PCCP
Physical Chemistry Chemical Physics

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: A. Schindl, R. R. Hawker, K. S. Schaffarczyk McHale, K. T.-C. Liu, D. C. Morris, A. Y. Hsieh, A. Gilbert, S. W. Prescott, R. S. Haines, A. K. Croft, J. B. Harper and C. M. Jäger, *Phys. Chem. Chem. Phys.*, 2020, DOI: 10.1039/D0CP04224B.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

View Article Online
DOI: 10.1039/D0CP04224B

Journal Name

ARTICLE

Controlling the outcome of S_N2 reactions in ionic liquids: From rational data set design to predictive linear regression models†

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Alexandra Schindl,[a] Rebecca R. Hawker,[b,c] Karin S. Schaffarczyk McHale,[b] Kenny T.-C. Liu,[b] Daniel C. Morris,[b,d] Andrew Y. Hsieh,[b] Alyssa Gilbert,[b] Stuart W. Prescott,[d] Ronald S. Haines,[b] Anna K. Croft,[a,*] Jason B. Harper,[b,*] and Christof M. Jäger[a,*]

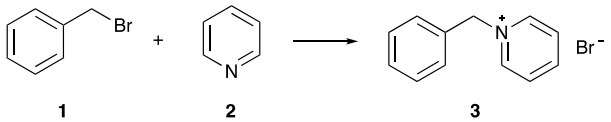
Rate constants for a bimolecular nucleophilic substitution (S_N2) process in a range of ionic liquids are correlated with calculated parameters associated with the charge localisation on the cation of the ionic liquid (including the molecular electrostatic potential). Simple linear regression models proved effective, though the interdependency of the descriptors needs to be taken into account when considering generality. A series of ionic liquids were then prepared and evaluated as solvents for the same process; this data set was rationally chosen to incorporate homologous series (to evaluate systematic variation) and functionalities not available in the original data set. These new data were used to evaluate and refine the original models, which were expanded to include simple artificial neural networks. Along with showing the importance of an appropriate data set and the perils of overfitting, the work demonstrates that such models can be used to reliably predict ionic liquid solvent effects on an organic process, within the limits of the data set.

Introduction

Ionic liquid solvents affect the reaction outcome of organic reactions differently to the more commonly-used molecular solvents.[1-6] Ionic liquids are composed of ions,[7, 8] leading electrostatic interactions to dominate, often affording favourable effects on reaction rates and increased selectivity in the formation of products. In comparison to molecular solvents, however, the origin of reaction outcomes in ionic liquids is less well understood.[2, 4, 9] In order to rationally select the appropriate ionic liquids as solvents for a given reaction, these solvent effects need to be both understood and, more importantly, predictable.

Systematic analysis of a variety of ionic liquids on a range of different reactions (examples include: alkyl substitution,[10-20] aromatic substitution,[21-30] condensation[31-33] and pericyclic[34-40] processes) has provided a greater understanding of the effects of ionic liquids on reaction outcomes. Analysis of similar reactions types has shown the solvent effects are predictable, with specific ionic liquids able to be chosen to control reaction outcome.[25, 26, 41] These analyses have identified some key interactions of the components of ionic liquids with species along the reaction coordinate,[3] involving both starting materials (aromatic systems[24-26, 29, 30] and lone-pairs[18, 19, 30, 31, 33, 42, 43]) and transition states.[16-20, 24]

An example of a reaction that has been examined extensively in ionic liquids is the bimolecular nucleophilic substitution (S_N2) reaction between benzyl bromide **1** and pyridine **2** (Scheme 1).[41-47] When the commonly-used ionic liquid 1-butyl-3-methylimidazolium *bis*(trifluoromethanesulfonyl)imide ([bmim][$N(SO_2CF_3)_2$]) was examined as a solvent in this reaction, a rate constant enhancement, *ca*. 2.5 times that of the molecular solvent, acetonitrile, was observed.[46] Determination of the activation parameters for this reaction provided insight into the microscopic origin of the rate constant enhancement;[46] the reaction was entropically driven and mediated by an interaction between the ionic liquid cation and the lone-pair on the nitrogen of the pyridine **2**. This interaction with the lone-pair of the pyridine **2** was confirmed through a combination of deconvolution studies and molecular dynamics simulations.[42]



1 2 3

Scheme 1. The bimolecular nucleophilic substitution reaction between benzyl bromide **1** and pyridine **2** to give *N*-benzyl pyridinium bromide **3**, which has been examined in ionic liquids and the molecular solvent, acetonitrile.

[a] *Department of Chemical and Environmental Engineering, University of Nottingham, NG7 2RD Nottingham, United Kingdom*
[b] *School of Chemistry, University of New South Wales, UNSW Sydney, 2052, Australia*
[c] *Present address: Department of Chemistry, University of Sheffield, S3 7HF Sheffield, United Kingdom*
[d] *School of Chemical Engineering, University of New South Wales, UNSW Sydney, 2052, Australia*

† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: Correlation of reaction outcome to Kamlet-Taft parameters using multivariate regression analysis; molecular electrostatic potentials for all of the cations considered; link to full set of calculated descriptors; linear correlation and dependency analysis of molecular descriptors; summary of regression models in this study; evaluation of efficacy of final models not included in main text; general experimental; synthesis of the ionic liquids used; stock solution compositions and rate constants from kinetic analysis. See DOI: 10.1039/x0xx00000x

Examination of a wide range of ionic liquids revealed the interaction involving the lone-pair on the pyridine **2** was dependent on the extent of charge localisation and steric availability of the cationic charged centre(s); the more accessible the charged centre(s) on the ionic liquid cation, the larger the rate constant for the reaction.[41, 44] Greater interaction between the ionic liquid cation and the pyridine **2** led to a more substantial enthalpic cost, with this cost overcome by an even larger entropic benefit, resulting in the observed increased rate constants. Ionic liquids with cations that contained electronegative atoms also gave larger rate constants than those with cations that did not,[41, 47] leading to the conclusion that the magnitude of the cationic charge density also has an effect on the rate constant observed. In order to determine the most favourable ionic liquids for this reaction, it was concluded that cations with accessible charges and/or containing electronegative atoms should be chosen.[41]

Whilst the qualitative understanding developed is useful and is similar to what is available in molecular solvents, the ideal would be a method to quantify the interaction identified and correlate it with the observed rate constants. Experimental methods, such as measuring diffusion constants,[17] may provide such information. Alternatively, correlation with solvent parameters, such as those developed by Kamlet and Taft, might be used to determine microscopic interactions; these solvent parameters give an indication of the properties of each component of the ionic liquid. The typically used Kamlet-Taft solvent parameters are α, the hydrogen bond donating ability of the solvent[48] (often associated with the cation of the ionic liquid); β, the hydrogen bond accepting ability of the solvent[49] (often associated with the anion of the ionic liquid); and π^*, the polarisability of the solvent[50] (associated with both constituents of the ionic liquid). These parameters have been correlated with reaction outcomes in ionic liquids previously[19-21, 51-53] and in this case multivariate regression analyses (as described by Welton *et al.*,[54] see ESI for discussion including Fig S1-S2) of the data presented in Table 1 (*vida infra*) with the limited solvent parameter data gave the most significant correlation with the α and π^* parameters, as below.§

$$\ln(k_2) = 4.35\alpha - 10.35\pi^* \qquad (2)$$

This correlation is reasonable and further shows the importance of interactions of the cations *and* demonstrates that the polarizability of the solvent has a (negative) effect on reaction outcome. However, the range of ionic liquids for which these data are reported is comparatively limited and determining them is non-trivial.[55] Such restrictions make their utility for the rational design of new solvents limited.‡

Alternatively, computational methods might provide a simple measure that could be used to choose an ionic liquid solvent. Based on the observations described above, the question arises whether reaction rate enhancements by different ionic liquids might be described simply by using the properties of the cations of the ionic liquids or, to be more precise, by investigating molecular surface properties that determine the nature of the interactions between the cations of the ionic liquid and pyridine **2**. If such a quantitative correlation is possible, regression models might be trained on these data with the potential to predict the solvent effects of ionic liquids without the need for large-scale experimental analysis, and opening the way for more extensive machine-learning approaches with large experimental data-sets.

