Rapid, Comprehensive Screening of Ionic Liquids towards Sustainable Applications

Vishwesh Venkatraman[a], Sigvart Evjen[a], Kallidanthiyil Chellappan Lethesh[a], Jaganathan Joshua Raj [b], Hanna K. Knuutila[c], and Anne Fiksdahl[a]

[a] Department of Chemistry, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway
[b] Centre of Research in Ionic Liquids (CORIL), Universiti Teknologi PETRONAS, Bandar Seri Iskandar 32610, Perak, Malaysia
[c] Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway

Abstract

A computational screening strategy applied to 8 million synthetically diverse ionic liquids, demonstrates its value for sustainable solvent design. In contrast to previous studies which have largely focused on single characteristics, important properties such as viscosity, density, thermal stability and toxicity, were estimated using machine learning. Experimental characterization of 15 compounds selected from the library reinforced the utility of the approach. Over 2600 ionic liquids, a majority of which have never been reported before, were investigated for applications in CO_2 and H_2S gas separation, and cellulose dissolution, using density functional theory calculations. Among these, around 250 low cytotoxicity and low viscosity ionic liquids merit further investigation.

1. Introduction

To facilitate the development of ionic liquids (ILs) for sustainable applications, rigorous yet expeditious approaches are desired. Advances in IL discovery and utilization have mainly been driven by laborious, time-consuming experimental testing, underpinned by knowledge of the underlying chemistry. Alternatively, properties of the IL can be estimated by computational methods[1], following which targeted synthesis can be performed. Previous reports employing computations have in large part been restricted to single properties and small systems.[2, 3, 4, 5, 6, 7, 8] The lack of a thorough screening however, leaves other relevant IL characteristics undetermined. For instance, detrimental behaviour such as low stability or high viscosity,

may not be discovered until later in the development process. Other areas of materials research have exploited the synergy between computational quantum-mechanical and thermodynamic approaches[9] to identify improved scintillator materials for γ-ray nuclear detection,[10] solid catalysts,[11] battery electrolyte solvents,[12, 13] and materials for CO_2 capture.[14]

Screening strategies based on machine learning (ML) are well-suited for the rapid evaluation of large libraries of chemical structures[15, 16, 17, 18] With the availability of data repositories such as ILThermo,[19] there has been a steep rise in the use of data-driven methods for modelling IL properties as evidenced by recent publications.[20, 21, 22, 23, 24, 25, 26] The application of statistical learning approaches requires a sufficiently large underlying data set of experimental or theoretically calculated properties, ideally covering a wide range of functionalities. As the amount of IL data increases by the day, the predictions can be expected to improve over time. Currently, those involving more computationally-intensive calculations, such as density functional theory (DFT) and the popular conductor-like screening model for real solvents (COSMO-RS) perform better.[27, 28, 29] However, these are typically applied in a low-throughput fashion with screening times being substantial. On the other hand, COSMO-RS can be used to calculate solute activity, a vital parameter when considering a solvent for a specific application. By using ML methods for an initial screening followed by more focused COSMO-RS calculations, large libraries may be rapidly assessed for desired uses.

Herein, we have employed a multi-property, high-throughput pipeline to facilitate task-specific ionic liquid (IL) discovery (see Figure 1). A combinatorial approach was used to enumerate structures for seven different cation scaffolds: ammonium, imidazolium, phosphonium, piperidinium, pyridinium, pyrrolidinium and sulphonium combined with a diverse set of anions (alkylsulphonates, phenolates, phosphates, triazolides, PF_6, BF_4) yielding close to 8 million combinations, the largest IL-based solvent screening to date. To enable rapid screening of the library, statistical structure-property models calculated for a number of properties of interest were used to prune the list to a reasonable number of promising candidates. Experimental investigations for selected ILs supported by quantum chemical calculations (based on the COSMO-RS theory) demonstrate the efficacy of the approach. Furthermore, a set of over 2600 ILs have been identified for potential applications in gas separation and cellulose dissolution. The presented approach can be readily expanded to other systems, allowing for more streamlined decision making.
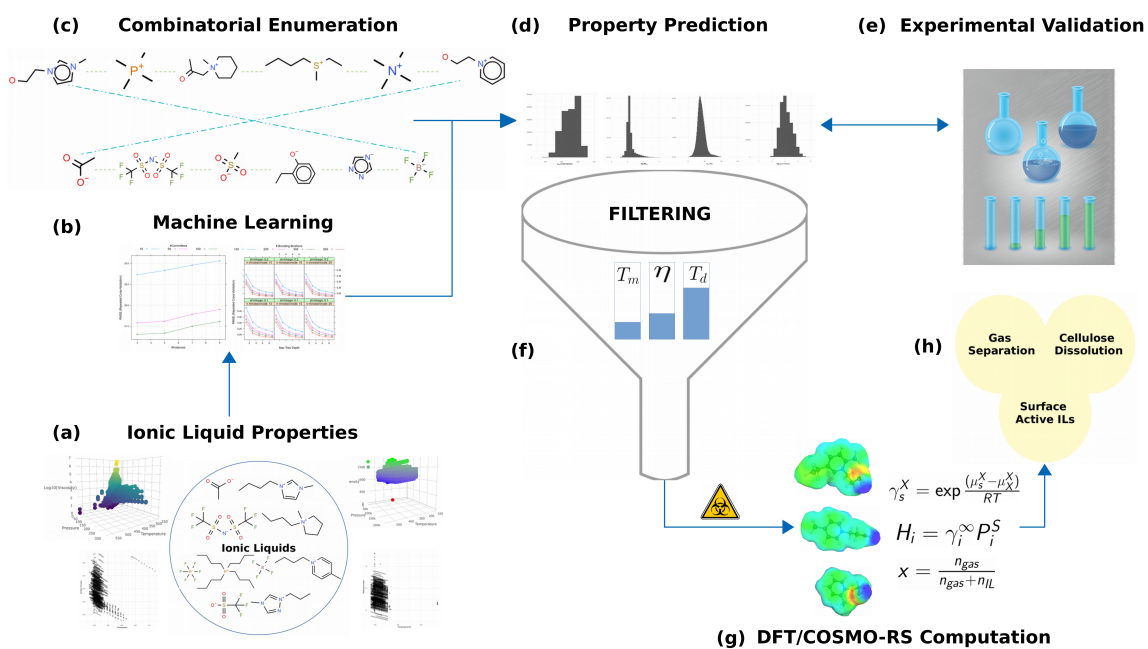
Figure 1: Overall concept of the presented approach: a) Data collection, b) Machine-learning calibration, c) Combinatorial library design and enumeration, d) Prediction of properties by ML, e) Experimental validation of selected candidates, f) Property-based filtering, g) Theoretical evaluation, h) Potential applications.

3

2. Methods

2.1. Data collection

Experimental data for 10 properties: melting point (T_m), glass transition temperature (T_g), thermal decomposition temperature (T_d), heat capacity (C_p), refractive index (n_D), density (ρ), viscosity (η), surface tension (γ), CO_2 solubility (x_{CO_2}) and cytotoxicity[30] towards the leukemia rat cell line IPC-81 $(\log_{10}(EC_{50})$ (in μM) were collated from literature, spanning more than 1500 references (Figure 1a). While a primary source for the data was the NIST ILThermo database,[31] values for T_m, T_d, T_g, and CO_2 solubility were taken from previously published scientific work.[32, 33, 24, 25, 26] In all cases, ILs composed of multiple cations or anions were excluded. Table S1 in the Supplementary Information(I) provides a summary of the collected data.

2.2. Designing the IL Library

Figure 2 shows the building blocks used in the generation of the library. Seven cation families and 38 anions were used to assemble a variety of ILs. Variations in the cationic cores (e.g. pyridinium), were chosen based on commercially available synthetic precursors. The substituents were selected based on alkyl halide reagents suitable for substitution reactions. The functionalities range from standard alkyl chains to glycols and amines[34] often used in task-specific ILs. The library presented herein covers a wide range of ILs of interest to researchers. The cation structures were enumerated using the software SmiLib.[35] A total of 208,268 cations were generated yielding $\sim$ 8 million ILs (see Table 1).

Table 1: The number of structures enumerated for each cation family. The last column gives the total number of cation-anion combinations.

