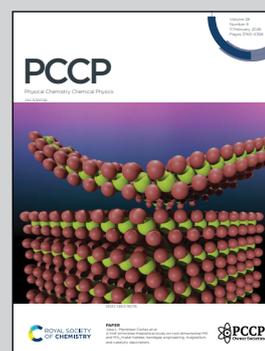**Showcasing research from Prof. Johansson's group at the Department of Physics, Chalmers University of Technology, Sweden.**

Ionic liquid conductivity models by symbolic regression

This work suggests a novel approach to model the ionic conductivity of ionic liquids (ILs), based on the machine learning (ML) method symbolic regression (SR) that finds analytical equations to describe data. Using molecular descriptors as model inputs, we find that combining SR with expressions derived from free volume theory gives symbolic models that agree well with experimental data, especially when considering more ideal ILs. The models we find are directly interpretable, which helps us understand the mechanism(s) governing ion transport in ILs better, showcasing the advantage of using SR over other ML techniques.

Image reproduced by permission of Isak Bengtsson from *Phys. Chem. Chem. Phys*., 2026, **28**, 3841.

As featured in:

See Isak Bengtsson and Patrik Johansson, *Phys. Chem. Chem. Phys*., 2026, **28**, 3841.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

Registered charity number: 207890

PCCP Owner Societies

Ionic liquid conductivity models by symbolic regression

Isak Bengtsson [ID] *[a] and Patrik Johansson [ID] [abc]

Organic solvents and fluorinated Li-salts is the basis of lithium-ion battery electrolytes, and it has remained unchanged for decades despite significant drawbacks such as thermal instability and high vapour pressure. One alternative is ionic liquid (IL) based electrolytes. However, the mechanism(s) that govern ion transport in IL based electrolytes, a property crucial for battery performance, is not yet fully understood. We here suggest a novel approach to model the ionic conductivity of ILs themselves; using symbolic regression (SR) to find analytical expressions derived from free volume theory (FVT). Using molecular descriptors as model inputs, we find several FVT-based models that show high correlations: $R^2 = 0.97$ and $R^2 = 0.94$ for the training and validation set, respectively, for an experimental dataset of 22 ILs measured in-house. Moving towards a significantly larger dataset, using data on 338 ILs from 125 publications, we find that our best model has a significantly higher spread in prediction accuracy but still shows appreciable performance for many ILs ($R^2 = 0.76$ and $R^2 = 0.73$ for the training and validation set, respectively). Overall, the FVT derived models perform best for "good" ILs, *i.e.* with well-dissociated ions, and worse for those ILs with strong ion–ion interactions. Using data from many publications impacts model performance, likely due to significant variations in *e.g.* impurities and dryness, as well as experimental set-ups and conditions.

1 Introduction

The discovery of new materials has always been a cornerstone of scientific and technological advancement. One of the most transformative technologies during the last decade, the lithium-ion battery (LIB), required a fundamental understanding of intercalation electrodes, electrolytes and electrode/electrolyte interphases before successful realisation.[1] Today it powers everything from mobile electronics to electric vehicles and is becoming increasingly important to efficiently make use of renewable energy. Improving and understanding the electrolyte remain one key challenge of LIBs, but the overall composition has in fact not changed much during the last two-three decades.[2] Yet, current LIB electrolytes, all relying on organic solvents, fluorinated Li-salts, and various tailored additives, suffer from thermal instability and high vapour pressures, amongst other drawbacks.[3–6]

One alternative is to increase the Li-salt concentration to *ca.* 3–5 M, *i.e.* highly concentrated electrolytes (HCEs), or to resort to ionic liquid (IL) based electrolytes.[7–9] These types of

electrolytes have shown to offer several advantages, such as improved high voltage performance, suppressed dendrite growth when using lithium metal anodes, and overall a higher level of safety.[10,11]

Ionic interactions will have strong influence on the electrolyte properties and the ion transport in these types of electrolytes. In conventional, *ca.* 1 M Li-salt in organic solvent based electrolytes, ion transport is mainly governed by a vehicular mechanism,[12] but in HCEs and IL-based electrolytes the solvation shells are less stable and not always even definable, why other conduction mechanisms come into play.[13–16] Looking at IL-based electrolytes in particular, the mechanism(s) that govern ion transport is not yet fully understood. One hypothesis is that ions migrate by jumping between voids.[17] This is supported by free volume theory (FVT), which has been used to describe transport phenomena in various glass formers, including ILs.[18–20]

According to FVT, voids appear in liquids due to thermal redistribution of a free volume v_f. Given a critical volume v_f^* and a factor γ, accounting for overlapping voids, molecules diffuse when a nearby void reaches a critical value γv_f^*. The diffusion coefficient of the molecules is related to the free volume with a temperature dependence according to

$$D \propto \sqrt{T} \exp\left(-\frac{\gamma v_f^*}{v_f}\right). \quad (1)$$

[a] Department of Physics, Chalmers University of Technology, SE-41296, Göteborg, Sweden. E-mail: isak.bengtsson@chalmers.se

[b] Department of Chemistry – Ångström, Uppsala University, SE-75237 Uppsala, Sweden

[c] ALISTORE-ERI, FR CNRS 3104, Hub de l'Energie, 80039 Amiens, France

This journal is © the Owner Societies 2026

Phys. Chem. Chem. Phys., 2026, **28**, 3841–3849 | **3841**

Combining this with the Nernst–Einstein equation for conductivity:

$$\sigma = \frac{nq^2 D}{k_B T} \qquad (2)$$

for a concentration n of charge carriers with charge q (k_B represents Boltzmann's constant), ionic conductivity can be related to the free volume according to:

$$\sigma \propto \frac{nq^2}{k_B \sqrt{T}} \exp\left(-\frac{\gamma v_f^*}{v_f}\right). \qquad (3)$$

This expression assigns an explicit temperature dependence to the ionic conductivity by the factor $1/\sqrt{T}$, but the free volume v_f will also be a function of temperature.

Experimentally and at a macroscopic level, the electrolyte is indeed commonly characterised mainly by its ionic conductivity, in many cases for the simplicity in contrast to any attempt to fully account of its ion transport properties. For ILs there are some rules of thumb that relate (ion) transport dynamics to the molecular structure of the ingoing ions, but with claims that over 10^{18} ILs exist it has proven difficult to find general relations.[21] Considering how crucial adequate ion transport is for battery performance it is of considerable interest to understand structure–property relations in ILs, as well as all electrolytes, better. Here modelling approaches become indispensable.