The molecular electrostatic potential (MEP) and other local molecular properties derived from quantum chemical calculations have been proposed and tested for their applicability for predicting chemical reactivity.[56, 57] These properties have been used for QSAR (quantitative structure-activity relationship) and 3D-QSAR studies in chemistry and medicinal research, with direct impact on drug and agrochemical discovery.[58, 59] In a recent review, Brinck and Stenlid[60] summarize the status and capacity of the molecular surface property approach (MSPA) for the prediction of intermolecular interactions and reactivity beyond the molecular level. Going beyond the molecular surface, we have further recently demonstrated the usability of local molecular properties for describing charge transport in molecular organic electronic devices.[61, 62]

Models using MEPs as a basis to describe site interaction points have been developed specifically for acetonitrile,[63] and a more general approach to estimate solvent effects between solutes has been provided by Hunter.[64] This approach obtained hydrogen bond parameters, that are in congruence with Abraham's solvent parameters,[65] from the maxima and minima of calculated MEPs and thereby demonstrated, that there is 'a logarithmic relationship between the value of the surface electrostatic potential and the association constant for an intermolecular interaction'.[64] The model has recently been extended to identify parameters for, and interactions with, selected ions.[66-68] Similar to this, MEP maxima were used to describe the energetic ordering of equilibrium structures and the number of possible Lewis acid sites.[69]

In the study presented here, rather than considering directly the effects of varying surface properties on the reactivity of a given species, the direct interactions between the cations of the ionic liquid and pyridine **2** are examined as a surrogate for the entropically driven rate constant enhancement. These interactions are predominantly of electrostatic nature and therefore were expected to be described well by the molecular electrostatic potential. We investigate the local molecular properties of the cations of the different ionic liquids already considered experimentally as solvents for the reaction in Scheme 1, based on semiempirical quantum chemical calculations. Relationships between these properties and the rate constant data are then elaborated. Based on these relationships we train a regression model to predict the effect of other ionic liquids on the reaction kinetics and discuss critically the strengths and limitations of the predictions. Subsequently, we use the prediction model to suggest and test new ionic liquid combinations to further evaluate and improve the model.

View Article Online
DOI: 10.1039/D0CP04224B

Journal Name

ARTICLE

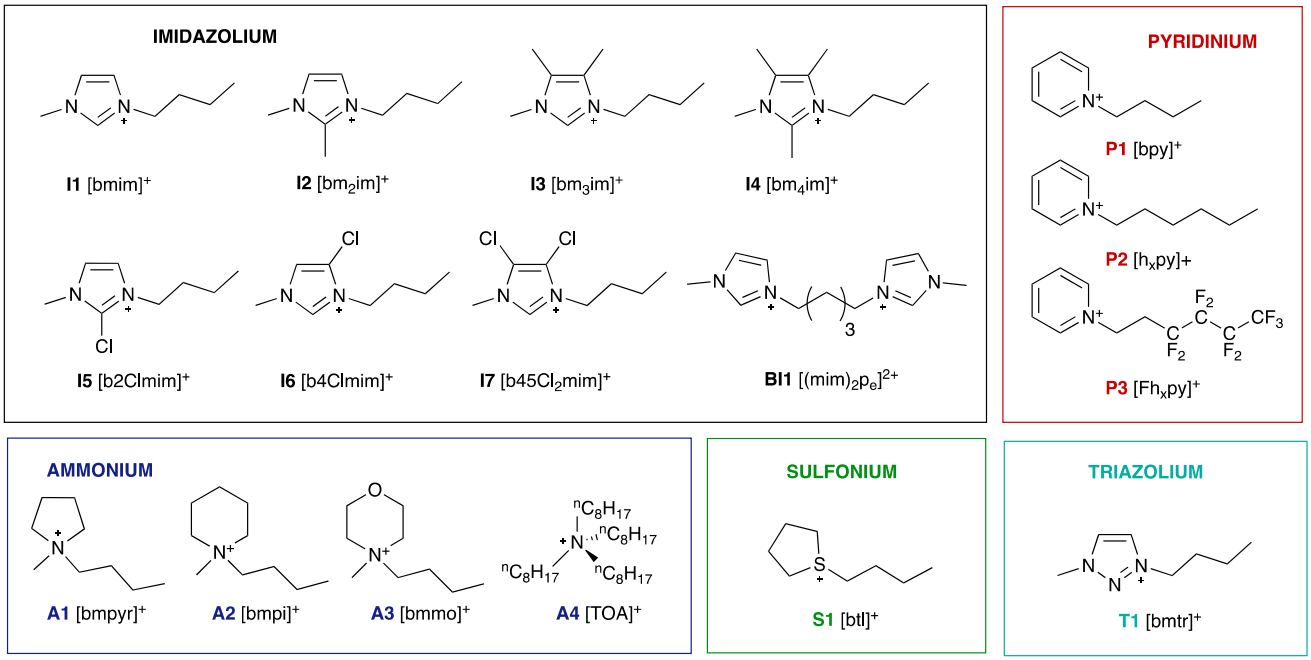


Figure 1. The cations in the ionic liquids considered initially, based on their effects on the reaction shown in Scheme 1; the counterion in all cases is *bis*(trifluoromethanesulfonyl)imide ($[N(SO_2CF_3)_2]^-$). The cations are numbered by group – **I** = imidazolium, **BI** = *bis*(imidazolium), **P** = pyridinium, **A** = ammonium, **S** = sulfonium and **T** = triazolium. Other nomenclature used throughout is based upon common nomenclature for the 1-alkyl-3-methylimidazolium series; members of this series are frequently abbreviated $[xmim]^+$, where *x* is the first letter of the name of the alkyl chain; a second letter is used when needed to distinguish systems. Abbreviations for other systems are determined in the same way.

Results and Discussion

Initial considerations

The rate constant data for the reaction shown in Scheme 1 at 295 K in mixtures containing the ionic liquids having the cations shown in Figure 1 are summarised in Table 1; these cations have been organised by structure to allow the effect of changes to structural features to be readily seen. Two key features are the importance of the extent of charge localisation as demonstrated, for example, through comparison of the effects of the $[Fh_xpy]^+$ **P3** and $[h_xpy]^+$ **P2** salts along with other cases such as $[bmmo]^+$ **A3** *cf.* $[bmpi]^+$ **A2** and $[b45Cl_2mim]^+$ **I7** *cf.* $[bm_3im]^+$ **I4**) and accessibility of the charge ($[btl]^+$ **S1**€ *cf.* $[bmpyr]^+$ **A1** *cf.* $[TOA]^+$ **A4**). Generally, no co-solvent is present for the data given; exceptions include the pyridinium systems (the $[Fh_xpy]^+$ **P3** and $[h_xpy]^+$ **P2** salts) where solubility issues limited the maximum amount of ionic liquid that could be added.¢

It should be noted that, in all cases, the anion is *bis*(trifluoromethanesulfonyl)imide ($[N(SO_2CF_3)_2]^-$); the solvent effects on the reaction in Scheme 1 are known for a wide range of ionic liquids based on this anion, at least in part because it is it is relatively non-coordinating (as measured by, for example, the Kamlet-Taft β parameter[70]) and hence the cation is relatively 'available', maximising the key interaction responsible for rate constant enhancement. Use of another anion might result in greater anion-cation interactions, reducing the observed solvent effects, along with introducing effects beyond simple ion-pairing (as has been observed previously[45]).

