Statistical models are able to predict ionic liquid viscosity across a wide range of chemical functionalities and experimental conditions, Mol. Syst. Des. Eng. 3 (2018) 253–263. https://doi.org/10.1039/C7ME00094D.
[33] Z.K. Koi, W.Z.N. Yahya, R.A.A. Talip, K.A. Kurnia, Prediction of the viscosity of imidazolium-based ionic liquids at different temperatures using the quantitative structure property relationship approach, New J. Chem. 43 (2019) 16207–16217. https://doi.org/10.1039/C9NJ03436F.

[34] H. Matsuda, H. Yamamoto, K. Kurihara, K. Tochigi, Computer-aided reverse design for ionic liquids by QSPR using descriptors of group contribution type for ionic conductivities and viscosities, Fluid Phase Equilibria. 261 (2007) 434–443. https://doi.org/10.1016/j.fluid.2007.07.018.
[35] K. Tochigi, H. Yamamoto, Estimation of Ionic Conductivity and Viscosity of Ionic Liquids Using a QSPR Model, J. Phys. Chem. C. 111 (2007) 15989–15994. https://doi.org/10.1021/jp073839a.
[36] J.A. Lazzús, G. Pulgar-Villarroel, A group contribution method to estimate the viscosity of ionic liquids at different temperatures, Journal of Molecular Liquids. 209 (2015) 161–168. https://doi.org/10.1016/j.molliq.2015.05.030.
[37] K. Paduszyński, U. Domańska, Viscosity of Ionic Liquids: An Extensive Database and a New Group Contribution Model Based on a Feed-Forward Artificial Neural Network, J. Chem. Inf. Model. 54 (2014) 1311–1324. https://doi.org/10.1021/ci500206u.
[38] K. Paduszyński, Extensive Databases and Group Contribution QSPRs of Ionic Liquids Properties. 2. Viscosity, Ind. Eng. Chem. Res. 58 (2019) 17049–17066. https://doi.org/10.1021/acs.iecr.9b03150.
[39] A.D. Boualem, K. Argoub, A.M. Benkouider, A. Yahiaoui, K. Toubal, Viscosity prediction of ionic liquids using NLR and SVM approaches, Journal of Molecular Liquids. 368 (2022) 120610. https://doi.org/10.1016/j.molliq.2022.120610.
[40] R.L. Gardas, J.A.P. Coutinho, A Group Contribution Method for Heat Capacity Estimation of Ionic Liquids, Ind. Eng. Chem. Res. 47 (2008) 5751–5757. https://doi.org/10.1021/ie800330v.
[41] M. Sattari, F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, Estimation of the Heat Capacity of Ionic Liquids: A Quantitative Structure–Property Relationship Approach, Ind. Eng. Chem. Res. 52 (2013) 13217–13221. https://doi.org/10.1021/ie401782n.
[42] Y. Zhao, S. Zeng, Y. Huang, R.M. Afzal, X. Zhang, Estimation of Heat Capacity of Ionic Liquids Using $S_{\sigma\text{-profile}}$ Molecular Descriptors, Ind. Eng. Chem. Res. 54 (2015) 12987–12992. https://doi.org/10.1021/acs.iecr.5b03576.
[43] A. Paternò, R. Fiorenza, S. Marullo, G. Musumarra, S. Scirè, Prediction of ionic liquid's heat capacity by means of their in silico principal properties, RSC Adv. 6 (2016) 36085–36089. https://doi.org/10.1039/C6RA05106E.
[44] X. Kang, X. Liu, J. Li, Y. Zhao, H. Zhang, Heat Capacity Prediction of Ionic Liquids Based on Quantum Chemistry Descriptors, Ind. Eng. Chem. Res. 57 (2018) 16989–16994. https://doi.org/10.1021/acs.iecr.8b03668.
[45] F. Yan, Y. Shi, Y. Wang, Q. Jia, Q. Wang, S. Xia, QSPR models for the properties of ionic liquids at variable temperatures based on norm descriptors, Chemical Engineering Science. 217 (2020) 115540. https://doi.org/10.1016/j.ces.2020.115540.
