Development and Application of the CL&Pol Polarisable Force Field for Ionic Liquid-Based Electrolytes

TAYLA LEE WILSON

Supervisor: Dr. Gerhard A. Venter

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

January, 2025

Department of Chemistry

Faculty of Science

University of Cape Town

Declaration:

1. I know that plagiarism is wrong. Plagiarism is to use another's work and to pretend that it is one's own.

2. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

3. This project is my own work.

4. I have included internet article, book, or other material references used for this project.

Signed: Signed by candidate

Date: Monday 27th January, 2025

Abstract

Ionic liquids (ILs) are a fascinating class of molecular systems due to diverse and promising applications, including the potential use as electrolyte systems, replacing traditional volatile organic electrolytes. The appeal of ILs for such applications lies in their favourable properties, such as high diffusivity, thermal and electrochemical stability, and low volatility. Molecular dynamics (MD) simulation is a powerful tool for studying the physical properties of liquids; however, whereas traditional solvents are typically well described using classical, fixed-charged force fields (FFs), the explicit inclusion of polarisation is essential for accurate description of IL dynamics. Consequently, the development and application of polarisable FFs for ILs is a necessary focus within IL research.

Therefore, this work's first aim was to extensively validate the recently developed Drude-based CL&Pol FF for six pure ILs and five alkali-earth-containing IL-electrolytes. Reproducibility and precision are not often addressed when MD simulation is used to calculate thermophysical properties, yet without quantification of uncertainty, the value of a validation study is questionable. Hence, statistically meaningful uncertainties were reported for all properties as a 95 % confidence interval of the mean over replicate simulations. The simulation protocol was further validated using well-known theoretical relationships (e.g., the Stokes-Einstein and Nernst-Einstein equations) to confirm the internal consistency of key calculated transport properties. The accuracy of calculated properties of pure ILs varied, with average errors as low as 1 % for density, 35 % for viscosity, and 50 % for conductivity. Most properties could be calculated with uncertainties of $\sim$20 %, while calculated conductivities had uncertainties of $\sim$50 %.

The second aim of this work involved the further development of the Van der Waals component of the intermolecular interaction. The CL&Pol Lennard-Jones (LJ) parameters are carried over from its fixed-charge predecessor, CL&P. The parameters are then adjusted with the prescribed scaling of the LJ well-depth (ε_{ij}) to remove the induction contribution, making it transferable to the Drude FF featuring explicit polarisation. While scaling of ε_{ij}

produces reasonably accurate transport properties, the resulting induction-free LJ potentials for interactions involving the alkali metals do not reproduce the ab initio exchange-dispersion potentials, producing theoretically unsatisfactory Van der Waals interactions. Furthermore, the existing OPLS ε_{ij} parameters for the alkali metals do not correlate with the strength of the ab initio dispersion interactions, warranting reconsideration of these parameters.

Thus, this work presents a stable and robust protocol for obtaining Van der Waals potential parameters compatible with a Drude FF without the need for scaling, based on first-principles resolution of the dispersion and exchange components of the potential using Symmetry-Adapted Perturbation Theory (SAPT). However, while overall interaction energies can be accurately obtained with small basis sets, this is due to error cancellation in the components. Consequently, the f_{2+}SAPT0 complete basis set (CBS) methodology was developed using the extended AHB21$\times$5 dataset of anion-neutral dimers. This method uses basis set extrapolation and scaling of SAPT0 components to provide dispersion and exchange components with an average deviation of $\sim$10 % from higher order SAPT2+/CBS equivalents. Finally, investigation into various potentials showed that Halgren's buffered LJ potential describes the exchange and dispersion interactions more consistent with the equivalent f_{2+}SAPT0/CBS components, particularly at short range, than is offered by the CL&Pol scaled LJ potential.

Acknowledgements

I would first like to thank my supervisor, Dr. Gerhard Venter, for the guidance, mentorship and many invaluable discussions over the past six years. My thanks is extended to the University of Cape Town Science Faculty for providing funding for this project. Acknowledgement is given to the Centre for High-Performance Computing (CHPC), South Africa and the University of Cape Town's ICTS High Performance Computing team, for providing computational resources to this research project. To my dear friend Sarah, thank you for your unwavering motivation and for being the ultimate hype girl. Josh, from listening to me go on about my research, to helping cheer me up after a tough day — thank you for everything. I love you. I would like to express my deepest gratitude to my brother Bradlee, and my parents Glisson, Lynne and Shelagh, for all the love and motivation throughout my PhD. Dad, I could not have made it this far without your support. Wherever you were in the world, you were able to provide the most invaluable advice, encouragement and inspiration. Thank you for helping me achieve this dream!

Contents

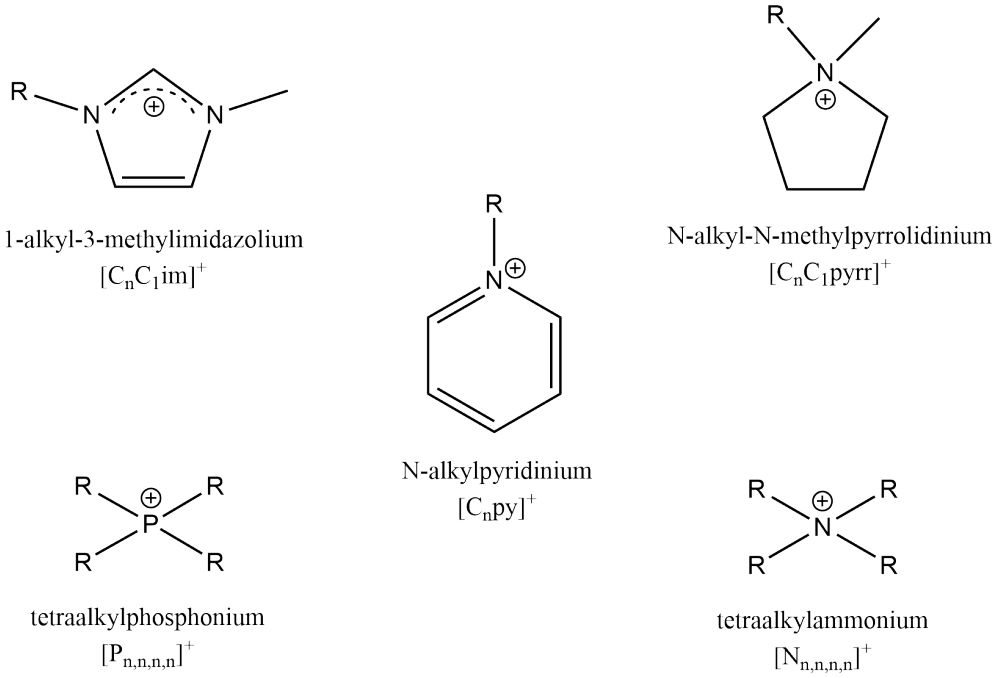
Chapter 1

Introduction

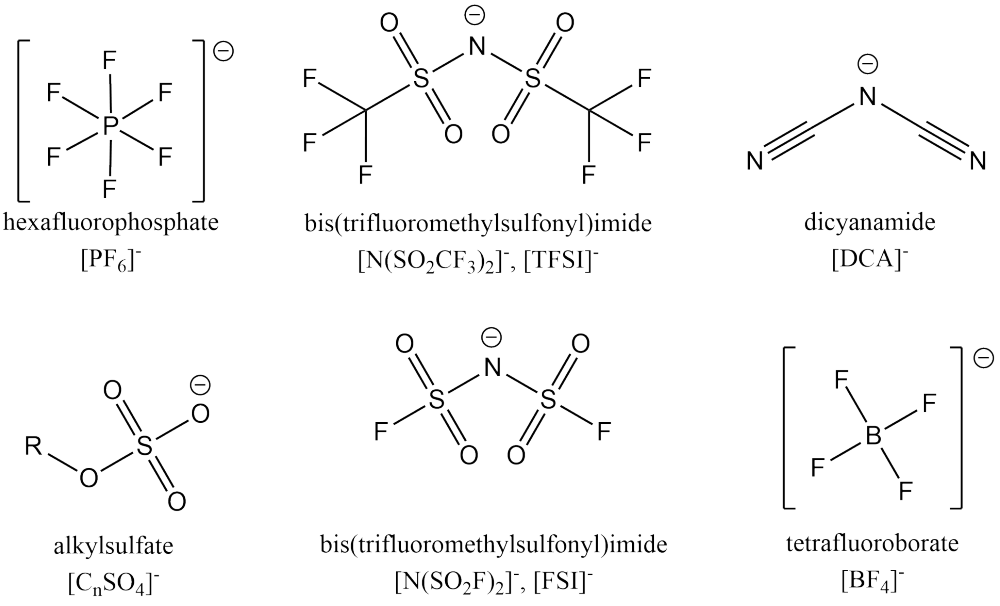
1.1 Introduction to ionic liquids

Liquids are ubiquitous in daily life and regularly encountered. But what are *ionic* liquids? Ionic liquids (ILs) are defined, practically, as salts that are liquid at *ambient temperatures*.[1] This subtle temperature-based classification is what distinguishes ILs from crystalline salts, such as NaCl (for which the melting point is ∼800 °C). A further distinction can be made between ILs as "ambient-temperature" or "room-temperature" ILs, the latter (RTILs) being identified by melting points at temperatures lower than 100 °C.[2,3] This distinction is rather arbitrary, but literature often refers to RTILs, thus warranting a definition of this term. RTILs can be further described, based on the evolution of the field of research, as typically consisting of a large, asymmetric, nitrogen-containing organic cation paired with either an organic or inorganic anion.[4] This is a somewhat general classification that best defines an "ionic liquid" as explored in this work, and it should be pointed out that the definitions of, and terms "RTIL" and "IL" are often used interchangeably, but shall be referred to solely as ILs throughout this thesis. Regarding the naming of ILs, these are often long and are thus commonly abbreviated, generally consistent across references; however, these abbreviations do often vary. For example, the popular 1-ethyl-3-methylimidazolium cation

has been abbreviated as $[C_2C_1im]^+$, $[C_2MIM]^+$ or $[emim]^+$ previously.[5–8] In this thesis, the abbreviations suggested by Welton[5] are used. Common cation cores and anions, of which some feature in this work, are shown in Figure 1.1 to provide example abbreviations.



(a) Cations



(b) Anions

Fig. 1.1. Examples of typical cation cores (top) and anions (bottom) used for RTILs.

1.1.1 A brief history

There are clear nuances associated with ILs, and it is thus worthwhile to understand the history that led to such an intricate and interesting field of research. The "first" IL, ethylammonium nitrate, was synthesised in 1914 by Paul Walden,[9] revealing that salts consisting of large molecular cations have significantly lowered melting points than those containing atomic cations — a fundamental property defining ILs. It was only in 1951 when the true potential of such low-melting salts was recognised by Hurley and Wier:[10] ILs consisting of a pyridinium-based cation and a halide ion, mixed with aluminium chloride, were used as electrolytes for electroplating. IL and aluminium chloride systems were then largely used, not only as electrolytes but as solvents for synthesis and spectroscopy over the next few decades.[11–14] In the 1980s, Cooper and Angell[15] synthesised tetrachloroaluminate-containing ILs with large organic cations for use in batteries and metal electrodeposition, further demonstrating the role the large cation plays in reducing the melting point. While these systems moved away from mixtures containing aluminium chloride which is extremely reactive in water, the tetrachloroaluminate anion is also sensitive to hydrolysing to hydrochloric acid in the presence of moisture.[16] Therefore, an extremely important advance in the history of ILs was the synthesis of air- and water-stable ILs comprised of the $[C_2C_1im]^+$ cation with tetrafluoroborate ($[BF_4]^-$), nitrate ($[NO_3]^-$) and acetate ($[Ac]^-$) anions in 1992 by Wilkes and Zaworotko.[17] The development of such systems lead to an increased interest in ILs and has driven research into the many potential applications in synthesis and catalysis.[5,18–21] The wide spectrum of application of IL systems is, fundamentally, a result of the interesting physical properties of these species.

1.1.2 Physical properties of ionic liquids

ILs possess largely desirable properties such as low melting points and vapour pressures, high conductivity and diffusivity, as well as significant electrochemical and thermal stabil-

ity. These properties have led to the exploration of ILs as solvents for applications including metal electrodeposition,[22 23–25] batteries,[22,26–29] fuel cells[30–32] and supercapacitors.[33–38] As the scope of this work is in the context of electrolytes, an overview of the relevant properties will be given in this section.

1.1.2.1 Melting point

ILs are, by definition, low-melting salts; thus, the melting point is a fundamental property to quantify. The melting point, T_{m}, can be described as the lower limit of the liquid range, which is the temperature span between crystallization and thermal decomposition of a liquid — the temperature range within which the salt stays in a liquid state. T_{m} specifically refers to the temperature at which the solid phase loses its long-range order. For ILs, this definition can be less straightforward. Unlike many conventional salts, ILs often do not exhibit a clear melting point. Instead, many ILs form a supercooled liquid or undergo a glass transition, where the material transitions from a solid-like state to a highly viscous, amorphous liquid. The temperature at which a glass is formed is referred to as the glass transition temperature, T_{g}. It is important to be aware of both T_{m} and T_{g} for ILs, as measuring or computing liquid properties for a system at a temperature corresponding to a glass, can be problematic. Furthermore, for many ILs, particularly those with weakly coordinating anions like $[\mathrm{TFSI}]^-$ or $[\mathrm{BF_4}]^-$, thermal decomposition occurs before a true melting point is observed. Thus, the decomposition temperature is also a significant parameter when considering ILs for high-temperature applications.

Although typically not very easily determined, the melting points of ILs are typically lower than that of inorganic salts. T_{m} is influenced by the properties of both the solid and liquid phases, but it primarily reflects the interactions in the crystal phase, which are related to the free energy, G. When there are two possible states, the one with the lower free energy is preferred because it is more stable. Below T_{m}, the free energy of the crystalline phase, G_{c} is lower than that of the liquid phase G_{l}, while above T_{m}, the reverse

is true. Thus, T_{m} is the point at which the two phases are in equilibrium ($G_{\mathrm{c}} = G_{\mathrm{l}}$, and $\Delta G = 0$). Given the Gibbs-Helmholtz equation,

$$\Delta G = \Delta H - T\Delta S, \tag{1.1}$$

and given that T_{m} is equal to the temperature of fusion, T_{fus}, T_{fus} can be expressed as

$$T_{\mathrm{m}} = T_{\mathrm{fus}} = \frac{\Delta_{\mathrm{fus}} H}{\Delta_{\mathrm{fus}} S}. \tag{1.2}$$

Therefore, by decreasing $\Delta_{\mathrm{fus}} H$ or increasing the entropy of fusion $\Delta_{\mathrm{fus}} S$, T_{m} can be decreased. $\Delta_{\mathrm{fus}} H$ can be decreased by decreased the strength of the intermolecular interactions of the solid, in turn destabilising the crystal lattice. $\Delta_{\mathrm{fus}} S$ reflects the disorder of the crystalline state and increases with configurational degrees of freedom. Entropy, S, can be expressed as

$$S = k_{\mathrm{B}} \ln (W), \tag{1.3}$$

where W is the number of microstates, representing the different ways ions can be arranged to achieve the same configuration. As the number of uniquely positioned atoms increases, W decreases, leading to lower entropy in the solid phase. However, the entropy of the liquid phase remains unchanged regardless of the crystallization process, meaning that $\Delta_{\mathrm{fus}} S$ is primarily influenced by the entropy of the crystal phase. Therefore, when the crystal state consists of ions with widely varying positions, the ions exhibit lower entropy in the crystalline state and a greater increase in entropy upon melting. As a result, more configurational freedom leads to a larger $\Delta_{\mathrm{fus}} S$.

1.1.2.2 Intermolecular interactions

As highlighted earlier, the melting points of ILs are much lower than molten salts; this is due in part to the distinctive structural features of these species, but is mostly driven by the intermolecular interactions present. Attractive forces present in ILs include Coulomb (electrostatic) forces, van der Waals interactions, hydrogen bonding, as well as π-π interactions in some species. Of these forces, the electrostatic force, F, makes the largest contribution to the intermolecular interactions, quantified using Coulomb's Law:

$$F = \frac{kq_1q_2}{r^2}, \tag{1.4}$$

where k is Coulomb's constant, q_1 and q_2 are the charges of the interacting species, and r is the separation distance between the centres of the charges. The electrostatic force is inversely dependent on the distance of charge separation, which will be greater for larger molecules. ILs consisting of larger molecular ions, such as those depicted in Figure 1.1, will experience weaker electrostatic interactions than smaller molten salts, decreasing the melting point of these species. Weakening the electrostatic interactions is therefore a desired goal for the design of ILs. These interactions can also be reduced through charge delocalisation: When highly electronegative atoms are present in a molecule (or molecular ion) inductive effects can occur — the electronegative atoms (ie., fluorine in bis(trifluoromethane)sulfonimide)) will draw charge away from the central, charged atom, delocalising the net charge. Additionally, charge delocalisation can be driven by resonance stabilisation, leading to the spread of charge over multiple atoms.

Furthermore, electrostatic interactions can be weakened through shielding effects that are achieved in ILs with the presence of non-polar, alkyl chains attached to the cation core, decreasing the intermolecular contact between charged regions. This would suggest that longer alkyl chains are favourable for decreasing the melting point; however, when the alkyl chain is further than 12 Å (approx. seven carbons) from the cation core, melt-

ing points tend to increase with as the alkyl chain extends past this length. López-Martin et al.[39] illustrate this behaviour by classifying the regions of an IL[I] to be 1) the charge rich region, localised on the imidazolium core; 2) the symmetry breaking region, the park of the alkyl chain that extends from the core to a distance of 12 Å; and 3) the hydrophobic region, the section of the alkyl chain extending beyond the symmetry-breaking region. These regions are illustrated in Figure 1.2. If the alkyl chain length falls within the symmetry-breaking region it enhances the shielding effects, as well as increases the ion asymmetry, as seen in Figure 1.2. Removing symmetry decreases the packing efficiency, resulting in less stable crystal latices and lower melting points. When the chain length increases beyond the symmetry-breaking region, falling within the hydrophobic region, the dispersion effects begin to have a greater effect than that of the charge shielding, once again *increasing* the strength of the intermolecular interactions. It is worth noting that alkyl chain length also affects the aggregation of polar and non-polar structural domains. Using a set of imidazolium-based ILs, Canongia Lopes and Pádua[40] confirmed with molecular dynamics simulations that when the alky chain has four or fewer carbons, non-polar domains exist in dispersed clusters arranged in the polar domains; however, when the alkyl chain is longer, the non-polar regions are no longer dispersed and form a continuous network. These studies both draw a pertinent connection between structure and phase and as such, probing ILs that have alkyl chains that can be considered symmetry-breaking can be considered a fundamentally important physical feature to take into consideration when designing or testing ILs for solvent-based applications.

1.1.2.3 Transport properties

Transport properties are an important aspect to understand for ILs as they describe the movement of ions. Among the various ions studied, lithium ions (Li+) are of particular interest due to their central role in lithium-ion batteries and other electrochemical tech-

I López-Martin et al.[39] define these regions for an imidazolium-based IL, but the concept is transferable to all ILs with alkyl chains, as only the structure of the charge-rich region differs, not the presence of it.

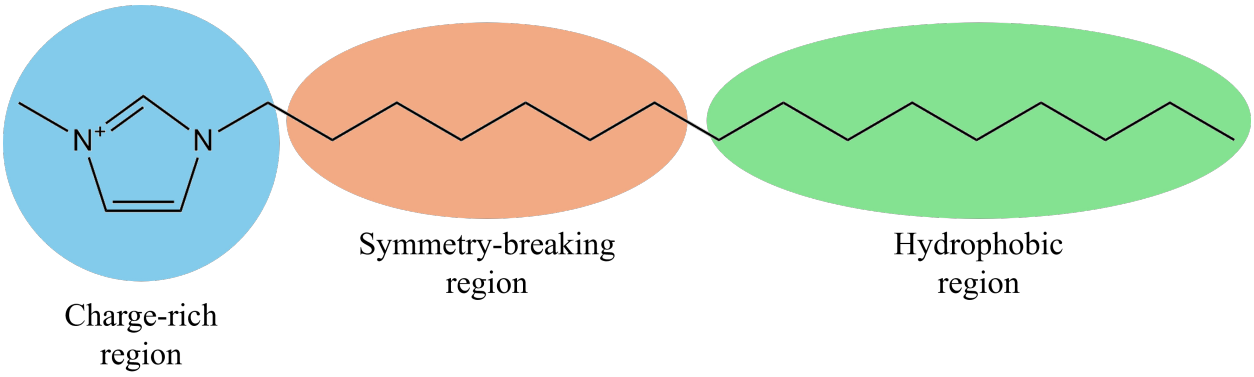


Fig. 1.2. Structural regions of asymmetric cations defined by López-Martin et al.,[39] illustrated for the 1-hexadecyl-3-methylimidazolium cation.

nologies. Understanding the unique transport behaviors of Li+, including its solvation dynamics and mobility in ILs, provides valuable insights into optimizing IL-based systems for practical applications.The three most commonly investigated transport properties for ILs are conductivity, viscosity, and diffusivity. These properties are of particular interest because they can all be interpreted as responses to applied forces. Conductivity reflects the effect of an electric field on ion movement, viscosity quantifies the response to an applied shear force on flow, and diffusivity represents the response to a gradient in chemical potential, which acts as a force. Additionally, these responses are related to the frictional forces that restrict ion movement. Therefore, these three properties provide a theoretical framework for quantifying ion movement.[1]

Conductivity To understand the importance of conductivity in IL, it is necessary to contextualise it in terms of other useful electrochemical properties, as these are used to determine the suitability of an IL for electrolyte applications. One such key property used to quantify the electrochemical stability of ILs is the electrochemical window (ECW). The ECW represents the range of potentials over which an electrolyte can effectively facilitate oxidation and reduction reactions without undergoing undesirable side reactions. Compared to traditional aqueous solutions, ILs exhibit significantly larger ECWs due to the stability of the ion components. The enhanced stability stems from the fact that ILs can

withstand higher oxidative and reductive potentials (higher oxidative and reductive limits) than traditional electrolytes, making them ideal solvents for applications where high stability is required, such as in energy storage devices and electrochemical sensors.[41]

Beyond just the ECW, the electrochemical performance of an IL as an electrolyte is largely dependent on its ability to facilitate the transport of ions between electrodes. Ionic conductivity is a measure of this and more so, is a crucial parameter used to understand the efficiency and effectiveness of electrochemical processes. The conductivity of ILs is typically $\sim$1 mS/cm, approximately one order smaller than that of aqueous salt solutions used in metal-air batteries, or the organic solvents in lithium-ion batteries.[42] However, the thermal stability of ILs is a significant factor driving research into ILs as electrolytes. Therefore, comparable to traditional electrolytes, ILs still possess relatively high conductivity for high-temperature applications.[1] Ionic conductivity not only reflects the rate of ion transport but also provides insights into the aggregation behaviour of ions within the IL. The aggregation of ions can significantly impact conductivity performance. In particular, the degree of ion aggregation affects the accessibility and mobility of charge carriers within the IL. Techniques such as electrochemical impedance spectroscopy and NMR experiments are commonly used to measure and characterise ionic conductivity.[43] While electrochemical impedance spectroscopy provides a direct measurement of the "true" conductivity, NMR experiments allow for the calculation of Nernst-Einstein (NE) conductivity based on ion diffusion. It's important to note that NE conductivity tends to be higher than true conductivity, primarily due to the presence of ion aggregation within the IL.[44] By quantifying the disparity between the NE conductivity and the true conductivity, one can gain valuable insights into the extent of ion aggregation, which is essential for optimising the performance of IL-based electrolyte systems. Overall, understanding and enhancing electrochemical properties such as ECW and ionic conductivity are necessary for harnessing the full potential of ILs in electrochemical applications.

Viscosity While ILs possess numerous favourable properties for electrolyte applications, such as high electrochemical stability and ionic conductivity, the typically higher viscosities of these species present a notable challenge. Viscosity, a measure of a liquid's resistance to flow, is directly related to the other transport properties, in turn affecting these. Therefore, addressing the issue of high viscosity in ILs is necessary for optimising IL performance in ion transport applications.[1]

ILs generally exhibit viscosities significantly higher than those of traditional organic solvents. ILs most commonly explored for electrolyte applications have viscosities ranging between that of sunflower oil (48.8 cP)[45] and maple syrup (138 cP).[46] To provide context, IL viscosities are typically more than ten times greater than that of water (1 cP), which is already significantly more viscous than common organic solvents such as benzene (0.65 cP) and acetone (0.318 cP).[47]

It is not surprising that ILs possess large viscosities as these species often consist of large, strongly interacting ions. Accurate determination of IL viscosity is essential for understanding this property and identifying strategies to reduce it. Lowering viscosity can lead to improvements in transport properties, including enhanced conductivity and ion diffusivity, which are critical for effective electrolyte performance. Therefore, research efforts aimed at reducing IL viscosity are necessary for unlocking the full potential of IL-based electrolyte systems.[1] Various approaches can be explored to decrease IL viscosity, most fundamentally being modifications to the constituent ions. By systematically investigating the factors influencing IL viscosity further advances in the application of ILs in electrochemical technologies can be achieved. Ultimately, a balance between favourable electrochemical properties and manageable viscosity will be key to implementing IL-based electrolyte systems in various applications

Ion diffusivity The transport of the ion in an IL is commonly assessed using two measures: diffusion coefficients and ion transport numbers. Diffusion coefficients provide a

measure of the rate of diffusion of an ion in solution, and ion transport numbers indicate the charge carried by an ion when an electric field is applied, which in an electrolyte would indicate the ability of an ion to carry charge. Diffusion coefficients can be reported for Fickian diffusion or self-diffusion, the latter of which is most commonly reported for ILs. Fickian diffusion is measured using electrochemical methods,[43] requiring a concentration gradient to be present, whereas self-diffusion is measured with NMR methods,[48] without a concentration gradient. Conveniently, these diffusion coefficients are typically quite similar, but given that self-diffusion is typically reported, the discussion shall be centred on this.

Self-diffusion coefficients for molten salts (ie., KNO_3) are of the order of 10^{-9} m^2/s, whereas those for pure ILs fall in the order of 10^{-11} m^2/s. Therefore, this is once again a property of ILs that reduction can improve the transport properties of pure ILs. If the self-diffusion coefficients of both the cation, $D_{s,+}$, and anion, $D_{s,+}$, are known, then the cation and anion transport numbers, t_+ and t_- respectively, can be obtained:[49,50]

$$t_+ = \frac{D_{s,+}}{D_{s,+} + D_{s,-}} \tag{1.5}$$

and

$$t_- = \frac{D_{s,-}}{D_{s,+} + D_{s,-}}. \tag{1.6}$$

As such, accurately obtaining sef-diffusion coefficients from simulation allows for ample interpretation of ion mobility and charge-carrying potential, both important qualities for electrolytes. It should also be noted that while conductivity and viscosity are bulk properties, diffusion is a statistical property, that being it is dependent on each ion individually. Furthermore, there are fundamental relationships that exist between these properties, explored in Chapter 2. As such, these relationships can be used to determine the correlation between bulk and statistical transport properties, allowing for a validation of these properties.