For the computational assessment of the relationship of molecular properties and reaction kinetics, all of the cations shown in Figure 1 (with rate constant data for the corresponding ionic liquids shown in Table 1) were optimised with the semiempirical Hamiltonian AM1[55] and the local molecular properties calculated. Besides the molecular electrostatic potential (MEP; for example see Figure 2a), for complete data see Table S1 in ESI),[56] the local ionization potential,[57] the local electron affinity (E_{AL}),[58] the local hardness,[58] local electronegativity,[58] and dipolar density[59] have also been calculated, in addition to global properties such as polarizability,[60] dipole moment, and globularity.[61] From these properties, characteristic descriptors have been extracted on

Table 1. The bimolecular rate constants for the reaction shown in Scheme 1 at 295 K in reaction mixtures containing the specified ionic liquid at the mole fraction specified in acetonitrile. Uncertainties are reported as the standard deviation of at least three replicate experiments.

Cation	Ionic liquid	χ_{SALT}	k_2 / 10^{-4} L mol^{-1} s^{-1}
I1	[bmim][$N(SO_2CF_3)_2$][46]	0.86	14.5 ± 0.5[a]
I2	[bm_2im][$N(SO_2CF_3)_2$][44]	0.88	14.9 ± 1.3[a]
I3	[bm_3im][$N(SO_2CF_3)_2$][44]	0.85	6.70 ± 0.62
I4	[bm_4im][$N(SO_2CF_3)_2$][44]	0.85	5.4 ± 1.3[b]
I5	[b2Clmim][$N(SO_2CF_3)_2$][47]	0.86	13.1 ± 0.7
I6	[b4Clmim][$N(SO_2CF_3)_2$][47]	0.86	11.9 ± 1.2
I7	[b45Cl_2mim][$N(SO_2CF_3)_2$][47]	0.85	17.0 ± 0.3
BI1	[$(mim)_2p_e$][$N(SO_2CF_3)_2]_2$[41]	0.78	22.0 ± 1.1
P1	[bpy][$N(SO_2CF_3)_2$][41]	0.85	18.3 ± 1.0
P2	[$h_x$py][$N(SO_2CF_3)_2$][41]	0.36¢	13.4 ± 0.3
P3	[Fh_xpy][$N(SO_2CF_3)_2$][41]	0.36¢	26.3 ± 0.7
A1	[bmpyr][$N(SO_2CF_3)_2$][44]	0.86	17 ± 1[a]
A2	[bmpi][$N(SO_2CF_3)_2$][41]	0.85	15.1 ± 0.2
A3	[bmmo][$N(SO_2CF_3)_2$][41]	0.70	32.8 ± 3.2
A4	[TOA][$N(SO_2CF_3)_2$][44]	0.74	2.9 ± 0.4[b]
S1	[btl][$N(SO_2CF_3)_2$][41]	0.86	38 ± 10
T1	[bmtr][$N(SO_2CF_3)_2$][41]	0.86	17.4 ± 1.6

[a]Data interpolated from an Eyring analysis. Uncertainties are compounded from uncertainties in activation parameters.
[b]Data extrapolated from an Eyring analysis. Uncertainties are compounded from uncertainties in activation parameters.

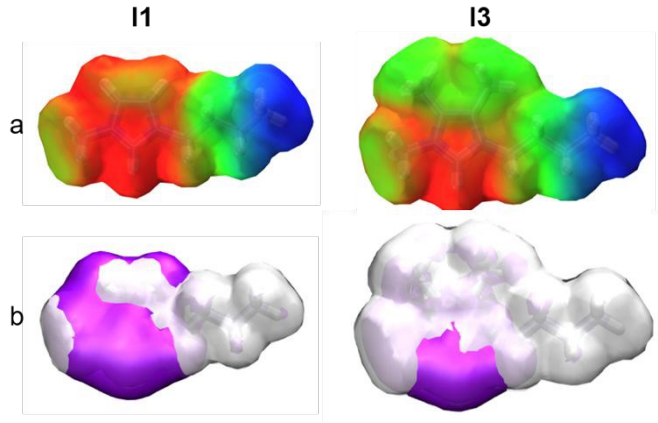


Figure 2. Representative Molecular Electrostatic potential (MEP) plots for compounds **I1** and **I3**. a) MEP (colour range blue-green-red, +50 to +110 kcal mol^{-1}) mapped onto the molecular electronic isodensity surface (isovalue 0.02 *e* Å^{-3}); b) MEP isosurface for values >100 kcal mol^{-1} represented superimposed with the electronic isodensity surface (white) indicating the area representing the MEParea100 descriptor.

the molecular surface as defined in the program Parasurf.[63] A natural atomic orbital point charge (NAO-PC) model was obtained[64] that describes the electrostatics of non-hydrogen atoms (valence-only *s*- and *p*-basis set) using the semi-empirical Hamiltonian AM1. These data include maximum, minimum, average and range values of the selected properties. (For a link to the full set of calculated descriptors, see the ESI page S13.)

Based on the relatively small experimental sample size, we subsequently used the whole data set of the properties for the 17 cations to train regression models using the available local properties. In the first instance, a regression analysis was performed to find non-linear correlations between those experimentally-obtained rate constant data and the molecular descriptors, and deriving F-test and Mutual Information (MI)-test values, which show linear dependency and non-linear dependency, respectively (see Figure S3 and Table S3). The cross correlation of the data, perhaps unsurprisingly, showed higher correlation within the groups of MEP and E_{AL} variables, but also correlations between E_{AL} and MEP descriptors and between dipolar density (dipden) and MEP range.

Simple linear regression models

Following our hypothesis that the electrostatic interactions between the substrate and the cations of the ionic liquid were of significant importance, we trained polynomial and linear regression models on single or combinations of the available MEP data on all available 17 cations. Not surprisingly, polynomial fits result in good R^2 values, however, showed clear signs of overfitting.

In order to prevent similar overfitting for linear regression models, the number of chosen descriptors were limited to three. Initial models trained showed reasonable accuracy with one characteristic outlier $[btl]^+$ (**S1**). Along with [btl] being a notable outlier from the model, it was the only IL that broke down under the experimental conditions.[41] As such, the experimental data for the reaction in this ionic liquid needs to be treated with caution. A first linear regression model excluding **S1** for the training (model *A*£) resulted in a good fit (R^2 = 0.87; MAE = 2.19 × 10^{-4} L mol^{-1} s^{-1}, Figure 3), verifying the importance of the electrostatic interactions between the cation of the ionic liquid and the pyridine **2**, which lead to the previously observed entropically driven rate constant enhancement.[46] (See Table S4, ESI for all models tested and compared initially.)

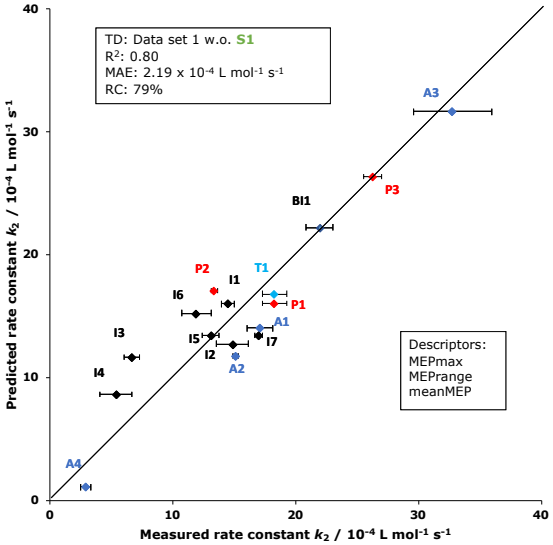


Figure 3. Correlation of linear regression model *A* with experimentally observed kinetic data, using the descriptors MEPmax, MEPrange and meanMEP. TD: Training data; MAE: Mean absolute error; RC: Rank correlation. Line shown is y = x, added as a guide only.