[46] Q. Dong, C.D. Muzny, A. Kazakov, V. Diky, J.W. Magee, J.A. Widegren, R.D. Chirico, K.N. Marsh, M. Frenkel, ILThermo: A Free-Access Web Database for Thermodynamic Properties of Ionic Liquids, J. Chem. Eng. Data. 52 (2007) 1151–1159. https://doi.org/10.1021/je700171f.
[47] A.F. Kazakov, J.W. Magee, R.D. Chirico, E. Paulechka, V. Diky, C.D. Muzny, K. Kroenlein, M. Frenkel, NIST Standard Reference Database 147: NIST Ionic Liquids Database - (ILThermo), (2022). https://ilthermo.boulder.nist.gov/.

[48] H. Benimam, C.S. Moussa, M. Hentabli, S. Hanini, M. Laidi, Dragonfly-Support Vector Machine for Regression Modeling of the Activity Coefficient at Infinite Dilution of Solutes in Imidazolium Ionic Liquids Using σ-Profile Descriptors, J. Chem. Eng. Data. 65 (2020) 3161–3172. https://doi.org/10.1021/acs.jced.0c00168.
[49] A. Baghban, M.N. Kardani, S. Habibzadeh, Prediction viscosity of ionic liquids using a hybrid LSSVM and group contribution method, Journal of Molecular Liquids. 236 (2017) 452–464. https://doi.org/10.1016/j.molliq.2017.04.019.
[50] Z.K. Koi, W.Z.N. Yahya, K.A. Kurnia, Prediction of ionic conductivity of imidazolium-based ionic liquids at different temperatures using multiple linear regression and support vector machine algorithms, New J. Chem. 45 (2021) 18584–18597. https://doi.org/10.1039/D1NJ01831K.
[51] P. Dhakal, J. Shah, A Generalized Machine Learning Model for Predicting Ionic Conductivity for Ionic Liquids, Mol. Syst. Des. Eng. (2022). https://doi.org/10.1039/D2ME00046F.
[52] V. Venkatraman, S. Evjen, H.K. Knuutila, A. Fiksdahl, B.K. Alsberg, Predicting ionic liquid melting points using machine learning, Journal of Molecular Liquids. (2020) 114686. https://doi.org/10.1016/j.molliq.2020.114686.
[53] T. Le, V.C. Epa, F.R. Burden, D.A. Winkler, Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties, Chem. Rev. 112 (2012) 2889–2919. https://doi.org/10.1021/cr200066h.
[54] F. Philippi, D. Rauber, K.L. Eliasen, N. Bouscharain, K. Niss, C.W.M. Kay, T. Welton, Pressing matter: why are ionic liquids so viscous?, Chem. Sci. 13 (2022) 2735–2743. https://doi.org/10.1039/D1SC06857A.
[55] A. Klamt, Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena, J. Phys. Chem. 99 (1995) 2224–2235. https://doi.org/10.1021/j100007a062.
[56] A. Klamt, G. Schüürmann, COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, J. Chem. Soc., Perkin Trans. 2. (1993) 799–805. https://doi.org/10.1039/P29930000799.
[57] A. Klamt, F. Eckert, W. Arlt, COSMO-RS: An Alternative to Simulation for Calculating Thermodynamic Properties of Liquid Mixtures, Annu. Rev. Chem. Biomol. Eng. 1 (2010) 101–122. https://doi.org/10.1146/annurev-chembioeng-073009-100903.
[58] Y. Zhao, Y. Huang, X. Zhang, S. Zhang, A quantitative prediction of the viscosity of ionic liquids using Sσ-profile molecular descriptors, Phys. Chem. Chem. Phys. 17 (2015) 3761–3767. https://doi.org/10.1039/C4CP04712E.
[59] L. Cao, P. Zhu, Y. Zhao, J. Zhao, Using machine learning and quantum chemistry descriptors to predict the toxicity of ionic liquids, Journal of Hazardous Materials. 352 (2018) 17–26. https://doi.org/10.1016/j.jhazmat.2018.03.025.
[60] F. Gharagheizi, P. Ilani-Kashkouli, A.H. Mohammadi, D. Ramjugernath, D. Richon, Development of a group contribution method for determination of viscosity of ionic liquids at atmospheric pressure, Chemical Engineering Science. 80 (2012) 326–333. https://doi.org/10.1016/j.ces.2012.06.045.