1.1.2.4 Transport of alkali ions in ionic liquids

Among alkali ions, lithium ion (Li^+) transport in ILs is of particular interest due to its crucial role in energy storage applications, such as lithium-ion batteries.[51] The unique properties of Li^+ in ILs, including its strong solvation and complex transport behavior, make it an ideal case study for understanding ion dynamics in these systems. Lithium ion transport in ILs is characterized by distinct solvation and transport properties that set it apart from other alkali ions and conventional electrolytes.[52,53] Due to the strong coordination of Li^+ with oxygen-containing anions such as bis(trifluoromethanesulfonyl)imide ([TFSI]) and bis(fluorosulfonyl)imide ([FSI]), the ion forms tightly bound solvation complexes that significantly affect its mobility.[51,54] MD simulations consistently reveal sharp peaks in radial distribution functions (RDFs) corresponding to Li^+ -O interactions, indicating localized and strong solvation shells.[55] These solvation shells result in an expanded effective hydrodynamic radius for Li^+ , slowing its diffusion and contributing to low self-diffusion coefficients, typically on the order of 10^{-11} m^2/s.[51,56]

One of the most intriguing phenomena observed in ILs is the negative transference number of Li^+, which arises from electrostatic drag effects caused by the tight coordination with the solvating anions.[52,57] Under an applied electric field, these anions move counter to the direction of current, reducing the net movement of Li^+ and generating complex transport dynamics.[58] MD simulations further highlight the role of ion pairing and aggregation, where Li^+ and its solvating anions form clusters that reduce the number of free charge carriers.[52,59]

Despite these insights, MD simulations often underestimate the experimentally observed transport properties due to the exclusion of phenomena such as the solid electrolyte interphase (SEI), and polarisation effects: When evaluating the measured conductivity of ILs, it is crucial to consider the formation of surface species and the SEI. The polarisation effects within the IL during operation, including ion ordering around electrodes and charge screening, contribute to the generation of over-potentials that are often not captured in MD

simulations. This becomes significant as MD simulations of ILs cannot account for the interfacial phenomena that occur in real conditions, which may explain why MD simulations of ”pure” systems do not consistently align with experimental results.[60,61]

1.2 Ionic liquid applications

To begin, the diverse range of potential IL applications arises from the numerous possible structural variations, which lead to differing chemical behaviours. Consequently, ILs have many potential uses, and while the focus here is on ILs as electrolytes, it is worthwhile to highlight some other notable applications.

ILs have been explored as substitutes for traditional organic solvents due to their favourable low volatilities, qualifying these species as viable “green solvents”. The diverse possible compositions of ILs, resulting from various possible modifications, have led to numerous solvation applications. The ionic nature facilitates strong electrostatic interactions both among the constituent ions and between the IL and solute molecules. Functional groups on the cations and anions aid in dissolving polar organic compounds, while long alkyl chains on cations enhance the dissolution of non-polar organic compounds.[1]

One prominent application of ILs is their use in biomass dissolution and swelling applications. Biomass, such as cellulose, chitin and chitosan, serves as a renewable resource for various purposes such as biofuels and biomaterials. ILs have been proposed for extracting and purifying active compounds from biomass, offering a more efficient and environmentally friendly alternative.[62] In biphasic catalysis, ILs can be utilized to facilitate organic product extraction and enable the reuse of the IL solvent.[63] Additionally, ILs exhibit solubility for gases like carbon dioxide, nitrogen, and methane, with a particular focus on capturing carbon dioxide to reduce greenhouse gas emissions.[64]

Energy storage is another emerging field of IL research and application. The high ionic conductivity and electrochemical stability of ILs make them suitable as electrolytes

in high-energy electrochemical devices such as lithium, sodium, magnesium, and metal-air batteries.[20,65,66] ILs have been employed in fuel cells for proton transport, expanding the selection of electrolytes beyond typical aqueous solutions. The large electrochemical window and thermal stability of ILs are advantageous for thermoelectric cells, broadening the temperature range for heat sources.[66]

ILs also serve as effective solvents for material synthesis, especially in the preparation of nanomaterials. The high electrochemical window and ionic conductivity enable the electrodeposition of nanostructures using IL-based electrolytes. Furthermore, ILs can stabilise nanoparticles during synthesis with long alkyl chains on the cation and specific selection of functional groups in both the cation and anion.[67]

While these applications are all of interest in their own right, given that the scope of this thesis places focus on ILs as electrolytes, a more detailed discussion of the uses and considerations of this application is presented here.

1.2.1 Ionic liquids-based electrolytes

While transport properties of ILs are typically less favourable than those of traditional aqueous electrolytes, it is the *thermal* properties of ILs that have generated significant interest in this application. The high thermal stability and negligible volatilities, along with other attractive electrolyte properties such as large ECWs, have led to the proposal of IL-based electrolytes for use in devices such as lithium-, sodium- and magnesium-ion batteries.[20,65,66]

The physicochemical properties of IL-based electrolytes are largely dependent on the choice of cation and anion. When designing IL-based electrolytes, several factors related to the anion structure must be considered. As mentioned earlier, reducing overall intermolecular interactions is crucial in the effective design of IL-based electrolytes. Therefore, the often-encountered concept of a “weakly-coordinating anion” is important. Such anions have greater charge delocalisation leading to overall weaker electrostatic interaction than

small, less-diffuse anions with highly localised charge, as well as a high degree of charge delocalisation and peripheral groups with low basicity.[68,69] The class of imide-type anions has received significant attention in the context of IL-based electrolytes as the combination of these anions with pyrrolidinium-based cations has desired transport and electrochemical properties.[21,70,71] The bis(trifluoromethanesulfonyl)imide ($[TFSI]^-$) anion is arguably the most widely-used imide-based anion. Pyrrolidinium-based $[TFSI]^-$ ILs were introduced by MacFarlane et al.[72] showing notably low melting points. This led to further investigation into $[TFSI]^-$-based ILs,[73,74] revealing that these ILs possess higher conductivity and greater liquid ranges than previous ILs based on smaller, simpler, anions (ie. hexafluorophosphate ($[PF_6]^-$).[72,75] Following this, the first studies of $[TFSI]^-$ ILs as electrolytes for electrochemistry of lithium ions (Li^+) were carried out.[76,77] The use of $[TFSI]^-$ ILs has shown to improve factors of lithium-battery technology such as the electrochemical cycling of Li^+;[78,79] however, the ion transport is typically not optimal and the solid electrolyte interface layers formed often do not result in long-term stability of the Li^+ electrode.[71] Consequently, recently focus has been placed on improving the Li^+ electrode behaviour. ILs containing the bis(fluorosulfonyl)imide, $[FSI]^-$, anion were later proposed as metal salts formed with this anion possess anti-corrosive properties.[80] The conductivity of the [Li][FSI] salt was shown to be greater than that of [Li][TFSI][81] and furthermore, [Li][FSI] formed significantly more stable solid electrolyte interfaces, comparable to that of $[Li][PF_6]$.[81,82] Primarily attributed to the lower viscosity of the $[FSI]^-$-based ILs, these systems show wider liquid ranges as well as improved transport properties compared to the equivalent $[TFSI]^-$ systems.[59,82,83] Unfortunately, the $[FSI]^-$-based ILs showed much lower thermal stability than seen with the $[TFSI]^-$ anion, indicating that the trifluoromethyl groups result in greater stability than a terminal fluorine atom.

At this point, only symmetric anions have been discussed. The group of Matsumoto et al.[84] synthesised the (fluorosulfonyl)(trifluorosulfonylimide) anion, $[FTFSI]^-$, for which impressive physical and electrochemical properties were noted for ILs containing this an-

ion. Structurally, $[FTFSI]^-$ sits between that of $[TFSI]^-$ and $[FSI]^-$, but due to the asymmetry, ILs based on this anion possess significantly larger liquid ranges than the symmetric counterparts, while still maintaining high ionic conductivity — ideal for battery electrolytes.[85] This observation highlights the importance of ion symmetry (or favoured lack thereof) concerning liquid ranges. It is therefore unsurprising that there are further examples of synthesis of asymmetric anions for ILs formed through the substitution of one terminal end of $[TFSI]^-$: The 2,2,2-trifuoromethanesulfonyl-N-cyano imide ($[TFSAM]^-$)[86] and 2,2,2-trifluoro-N-(trifluoromethylsulfonyl)acetamide ($[TSAC]^-$)[87] $[TSAC]^-$ anions were shown to have similar physicochemical properties to $[FTFSI]^-$[88,89] but show inferior electrochemical stability.[90]

Concerning electrolytes, consideration has also been directed towards alignment with environmentally safer and more sustainable alternatives with respect to the charge-carrying ions. IL-based electrolytes containing alkali-metal ions are a worthy landscape to explore, for which targeted research has been carried out with promising outcomes.[54,91–95] Concerning charge-carrying ions, Li-ion technology has been the forerunner in rechargeable batteries, with a range of important applications including portable electronic devices and electric vehicles.[96–98] The increasing demand for Li, paired with the uneven global distribution, has resulted in an inevitable shortage of Li and an increase in the cost associated with Li-ion batteries.[99] Recently, sodium (Na)[100,101] and potassium (K) ion batteries[102,103] have been considered as post-Li technologies, due to a greater global abundance and thus lower associated costs.[99,104,105]

IL-electrolytes have been coined "beyond-lithium" chemistries, due to the ability to form electrolytes based on other ion systems, as mentioned above. Sustainably, Na-ion batteries are the most appealing due to their abundance and, therefore reduced cost. While larger than Li, Na-based systems still possess promising electrolyte properties.[106–109] However, Na-ion batteries have limited energy density due to the high reduction potential of

Na, thus making these systems unsuitable for applications requiring large energy density (ie., in electric vehicles).[110]

The cost downfall of Li-ion batteries coupled with the limited applications of Na-ion batteries fueled the consideration of K-ion batteries. The reduction potential of these systems is comparable to those containing Li,[110] and the high abundance results in lower costs.[111] While these properties are favourable, challenges facing K-ion batteries include relatively low ion diffusivity[112] and poor heat dissipation.[113] Nevertheless, the research into beyond-lithium technologies is a promising avenue, with a large degree of untapped potential. Therefore, the ability to understand the interactions and transport properties of these alkali metals in ILs is an important step towards improving IL-based electrolyte technology.

1.3 Molecular dynamics simulations of ionic liquids

Molecular dynamics (MD) simulation is an important tool with which the structure and transport properties of ILs (and other liquid-phase systems) can be studied. An overview of previous literature in which MD simulations of ILs have been carried out, classified as outlined in Figure 1.3, is presented here. The technical details and background theory of MD simulation are given in Chapter 2.

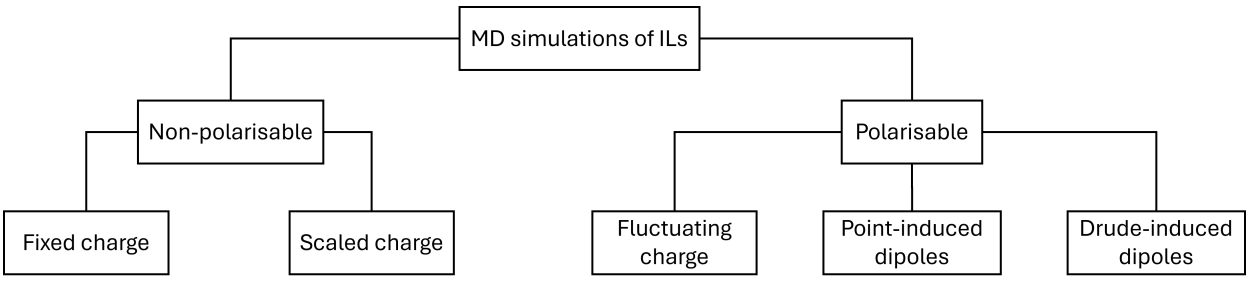


Fig. 1.3. Classification of molecular dynamic simulations, based on the type of force field employed.

This literature review also serves to draw attention to 1) the variation in the potentials used to describe the Van der Waals interactions 2) and the parameters used in different FFs. Furthermore, this section provides familiarity with the typical simulation methodologies followed.

1.3.1 Non-polarisable force fields

1.3.1.1 Fixed-charge force fields

Non-polarisable MD simulations use atomistic FFs in which the atoms are modelled as particles with fixed charges. ILs have been simulated using non-polarisable FFs including the All-Atom Optimized Potentials for Liquid Simulations[114] (OPLS-AA),[115,116] General AMBER Force Field[117] (AMBER/GAFF) and the CHARMM general Force Field.[118] IL-specific non-polarisable FFs have been developed: Hanke et al.[119] were the first to report an IL-specific non-polarisable FF in which potentials for imidazolium-based salts were developed. The dispersion and repulsion terms were described with Buckingham potentials, using parameters either taken from the group of Williams,[120–122] or derived from a 12-6 LJ potential. This model was used to simulate dimethylimidazolium chloride ($[C_1C_1im][Cl]$) and dimethylimidazolium hexafluorophosphate ($[C_1C_1im][PF_6]$). Using approximately five independent simulation trajectories for each of the two systems, average radial distribution functions, densities and self-diffusivity were computed. No comparison to the experiment was made in this work and as such, no comment on FF accuracy was made. Following this, the FF of Morrow and Maginn[123] was developed, using a combination of ab initio and existing parameters (LJ parameters) from the CHARMM 22 protein FF. This was applied to 1-butyl-3-methylimidazolium hexafluorophosphate ($[C_4C_1im][PF_6]$), for which thermodynamic properties were calculated, shown to be within good agreement of experimental values. While self-diffusion coefficients were computed, there was no comparison to experiment available. Liu et al.[124] later refined the AMBER[125] FF and modified bond, torsion

and Van der Waals parameters for the 1-alkyl-3methylimidazolium cations (alkyl = methyl, ethyl and butyl), with the tetrafluoroborate ($[BF_4]^-$), $[PF_6]^-$ and Cl^- anions. This was further extended by Chaban et al.[126] by refining the partial charges of the cation atoms. Later, the group of Canongia Lopes et al.[127] developed one of the most widely used FFs for IL simulation — CL&P. This FF is based on the OPLS-AA[115] FF, extended to include bonded and non-bonded parameters for imidazolium, pyridinium, ammonium and phosphonium-based cations.[40] During the initial development, the FF was validated by simulating 14 imidazolium-based ILs with small symmetrical anions and calculating densities. The results showed good agreement with experimental values, with an average deviation of ~ 3 % from experimentally measured densities. It is worth noting that the CL&P FF makes use of the 12-6 Lennard Jones potential to describe the Van der Waals interactions. The FF was later extended to the commonly used, previously unparameterized triflate ($[OTf]^-$) and $[TFSI]^-$ anions,[128] as well as tri-substituted imidazolium cations.[129] The CL&P FF was designed such that the parameters are transferable between ions based on the same base structure (ie., between imidazolium cations with varying alkyl chain lengths, or for common terminal groups in different anions) and can be used in combination with already existing OPLS-AA parameters. Further refinement of this FF was carried out by Bhargava and Balasubramanian[130] for the $[C_4C_1im][PF_6]$ IL, allowing for densities to be predicted within 1.4 % of experiment. Zhao et al.[131] modified the FF for $[C_4C_1im][TFSI]$ with the introduction of an effective dielectric constant to reduce the electrostatic interactions, a reduction in the radius of the imidazolium ring hydrogen atoms, and the LJ parameters of the oxygen and fluorine atoms of the anion are optimized to reproduce experimental properties. A considerable improvement over the original model of Canongia Lopes et al. was reported. Chaban and Voroshylova[132] then developed a parameterization procedure which accounts for specific cation-anion interactions in the liquid phase, and provide a systematic refinement of the CL&Pol model for the N-butyl-N-methylpyrrolidinium cation ($[C_4C_1pyr]^+$) paired with $[TFSI]^-$, $[FSI]^-$, $[PF_6]^-$, Cl^-, triflate ($[OTf]^-$) and dicyanamide

([DCA]$^-$) anions. For a further detail of the development of CL&P, see Reference[133] and the references within. In the context of application to IL-based electrolytes: Hakim et al.[134] applied the CL&P FF to study [C_2C_1im][FSI] as well as the N-methyl-N-propylpyrollidinium ([C_3C_1pyr]$^+$) cation paired with [FSI]$^-$, both containing the [Na][FSI] salt. Additionally, the system consisting of an equal mixture of [K][FSI] and [Na][FSI] was included in the study. The authors stated that the metal cation parameters were taken from the CL&P FF; however, only Li^+ and Na^+ parameters were available in the FF files supplied by Padua et al. as of the submission of this thesis, and no statement is made regarding the source of the parameters used for K^+. Here, transport (self-diffusivity and ionic conductivity) and structural (density) properties were obtained. The authors highlight the ability of the fixed-charge FF to accurately reproduce the experimental density and conductivity across the investigated systems; however, the large error relative to experiment for conductivity of $\sim$80 % conflicts with the low error of $\sim$4 % for calculated densities. Diffusion coefficients were also computed here, but no comparison was made due to a lack of available experimental measurements.

1.3.1.2 Scaled-charge force fields

While fixed-charge potentials can give a good description of IL structure, the overarching theme across simulation of ILs with non-polarisable FFs is the significant slow-down in calculated transport properties.[126,127,130,135–139] This is attributed to the incomplete description of the interactions, specifically the polarisation effects present in IL systems. An approach to overcome the slow-down of dynamics is through the scaling of the atomic partial charges, motivated by an incomplete description of the interactions (polarisation effects). The first study in which scaled charges were used is that of Morrow and Maginn[123] Using quantum mechanical calculations they found that the cation and anion of [C_4C_1im][PF_6] have a net charge of ± 0.904. Partial charges were adjusted to reflect this and thermodynamic properties were computed showing good agreement with experiment.

Both the values of, and methods used to obtain scaling factors vary, but the factors are typically of similar magnitude: Scaling factors of 0.63–0.9 have been obtained from ab initio calculations of ion pairs.[140–143] Values ranging between 0.75–0.8 were obtained using the refractive indices and dielectric constants[131,144] Last, scaling factors of 0.7-0.9 have also been obtained based on performance in replicating experimental properties.[145–147] Charge scaling has also been introduced to the CL&P FF, where partial charges were scaled by 0.8 for the simulation of imidazolium and pyrrolidinium-based ILs. Densities, self-diffusion coefficients and viscosities were computed, showing that charge scaling improved both transport properties by an order of magnitude, but resulted in less accurate calculated densities.[148]

1.3.2 Polarisable force fields

A problem with scaled-charge FFs is while they do speed up dynamics and improve transport properties, structural properties are also affected, decreasing the structural accuracy to what is achieved with the original FF. Therefore, to accurately capture ion-ion interactions, one needs to include explicit polarisation. While polarisation can be represented with fluctuating charge models, fluctuating charge models have not received large attention in the IL modelling community and have only been applied to limited IL systems.[149,150] PFFs routinely applied to ILs make use of point-induced dipole and Drude-induced dipole models. The technical details of these PFFs are given in Chapter 2, but an overview of the recent literature in which these PFFs have been applied to ILs and IL-based electrolytes is presented next. Additionally, simulations using polarisable FFs require significantly longer computation time that fixed-charge FFs, making the simulation length an important point to consider. As such, these are highlighted in the following sections to provide context as to the typical trajectory lengths used.

1.3.2.1 Point-induced dipole force fields

Of the first polarisable FFs developed for ILs is the polarisable ion model of Yan et al.[151] The $[C_2C_1im][NO_3]$ IL was simulated and viscosities and self-diffusion coefficients were computed from a single 1.2 ns trajectory simulated at 400 K, showing a significant improvement over the values computed with the non-polarisable FF, compared to experiment. The polarisable ion model was further used by Salanne et al.[152] to simulate equimolar mixtures of $[C_2C_1im][AlCl_4]$ and $[C_2C_1im][Cl]$, again producing excellent agreement with experimental conductivities, viscosities and self-diffusion coefficients.

The AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) FF introduces explicit polarisation with atom-centered multipoles, up to quadrupoles, and the buffered Halgren potential — a modified version of the 14-7 Lennard-Jones potential — to describe the Van der Waals interactions,[153–156] with the inclusion of Thole damping functions to avoid the "polarisation catastrophe".[157] The AMOEBA-IL FF is a parameterization of AMOEBA for ILs, developed by the group of Cisneros et al.[158] AMOEBA-IL has been applied to various systems including dodecane mixed with the $[C_1C_1im][BF_4]$ and $[C_2C_1im][BF_4]$, where structural properties were determined from 10 ns trajectories at 300 K.[159] The potential electrolyte system of spyrocyclic pyrrolidinium tetrafluoroborate containing Li-ion salts was also simulated using AMOEBA-IL.[160] Here, the systems were heated to 600 K and a single 20 ns trajectory obtained for each, from which structural and thermodynamic properties were computed. More recently, AMOEBA-IL has been used for the simulation of a comprehensive set of asymmetric imidazolium-based cations paired with simple anions including Cl^-, Br^-, $[NO_3]^-$ and $[BF_4]^-$.[161] These systems were simulated for 10 ns at temperatures of 300–360 K. For these systems, densities and enthalpies of vaporization were estimated within 5 % and 12 % of experiment, respectively, while self-diffusion coefficients were underestimated by an order of magnitude.

The APPLE&P (Atomistic Polarizable Potential for Liquids, Electrolytes and Polymers) FF, developed by the group of Borodin,[162] also includes explicit polarisation through point

induced dipoles. APPLE&P was developed for imidazolium, pyrrolidinium, pyridinium, piperidinium, ammonium and phosphonium-based cations and a range of small symmetrical molecular anions, as well as the $[TFSI]^-$ and $[FSI]^-$ anions.[162] In the original FF, bonded parameters including bond lengths and angles are fit to reproduce quantum mechanics (QM) optimised geometries, angles are fit to QM angle distortions, and dihedral angles are parameterized against gas-phase potential energy surfaces. For non-bonded parameters, APPLE&P makes use of atomic partial charges determined from grid fitting to the molecular electrostatic potential. The repulsion-dispersion parameters are determined by fitting the Buckinham-based potential to IL transport properties including diffusion and conductivity. The APPLE&P FF has been used to simulate IL systems consisting of imidazolium and pyrrolidinium-based cations with the $[TFSI]^-$ anion mixed with [Li][TFSI] salts by Lesch et al.[163] For each system a trajectory was obtained of 60 ns in length, from which diffusion coefficients were computed within ∼10 % of experimental values. Borodin et al.[52] have also used APPLE&P to simulate IL-based electrolyte systems of$[C_4C_1pyr][TFSI]$ containing Li^+, Na^+, Mg^{2+} and Zn^{2+} salts with the $[TFSI]^-$ anion. The Li^+ and Zn^{2+} non-bonded parameters were taken from the authors' previously-developed parameters,[164,165] and parameters for the other two metal cations were developed.[52] Single simulations of ∼10–150 ns were conducted for each system to compute viscosity, self-diffusion and conductivity. Compared to available experimental measurements, the viscosity values showed large deviations, whereas conductivity and self-diffusion coefficients showed to be more consistent with experimental values.[II]

1.3.2.2 Drude polarisable force fields

The first application of a Drude oscillator model to ILs was the study of Schröder and Steinhauser[166] in which $[C_2C_1im][OTf]$ was simulated to investigate dielectric properties. Among Drude-polarisable FFs, two have been developed specifically for ILs. First, is SAPT-

II No values are reported and errors are inferred from the figures supplied by Reference 52.

FF of McDaniel and Schmidt[167] which has been parameterized for various imidazolium-based cations and anions including $[BF_4]^-$, $[PF_6]^-$, $[TFSI]^-$, $[FSI]^-$ and $[OTf]^-$. The bonded parameters of SAPT-FF were taken from the OPLS-AA FF, and non-bonded parameters were determined by matching the non-bonded terms to the equivalent energies obtained SAPT.[168] Here, the Van der Waals potential is based on the Born-Meyer potential. McDaniel et al.[169] used SAPT-FF to simulate systems consisting of the $[CnC_1im]^+$ (n = 2,4,6) cations paired with the $[BF_4]^-$ and $[PF_6]^-$ anions at 298 K and 353 K. Single 50 ns trajectories were generated for these systems, and an additional eight independent 50 ns for the $[C_4C_1im][BF_4]$ system to provide a measure of uncertainty. Densities were reproduced within $\sim$1.5 % of experimental values and enthalpies of vaporisation within $\sim$8 %. Self-diffusion coefficients were also computed, showing a larger deviation from experiment. The replicate trajectories indicated an uncertainty of $\pm$30 % for the calculation of diffusion coefficients. In a later study, McDaniel[170] simulated ILs consisting of the $[C_4C_1im]^+$ cation paired with the $[TFSI]^-$, $[FSI]^-$ and $[OTf]^-$ anions at 300 K. Again, densities, conductivities and enthalpies of vaporisations were computed from single 50 ns trajectories with errors relative to experiment of approx. 3 %, 13 % and 35 %, respectively.

The second IL-specific PFF is the CL&Pol FF Developed by the group of Goloviznina et al..[148] This FF can be used for a large set of cations and anions due to the nature of the FF being built from its non-polarisable predecessor, CL&P.[127] All bonded and non-bonded parameters are carried over from CL&P. To account for the introduction of Drude particles, a "fragment approach" is used to scale the Lennard-Jones potential to remove implicit induction in the potential, rather than parameterizing a Drude-compatible potential from scratch. More detail is given on the parameterization, with a focus on the scaling approach, in Chapter 2. The CL&Pol FF has been developed such that it is easily extendible to new systems and has been applied to ILs consisting of imidazolium, pyrrolidinium, pyridinium and ammonium-based anions with small symmetric anions, as well as imide-based anions.[148] In the initial development, self-diffusion coefficients and viscosities were com-

puted from a single 10 ns trajectory for each system. The calculated properties showed a significantly better agreement with the experimental values than both CL&P and sCL&P; however, viscosities were still estimated with large errors of $\sim$50 %. Philippi et al.[171] used CL&Pol to simulate $[C_4C_1im]^+$-based ILs with anions based on $[TFSI]^-$ to study the role of anion flexibility. A single 20 ns trajectory was used for each system to compute properties at either 298 or 333 K. As the focus of this work probed modification and structural design, little attention was given to transport properties. In a comparison to the CL&P FF, Zhang et al.[172] used CL&Pol to simulate poly(1-butyl-3-methyl-imidazolium hexafluorophosphate) at 600 K. A single 200 ns trajectory was obtained from which densities, cohesive energy densities and self-diffusion coefficients were obtained, for which no comparative experimental data is available. Wang and Maginn[173] have extended the model for eight ILs: $[C_6C_1im][Cl]$, $[C_4C_1im][BF_4]$, $[C_4C_1im][PF_6]$, $[C_2C_1im][BF_4]$, $[C_2C_1im][DCA]$, $[C_4Cpyr][DCA]$, $[C_2C_1im][TFSI]$ and $[C_4C_1pyr][TFSI]$. Simulations were carried out at 303 K to produce a single 30 ns trajectory for each system, from which density, self-diffusion coefficients and radial distribution functions were computed. The maximum deviation for density was 5 % and the mean absolute percentage error for self-diffusion coefficients was 49.5 %, where the diffusion was overpredicted for the imidazolium-based ILs and underpredicted for the pyrrolidinium-based ILs.