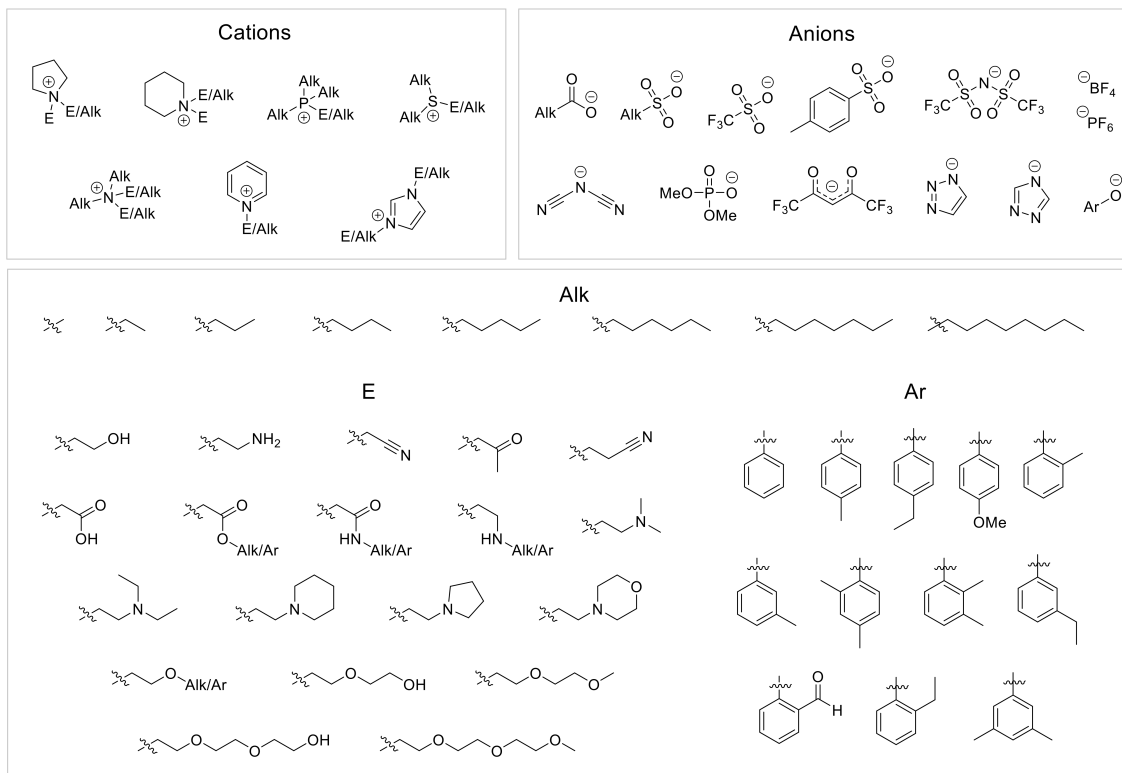| Cation	# Molecules	# Ionic Liquids
ammonium	179466	6819708
imidazolium	5460	207480
phosphonium	7914	300732
piperidinium	5460	207480
pyridinium	1040	39520
pyrrolidinium	5460	207480
sulphonium	3576	135888

Figure 2: The building blocks used for generating the library. The "E/Alk/Ar" represent the attachment points in the generation process.

2.3. Computational Modelling

Descriptor Calculation. The three-dimensional geometries of the molecules were generated using OpenBabel[36] and further optimized at the semi-empirical PM6 level using MOPAC.[37] For computational ease, we have chosen to retain only the lowest energy conformation. Molecular descriptors (HOMO/LUMO energies, polarizabilities, superdelocalizabilities, charge partial surface areas (CPSA) and geometrical indices) were calculated independently for each cation and anion using the software KRAKENX.[38] After eliminating missing values and excluding highly correlated ($R^2 > 0.95$) pairs of variables (of which only one variable in the pair was retained), each IL was represented by a numeric vector of close to 100-120 columns.

Machine Learning. The available data for each property, was divided into independent calibration (67%) and validation (33%) sets. For properties such as viscosity, heat capacity and cytotoxicity, the experimental values were log-transformed. The

models were generated using tree-based ensemble learning methods: random forests (RF), Cubist and gradient boosted regression (GBM) which have been shown to outperform other methods[25, 26]. A repeated 5-fold cross-validation was used to analyze the predictive ability of the models. For each property, the best performing model (among RF, Cubist and GBM) across both calibration and test data was determined based on evaluation metrics (squared coefficient of correlation (R^2), root mean square error ($RMSE$) and the mean absolute error (MAE)). The selected model was subsequently used for further predictions. In order to assess the reliability of the predictions, model-based bootstrapping (100 bootstrap models in this study) was used to provide quantitative measures of predictive uncertainty.[25] The statistical analysis was carried out using the software R.[39]

Quantum Chemistry. Selected combinations of the ions were subjected to DFT/COSMO geometry optimizations (using the quantum chemistry program ORCA[40]) according to the standard quantum chemical method for COSMO-RS i.e. the DFT functional B88-PW86 with a triple zeta valence polarized basis set[41] (TZVP) and the resolution of identity standard approximation. Values of temperature dependent viscosities and densities, heat capacities, and critical micelle concentrations (CMC) were calculated using COSMOtherm (parameterization set BP_TZVP_C30_01601).[42, 43] In all calculations, the ILs were treated as isolated ions. Gas solubilities for COSMO-RS calculations were derived from activity coefficients, Raoult's law and experimental data for the gases, assuming physical solvation. Ideal solubility of gases and partial pressure over the pure liquid were collected from literature.[44, 45, 46] Deviation in derived CO_2 solubilities was <4% from the different sources.

Relative Timings. The calculations were carried out on an Intel Xeon CPU E5-2687W at 3.40GHz with 32 cores and 192 GB RAM. The timings for the DFT calculations varied between 2 minutes to almost an hour for some cases, with average values around 30 minutes (using 16 cores). Semi-empirical MOPAC computations (spread over 4 cores) were considerably quicker with average times being less than 30 seconds with up to 2 minutes for some cases. Descriptor calculations were almost instantaneous taking less than a second per structure. The time taken to screen the entire library took about 25 minutes (0.2 milliseconds/per IL) using a single processor. Screening using COSMOtherm took around 3-4 seconds per IL.

2.4. Experimental Details

Details of the synthesis and characterization of the prepared ILs are described in Supplementary Information(I). Measurements of thermal decomposition temperature, T_d, were performed by a Netzch Libra TG 209 with a heating rate of 20

°C. T_g values were determined by a Netzch DSC 214 Polyma differential scanning calorimeter. Viscosity (η) and density (ρ) were measured using SVM 3000 density and viscosity meter. Refractive indices (n_D) were obtained from a PAL-RI refractometer from Atago. Determination of heat capacity (C_p) was performed with a Pyris1 from Perkin-Elmer. A tensiometer from Krüss was used to determine surface tension (γ). CO_2 solubility experiments were carried out in a Rubotherm apparatus from TA Instruments.

3. Results and Discussion

3.1. Machine Learning Performance

A summary of the fitting results for each of the properties studied is provided in Table 2. With the exception of T_m and T_g, models for the other properties yield correlations > 0.75 across both training and validation data sets. The best predictions were obtained for ρ and $\ln C_p$, while estimations for n_D were generally within 0.01 of the experimental values. The mean absolute deviations (MAE) and average absolute relative deviations ($AARD$) for properties such as T_m, T_g, $\log \eta$ and x_{CO_2} are considerably larger. For x_{CO_2}, values less than 0.10 ($\sim$ 2000 observations) are poorly predicted yielding an overall squared correlation of 0.20 which contributes to the higher AARD across both data sets. Attempts to improve the predictive ability of the models by adding other descriptors[47] based on fragment counts and other two-dimensional indices, yielded negligible changes. Factors affecting the performance include the limited amount of training data and the choice of the molecular encoding. For the latter, previous studies[48, 49, 16, 50] have shown that the marginal gains do not warrant the added complexity of using more variables. The presence of impurities and measurement protocols can also lead to considerable variations in the experimental values. Although schemes for including experimental uncertainties have been suggested,[51, 52] for the data sets used in this study, such values were not always available.

7

Table 2: Performance statistics of the ML models. Where applicable, the associated number of data points with respect to the calibration testing split is specified. The models are based on generalized boosted regression models (GBM), random forests (RF) and Cubist methods. Overall quantitative analysis of the model accuracies are summarized by the squared correlation (R^2), the root mean square error ($RMSE$), mean absolute error (MAE) and the average absolute relative deviation ($AARD$).