View Article Online
DOI: 10.1039/D0CP04224B

This model demonstrates a very good correlation, however it is one that needs to be treated with care. The different MEP descriptors used for the model are not independent (they have a high correlation with each other). Therefore, it is questionable how good the predictive power of the model is outside the examples used to train the model. Thus, we investigated the possibility of defining a new MEP descriptor that might be used to predict rate constants for this reaction more effectively when combined with other independent descriptors. Keeping in mind that the accessible charged area of the ionic liquid will be important for interactions with the pyridine **2**, we defined MEPsum*x*, MEPcount*x*, MEParea*x* (see Figure 2b)), MEPareasum*x*, and MEPtop25 (Table 2) as new MEP-based descriptors. These values were chosen as they all indirectly quantify a molecular surface area above a certain MEP value in different ways. Thus, they are linked to the area of highest electrostatic attraction to anions or nucleophiles. Different cut-off values for the new MEP-derived values have been tested for best correlation with the experimental kinetic data (see ESI, Table S5). The new MEP descriptors have then been combined with other descriptors (based on strong correlations between descriptors found in the MI-test and F-test) for further regression models (see ESI, Table S6 for full list of tested combinations). Best correlations for combinations were achieved using MEPtotal100 in combination with either the dipole density (dipden), an E_{AL}-derived descriptor, or an electronegativity-derived descriptor. The new linear regression models showed that besides ion **S1**, other singly-represented cation types in the data set were difficult to predict accurately. In particular, the bisimidazolium ion **BI1** shows very different characteristics in the new descriptors. Thus, final models have been trained without these ions, with the best model (model *D*) using MEPtotal100, EALbar, andENEGrange as descriptors (R^2 = 0.88, MAE = 1.45 × 10^{-4} L mol^{-1} s^{-1}, Figure 4; see ESI Tables S7 and S8 and Figures S4 and S5 for analysis of remaining models).

Table 2. Definition of MEP-based parameters introduced in this work.

Parameter	Description
MEPsum*x*	Sum of MEP values above a certain threshold *x*
MEPcount*x*	Total number of MEP values above a certain threshold *x*
MEParea*x*	Sum of corresponding areas of MEP values above a certain threshold *x*
MEPareasum*x*	MEP values above a certain threshold *x* and their corresponding areas are summed
MEPtop25	Sum of highest 25 MEP values

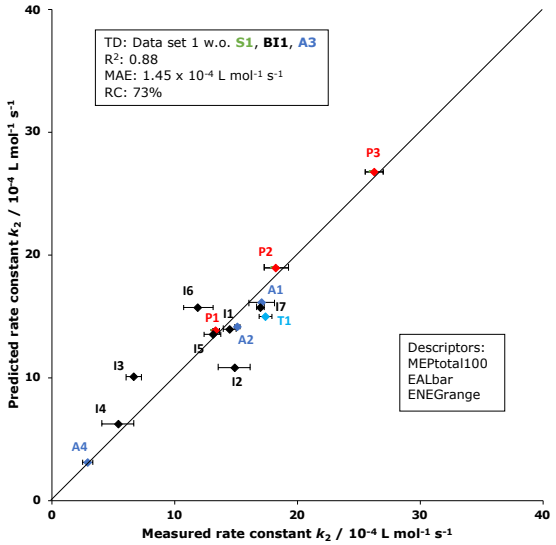


Figure 4. Correlation of linear regression model *D* with experimentally observed kinetic data, using the descriptors MEPtotal100, EALbar and ENEGrange. TD: Training data; MAE: Mean absolute error; RC: Rank correlation. Line shown is y = x, added as a guide only.

Expanding the data set and revision of the models

The above data was compiled from literature available at the start of this project on the solvent effects of ionic liquids on the reaction between benzyl bromide **1** and pyridine **2**. Whilst there was rational choice behind the structures of the solvents considered (particularly in the work by Hawker *et al*.[41]), ionic liquids containing cations with structures unlike other series were present. Hence, a set of ionic liquids based on the cations shown in Figure 6 were specifically considered and their effects on the reaction between benzyl bromide **1** and pyridine **2** were analysed. This set included two homologous series of ionic liquids; one with cations involving 1-alkyl-3-methylimidazolium salts (building on the [bmim][$N(SO_2CF_3)_2$] case introduced above) and another with doubly charged *bis*(imidazolium) cations (building on the [$(mim)_2p_e$][$N(SO_2CF_3)_2]_2$) case). Further, in the previous data set there was only one ionic liquid with a cation incorporating a fluorous side chain and only one ionic liquid with a non-cyclic cation; an additional member of each series has been included. In addition, it was of interest to incorporate different functionalities in the side chains of the ionic liquid, hence the introduction of the ether series. Finally, data on alternate types of cations that became available – particularly the lithium glyme solvate-based ionic liquids – provides a test for the versatility of the models determined. The rate constant data for the reaction shown in Scheme 1 at 295 K in mixtures containing ionic liquids shown in Figure 5 are summarised in Table 3. These data are, as previously, grouped by cation type.

Immediately apparent is the trend in the new data for the imidazolium series **I**; increasing the alkyl chain length on the cation decreases the rate constant for the process. These data are consistent with decreased interactions of the substrate **2** with the cation of the ionic liquid and will be discussed further elsewhere.[71] No equivalent trend is observed for the series of ionic liquids based on two linked imidazolium centres **BI**. The difference between the series suggests that the linking alkyl chain adds some restriction to the interactions involving the pyridine **2**; this is discussed in a separate report.[72] The additions to existing series provided data that was not unexpected; the greater accessibility of charge in the case of the ionic liquid based on cation **A5** results in a slight increase in rate constant compared to the **A4** case. Likewise, inductive removal of electron density increases charge density on the cation, thus

View Article Online
DOI: 10.1039/D0CP04224B

Table 3. The bimolecular rate constants for the reaction shown in Scheme 1 at 295 K in reaction mixtures containing the specified ionic liquid at the mole fraction specified. Uncertainties are reported as the standard deviation of at least three replicate experiments.

Cation	Ionic liquid	χ_{SALT}	k_2 / 10^{-4} L mol^{-1} s^{-1}
I8	[emim][$N(SO_2CF_3)_2$]	0.81	17.9 ± 1.1
I9	[$h_x$mim][$N(SO_2CF_3)_2$]	0.79	12.1 ± 1.0
I10	[omim][$N(SO_2CF_3)_2$]	0.80	10.5 ± 0.4
I11	[dmim][$N(SO_2CF_3)_2$]	0.80	9.40 ± 0.11
I12	[ddmim][$N(SO_2CF_3)_2$]	0.80	8.3 ± 0.6
I13	[F$h_x$mim][$N(SO_2CF_3)_2$]$_2$	0.80	38 ± 3
BI2	[$(mim)_2p_r$][$N(SO_2CF_3)_2$]$_2$	0.43$	35.4 ± 1.9
BI3	[$(mim)_2$b][$N(SO_2CF_3)_2$]$_2$	0.35$	7.4 ± 0.3
BI4	[$(mim)_2h_x$][$N(SO_2CF_3)_2$]$_2$	0.76	19.6 ± 1.9
BI5	[$(mim)_2h_p$][$N(SO_2CF_3)_2$]$_2$	0.22$	21 ± 4
BI6	[$(mim)_2$o][$N(SO_2CF_3)_2$]$_2$	0.74	15.3 ± 1.1
BI7	[$(mim)_2$n][$N(SO_2CF_3)_2$]$_2$	0.75	20.8 ± 2.2
BI8	[$(mim)_2$d][$N(SO_2CF_3)_2$]$_2$	0.74	19.6 ± 1.9
BI9	[$(mim)_2$u][$N(SO_2CF_3)_2$]$_2$	0.74	16.3 ± 1.9
BI10	[$(mim)_2$dd][$N(SO_2CF_3)_2$]$_2$	0.72	16.2 ± 0.9
A5	[mTOA][$N(SO_2CF_3)_2$]$_2$	0.76	4.4 ± 0.7
E1	[(eOm)mim][$N(SO_2CF_3)_2$]$_2$	0.81	21.4 ± 1.1
E2	[(eOm)py][$N(SO_2CF_3)_2$]$_2$	0.82	21.3 ± 0.2
E3	[(eOm)mmo][$N(SO_2CF_3)_2$]$_2$	0.82	23.1 ± 1.1
G1	[Li(G3)][$N(SO_2CF_3)_2$]	0.83	0.36 ± 0.01[73]
G2	[Li(G4)][$N(SO_2CF_3)_2$]	0.82	0.51 ± 0.11[73]

resulting in a rate constant increase on moving from the ionic liquid based on the non-fluorinated anion **I9** to that based on the fluorinated variant **I13**. The two additional series based on ether substituents **E** and glyme coordination compounds **G**, gave very similar data within these series. The values for the former series were comparable with data for the other ionic liquids considered, whilst the glyme-based systems gave notably lower rate constants; the origin of such was found to be a greater enthalpic cost offsetting the previously observed entropic benefit.[73]