Upon this base, we here aim to build models that relates the ionic conductivity to the molecular structures of the IL ions, as a stepping stone towards understanding the ion transport better and eventually constructing similar models for IL-based LIB electrolytes, i.e. Li-salt doped ILs.[22–24] Previous modelling of ion transport of ILs can roughly be divided into two approaches; those that use molecular dynamics (MD) simulations and extract transport properties from the dynamics,[25–28] and those that construct predictive, often analytical, models directly from a set of descriptors.[29–32] Recently however, there has been a significant shift towards machine learning (ML) methods, with models based on support vector machines (SVMs), random forests (RFs) and many versions of neural networks (NNs).[33–36] While NN approaches often generalise well and offer very accurate property predictions, they are generally not constrained by any physical laws and require large amounts of data. When training data is scarce, NNs are prone to overfitting, they may learn the training examples extremely well, yet fail to generalize to unseen data. Modern NNs are typically over-parameterized, giving them sufficient capacity to capture not only the meaningful structure of the problem but also incidental fluctuations, noise, or rare coincidences present in the training set. This primarily degrades their ability to interpolate reliably within the domain spanned by the available data. In contrast, difficulties with extrapolation arise even for well-regularized models, simply because they are asked to predict outside the region represented in the training set. Overfitted models can therefore appear highly accurate during training while failing to capture the genuine physical trends required for robust predictive performance. Physics-Informed NNs (PINNs) address these problems by incorporating known physics in the models, which has shown to improve predictive

performance for out-of-distribution data.[37,38] However, the mapping between input features and predictions remain complex. A viable modelling alternative, and the one we apply here, is to use symbolic regression (SR), a supervised ML method which tries to find mathematical expressions that best explain data i.e. analytical models that are directly interpretable and furthermore can be constrained to fulfil dimensional requirements and physical laws.

2 Methods

Overall, three steps were applied: collecting and constructing datasets, computing molecular descriptors, and running the actual SR searches.

2.1 Collecting and constructing datasets

Two different datasets were considered, one with only 22 different ILs, all provided by Solvionic for the study by Nilsson-Hallen et al.,[39] and one with 338 ILs collected from the database ILThermo.[40,41] The first dataset consists of only 176 data points, whilst the second has 3933. For simplicity and w.r.t. end-use for battery application only aprotic ILs were considered. For both datasets, the measurement records were grouped per IL, using their IUPAC names as identifiers. To facilitate the computation of molecular descriptors, the IUPAC name was parsed to SMILES strings, one for each ion. For details, see Note 1 in the SI.

For the first dataset, all ILs were of the highest available purity (99.9% or 99.5%) and used as received. The dataset included cations from imidazolium, ammonium, pyrrolidinium and piperidinium families, combined with the anions $[BF_4]^-$, $[PF_6]^-$, $[Tf]^-$, $[FSI]^-$ and $[TFSI]^-$. For each IL, the ionic conductivity was measured in steps of 10 K from 298.15 to 368.15 K using dielectric spectroscopy, yielding 10^{-1}–10^1 mS cm^{-1}. For experimental details, see Nilsson-Hallen et al.[39]

The second dataset is more diverse, with data originating from 125 publications and ILs based on no less than 169 cations and 66 anions. Although a majority of the ILs were based on imidazolium and pyridinium cations, or other ring structures with nitrogen atom(s), the dataset also included ILs based on cations with different forms, such as phosphonium, cyclopropenium and sulfonium cations. For several ILs, there were more than one publication source; only the one with the lowest reported ionic conductivities was included, on the assumption that those correspond to the purest ILs. The ionic conductivities covered a much wider range, 10^{-5}–10^2 mS cm^{-1} at temperatures between 210 and 571 K.

2.2 Computing molecular descriptors

We used the open-source toolkit RDKit,[42] which given a SMILES string can compute many molecular descriptors and take different conformations into account. The conformer generation was done using what is considered the best-performing freely available conformer generator,[43] the ETKDG method, based on a distance geometry algorithm that also leverages experimental torsion angles and uses chemical constraints.[44]

3842 | Phys. Chem. Chem. Phys., 2026, 28, 3841–3849

This journal is © the Owner Societies 2026

Table 1 The descriptors available for the SR search

Descriptor	Symbol	Unit
Temperature	T	K
van der Waals (vdW) volume	V	Å^3
Radius of gyration	R	Å
Inertial shape factor	Z	μ^{-1} Å^{-2}
Asphericity	A	N/A
Eccentricity	E	N/A
Spherocity index	ζ	N/A
Norm of dipole moment	μ	CÅ
Gini coefficient (of charge distribution)	G	N/A

We ran energy minimizations using the MMFF94 force field to obtain the lowest energy conformer, to subsequently be used for computing the molecular descriptors that require geometrical information.[45] To compute partial charges, we combined the RDKit with the open-source toolbox Open Babel[46] and used the charge equilibration method (Qeq), which has been shown to give charges that agree well with experimental dipole moments.[47] Finally, we intentionally restricted our set of descriptors to only those that carry a clear physical meaning, which besides molecular volume includes shape and charge distribution (Table 1). For details on how the descriptors are calculated, see Note 2 in the SI. To distinguish which ion a descriptor refers to, we used the subscripts c (cation) and a (anion).

2.3 Symbolic regression implementation and runs

The SR implementation used the open-source library PySR[48] and its multi-population evolutionary algorithm that also optimises for unknown scalar constants. In an evolutionary algorithm for SR, analytical expressions are represented by expression trees, with leaf nodes for the variables and internal nodes for the operators. An expression tree that computes the ionic conductivity (σ) as a function of three variables and three operators is schematically shown in Fig. 1. To progress, the algorithm suggests new expressions by applying mutation operators to the pool of existing expression trees and then computes fitness scores based on a loss function. To not have a growing population of expression trees, the fitness scores were compared in tournament selections where only the fittest individuals were kept. By running several populations in
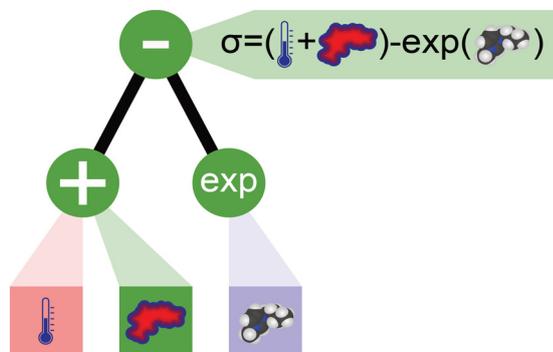

Fig. 1 A schematic illustration of an expression tree that computes the sum of two input variables and subtracts the exponent of a third one.

parallel, the algorithm becomes more efficient whilst also promoting a greater diversity in the found expressions.