Property	Method	Calibration					Test				
		N_{cal}	$RMSE$	MAE	$AARD(\%)$	R^2_{cv}	N_{test}	$RMSE$	MAE	$AARD(\%)$	R^2
T_m (° C)	RF	1486	44	14	39	0.67	726	45	33	97	0.66
T_g (° C)	Cubist	442	21	2	8	0.62	202	19	12	58	0.62
T_d (° C)	RF	833	39	12	5	0.77	455	35	25	10	0.80
$\log_{10}(\eta)$ (mPa.s)	GBM	4658	0.17	0.06	3.9	0.94	3994	0.35	0.23	14.2	0.76
ρ(kg/m³)	GBM	9224	3.85	2.61	0.22	0.99	7730	47.73	28.21	2.29	0.93
$\ln(C_p)$ (J/K/mol)	GBM	6320	0.042	0.012	3.4	0.99	2763	0.28	0.19	3.9	0.91
γ (N/m)	GBM	1863	0.001	0.0002	6.8	0.97	1117	0.004	0.0027	7.6	0.79
n_D	GBM	1646	0.006	0.002	0.15	0.97	1456	0.017	0.011	0.74	0.83
x_{CO_2}	GBM	6062	0.03	0.01	43	0.98	4825	0.09	0.06	163	0.86
$\log_{10}(EC_{50})$ (Rat cell line)	Cubist	157	0.52	0.05	10.9	0.79	70	0.40	0.30	16.3	0.86

Among the decision tree-based modelling schemes used, generalized boosted regression models were found to give slightly better results compared with the random forests and Cubist approaches. The reported performances for T_m, T_d, n_D and x_{CO_2} show only marginal differences compared with those published earlier by our group.[24, 33, 25, 26] For the other properties, the obtained results are in reasonable agreement with previous reports.[48, 53, 20, 54, 55] The descriptors contributing to the performance vary with the property being investigated. Analysis of the variable importance plots (see Figures F1-F2 in Supplementary Information(I)) reveals that the influential variables include a number of energy based descriptors (HOMO/LUMO energies, heat of formation), geometrical indices (volume, area) and charge based descriptors (CPSA, charge dipole) with respect to both the cation and anion. Variables such as the HOMO-LUMO gap and the heat of formation (HOF) reflect the stability of the IL and feature prominently for T_g and ρ. The nucleophilic delocalizabilities associated with cation/anion influence T_d to a large extent.[33] The electronic structural features captured by the atomic charges (most negative or most positive) and partial surface area descriptors influence intermolecular (nucleophilic-electrophilic) interactions and are likely to play crucial roles C_p, cytotoxicity and η. Geometrical indices such as the ovality/globularity impact both x_{CO_2}[24] and n_D[25], and reflect the ability of the molecule to adapt its shape with respect to the approaching reactant.

3.2. Predictions from Machine Learning

The ML-based predictive screening focused on primary properties of interest, namely T_m, T_d, T_g, η and ρ, all of which play significant roles in many applications. While the variations within a single family of cations differs considerably with functionality (see Figure 3), the anions have the greatest influence on the properties of the ILs (shown in Figure 4). Observed trends for the various properties were generally found to be in reasonable agreement with the underlying experimental data. Extremes in values are nonetheless seen for some cation-anion pairs. From Figure 3, it is clear that phosphonium ILs have higher thermal stabilities compared to the ammonium ILs[56, 57] which can be attributed to the difference in their thermal decomposition mechanisms.[58] In particular, weakly coordinating anions such as bistriflimide (NTf$_2$) and PF$_6$, typically possess high thermal stabilities[59] that are also reflected in the estimated values. On the other hand, the phenolate and triazolide-based ILs possessed lower T_d due to their alkaline nature.[60] Predicted values for T_m range between -25 to 155 °C. Within this range, ILs associated with acetate, triazolide, phosphate and phenolate anions have T_m ¡ 50 °C. The symmetrical BF$_4$ and PF$_6$ anions possess the highest melting points. Anions containing

9

fluorine, such as triflate, PF_6 and NTf_2 have higher densities than those without.(see Figure 4) Lower viscosities were seen for the sulphonium ILs due to the bigger size of the sulphur atom, followed by the imidazolium ILs (highly asymmetric cation). Less than a tenth of the ILs were predicted to have viscosities below 300 mPa.s and are populated by those containing hexafluoroacetylacetonate (hfac), acetate, phenolate and NTf_2 anions.
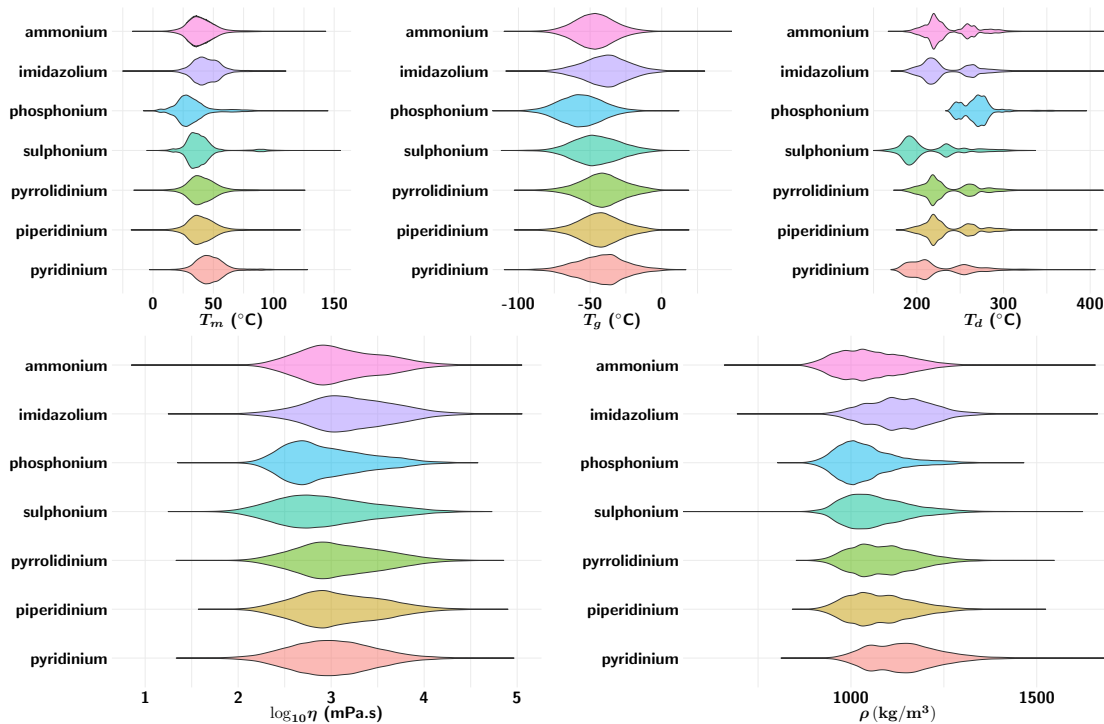

Figure 3: Plot shows the density distribution of the ML-predicted values for the seven cation families investigated. Calculated properties include melting point (T_m), glass transition T_g, thermal decomposition temperature (T_d), viscosity $(\log_{10} \eta)$ and density (ρ) at room temperature and standard pressure. The groups have been colourized for clarity.
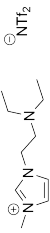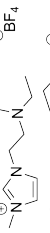
3.3. Validation of the Approach
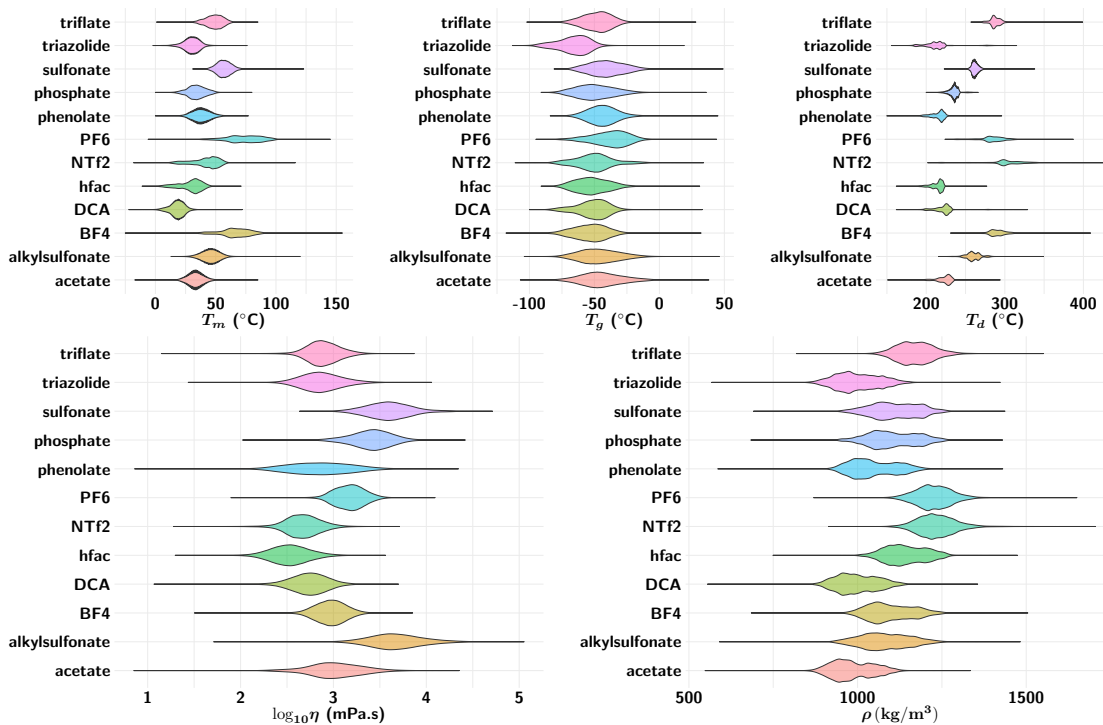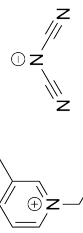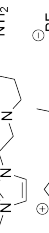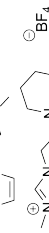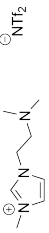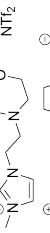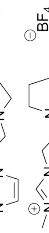
Figure 4: Plot shows the density distribution of the ML-predicted values for melting point (T_m), glass transition T_g, thermal decomposition temperature (T_d), viscosity ($\log_{10} \eta$)and density (ρ) at room temperature and standard pressure. The data has been colourized according to the different anion groups.