Model validation and new regression models

The new experimental results have been taken as test data for the first set of linear regression models. All of the cations in Figure 2 were optimised as described above, with the same series of parameters determined (see Table S2). As predicted, the first model only based on MEP descriptors predicts the new rate constant data poorly with, particularly, data associated with additional classes of molecular ions being inadequately predicted. Such poor prediction on incorporation of additional (physicochemical) features is well-known.[74, 75] The other models generated also did not predict the new rate constant data well (see Tables S7 and S8, ESI). All of these examples demonstrate how narrow the 'comfort zone' of such simple predictive models are, when adding different chemical functionalities.

Subsequently, the models were retrained on the complete data set and a six-fold random cross validation has been applied to test their predictability. This process revealed two important observations. Firstly, model *D* was the best performing (R^2 = 0.82, rank correlation of 88%, MAE = 2.88 x 10^{-4} L mol^{-1} s^{-1}, Figure 6). Its predictability is also of good quality with an R^2 of 0.69 for the cross validation. Secondly, all models showed poor predictability when bisimidazolium-based cations (the **BI** series) were included in the training set. This finding suggests that the interactions involving these cations might not be easily predictable based on parameters derived from singly charged cations; this point is elucidated further elsewhere.[72] Particularly, this work indicates that the model breaks down when effects beyond the electrostatics modelled by the parameters introduced here become important in determining the interaction between pyridine **2** and the cation of the ionic liquid, and hence the observed solvent effects.

Based on the new data, we have also evaluated new combinations of predictors that might work better for the chemically more diverse data set and additionally tested simple forward propagation artificial neural networks (ANNs), including cross-validations for all models. As summarized in the ESI (Table S7 and S8, Figures S6-S14), none of the new models was significantly superior to model *D* (Figure 6a). However, a neural network model based on the descriptors of model *D*, referred to here as model *D*', led to even better predictability (Figure 6b). Without the **BI** series, a R^2 of 0.89 was achieved, and even when including the bisimidazolium ions the R^2 remains at a high value of 0.82 together with a rank correlation of 90%. However, when examining the data closely it is apparent that the **BI** series is less correlated than the other series (for this series alone: R^2 = 0.47, rank correlation = 28%, Figure 6c); this is in contrast to the correlation for the **I** series which improved notably on going from model *D* to *D*' (from R^2 = 0.65, rank correlation = 80%, Figure 4, to R^2 = 0.71, rank correlation = 90%, Figure 6b). It should be noted that this specific neural network model was also more thoroughly validated against signs of overfitting (see Table S7, ESI) and that other models with a higher number of predictors very quickly showed overfitting.

It is worth considering why the ionic liquids based on the bisimidazolium cations (the **BI** series) are not modelled well. Whilst the values of the local properties (particularly those that are MEP related) are outside the range of the other ions, it is likely that the linked nature of the di-cation systems is important. Particularly, the results would be that interactions with the imidazolium centres are unlikely to be independent and hence features not incorporated in models here (such as the conformation about the linker) need to be considered. This argument is supported by recent analyses of temperature dependent kinetic data in these solvent systems.[72]

Conclusions

In the first instance, the work described here has substantially expanded the range of ionic liquids that have been considered for the impact of solvent effects on the rate constant of a representative S_N2 organic process. Importantly, these data provide a greater understanding of the effects of systematic changes to the cation structure, along with how the incorporation of new functionalities impacts on these solvent effects. Further, this expanded data set can be further applied to the development of models for predicting these effects.

View Article Online
DOI: 10.1039/D0CP04224B

Journal Name

ARTICLE

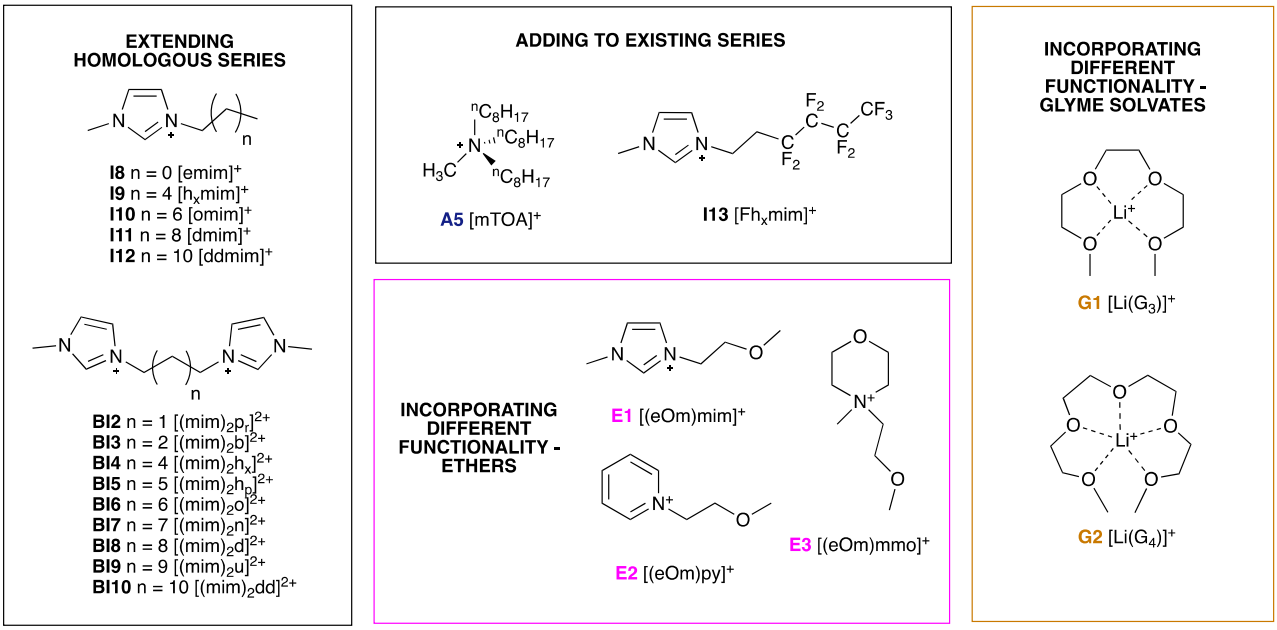


Figure 5. The cations in the ionic liquids considered in the second portion of this work, based on their effects on the reaction shown in Scheme 1; the counterion in all cases is *bis*(trifluoromethanesulfonyl)imide ($[N(SO_2CF_3)_2]^-$). The cations are numbered by group and are continued from Figure 1; different groups included here are – E = ethers and G = glyme. Once again, other nomenclature is based on the $[xmim]^+$ series, noting that standard abbreviations for the glyme solvates are also used. The representations of the glyme are what is commonly used though it should be noted that they can also exist either as complexes where multiples glymes complex to multiple lithium centres or as larger clusters.[76]