The ultimate aim was to find generic functions that model the ionic conductivity as a function of temperature (T), a set of molecular descriptors (x_i), and numerical constants (p_i). To find expressions that are consistent with FVT and eqn (3), we constructed a template function:

$$\sigma = \frac{f(x_1, x_2, \ldots)}{\sqrt{T}} \exp\left(-\frac{g(x_1, x_2, \ldots)}{h(T, x_1, x_2, \ldots)}\right) \quad (4)$$

for which SR is used to find the functions f, g and h. We treat the free volume in ILs as a more abstract concept than a single variable v_f and use the different molecular shape and charge descriptors (Table 1) to express the unknown functions. We choose to use an expression derived from FVT as our template function as we want to enforce physically motivated models that fit with the hypothesis of an ion-hopping mechanism governing ion transport. In principle any type of template function could be used, for instance the semi-empirical Vogel–Fulcher–Tammann equation. When searching for the functions, only operators $\mathcal{O} \in \{+, -, *, \div, \wedge\}$ are allowed in the expressions, and we define a complexity C as the sum of the number of operators, variables and constant scalars used. To not render overly complicated expressions we set $C_{max} = 30$.

Since the ionic conductivity has a strong temperature dependency, often across several orders of magnitude, it is beneficial to use a loss function that handles relative errors when computing the fitness of an expression. Therefore, we transformed our targets (y) and predictions ($\hat{y}$) to logspace before computing the root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=0}^{N}(y_i - \hat{y}_i)^2}, \quad (5)$$

where the sum runs over the N data points. We also computed the RMSE across individual ILs and added an average of those RMSEs to the loss. This helps to ensure that the models capture the temperature dependency of ionic conductivity well.

To further promote useful and physically motivated models, we made use of dimensional constraints. For ionic conductivities measured in mS cm^{-1}, eqn (4) must have units mS cm^{-1} K$^{0.5}$ whilst functions f and g must divide to a dimensionless quantity. To promote dimensionally consistent expressions, we penalize those expressions that do not meet the requirements on f, g and h by adding an additional term $\beta = 10^4$ to the loss. To not be too restrictive about the search space, however, we do allow scalar constants to have any dimensionality.

We ran PySR in a distributed setting using 32 cores for 72 hours. To ensure that the SR search does not get stuck in a local optimum of the search space, we ran 8 separate instances of the SR search in parallel. This will ideally generate a large pool of candidate expressions with various complexities. To find the model that best balances accuracy and complexity, we ran a model selection algorithm that assigns scores to the discovered expressions. By computing a cost ε for each model, the score function first selects the subset of models that have

the lowest cost for a given complexity (C). Then, a second subset of n models that have a cost equal or within a maximum 5% of the lowest cost achieved are considered. For these n models, a score (S) was computed by taking the derivative of the negative logarithmic cost with respect to complexity:

$$S = -\frac{\partial}{\partial C (\log \varepsilon)}. \quad (6)$$

The highest scoring model was then considered the best. To align with our desire to promote models that are performant when it comes to capturing the temperature dependency of individual ILs, the cost was defined as the average of RMSEs computed across every IL in the dataset.

3 Results and discussion

We start by showing and discussing the performance of the two best models, one for (i) – the in-house dataset – and one for (ii) – the large dataset collected from ILThermo – before we continue to explore how the models compare and perform for different types of ILs.

3.1 In-house dataset

For this dataset, our best model has a complexity $C = 27$ and is given by:

$$\sigma = \frac{\left(V_c p_0 - A_c p_1 ((G_c - p_2)(E_a + \zeta_c - p_3))^{-1}\right) e^{-\frac{V_c p_4 - \mu_a}{T p_5 - p_6}}}{\sqrt{T}} \quad (7)$$

and is thus a function of many properties: asphericity (A), van der Waals volume (V), Gini coefficient (G), eccentricity (E), spherocity index (ζ), norm of the dipole moment (μ) and the temperature (T). The numerical values p_i are all positive and given in Table S1. The model predictions (σ_{SR}) show a good agreement with the experimental targets (σ_{exp}), both for the training ($R^2 = 0.97$) and validation ($R^2 = 0.94$) set (Fig. 2). The similar prediction accuracies indicate that the model does not overfit. Furthermore, the model predictions for four different ILs [N$_{1116}$][TFSI] (orange), [C$_6$C$_1$Im][PF$_6$] (light blue), [Pyr$_{14}$]-[OTf] (dark blue) and [N$_{2223}$][FSI] (green) show how it captures the temperature dependencies well across the dataset (Fig. 3), with smooth predictions across the temperature ranges, another sign that it does not overfit. The latter can partly be attributed to the comparatively limited expressiveness of an analytical model. Combining the SR search with the template function in eqn (4) helps to find physically motivated expressions – one of the major reasons for choosing SR over NNs, which often overfit in the limited data regime as the number of model parameters in a NN often will exceed the number of data points in the training set, increasing the risk to memorize data without capturing true correlations.[49]

As for the expression that constitute the found model (eqn (7)), we highlight that the molecular volume of the cation is larger than the norm of the dipole moment of the anion for all ILs in the dataset. This means that the exponential captures the expected positive correlation between ionic conductivity
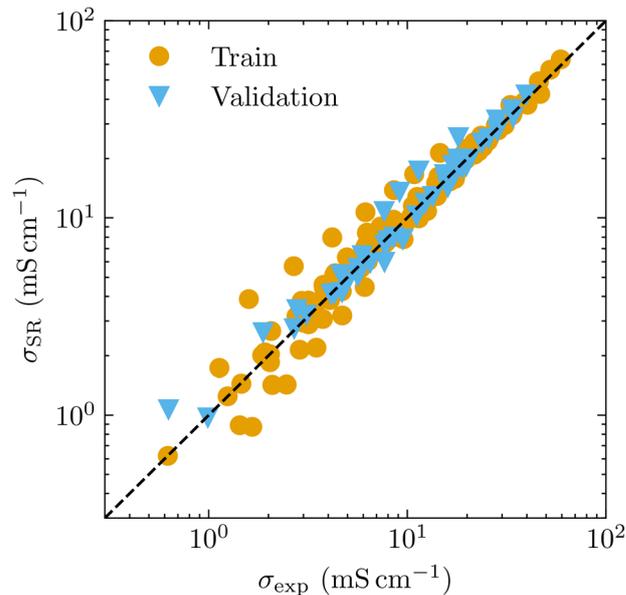