11

Table 3: Experimental values for x_{CO_2}, T_g, T_d (10%), η and ρ, C_p, γ and n_d. While η and ρ, C_p, γ and n_d were measured at 20 °C, x_{CO_2} was measured at 25 °C and 0.1 MPa. A "-" indicates that the values were not determined/measured. [a] - value measured at 30 °C.
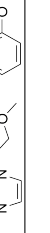
Ionic Liquid	Structure	x_{CO_2}	T_g	T_d	η	ρ	C_p	γ	n_D
01		-	-87	251	43.32	1141.4	317.89	0.0417	>1.5300
02		-	-84	328	130.85	1380.0	-	-	1.4363
03		0.018	-76	299	391.70	1167.2	-	-	1.4403
04		-	-68	327	194.10	1410.1	-	-	1.4490
05		0.017	-64	267	-	-	-	-	1.4353
06		-	-52	280	1848.40	1202.9	404.88	0.0353	1.4565
07		-	-46	325	-	-	-	-	1.4575
08		-	-76	298	-	-	-	-	1.4347
09		0.032	-61	333	193.64	1456.7	595.08	0.0255	1.4512
10		0.033	-68	350	-	-	-	-	1.4485
11		0.014	-57	279	-	-	-	-	1.4551
12		0.503	-68	205	119.50^a	1094.8	-	-	1.5206
13		0.494	-70	211	240.50	1082.8	-	-	1.5214
14		0.488	-68	179	595.50	1060.4	-	-	1.5063
15		-	-59	165	925.88	1132.7	341.84	0.0359	>1.5300

Using the predicted values as a guideline, we synthesized 15 compounds to provide a broad examination of the library. The values were further compared with COSMO-RS calculations (Figure 1e). Cations were selected from three of the classes combined with six different types of anions. Readily available functionalities for which physicochemical properties have not previously been reported, were chosen. The synthesized ILs were selected to span the predicted property range from low to high for T_d, T_g η, x_{CO_2}, ρ and n_D. Structures of the synthesized ILs and associated experimental values are shown in Table 3. The experimental viscosities ranged from 43-1848 mPa.s, T_d from 165-350 °C and ρ from 1060-1457 kg/m^3. A detailed comparison of the ML and corresponding COSMO-RS values is provided in Tables S2-S10 of Supplementary Information(I).

The machine learning predictions and experimental T_g results showed a mean average error (MAE) of 8 °C. The model cannot determine whether an IL will have a glass transition or melting point, and therefore predicts both for all structures. The T_m calculations were not compared with experimental results as none of the synthesized compounds were observed to have a melting point. Decomposition temperatures, T_d, were less accurately predicted with a MAE of 25 °C. ILs with n_D above 1.53 were not reported because these values were outside the range of the apparatus used, but for those measured, values were found to be in line with computational estimates. The predicted trends for density followed the experimental values quite closely, with ML predictions being only marginally poorer in comparison with COSMO-RS. The temperature dependency of the density was well encompassed by the COSMO-RS, having a standard slope deviation of $\sim 5\%$. Although temperature related trends for the heat capacity were not captured accurately, squared correlations > 0.80 were observed, mirroring the performance on the validation set. Rather encouragingly, mean absolute deviations of less than 0.01 were obtained for the surface tension predictions.

Comparison of the experimental and predicted viscosities showed variable agreement (see Table 3). For some compounds, both the ML and COSMO-RS predictions were found to have AARD of $<25\%$. In other cases, the predicted viscosities were significantly under- or overestimated, particularly for those with high viscosities (> 400 mPa.s). Viscosity predictions are generally more challenging because experimental values for the same commercially available compound can vary by more than 30% between studies, depending on experimental procedure and purity.[61, 62, 63, 64] The ML model does not account for any uncertainty in the measurements and is therefore susceptible to discrepancies in the underlying data.

The triazolide (IL-12 to IL-14) CO_2 solubility predictions compared well with the experimental values, despite the ML model not being trained on systems of similar

basicity. Corresponding COSMO-RS predictions showed higher deviations. For the amino-functionalized ILs, the ML estimations were less precise. The training data used as the foundation for the CO_2 predictions contains both physical and chemically absorbing ILs. In most cases, the physical ILs contained no hetero-atoms outside the cationic core, whereas the chemically absorbing structures contained one or more hetero-atoms. As the ML model incorporates the electronegativity/electron density in the calculations, it merely considers the presence of a nitrogen in the imidazolium side chains. The model is unable to distinguish between a primary (chemical absorption) or tertiary amine (physical absorption), and thus incorrectly assumes that ILs containing tertiary amines will be comparable to previously reported primary amines. This result demonstrates an inherent weakness with the use of ML that has to be taken into account when predicting properties and relationships. The COSMO-RS methodology in comparison, correctly identified the amino-functionalized ILs as physical solvents. The models have, in general yielded encouraging results for a majority of the properties investigated, and can serve as a useful guide for the prospective design of task specific ILs. More importantly, the rapid responses provided by data-driven ML make this approach ideal for initial screening of vast libraries. While on average, COSMO-RS performed slightly better than the ML models, the speed-up offered by the latter is considerable.

4. Potential Ionic Liquid Applications

In this section, we highlight potential areas, including CO_2 and H_2S separation, cellulose dissolution and extraction technologies, where the designed ILs can be gainfully employed. Many applications require ILs to be liquid at low temperatures and be thermally stable. High viscosity ILs are likely to increase pumping costs and reduce mass-transfer rates, rendering them unsuitable for most processes. Based on these recommendations, the number of ILs for further investigation was reduced to 2634 following the cut-off criteria: $T_m < 35$ °C, $T_d > 150$ °C and room-temperature viscosity $\eta < 150$ mPa.s (see heatmaps in Figure 1f - 1h). Calibration trends for the ML-based viscosity model suggested that the relative deviations from experimental data were smaller at lower viscosities. Ensuing analysis revealed that nearly 75% of the short-listed ILs were found to have COSMO-RS calculated viscosities below 200 mPa.s. The selected ILs (many of which have not been investigated before) were analyzed using COSMO-RS, wherein activity coefficients at infinite dilution γ^{inf} of molecular solutes in ILs and critical micelle concentrations were computed.[65] The γ^{inf} is often considered as a qualitative measure for the dissolving power of the solvent.[66, 67, 68, 69] The computed properties (using COSMO-RS) for all 2634 ILs are provided in Supplementary Information (II).
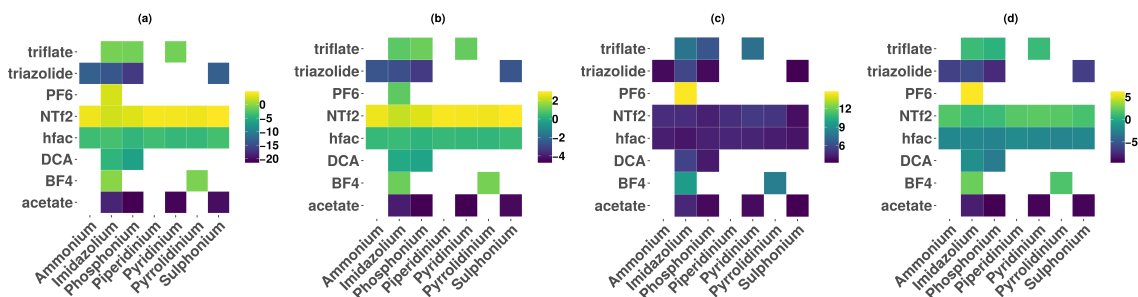
Figure 5: Heatmaps showing the COSMO-RS predicted natural logarithms of activity coefficients at infinite dilution ($\ln \gamma^\infty$ at 25 °C) for (a) microcrystalline cellulose (b) water (c) dodecane (d) naphthenic acid in the 2600 ILs. Each tile is coloured according to the average of the values of $\ln \gamma^\infty$ computed for the cation-anion pair. Missing tiles indicate that the cation-anion combination did not feature in the shortlist. The mole fraction of cellulose was set to 0.5, the mole fractions of IL cations and anions set to 0.25 each.