[61] D.K. Mital, P. Nancarrow, T.H. Ibrahim, N. Abdel Jabbar, M.I. Khamis, Ionic Liquid Melting Points: Structure–Property Analysis and New Hybrid Group Contribution Model, Ind. Eng. Chem. Res. 61 (2022) 4683–4706. https://doi.org/10.1021/acs.iecr.1c04292.

[62] J.A. Lazzús, A group contribution method to predict the melting point of ionic liquids, Fluid Phase Equilibria. 313 (2012) 1–6. https://doi.org/10.1016/j.fluid.2011.09.018.
[63] Z. Song, H. Shi, X. Zhang, T. Zhou, Prediction of CO_2 solubility in ionic liquids using machine learning methods, Chemical Engineering Science. 223 (2020) 115752. https://doi.org/10.1016/j.ces.2020.115752.
[64] S. Seo, L.D. Simoni, M. Ma, M.A. DeSilva, Y. Huang, M.A. Stadtherr, J.F. Brennecke, Phase-Change Ionic Liquids for Postcombustion CO_2 Capture, Energy Fuels. 28 (2014) 5968–5977. https://doi.org/10.1021/ef501374x.
[65] S. Seo, M.A. DeSilva, H. Xia, J.F. Brennecke, Effect of Cation on Physical Properties and CO_2 Solubility for Phosphonium-Based Ionic Liquids with 2-Cyanopyrrolide Anions, J. Phys. Chem. B. 119 (2015) 11807–11814. https://doi.org/10.1021/acs.jpcb.5b05733.
[66] L. Sun, O. Morales-Collazo, H. Xia, J.F. Brennecke, Effect of Structure on Transport Properties (Viscosity, Ionic Conductivity, and Self-Diffusion Coefficient) of Aprotic Heterocyclic Anion (AHA) Room Temperature Ionic Liquids. 2. Variation of Alkyl Chain Length in the Phosphonium Cation, J. Phys. Chem. B. 120 (2016) 5767–5776. https://doi.org/10.1021/acs.jpcb.6b03934.
[67] S. Seo, Chemically Tunable Ionic Liquids with Aprotic Heterocyclic Anions for CO_2 Separation, Doctor of Philosophy, The University of Notre Dame, 2014.
[68] C. Shi, M. Quiroz-Guzman, A. DeSilva, J.F. Brennecke, Physicochemical and Electrochemical Properties of Novel Ionic Liquids Containing Aprotic Heterocyclic Anions Doped with Lithium Salts, ECS Trans. 50 (2013) 309. https://doi.org/10.1149/05011.0309ecst.
[69] B.E. Gurkan, Kinetics of CO_2 Absorption of Reactive Ionic Liquids, Doctor of Philosophy, The University of Notre Dame, 2011.
[70] Y. Huang, Thermal Stability and Phase Transitions of Ionic Liquids, Doctor of Philosophy, The University of Notre Dame, 2013.
[71] C. Shi, Development of Ionic Liquids for Li-Ion Battery Applications, University of Notre Dame, 2013.
[72] L. Sun, Investigation of Ionic Liquids for Electrochemical Applications, University of Notre Dame, 2016.
[73] O. Nordness, O. Morales-Collazo, J.F. Brennecke, Uncommon Behavior of Tetra-alkyl-phosphonium 2-Cyano-pyrrolide Ionic Liquids + Glycerol and Triethanolamine Systems, J. Chem. Eng. Data. 65 (2019) 373–384. https://doi.org/10.1021/acs.jced.9b00769.
[74] T. Song, M.J. Lubben, J.F. Brennecke, Solubility of argon, krypton and xenon in ionic liquids, Fluid Phase Equilibria. 504 (2020) 112334. https://doi.org/10.1016/j.fluid.2019.112334.
[75] K. Oster, P. Goodrich, J. Jacquemin, C. Hardacre, A.P.C. Ribeiro, A. Elsinawi, A new insight into pure and water-saturated quaternary phosphonium-based carboxylate ionic liquids: Density, heat capacity, ionic conductivity, thermogravimetric analysis, thermal conductivity and viscosity, The Journal of Chemical Thermodynamics. 121 (2018) 97–111. https://doi.org/10.1016/j.jct.2018.02.013.