Fig. 2 Measured ionic conductivities *vs.* the model predictions computed using eqn (7), found by SR applied to the inhouse dataset. Orange circles indicate training data, whilst blue triangles represent validation data.
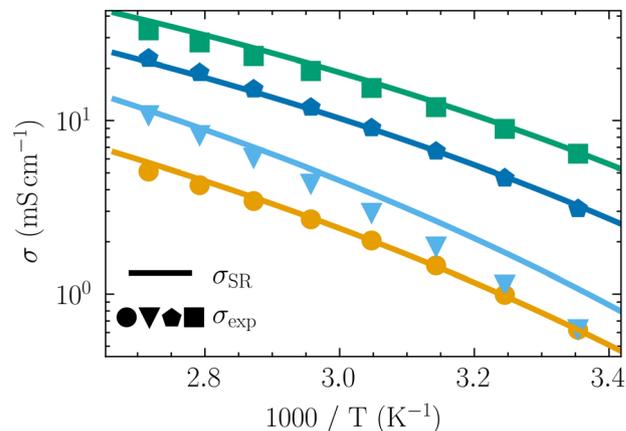

Fig. 3 Measured ionic conductivities (markers) and model predictions (solid lines) found using eqn (7) for a selection of four ILs plotted against temperature.

and temperature, and also suggests that ILs with larger cations have lower conductivities, which is consistent with structural diffusion. Furthermore, the model indicates that the molecular structure is important in determining the magnitude of the ionic conductivity, as molecular asymmetry is captured by the asphericity, eccentricity, and spherocity index in the pre-factor. As a higher asphericity indicates a more asymmetric ion, a higher eccentricity a more elongated ion, and a higher spherocity index a more spherical ion, the model suggests that ILs with bulkier and more asymmetric ions have lower ionic conductivities. It is a rather complex interplay, though, as ILs with more elongated anions and a high charge delocalization on the cation (by the Gini coefficient), appears to give higher

3844 | *Phys. Chem. Chem. Phys.*, 2026, **28**, 3841–3849

This journal is © the Owner Societies 2026

conductivities. While the size of the dataset limits the extrapolation possible, the excellent prediction accuracy indicates that an ion-hopping mechanism is responsible for the ion transport in these ILs.

3.2 Large dataset

For this dataset, our best model is given by:

$$\sigma = \frac{p_0 p_1^{E_c} e^{-\frac{R_c p_2 - R_a p_3 + p_4}{T - G_a^{A_a P_5} p_6 - p_7}}}{\sqrt{T}}. \tag{8}$$

This model has a complexity $C = 25$ and is a function of the properties: eccentricity (E), radius of gyration (R), Gini coefficient (G), asphericity (A) and temperature (T). The numerical values p_i are all positive and given in Table S2. The spread in prediction accuracy is significant for both the training ($R^2 =$ 0.76) and validation ($R^2 = 0.73$) set, although the intensity peak that indicates where most data points lie is centred around the ideal prediction line (Fig. 4 and Fig. 5). Examples of ILs that the model capture the ionic conductivity trend well for include [P_{66614}][Hex] (orange), [N_{2228}][TFSI] (light blue), [BMIM][PF_6] (green) and [S_{221}][TFSI] (dark blue) (Fig. 6). This indicates that the model has generalised to a relatively diverse set of ILs, here displaying good performance for both phosphonium-, ammonium-, imidazolium- and sulfonium-based ILs. Yet, the model struggles to cover all ILs in the dataset. This could indicate that FVT is not a suitable description for all ILs, or that the molecular descriptors we use are insufficient to describe the shape and size of the voids that drive molecular diffusion. It should be noted, though, that the relation for ionic conductivity as derived from FVT (eqn (3)), only is a proportional one, not aimed to predict any absolute values. If the model can capture the temperature dependency of ionic conductivity reasonably well for an IL, that could still be indicative of an ion transport governed by the formation of sufficiently

Fig. 5 Measured ionic conductivities vs. the model predictions for the validation set, computed using eqn (8) found by applying SR to the large dataset.

Fig. 6 Measured ionic conductivities (markers) and model predictions (solid lines) found using eqn (8) for a selection of four ILs plotted against temperature.

large voids, even if that model would have significant deviations in absolute terms.

The expression that constitutes the model (eqn (8)) looks quite different as compared to the one found for the in-house dataset (eqn (7)). The model indicates that cations with more elongated shapes tend to yield ILs with higher ionic conductivities, which is somewhat counter-intuitive, as one might expect asymmetric and/or bulky ions to diffuse more slowly. However, elongation is a different structural attribute from overall asymmetry, and ions that are extended along a single axis may, in fact, diffuse more readily despite not being particularly symmetric. Why only the eccentricity appears in the pre-factor is hard to say, but it is possible that our upper bound on the model complexity, $C_{max} = 30$, do not leave much room for the SR search to explore complex expressions in both the pre-factor and the exponential. For the in-house dataset the best model found yields accurate predictions despite featuring a simple
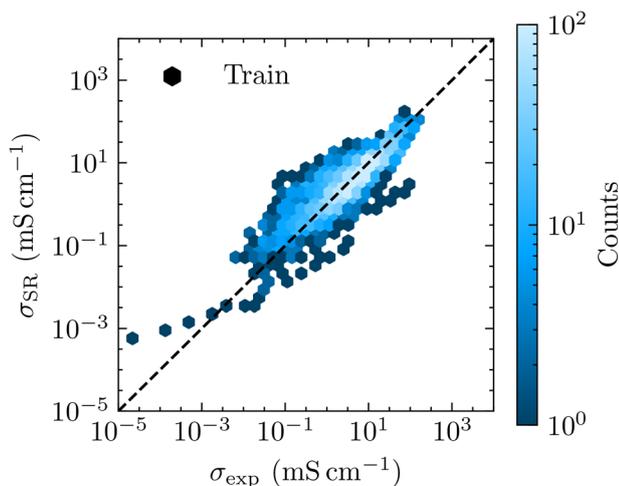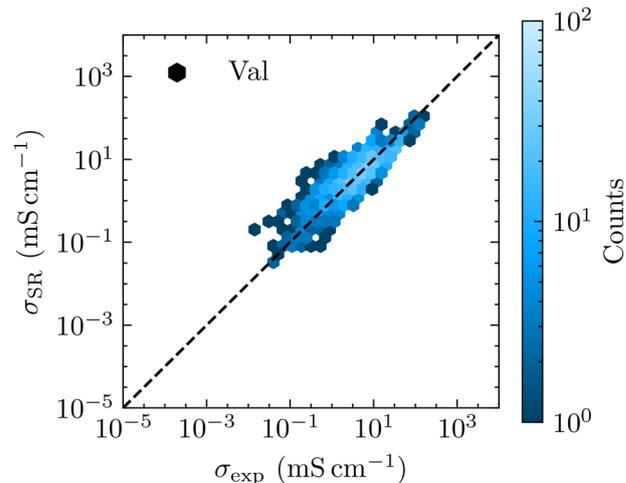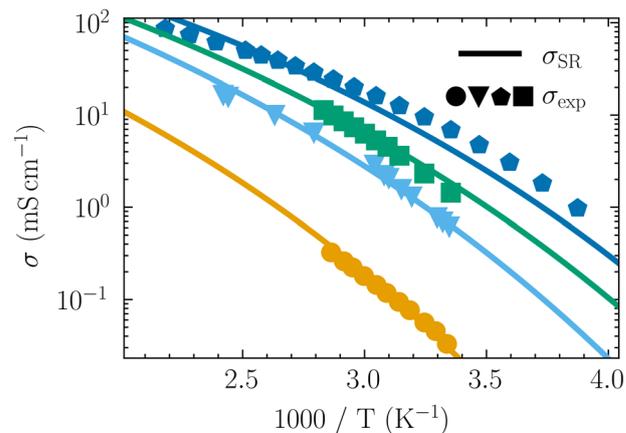