Cellulose Dissolution and Absorptive Cooling. The computed activity coefficients of microcrystalline cellulose (MCC) and water in a series of ILs are shown in Figures 5a and 5b. The computed values for each cation-anion family were averaged and plotted as a heatmap where darker colours indicate suitable task-specific ILs. Interestingly, the dissolution capabilities of triazolide ILs appear are comparable to those of acetate ILs (Figure 5a). A subsequent experimental study carried out for one of the triazolide ILs (entry 14, Table 3), showed successful dissolution of 2 wt% MCC at 20 °C in 48 h. This was further confirmed by thermogravimetric analysis (TGA) of the dissolved cellulose after regeneration, (see Supporting Information).[70] Although butylpiperidinium ILs are not considered to be the best cations for cellulose dissolution,[71], the results suggest that further improvements are possible. The experimental results and COSMO-RS calculations show that triazolate-based ILs are promising cellulose-dissolving solvents at low temperature. Interestingly, a correlation can be observed for the activity coefficient of cellulose (Figure 5a) and water (Figure 5b) in ILs. Both water and cellulose activities appear to be closely linked to IL H-bond affinity. Triazolide and acetate ILs are potential dehydration or absorption cooling agents (Figure 5b). Viscosity predictions can help with selecting cation-anion pairs with higher cellulose dissolution and dehydration rates.

Gas Separation. The synthesized ILs containing tertiary amines can be used as gas sweetening agents.[72] The negligible vapour pressure and higher chemical stability compared to traditional amine absorbents mitigate solvent loss, and therefore make the ILs interesting alternatives. Furthermore, process intensification is an important driver in the development of gas treatment processes. One such example is the

combined H_2S and water removal, where these compounds are removed simultaneously in one separation unit with the same solvent.[73] Amino-functionalized acetate and dicyanamide (DCA) ILs, short-listed by the ML models, have been identified as interesting solvent candidates for this application. Figure 6 shows a plot of the calculated gas (CH_4, H_2S and CO_2) solubilities.

High pressure applications have been the primary focus of IL-based physical CO_2 separation technology.[74] The triazolide ILs identified by ML react chemically with CO_2 both as neat ILs and in aqueous solution. Thus, they are feasible for CO_2 absorption as they show competitive CO_2 capacities to commercial amine solvents.[72] The high IL decomposition temperatures suggest that they may be useful even at process temperatures. Although the viscosities at 25 °C are somewhat high, the solvents could still be useful for processes at elevated temperatures ($¿$ 70 °C) or high pressure applications where the increased CO_2 partial pressure compensates for the viscosity.[75]

Selective removal of H_2S from gas streams containing H_2S and CO_2 is a potential application for ILs. Traditionally aqueous tertiary amines are used, and the selectivity is mostly based on the faster absorption kinetics of H_2S compared to CO_2.[72] COSMO-RS predictions for the ML-filtered solvents gave up to 16 times higher selectivity towards H_2S compared to CO_2. The selectivities were significantly higher than those seen for aqueous N-methyldiethanolamine, where the equilibria loadings for CO_2 and H_2S are comparable.[76] Similarly, for natural gas sweetening and biogas upgrading processes, the predicted methane solubility can be used to exclude uneconomical solvent candidates, since high methane solubility into the IL will give unwanted methane loss.

Another interesting gas purification technology is the incorporation of ILs in membranes to achieve high CO_2 permeance and selectivity. Several of the studied ILs have potential application in gas separating membranes. A low viscosity is essential in IL-membrane systems to increase diffusivity, and the prediction models applied allow for targeted selection of membrane materials.[77] Interestingly, high viscosity ILs are applicable in IL-supported membranes.[78]

Extraction technologies. The COSMO-RS calculated activity coefficients of dodecane and naphthenic acid in ionic liquids are displayed in Figures 5c and 5d, respectively. The two compounds, dodecane and naphthenic acid were chosen to represent hydrocarbons[79] and organic acids[80], respectively. NTF_2-based ILs possess low water affinities (Figure 5b) and high dodecane solubilities, making them well suited for hydrocarbon extraction. Acetate and triazolide ILs exhibit higher selectivity towards naphthenic acid, compared with dodecane, and may be used to extract acids from oils and emulsions. From the COSMO-RS CMC predictions (see Supplementary

16

Information (II)), PF_6 was found to have the highest critical micelle concentration, in line with previous reports.[81] However, BF_4-based ILs are better suited for preventing or disrupting emulsions in petroleum-water systems. By increasing the critical aggregation point in crude oils, asphaltene and paraffin crystallization in pipelines may be inhibited.

Drug delivery and toxicity. Other avenues for investigation include new non-toxic, biodegradable ILs for drug delivery.[82] By screening the toxicity of candidate molecules, harmful ILs can be omitted in the early stages of investigation. The CMC predictions allow for the selection of ILs with the desired influence on emulsion properties, improving drug solubility and enhancing topical and transdermal delivery.[83] By incorporation of toxicity estimates for all ILs, an initial evaluation of the environmental impact can be assessed. Our analysis of the cytotoxicity predictions reveals that only 0.23% of the 8 million ILs expressed low to moderate toxicity (using the criterion of low IL cytotoxicity: $\log_{10}$ EC_{50} ¿ 3.4 μM).[53] The high prevalence of toxic structures emphasizes the need to consider the hazard potential. For the shortlisted ILs, around 250 compounds are expected to have low hazardous effects (see Figure 7). The library can be modified to include more toxicologically favourable structural elements paving the path for environment-friendly and sustainable ILs.

As demonstrated, the big data approach provides a broad and rational examination of assumptions and possibilities surrounding ILs. More importantly, the use of ML-models to screen millions of compounds based on multiple solvent properties of relevance is a powerful tool, enabling selection of only the most promising structures for experimental consideration.

5. Conclusions

We have presented a rapid, large-scale virtual screening approach to identify promising task-specific ionic liquid (IL) solvents. In the first step, machine learning models were trained on with available experimental data for 10 different IL properties of interest. The models were subsequently applied to a large library of 8 million cation-anion pairs that span diverse chemical scaffolds. The validity of the approach was confirmed both theoretically (using quantum chemistry-based COSMO-RS) and experimentally for 15 ILs. While predictions for properties such as n_D, T_d, C_p and ρ generally reproduced trends observed during calibration, results for others were moderate, but significant enough to justify their use. Property-based filters focusing on T_m, T_d and η yielded a little over 2600 potentially interesting ILs, for which applications in cellulose dissolution and gas separation have been identified. To the best of our knowledge, the exercise carried out herein is the largest ever reported for

17

these compounds. The approach can be used as a framework for targeted screening towards a multitude of applications, for both IL and non-IL systems. Such data-driven filtering can be used to drastically reduce the time and expenditure associated with solvent selection. We nonetheless advocate caution in terms of their use, given that the predictive ability and generalizability of such approaches cannot always be guaranteed.[84]

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

Author contributions

V.V, S.E. and A.F. designed the study. V.V. performed the data curation, developed the models and performed the calculations. S.E. proposed the molecular library, carried out synthesis and experimental measurements. K.C.L. performed synthesis. J.J.R. measured experimental data. A.F. and H.K.K. contributed to the writing of the paper.

References

[1] E. I. Izgorodina, Towards large-scale fully ab initio calculations of ionic liquids, Phys. Chem. Chem. Phys. 13 (2011) 4189–4207. doi:10.1039/C0CP02315A.

[2] O. Borodin, G. D. Smith, Structure and dynamics ofN-methyl-n-propylpyrrolidinium bis(trifluoromethanesulfonyl)imide ionic liquid from molecular dynamics simulations, J Phys. Chem. B 110 (23) (2006) 11481–11490. doi:10.1021/jp061593o.

[3] O. Borodin, Polarizable force field development and molecular dynamics simulations of ionic liquids, J Phys. Chem. B 113 (33) (2009) 11463–11478. doi:10.1021/jp905220k.

[4] S. P. Ong, O. Andreussi, Y. Wu, N. Marzari, G. Ceder, Electrochemical windows of room-temperature ionic liquids from molecular dynamics and density functional theory calculations, Chem. Mater. 23 (11) (2011) 2979–2986. doi:10.1021/cm200679y.

[5] Y. Zhang, C. Shi, J. F. Brennecke, E. J. Maginn, Refined method for predicting electrochemical windows of ionic liquids and experimental validation studies, J. Phys. Chem. B 118 (23) (2014) 6250–6255. doi:10.1021/jp5034257.

[6] S. Pandian, S. Raju, K. S. Hariharan, S. M. Kolake, D.-H. Park, M.-J. Lee, Functionalized ionic liquids as electrolytes for lithium-ion batteries, J. Power Sources 286 (2015) 204–209. doi:10.1016/j.jpowsour.2015.03.130.

[7] K. Karu, A. Ruzanov, H. Ers, V. Ivaništšev, I. Lage-Estebanez, J. G. de la Vega, Predictions of physicochemical properties of ionic liquids with DFT, Computation 4 (4) (2016) 25. doi:10.3390/computation4030025.

[8] N. V. Ilawe, J. Fu, S. Ramanathan, B. M. Wong, J. Wu, Chemical and radiation stability of ionic liquids: A computational screening study, J Phys Chem. C 120 (49) (2016) 27757–27767. doi:10.1021/acs.jpcc.6b08138.