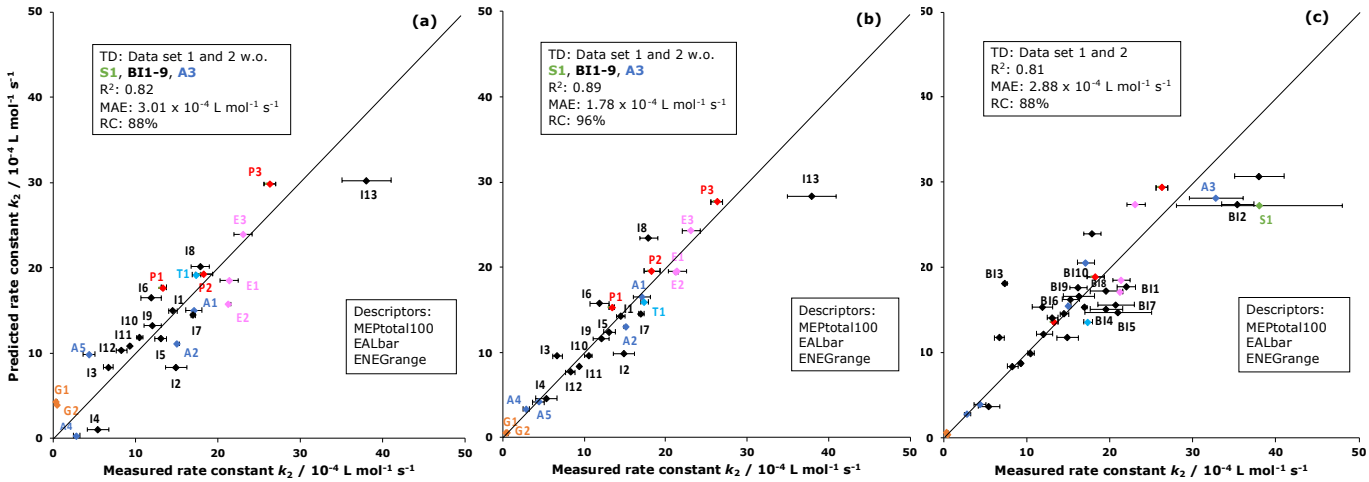


Figure 6. Correlation of the linear regression model *D* (a) and the artificial neural network model *D'* (b, c) with experimentally observed kinetic data, with data sets and descriptors as listed. TD: Training data; MAE: Mean absolute error; RC: Rank correlation. Lines shown are y = x, added as a guide only. For clarity, only ions not present in (b) are listed in (c)

Simple regression models have been critically investigated and assessed as predictors of solvent effects and shown to deliver very promising results. The initial qualitative assumption that the accessibility of high charge density on the cation of the ionic liquid critically influences the rate constant of the reaction, shown in Scheme 1, holds when evaluated quantitatively for this initial data set. Local molecular properties that include this and other information influencing intermolecular interactions

can successfully be used to train regression models, predicting the reaction kinetics in different ionic liquids.

Perhaps unsurprisingly, the linear regression models fail to predict rate constant data well for systems outside their 'prediction comfort zone.' The relatively small data set available at the beginning of this study was insufficient to train a robust predictive model. Further, the dataset design needs to accommodate a wide chemical variability, which we tried to achieve by targeted expansions to the initial data set and we contend should be considered explicitly in future studies, noting that care should be taken to ensure the microscopic origins of the rate constant enhancement do not change with the expansion of the data set. The expanded data set used here was found to be amenable to use of an artificial neural network, and retained excellent predictability even with the incorporation of series that were not modelled well by linear regression methods.

This study also shows again how easy predictive regression models can tend to overfitting. Careful cross validation is therefore a vital inclusion to any process. It also demonstrated again the difficulty when dealing with an imbalanced data set with few data points to predict but many predictors. In order to reduce the tendency of overfitting a few predictors need to be selected for training the regression model. In this study three predictors were ideal for linear regression models that tended much less to overfitting.

It is worth finishing with a consideration of what these models allow. Consider the final example of the artificial neural network: it allows the rate constant for the process shown in Scheme 1 to be determined with a mean difference of < 10% between the predictions from the model and experimental data, allowing the effect of a given ionic liquid to be effectively predicted. Further, the ranking efficacy is such that whether one ionic liquid might result in a greater rate constant than another can be predicted with 90% accuracy. This outcome is important as it allows a 'choice' of ionic liquid to begin to be made with rational selection, based on computable parameters for specific reactions. The clear opportunity to apply this to other reactions and systems, and compare models to uncover fundamentals about the underlying processes is tantalising, especially combined with new opportunities in high-throughput ionic liquid chemistry.[77]

Methodology

Computational methods

All cations in this study were geometry optimized using the semiempirical Hamiltonian AM1 with the program EMPIRE (lithium ion-containing structures used PM3).[78-80] The local molecular properties were then calculated using the program Parasurf[63] and the Marching-cube algorithm to generate surfaces with the local molecular properties.[81] Calculated descriptors included MEP (molecular electrostatic potential), IEL (local ionisation energy), EAL (local electron affinity) and polarisability (aL). The newly defined MEP-based descriptors (MEPtop25, MEPcount, MEPsum, MEParea and MEPareasum) were generated using in-house written python scripts.

For initial correlation analysis of the descriptors, Python Packages numpy, pandas, matplotlib and seaborn were used to define a correlation matrix based on all descriptors (see Figure S3, ESI). Python packages were also used to perform the F-test and mutual information MI-test. Selected descriptors are shown in Table S3. For the F-test the correlation between each regressor and the target is computed as ((X[:, i] - mean(X[:, i])) * (y - mean_y)) / (std(X[:, i]) * std(y)), where X is defined as the set of regressors that will be tested sequentially [{array-like, sparse matrix} shape = (n_samples, n_features)] and Y as the data matrix [array of shape(n_samples)]. The function used for the mutual information test relies on nonparametric methods based on entropy estimation from k-nearest neighbors distances as previously described.[82, 83]

KNIME (KoNstanz Information MinEr)[84] was used for all linear, polynomial, and artificial neural network (ANN)-based regression models. To prevent overfitting based on the imbalance between the number of descriptors and targets, the number of descriptors was limited to 3-5 and selected based on the preceding correlation analysis. Polynomial regression models were disregarded after demonstrating strong overfitting tendencies. Linear regression models were trained on the whole first data set without explicit cross validation due to the small number of target values. When data was trained including the second data set, a 6-fold random cross-validation was applied to all models.

Artificial neural network (ANN)-based models were trained using a multilayer feedforward neural network applying the RProp algorithm as implemented in KNIME. Different combinations of the number of hidden layers (between 1 and 3), number of hidden neurons (4 to 20) and number of earning iterations (50 to 1000) have been tested and optimized along 6-fold random cross validation to balance prediction and fitness of prediction of the model. This optimization has been carried out for model D' and has been repeated with narrower parameter variation for the other models.

Experimental methods

Benzyl bromide **1** and pyridine **2** were distilled[85] and stored over activated molecular sieves (3 Å) at 4 °C prior to use. Acetonitrile distilled[85] and stored over activated molecular sieves (3 Å) at room temperature under nitrogen. All other chemicals were purified through literature methods[85] and used immediately. The ionic liquids used were generally prepared through the *N*-alkylation of the appropriate precursor, followed by anion metathesis (complete synthetic procedures are given in the ESI).

All kinetic measurements were obtained by monitoring reaction progress using ^{1}H NMR spectroscopy with either a Bruker Avance III 400, 500 or 600 NMR spectrometer equipped with either a BBO, BBFO or TBI probe; results were shown to be reproducible regardless of either the spectrometer or the probe used. The temperature of the NMR spectrometer was calibrated using a thermocouple containing ethanol.

View Article Online DOI: 10.1039/D0CP04224B

The reaction mixtures for the kinetic studies were prepared under pseudo first order conditions such that they contained at least a 10-fold excess of pyridine **2** relative to benzyl bromide **1**. Reaction progress was monitored to >95% completion using ^{1}H NMR spectroscopy, following depletion of the signal corresponding to the benzylic protons of benzyl bromide **1** at approximately 4.0 ppm. The exception was the kinetic data determined with the ionic liquid [Fh_xmim][$N(SO_2CF_3)_2$] where reaction progress was monitored (again to >95% completion) using ^{1}H NMR spectroscopy following formation of the signal corresponding to the benzylic protons of the product **3** at approximately 5.6 ppm.