Fig. 4 Measured ionic conductivities vs. the model predictions for the training set, computed using eqn (8) found by applying SR to the large dataset.

This journal is © the Owner Societies 2026

Phys. Chem. Chem. Phys., 2026, **28**, 3841–3849 | **3845**

expression in the exponential, but the considered ILs were limited both in number and diversity. When the SR search for the large dataset instead finds an expression with a more complex exponential term, this suggests that the conductivity response to temperature can vary significantly between different ILs and that the radius of gyration of both the cations and the anions influence this, but with opposite signs, which suggests that whilst bulkier cations lower the conductivity, large anions increase it. This has previously been attributed to the fact that larger anions render weaker interionic interactions.[50] The denominator in the exponential features a measure of anion asymmetry, the asphericity, combined in the same term as a description of charge delocalization in the anion, the Gini coefficient. Higher asphericity indicates a more asymmetric anion and a higher Gini coefficient corresponds to a more imbalanced charge distribution. As this term increases, the denominator becomes smaller, which corresponds to a lower ionic conductivity. Since a more imbalanced charge distribution should make it easier for the anion to form stronger ionic interactions, thus lowering the ion mobility, it is reasonable that anions with higher Gini coefficients would give ILs with lower ionic conductivities. Overall, this suggests that both the shape and charge distribution of the anion play important roles in determining the ionic conductivity of ILs. We also emphasize that both the numerator and the denominator in the exponential remain positive for all ILs in the dataset, ensuring the expected positive correlation between temperature and ionic conductivity.

To better understand where the model fails, we categorize ILs based on both their cation and anion, averaging the RMSE (computed per IL across a temperature range, in logspace) for each category (Fig. 7). The combination of a triazolium cation and a sulfonate anion shows the highest error, owing to a single IL: [₄MPrTr][Tos]. In the original study providing the

experimental data, the authors do note that the tosylate anion gave an IL with a significantly higher relative viscosity and lower ionic conductivity as compared to analogous ILs based on other anions, such as $[NO_3]^-$ and $[BF_4]^-$. They attribute this to strong π–π interactions between the aromatic rings of the tosylate anion and the triazolium cation.[51] Considering how our model predicts the ionic conductivities of other ILs in the study to a much higher accuracy, it appears that it fails to capture these strong interactions. This could also explain why the model performs worse for ILs based on morpholinium, which can form strong ionic interactions as the ether groups act as dipole sites. In line with the observed error, this effect is particularly strong for polarizable anions such as dicyanamide, a pseudo-halide. Similarly, both the phosphonium-carbanion and the phosphonium-halide/pseudohalide ILs report comparatively low ionic conductivities, attributed to strong association, and these are two other example of where our model fails to make accurate predictions.[52,53] Strong interactions do render poor ILs, often classified by the amount of deviation from ideal behaviour in a Walden plot, commonly attributed to the formation of neutral ion-pairs.[54] Cations with long alkyl chains can have similar effects, as dispersive forces may create a meso-scopic structure with small, localized charge regions separated by neutral domains. This could be another reason for why our model fails for the morpholinium group, which include cations with alkyl chains of up to nine carbons.[55]

As our FVT-inspired model builds on the Nernst–Einstein relation (eqn (2)) that assumes non-interacting ions, our model should in principle, by design, be less accurate for all ILs with strong interactions. Although the SR approach allows for considerable corrections to eqn (3) using a more general *Ansatz* (eqn (4)), it is likely quite hard to have a single model accurately capture the ionic conductivity of ILs that appear in very
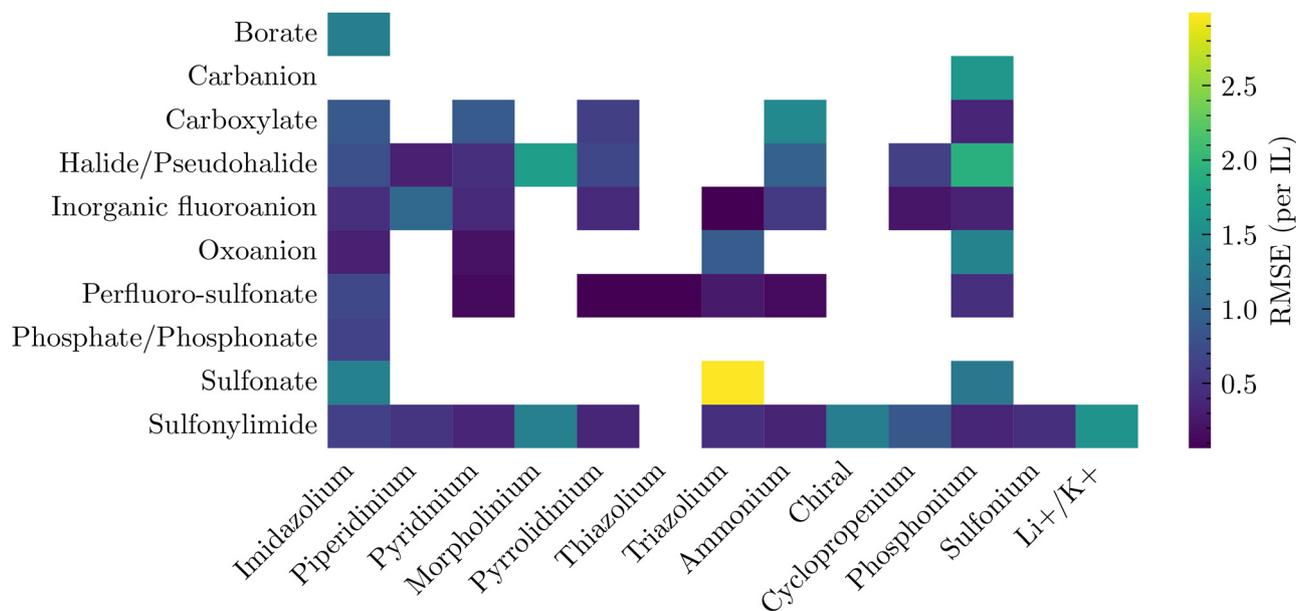