[9] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, Nature Mater. 12 (3) (2013) 191–201. doi:10.1038/nmat3568.

[10] C. Ortiz, O. Eriksson, M. Klintenberg, Data mining and accelerated electronic structure theory as a tool in the search for new functional materials, Comput. Mater. Sci. 44 (4) (2009) 1042–1049. doi:10.1016/j.commatsci.2008.07.016.

[11] J. K. Nørskov, T. Bligaard, J. Rossmeisl, C. H. Christensen, Towards the computational design of solid catalysts, Nature Chem. 1 (1) (2009) 37–46. doi:10.1038/nchem.121.

[12] M. Korth, Large-scale virtual high-throughput screening for the identification of new battery electrolyte solvents: evaluation of electronic structure theory methods, Phys. Chem. Chem. Phys. 16 (17) (2014) 7919–7926. doi:10.1039/c4cp00547c.

[13] T. Husch, M. Korth, Charting the known chemical space for non-aqueous lithium–air battery electrolyte solvents, Phys. Chem. Chem. Phys. 17 (35) (2015) 22596–22603. doi:10.1039/c5cp02937f.

[14] M. T. Dunstan, A. Jain, W. Liu, S. P. Ong, T. Liu, J. Lee, K. A. Persson, S. A. Scott, J. S. Dennis, C. P. Grey, Large scale computational screening and experimental discovery of novel materials for high temperature CO2 capture, Energy Environ. Sci. 9 (4) (2016) 1346–1360. doi:10.1039/c5ee03253a.

[15] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, G. Ceder, Finding nature's missing ternary oxide compounds using machine learning and density functional theory, Chem. Mater. 22 (12) (2010) 3762–3767. doi:10.1021/cm100795d.

[16] T. Le, V. C. Epa, F. R. Burden, D. A. Winkler, Quantitative structure-property relationship modeling of diverse materials properties, Chem. Rev. 112 (5) (2012) 2889–2919. doi:10.1021/cr200066h.

[17] K. A. Phillips, J. F. Wambaugh, C. M. Grulke, K. L. Dionisio, K. K. Isaacs, High-throughput screening of chemicals as functional substitutes using structure-based classification models, Green Chem. 19 (4) (2017) 1063–1074. doi:10.1039/c6gc02744j.

[18] K. Tran, Z. W. Ulissi, Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and h2 evolution, Nat. Catal. 1 (9) (2018) 696–703. doi:10.1038/s41929-018-0142-1.

[19] 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 (4) (2007) 1151–1159. doi:10.1021/je700171f.

[20] 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 (5) (2014) 1311–1324. doi:10.1021/ci500206u.

[21] A. Rybinska-Fryca, A. Sosnowska, T. Puzyn, Prediction of dielectric constant of ionic liquids, J. Mol. Liq. 260 (2018) 57–64. doi:10.1016/j.molliq.2018.03.080.

[22] L. Cao, P. Zhu, Y. Zhao, J. Zhao, Using machine learning and quantum chemistry descriptors to predict the toxicity of ionic liquids, J Hazard. Mater. 352 (2018) 17–26. doi:10.1016/j.jhazmat.2018.03.025.

[23] 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. Sys. Des. Eng. 3 (1) (2018) 253–263. doi:10.1039/c7me00094d.

[24] V. Venkatraman, B. K. Alsberg, Predicting CO2 capture of ionic liquids using machine learning, J. CO2 Util. 21 (2017) 162–168. doi:10.1016/j.jcou.2017.06.012.

[25] V. Venkatraman, S. Evjen, H. K. Knuutila, A. Fiksdahl, B. K. Alsberg, Predicting ionic liquid melting points using machine learning, J. Mol. Liq. 264 (2018) 318–326.

[26] V. Venkatraman, J. J. Raj, S. Evjen, K. C. Lethesh, A. Fiksdahl, In silico prediction and experimental verification of ionic liquid refractive indices, J. Mol. Liq. 264 (2018) 563–570. doi:10.1016/j.molliq.2018.05.067.

[27] A. Klamt, F. Eckert, COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids, Fluid Phase Equilib 172 (1) (2000) 43–72. doi:10.1016/s0378-3812(00)00357-5.

[28] K. A. Kurnia, S. P. Pinho, J. A. P. Coutinho, Designing ionic liquids for absorptive cooling, Green Chem. 16 (8) (2014) 3741. doi:10.1039/c4gc00954a.

[29] X. Liu, Y. Nie, Y. Liu, S. Zhang, A. L. Skov, Screening of ionic liquids for keratin dissolution by means of COSMO-RS and experimental verification, ACS Sustain. Chem. Eng. 6 (12) (2018) 17314–17322. doi:10.1021/acssuschemeng.8b04830.

[30] M. H. Fatemi, P. Izadiyan, Cytotoxicity estimation of ionic liquids based on their effective structural features, Chemosphere 84 (5) (2011) 553–563. doi:10.1016/j.chemosphere.2011.04.021.

[31] A. Kazakov, J. Magee, R. Chirico, E. Paulechka, V. Diky, C. Muzny, K. Kroenlein, M. Frenkel, Nist standard reference database 147: Nist ionic liquids database - (ilthermo), http://ilthermo.boulder.nist.gov/ (2017).

[32] S. Zhang, X. Lu, Q. Zhou, X. Li, X. Zhang, S. Li, Ionic Liquids Physicochemical Properties, Elsevier, Amsterdam, 2009.

[33] V. Venkatraman, B. K. Alsberg, Quantitative structure-property relationship modelling of thermal decomposition temperatures of ionic liquids, J. Mol. Liq. 223 (2016) 60–67. doi:10.1016/j.molliq.2016.08.023.

[34] E. D. Bates, R. D. Mayton, I. Ntai, J. H. Davis, CO2capture by a task-specific ionic liquid, J. Am. Chem. Soc. 124 (6) (2002) 926–927. doi:10.1021/ja017593d.

[35] A. Schüller, V. Hähnke, G. Schneider, SmiLib v2.0: A java-based tool for rapid combinatorial library enumeration, Mol. Inf. 26 (3) (2007) 407–410. doi:10.1002/qsar.200630101.

[36] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open babel: An open chemical toolbox, J. Cheminf. 3 (1) (2011) 33.

[37] J. J. P. Stewart, Mopac2016, stewart Computational Chemistry, Colorado Springs, CO, USA, (http://OpenMOPAC.net) (2016).

[38] V. Venkatraman, B. K. Alsberg, Krakenx: software for the generation of alignment-independent 3d descriptors, J. Mol. Model. 22 (4) (2016) 1–8.

[39] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2017).
URL https://www.R-project.org/

[40] F. Neese, F. Wennmohs, Orca version 3.0.3, max Planck Institute for Chemical Energy Conversion, Germany (2015).

[41] F. Weigend, R. Ahlrichs, Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy, Phys. Chem. Chem. Phys. 7 (18) (2005) 3297.

[42] F. Eckert, A. Klamt, Cosmotherm version c3.0, release 16.01, cOSMOlogic GmbH & Co. KG, Leverkusen, Germany (2015).

[43] A. Klamt, Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena, J. Phys. Chem. 99 (7) (1995) 2224–2235. doi:10.1021/j100007a062.

[44] Gas encyclopedia air liquide, https://encyclopedia.airliquide.com/hydrogen-sulfide, accessed: 01-Mar-2019.

[45] Gas encyclopedia air liquide, https://encyclopedia.airliquide.com/carbon-dioxide, accessed: 01-Mar-2019.

[46] M. P. John, R. Lichtenthaler, E. G. de Azevedo, Molecular Thermodynamics of Fluid-Phase Equilibria, 3rd Edition, Prentice Hall, New Jersey, 1999.

[47] G. Landrum, Rdkit: Open-source cheminformatics software, https://www.rdkit.org/ (2018).

[48] A. Varnek, N. Kireeva, I. V. Tetko, I. I. Baskin, V. P. Solov'ev, Exhaustive qspr studies of a large diverse set of ionic liquids: How accurately can we predict melting points?, J Chem. Inf. Model. 47 (3) (2007) 1111–1122. doi:10.1021/ci600493x.

[49] J. A. P. Coutinho, P. J. Carvalho, N. M. C. Oliveira, Predictive methods for the estimation of thermophysical properties of ionic liquids, RSC Adv. 2 (19) (2012) 7322. doi:10.1039/c2ra20141k.

[50] G. Yu, D. Zhao, L. Wen, S. Yang, X. Chen, Viscosity of ionic liquids: Database, observation, and quantitative structure-property relationship analysis, AIChE J. 58 (9) (2012) 2885–2899. doi:10.1002/aic.12786.

[51] W. M. Czarnecki, I. T. Podolak, Machine learning with known input data uncertainty measure, in: K. Saeed, R. Chaki, A. Cortesi, S. Wierzchoń (Eds.), Computer Information Systems and Industrial Management, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 379–388.