Acknowledgements

JBH and his co-workers at UNSW acknowledge financial support from the Australian Research Council Discovery Project Funding Scheme (Projects DP130102331 and DP180103682). KTCL and AG also acknowledge the Australian Government for the award of Research Training Program Scholarships. The support of the Mark Wainwright Analytical Centre and the UNSW NMR facility are gratefully acknowledged. AS acknowledges financial support from the BBSRC/EPSRC Synthetic Biology Research Centre – Nottingham (BB/L013940/1) and the University of Nottingham. CMJ acknowledges FJJ, BAJ, and MVJ for making the write-up of this manuscript under lockdown conditions an especially eventful endeavour. Without EKJ this work would be in pieces. CMJ acknowledges the UoN collaboration fund for travel support to deepen this collaboration with JBH.

Conflicts of interest

The authors declare no conflicts of interest.

Notes and references

§ Subsequent analysis incorporating rate constant data determined for ionic liquids based on cations shown in Figure 6 expanded the data set from seven to eleven, however the same parameters were shown to give the most significant correlation with effectively the same relative weighting. For details, see ESI.

‡ It is noted that the computational prediction of Kamlet-Taft parameters has been reported.[86]

€ The sulfonium ion **S1** has been reported to breakdown slowly under the conditions of the reaction shown in Scheme 1, however, this breakdown was sufficiently slow that a rate constant for the reaction between reagents **1** and **2** could still be readily determined.[41]

¢ The co-solvent in these cases is acetonitrile. Given the demonstrated dependence of the rate constant on the proportion of ionic liquid in the reaction mixture for the reaction shown in Scheme 1,[41, 43] it would be expected that the rate constant would not vary significantly if further ionic liquid were added in these cases. Hence, all of the data listed is used in the subsequent analysis.

£ The models developed are labelled model *A*, *B*, *C*, *etc.* Note that not all models are discussed explicitly in the manuscript; for full details of each model and assessment of their efficacy, see ESI, particularly Figures S4-S14.

$ As per footnote ¢, the co-solvent is acetonitrile and it is not anticipated that the effect of adding further ionic liquid would be significant.

1. J. P. Hallett and T. Welton, *Chem. Rev.*, 2011, **111**, 3508-3576.
2. A. Gilbert, R. S. Haines and J. B. Harper, in *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, ed. J. Reedjik, Elsevier, Waltham, MA, 2018, DOI: 10.1016/B978-0-12-409547-2.14212-X, pp. doi:10.1016/B1978-1010-1012-409547-409542.414212-X.
3. S. T. Keaveney, R. S. Haines and J. B. Harper, *Pure Appl. Chem.*, 2017, **89**, 745-757.
4. S. T. Keaveney, R. S. Haines and J. B. Harper, in *Encyclopedia of Physical Organic Chemistry*, ed. U. Wille, Wiley, 2017, vol. 2, ch. 27.
5. D. R. MacFarlane, M. Kar and J. M. Pringle, *Fundamentals of Ionic Liquids*, Wiley-VCH, Weinheim, Germany, 2017.
6. R. R. Hawker and J. B. Harper, *Adv. Phys. Org. Chem.*, 2018, **52**, 49-85.
7. D. R. MacFarlane, A. L. Chong, M. Forsyth, M. Kar, R. Vijayaraghavan, A. Somers and J. M. Pringle, *Faraday Discuss.*, 2018, **206**, 9-28.
8. C. Reichardt and T. Welton, in *Solvents and Solvent Effects in Organic Chemistry*, Wiley, Weinheim, Germany, 2010, ch. 8, pp. 509-548.
9. H. M. Yau, S. T. Keaveney, B. J. Butler, E. E. L. Tanner, M. S. Guerry, S. R. D. George, M. H. Dunn, A. K. Croft and J. B. Harper, *Pure Appl. Chem.*, 2013, **85**, 1979-1990.
10. N. L. Lancaster, T. Welton and G. B. Young, *J. Chem. Soc., Perkin Trans. 2*, 2001, 2267-2270.
11. N. L. Lancaster, P. A. Salter, T. Welton and G. B. Young, *J. Org. Chem.*, 2002, **67**, 8855-8861.
12. N. L. Lancaster and T. Welton, *J. Org. Chem.*, 2004, **69**, 5986-5992.
13. L. Crowhurst, N. L. Lancaster, J. M. Pérez Arlandis and T. Welton, *J. Am. Chem. Soc.*, 2004, **126**, 11549-11555.
14. N. L. Lancaster, *J. Chem. Res.*, 2005, **2005**, 413-417.
15. B. Y. W. Man, J. M. Hook and J. B. Harper, *Tetrahedron Lett.*, 2005, **46**, 7641-7645.
16. H. M. Yau, S. A. Barnes, J. M. Hook, T. G. A. Youngs, A. K. Croft and J. B. Harper, *Chem. Commun.*, 2008, 3576-3578.
17. S. T. Keaveney, K. S. Schaffarczyk McHale, J. W. Stranger, B. Ganbold, W. S. Price and J. B. Harper, *ChemPhysChem*, 2016, **17**, 3853-3862.
18. K. S. Schaffarczyk McHale, R. S. Haines and J. B. Harper, *ChemPlusChem*, 2018, **83**, 1162-1168.
19. A. Gilbert, R. S. Haines and J. B. Harper, *Org. Biomol. Chem.*, 2019, **17**, 675-682.
20. A. Gilbert, G. Bucher, R. S. Haines and J. B. Harper, *Org. Biomol. Chem.*, 2019, **17**, 9336-9342.
21. M. Gazitua, R. A. Tapia, R. Contreras and P. R. Campodonico, *New J. Chem.*, 2014, **38**, 2611-2618.
22. J. Alarcón-Espósito, R. Contreras, R. A. Tapia and P. R. Campodónico, *Chem. Eur. J.*, 2016, **22**, 13347-13351.
23. F. D'Anna, V. Frenna, R. Noto, V. Pace and D. Spinelli, *J. Org. Chem.*, 2006, **71**, 5144-5150.
24. R. R. Hawker, M. J. Wong, R. S. Haines and J. B. Harper, *Org. Biomol. Chem.*, 2017, **15**, 6433-6440.
25. R. R. Hawker, R. S. Haines and J. B. Harper, *Org. Biomol. Chem.*, 2018, **16**, 3453-3463.
26. R. R. Hawker, R. S. Haines and J. B. Harper, *J. Phys. Org. Chem.*, 2018, **31**, e3862.
27. I. Newington, J. M. Perez-Arlandis and T. Welton, *Org. Lett.*, 2007, **9**, 5247-5250.
28. M. Gazitúa, R. A. Tapia, R. Contreras and P. R. Campodónico, *New J. Chem.*, 2018, **42**, 260-264.
29. S. G. Jones, H. M. Yau, E. Davies, J. M. Hook, T. G. A. Youngs, J. B. Harper and A. K. Croft, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1873-1878.
30. K. S. Schaffarczyk McHale, R. S. Haines and J. B. Harper, *ChemPlusChem*, 2019, **84**, 465-473.
31. S. T. Keaveney, K. S. Schaffarczyk McHale, R. S. Haines and J. B. Harper, *Org. Biomol. Chem.*, 2014, **12**, 7092-7099.
32. S. T. Keaveney, R. S. Haines and J. B. Harper, *Org. Biomol. Chem.*, 2015, **13**, 3771-3780.
33. S. T. Keaveney, J. B. Harper and A. K. Croft, *ChemPhysChem*, 2018, **19**, 3279-3287.
34. T. Fischer, A. Sethi, T. Welton and J. Woolf, *Tetrahedron Lett.*, 1999, **40**, 793-796.
35. A. Aggarwal, N. L. Lancaster, A. R. Sethi and T. Welton, *Green Chem.*, 2002, **4**, 517-520.
36. A. Vidis, C. A. Ohlin, G. Laurenczy, E. Kusters, G. Sedelmeier and P. J. Dyson, *Adv. Synth. Cat.*, 2005, **347**, 266-274.
37. H. M. Yau, S. J. Chan, S. R. D. George, J. M. Hook, A. K. Croft and J. B. Harper, *Molecules*, 2009, **14**, 2521-2534.
38. S. R. D. George, G. L. Edwards and J. B. Harper, *Org. Biomol. Chem.*, 2010, **8**, 5354-5358.
39. N. D. Khupse and A. Kumar, *J. Phys. Chem. A*, 2011, **115**, 10211-10217.
40. S. T. Keaveney, R. S. Haines and J. B. Harper, *ChemPlusChem*, 2017, **82**, 449-457.
41. R. R. Hawker, R. S. Haines and J. B. Harper, *Chem. Commun.*, 2018, **54**, 2296-2299.
42. H. M. Yau, A. K. Croft and J. B. Harper, *Faraday Discuss.*, 2012, **154**, 365-371.
43. K. S. Schaffarczyk McHale, R. R. Hawker and J. B. Harper, *New J. Chem.*, 2016, **40**, 7437-7444.
44. E. E. L. Tanner, H. M. Yau, R. R. Hawker, A. K. Croft and J. B. Harper, *Org. Biomol. Chem.*, 2013, **11**, 6170-6175.
45. S. T. Keaveney, D. V. Francis, W. Cao, R. S. Haines and J. B. Harper, *Aust. J. Chem.*, 2015, **68**, 31-35.
46. H. M. Yau, A. G. Howe, J. M. Hook, A. K. Croft and J. B. Harper, *Org. Biomol. Chem.*, 2009, **7**, 3572-3575.
47. R. R. Hawker, J. Panchompoo, L. Aldous and J. B. Harper, *ChemPlusChem*, 2016, **81**, 574-583.
48. R. W. Taft and M. J. Kamlet, *J. Am. Chem. Soc.*, 1976, **98**, 2886-2894.
49. R. W. Taft and M. J. Kamlet, *J. Am. Chem. Soc.*, 1976, **98**, 377-383.
50. M. J. Kamlet, J. L. Abboud and R. W. Taft, *J. Am. Chem. Soc.*, 1977, **99**, 6027-6039.