Fig. 7 A heatmap showing how the average RMSE (computed per IL across a temperature range, in logspace) differ across different types of ILs. In the vertical direction the ILs are classified based on their anion category, and on the horizontal based on their cation type.
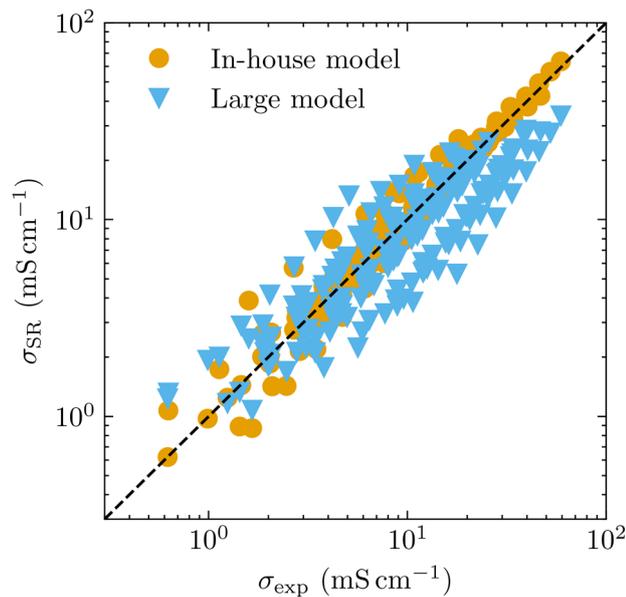
3846 | *Phys. Chem. Chem. Phys.*, 2026, **28**, 3841–3849

This journal is © the Owner Societies 2026

Fig. 8 Measured ionic conductivities *vs.* the model predictions for the in-house dataset using eqn (7) (orange circles, found by SR applied to the in-house dataset), and eqn (8) (blue triangles, found by SR applied on the large dataset).

different regions of a Walden plot, it will be more accurate for ILs closer to the ideal Walden line. The ILs with the small lithium and potassium cations are also very far from ideal ILs; most ILs consist of bulky organic cations.

Although it might seem discouraging that the model has not managed to accurately describe the ionic conductivity of all ILs in the dataset, it is important to reflect on what this means. Clearly, searching for expressions inspired by FVT using SR do not yield a model than generalise across the entire dataset. However, the model still works very well for many ILs, with less accurate predictions primarily observed for non-ideal ones. This suggests that the SR approach can capture fundamental correlations that apply to the more idealised systems.

To strengthen this belief, we explore how all eight found models of complexity $C = 25$ compare. Our SR implementation is inherently stochastic and there is no guarantee that different runs will yield similar models, especially if there would be no clear underlying relationships in the data. Yet, looking at how the models differ in their predictions, we find that they agree to a high extent, both for very accurate predictions such as the ones for [HPy][TFSI], and for considerably less accurate examples like [4MPrTr][Tos] (Fig. 9). Instead of a case where each SR search overfit to different chemical motifs, the discovered models converge to make similar predictions. Importantly, this shows that our models capture fundamental correlations and is a testament to the robustness of the SR approach.

To further test if the model from the large dataset has captured the relevant physics we cross-test it on the smaller in-house dataset and compare *vs.* the model represented by eqn (7). Unfortunately, it does not achieve the same prediction accuracy and tends to underestimate the ionic conductivities (Fig. 8). A possible/plausible explanation is that although the model performs best for more ideal ILs, the exposure to less ideal ILs during training has given it suboptimal performance across the board. Indeed, the loss function guiding the SR search punishes models that completely ignore a subset of ILs, and instead rewards those that display decent performance for many ILs. Especially, as discussed above, the large dataset contained several ILs with strong interionic interactions and low ionic conductivities, why it is reasonable to expect that the SR search tried to find models that adapted at least somewhat to these data, which could manifest itself as underestimations of the ionic conductivities.

On a more general note, the difficulties of dealing with experimental data sourced from 125 different publications should not be underestimated. In many cases, the database used to construct the dataset, ILThermo, has reported ionic conductivities that are multiple times or even magnitudes higher than other data for the same IL. These differences can likely be explained by variations in sample purity or dryness,
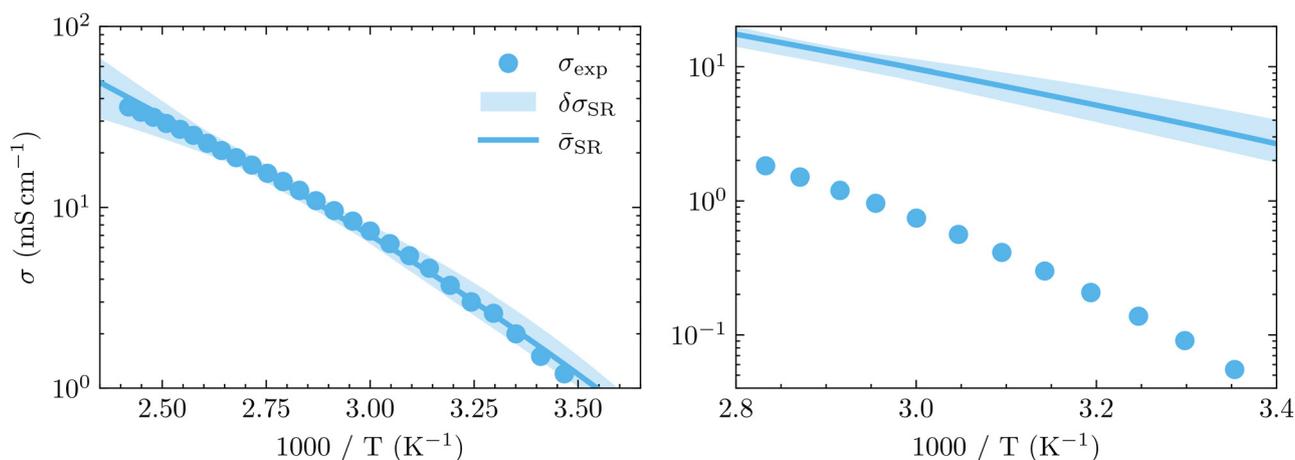