The CL&Pol FF has also been extended to the protic ionic liquid ethylammonium nitrate, a deep eutectic solvent consisting of a mixture of choline chloride and ethylene glycol, and the electrolyte system of $[C_2C_1im][TFSI]/[Li][TFSI]$.[174,175] The significance of this extension is the implementation of Tang-Toennies (TT) damping functions to deal with the short-range problems of small, highly charged ions or strong hydrogen bonding as present in these systems, as they are subject to over-correlation between DPs. Once again, a single 10 ns trajectory for each system was used to compute densities, self-diffusion coefficients (no experimental data available) and viscosities. Compared to experiment, the viscosities still showed large deviations, while densities were computed with high accuracy,

and the improvements of the TT functions were demonstrated. This provided the framework for sound simulation of IL-based electrolytes. As such, the CL&Pol FF has emerged as a desirable FF for the simulation of IL systems due to its easily extendable nature. While the TT damping was recommended in the 2021 work of Goloviznina et al.,[174] the CL&Pol FF was used by Massaro et al.[176] in 2020 where the dynamics of the N-methyl-N-propylpyrrolidinium bis(fluorosulfonyl)imide IL mixed with varying concentrations of sodium bis(fluorosulfonyl)imide salts ([C_3C_1pyr][FSI]/[Na][FSI]) was investigated. The systems were simulated at 353 K and a single 8 ns trajectory was obtained for each system. While no TT damping was applied, CL&Pol still outperformed both CL&P and sCL&P in reproducing densities and self-diffusion coefficients. It is worth highlighting that the non-bonded metal parameters were taken from Yu et al.,[177] rather than using OPLS parameters, consistent with the existing parameters used by CL&P. The work of Kubisiak et al.[178] used a combination of parameters with a Drude-based FF, where the non-bonded parameters are taken from CL&P, but atomic polarisabilities and charges are adapted from APPLE&P. It is worth noting that this work was published before the development of CL&Pol and as such the prescribed parameterization protocol involving scaling of LJ parameters was not available yet. Here, the [C_2C_1im][TFSI]/[M][TFSI] (M = Li^+ and Na^+) were simulated at 333 K, for a notably long simulation time of 1100 ns using the parameters of Jensen[179] for the metal cations.

The literature here thus highlights the variability in the length of the simulation trajectory used to compute properties. Furthermore, a lack of reporting of uncertainty is seen more often than not, removing any measure of reproducibility. Moreso, the literature highlights that the studies looking at FF performance typically focus on the accuracy of the properties calculated, relative to experiment. Rather, as pointed out by Maginn et al.,[180] a focus should be placed on the reliability of the simulation protocol. Additionally, the literature outlined above indicates the variability in potentials and parameters used to describe

the induction, dispersion and exchange interactions, which can often lead to inconsistent use of parameters or a lack of indication of where the parameters are taken from.

1.4 Aims and objectives

This work aimed to make several contributions to the field of polarisable MD simulations. Given the importance of quantifying the uncertainty of MD simulations,[180] coupled with the clear lack of uncertainty reported for previous simulations of IL, this project aimed to develop a robust and reliable simulation protocol using the CL&Pol FF. Secondly, the description of the Van der Waals interactions in the framework of a polarisable FF varies based on the choice of potential and parameters. Thus, this inspired the second aim of this work: to provide a stable and robust protocol for obtaining Van der Waals potential parameters compatible with a Drude FF without the need for scaling, based on first-principles resolution of the dispersion and exchange components. The above aims were achieved with the following objectives:

1. Apply a well-defined and cost-effective simulation protocol for the calculation of structural and transport properties of ILs and IL-based electrolytes;

2. Validate the accuracy of the CL&Pol FF and provide a measure uncertainty associated with calculated properties;

3. Develop a method for obtaining cost-effective interaction energy components, void of basis-set incompleteness, based on Symmetry-Adapted Perturbation Theory, to be used to

4. Improve the CL&Pol description of the dispersion and exchange interactions with physically-motivated parameters, compatible with IL-electrolyte systems.

1.5 Thesis outline

The findings of this thesis are presented in two parts, each pertaining to a major aim of the thesis. The first aim was to validate the CL&Pol FF. This required the development of a robust simulation protocol, from which the calculation of various structural and transport properties can be obtained. Chapter 3 lays the foundation of this by detailing the property calculation methodologies employed throughout the research. This chapter provides a comprehensive guide to the methodology and algorithms used to compute properties for the systems studied in the subsequent chapters.

In Chapter 4, the CL&Pol FF is applied to a systematically chosen set of ILs consisting of imidazolium and pyrrolidinium-based cations paired with the imide-type anions. The validation process involved comparing the simulated properties with experimental data to ensure that the force field can accurately predict the behaviour of these ILs. The precision of the calculated properties is indicated by reporting a 95 % confidence interval of the mean. The results presented in this chapter demonstrate the reliability of the CL&Pol FF in modelling ILs, thereby establishing a solid basis for further applications.

Chapter 5 extends the methodology to a broader range of systems where mixtures of ILs and alkali-metal salts are used to form IL-based electrolytes. The properties of these systems are calculated using the established simulation protocol. Additionally, the effect of polarisation on the properties of these systems is thoroughly investigated, providing new insights into how polarisation influences the behaviour of IL-based electrolytes.

The second aim was to develop robust non-bonded parameters for a Van der Waals potential, compatible with the CL&Pol Drude framework. This first required the development of a reliable parameterization protocol. Creating physically meaningful parameters required a method to obtain highly accurate energy components against which the Van der Waals potential could be parameterized. Chapter 6 presents the development of a methodology to compute highly accurate SAPT energy components, void of basis-set incompleteness errors.

Chapter 7 then applies this methodology to obtain the exchange and dispersion energy components, against which the induction-free Van der Waals potential could be parameterized. This was first applied to metal cations and then extended to include parameters for simple IL components.

Chapter 2

Theoretical background

Chapter 2 provides the fundamental theory used in the molecular dynamics simulation and quantum chemistry computations applied in this work. First, molecular dynamics simulations, with explicit details of the relevant theory are presented. The section on classical force fields also includes an overview of the classical description of intermolecular interaction. This is followed by a general overview of the quantum mechanics theory applied throughout. Last, the Symmetry-Adapted Perturbation Theory formalism is detailed, which provides the basis of the parameterization work carried out in this thesis.

2.1 Molecular dynamics

Molecular dynamics (MD) simulations use Newton's equations of motion to solve the time evolution of a system of atoms or molecules. The velocity and positions of the elements of the system can be obtained from Newton's second law if the total force on the system is known, which can be determined from the negative gradient of the potential energy of the system. As such, an accurate expression for the potential energy is required, described in MD simulation with a force field. Thus, it is the chosen force field that will govern the bonded and non-bonded interactions within a simulated system, represented as a simulation box (typically cubic) containing thousands of atoms. While this provides a reasonable

representation of the interactions, a box of finite size will inevitably have artificial boundaries, affecting the behaviour of atoms when reaching this boundary. As such, techniques have been developed to correct for this, allowing for the simulation of an infinite periodic lattice. Thus, the trajectories obtained from such MD simulations can be used to determine thermodynamic properties, and using statistical mechanics, measurable macroscopic properties can be obtained. Therefore, based on the required properties, an appropriate statistical ensemble in which macroscopic properties are kept constant must be generated. Typical ensembles used for MD simulations include the canonical ensemble with a fixed number of atoms, volume and temperature (*NVT*); the microcanonical ensemble, with a fixed number of atoms, volume and energy (*NVE*); and the isobaric isothermal ensemble, with a fixed number of atoms, pressure, and temperature (*NPT*).

The following sections give an overview of the aspects required for equilibrium molecular dynamics (MD) simulation.[181–183]

2.1.1 Solving the equations of motion

By applying Newton's second law of motion, a trajectory can be obtained containing all atomic positions and velocities at each time step in the simulation. Trajectories of the simulation are obtained by solving the following:

$$m_i \ddot{\mathbf{r}}_i = \mathbf{f}_i, \tag{2.1}$$

where $\ddot{\mathbf{r}}$ is the second time derivative of position (acceleration) and $\mathbf{f}_i$ is the force exerted by particle i, given by

$$\mathbf{f}_i = -\boldsymbol{\nabla}_{r_i} \boldsymbol{\mathcal{V}}, \tag{2.2}$$

with $\boldsymbol{\mathcal{V}}$ being the potential energy of the system. Equations 2.1 and 2.2 are solved using a *finite difference approach*.

All algorithms that integrate the equations of motion with the finite difference approach assume that change in position, velocity, acceleration and other dynamic properties between a time t and $t + \delta t$, where δt is the chosen timestep, can be approximated from a truncated Taylor series expansion. The *velocity-Verlet algorithm*[184] is the most widely used of such integration methods, where the position and velocity are updated as

$$\dot{\mathbf{r}}(t + \frac{1}{2}\delta t) = \ddot{\mathbf{r}}(t) + \frac{1}{2}\delta t\ddot{\mathbf{r}}(t), \tag{2.3}$$

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t\dot{\mathbf{r}}(t + \frac{1}{2}\delta t) \tag{2.4}$$

and

$$\ddot{\mathbf{r}}(t + \delta t) = \ddot{\mathbf{r}}(t + \frac{1}{2}\delta t) + \frac{1}{2}\delta t\ddot{\mathbf{r}}(t + \delta t) \tag{2.5}$$

If $\boldsymbol{\mathcal{V}}$ is known, Equation 2.2 can be used to obtain $\mathbf{f}_i$ and then Equation 2.1 can be used to calculate $\ddot{\mathbf{r}}$, from which a time-dependent equation of motion, describing the evolution of atomic positions can be constructed. Therefore, MD simulations require expressions that describe the intra- and intermolecular components of $\boldsymbol{\mathcal{V}}$, referred to as a *force field*.

2.1.2 Force fields

The functional form of a force field (FF) requires atom-specific parameters which are derived from either experimental data or quantum mechanical calculations and are typically developed to allow for transferability between atoms of the same type (same chemical environment) in different molecules. The FFs implemented in this work are based on the

OPLS FF,[114,185] taking the functional form of

$$\begin{aligned}\boldsymbol{\mathcal{V}} =& U_{\text{bonds}} + U_{\text{angles}} + U_{\text{torsions}} + U_{\text{non-bonded}} \\ =& \sum_{ij}^{\text{bonds}} \frac{k_{\text{r},ij}}{2} \left(r_{ij} - r_{0,ij}\right)^2 + \sum_{ijk}^{\text{angles}} \frac{k_{\theta,ijk}}{2} \left(\theta_{ijk} - \theta_{0.ijk}\right)^2 + \sum_{ijkl}^{\text{torsions}} \sum_{n} k_{\psi,ijkl} \left[\cos(n\psi_{ijkl} + \delta_n) + 1\right] \\ &+ \sum_{i}^{\text{atoms}} \sum_{j>i} f_{ij} \left(U_{\text{vdW}} + U_{\text{elec}}\right).\end{aligned} \tag{2.6}$$

The sum of the energy over all bonds, U_{bonds}, is calculated for every pair of directly bonded atoms i and j, using the harmonic potential shown, where $k_{\text{r},ij}$ is the bond-force constant and r_{ij} and $r_{0,ij}$ are the reference and equilibrium bond lengths, respectively. The energy due to angle changes, U_{angles}, between bonded atoms i, j and k is calculated over all angles, with the angle-force constant $k_{\theta,ijk}$ using the reference angle, θ_{ijk}, and equilibrium angle, $\theta_{0,ijk}$. The last of the bonded contributions is the energy due to torsional motion, U_{torsion}, where the rotation-barrier height is indicated by $k_{\psi,ijkl}$. The multiplicity, n, is the number of minima in a single rotation around the *j*–*k* bond and δ_{n} is the torsion angle where ψ_{ijkl}, is a minimum.

The remaining contribution to $\boldsymbol{\mathcal{V}}$ is the energy due to non-bonded interactions, $U_{\text{non-bonded}}$, which has both inter- and intramolecular contributions. The classical non-bonded interaction consists of short-range (or steric) repulsion, electrostatic interaction, induction, and dispersion.[186] Short-range repulsion results from the overlap of electron clouds at small separation distances, which leads to a strong repulsive force driven by the exchange interaction that forms the basis of the Pauli exclusion principle. Because the overlap between two electron clouds decreases exponentially with increasing separation between the atoms, the most physically realistic description of short-range exchange

repulsion requires a functional form such as the Born-Mayer potential,[187]

$$U_{\text{BM}} = A \exp(-B r_{ij}) \tag{2.7}$$

in which A and B are empirical constants. However, for computational convenience, the exponential function can be avoided by using a function that is proportional to $\frac{1}{r^n}$, with $n \geq 12$. Electrostatic interactions result from the static charge distributions of interacting molecules. Full expressions for every contributing term can be obtained from the multipole expansion.[186] In systems consisting of charged polar molecules, the leading terms are charge-charge, charge-dipole, charge-quadrupole, etc. interactions. However, the first nonzero contribution in a system of charge-neutral polar molecules is the dipole-dipole interaction energy,[186,188]

$$U_{\text{dipole-dipole}} = -\frac{\mu_i \mu_j}{4\pi\epsilon_0 r_{ij}^3} \left(2\cos\theta_i \cos\theta_j - \cos\theta_i \cos\theta_j \cos\varphi\right), \tag{2.8}$$

where μ_i and μ_j are the dipole moments of molecules i and j, ε_0 is the vacuum permittivity, and θ_i, θ_j and φ are angles describing the orientation of the dipole moment vectors. However, in the condensed state, the orientation of molecules changes rapidly, and the angle-averaged interaction energy, also known as the Keesom interaction, provides a better description of the dipole-dipole interaction:[188]

$$U_{\text{Keesom}} = -\frac{2}{3} \frac{\mu_i \mu_j}{(4\pi\epsilon_0)^2 r_{ij}^6 kT}, \tag{2.9}$$

where k_B is the Boltzmann constant and T the temperature at which the Boltzmann-weighted averaging of the angles is done. Induction effects result from the distortion of a molecule in response to the electric field of all surrounding molecules and are always attractive. The angle-averaged dipole-induced dipole interaction energy, also known as the Debye interaction, between two polar molecules possessing permanent dipole moments μ_i

and μ_j, and polarisabilities α_i and α_j, is[188]

$$U_{\text{Debye}} = -\frac{\left[\mu_i^2\alpha_i + \mu_j^2\alpha_i\right]}{\left(4\pi\epsilon_0\right)^2 r_{ij}^6}. \tag{2.10}$$

Lastly, the dispersion energy is an attractive interaction as a result of the temporary fluctuations in electron distributions and will be present in the interaction between polar or nonpolar molecules. Dispersion can be quantified using the London formula,[186,188]

$$U_{\text{London}} = -\frac{3}{2}\frac{I_i I_j}{\left(I_i + I_i\right)}\frac{\alpha_i\alpha_j}{\left(4\pi\epsilon_0\right)^2 r_{ij}^6}, \tag{2.11}$$

where I_i and I_j are the first ionization potentials of molecules i and j, respectively. Note that Equations 2.9, 2.10 and 2.11 share a $\frac{1}{r^6}$-dependence on distance and can therefore be unified into a single expression that is collectively referred to as the *Van der Waals interaction*.

In a classical force field, the short-range exchange repulsion and dispersion contributions are typically grouped together into one potential function, one of the most popular that is also used in OPLS-type force fields, is the 12-6 Lennard-Jones potential,

$$U_{\text{LJ}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \tag{2.12}$$

in which σ_{ij} is the interatomic distance corresponding to a zero potential, and ϵ_{ij} is the minimum energy, or well depth, of the potential. The first term represents short-range exchange repulsion, and the functional form of the last term is based on the $\frac{1}{r^6}$-dependence of the dispersion energy. Note that by construction, the second term formally captures only the dispersion interaction (Equation 2.11), but if the values of ϵ_{ij} and σ_{ij} are empirically determined in a fixed-charge, non-polarisable force field, may implicitly account for the Debye interaction (Equation 2.10) as well. Consequently, the LJ potential is often referred to as a *Van der Waals potential*, which includes exchange repulsion in this definition. If

the potential is evaluated between atoms of different type, the OPLS FF implements the following mixing rules, based on geometric averaging, are used to obtain these values such that:

$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j} \tag{2.13}$$

and

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}, \tag{2.14}$$

where the σ and ϵ values that are averaged are atom-specific parameters obtained through parameterization. All remaining electrostatic interactions are typically calculated using an atom-centered expression of Coulomb's Law,

$$U_{\text{Coul}} = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}, \tag{2.15}$$

where q_i and q_j are the point charges on atoms i and j, respectively. The summation over all atom-atom pairs represents a discrete version of the multipole expansion that accounts for charge-charge (or monopole-monopole) interaction between non-neutral molecules, as well the coupling between higher-order multipoles. The initial atomic charges determined for the OPLS FF were either obtained from quantum mechanical computations or empirical fitting against experimental data.[114,185] Finally, the scaling factor f_{ij} in Equation 2.6 is dependent on the separation between atoms i and j: The Coulomb and LJ interactions between atoms separated by one or two bonds, referred to as 1-2 and 1-3 interactions respectively, are excluded in the potentials ($f_{ij} = 0$), and interactions between atoms separated by three bonds (1-4 interactions) are reduced by half $f_{ij} = 0.5$. All other interactions are left unscaled ($f_{ij} = 1$).

2.1.3 Solving long-range interactions

To realistically simulate a liquid, upward of several hundred molecules would need to be included in a simulation box, while this is not feasible and finite boundary conditions are unavoidable, *periodic boundary conditions* are typically employed: A simulation cell (box), containing the molecules to be simulated is generated, and then duplicated surrounding the initial (central) simulation cell in all direction, and then duplicated again surrounding the next existing cells, and so on, generating periodic images of the central cell. Every cell has the same atoms moving and interacting identically. As such, when a molecule leaves a boundary of the central cell, its image will re-enter the cell from the neighbouring image cell. This process is referred to as *imaging*.

The intermolecular interactions are described by the LJ potential (Equation 2.12) and the Coulomb potential (Equation 2.15). Considering the power of the distance dependence of each potential, the LJ potential will quickly approach zero as r_{ij} increases, whereas the Coulomb interactions will still be present at long range. As such, reference to long-range interactions is only considering electrostatics. The long-range interactions are problematic for MD simulations, as to be able to truly describe interactions acting over an infinite range would require an infinite number of simulation cells. Realistically, most interactions become insignificant over a large enough finite range, and a large enough simulation box could be sufficient to describe the long-range interactions. However, increasing the box size quickly increases the computational resources required and becomes rapidly unfeasible. That being if all pair-wise interactions are accounted for explicitly, MD simulations will typically scale by $\mathcal{O}(n^2)$.

Electrostatic interactions are best treated with the *Ewald summation*.[189–191] The basis of this approach is to consider the interaction of one molecule with all other molecules in the box as well as with all molecules in an infinite number of periodic images. The

electrostatic energy can then be written as

$$\boldsymbol{\mathcal{V}}^{qq} = \frac{1}{4\pi\varepsilon_0} \sum_{\mathbf{m}\in\mathbb{Z}^3}^{\infty} \left(\sum_{i,j>1}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{m}\mathrm{L}|} \right), \tag{2.16}$$

with L being the length of the cubic box and $\mathbf{m}$L being the position at $(m_x\mathrm{L}, m_y\mathrm{L}, m_x\mathrm{L})$. This summation has an extremely slow convergence, and the Ewald method transforms $\boldsymbol{\mathcal{V}}^{qq}$ to ensure rapid and absolute convergence. The idea here is to build an infinite system: A large sphere of periodic boxes is generated and placed in a perfect conductor where the dielectric constant $\varepsilon = \infty$. Then, the continuous charge distribution around a fixed charge q_i at position $\mathbf{r}$ can be expressed as

$$\rho_i^q(\mathbf{r}) = q_i \kappa^3 \exp\left(-\kappa^2 r^2\right)/\pi^{3/2}, \tag{2.17}$$

where κ is a parameter controlling the width of the distribution. Next, a charge distribution of the same magnitude as $\rho_i^q(\mathbf{r})$ and the *same* sign as q_i, referred to as a "cancelling distribution", is added to cancel the potential due to the original set of charges. Then, the interaction between the cancelling distribution and the original charge at position $\mathbf{r}$ (the self term) must be subtracted. This leads to a potential, $\boldsymbol{\mathcal{V}}^{qq}(\varepsilon = \infty)$, consisting of a real space contribution,

$$\boldsymbol{\mathcal{V}}_1 = \frac{1}{4\pi\varepsilon_0} \sum_{i,j>1}^{N} q_i q_j \sum_{\mathbf{m}\in\mathbb{Z}^3}^{\infty} \frac{\mathrm{erfc}(\kappa|\mathbf{r}_{ij} + \mathbf{m}L|)}{|\mathbf{r}_{ij} + \mathbf{m}L|}, \tag{2.18}$$

where

$$\mathrm{erfc}\,(x) = \left(2/\pi^{1/2}\right) \int_x^{\infty} \exp\left(-t^2\right) \mathrm{d}t; \tag{2.19}$$

and a reciprocal contribution

$$\boldsymbol{\mathcal{V}}_2 = \frac{1}{4\pi\varepsilon_0} \sum_{i,j>1}^{N} q_i q_j \frac{1}{\pi L^3} \sum_{k\neq 0} \frac{4\pi^2}{k^2} \exp\left(-k^2/4\kappa^2\right) \left|\hat{\rho}^q(\mathbf{k})\right|^2, \tag{2.20}$$

with the Fourier transform of the charge density defined as

$$\hat{\rho}^q(\mathbf{k}) = \sum_{i=1}^{N} q_i \exp\left(-\mathrm{i}\mathbf{k}\cdot\mathbf{r}_i\right) \tag{2.21}$$

and the self term

$$\boldsymbol{\mathcal{V}}_3 = \frac{\kappa}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \tag{2.22}$$

such that

$$\boldsymbol{\mathcal{V}}^{qq}(\varepsilon_\mathrm{s} = \infty) = \boldsymbol{\mathcal{V}}_1 + \boldsymbol{\mathcal{V}}_2 - \boldsymbol{\mathcal{V}}_3. \tag{2.23}$$

Lastly, as the problem being solved is the long-range electrostatic interactions, one must consider the infinite limit this would be between the charges in a vacuum ($\varepsilon_\mathrm{s} = 1$). The potential defined thus far has been for a system in a conductor; however, at the limit of a very large sphere, the following relationship can be applied:

$$\boldsymbol{\mathcal{V}}^{qq}(\varepsilon_\mathrm{s} = \infty) = \boldsymbol{\mathcal{V}}^{qq}(\varepsilon_\mathrm{s} = 1) - \frac{2\pi}{3L^3}\left|\sum_i q_i \mathbf{r}_i\right|^2 .^{192} \tag{2.24}$$

Thus, the full expression for the electrostatic energy using the Ewald summation is given as

$$\begin{aligned}\boldsymbol{\mathcal{V}}^{qq}(\varepsilon = 1) =& \frac{1}{4\pi\varepsilon_0} \sum_{i,j>1}^{N} q_i q_j \left(\sum_{\mathbf{m}\in\mathbb{Z}^3}^{\infty} \frac{\mathrm{erfc}(\kappa|\mathbf{r}_{ij} + \mathbf{m}L|)}{|\mathbf{r}_{ij} + \mathbf{m}L|} \right. \\ &\left. + \frac{1}{\pi L^3} \sum_{k\neq 0} \frac{4\pi^2}{k^2} \exp\left(-k^2/4\kappa^2\right) \left|\hat{\rho}^q(\mathbf{k})\right|^2 \right) \\ &- \frac{\kappa}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 + \frac{2\pi}{3L^3} \left|\sum_{i=1}^{N} q_i \mathbf{r}_i\right|^2 .\end{aligned} \tag{2.25}$$

If a sharp charge distribution (large value of κ) is used, the first term of the sum will only be dependent on the original unit cell. Since the first term (real space) is only the minimum sum, enough k-vectors must be included such that the total sum will result in

the correct energy. Therefore, the second term is referred to as the long-range (k-space) contribution to the electrostatic energy.

The computational cost can be reduced for the Ewald summation by using methods such as Particle-Mesh Ewald summation (PME) where particles are mapped onto a grid and the potential is determined by solving the Poisson equation. An issue with the PME method is that it introduces significant errors for close-range force calculations. This can be remedied with the Particle-Particle-Particle-Mesh Ewald summation (PPPM) where the short-range forces are calculated from particle-particle interactions, and the PME method is only applied for long-range interactions. With these methods the Poisson equation is solved using a fast Fourier transform, increasing computational efficiency from $\mathcal{O}(n^2)$ to $\mathcal{O}(n\log(n))$.

2.1.4 Polarisable force fields

A limitation of FFs implemented only as described in Section 2.1.2, is that atoms are treated as fixed point charges and there is no account of polarisation effects, such that there is no change in charge as a response to the local environment of the atoms. Scaled-charge FFs do provide an implicit account of polarisation by reducing the magnitude of atomic charges, approximating the dampening of electrostatic interactions caused by polarisation. However, to accurately reproduce polarisation effects, these need to be included explicitly. Polarisable FFs typically represent polarisation with fluctuating charge,[193–195] point-induced dipoles[153,154,183] or Drude-induced dipoles,[196–199] with the fundamental differences between these models and that of classical FFs illustrated in Figure 2.1.

2.1.4.1 Fluctuating charge model

The fluctuating charge model[193–195] introduces explicit polarisation by allowing the partial charges of atoms to be responsive to environmental changes. For an atom i, the energy

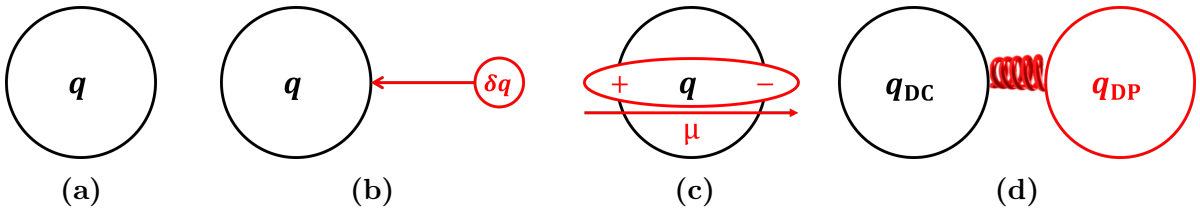


Fig. 2.1. Illustrations of charge description implemented in the **(a)** fixed charge, **(b)** fluctuating charge, **(c)** point-induced dipole and **(d)** Drude-induced dipole force field models. Here, q is the atomic partial charge, δq is the change in q at each minimization, μ is the induced dipole, and q_{DC} and q_{DP} are the Drude core and Drude particle charges, respectively; further definition is given in Sections 2.1.4.1, 2.1.4.2 and 2.1.4.3.

needed to create a partial charge, q_i, can be written as a Taylor series expansion such that

$$U(q_i) = U_i(0) + \chi_i^0 q_i + \frac{1}{2} J_{ii} q_i^2, \tag{2.26}$$

where $U_i(0)$ is the ground state energy of i and χ_i^0 is the corresponding electronegativity, defined as[200]

$$\chi_i = \frac{\partial U_i(0)}{\partial q_i}, \tag{2.27}$$

and $\frac{1}{2} J_{ii}$ is the hardness.[201] Therefore, the total potential energy of a system of N atoms can be written as

$$U_{\text{pot}} = \sum_{i=1}^{N} \left(U_i(0) + \chi_i^0 q_i + \frac{1}{2} J_{ii} q_i^2 \right) + \sum_{i=1}^{N} \sum_{j>i}^{N} J_{ij}(r_{ij}) q_i q_j. \tag{2.28}$$

Using the relationship between electronegativity and potential energy, the partial charge of i is then computed through the minimization of χ_i, such that

$$\chi_i = \frac{\partial U_{\text{pot}}}{\partial q_i} = \chi_i^0 + J_{ii} q_i + \sum_{i=1}^{N} \sum_{j>i}^{N} J_{ij}(r_{ij}) q_j = 0. \tag{2.29}$$

Therefore, each time U_{pot} is updated (the system environment changes) χ_i can be minimized, with the overall charge remaining constant, controlled through constraining the

system such that

$$\sum_{i}^{N} q_i = q_{\text{tot}}. \tag{2.30}$$

Thus, a fluctuation in charge, while keeping the overall charge fixed, is created in the system as a result of interaction with the local environment.