View Article Online
DOI: 10.1039/D0CP04224B

51. R. Bini, C. Chiappe, V. L. Mestre, C. S. Pomelli and T. Welton, *Org. Biomol. Chem.*, 2008, **6**, 2522-2529.
52. A. Habibi-Yangjeh, Y. Jafari-Tarzanag and A. R. Banaei, *Int. J. Chem. Kin.*, 2009, **41**, 153-159.
53. P. Pavez, D. Millán, J. I. Morales, E. A. Castro, C. López and J. G. Santos, *J. Org. Chem.*, 2013, **78**, 9670-9676.
54. M. A. A. Rani, A. Brant, L. Crowhurst, A. Dolan, M. Lui, N. H. Hassan, J. P. Hallett, P. A. Hunt, H. Niedermeyer, M. Perez-Arlandis, M. Schrems, T. Welton and R. Wilding, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16831-16840.
55. S. Spange, R. Lungwitz and A. Schade, *J. Mol. Liq.*, 2014, **192**, 137-143.
56. P. Politzer and J. S. Murray, in *Reviews in Computational Chemistry*, eds. K. B. Lipkowitz and D. B. Boyde, Wiley, Hoboken, 1991, DOI: https://doi.org/10.1002/9780470125793.ch7, pp. 273-312.
57. B. Ehresmann, B. Martin, A. H. C. Horn and T. Clark, *J. Mol. Model.*, 2003, **9**, 342-347.
58. P. Gedeck, C. Kramer and P. Ertl, *Prog. Med. Chem.*, 2010, **49**, 113-160.
59. A. E. Kerdawy, S. Güssregen, H. Matter, M. Hennemann and T. Clark, *J. Chem. Inf. Model.*, 2013, **53**, 1486-1502.
60. T. Brinck and J. H. Stenlid, *Adv. Theory Simul.*, 2019, **2**, 1800149.
61. C. M. Jäger, T. Schmaltz, M. Novak, A. Khassanov, A. Vorobiev, M. Hennemann, A. Krause, H. Dietrich, D. Zahn, A. Hirsch, M. Halik and T. Clark, *J. Am. Chem. Soc.*, 2013, **135**, 4893-4900.
62. T. Bauer, C. M. Jäger, M. J. T. Jordan and T. Clark, *J. Chem. Phys.*, 2015, **143**, 044114.
63. X. Grabuleda, C. Jaime and P. A. Kollman, *J. Comp. Chem.*, 2000, **21**, 901-908.
64. C. A. Hunter, *Chem. Sci.*, 2013, **4**, 1687-1700.
65. M. H. Abraham and J. A. Platts, *J. Org. Chem.*, 2001, **66**, 3484-3491.
66. S. J. Pike and C. A. Hunter, *Org. Biomol. Chem.*, 2017, **15**, 9603-9610.
67. S. J. Pike, J. J. Hutchinson and C. A. Hunter, *J. Am. Chem. Soc.*, 2017, **139**, 6700-6706.
68. S. J. Pike, E. Lavagnini, L. M. Varley, J. L. Cook and C. A. Hunter, *Chem. Sci.*, 2019, **10**, 5943-5951.
69. S. Scheiner, *J. Comp. Chem.*, 2018, **39**, 500-510.
70. S. Spange, R. Lungwitz and A. Schade, *J. Mol. Liq.*, 2012, **192**, 137-143.
71. D. C. Morris, S. W. Prescott and J. B. Harper, *unpublished work*.
72. K. T.-C. Liu, R. S. Haines and J. B. Harper, *Org. Biomol. Chem.*, 2020, doi: 10.1039/D1030OB01500H
73. K. S. Schaffarczyk McHale, M. J. Wong, A. K. Evans, A. Gilbert, R. S. Haines and J. B. Harper, *Org. Biomol. Chem.*, 2019, **17**, 9243-9250.
74. M. Glavatskikh, J. Leguy, G. Hunault, T. Cauchy and B. D. Mota, *J. Cheminformatics*, 2019, **11**, 69.
75. A. K. Croft and M. K. Foley, *Org. Biomol. Chem.*, 2008, **6**, 1594-1600.
76. W. Shinoda, Y. Hatanaka, M. Hirakawa, S. Okazaki, S. Tsuzuki, K. Ueno and M. Watanabe, *J. Chem. Phys.*, 2018, **148**, 193809.
77. A. Zhu, L. Li, C. Zhang, Y. Shen, M. Tang, L. Bai, C. Du, S. Zhang and J. Wang, *Green Chem.*, 2019, **21**, 307-313.
78. M. Hennemann and T. Clark, *J. Mol. Model.*, 2014, **20**, 2331.
79. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**.
80. J. J. P. Stewart, *J. Comp. Chem.*, 1989, **1989**, 209-220.
81. W. Heiden, T. Goetze and J. Brickmann, *J. Comp. Chem.*, 1993, **14**, 246-250.
82. A. Kraskov, H. Stogbauer and P. Grassberger, *Phys. Rev. E*, 2004, **69**, 066138.
83. B. C. Ross, *PLoS ONE*, 2014, **9**, e87357.
84. M. R. Berthold, N. Cebron, F. Dill, T. R. Gabriel, T. Kötter, T. Meinl, P. Ohl, K. Thiel and B. Wiswedel, *SIGKDD Explor.*, 2009, **11**, 26-31.
85. W. L. F. Armarego and C. L. L. Chai, *Purification of Laboratory Chemicals*, Butterworth-Heinemann, Oxford, 2013.
86. H. Niedermeyer, C. Ashworth, A. Brandt, T. Welton and P. A. Hunt, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11566-11579.

View Article Online
DOI: 10.1039/D0CP04224B

An iterative, combined experimental and computational approach towards predicting reaction rate constants in ionic liquids is presented.