Fig. 9 The model prediction differences across eight independently discovered SR models of complexity $C = 25$, for an IL the model gives accurate predictions for (left, [HPy][TFSI]) and an IL that the model fails to describe (right, [4MPrTr][Tos]). The shaded areas give the prediction ranges, the solid lines represent the mean predictions and the circles are experimental data.

This journal is © the Owner Societies 2026

Phys. Chem. Chem. Phys., 2026, **28**, 3841–3849 | **3847**

which is known to have a considerable effect on the physical properties of ILs.[56,57] When several publications had data on the ionic conductivity of a specific IL, our rule of always choosing the data series with the lowest reported ionic conductivities could help filter out less pure ILs, but for many there is only one publication and set of data.

4 Conclusions

FVT and SR have been used to find a set of analytical models that describe the temperature-dependent ionic conductivity of ILs, with molecular descriptors related to the shape and charge distribution of the ions as model inputs. For a dataset collected under controlled measurement conditions, SR proved itself as a performant modelling alternative, giving an analytical model with a good prediction accuracy across all 22 considered ILs without overfitting. This highlights the advantage of using SR in the limited data regime where conventional ML methods often overfit.

Moving to a considerably larger dataset collected from the database ILThermo, SR found symbolic models that could capture the temperature dependency of many ILs, but with a relatively high spread in prediction accuracy. Grouping the ILs based on their constituent ions, we observe that the model performs better for more ideal ILs. When ionic interactions become stronger, or long alkyl chains create mesoscopic ordering with large neutral domains, the model predictions become less accurate. It appears difficult to capture the conductivity behaviour of ILs with very different amounts of ion pairing in a single model, at least when enforcing an expression consistent with FVT. It is likely that the diversity of the two datasets had an impact on the model performance even for more ideal ILs, this as the best model for the larger dataset tends to underestimate the ionic conductivities of the smaller dataset and does not achieve the same performance as the first model. Allowing for a model with multiple terms, each one consistent with the expression in eqn (3), could perhaps allow the model to generalise better. This would, however, come at the expense of model complexity.

The lack of control we had over the large dataset is also believed to be a contributing factor to the spread in prediction accuracy. As even trace amounts of water or other contaminants are known to affect the ionic conductivity, the model performance should in practice be limited by the variation in sample purity. This further points towards controlled, in-house measurements, as the ideal setting for SR. However, there is a strong argument for reusing data when possible, and in future studies it would be interesting to address the issues that arise from dealing with data from multiple sources.

Finally, the present models and the SR approach we have designed are most useful stepping-stones towards also modelling the ionic conductivities of IL-based electrolytes for LIBs, but as we then must expect strong ionic interactions by the introduced lithium cations there is at the same time no simple blue-print solution.

Author contributions

I. B.: investigation, data curation, methodology, software, formal analysis, visualization, writing – original draft. P. J.: conceptualization, supervision, validation, writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data that support the findings of this study are available at Researchdata.se at https://doi.org/10.71870/11vs-fb95. This includes both experimental data, ionic conductivity *vs* temperature for all considered ILs, and the computed molecular descriptors. The descriptors were computed using the open-source toolkits RDKit[42] and OpenBabel.[46] The modelling method was based on SR, implemented by the open-source library PySR.[48] To enable the use of dimensional constraints for the FVT-inspired template function, minor modifications to PySR and its backend SymbolicRegression.jl had to be made. These are provided as forks to the original repositories at https://github.com/ibengtsson/PySR and https://github.com/ibengtsson/SymbolicRegression.jl.

Supplementary information (SI): the SI discusses how chemical names were parsed to SMILES strings, provides definitions for the molecular descriptors used, and contains tables with the numerical parameters found for the analytical models. See DOI: https://doi.org/10.1039/d5cp04143k.

Acknowledgements

References

1 C. A. Vincent, *Solid State Ionics*, 2000, **134**, 159–167.
2 Y. S. Meng, V. Srinivasan and K. Xu, *Science*, 2022, **378**, 1065.
3 A. Hammami, N. Raymond and M. Armand, *Nature*, 2003, **424**, 635–636.
4 J. B. Goodenough and Y. Kim, *Chem. Mater.*, 2010, **22**, 587–603.
5 K. Xu, *Chem. Rev.*, 2014, **114**, 11503–11618.

3848 | *Phys. Chem. Chem. Phys.*, 2026, **28**, 3841–3849

This journal is © the Owner Societies 2026

6 Y. Chen, Y. Kang, Y. Zhao, L. Wang, J. Liu, Y. Li, Z. Liang, X. He, X. Li, N. Tavajohi and B. Li, *J. Energy Chem.*, 2021, **59**, 83–99.

7 Y. Yamada and A. Yamada, *J. Electrochem. Soc.*, 2015, **162**, A2406.

8 O. Borodin, J. Self, K. A. Persson, C. Wang and K. Xu, *Joule*, 2020, **4**, 69–100.

9 M. Watanabe, M. L. Thomas, S. Zhang, K. Ueno, T. Yasuda and K. Dokko, *Chem. Rev.*, 2017, **117**, 7190–7239.

10 Y. Yamada, J. Wang, S. Ko, E. Watanabe and A. Yamada, *Nat. Energy*, 2019, **4**, 269–280.

11 K. Liu, Z. Wang, L. Shi, S. Jungsuttiwong and S. Yuan, *J. Energy Chem.*, 2021, **59**, 320–333.

12 Y. Aihara, K. Sugimoto, W. S. Price and K. Hayamizu, *J. Chem. Phys.*, 2000, **113**, 1981–1991.

13 E. Flores, G. Åvall, S. Jeschke and P. Johansson, *Electrochim. Acta*, 2017, **233**, 134–141.

14 F. Lundin, L. Aguilera, H. W. Hansen, S. Lages, A. Labrador, K. Niss, B. Frick and A. Matic, *Phys. Chem. Chem. Phys.*, 2021, **23**, 13819–13826.