2.1.4.2 Point-induced dipole model

The point-induced dipole model[153,154,183] introduces polarisation by replacing the fixed partial charge q_i with an induced dipole, $\boldsymbol{\mu}_i$, calculated as

$$\boldsymbol{\mu}_i = \boldsymbol{\alpha}_i \cdot \mathbf{E}_i, \tag{2.31}$$

where $\boldsymbol{\alpha}_i$ is the atomic polarisability tensor and $\mathbf{E}_i$ is the electric field at the site. The electric field of a dipole will have contributions from the permanent charges, $\mathbf{E}_i^0$, and the other induced dipoles in the system such that

$$\mathbf{E}_i = \mathbf{E}_i^0 - \sum_{j \neq i} \mathbf{T}_{ij} \boldsymbol{\mu}_{ij}. \tag{2.32}$$

Here, $\mathbf{T}_{ij}$ is the dipole field tensor which describes the interaction between induced dipoles, calculated as

$$\mathbf{T}_{ij} = \frac{1}{\mathbf{r}_{ij}^3}\mathbf{I} - \frac{3}{\mathbf{r}_{ij}^5} \begin{pmatrix} x^2 & xy & xz \\ yx & y^2 & yz \\ zx & zy & z^2 \end{pmatrix}, \tag{2.33}$$

with $\mathbf{I}$ the identity matrix, $\mathbf{r}_{ij}$ the distance between atoms i and j, and (x, y, z) the components of the Cartesian vector between i and j. The energy of the induced dipoles, U_{ind}, can then be expressed in terms of three contributions:

$$U_{\text{ind}} = U_{\text{stat}} + U_{\mu\mu} + U_{\text{pol}}. \tag{2.34}$$

The first contribution, U_{stat}, is due to the interaction of N induced dipoles with the static fields of the surrounding permanent charges at each charged site,

$$U_{\text{stat}} = -\sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \boldsymbol{E}_i^0. \tag{2.35}$$

The energy due to the induced dipoles interacting with each other, $U_{\mu\mu}$, is calculated as

$$U_{\mu\mu} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j>i} \boldsymbol{\mu_i} \cdot \mathbf{T}_{ij} \cdot \boldsymbol{\mu}_j, \tag{2.36}$$

and the energy due to polarisation, U_{pol}, is

$$U_{\text{pol}} = \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{\mu}_i \cdot \mathbf{E}_i. \tag{2.37}$$

Therefore, obtaining atomic polarisabilities enables the introduction of point-induced dipoles, which can theoretically be expanded to include higher-order multipoles.

2.1.4.3 Drude-induced dipole model

The Drude-induced dipole model, or Drude oscillator model[196–199] describes polarisation by replacing the single point charge of atom i at $\mathbf{r}_i$ with a pair of point charges connected by a harmonic spring with an equilibrium distance of zero. By convention, the pair consists of a Drude core (DC) at $\mathbf{r}_i$ joined to an oppositely-charged Drude particle (DP). The magnitude of the DP charge, $q_{\text{D},i}$ is determined from the atomic polarisability of i, α_i,

$$\alpha_i = \frac{q_{\text{D},i}^2}{k_{\text{D}}}, \tag{2.38}$$

where k_{D} is the force constant of the harmonic spring and is typically set to 4184 kJ Å^2/mol for all atoms. The DC charge, q_i, is then calculated as the difference between the original charge of atom i, q_i', and $q_{\text{D},i}$. The values of q_i and $q_{\text{D},i}$ remain fixed and polarisation results

due to the displacement of the DP from the DC by a distance $\mathbf{d}_i$, inducing a dipole moment,

$$\boldsymbol{\mu}_i = q_{\mathrm{D},i}\mathbf{d}_i. \tag{2.39}$$

This is illustrated for a nitrogen atom within the 1,3-dimethylimidazolium cation in Figure 2.2.

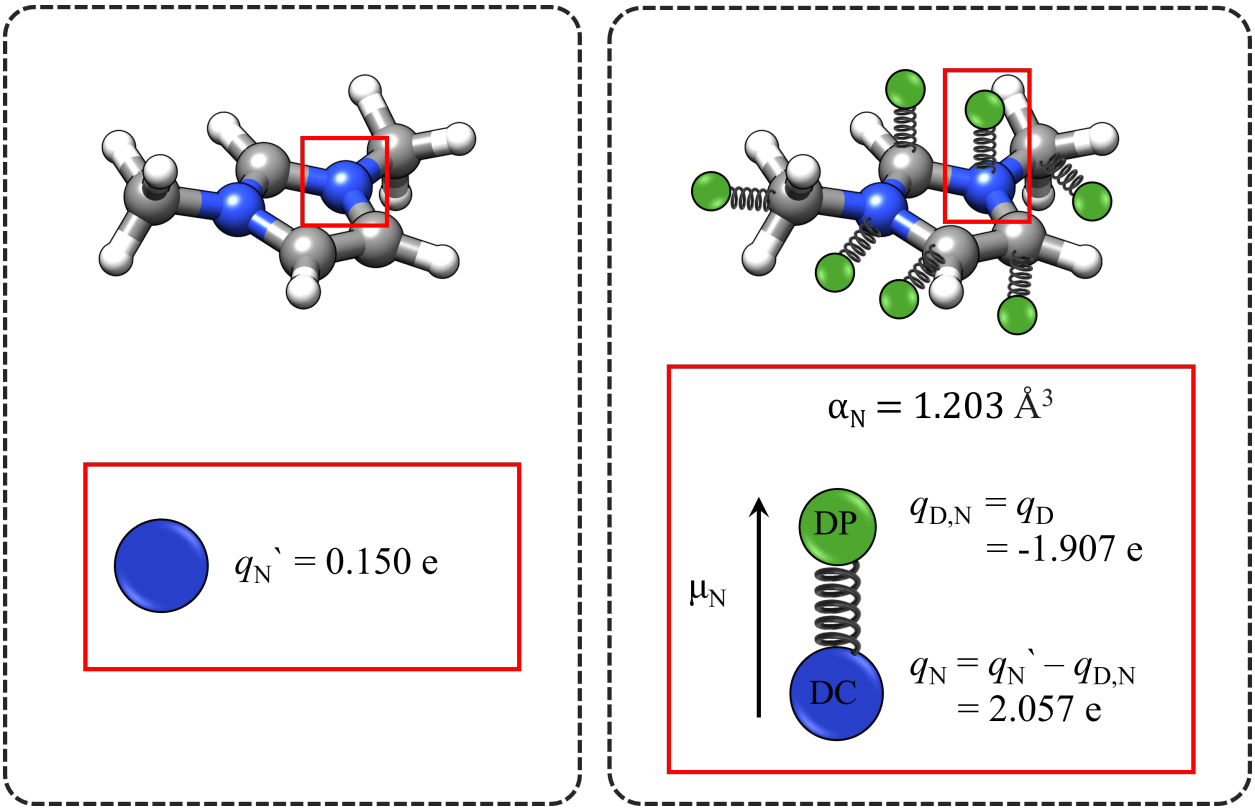


Fig. 2.2. Illustration of the transformation of a non-polarisable FF (left) to a Drude-polarised FF (right) for a nitrogen atom in the 1,3-dimethylimidazolium cation. Here the atomic polarisability, α_N, the charges on the non-polarised nitrogen, q'_N, Drude core (DC), q_N, and Drude particle (DP), $q_{D,N}$, are taken from the Cl&Pol polarisable FF.[148] The direction of the resulting induced dipole is indicated by μ_N.

The potential energy of the induced dipole can be represented as in Equation 2.34, repeated here:

$$U_{\mathrm{ind}} = U_{\mathrm{stat}} + U_{\mu\mu} + U_{\mathrm{pol}}. \tag{2.34}$$

In this model, U_{stat} is determined as

$$U_{\text{stat}} = \sum_{i=1}^{N} (q_i - q_{\text{D},i}) \left[\mathbf{r}_i \cdot \mathbf{E}_i^0 - (\mathbf{r}_i + \mathbf{d}_i) \cdot \mathbf{E}_{\text{D},i}^0 \right], \tag{2.40}$$

with $\mathbf{E}_i^0$ and $\mathbf{E}_{\text{D},i}^0$ being the static fields at the DC and DP sites, respectively. The dipole-dipole interaction energy is the sum of the charge-charge (Coulomb) interactions at each charged site:

$$U_{\mu\mu} = \sum_{i}^{N} \sum_{j \neq i} \left[\frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{q_{\text{D},i} q_{\text{D},j}}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{d}_i - \mathbf{d}_j|} - \frac{q_{\text{D},i} q_j}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{d}_j|} - \frac{q_i q_{\text{D},j}}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{d}_i|} \right]. \tag{2.41}$$

Here, the polarisation contribution, U_{pol}, is due to the energy of the harmonic oscillator between the pair of point charges,

$$U_{\text{pol}} = \frac{1}{2} \sum_{i=1}^{N} k_{\text{D}} |\mathbf{d}_i|^2. \tag{2.42}$$

The equations of motion can then be integrated for the DPs to obtain the trajectories using a *dynamic approach*, in which the DP is assigned a small mass, typically set to 0.4 amu, that will respond to the degrees of freedom of the DC. The mass of the DP does not have any physical interpretation and can be seen as a parameter required for numerical integration.[197] To ensure that the trajectory approaches that of fully relaxed Drude dipoles — that the kinetic energy of the system is correctly distributed into the degrees of freedom — a *temperature-grouped dual thermostat* is employed, in which the DPs are kept at a low temperature, typically 1 K,[202] and the DCs are thermostated at the required temperature. The dual temperature-grouped Nosé-Hoover thermostat is widely implemented, in which all non-DPs are thermostatted using the following:

$$m_i \frac{\partial^2 \mathbf{R}_i}{\partial t^2} = \mathbf{F}_i - m_i \frac{\partial \mathbf{R}_i}{\partial t} \frac{\partial \eta}{\partial t}, \tag{2.43}$$

where for the DC-DP pair i, m_i is the total mass ($m_{\text{DC}} + m_{\text{DP}}$), $\mathbf{R}_i$ is the centre of mass (COM), $\mathbf{F}_i$ is the force acting on the COM, and η is the specific energy of the heat bath. At the same time, a second thermostat is applied to the DPs such that

$$m_{\text{D},i}\frac{\partial^2 \mathbf{d}_i}{\partial t^2} = \mathbf{F}_{\text{D},i} - m_{\text{D},i}\frac{\partial \mathbf{d}_i}{\partial t}\frac{\partial \eta_{\text{D}}}{\partial t}, \tag{2.44}$$

where the subscript D refers to the reduced value of the corresponding variable in Equation 2.43 for the DP.

It has been found that dual thermostat schemes, commonly used in modern MD codes for computational efficiency, tend to underestimate dynamic properties such as diffusion coefficients and ionic conductivities in ILs by 50% or more compared to self-consistent (SCF) or adiabatic optimization of Drude oscillator positions.[197] This discrepancy arises because dual thermostat methods approximate the positions of Drude oscillators near 0 K rather than achieving a fully optimized distribution, which affects the accuracy of predicted dynamical properties. Although the dual-thermostat approach, using an extended Lagrangian, is computationally efficient, its limitations in accurately modelling dynamic behavior have been noted by McDaniel[170] While more accurate, SCF optimization is not available in any open-source MD software. As such, this technique is not used for any MD simulations in this work, and the focus is placed on efficiency and accessibility of simulations performed using the tools available. Thus, all simulations in this thesis use the extended Lagrangian approach to optimise DP positions.

2.1.5 The CL&Pol force field

In this work, the CL&Pol Drude polarisable FF is used, taking the functional form of[133]

$$E_{\text{tot}} = E_{\text{bonded}} + E_{\text{non-bonded}}, \tag{2.45}$$

where the bonded contribution, E_{bonded}, is expressed as

$$\begin{aligned} E_{\text{bonded}} = &\sum_{ij}^{\text{bonds}} \frac{k_{\text{r},ij}}{2} \left(r_{ij} - r_{0.ij}\right)^2 + \sum_{ijk}^{\text{angles}} \frac{k_{\theta,ijk}}{2} \left(\theta_{ijk} - \theta_{0.ijk}\right)^2 \\ &+ \sum_{ijkl}^{\text{dihedrals}} \sum_{m=1}^{4} \frac{V_{\text{m},ijkl}}{2} \left[1 + (-1)^m \cos(m\eta_{ijkl})\right]. \end{aligned} \tag{2.46}$$

The non-bonded terms consist of contributions from the LJ, E_{LJ}, and Coulomb, $E_{\text{coul}}^{\text{NP}}$,[I] potentials,

$$\begin{aligned} E_{\text{non-bonded}} &= f_{ij} \left\{E_{\text{LJ}} + E_{\text{coul}}^{\text{NP}}\right\} \\ &= f_{ij} \left\{\sum_{i}^{\text{atoms}} \sum_{j \neq i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i}^{\text{atoms}} \sum_{j \neq i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}\right\} \end{aligned} \tag{2.47}$$

Here, the f_{ij} scaling factor is implemented as in Section 2.1.2.

Polarisation is included in the FF by replacing each heavy atom with a Drude oscillator, as described in Section 2.1.4.3. Hydrogen atoms are not treated as polarisable and as such, the hydrogen polarisabilities are merged onto the heavy atom to which they are bonded. The DPs are given a mass of 0.4 au and the mass of the DC is taken as the difference between that of the heavy atom and the DP. DPs are taken to be negatively charged, and a force constant, k_D, of 4184 kJ/mol is used from which DP partial charges, q_D, are determined using atomic polarisabilities,[203] α, such that $\alpha = q_D^2/k_D$.[197] The partial charge of the DC is taken as the difference between the total atomic charge and the DP partial charge. The addition of Drude oscillators requires damping of the short-range electrostatics which can be interpreted as the physically accurate smearing of charge at short range.[175] In practice, this avoids over-correlation between induced dipoles leading to the "polarisation catastrophe" where the DP of one oscillator is captured by the DC of

[I] the NP superscript indicates non-polarisable electrostatics and is for convenience of notation

another. As such, the Coulomb potential of the CL&Pol is expressed as

$$E_{\text{coul}} = E_{\text{coul}}^{\text{NP}} + \frac{1}{4\pi\epsilon_0}\left[\sum_{i}^{\text{atoms}}\sum_{j\neq i} S_{ij}\frac{q_{D,i}q_{D,j}}{r_{ij}} + \sum_{i}^{\text{atoms}}\sum_{j} T_{ij}\frac{q_i q_{D,j}}{r_{ij}}\right]. \tag{2.48}$$

First, the short-range interactions between DPs are damped by the Thole damping function,[157,204] S_{ij},

$$S_{ij} = 1 - (1 + \frac{s_{ij}r_{ij}}{2})\exp(-s_{ij}r_{ij}), \tag{2.49}$$

with

$$s_{ij} = \frac{a_{ij}}{\left(\sqrt{\alpha_i\alpha_j}\right)^{1/3}}, \tag{2.50}$$

where a_{ij} takes a universal value of 2.6. Second, an adaptation of the Tang-Toennies damping function,[205] T_{ij}, is applied to short-range interactions between DP and DC of densely charged ions (ie., alkali-metal cations):

$$T_{ij} = 1 - c_{ij}\exp(-b_{ij}r_{ij})\sum_{k}^{4} = 0\frac{(b_{ij}r_{ij})^k}{k!}, \tag{2.51}$$

where b_{ij} and c_{ij} are 4.5 and 1.0, respectively.

Since the LJ σ_{ij} and ε_{ij} parameters are not reparemeterized and carried directly over from the CL&P FF, the attractive part of the LJ potential already includes the induction energy, and as such this contribution needs to be removed. Pádua[206] developed the "fragment approach" which is used to obtain values of k_{ij}, by which ε_{ij} is scaled to effectively remove the induction contribution. While the subscripts i and j refer to the atoms involved in the interaction, k_{ij} is not determined for atom-atom specific interactions, but rather for the interaction of representative fragments containing atoms of type i and j. That being, all interactions between all atoms in two defined fragments will be scaled by the same value of k_{ij}, and atom-atom interactions within a fragment are not scaled. Using the fragment approach allows for transferability, as new k_{ij} factors do not need to be determined if the ions can be built out of previously defined fragments. ILs are typically represented

by three fragments: a cation core, a representative alkyl chain, and an anion. Throughout development, several variations of the CL&Pol FF[148] were tested, two of which scale the LJ ε parameter and are outlined here:

1. *CSDrude*, where ε is scaled by $k_{ij}^{\text{SAPT2+}}$, calculated using the dispersion and induction energy components obtained from a SAPT2+ energy decomposition of the representative fragments for i and j:

$$k_{ij}^{\text{SAPT2+}} = \frac{E_{\text{disp}}}{E_{\text{disp}} + E_{\text{ind}}}, \tag{2.52}$$

   where E_{disp} and E_{ind} are the dispersion and induction energies, respectively. These are calculated using a Symmetry-Adapted Perturbation Theory (SAPT),[207] energy decomposition, at the SAPT2+/aug-cc-pVDZ level, of a dimer consisting of the representative fragments containing i and j.

2. *KCSDrude*, where ε between dimers i and j is scaled by, which requires the calculation of individual ion-dipole moments, μ_{ij}, and dimer equilibrium distances between the COMs, r_{ij}, and molecular polarisabilities, α_i and α_j. These values are then used to obtain according to the following predictive scheme:

$$k_{ij}^{\text{pred}} = \Big(1 + 0.25 r_{ij}^2 \frac{Q_i^2\alpha_j + Q_j^2\alpha_i}{\alpha_i\alpha_j} + 0.11 \frac{\mu_i^2\alpha_j + \mu_j^2\alpha_i}{\alpha_i\alpha_j}\Big). \tag{2.53}$$

   This requires only the charge, Q_n, polarisability, α_n, and dipole moment, μ_n for each fragment n, as well as the equilibrium distance separation of the fragments, r_{ij}, rather than a costly SAPT energy decomposition. For fragments representing i and j, only calculation of r_{ij} is required, calculatable from the geometry optimised pair. Calculation of α_n for each fragment n is not required as the `scaleLJ` script provided by Goloviznina et al.[148] calculates these values.

Note that the non-bonded LJ interaction between atoms that are separated by three bonds (1-4 interactions) is coupled to the corresponding torsional profile in the bonded contribution to the FF energy. In practice, the torsional parameters must be parameterized so that the combination of the 1-4 interaction and the pure torsional profile matches the target quantum mechanical profile. If the torsional parameters are carried over from the CL&P FF, the quality of the fit may be affected when scaling the LJ parameters.

2.2 Quantum Mechanics background

Quantum mechanics (QM) is a fundamental theory that describes the behaviour of particles at the atomic and subatomic levels. Central to quantum chemistry, it underpins our understanding of molecular interactions and reactions. Within this framework, various computational methods have been developed to solve the Schrödinger equation for molecular systems. The Hartree-Fock method provides the most basic approximate solution by considering electron interactions as a wavefunction represented by a Slater determinant; though this often underestimates the true energy of the system due to neglect of electron correlation. Electron correlation energy accounts for the interactions between electrons that are not captured by Hartree-Fock, and sophisticated methods like Coupled Cluster theory can accurately include these effects, making it one of the most precise techniques for electronic structure calculations. Additionally, Density Functional Theory offers a different approach by using electron density rather than wavefunctions, providing a balance between accuracy and computational efficiency, making it widely used in both academic and industrial research for studying the electronic properties of molecules and materials. The following section gives background into the QM theory forming the basis of the methods used throughout this work.[179,182]

2.2.1 Hartree-Fock theory

Fundamentally, QM describes atoms and molecules using a wavefunction, Ψ, for which the energy, E can be obtained through the solution of the time-independent Schrödinger Equation,

$$\hat{H}\mathbf{\Psi} = E\mathbf{\Psi}, \tag{2.54}$$

where **H** is the Hamiltonian operator, corresponding to the total energy of the system—the summation of the kinetic and potential energy of the nuclei and electrons. If the Born-Oppenheimer approximation is applied, the nuclei positions are considered fixed and the energy of the system becomes dependent on only electronic motion and nuclear positions, but not nuclear momentum. Equation 2.54 can be expressed as the *electronic* Schrödinger Equation,

$$\hat{H}_e\Psi_e(\mathbf{r},\mathbf{R}) = E_e\Psi_e(\mathbf{r},\mathbf{R}), \tag{2.55}$$

where **r** is the electronic coordinates and **R** is the position of the nuclei. The electronic Hamiltonian operator, $\hat{H}_e$, is composed of contributions from the kinetic and potential energy operators. The kinetic energy operator, $\hat{T}$, is defined as

$$\hat{T} = -\frac{\hbar^2}{2}\sum_i \frac{1}{m_i}\nabla_i^2, \tag{2.56}$$

where the Laplacian operator, concerning electron coordinates, is expressed as

$$\nabla_i^2 = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2}. \tag{2.57}$$

The potential energy operator, $\hat{V}$, is expressed as

$$\hat{V} = V(\mathbf{r}) = -\frac{1}{4\pi\epsilon_0}\left[\sum_i^{N_{\text{elec}}}\sum_A^{N_{\text{nuclei}}}\frac{Z_A e^2}{|\mathbf{r}_i - \mathbf{R}_A|} - \sum_i^{N_{\text{elec}}}\sum_{j>i}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_A^{N_{\text{nuclei}}}\sum_{B>A}\frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}\right]. \tag{2.58}$$

Here e is the charge of an electron, Z_{A} and Z_{B} are the charges on nuclei A and B, respecively and $\mathbf{R}_{\mathrm{A}}$ is the coordinates of A and $\mathbf{R}_{\mathrm{B}}$ is the coordinates B. The Schrödinger Equation can only be solved exactly for single-electron systems and as such, numerical, approximate methods are required to obtain a solution for a many-electron system. The antisymmetry principle states that a wave function must be antisymmetric concerning electron-coordinate interchange, which can be achieved by describing the wavefunction with a Slater Determinant (SD). Considering a polyatomic system with N electrons and N molecular orbitals (from ψ_1 to ψ_N), the normalized wavefunction can be written as

$$\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) = \frac{1}{\sqrt{N}} \begin{vmatrix} \psi_1(\mathbf{r}_1) & \psi_1(\mathbf{r}_2) & \cdots & \psi_1(\mathbf{r}_N) \\ \psi_2(\mathbf{r}_1) & \psi_2(\mathbf{r}_2) & \cdots & \psi_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_N(\mathbf{r}_1) & \psi_N(\mathbf{r}_2) & \cdots & \psi_N(\mathbf{r}_N) \end{vmatrix}. \tag{2.59}$$

The energy of the SD is expressed as

$$E = \sum_{i}^{N_{\text{elec}}} \langle \psi_i | \hat{h}_i | \psi_i \rangle + \frac{1}{2} \sum_{ij}^{N_{\text{elec}}} (\langle \psi_| \hat{J}_i | \psi_j \rangle - \langle \psi_j | \hat{K}_i | \psi_j \rangle), \tag{2.60}$$

where $\hat{h}_i$ is the one-electron Hamiltonian operator, describing the kinetic and potential energy of electron i in the presence of all nuclei, and $\hat{J}$ and $\hat{K}$ are the Coulomb and exchange operators, respectively. The variational principle states that an approximate wavefunction will have an energy equal to or above the exact energy, whereas the two will only be equal when the wavefunction is exact. To optimise the molecular orbitals (MOs) for which the energy is a minimum, the variation in energy, δE,expressed as

$$\delta E = \sum_{i}^{N_{\text{elec}}} (\langle \delta\psi_i | \hat{F}_i | \psi_i \rangle + \langle \psi_i | \hat{F}_i | \delta\psi_i \rangle) \tag{2.61}$$

is minimized, where the Fock operator, $\hat{F}_i$, given as

$$\hat{F}_i = \hat{h}_i + \sum_{j}^{N_{\text{elec}}} (\hat{J}_j - \hat{K}_j). \tag{2.62}$$

The Fock operator $\mathbf{F}_i$ describes the kinetic energy, attraction to all nuclei, and repulsion from all other electrons for each electron in MO i. From the minimization of Equation 2.61, the Hartree-Fock equations are obtained:

$$\hat{F}_i \psi_i = \varepsilon_i \psi_i, \tag{2.63}$$

for which

$$\hat{F}_i = \hat{h}_i + \hat{v}_i^{HF} \tag{2.64}$$

and

$$\hat{v}_i^{HF} = \sum_{j}^{N} \hat{J}_{i,j} - \hat{K}_{i,j}. \tag{2.65}$$

The eigenvalues of the Fock operator are the MO energies, from which the total energy can be obtained. Each MO ψ_i, can be expanded expressed in terms of atomic orbitals ϕ, such that

$$\psi_i = \sum_{k}^{N_{\text{AO}}} c_{ki} \phi_k. \tag{2.66}$$

for which N_{AO} is the number of atomic orbitals and c_{ki} is the MO coefficient, which can be solved using variation theorem. The atomic orbitals are written as a set of mathematical functions referred to as a basis set, and using a linear combination of atomic orbitals (LCAO), Equation 2.63 can be reformulated into the Roothaan-Hall equations,

$$\mathbf{FC = SC}\varepsilon \tag{2.67}$$

where **F** is the Fock matrix, **C** is the matrix of MO coefficients, ε is a diagonal matrix of orbital energies and **S** is the overlap matrix. Based on the Variational Theorem, c_{ki} in

Equation 2.66 should be the set of coefficients that result in a minimum in the variation of energy. Here, **C** is the $K \times K$ matrix containing the c_{ki} coefficients, and the elements of the Fock matrix are thus dependent on the MO coefficients. The Fock matrix is therefore dependent on its own solution and is solved iteratively using the self-consistent field procedure, from which the orbital energies can be solved and used to calculate the Hartree-Fock energy.

While the Hartree-Fock method can provide an almost complete solution to the Schrödinger equation when a sufficiently large basis set is used, the correlation between electrons is not considered with this approximation. This results in a deficit in the calculated energy, where the difference between the true energy, E, and the Hartree-Fock energy, E^{HF}, is referred to as the electron correlation energy,[208] denoted here as E_{corr}. Three common approaches to correct calculated energies to account for E_{corr}, used in this work, are discussed below.

2.2.2 Møller-Plesset perturbation method

In general, perturbation theory defines $\hat{H}$ as consisting of a reference operator, $\hat{H}_0$, and a perturbation, $\hat{H}'$,

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}', \tag{2.68}$$

where $\hat{H}_0$ has a closed solution, as in Equation 2.54, and the perturbed Schrödinger equation can be written as

$$\hat{H}\Psi = W\Psi. \tag{2.69}$$

Both the perturbed energy, W, and Ψ can be written as a power series of the perturbation parameter, λ, as

$$W = \lambda^0 W_0 + \lambda^1 W_1 + \lambda^2 W_2 + \lambda^3 W_3 + ... \tag{2.70}$$

and

$$\Psi = \lambda^0\Psi_0 + \lambda^1\Psi_1 + \lambda^2\Psi_2 + \lambda^3\Psi_3 + ..., \tag{2.71}$$

respectively. In both Equations 2.70 and 2.71 the superscript represents the order of correction to the wavefunction and energy. In Møller-Plesset (MP) perturbation theory, $\hat{H}_0$ is taken as the sum over Fock operators, such that

$$\hat{H}_0 = \sum_{i=1}^{N_{\text{elec}}} \hat{F}_i, \tag{2.72}$$

from which the perturbation operator can be deduced as

$$\hat{H}' = \hat{H} - \hat{H}_0 = \sum_{i=1}^{N_{\text{elec}}}\sum_{j>1}^{N_{\text{elec}}} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i=1}^{N_{\text{elec}}}\sum_{j=1}^{N_{\text{elec}}} \left\langle \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \right\rangle. \tag{2.73}$$

Expressions for the various orders of correction to the energy and wavefunction can be obtained by comparing terms with the same power of lambda, and recognizing that in the absence of the perturbation,

$$\hat{H}_0\Psi_0 = E_0\Psi_0. \tag{2.74}$$

Consequently, the zeroth-order energy is only a function of the HF reference wavefunction, Ψ_0, and is simply the sum of MO energies,

$$W_0 = \langle\Psi_0|\hat{H}_0|\Psi_0\rangle = \sum_{i=1}^{N_{\text{elec}}} \varepsilon_i. \tag{2.75}$$

The first-order correction is then determined as

$$W_1 = \langle\Psi_0|\hat{H}'|\Psi_0\rangle = -\sum_{i=1}^{N_{elec}}\sum_{j>i}^{N_{elec}} \left\langle \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \right\rangle, \tag{2.76}$$

and as such, the first-order MP energy, E^{MP1}, is the sum of W_0 and W_1, such that Equation 2.75 added to Equation 2.76 is E^{HF}. Therefore, electron configuration energy is only included when second and further higher-order MP energy contributions are considered.