[76] B.F. Goodrich, J.C. de la Fuente, B.E. Gurkan, D.J. Zadigian, E.A. Price, Y. Huang, J.F. Brennecke, Experimental Measurements of Amine-Functionalized Anion-Tethered Ionic Liquids with Carbon Dioxide, Ind. Eng. Chem. Res. 50 (2011) 111–118. https://doi.org/10.1021/ie101688a.

[77] T.B. Lee, S. Oh, T.R. Gohndrone, O. Morales-Collazo, S. Seo, J.F. Brennecke, W.F. Schneider, CO_2 Chemistry of Phenolate-Based Ionic Liquids, J Phys Chem B. 120 (2016) 1509–1517. https://doi.org/10.1021/acs.jpcb.5b06934.
[78] M. Quiroz-Guzman, D.P. Fagnant, X.-Y. Chen, C. Shi, J.F. Brennecke, G.S. Goff, W. Runde, Synthesis and characterization of the thermodynamic and electrochemical properties of tetra-alkyl phosphonium oxalate ionic liquids, RSC Adv. 4 (2014) 14840–14846. https://doi.org/10.1039/C4RA01467G.
[79] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 16, (2016).
[80] COSMOthermX, (2019). https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/solvation-chemistry/biovia-cosmotherm/.
[81] F. Eckert, A. Klamt, Fast solvent screening via quantum chemistry: COSMO-RS approach, AIChE Journal. 48 (2002) 369–385. https://doi.org/10.1002/aic.690480220.
[82] A. Klamt, V. Jonas, T. Bürger, J.C.W. Lohrenz, Refinement and Parametrization of COSMO-RS, J. Phys. Chem. A. 102 (1998) 5074–5085. https://doi.org/10.1021/jp980017s.
[83] M. Diedenhofen, A. Klamt, COSMO-RS as a tool for property prediction of IL mixtures—A review, Fluid Phase Equilibria. 294 (2010) 31–38. https://doi.org/10.1016/j.fluid.2010.02.002.
[84] J. Palomar, J.S. Torrecilla, V.R. Ferro, F. Rodríguez, Development of an a Priori Ionic Liquid Design Tool. 1. Integration of a Novel COSMO-RS Molecular Descriptor on Neural Networks, Ind. Eng. Chem. Res. 47 (2008) 4523–4532. https://doi.org/10.1021/ie800056q.
[85] C.A. Faúndez, R.A. Campusano, J.O. Valderrama, Misleading results on the use of artificial neural networks for correlating and predicting properties of fluids. A case on the solubility of refrigerant R-32 in ionic liquids, Journal of Molecular Liquids. 298 (2020) 112009. https://doi.org/10.1016/j.molliq.2019.112009.
[86] D.M. Makarov, Yu.A. Fadeeva, L.E. Shmukler, I.V. Tetko, Beware of proper validation of models for ionic Liquids!, Journal of Molecular Liquids. 344 (2021) 117722. https://doi.org/10.1016/j.molliq.2021.117722.
[87] D.M. Makarov, Y.A. Fadeeva, L.E. Shmukler, I.V. Tetko, Machine learning models for phase transition and decomposition temperature of ionic liquids, Journal of Molecular Liquids. 366 (2022) 120247. https://doi.org/10.1016/j.molliq.2022.120247.
[88] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research. 12 (2011) 2825–2830.

[89] R.L. Gardas, J.A.P. Coutinho, A group contribution method for viscosity estimation of ionic liquids, Fluid Phase Equilibria. 266 (2008) 195–201. https://doi.org/10.1016/j.fluid.2008.01.021.
[90] A.A. Strechan, A.G. Kabo, Y.U. Paulechka, A.V. Blokhin, G.J. Kabo, A.S. Shaplov, E.I. Lozinskaya, Thermochemical properties of 1-butyl-3-methylimidazolium nitrate, Thermochimica Acta. 474 (2008) 25–31. https://doi.org/10.1016/j.tca.2008.05.002.
[91] Y.U. Paulechka, A.G. Kabo, A.V. Blokhin, G.J. Kabo, M.P. Shevelyova, Heat Capacity of Ionic Liquids: Experimental Determination and Correlations with Molar Volume, J. Chem. Eng. Data. 55 (2010) 2719–2724. https://doi.org/10.1021/je900974u.