15 G. Åvall, J. Wallenstein, G. Cheng, K. L. Gering, P. Johansson and D. P. Abraham, *J. Electrochem. Soc.*, 2021, **168**, 050521.

16 R. Andersson, F. Årén, A. A. Franco and P. Johansson, *J. Electrochem. Soc.*, 2020, **167**, 140537.

17 A. W. Taylor, P. Licence and A. P. Abbott, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10147–10154.

18 A. K. Doolittle, *J. Appl. Phys.*, 1951, **22**, 1471–1475.

19 M. H. Cohen and D. Turnbull, *J. Chem. Phys.*, 1959, **31**, 1164–1169.

20 W. Beichel, Y. Yu, G. Dlubek, R. Krause-Rehberg, J. Pionteck, D. Pfefferkorn, S. Bulut, D. Bejan, C. Friedrich and I. Krossing, *Phys. Chem. Chem. Phys.*, 2013, **15**, 8821–8830.

21 N. V. Plechkova and K. R. Seddon, *Chem. Soc. Rev.*, 2007, **37**, 123–150.

22 J.-W. Park, K. Ueno, N. Tachikawa, K. Dokko and M. Watanabe, *J. Phys. Chem. C*, 2013, **117**, 20531–20541.

23 M. Kerner, N. Plylahan, J. Scheers and P. Johansson, *Phys. Chem. Chem. Phys.*, 2015, **17**, 19569–19581.

24 G. A. Elia, U. Ulissi, S. Jeong, S. Passerini and J. Hassoun, *Energy Environ. Sci.*, 2016, **9**, 3210–3220.

25 S. Tsuzuki, H. Tokuda, K. Hayamizu and M. Watanabe, *J. Phys. Chem. B*, 2005, **109**, 16474–16481.

26 S. Tsuzuki, *ChemPhysChem*, 2012, **13**, 1664–1670.

27 O. Borodin, *J. Phys. Chem. B*, 2009, **113**, 12353–12357.

28 H. Liu and E. Maginn, *ChemPhysChem*, 2012, **13**, 1701–1707.

29 A. P. Abbott, *ChemPhysChem*, 2004, **5**, 1242–1246.

30 A. P. Abbott, *ChemPhysChem*, 2005, **6**, 2502–2505.

31 H. Matsuda, H. Yamamoto, K. Kurihara and K. Tochigi, *Fluid Phase Equilib.*, 2007, **261**, 434–443.

32 K. Tochigi and H. Yamamoto, *J. Phys. Chem. C*, 2007, **111**, 15989–15994.

33 K. Baran and A. Kloskowski, *J. Phys. Chem. B*, 2023, **127**, 10542–10555.

34 I. Baskin, A. Epshtein and Y. Ein-Eli, *J. Mol. Liq.*, 2022, **351**, 118616.

35 C. Song, C. Wang, F. Fang, G. Zhou, Z. Dai and Z. Yang, *J. Chem. Eng. Data*, 2024, 4310–4319.

36 A. Racki and K. Paduszyski, *J. Chem. Inf. Model.*, 2025, **65**, 3161–3175.

37 M. Raissi, P. Perdikaris and G. E. Karniadakis, *J. Comput. Phys.*, 2019, **378**, 686–707.

38 Y. Zhu, N. Zabaras, P.-S. Koutsourelakis and P. Perdikaris, *J. Comput. Phys.*, 2019, **394**, 56–81.

39 J. Nilsson-Hallen, B. Ahlström, M. Marczewski and P. Johansson, *Front. Chem.*, 2019, **7**, 126.

40 A. Kazakov, J. Magee, R. Chirico, E. Paulechka, V. Diky, C. Muzny, K. Kroenlein and M. Frenkel, https://ilthermo.boulder.nist.gov.

41 Q. Dong, C. D. Muzny, A. Kazakov, V. Diky, J. W. Magee, J. A. Widegren, R. D. Chirico, K. N. Marsh and M. Frenkel, *J. Chem. Eng. Data*, 2007, **52**, 1151–1159.

42 RDKit: Open Source Cheminformatic, https://www.rdkit.org.

43 N.-O. Friedrich, C. de Bruyn Kops, F. Flachsenberg, K. Sommer, M. Rarey and J. Kirchmair, *J. Chem. Inf. Model.*, 2017, **57**, 2719–2728.

44 S. Riniker and G. A. Landrum, *J. Chem. Inf. Model.*, 2015, **55**, 2562–2574.

45 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519.

46 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminf.*, 2011, **3**, 33.

47 A. K. Rappe and W. A. Goddard, *J. Phys. Chem.*, 1991, **95**, 3358–3363.

48 M. Cranmer, *arXiv*, 2023, preprint, DOI: 10.48550/arXiv.2305.01582.

49 X. Ying, *J. Phys.: Conf. Ser.*, 2019, **1168**, 022022.

50 J. Leys, R. N. Rajesh, P. C. Menon, C. Glorieux, S. Longuemart, P. Nockemann, M. Pellens and K. Binnemans, *J. Chem. Phys.*, 2010, **133**, 034503.

51 U. G. Brauer, A. T. De La Hoz and K. M. Miller, *J. Mol. Liq.*, 2015, **210**, 286–292.

52 X.-L. Wu, X.-Y. Sang, Z.-M. Li and D.-J. Tao, *J. Mol. Liq.*, 2020, **312**, 113405.

53 A. García, L. C. Torres-González, K. P. Padmasree, M. G. Benavides-Garcia and E. M. Sánchez, *J. Mol. Liq.*, 2013, **178**, 57–62.

54 W. Xu, E. I. Cooper and C. A. Angell, *J. Phys. Chem. B*, 2003, **107**, 6170–6178.

55 O. Russina, R. Caminiti, A. Triolo, S. Rajamani, B. Melai, A. Bertoli and C. Chiappe, *J. Mol. Liq.*, 2013, **187**, 252–259.

56 K. R. Seddon, A. Stark and M.-J. Torres, *Pure Appl. Chem.*, 2000, **72**, 2275–2287.

57 C. Ma, A. Laaksonen, C. Liu, X. Lu and X. Ji, *Chem. Soc. Rev.*, 2018, **47**, 8685–8720.

This journal is © the Owner Societies 2026

Phys. Chem. Chem. Phys., 2026, **28**, 3841–3849 | **3849**