To obtain the second-order MP correction, W_2, all matrix elements of $\hat{H}'$ between the ground state and all doubly-excited states are required, which is the result of exciting an electron from each occupied (occ) orbital i and j to the corresponding virtual (vir) orbitals a and b, where each excited state is only counted once. This leads to the the second order MP energy, E^{MP2},

$$\begin{aligned} E^{\text{MP2}} &= E^{\text{MP1}} + W_2 \\ &= E^{\text{MP1}} + \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} \frac{\langle \Psi_0 | \mathbf{H}' | \Psi_{ij}^{ab} \rangle \langle \Psi_{ij}^{ab} | \mathbf{H}' | \Psi_0 \rangle}{E_0 - E_{ij}^{ab}} \\ &= \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} \frac{\left(\langle \psi_i \psi_j | \psi_a \psi_b \rangle - \langle \psi_i \psi_j | \psi_b \psi_a \rangle \right)^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}. \end{aligned} \tag{2.77}$$

The perturbation can be extended to include higher-order corrections. The second-order (MP2),[209] third-order (MP3)[210] and fourth-order (MP4)[211] are routinely used and implemented in many widely used computational packages. While higher-order calculations are theoretically possible, these are not typically used due to the large computational cost associated with them.

2.2.3 Coupled Cluster

The MP method includes all types of corrections up to a given order, whereas the Coupled Cluster (CC) method includes corrections of a given type up to an infinite order. The CC wavefunction is defined in terms of $\mathbf{\Psi}_0$

$$\mathbf{\Psi}_{\text{CC}} = \mathrm{e}^{\hat{T}} \mathbf{\Psi}_0, \tag{2.78}$$

where the operator $e^{\hat{T}}$ is

$$\mathrm{e}^{\hat{T}} = \sum_{k=0}^{\infty} \frac{1}{k!} \hat{T}^k. \tag{2.79}$$

The excitation operator, $\hat{T}$, is defined as

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ... + \hat{T}_{N_{\text{elec}}}, \tag{2.80}$$

where $\hat{T}_i$ generates excited states to the ith degree, such that

$$\hat{T}_1 \mathbf{\Psi}_0 = \sum_{i}^{\text{occ}} \sum_{a}^{\text{vir}} t_i^a \mathbf{\Psi}_i^a, \tag{2.81}$$

$$\hat{T}_2 \mathbf{\Psi}_0 = \sum_{i<j}^{\text{occ}} \sum_{a<b}^{\text{vir}} t_{ij}^{ab} \mathbf{\Psi}_{ij}^{ab}, \tag{2.82}$$

and so on, where t is the expansion coefficient (amplitude). Using the CC wavefunction, the Schrödinger equation is written as

$$\hat{H}_{\text{e}}^{\hat{T}} \mathbf{\Psi}_0 = E_{\text{CC,e}}^{\hat{T}} \mathbf{\Psi}_0. \tag{2.83}$$

If we treat the Hamiltonian operator in Equation 2.83 with a similarity transformation, such that

$$\mathrm{e}^{-\hat{T}} \hat{H}_{\text{e}}^{\hat{T}} \mathbf{\Psi}_0 = E_{\text{CC}} \mathbf{\Psi}_0, \tag{2.84}$$

then multiply through by $\mathbf{\Psi}_0^*$ and integrating, the CC energy is then obtained as

$$E_{\text{CC}} = \langle \mathbf{\Psi}_0 | \mathrm{e}^{-\hat{T}} \hat{H}_{\text{e}}^{\hat{T}} | \mathbf{\Psi}_0 \rangle. \tag{2.85}$$

If all operators up to $\hat{T}_{N_{\text{elec}}}$ are considered then all excited states are described. However, this is impossible as the operator $\hat{T}$ is truncated, as such arriving at an approximate solution. The lowest improvement over the E^{HF} is truncating $\hat{T}$ at $\hat{T}_2$, such that single and

double excitations are included, to produce the Coupled Cluster with Singles and Doubles (CCSD) model. Including triple excitations (CCSDT) leads to the next level of improvement to the model, but at a significantly greater computational cost. As such, a common approach is to use the CCSD method and include triples through perturbation theory; this method is referred to as Coupled Cluster with Singles and Doubles and perturbative Triples, CCSD(T), and provides comparable accuracy to the full CCSDT method.

2.2.4 Density Functional Theory

Another way to introduce electron correlation is with Density Functional Theory (DFT). It has been shown thus far that the energy of a system can be solved if the wavefunction of a system is known; however, as the number of electrons increases, the complexity of the wavefunction becomes greater and as such, the solution to the Schrödinger equation can quickly become unobtainable. In DFT, the many-body problem of interacting electrons is simplified by expressing the system's energy as a functional of the electron density, rather than the wavefunction. Electron correlation is inherently included through the exchange-correlation functional, which encapsulates both exchange effects and correlation energies. This functional, while not exactly known, is approximated using various methods, allowing DFT to balance computational efficiency with reasonable accuracy. Hohenberg and Kohn[212] showed that the ground state energy is a unique function of the ground state electron density — the latter of which only depends on three coordinates, independent of the number of electrons. Thus, DFT allows for the energy to be expressed in terms of electron density, ρ as

$$E[\rho] = T[\rho] + E_{\text{en}}[\rho] + E_{\text{ee}}[\rho]. \tag{2.86}$$

Equation 2.86 includes contributions from kinetic energy, denoted as $T[\rho]$, potential energy consisting of nuclei-electron attraction, $E_{\text{en}}[\rho]$, and electron-electron repulsion, $E_{\text{ee}}[\rho]$,

where

$$E_{\text{en}}[\rho] = -\sum_{A}^{N_{\text{nuclei}}} \int \frac{Z_A(\mathbf{R}_A)\rho(\mathbf{r})}{|\mathbf{R}_A - \mathbf{r}|}\mathrm{d}\mathbf{r}, \tag{2.87}$$

with $\rho(\mathbf{r})$ being the electron density at the coordinates $\mathbf{r}$. As with HF theory, $E_{\text{ee}}[\rho]$ consists of contributions from the Coulomb, $J[\rho]$, and exchange, $K[\rho]$, energies, where the latter is given by the classical expression

$$J[\rho] = \frac{1}{2}\int\int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\mathrm{d}\mathbf{r}\ \mathrm{d}\mathbf{r}'. \tag{2.88}$$

Expression for $T[\rho]$ and $K[\rho]$ were first approximated by Dirac, Fermi, and Thomas[213,214] with the assumption of a single-electron, non-interacting gas. While the fundamental idea of DFT is to make an "orbital-free" approximation based solely on electron density, it is difficult to obtain an accurate functional for $T[\rho]$ many-electron systems. The later description by Kohn and Sham (KS)[215] offered an improvement on the initial approximation such that orbital effects are included: One evaluates the kinetic energy using a fictitious non-interacting wavefunction that has the same electron density as the system of interacting electrons being evaluated. Therefore, the exact kinetic energy can be calculated using the exact density, ρ_{exact}, as

$$T[\rho_{\text{exact}}] = \sum_{i=1}^{\infty} n_i \left\langle \psi_i^{\text{NO}} \right| -\frac{1}{2}\nabla^2 \left| \psi_i^{\text{NO}} \right\rangle, \tag{2.89}$$

where the natural orbitals, ψ_i^{NO}, are the infinite set of eigenvectors of the exact density matrix,

$$\rho_{\text{exact}} = \sum_{i=1}^{\infty} n_i \left| \psi_j^{\text{NO}} \right|^2, \tag{2.90}$$

with n_i being the orbital occupation numbers such that

$$N_{\text{elec}} = \sum_{i=1}^{\infty} n_i. \tag{2.91}$$

As ρ_{exact} is not known, it is approximated as

$$\rho_{\text{exact}} \approx \rho_{\text{approx}} = \sum_{i=1}^{N_{\text{elec}}} |\psi_i|^2. \tag{2.92}$$

The KS model is based on non-interacting electrons and, while small, this assumption creates a deviation between the calculated and exact kinetic energy. With DFT, this excluded energy is absorbed into the exchange-correlation energy, $E_{\text{XC}}[\rho]$, which also includes $K[\rho]$, and Equation 2.86 can be rewritten generally as

$$E_{\text{DFT}}[\rho] = T_{\text{KS}}[\rho] + E_{\text{en}}[\rho] + J[\rho] + E_{\text{XC}}[\rho], \tag{2.93}$$

in which

$$T_{KS}[\rho] = \sum_{i=1}^{N} \langle \psi_i | - \frac{1}{2}\nabla^2 | \psi_i \rangle. \tag{2.94}$$

As such, using Equation 2.93, the energy can be solved as a functional of the density.

2.3 Symmetry-Adapted Perturbation Theory

Accurately describing the strength and nature of non-covalent interactions in a chemical system is essential, and the development of such techniques is an important task within the field of quantum chemistry.[216] The gold standard for the calculation of the interaction energy, E_{int}, against which the accuracy of other methods is often benchmarked, is usually the coupled-cluster method with single, double and perturbative triple excitations, CCSD(T),[217] extrapolated to the complete basis set (CBS) limit.[218–220] Although using routine use of this method would be ideal, the much-desired accuracy comes at a steep computational cost, with CCSD(T) traditionally scaling as $\mathcal{O}(N^7)$ with system size N. Furthermore, while CCSD(T)/CBS can be used to calculate supermolecular interaction energies with high accuracy, this computation does not give insight into the nature of the interac-

tion. For this purpose, energy decomposition analysis (EDA) methods have been developed[221–226] that decompose supermolecular E_{int} into physically meaningful components, typically consisting of electrostatics, exchange-repulsion (Pauli repulsion), polarisation (or induction), charge-transfer (or charge delocalization), and dispersion.[227,228]

An attractive alternative to a supermolecular approach is Symmetry-Adapted Perturbation Theory (SAPT),[168] in which E_{int} is calculated as the perturbative sum of physically motivated terms. The SAPT E_{int} provides an unambiguous description of the components, based on first principles, for which the formalism and definitions have been detailed below. SAPT describes E_{int} as the sum of contributions from (1) electrostatics, the Coulomb interaction between the charge density of isolated molecules; (2) induction, the stabilisation resulting from polarisation between two molecules; (3) dispersion, an attractive interaction resulting from instantaneous electronic correlation between molecules; and (4) exchange, a short-range repulsive force as a result of the Pauli exclusion principle.[229] In recent years, SAPT has become widely popular for obtaining interaction energy components and has been implemented based on either a wavefunction[207] (WFT) or density functional theory (DFT) treatment of monomers; the latter was independently developed as SAPT-DFT or SAPT(DFT).[230] The sections below outline the SAPT formalism and how the various levels of SAPT, used in this work, have been defined.

2.3.1 The SAPT Hamiltonian

Using the QM methods discussed above, the interaction energy, E_{int}, between two molecules A and B can be calculated using a supermolecular approach as

$$E_{\text{int}} = E_{\text{AB}} - (E_{\text{A}} + E_{\text{A}}). \tag{2.95}$$

SAPT on the other hand, provides a way of obtaining E_{int} without computing the total energy of the dimer system and each monomer.

The SAPT formalism begins with partitioning the Hamiltonian operator of the dimer into contributions from each monomer and the interaction between them:

$$\hat{H} = \hat{H}_{\mathrm{A}} + \hat{H}_{\mathrm{B}} + \hat{V}. \tag{2.96}$$

The intermolecular operation operator, $\hat{V}$, can be considered a perturbation to the Hamiltonian of the non-interacting monomers. Assuming that the electronic Schrödinger equation can be written exactly and independently for each monomer, that being all electronic correlation is calculated, the zeroth-order Hamiltonian for the non-interacting monomers, $\hat{H}_0$, can be expressed as

$$H_0 = \hat{H}_{\mathrm{A}} + \hat{H}_{\mathrm{B}}. \tag{2.97}$$

The corresponding eigenfunction is expressed as

$$\Psi^{(0)} = \Psi^{(0)}_{\mathrm{A}}(1, ..., N_{\mathrm{A}}) \cdot \Psi^{(0)}_{\mathrm{B}}(1 + N_{\mathrm{A}}, ..., N_{\mathrm{B}} + N_{\mathrm{A}}), \tag{2.98}$$

where N_{A} and N_{B} are the number of electrons in monomers A and B, respectively. The wavefunction given in Equation 2.98 is therefore antisymmetric regarding electron exchange *within* each monomer, but not *between* monomers. Obtaining the energy using this description ($\hat{H}_0 = \hat{H}_A + \hat{H}_B$ and **V** as the perturbation) which is known as the *polarisation approximation*, is only valid for a system consisting of, at most, two electrons. The "intermolecular" Pauli exclusion principle must then be accounted for with the inclusion of a *symmetry-adapted* permutation in the Hamiltonian in Equation 2.96, with the inclusion of an electron antisymmetrizer *A*, which in turn includes an exchange correction into SAPT.

After applying the perturbation in orders of $\hat{V}$, together with the antisymmetrizer, the first-order energy contribution contains electrostatic interaction between monomer charge densities and the first-order exchange contribution. The first-order electrostatic interaction

energy is given by[229]

$$E^{(1)}_{\text{elst}} = \langle \Psi^{(0)}_{\text{A}} \Psi^{(0)}_{\text{B}} | V | \Psi^{(0)}_{\text{A}} \Psi^{(0)}_{\text{B}} \rangle. \tag{2.99}$$

The second-order contribution includes the induction, dispersion and second-order exchange contributions, which separate into an exchange-induction and exchange-dispersion contribution. Induction further separates into the polarization of molecule A due to the electric field generated by molecule B,[229]

$$E^{(2)}_{\text{ind,B}\to\text{A}} = \sum_{i\neq 0} \frac{|\langle \Psi^{(0)}_{\text{A}} \Psi^{(0)}_{\text{B}} | V | \Psi^{(i)}_{\text{A}} \Psi^{(0)}_{\text{B}} \rangle|^2}{E^{(0)}_{\text{A}} - E^{(i)}_{\text{A}}}, \tag{2.100}$$

and another term where the electric field of A polarizes B,

$$E^{(2)}_{\text{ind,A}\to\text{B}} = \sum_{j\neq 0} \frac{|\langle \Psi^{(0)}_{\text{A}} \Psi^{(0)}_{\text{B}} | V | \Psi^{(0)}_{\text{A}} \Psi^{(j)}_{\text{B}} \rangle|^2}{E^{(0)}_{\text{B}} - E^{(j)}_{\text{B}}}. \tag{2.101}$$

In these expressions, $\Psi^{(i)}_{\text{A}}$ and $E^{(i)}_{\text{A}}$ are the excited eigenfunctions and eigenvalues of operator $\hat{H}_{\text{A}}$, and $\Psi^{(j)}_{\text{B}}$ and $E^{(j)}_{\text{B}}$ are the excited eigenfunctions and eigenvalues of operator $\hat{H}_{\text{B}}$. The dispersion energy is expressed as[229]

$$E^{(2)}_{\text{disp}} = \sum_{i,j\neq 0} \frac{|\langle \Psi^{(0)}_{\text{A}} \Psi^{(0)}_{\text{B}} | V | \Psi^{(i)}_{\text{A}} \Psi^{(j)}_{\text{B}} \rangle|^2}{E^{(0)}_{\text{A}} + E^{(0)}_{\text{B}} - E^{(i)}_{\text{A}} - E^{(j)}_{\text{B}}}. \tag{2.102}$$

As such, at the second order, the interaction energy includes contributions from electrostatics, exchange, induction and dispersion.

2.3.2 Many-body SAPT

Practically, the exact intramolecular electron correlation cannot be accounted for, and the definition of the zeroth-order Hamiltonian needs further consideration. Using the many-

body SAPT technique, the Hamiltonian in Equation 2.96 is replaced by

$$\hat{H} = (\hat{F}_{\mathrm{A}} + W_{\mathrm{A}}) + (\hat{F}_{\mathrm{B}} + W_{\mathrm{B}}) + \hat{V} \tag{2.103}$$

where the Hamiltonian of each monomer is expressed as the sum of the monomer Fock operator and a fluctuation potential. One can then do a double perturbation and calculate $E_{\mathrm{int}}^{(mn)}$ where m indicates the order of intermonomer correction (the order of $\hat{V}$) and n indicates the intramonomer correlation (the order of $\hat{W}$). Therefore, the simplest many-body SAPT method is SAPT0 where

$$E_{\mathrm{int}}^{\mathrm{SAPT0}} = E_{\mathrm{elst}}^{(10)} + E_{\mathrm{exch}}^{(10)} + E_{\mathrm{ind,resp}}^{(20)} + E_{\mathrm{exch\text{-}ind,resp}}^{(20)} + E_{\mathrm{disp}}^{(20)} + E_{\mathrm{exch\text{-}disp}}^{(20)} + \delta_{\mathrm{HF}}^{(2)}. \tag{2.104}$$

Here, the subscript "resp" indicates that the orbital relaxation of one monomer due to the electrostatic potential of the other monomer is included, and the $\delta_{\mathrm{HF}}^{(2)}$ term accounts for all higher-order induction and exchange-induction contributions not included in SAPT0, calculated as

$$\delta_{\mathrm{HF}}^{(2)} = E_{\mathrm{int}}^{\mathrm{HF}} - E_{\mathrm{elst}}^{(10)} - E_{\mathrm{exch}}^{(10)} - E_{\mathrm{ind,resp}}^{(20)} - E_{\mathrm{exch\text{-}ind,resp}}^{(20)}, \tag{2.105}$$

where $E_{\mathrm{int}}^{\mathrm{HF}}$ is the counterpoise-corrected Hartree-Fock interaction energy.[231] Including the $\delta_{\mathrm{HF}}^{(2)}$ correction is sometimes referred to as *hybrid* SAPT, opposed to *pure* SAPT where every term follows from perturbation theory.

The accuracy of the SAPT0 interaction energy can be improved by including (1) intramolecular correlation (*W*) and (2) higher-order intermolecular contributions. The inclusion of (1) leads to the SAPT2 and SAPT2+ methods:

$$E_{\mathrm{int}}^{\mathrm{SAPT2}} = E_{\mathrm{int}}^{\mathrm{SAPT0}} + E_{\mathrm{elst,resp}}^{(12)} + E_{\mathrm{exch}}^{(11)} + E_{\mathrm{exch}}^{(12)} + {}^{t}E_{\mathrm{ind}}^{(22)} + {}^{t}E_{\mathrm{exch\text{-}ind}}^{(22)} \tag{2.106}$$

and

$$E_{\mathrm{int}}^{\mathrm{SAPT2+}} = E_{\mathrm{int}}^{\mathrm{SAPT2}} + E_{\mathrm{disp}}^{(21)} + E_{\mathrm{disp}}^{(22)}. \tag{2.107}$$

The inclusion of (2) forms the SAPT2+(3) and SAPT2+3 methods:

$$E_{\text{int}}^{\text{SAPT2+(3)}} = E_{\text{int}}^{\text{SAPT2+}} + E_{\text{elst,rep}}^{(13)} + E_{\text{disp}}^{(30)} \tag{2.108}$$

and

$$E_{\text{int}}^{\text{SAPT2+3}} = E_{\text{int}}^{\text{SAPT2+(3)}} + E_{\text{exch-disp}}^{(30)} + E_{\text{ind-disp}}^{(30)} + E_{\text{exch-ind-disp}}^{(30)} - \delta_{\text{HF}}^{(2)} + \delta_{\text{HF}}^{(3)}, \tag{2.109}$$

where

$$\delta_{\text{HF}}^{(3)} = E_{\text{int}}^{\text{HF}} - E_{\text{elst}}^{(10)} - E_{\text{ind,resp}}^{(20)} - E_{\text{exch-ind,resp}}^{(20)} - E_{\text{ind,resp}}^{(30)} - E_{\text{exch-ind,resp}}^{(30)}. \tag{2.110}$$

The superscript t in Equation 2.106 indicates that this is the true second-order correlation, not already included for in $E_{\text{ind,resp}}^{(20)}$, and is obtained as

$${}^{t}E_{\text{exch-ind}}^{(22)} = {}^{t}E_{\text{ind}}^{(22)} \left(\frac{E_{\text{exch-ind}}^{(20)}}{E_{\text{ind}}^{(20)}} \right). \tag{2.111}$$

Furthermore, the $E_{\text{ind,resp}}^{(30)}$ correction introduced in Equation 2.110 is not calculated explicitly, but is determined as

$$E_{\text{exch-ind,resp}}^{(30)} \approx E_{\text{exch-ind}}^{(30)} \cdot \frac{E_{\text{ind,resp}}^{(30)}}{E_{\text{ind,resp}}^{(30)}}. \tag{2.112}$$

2.3.3 Further improvements to SAPT

Several further improvements to the computation of the SAPT interaction energy can be made, including the use of density fitting,[232,233] CC amplitudes,[231,234] and the extension of the closed-form solution of all exchange-containing terms to the second order. The relevant improvements applied to this work are discussed here.

In terms of computational efficiency, density fitting can be implemented, for which four-index quantities (two-electron integrals and dispersion amplitudes) are approximated as

the linear combination of three-index quantities.[232] This offers a significant improvement, as it largely reduces the memory required for SAPT0 computations, allowing for SAPT to be applicable to larger systems. This also improves the scaling of some higher-order terms in the many-body SAPT levels; however, for SAPT2+ and higher, the terms affected by density fitting are the most memory-intensive part of the calculation, and further techniques are required for a significant speed-up.[229]

The accuracy of SAPT can be improved with a coupled-cluster doubles (CCD) treatment of dispersion.[235] Due to the nature of SAPT, the dispersion terms and the influence of intramonomer correlation on dispersion are only treated in a perturbative manner. However, using the CCD+ST(CCD) approach of Williams et al.,[234] the dispersion amplitudes can be treated analogously to the CCD approach, allowing for a coupled treatment of the electron correlation effects, which includes higher order excitations. At the SAPT2+ level, the second-order dispersion is then expressed as:[233]

$$\begin{aligned} E_{\text{disp}}^{\text{SAPT2+(CCD)}} = E_{\text{exch-disp}}^{(20)} + E_{\text{disp}}^{(2)}[\text{CCD}] \\ + E_{\text{disp}}^{(22)}[\text{S(CCD)}] + E_{\text{disp}}^{(22)}[\text{T(CCD)}], \end{aligned} \tag{2.113}$$

in which CCD dispersion and perturbative corrections due to singles and triples are added to exchange-dispersion. Parrish et al.[235] highlight that this approach adds five sets of iterative equations, each scaling as $\mathcal{O}(N^6)$, making the calculation of the CCD dispersion the rate-limiting step. It is also noted that the additional cost of higher orders of SAPT is negligible compared to the cost associated with CCD-calculated dispersion. Thus, while improving the description of dispersion, the computational cost may outweigh the need for greater accuracy.

An additional correction can be made to account for third-order and higher coupling between induction and dispersion terms that are not included in the $\delta_{\text{HF}}^{(2)}$ correction (Equation 2.106). This correction, δE_{MP2}, is obtained from a supermolecular MP2 calculation

such that

$$\delta E_{\text{MP2}} = E_{\text{int}}^{\text{MP2}} - E_{\text{int}}^{\text{SAPT2}}. \tag{2.114}$$

2.3.4 The single-exchange approximation

With the SAPT formalism, the "symmetry-adapted" nature refers to the antisymmetrization of the wavefunctions of each monomer, giving rise to the exchange terms. The formulae used to compute exchange can become complicated with an increase in order of perturbation, as the exchange between all electrons needs to be considered. As such, only the first-order exchange, $E_{\text{exch}}^{(10)}$, is computed with the full form of the antisymmetrizer, while higher-order exchange terms are typically calculated with the single-exchange approximation (SEA).[207] The SEA does perform well but is known to break down at shorter distances for higher-order terms. As such, Parker et al.[233] recommend the scaling of approximated terms by a scaling factor $p_{\text{EX}}(\alpha)$, determined using the ratio of the non-approximated $E_{\text{exch}}^{(10)}$ to its approximated equivalent, $E_{\text{exch}}^{(10)}(S^2)$, such that

$$p_{\text{EX}}(\alpha) = \left(\frac{E_{\text{exch}}^{(10)}}{E_{\text{exch}}^{(10)}(S^2)} \right)^{\alpha}, \tag{2.115}$$

where $\alpha = 1$ should be used, but it has been shown that $\alpha = 3$ is more appropriate for more weakly bound systems.[233] While typically obtained with the SEA, the closed-form solution, using the full form of the antisymmetrizer, has been extended and implemented for the second order $E_{\text{exch-ind,resp}}^{(20)}$[236] and $E_{\text{exch-disp}}^{(20)}$[237] terms. This allows for an improved accuracy of the higher-order exchange terms.

2.3.5 Comparing a classical and quantum mechanical description of interaction energy components

While classical models provide an approximate description of the components contributing to the intermolecular forces, they fail to account for the complexities of quantum mechanical interactions. For example:

- **Electrostatics**: In classical FFs, electrostatic interactions are described using fixed charges and Coulomb's law, assuming point charges. In contrast, SAPT computes electrostatic energy based on charge densities, which leads to a more accurate representation of real systems.

- **Steric (repulsive) interactions**: Classical FFs use empirical terms, such as the exchange-repulsion term in the Lennard-Jones potential, to model steric effects. Quantum mechanically, steric repulsion arises naturally from the exchange-repulsion term in SAPT, linked to wavefunction antisymmetry.

- **Dispersion forces**: These are modelled in classical potentials using an empirical attractive term. However, dispersion forces are inherently quantum mechanical and arise from electron correlation effects, which SAPT explicitly calculates.

Therefore, using the SAPT energy components for FF development will not only enhance the accuracy of molecular dynamics simulations but also bridge the gap between classical and quantum mechanical descriptions of intermolecular interactions, laying the groundwork for future advancements in the field of molecular simulations.

2.4 Comparative overview of methods: advantages, limitations and applicability

The theoretical techniques discussed in this chapter offer diverse capabilities for understanding the properties and behaviors of ILs. However, each method has unique advantages, disadvantages, and areas of optimal application, making it important to contextualize the strengths and limitations.

MD simulations serve as a foundational tool for exploring the atomistic and dynamic properties of ILs. The strength of this method is the ability to predict macroscopic transport properties, such as diffusion coefficients and ionic conductivities, through the analysis of time-resolved particle trajectories. However, the accuracy of MD simulations is fundamentally dependent on the FF employed. Classical FFs, while computationally efficient, do not accurately capture polarisation effects, limiting the reliability when applied to systems with significant electronic effects. MD simulations are well-suited for studying structural organization and large-scale transport phenomena, particularly in complex and heterogeneous IL systems. Polarisable FFs improve upon classical approaches by incorporating electronic polarisation effects, particularly important in systems dominated by strong ion-ion or ion-solvent interactions. This allows for more accurate predictions of properties such as ion pairing, solvation dynamics, and ionic mobility. Nevertheless, the increased computational expense of polarisable FFs makes the use challenging for large systems or extended simulation timescales. Despite this limitation, these methods are indispensable for IL systems where polarisation plays a pivotal role.