[52] A. Jha, A. Chandrasekaran, C. Kim, R. Ramprasad, Impact of dataset uncertainties on machine learning model predictions: the example of polymer glass transition temperatures, Model. Simul. Mater. Sci. 27 (2) (2019) 024002. doi:10.1088/1361-651x/aaf8ca.

[53] J. S. Torrecilla, J. Palomar, J. Lemus, F. Rodríguez, A quantum-chemical-based guide to analyze/quantify the cytotoxicity of ionic liquids, Green Chem. 12 (1) (2010) 123–134. doi:10.1039/b919806g.

[54] A. Rybinska, A. Sosnowska, M. Barycki, T. Puzyn, Geometry optimization method versus predictive ability in QSPR modeling for ionic liquids, J Comput. Aided Mol. Des. 30 (2) (2016) 165–176. doi:10.1007/s10822-016-9894-3.

[55] A. Rostami, A. Hemmati-Sarapardeh, A. Karkevandi-Talkhooncheh, M. M. Husein, S. Shamshirband, T. Rabczuk, Modeling heat capacity of ionic liquids using group method of data handling: A hybrid and structure-based approach, Int. J. Heat Mass Tran. 129 (2019) 7–17. doi:10.1016/j.ijheatmasstransfer.2018.09.057.

[56] M. Kermanioryani, M. I. A. Mutalib, Y. Dong, K. C. Lethesh, O. B. O. B. Ghanem, K. A. Kurnia, N. F. Aminuddin, J.-M. Leveque, Physicochemical properties of new imidazolium-based ionic liquids containing aromatic group, J Chem. Eng. Data 61 (6) (2016) 2020–2026. doi:10.1021/acs.jced.5b00983.

[57] J. J. Raj, C. D. Wilfred, S. N. Shah, M. Pranesh, M. A. Mutalib, K. C. Lethesh, Physicochemical and thermodynamic properties of imidazolium ionic liquids with nitrile and ether dual functional groups, J Mol. Liq. 225 (2017) 281–289. doi:10.1016/j.molliq.2016.11.049.

[58] S. Sowmiah, V. Srinivasadesikan, M.-C. Tseng, Y.-H. Chu, On the chemical stabilities of ionic liquids, Molecules 14 (9) (2009) 3780–3813. doi:10.3390/molecules14093780.

[59] C. Maton, N. D. Vos, C. V. Stevens, Ionic liquid thermal stabilities: decomposition mechanisms and analysis tools, Chem. Soc. Rev. 42 (13) (2013) 5963. doi:10.1039/c3cs60071h.

[60] S. N. Shah, K. C. Lethesh, M. I. A. Mutalib, R. B. M. Pilus, Evaluation of thermophysical properties of imidazolium-based phenolate ionic liquids, Ind. Eng. Chem. Res. 54 (14) (2015) 3697–3705. doi:10.1021/ie505059g.

[61] F. S. Jerome, J. T. Tseng, L. T. Fan, Viscosities of aqueous glycol solutions, J. Chem. Eng. Data 13 (4) (1968) 496–496. doi:10.1021/je60039a010.

[62] C. M. Kinart, M. Maj, W. J. Kinart, Study on the internal structure of some ethylene glycols with sulfolane binary mixtures by means of measuring their viscosities atT= 303.15 k, Phys. Chem. Liq. 47 (5) (2009) 487–494. doi:10.1080/00319100802256069.

[63] N. G. Tsierkezos, I. E. Molinou, Thermodynamic properties of water + ethylene glycol at 283.15, 293.15, 303.15, and 313.15 k, J. Chem. Eng. Data 43 (6) (1998) 989–993. doi:10.1021/je9800914.

[64] T. M. Aminabhavi, B. Gopalakrishna, Density, viscosity, refractive index, and speed of sound in aqueous mixtures of n,n-dimethylformamide, dimethyl sulfoxide, n,n-dimethylacetamide, acetonitrile, ethylene glycol, diethylene glycol, 1,4-dioxane, tetrahydrofuran, 2-methoxyethanol, and 2-ethoxyethanol at 298.15 k, J. Chem. Eng. Data 40 (4) (1995) 856–861. doi:10.1021/je00020a026.

[65] K. Paduszyński, 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. 19 (19) (2017) 11835–11850. doi:10.1039/c7cp00226b.

[66] J. Kahlen, K. Masuch, K. Leonhard, Modelling cellulose solubilities in ionic liquids using COSMO-RS, Green Chem. 12 (12) (2010) 2172. doi:10.1039/c0gc00200c.

[67] A. Casas, S. Omar, J. Palomar, M. Oliet, M. V. Alonso, F. Rodriguez, Relation between differential solubility of cellulose and lignin in ionic liquids and activity coefficients, RSC Adv. 3 (10) (2013) 3453. doi:10.1039/c2ra22800a.

[68] K. A. Kurnia, M. V. Quental, L. M. N. B. F. Santos, M. G. Freire, J. A. P. Coutinho, Mutual solubilities between water and non-aromatic sulfonium-, ammonium- and phosphonium-hydrophobic ionic liquids, Phys. Chem. Chem. Phys. 17 (6) (2015) 4569–4577. doi:10.1039/c4cp05339g.

[69] Y.-R. Liu, K. Thomsen, Y. Nie, S.-J. Zhang, A. S. Meyer, Predictive screening of ionic liquids for dissolving cellulose and experimental verification, Green Chem. 18 (23) (2016) 6246–6254. doi:10.1039/c6gc01827k.

[70] R. P. Swatloski, S. K. Spear, J. D. Holbrey, R. D. Rogers, , J Am. Chem. Soc. 124 (18) (2002) 4974–4975. doi:10.1021/ja025790m.
URL https://doi.org/10.1021%2Fja025790m

[71] M. Isik, H. Sardon, D. Mecerreyes, Ionic liquids and cellulose: Dissolution, chemical modification and preparation of new cellulosic materials, Int. J. Mol. Sci. 15 (7) (2014) 11922–11940. doi:10.3390/ijms150711922.

[72] A. L. Kohl, R. B. Nielsen, Alkanolamines for hydrogen sulfide and carbon dioxide removal, in: Gas Purification, Elsevier, 1997, pp. 40–186. doi:10.1016/b978-088415220-0/50002-1.

[73] E. Skylogianni, M. I. Lilleng, H. Knuutila, Combined hydrogen sulfide removal and hydrate control for subsea application: Simulation study, in: 2016 Techno-Ocean (Techno-Ocean), IEEE, 2016. doi:10.1109/techno-ocean.2016.7890757.

[74] M. Ramdin, T. W. de Loos, T. J. Vlugt, State-of-the-art of CO2 capture with ionic liquids, Ind. Eng. Chem. Res. 51 (24) (2012) 8149–8177. doi:10.1021/ie3003705.

[75] M. T. Mota-Martinez, J. P. Hallett, N. M. Dowell, Solvent selection and design for CO2 capture – how we might have been missing the point, Sustain. Energ. Fuels 1 (10) (2017) 2078–2090. doi:10.1039/c7se00404d.

[76] F. Y. Jou, A. E. Mather, F. D. Otto, Solubility of hydrogen sulfide and carbon dioxide in aqueous methyldiethanolamine solutions, Ind. Eng. Chem. Process. Des. Dev. 21 (4) (1982) 539–544. doi:10.1021/i200019a001.

[77] L. Lozano, C. Godínez, A. de los Ríos, F. Hernández-Fernández, S. Sánchez-Segado, F. Alguacil, Recent advances in supported ionic liquid membrane technology, J. Membrane Sci. 376 (1-2) (2011) 1–14. doi:10.1016/j.memsci.2011.03.036.

[78] J. Wang, J. Luo, S. Feng, H. Li, Y. Wan, X. Zhang, Recent development of ionic liquid membranes, Green Energy Environ. 1 (1) (2016) 43–61. doi:10.1016/j.gee.2016.05.002.

[79] S. N. Shah, M. Ismail, M. I. A. Mutalib, R. B. M. Pilus, L. K. Chellappan, Extraction and recovery of toxic acidic components from highly acidic oil using ionic liquids, Fuel 181 (2016) 579–586. doi:10.1016/j.fuel.2016.05.041.

[80] S. N. Shah, L. K. Chellappan, G. Gonfa, M. I. A. Mutalib, R. B. M. Pilus, M. A. Bustam, Extraction of naphthenic acid from highly acidic oil using phenolate based ionic liquids, Chem. Eng. J. 284 (2016) 487–493. doi:10.1016/j.cej.2015.09.017.

[81] J. A. Murillo-Hernández, I. García-Cruz, S. López-Ramírez, C. Duran-Valencia, J. M. Domínguez, J. Aburto, Aggregation behavior of heavy crude oil-ionic liquids solutions by fluorescence spectroscopy, Energy Fuels 23 (9) (2009) 4584–4592. doi:10.1021/ef9004175.