QM methods offer extreme precision and insight into the electronic structure of ILs. Techniques such as DFT and wavefunction-based methods enable researchers to model molecular geometries, electronic properties, and interaction energies with high accuracy. DFT is a widely used QM approach due to its balance between computational efficiency and accuracy, making it suitable for medium-sized systems. However, it can struggle with dis-

persion forces, which are crucial in ILs, unless dispersion corrections are included. On the other hand, wavefunction-based methods such as MP2 and CC provide highly accurate results, particularly for systems dominated by strong electronic effects or subtle interactions like charge transfer. These methods, however, are computationally intensive and generally applied to smaller systems or clusters. SAPT is an advanced QM method specifically tailored to decompose interaction energies into physically meaningful components such as electrostatics, dispersion, induction, and exchange-repulsion. This capability is particularly valuable for parameterizing FFs and for understanding fundamental non-covalent interactions in ILs. SAPT provides unmatched clarity in elucidating the nature of ion-ion and ion-solvent interactions, but, like other QM methods, its application is constrained to small molecular systems due to high computational costs.

In summary, MD simulations provide an accessible framework for investigating transport and macroscopic properties of ILs, while polarisable FFs extend this approach to systems requiring greater accuracy in representing electronic effects. Quantum mechanical methods, though computationally intensive, are indispensable for precise interaction studies and parameter development.

Part I

Calculating Thermodynamic and Transport Properties of Ionic Liquids and Ionic Liquid-Based Electrolytes

Chapter 3

Property calculation methodology

Favourable transport properties of ionic liquids (ILs) such as high diffusivity, and high ionic conductivity are important for peak electrochemical performance, and an in-depth understanding of the structural features that control these properties is key to the development of improved electrolytes. Designing ILs for such an application (as potential electrolytes) requires a cation-anion combination that will yield the required properties. Therefore, an understanding of structure-property relationships — how structural changes will influence physical properties — is required.[238–240]

Molecular dynamics (MD) simulation can complement experimental research to understand how thermodynamic and transport properties of ILs are related to each other and influenced by the underlying molecular structure. Simulation can also be invaluable for predicting properties that might otherwise be lacking when studying the effect of systematic structural changes.[171,241] However, the statistical uncertainties in MD-calculated properties of ILs are generally greater than alluded to in literature,[169] and quantification of the precision of calculated properties is essential to provide a validation study.[180] Therefore, it is necessary to develop a protocol which allows for the reproducibility of MD simulations of highly-viscous ILs and to furthermore provide a framework to characterise the uncertainty of calculated properties, using the CL&Pol FF.

3.1 Statistical measures of error and uncertainty

The Ergodic hypothesis makes the assumption that an MD trajectory will, over sufficiently long simulation times, explore all accessible microstates at equilibrium. This principle is foundational for MD simulations, where time averages from a single trajectory are assumed to represent ensemble averages. However, practical limitations often prevent exhaustive sampling of the energy landscape, leading to incomplete equilibrium distributions and statistical uncertainties in calculated properties.[242]

The importance of providing a meaningful estimate of uncertainty in any simulated result has been discussed before.[180] As pointed out by McDaniel et al.,[169] the statistical uncertainties in MD-calculated properties of ILs are generally greater than alluded to in literature. Therefore, while a measure of error and deviation is provided for all calculated values, uncertainty is characterised in each calculated property using the 95 % confidence interval for the mean, calculated over five or ten independent simulations.

3.1.1 Quantifying error and deviation

The average absolute relative deviation (AARD) is the average deviation across a calculated value, $\hat{y}_i$ from a reference value, y_i, assessed over a set of n points, determined as

$$\text{AARD \%} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100. \tag{3.1}$$

Where experimental data is available, y_i is the corresponding measured value of the calculated $\hat{y}_i$ and the AARD is then referred to as a mean absolute percentage error (MAPE).

3.1.2 95 % confidence intervals

Where properties are reported as the average ($\bar{x}$) over the n trajectories used for the given simulation type, the uncertainty is indicated with the 95 % confidence interval (CI), calcu-

Table 3.1: t-distribution based on the degrees of freedom $(n-1)$ for over which a mean value is calculated.

$\mathbf{n-1}$	$\mathbf{t_{n-1}}$
4	2.78
5	2.57
6	2.45
7	2.37
8	2.31
9	2.26

lated as

$$95\ \%\ \text{CI} = \bar{x} \pm t_{n-1} \times \frac{s}{\sqrt{n}}. \tag{3.2}$$

This requires the standard deviation of the mean, s,

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}, \tag{3.3}$$

and the value of t_{n-1}, taken from a t-distribution, indicated in Table 3.1.

3.2 Calculation methodology

3.2.1 Density

Densities were calculated using the *NPT* trajectories, of which the first 4 ns were discarded as equilibration. The density, ρ, is calculated from the average volume recorded every 10 ps from the post-equilibration trajectory.

3.2.2 Self-diffusion coefficients

The self-diffusion coefficient, D_s, was calculated using the well-known Einstein relation,[181]

$$D_\text{s} = \frac{1}{6} \lim_{t\to\infty} \frac{d}{dt} \langle [\mathbf{r}(t) - \mathbf{r}(0)]^2 \rangle, \tag{3.4}$$

where the average mean-squared displacement (MSD) is calculated as

$$\left\langle \left[\mathbf{r}(t) - \mathbf{r}(0)\right]^2 \right\rangle = \frac{1}{t_s - t + 1} \sum_{t_0=0}^{t_s - t} \frac{1}{N} \sum_{i=1}^{N} \left[\boldsymbol{r}_i(t_0 + t) - \boldsymbol{r}_i(t_0)\right]^2 . \tag{3.5}$$

Here, $\mathbf{r}_i$ is the centre-of-mass coordinates of ion i, t_s is the total number of time steps, t_0 is the time origin at each step considered and N is the total number of molecules in the trajectory. The MSD is determined every 1 ps for each unique ion in the *NVT* trajectories. While the MSD is computed over the entire trajectory, only the linear region is used for the calculation of D_s. To ensure that only the linear region was used, the following protocol was followed: A plot of the natural logarithms of $t - t_0$ and MSD was obtained. The slope of the plot, β, is calculated to determine where the power of the relationship between the two is linear, ie., where $\beta = 1$. While a slope of 1 indicates the linear region, this can be a rather strict criterion, reducing the number of points included in the fit. Therefore, a tolerance is introduced, and the linear region is determined where the following criteria are met:

$$\beta = \frac{\mathrm{d}}{\mathrm{d}t} \frac{\ln \mathrm{MSD}}{\ln t} = 1 \pm 0.15 \tag{3.6}$$

The MD simulations carried out in this work make use of periodic boundary conditions (PBC) and as such, an inherent error exists due to the underprediction of hydrodynamic self-interactions in the finite (periodic) system. Diffusion coefficients calculated from finite cubic simulations can be corrected to reproduce those of an infinite (non-periodic) system using the correction of Yeh and Hummer,[243,244]

$$D_\mathrm{s}^{\infty} = D_\mathrm{s}^{\mathrm{PBC}} + \Delta D_\mathrm{s}^{\mathrm{YH}} = D_\mathrm{s}^{\mathrm{PBC}} + \frac{\zeta k_B T}{6\pi\eta L}. \tag{3.7}$$

Here, $D_\mathrm{s}^{\mathrm{PBC}}$ and D_s^{∞} are the values of the PBC-calculated and corrected values of D_s, respectively, k_B is the Boltzmann constant, T is the temperature, η is the shear viscosity, L is the length of the simulation box and the constant ζ is equal to 2.837297.[243] In this work,

the average diffusion coefficients for each ion in a system are corrected with Equation 3.7 using the average value of η of the total system, obtained as described below in Section 3.2.3.

3.2.3 Viscosity

The shear viscosity, η, was obtained using the Green-Kubo expression,[181]

$$\eta = \frac{V}{6k_{\mathrm{B}}T} \sum_{\alpha \leq \beta} \int_0^\infty \langle \bar{P}_{ab}(t) \cdot \bar{P}_{ab}(0) \rangle \, \mathrm{d}t, \tag{3.8}$$

where V is the box volume and the angular brackets indicate the ensemble average of the autocorrelation function (S_{ACF}) of the ab elements of the pressure tensor, $\bar{P}_{\alpha\beta}$. To improve the quality of the pressure tensor data used, the average is taken over the off-diagonal elements (where $\bar{P}_{ab} = \bar{P}_{ba}$ for all (x, y, z)) and the following modified diagonal elements:[245,246]

$$\begin{aligned} \bar{P}_{xx} &= \tfrac{1}{2} \left(P_{xx} - P_{yy} \right) \\ \bar{P}_{yy} &= \tfrac{1}{2} \left(P_{yy} - P_{zz} \right) \\ \bar{P}_{zz} &= \tfrac{1}{2} \left(P_{xx} - P_{zz} \right) \end{aligned} \tag{3.9}$$

A known issue with the calculation of η is that long-time oscillations in S_{ACF} produce noisy running integrals. One approach is to use more shorter simulations over fewer longer simulations, as to remove noise through averaging.[180,247] With respect to finding the plateau value of the running integral of S_{ACF}, several approaches can be applied, and were tested, with the comparison given in Section 4.5.3.

3.2.4 Ionic conductivity

The ionic conductivity can be obtained as the Nernst-Einstein (NE) ionic conductivity (σ_{NE}) calculated as[248,249]

$$\begin{aligned}\sigma_{\text{NE}} &= \frac{e^2}{6k_BTV}\lim_{t\to\infty}\frac{d}{dt}\sum_{k}^{M} z_k^2\Big\langle\big[\mathbf{r}(t)-\mathbf{r}(0)\big]^2\Big\rangle_k \\ &= \frac{e^2}{6k_BTV}\sum_{k}^{M} n_k x_k^2 D_{\text{s},k}\end{aligned} \tag{3.10}$$

where M is the number of different ion types and z_k, n_k and $D_{\text{s},k}^{\infty}$ are the charge, number of ions and self-diffusion coefficient of ion type k, respectively, in the simulation ensemble. The NE conductivity is thus entirely dependent on the relative motion of ions of the same type and fails to account for the correlation between ions of different types. To obtain the "true" ionic conductivity, accounting for ion correlation, the Einstein-Helfand (EH) conductivity, σ_{EH}, can be obtained using the Einstein relation,

$$\sigma_{\text{EH}} = \frac{e^2}{6k_BTV}\lim_{t\to\infty}\frac{d}{dt}\sum_{k,l} z_k z_l\Big\langle\big[\mathbf{r}_k(t)-\mathbf{r}_k(0)\big]\cdot\big[\mathbf{r}_l(t)-\mathbf{r}_l(0)\big]\Big\rangle, \tag{3.11}$$

where e is the elementary charge, z_i and z_j are the charge numbers of ions i and j, and the summation is over all N ions in the simulation. The angular brackets here contain the collective MSD. That being, the MSD is considered between all ions, not just ions of the same type.

The σ_{NE} and σ_{EH} values are calculated from the MSD and collective MSD, respectively, where the linear region (for both) is required. As σ_{EH} is computed using diffusion coefficients, which are taken from a suitably linear region, this requires no further consideration. To determine the appropriate region required to calculate σ_{NE}, the linear region of the collective MSD is determined using the same approach as for D_{s}, using the criteria analogous to Equation 3.6. Conductivities are calculated using the NVT trajectories.

3.2.5 Enthalpies of vaporisation

The molar enthalpy of vaporisation at a given temperature, $\Delta_{\text{vap}}H(T)$, can be calculated as

$$\Delta_{\text{vap}}H(T) = U(\text{g}) - \frac{U(\text{l})}{N} + P\big(V(\text{g}) - V(\text{l})\big), \tag{3.12}$$

where $U(\text{g})$ and $U(\text{l})$ are the molar potential energies of the gas and liquid phases, respectively, P is the pressure of 1 bar and $V(\text{g})$ and $V(\text{l})$ are the gas and liquid phase molar volumes, respectively, at the given temperature T. As $V(\text{g})$ is significantly larger than $V(\text{l})$, the latter can be considered negligible. Furthermore, if the vapour is described using the ideal gas law then Equation 3.12 can be expressed as

$$\Delta_{\text{vap}}H = U(\text{g}) - \frac{U(\text{l})}{N} + RT. \tag{3.13}$$

The values of $U(\text{l})$ are obtained from the NVT trajectories of the system being investigated, and $U(\text{g})$ is taken from an NVT trajectory for a representative "molecule". That being, for a neat IL, $U(\text{g})$ will be that of a single cation-anion pair.

3.2.6 Ion-pair lifetimes

In an IL consisting of ions α and β, a tagged ion pair (IP) is identified as the closest ion α_i around a central ion β_j, or vice versa. For every ion pair (with either type α or β in the centre), a function $p(t)$ is defined that is equal to one if α_i is the ion closest to β_j at time t and zero otherwise. The correlation function $C(t)$ is then calculated as[250,251]

$$C(t) = \frac{\langle \delta p(0)\delta p(t)\rangle}{\langle \delta p^2\rangle} = \frac{\langle p(0)p(t)\rangle - \langle p\rangle^2}{\langle p\rangle - \langle p\rangle^2} \approx \frac{\langle p(t)p(0)\rangle}{\langle p\rangle}, \tag{3.14}$$

where the angle brackets denote the ensemble average over all IPs at all considered time points. The average number of ion pairs is denoted as $\langle p\rangle$. This correlation function expresses the probability that a pair of tagged ions is intact at time t, given that it was

intact at time zero. The relaxation or IP lifetime, τ_{IP}, is defined as

$$\tau_{\text{IP}} = \int_0^\infty C(t)dt, \tag{3.15}$$

and can be evaluated analytically by fitting $C(t)$ to a multiple-exponential function,[251]

$$C(t) = \sum_i a_i e^{-t/b_i}, \tag{3.16}$$

where a_i and b_i are fitting constants.

As highlighted by Zhang and Maginn,[252] discrete IPs do not exist in a liquid, and one should rather consider the lifetime of an ion cage (IC) as a time-based measure of the cation-anion interaction.[252] The lifetime of an IC expresses the persistence of a static solvation shell of anions around a central cation, or vice versa, and can be calculated similarly to the IP lifetime using an adapted definition of $p(t)$.[251,252] However, τ_{IP} was shown to correlate linearly with IC lifetimes[252] and consequently, while the concept of an IP in the liquid should not be taken too literally, τ_{IP} can be used equally well to gain insight into the strength of ion-ion interaction. Average IP lifetimes were calculated from the NVT trajectories using the PyLAT software package,[253] in which the first picosecond of the NVT trajectory was excluded when fitting to $C(t)$. Eqn 3.16 was used to calculate the relaxation time with the sum expanded to three exponential functions.

Chapter 4

Validation of the CL&Pol force field

4.1 Introduction

In this chapter, molecular dynamics (MD) simulations are used to validate the accuracy of the CL&Pol force field (FF) by calculating a comprehensive set of thermodynamic and transport properties of ionic liquids (ILs) consisting of the 1-butyl-3-methylimidazolium ($[C_4C_1im]^+$) and *N*-butyl-*N*-methylpyrrolidinium ($[C_4C_1pyr]^+$) cations paired with the bis(trifluoromethanesulfonyl)imide ($[TFSI]^-$), bis(fluorosulfonyl)imide ($[FSI]^-$) and (fluorosulfonyl)(trifluoromethanesulfonyl)imide ($[FTFSI]^-$) anions, illustrated in Figure 4.1. These simulations are used to calculate densities, self-diffusion coefficients, ionic conductivities, viscosities, enthalpies of vaporisation and ion-pair lifetimes. In addition, replicate simulations are used to quantify uncertainty of the calculated properties. The accuracy of the FF is assessed by comparing the calculated properties to experimental measurements and the precision of the simulation methodology, which is often overlooked in validation studies, is quantified by reporting the 95 % confidence intervals (CIs) for the mean, calculated using the independent replicate trajectories. The calculation methodology for properties and uncertainty measurement presented in Section 3 is applied, for which an internal validation is presented to highlight the robustness of the methods used.

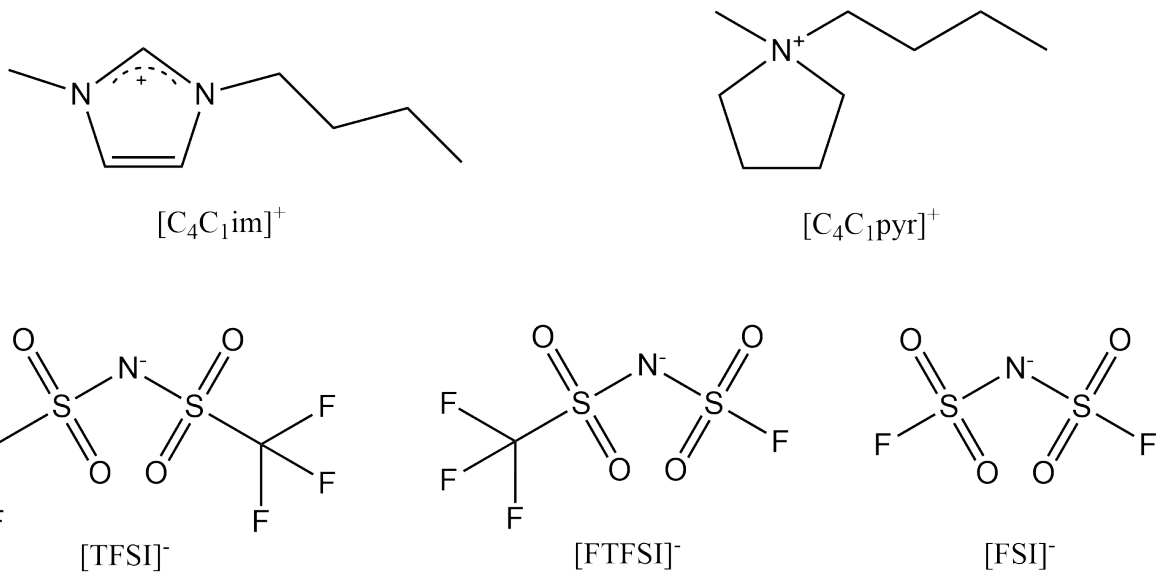


Fig. 4.1. Structural representations of the 1-butyl-3-methylimidazolium ($[C_4C_1im]^+$) and *N*-butyl=*N*-methylpyrrolidinium ($[C_4C_1pyr]^+$) cations and bis(trifluoromethylsulfonyl)imide ($[TFSI]^-$), bis(fluorosulfonyl)imide ($[FSI]^-$) and (fluorosulfonyl)(trifluoromethylsulfonyl)imide ($[FTFSI]^-$) anions forming the six ILs studied in this chapter.

4.2 Force Field Details

The CL&Pol FF was implemented as described in Section 2.1.5. The Lennard-Jones (LJ) σ_{ij} parameter values are scaled by 0.985 due to a known deficit in CL&Pol calculated densities without σ_{ij} scaling,[148] and the LJ ε_{ij} parameters are scaled to remove the induction contribution, using the fragment approach framework, developed by Goloviznina et al.[148] The representative fragments, required to obtain the ε_{ij} scaling factors, k_{ij}, used in this work are: (i) cation-head fragments: 1-ethyl-3-methylimidazolium ($[C_2C_1im]^+$) and *N*,*N*-dimethylpyrrolidinium ($[C_1C_1pyr]^+$); (ii) anion fragments: $[TFSI]^-$, $[FSI]^-$ and $[FTFSI]^-$ and (iii) neutral fragment: ($[C_4H_{10}]$), used to represent the alkyl side chain in each IL. The $[FTFSI]^-$ anion had not been previously parameterized with the CL&Pol FF and as such, no values of k_{ij}^{SAPT} or the required data for the calculation of k_{ij}^{pred} values were available in the supplied FF files. Therefore, the required fragment pairs were generated

and optimised, and both k_{ij}^{SAPT} and k_{ij}^{pred} were computed, for each fragment pair, to provide a comparison of the two (see Section 4.3 below).

4.3 Comparison of calculation procedures used to obtain k_{ij}

As detailed in Section 2.1.5, there are two variations of CL&Pol which can be used to determine the scaling factor k_{ij}. The two approaches are repeated here for convenience:

1. *CSDrude*, where the scale factor is determined using the dispersion and induction energy components of a Symmetry-Adapted Perturbation Theory (SAPT)[168] energy decomposition, computed with the SAPT2+/aug-cc-pVDZ method:

$$k_{ij}^{\text{SAPT}} = \frac{E_{\text{disp}}}{E_{\text{disp}} + E_{\text{ind}}} \tag{2.52}$$

2. *KCSDrude*, where the scale factor between fragments i and j is determined from the empirical scheme:

$$k_{ij}^{\text{pred}} = \left(1 + 0.25 r_{ij}^2 \frac{Q_i^2\alpha_j + Q_j^2\alpha_i}{\alpha_i\alpha_j} + 0.11 \frac{\mu_i^2\alpha_j + \mu_j^2\alpha_i}{\alpha_i\alpha_j}\right), \tag{2.53}$$

The $KCSDrude$ FF is the final version implemented by Goloviznina et al.[148] in subsequent work. While the initial development of the fragment approach shows that the k_{ij}^{pred} values are comparable to the k_{ij}^{SAPT} values, not all fragment pairs used in this work were used in the original development study and as such, a comparison including the new fragments was warranted.

The protocol followed to obtain k_{ij}^{pred} values, as well as a comparison of k_{ij}^{pred} and k_{ij}^{SAPT}, is presented below for the fragment pairs for which scale factors were not available:

$[C_1C_1pyr]^+\cdots[FSI]^-$, $[C_1C_1pyr]^+\cdots[FTFSI]^-$, $[C_2C_1im]^+\cdots[FTFSI]^-$, $[C_4H_{10}]\cdots[FSI]^-$ and $[C_4H_{10}]\cdots[FTFSI]^-$.

4.3.1 Ion-pair geometries

A conformer search for each fragment pair, as listed above, was carried out with ABCluster version 3.0.[254–256] The GFN2-xTB[257] Hamiltonian was employed, using flexible monomers with a maximum of 500 energy evaluations and an energy gap of 0.0001 au, such that conformers within this energy gap were treated as identical. The three lowest energy conformers were then saved for optimisation by DFT. Following optimisation, if conformers appeared structurally only one of the "duplicates" was used for further analysis. The optimised conformers obtained for each fragment are shown in Figure 4.2. The isolated $[FTFSI]^-$ anion and the chosen ABCluster-generated fragment pairs were optimised, as done by Goloviznina et al.,[148] using the Gaussian 16 software package[258] and the B97[259] density functional method with Grimme's dispersion and Becke-Johnson damping (D3BJ),[260] and the cc-pVDZ basis set. The lowest energy optimised structure for each dimer was then used for subsequent calculations. From the ion geometry optimisation, the dipole moment of $[FTFSI]^-$ was determined, using the origin at the centre of nuclear charge, as 3.5902 Debye. All other dipole moments used in this work were taken from the supplied CL&Pol FF parameter files.[148]

4.3.2 Calculation and comparison of $k_{ij}^{\mathbf{pred}}$ and $k_{ij}^{\mathbf{SAPT}}$

For the five fragment pairs listed above, r_{ij} values were calculated for each optimised conformer. The k_{ij}^{pred} values were then determined using Equation 2.53, and are given in Table 4.1. SAPT energy decompositions were carried out using the Psi4 software package[261,262] for each of the ion pairs above with the SAPT2+ approximation and aug-cc-pVDZ basis set, with each SAPT2+ energy term defined as given by Jeziorski et al.[207] From the SAPT

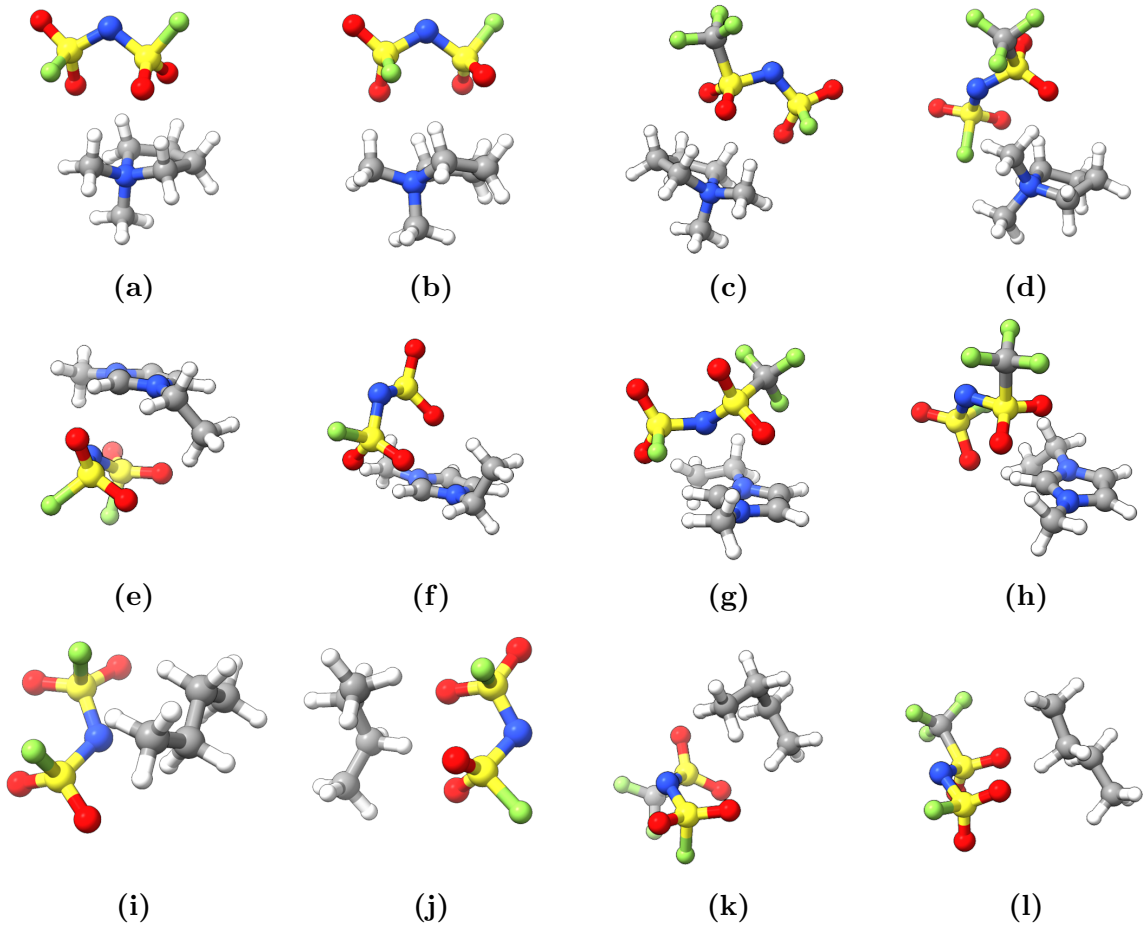


Fig. 4.2. Conformers used for calculation of k_{ij}^{pred} and k_{ij}^{SAPT} for the $[C_1C_1pyr]^+\cdots[FSI]^-$ (**a** and **b**), $[C_1C_1pyr]^+\cdots[FTFSI]^-$ (**c** and **d**), $[C_2C_1im]^+\cdots[FTFSI]^-$ (**e** and **f**), $[C_4H_{10}]\cdots[FSI]^-$ (**g** and **h**) and $[C_4H_{10}]\cdots[FTFSI]^-$ (**i** and **j**) fragment pairs

computations, the induction energy, E_{ind}, and dispersion energy, E_{disp}, were obtained and k_{ij}^{SAPT} values were calculated according to Equation 2.52 and are reported in Table 4.2.

The values presented in Table 4.2 show that the values of k_{ij}^{pred} are typically lower than k_{ij}^{SAPT}, except for the $[C_2C_1im]^+\cdots[FSI]^-$ pair. This suggests that using the k_{ij}^{pred} scaling factors could underestimate the LJ contribution and faster dynamics could be expected; given that the k_{ij}^{SAPT} values would be expected to produce more accurate results . However, the k_{ij}^{pred} values are less sensitive to the conformer used, as the COM-distances are fairly consistent regardless of conformer. Furthermore, given that the k_{ij}^{pred} values calculated here deviate from the corresponding k_{ij}^{SAPT} values by a similar magnitude as those previously

Table 4.1: Values of equilibirum centre-of-mass distances for each conformer considered for fragments i and j, r_{ij}, with the corresponding calculated values of k_{ij}^{pred} and k_{ij}^{SAPT} for each.

i	j	Conformer	r_{ij}	k_{ij}^{pred}	k_{ij}^{SAPT}
$[C_2C_1im]^+$	$[FTFSI]^-$	1	3.858	0.58	0.73
		2	4.314	0.55	0.65
$[C_2C_1im]^+$	$[FSI]^-$	1	3.77	0.61	0.62
		2	3.377	0.66	0.6
$[C_1C_1pyr]^+$	$[FTFSI]^-$	1	4.499	0.51	0.5
		2	4.5397	0.5	0.62
$[C_1C_1pyr]^+$	$[FSI]^-$	1	4.267	0.54	0.6
		2	4.182	0.55	0.68
$[C_4H_{10}]$	$[FTFSI]^-$	1	4.55	0.67	0.71
		2	3.947	0.71	0.77
$[C_4H_{10}]$	$[FSI]^-$	1	4.13	0.7	0.69
		2	3.865	0.74	0.74

reported by Goloviznina et al.[148] where k_{ij}^{pred} was used, the k_{ij}^{pred} values determined here were deemed sufficient and are used in this work. For consistency, the final values of r_{ij} and subsequently k_{ij}^{pred} were determined for the lowest energy conformer in each instance. All values of k_{ij}^{pred} used in this work are given in Table 4.2.

4.4 Simulation details

All molecular dynamics (MD) simulations were carried out using the LAMMPS software package.[263] Periodic cubic boxes containing 150 ion pairs with side lengths of 45 Å[I] were generated with Packmol,[264] for which the required input and data files were generated with the `fftool` utility.[265] Hydrogen-containing bonds were constrained using the SHAKE algorithm with a maximum of 20 iterations and a tolerance of 1.0×10^{-4}. A 12 Å cutoff was applied to the LJ and real-space Coulombic pair potentials and a long-range tail correction was included for total energy and pressure.[266] For all pair interactions between two Drude particles (DPs) within the 12 Å cutoff, the Thole damping function was used

I Using 150 ion pairs in initial boxes of 45 Å were chosen as this gave the initial configurations densities of approximately that of the experimental values (ranging from 1.301 to 1.436 g/mol).

Table 4.2: Values of equilibirum centre-of-mass distances for fragments i and j, r_{ij}, with the corresponding calculated values of k_{ij}^{pred} for each fragment pair used in this work. Values marked with an asterisk were not calculated in this work, and were reported by Goloviznina et al.[148]

i	j	r_{ij}/ Å	k_{ij}^{pred}
$[C_2C_1im]^+$	$[TFSI]^-$	4.314*	0.55
$[C_2C_1im]^+$	$[FTFSI]^-$	3.858	0.58
$[C_2C_1im]^+$	$[FSI]^-$	3.377*	0.66
$[C_1C_1pyr]^+$	$[TFSI]^-$	4.725*	0.50
$[C_1C_1pyr]^+$	$[FTFSI]^-$	4.499	0.51
$[C_1C_1pyr]^+$	$[FSI]^-$	4.267	0.54
$[C_4H_{10}]$	$[TFSI]^-$	4.436*	0.69
$[C_4H_{10}]$	$[FTFSI]^-$	3.947	0.71
$[C_4H_{10}]$	$[FSI]^-$	3.865	0.74

with a damping parameter of 2.6.[157,204] The particle-particle particle-mesh (PPPM) Ewald summation was used to evaluate long-range electrostatics, with an accuracy of 1×10^{-5} and neighbour lists were built with a skin distance of 3.0 Å. A timestep of 1 fs was used and the equations of motion were integrated using the Verlet integrator. A temperature-grouped Nosé–Hoover thermostat was used to allow better equipartitioning of the kinetic energy.[174,267] Molecular temperature was set at 298.15 K with a relaxation time of 100 fs and the Drude oscillator temperature was set to 1 K with a relaxation time of 20 fs. For *NPT* simulations, the Nosé–Hoover barostat was applied with pressure set to 1 bar, relaxed every 1 ps and the volume of the box was saved every 10 ps to calculate the average density. For *NVT* simulations, the six unique components of the pressure tensor (diagonal and non-diagonal) were saved every 5 fs and configurations were saved every 1 ps. The details of the specific simulation types used in this work are outlined below.

4.4.1 Liquid-phase simulations:

Two types of liquid-phase simulations, **L1** and **L2**, were carried out, both giving a total of 50 ns of production simulation time. The **L1** simulation protocol is outlined below, with the **L2** protocol indicated in parentheses.

1. For each IL, 5 (10) independent initial configurations were generated;

2. For each initial configuration, an energy minimization was carried out with a maximum of 1000 iterations of force/energy evaluations. The minimization criteria included an energy change between iterations less than 1.0×10^{-4} (unitless) or a maximum force on any atom less than 1.0×10^{-6} newton;

3. Each system was then heated for 500 ps, from 100 K to 298.15 K, in the *NVT* ensemble;

4. The systems were then equilibrated for 5 ns (2 ns) in the *NPT* ensemble;

5. For each replicate trajectory in step 4, the density was averaged over the last 3 (1) ns of the simulation. The final saved configuration was scaled to give this density and the NVT runs started with this initial configuration.

4.4.2 Gas-phase simulations:

1. For each IL, a single ion pair in a simulation box with side dimensions of 30 Å was generated;[II]

2. An energy minimization as for the **L1**/**L2** simulations was performed;

3. 10 ns production trajectories in the *NVT* ensemble were generated.

4.5 Calculated structural and transport properties

MD simulations were carried out for the six ILs studied here, following the methodology given in Section 4.4. All properties were calculated using the protocols outlined in Section 3.2, and all calculated values are summarised in Table 4.3 and a discussion of these properties follows. Properties were calculated for each trajectory and are reported as the

II It can be assumed that the gas phase predominantly consists of ion pairs.[268–271]

average, with the 95 % CI of the mean as the uncertainty, determined as described in Section 3.1.2.

When comparing simulation results to experimental data, potential error in the experimental measurements must also be considered. Experimental values are often subject to variability due to factors such as contamination of the IL by water from incomplete drying or exposure to atmospheric moisture. Additionally, impurities such as halide ions or residual solvents may inadvertently alter the physical properties being measured.[61] Such issues can lead to discrepancies between reported values and the actual behavior of the system. The precision of experimental methods and the careful handling of IL samples can vary significantly between studies, further contributing to variability in the data. As a result, when benchmarking simulations against experimental results, it is important to consider the associated experimental uncertainties and recognize that the "true" system being measured may deviate from the simulated system. As such, a 95 % CI is also provided for the experimental values. The experimental 95 % CIs for the ILs used in this work are taken from ILThermo.[60,272]

4.5.1 Density

The average density, ρ, for each IL is given in Table 4.3 and illustrated in Figure A.1. It is worth noting that the standard deviation of the density for each individual simulation is $\sim 0.007\,\mathrm{g/cm^3}$, showing little variation in the density over the trajectory region used to determine an average density.[III] The calculated densities are in excellent agreement with experimental data,[95,273,279,282] where available, with a mean absolute percentage error (MAPE) of 1.1 % and 95 % CIs comparable to those estimated for experiment; however, the density of $[C_4C_1im][FSI]$ is noticeably overestimated with a relative error of 2.7 %. The previous application of the CL&Pol FF to $[C_4C_1im][TFSI]$[53,148,283] and $[C_4C_1pyr][TFSI]$[148]

III Exemplar plots of the change in simulation density are given Appendix A.

Table 4.3: Calculated and experimental (where available) physical and transport properties for the ILs simulated in this work at 298.15 K. Values in parentheses give the 95 % confidence interval for the mean, calculated in this work for the computed values, and reported from ILThermo[60,272] for experimental values.

IL	ρ/g/cm^3		$D_{s,+}/10^{-11}$ m^2/s		$D_{s,-}/10^{-11}$ m^2/s		η/cP		σ_{EH}/S/m		$\Delta_{vap}H$/kJ/mol	
	Calc.	Exp.	Calc.	Exp.	Calc.	Exp.	Calc.	Exp.	Calc.	Exp.	Calc.	Exp.
[C_4C_1pyr][TFSI]	1.373	1.395[273]	2.4	1.4[274] / 1.8[275]	2.4	1.0[274] / 1.4[275]	27	78[273,276]	0.29	0.28[75]	124.0	143.9[277]
	(0.002)	(0.001)	(0.1)	(1.4)	(0.2)	(1.4)	(12)	(0.6)	(0.03)	(0.01)	(0.3)	(3.9)
[C_4C_1pyr][FTFSI]	1.350	1.357[95]	1.8	2.4[a]	1.8	2.3[a]	41	51[95]	0.27	0.44[95]	134.6	-
	(0.001)	(0.001)	(0.2)		(0.2)		(5)	(0.4)	(0.1)	(0.04)	(0.5)	
[C_4C_1pyr][FSI]	1.311	1.301[273]	0.8	-	1.0	-	46	53[273]	0.14	0.62[278]	141.6	147.3[277]
	(0.003)	(0.001)	(0.2)		(0.2)		(7)	(0.4)	(0.1)	(0.025)	(1.2)	(2)
[C_4C_1im][TFSI]	1.437	1.436[279]	3.7	2.8[280]	3.3	2.2[280]	24	51[281]	0.56	0.39[280]	123.2	130.9[277]
	(0.001)	(0.002)	(0.5)	(1.4)	(0.5)	(1.4)	(3)	(5.4)	(0.2)	(0.013)	(0.3)	(5)
[C_4C_1im][FTFSI]	1.416	-	2.5	-	2.3	-	33	-	0.42	-	128.7	-
	(0.001)		(0.3)		(0.4)		(4)		(0.1)		(0.5)	
[C_4C_1im][FSI]	1.387	1.350[282]	2.1	-	2.0	-	38	33[282]	0.19	0.82[282]	138.3	133.7[277]
	(0.001)	(0.003)	(0.2)	(1.3)	(0.1)	(1.3)	(11)	(1.4)	(0.1)	(0.01)	(0.4)	(5)

a Estimated at 298.15 K using a VFT fit based on data extracted from Fig. 7 in Ref. 85. No experimental uncertainties available on ILThermo for these values.

gave slightly underestimated values of ρ ($\sim 2\,\%$). While a similar underestimation is seen for [C_4C_1pyr][TFSI] a slight overestimation is seen here for [C_4C_1im][TFSI].

The calculated density depends significantly on the values of σ_{ij} in the LJ potential, and it should be noted that these parameters for the F and O atoms in the $[TFSI]^-$ and $[FSI]^-$ anions were recently adjusted to improve a systematic deviation in the densities predicted by CL&P.[284] Klajmon and Červinka[285] reinvestigated the scaling of the σ_{ij} values for a selection of $[C_2C_1im]^+$-based ILs and found improved densities with the original, unscaled CL&P σ_{ij} values. In subsequent work by the same authors,[286] unscaled σ_{ij} values gave densities for [C_4C_1pyr][TFSI] in excellent agreement with experiment; however, the values for [C_4C_1im][TFSI] remained slightly overestimated with a higher error than reported here, despite their use of unscaled σ_{ij} values. It is noted that Klajmon and Červinka used k_{ij}^{SAPT} parameters, whereas k_{ij}^{pred} values were used in this work.

The calculated values of ρ follow the experimentally measured order such that $\rho([TFSI]^-) > \rho([FTFSI]^-) > \rho([FSI]^-)$ for both cations considered. The densities of the imidazolium ILs are consistently greater than those of the pyrrolidinium ILs, as with experiment, indicating that the small errors present do not affect the ability of the FF to accurately reproduce this trend.

4.5.2 Self-diffusion coefficients

Self-diffusion coefficients corrected for finite size effects (see Section 3.2.2) for the cation, $D_{\mathrm{s},+}^{\infty}$, and anion, $D_{\mathrm{s},-}^{\infty}$ in each system were calculated. These values are summarised in Table 4.3 and compared to each other and the experimental values in Figure 4.3. The self-diffusion coefficients were computed from the mean-squared displacements (MSDs) over 3–5 ns. The plots of the β values (see Equation 3.6) are given in Appendix A, from which it is determined that a sufficiently linear region of the MSD is used to compute diffusivity. The self-diffusion coefficients of the ions in several ILs have been measured using pulsed field gradient spin echo (PGSE) NMR;[275,280,287] however, experimental values at 298.15 K

or VFT (Vogel–Fulcher–Tammann) parameters for interpolation to 298.15 K are only available for the ILs studied here that are based on $[TFSI]^-$. For $[C_4C_1pyr][TFSI]$, the errors in $D^{\infty}_{s,+}$ and $D^{\infty}_{s,-}$ are ∼50 % and ∼100 % respectively, whereas for $[C_4C_1im][TFSI]$ the errors are ∼32 % and ∼50 %; self-diffusivity is therefore overestimated for both $[C_4C_1pyr][TFSI]$ and $[C_4C_1im][TFSI]$. The 95 % CIs are generally less than 20 %, with the cation and anion coefficients consistently showing similar unceratinties. An overestimation of the diffusivity in $[C_4C_1im][TFSI]$ was also seen by Goloviznina et al.[148] in the benchmark parameterization of CL&Pol.

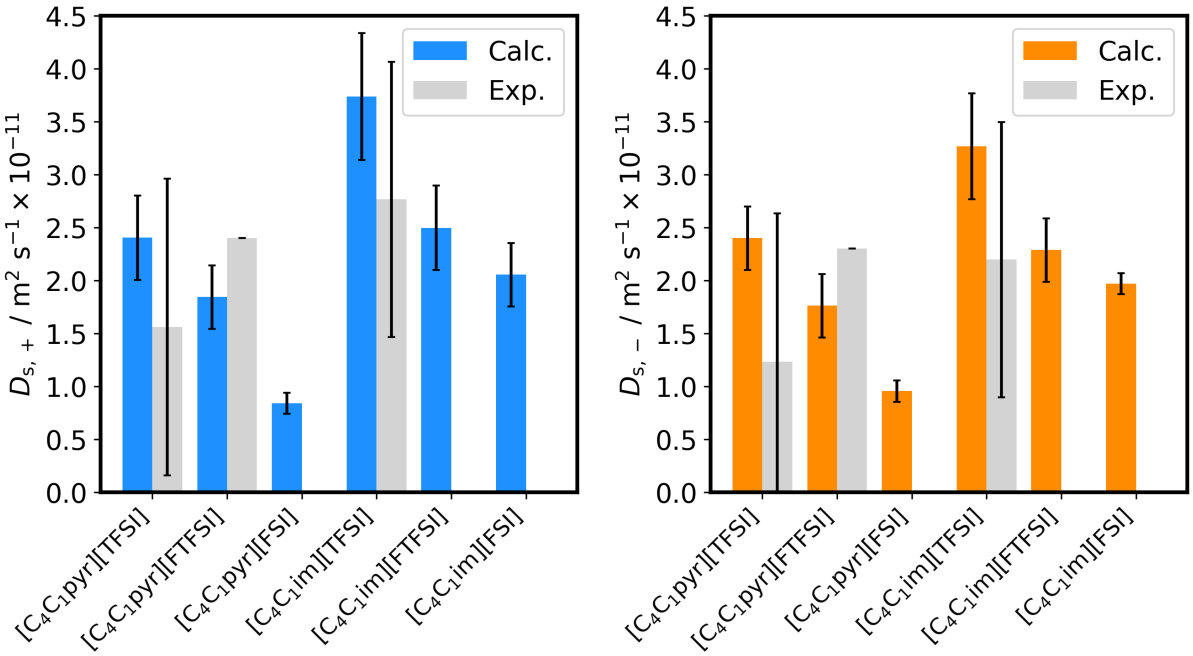


Fig. 4.3. Calculated and available experimental[274,275,280] self-diffusion coefficients of the cation (left) and anion (right). The error bars show the 95 % CI for the mean when calculating self-diffusivity using the L1 simulations.

As seen in Figure 4.3, the trends in $D^{\infty}_{s,+}$ and $D^{\infty}_{s,-}$ are similar, highlighting the direct influence of the anion and the general influence of the cation on diffusivity—ILs containing $[C_4C_1im]^+$ exhibit higher diffusion coefficients than those containing $[C_4C_1pyr]^+$. These cations have similar molar masses (142.3 g/mol and 139.2 g/mol) and are structurally similar in that both consist of a 5-membered cyclic core with methyl and butyl side chains;

however, the imidazolium core is aromatic and planar and, as such, $[C_4C_1im]^+$ exhibits a "slicing" motion due to the flat, board-like structure, leading to faster diffusion than the $[C_4C_1pyr]^+$ cation.[288–290] Due to the limited availability of measured values at 298.15 K, Table 4.4 shows experimental self-diffusion coefficients for the full series of $[C_4C_1pyr]^+$-based ILs, but at at a higher temperature of 305 K,[291] and the MD-calculated values for all four ILs based on $[TFSI]^-$ and $[FSI]^-$ using two polarisable FFs that are comparable to CL&Pol: APPLE&P (at 298 K) developed by Borodin[248] and the physically-motivated SAPT-based FF of McDaniel (at 300 K).[169,170] The $[C_4C_1pyr][TFSI]$ self-diffusion coefficients reported by Borodin are closer to the measured values than what was found using the CL&Pol FF in this work, although still slightly underestimated. The values for $[C_4C_1im][TFSI]$ reported by Borodin show excellent agreement with the measured values, whereas McDaniel underestimates these (at a slightly higher temperature). Regarding the qualitative trend, the measured data given in Table 4.4 shows an increase in the average self-diffusion as the anion decreases in size from $[TFSI]^-$ to $[FSI]^-$, while the values calculated here show the opposite trend with both cations. Although this trend disagrees with the measured data, it is noted that MD simulations of McDaniel[170] also gave lower self-diffusivity in $[C_4C_1im][FSI]$ compared to $[C_4C_1im][TFSI]$. In general, the values of $D^{\infty}_{s,-}$ are lower than the values of $D^{\infty}_{s,+}$ for all ILs.

Table 4.4: Calculated self-diffusion coefficients (in $10^{-11}m^2/s$) at 298.15 K from this work compared to MD-calculated values from literature using polarizable FFs at 298 K[248] and 300 K,[170] and experimental values measured using PGSE NMR at 305 K[291]

IL	This work (298.15 K)		MD[248] (298 K)		MD[170] (300 K)		PGSE NMR (305 K)	
	$D_{s,+}$	$D_{s,-}$	$D_{s,+}$	$D_{s,-}$	$D_{s,+}$	$D_{s,-}$	$D_{s,+}$	$D_{s,-}$
$[C_4C_1pyr][TFSI]$	2.4 ± 0.1	2.4 ± 0.2	1.2	1.0			2.5[291]	[291]
$[C_4C_1pyr][FTFSI]$	1.8 ± 0.2	1.8 ± 0.2					3.1[a]	3.0[a]
$[C_4C_1pyr][FSI]$	0.8 ± 0.2	1.0 ± 0.2					3.3[291]	4.0[291]
$[C_4C_1im][TFSI]$	3.7 ± 0.5	3.3 ± 0.5	2.9	2.0	2.0[b]	1.7[b]		
$[C_4C_1im][FTFSI]$	2.5 ± 0.3	2.3 ± 0.4						
$[C_4C_1im][FSI]$	2.1 ± 0.2	2.0 ± 0.1			1.6[b]	1.3[b]		

[a] Estimated at 305 K using a VFT fit based on data extracted from Fig. 7 in Ref. 85.
[b] Not corrected for finite-size effects due to the estimated statistical uncertainty being of similar order.

4.5.3 Viscosity

Viscosity is a collective property and as such, cannot be averaged over the number of molecules in the system, as with self-diffusion coefficients. Therefore, obtaining precise self-diffusion coefficients is much easier than precise viscosity estimates. It is thus recommended to use several shorter replicate trajectories than fewer longer trajectories to obtain viscosities.[180] As such, the L2 trajectories are used for the calculation of viscosity in this work. As shown in Figure 4.4, the although the calculated viscosities are underestimated

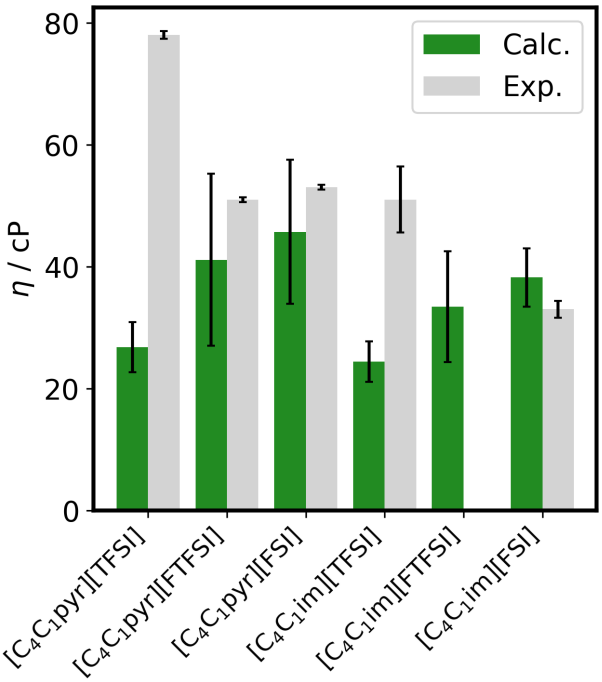


Fig. 4.4. Calculated and available experimental[95,273,276,281,282] shear viscosities. The error bars show the 95 % CI for the mean using the L2 simulations.

(except for $[C_4C_1im][FSI]$). For both the $[C_4C_1pyr]^+$ and $[C_4C_1im]^+$-based ILs, an inversion in the experimental trend as the anion increases in size from $[TFSI]^-$ to $[FSI]^-$ is seen. The overestimation of the diffusivity in both $[TFSI]^-$ ILs (vide supra) hints that the MD simulations should also give a fluidity that is too high, which is indeed the case and an underestimation of the viscosity of ∼60 % is seen for both ILs here. Additionally, the viscosities calculated in the benchmark parameterization of CL&Pol for $[C_4C_1pyr][TFSI]$ and

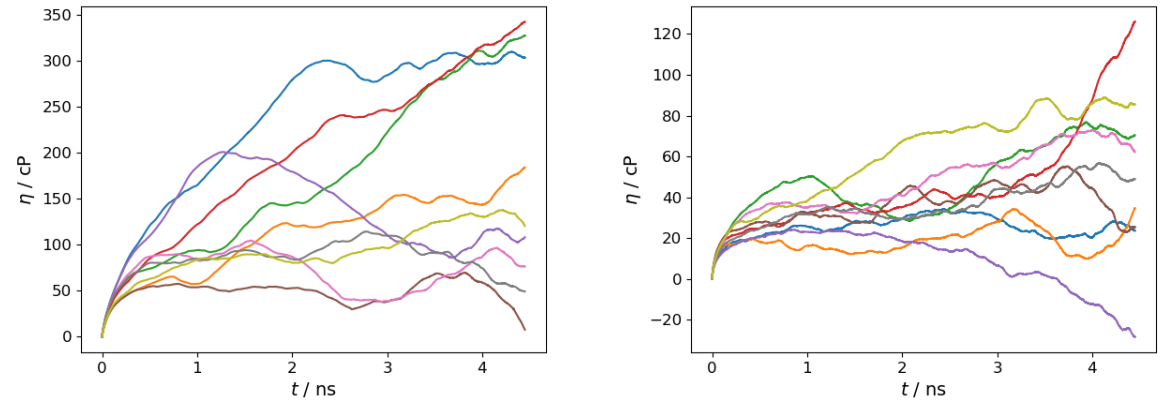
$[C_4C_1im][TFSI]$, although at higher temperatures of 323 K and 343 K, respectively, were also too low.

4.5.4 Validation of viscosity calculation method

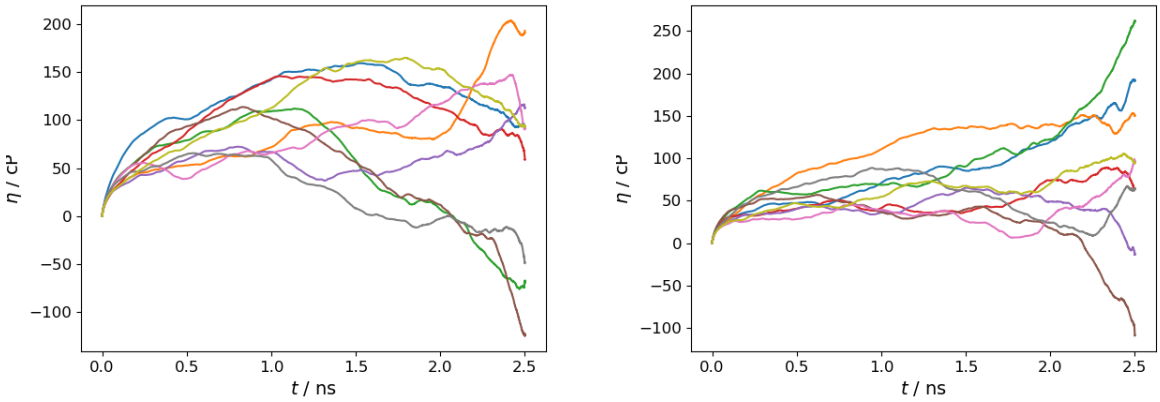
The difficulties associated with the calculation of viscosity, η, were highlighted in Section 3.2.3. As such, although the methodology presented in Section 3.2.3 is used in this and the following chapters, a thorough assessment of different approaches was tested to provide validation of the quality of calculated values of η.

PyLAT,[253] used for the calculation of ion-pair lifetimes (see Section 3.2.6), has the capability of calculating most transport properties explored in this chapter; however, due to the "blackbox" nature of the PyLAT tool, the transport properties were calculated using in-house tools for finer control of the fitting criteria. Nonetheless, using PyLAT for the calculation of η offered a comparison to the calculation methodology implemented here. PyLAT also requires multiple trajectories for the calculation of η, but bootstrapping is used, and the viscosity is reported as the average over all bootstrapping iterations. Ten bootstrapping iterations are typically performed, each sampling five trajectories, to which a double exponential function is fit to the average running integral. The least squares regression is weighted by the standard deviation at each point, such that only points where the standard deviation is less than 40 % of the running integral are included. The weighting function is too strict for the systems used here due to the large noise in the running integrals (see Figure 4.5) and as such, the criteria do not allow for the running integrals to be fit to the double exponential function. Therefore, the weighting function is removed to allow for a reasonable number of points to be included in the fit. The average viscosity obtained for each IL with bootstrapping, as well as the standard deviation over all iterations, is given in Table 4.5. The MAPE from experimental viscosities using this approach is 113 %.