[82] I. Marrucho, L. Branco, L. Rebelo, Ionic liquids in pharmaceutical applications, Annu. Rev. Chem. Biomol. Eng. 5 (1) (2014) 527–546. doi:10.1146/annurev-chembioeng-060713-040024.

[83] D. Monti, E. Egiziano, S. Burgalassi, P. Chetoni, C. Chiappe, A. Sanzone, S. Tampucci, Ionic liquids as potential enhancers for transdermal drug delivery, Int. J Pharm. 516 (1) (2017) 45 – 51. doi:https://doi.org/10.1016/j.ijpharm.2016.11.020.

[84] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559 (7715) (2018) 547–555. doi:10.1038/s41586-018-0337-2.
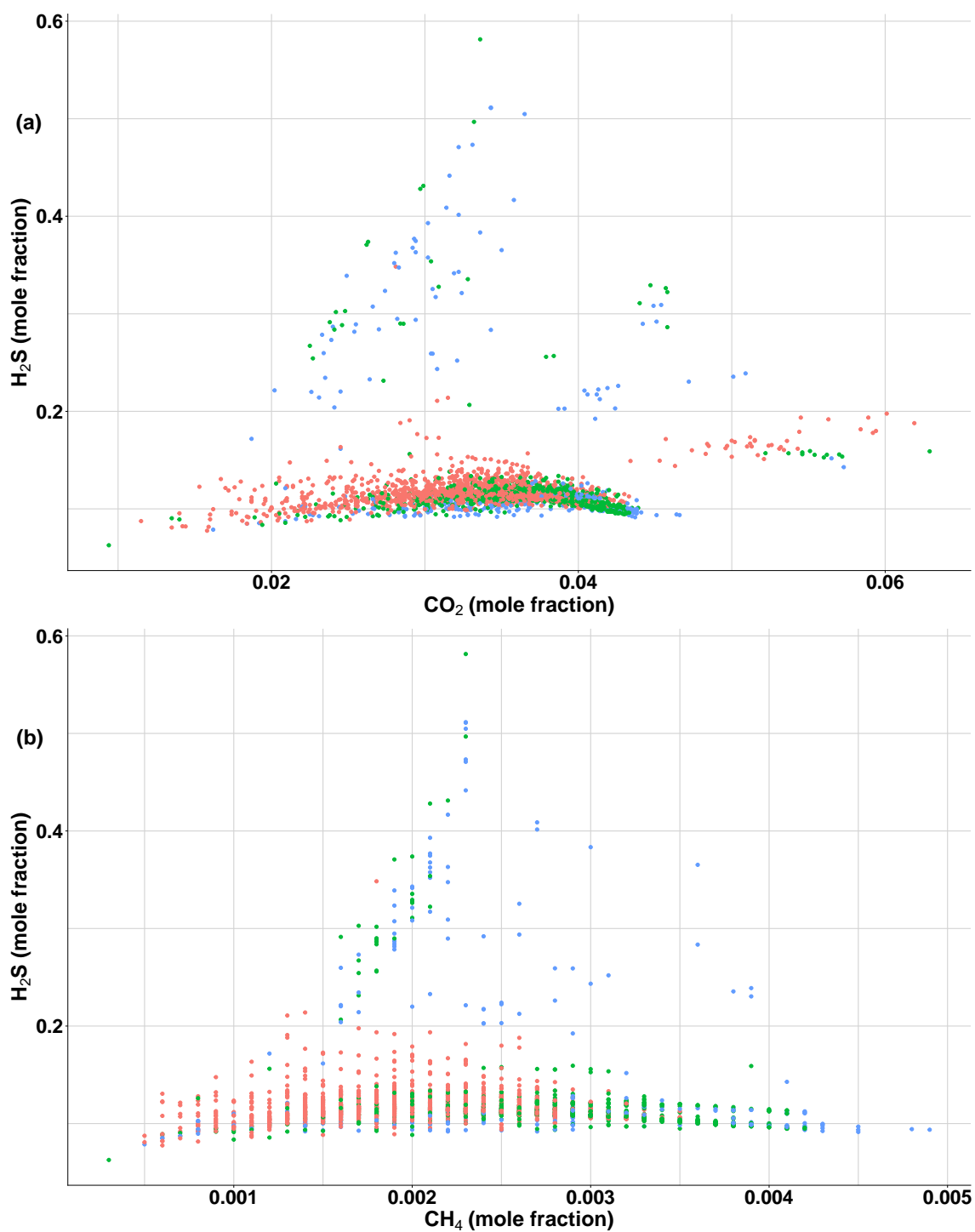
Figure 6: Calculated gas solubilities derived from COSMO-RS based activity coefficients for (a) CO_2 versus H_2S and (b) CH_4 versus H_2S in the 2634 short-listed ILs. The points have been coloured according to the COSMO-RS calculated viscosities (mPa.s at 25 °C) as follows: $\eta < 150$ - red, $150 < \eta < 300$ - green and $\eta > 300$ - blue.
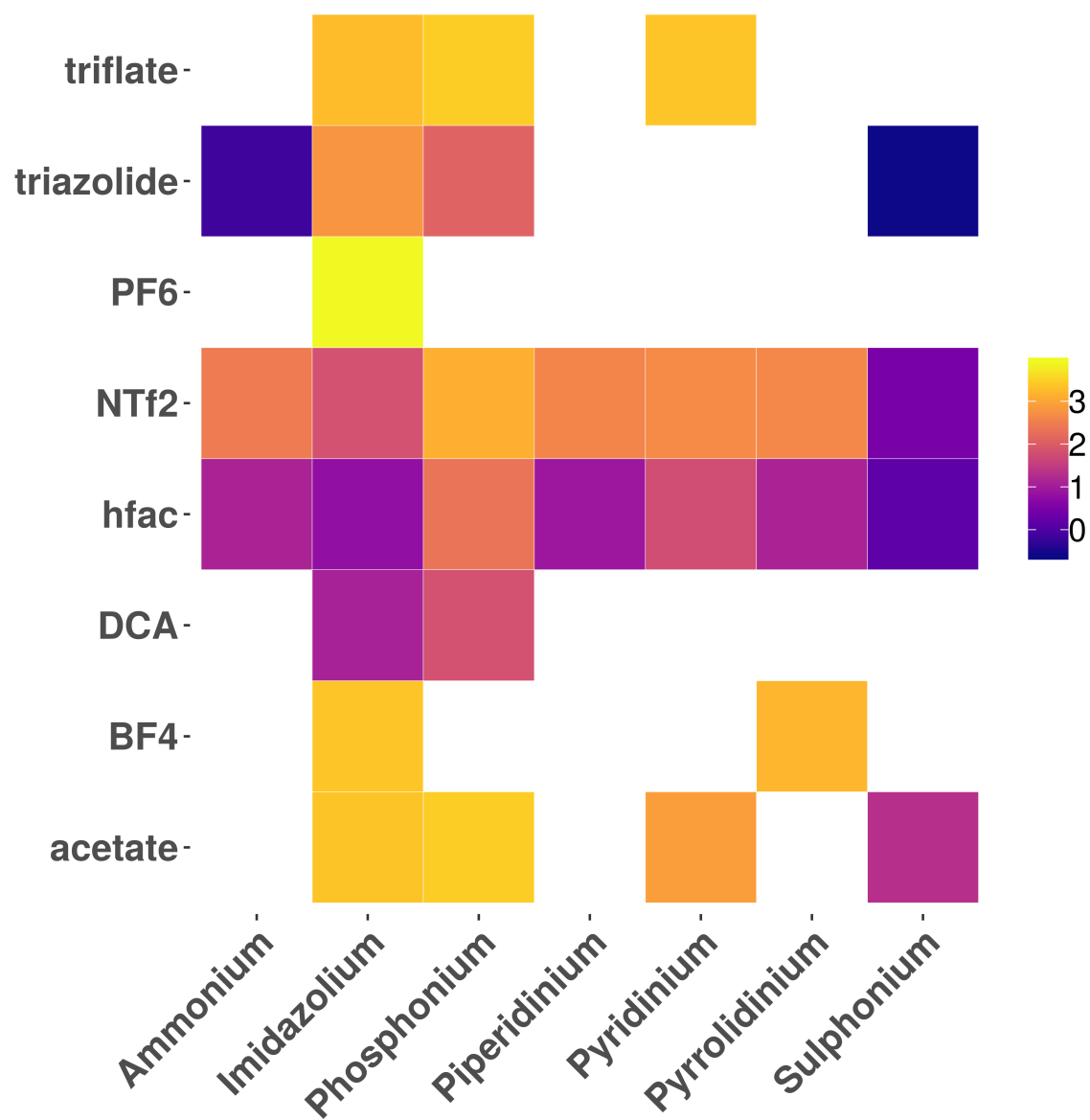
Figure 7: Predicted cytotoxicities $\log_{10}$ EC$_{50}$ IPC-81 values for the short listed ILs. Each tile is coloured according to the mean of the $\log_{10}$ EC$_{50}$ values associated for the cation-anion pair.