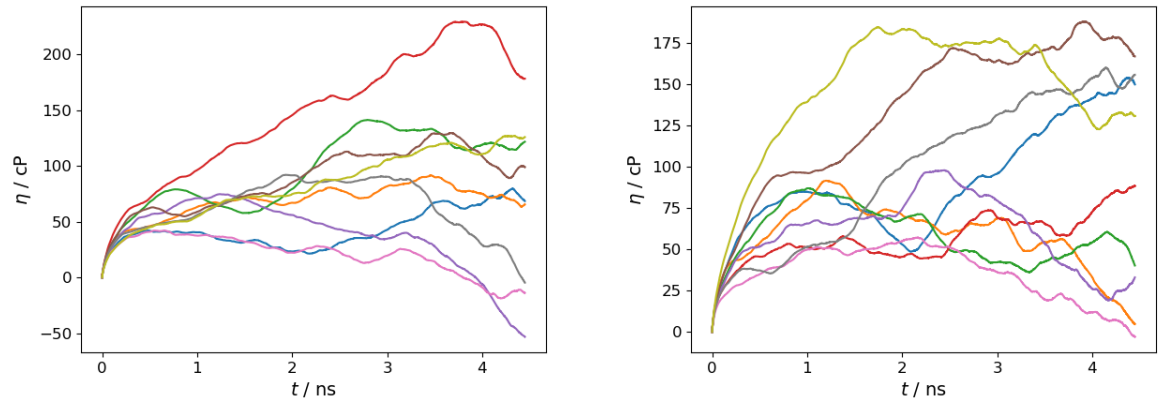
A second approach tested was to determine the average running integral over all trajectories of the L2 simulations for each IL. The average running integrals shown in Figure



(a) $[C_4C_1pyr][TFSI]$ **(b)** $[C_4C_1im][TFSI]$



(c) $[C_4C_1pyr][FTFSI]$ **(d)** $[C_4C_1im][FTFSI]$



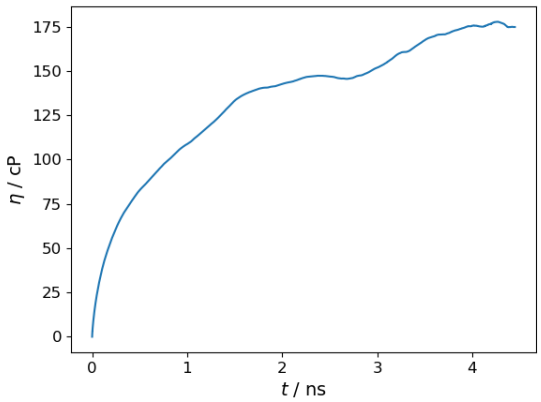
(e) $[C_4C_1pyr][FSI]$ **(f)** $[C_4C_1im][FSI]$

Fig. 4.5. Raw running integrals for all trajectories used in the L2 simulations for each IL used in this work.

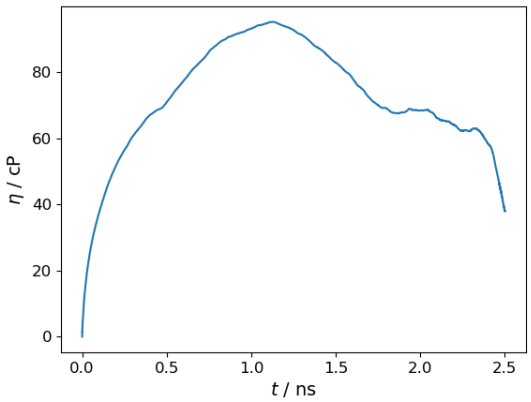
Table 4.5: Experimental and average viscosity (η) obtained for each IL using each method tested: (A) using PyLAT, (B) fitting the average integral to a double exponential function, (C) fitting the average integral with the hybrid method and (D) the Hybrid method as detailed in Section 3.2.3. All values are given in units of cP and confidence intervals, where applicable, are given in parentheses.

IL	(A)	(B)	(C)	(D)	Exp.
$[C_4C_1pyr][TFSI]$	92 (21)	72	32	27 (16)	78
$[C_4C_1pyr][FTFSI]$	120 (25)	70	31	41 (7)	51
$[C_4C_1pyr][FSI]$	167 (12)	130	44	46 (9)	53
$[C_4C_1im][TFSI]$	43 (17)	33	25	24 (5)	51
$[C_4C_1im][FTFSI]$	73 (7)	50	24	33 (6)	n/a
$[C_4C_1im][FSI]$	93 (36)	69	58	38 (14)	33
MAPE	113 %	67 %	48 %	33 %	

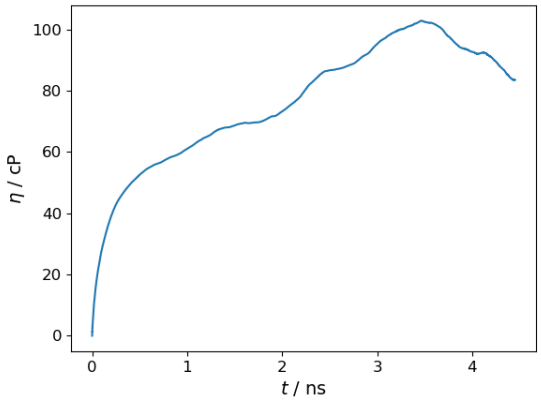
4.6 were then fit to the double exponential function, rather than the individual trajectories and the calculated viscosities are also presented in Table 4.5. The MAPE from experimental viscosities using this approach is 67 %. A hybrid methodology was also applied to the average running integral of each L2 trajectory, such that the fitting procedure detailed in Section 3.2.3 was applied to the average integral, rather than the integrals of the individual trajectories, and the η values determined, for which the calculated values are presented in Table 4.5, for which the MAPE is 48 %. The values obtained from the average running integral are comparable to those taken from the hybrid approach, with only $[C_4C_1im][FSI]$ having a significantly different value; however, the approach of using the average integral has a downfall, as with fitting the double exponential to the average integral, that no measure CI can be quantified. Therefore, the hybrid approach used to calculate the values presented in Table 4.3 as it produces the lowest relative deviation from experiment (MAPE = 33 %), as well as having a noticeably lower standard deviation compared to that of the bootstrapping approach in PyLAT. Table 4.5 indicates that the hybrid approach most accurately represents the experimental trends observed. Lastly, it is worth noting that, as seen in Table 4.5, the PyLAT bootstrapping approach and the fit of the average integral tend to overpredict viscosities, whereas the hybrid approach tends to underpredict viscosities.



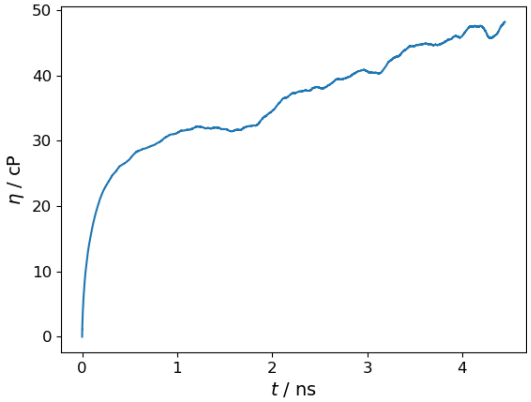
(a) $[C_4C_1pyr][TFSI]$



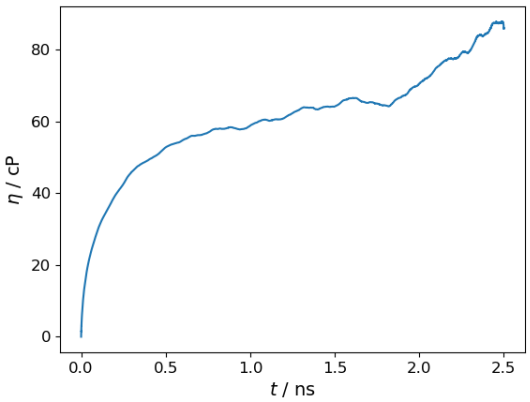
(b) $[C_4C_1im][TFSI]$



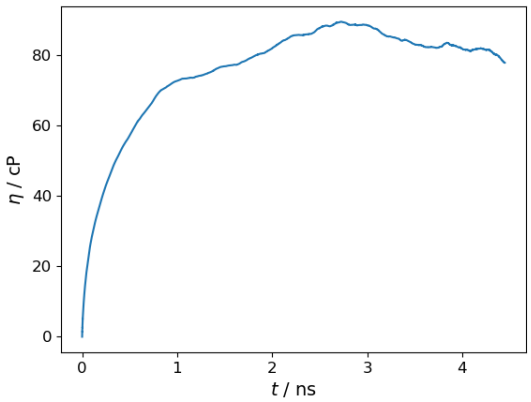
(c) $[C_4C_1pyr][FTFSI]$



(d) $[C_4C_1im][FTFSI]$



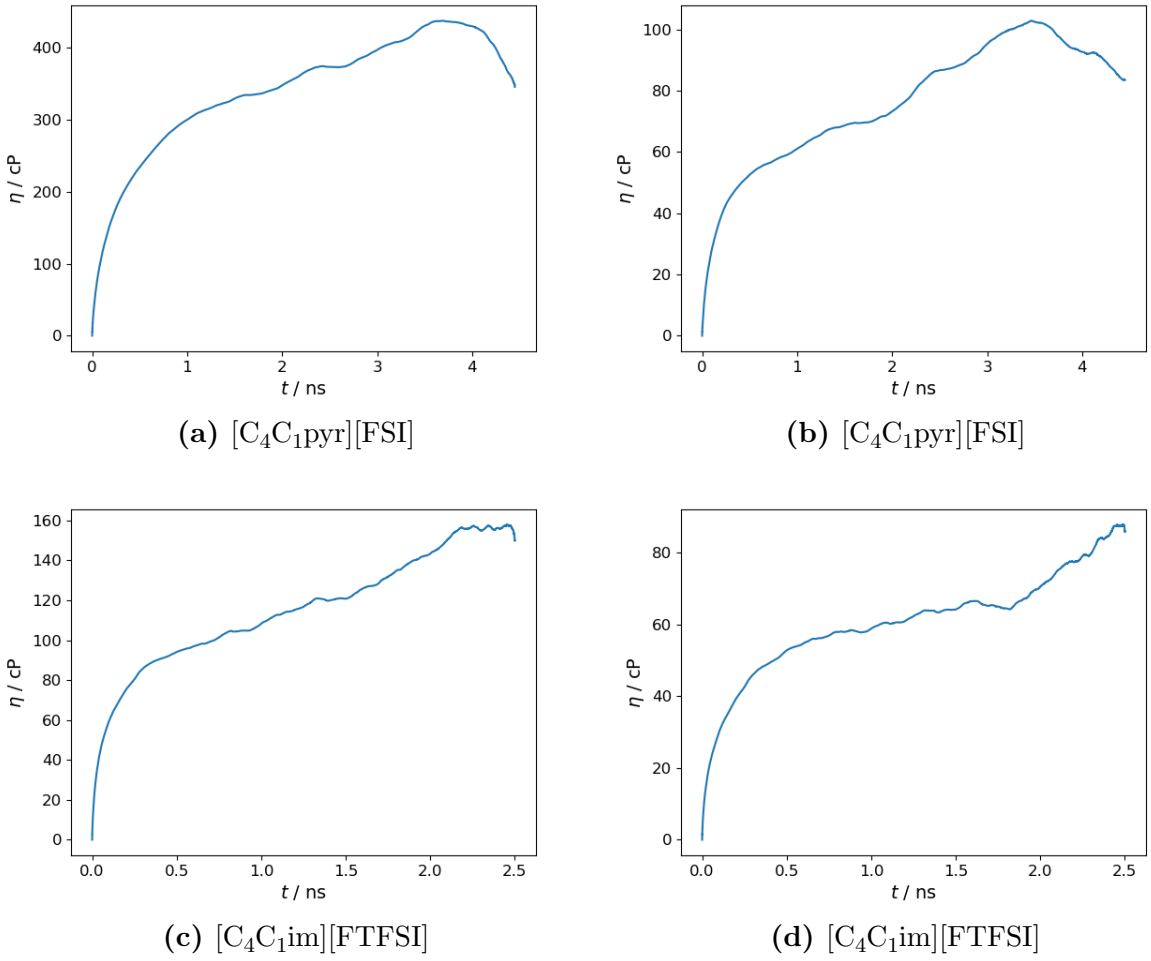
(e) $[C_4C_1pyr][FSI]$



(f) $[C_4C_1im][FSI]$

Fig. 4.6. Average running integrals over all trajectories used in the L2 simulations for each IL used in this work.

An additional 30 trajectories were produced for the $[C_4C_1pyr][FSI]$ and $[C_4C_1im][FTFSI]$ ILs to assess whether the inclusion of more trajectories would reduce the noise of the average integral. While the values do differ, the shape of the average integral does not improve significantly, and the issues present when using 10 trajectories persist (See Figure 4.7). As can be seen in Table 4.6, the use of the additional trajectories



(a) $[C_4C_1pyr][FSI]$ **(b)** $[C_4C_1pyr][FSI]$

(c) $[C_4C_1im][FTFSI]$ **(d)** $[C_4C_1im][FTFSI]$

Fig. 4.7. Average running integral using 40 trajectories (left) and 10 trajectories (right) for the $[C_4C_1pyr][FSI]$ and $[C_4C_1im][FTFSI]$ ILs.

does not improve the results produced using 10 trajectories. Consequently, using 10 trajectories is sufficient based on the results produced and the large additional computations that are required for the recommended 40 trajectories. Therefore, using the hybrid ap-

proach outlined in Section 3.2.3 is more suitable for systems, such as those analysed here, where the running integrals produced do not converge (as typically expected).

Table 4.6: Average viscosity determined for the $[C_4C_1pyr][FSI]$ and $[C_4C_1im][FTFSI]$ ILs using N=10 and N=40 independent trajectories. 95 % CIs are indicated in parentheses.

N	$[C_4C_1pyr][FSI]$	$[C_4C_1im][FTFSI]$
10	42 (8)	32 (7)
40	50 (13)	31 (7)

Based on the assessment carried out here, the Hybrid approach, as presented in Section 3.2.3, is the most statistically sound approach for the calculation of η and the choice of 10 trajectories is sufficient as the slightly improved convergence of the integrals with four times the number of trajectories is not significant to warrant the addition computations.

4.5.5 Conductivity

The Einstein-Helfand (EH) relation is used to calculate the ionic conductivity, σ_{EH}, as outlined in Section 3.2.4. Conductivity, like viscosity, is a collective property that depends on the combined displacement of all ions at every time interval in the trajectory; consequently, unless one utilises very long simulation times or preferably a large number of independent, shorter trajectories,[292] the uncertainty in the calculation will typically be large.[169] The use of the L1 simulations leads to an average uncertainty across all ILs of 37 %, with the lowest being 22 % for $[C_4C_1im][FTFSI]$ and the highest being 48 % for $[C_4C_1im][FSI]$. This is similar to the uncertainty of 30–40 % reported by McDaniel et al.[169] for the conductivity of $[C_4C_1im][BF_4]$ calculated from 10 independent 50 ns trajectories using a polarisable FF.[169]

Figure 4.8 illustrates that for ILs with a common cation, the experimental ionic conductivities increase with a decrease in the anion size. Calculated conductivities show the opposite trend for both cations considered. This trend inversion was also observed for the calculated viscosities and self-diffusion coefficients. Although conflicting with the measured trend, it is satisfying that the calculated fluidity is consistently reflected in all three inde-

pendently calculated transport properties. A further indication of the ability to reproduce trends is that all $[C_4C_1pyr]^+$-based ILs have a calculated conductivity that is lower than the corresponding $[C_4C_1im]^+$-based IL with the same anion, which is consistent with the trend in the measured values. In this context, it is interesting to note that the SAPT-based polarisable FF of McDaniel[170] also gave a conductivity for $[C_4C_1im][FSI]$ that is lower (0.3 S/m, but at 300 K) than the corresponding measured value of 0.85 S/m.[282] Compared to the available experimental measurements, the MAPE for the series of $[C_4C_1pyr]^+$ and $[C_4C_1im]^+$-based ILs are 40 % and 60 %, respectively.

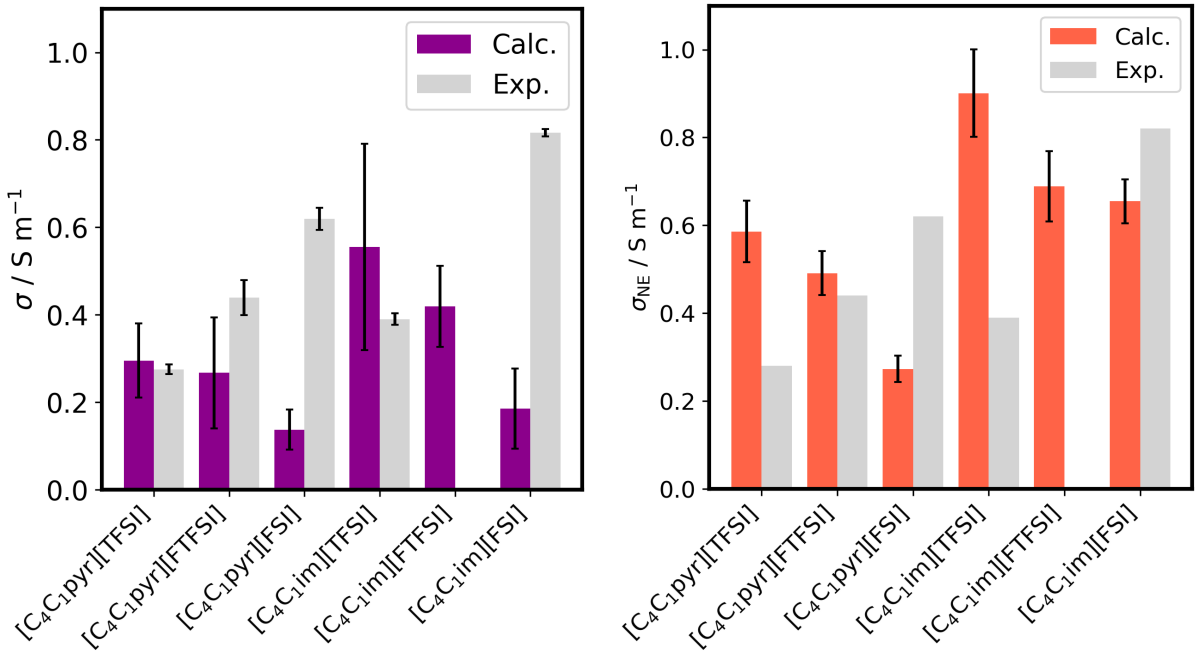


Fig. 4.8. Available experimental[75,95,278,280,282] and calculated conductivities, using Equation 3.11 (left) and Equation 3.10 (right). The error bars show the 95 % CI for the mean, calculated for the true conductivity (left) and the self-diffusion coefficients (right) using the L1 simulations.

If the correlated motion of different ions of the same (cation-cation or anion-anion) and opposite (cation-anion) charge is ignored, the cross-terms in the collective mean-square displacement fall away and the ideal, or Nernst-Einstein (NE), ionic conductivity, σ_{NE}, can be calculated using Equation 3.10. Figure 4.8 (right) shows the results of using Equation 3.10, with the calculated self-diffusion coefficients corrected for finite-size effects. It is

pointed out here that while self-diffusion coefficients show a strong dependency on system size,[243] the conductivity calculated using the collective displacement does not,[292,293] thus making this a more reliable quantity for smaller simulation boxes. However, the comparably lower uncertainties of self-diffusion coefficients compared to that of σ_{EH}, indicate the appeal of calculating σ_{NE}. The uncertainties for σ_{NE} were estimated by propagating the uncertainties in $D^{\infty}_{\text{s},+}$ and $D^{\infty}_{\text{s},+}$ and are consequently much lower than when using Equation 3.11; decreasing from 40 % to 10 % on average, but with the caveat of ignoring the effects of ion coupling. A comparison of the two calculations shows that the trend is qualitatively maintained between both methods.The relationship between the true (or real) and ideal conductivity, or strictly speaking the true (Λ) and ideal *molar* conductivity,[294,295] can be quantified using a deviation parameter, Δ, which accounts for the ion correlations that are absent in the simple Nernst-Einstein equation,[295,296]

$$\begin{aligned}\Lambda &= \frac{F^2}{RT}\left(z_+^2 D_{\text{s},+} + z_-^2 D_{\text{s},-}\right)(1-\Delta) \\ &= \Lambda_{\text{NE}}\,(1-\Delta)\,,\end{aligned} \tag{4.1}$$

where R and F are the gas and Faraday constants, respectively. The molar conductivity is given by

$$\Lambda = \frac{\sigma}{c} = \frac{\sigma M}{\rho}, \tag{4.2}$$

where c is concentration, M is the molar mass and ρ is the density. The deviation parameter can also be expressed in terms of the "ionicity", Y, which is the ratio of the true molar conductivity and that calculated from ionic self-diffusion coefficients,[280]

$$\Delta = 1 - \text{Y} = 1 - \frac{\Lambda_{\text{EH}}}{\Lambda_{\text{NE}}} = 1 - \frac{\sigma_{\text{EH}}}{\sigma_{\text{NE}}}. \tag{4.3}$$

Ionicities that are less than one have been interpreted as the result of the formation of ion pairs or neutral aggregates;[44,297,298] however, this interpretation is controversial,[299] and other mechanisms include correlations between ionic velocities[299,300] and/or charge

transfer.[301–303] For our purposes, ionicity captures the relationship between conductivity and the average ionic self-diffusion coefficients and thus provides a means to validate the internal consistency of these calculated properties. The calculated values of Y are shown in Table 4.7, along with the corresponding molar ionic conductivities calculated using the simulation densities shown in Table 4.3.

Table 4.7: Calculated true (Λ_{EH}) and ideal (Λ_{NE}) molar conductivities and corresponding ionicities (Y) obtained for each IL. All molar conductivities are reported in units of $\mathrm{S\,cm^2/mol}$

IL	Λ_{EH}	Λ_{NE}	Y
$[C_4C_1pyr][TFSI]$	0.9 ± 0.3	1.8 ± 0.2	0.50
$[C_4C_1pyr][FTFSI]$	0.7 ± 0.4	1.4 ± 0.1	0.54
$[C_4C_1pyr][FSI]$	0.3 ± 0.1	0.7 ± 0.1	0.50
$[C_4C_1im][TFSI]$	1.6 ± 0.7	2.6 ± 0.3	0.62
$[C_4C_1im][FTFSI]$	1.1 ± 0.2	1.8 ± 0.2	0.61
$[C_4C_1im][FSI]$	0.4 ± 0.2	1.5 ± 0.1	0.28

Tokuda et al.[44] have reported ionicities in the range 0.5–0.8 for typical aprotic ILs at 303 K, including $[C_4C_1im][TFSI]$ and $[C_4C_1pyr][TFSI]$ with $\mathrm{Y} = 0.61$ and 0.70, respectively. Therefore, the calculated ionicity of 0.62 for $[C_4C_1im][TFSI]$ compares very well with the expected value, but the calculated ionicity of 0.50 for $[C_4C_1pyr][TFSI]$ is too low. Both self-diffusivity and conductivity are underestimated for the latter IL, but one must consider that the calculated values of σ_{NE} $[C_4C_1pyr][FTFSI]$ have the largest uncertainty of $\sim$50 %, which could be also contributing to the break in expected trend. Similarly, the calculated comparably low ionicity of $[C_4C_1im][FSI]$ is also likely due to the again $\sim$50 % in σ_{NE} for this IL.

The empirical Walden rule relates the molar ionic conductivity and viscosity of ILs such that[304,305]

$$\Lambda\eta = k, \tag{4.4}$$

where k is a temperature-dependent constant. Based on this rule, Walden plots, which are $\log_{10}$-$\log_{10}$ plots of molar conductivity (in units of $\mathrm{S\,cm^2/mol}$) and fluidity (η^{-1}, with viscosity in units of 100 mPa s, or poise), have been used to categorize the conductivity be-

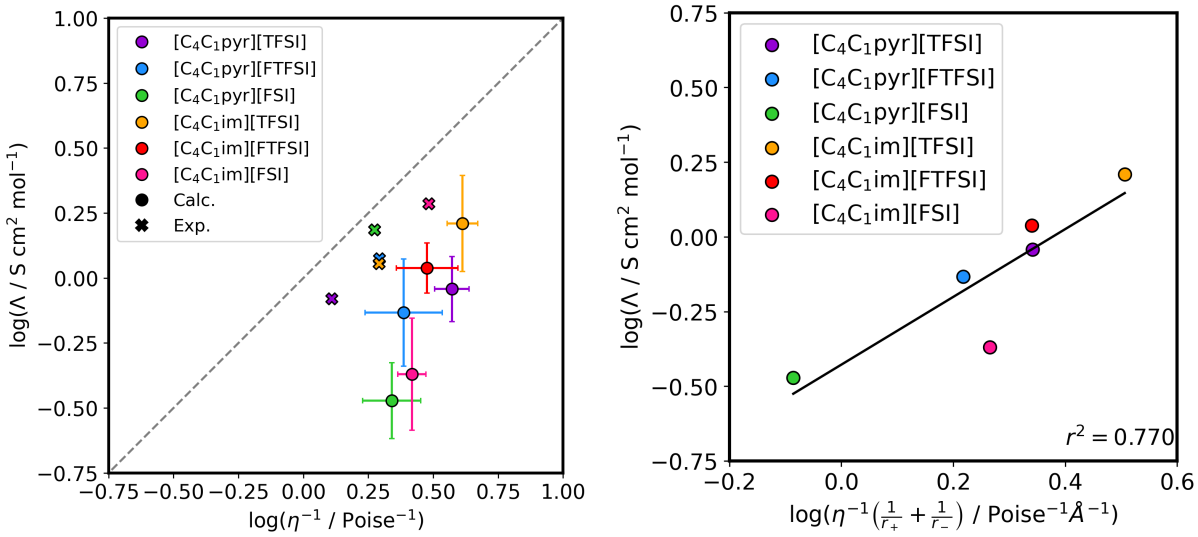


Fig. 4.9. The left panel shows Walden plots combining experimental (crosses) and computed (circles) molar conductivity and viscosity data, relative to the arbitrary reference line (dashed grey) of unit slope.[308] Propagated uncertainties in the calculated data are shown. The right panel shows the adjusted Walden plot[297] shown in Equation 4.6 that takes into account the calculated Stokes or hydrodynamic ionic radii in each liquid. r^2 is the coefficient of determination.

haviour of ILs concerning an arbitrary reference line with a slope of 1, passing through the origin.[305–307] This line has been likened to a Walden plot for dilute aqueous KCl that represents an ideal electrolyte solution[308] even though it has been pointed out that the data for infinitely diluted KCl has a slope of 0.87.[309] Figure 4.9 (left) compares the Walden plots using measured and calculated data shown in Table 4.3. Simulation agrees with measurement that classifies these systems as "good" ILs with Walden products below the reference line.[305] Linear relationships with similar slopes are apparent in both the measured and calculated conductivity-viscosity data but the intercepts differ for [C_4C_1pyr][TFSI] and [C_4C_1im][FSI], which lie well below the trend formed by the remaining four ILs. It has already been pointed out that the calculated ionicities (see Table 4.7) of these two ILs are noticeably different from the others (and what is expected based on measurement), and therefore this discrepancy also influences the Walden product for these ILs. However, the propagated uncertainties (shown in Figure 4.9) are large, and caution should be

taken against using this as the only evidence to conclude that the conductivity behaviour of $[C_4C_1pyr][TFSI]$ and $[C_4C_1im][FSI]$ is different from that of the other ILs.

The correlation between conductivity and viscosity, expressed as the Walden rule in Equation (4.4), may be further refined by taking into account differences in the effective size of migrating ions. The role of ion size is apparent in the Stokes-Einstein (SE) equation,

$$D_{\mathrm{s}} = \frac{k_{\mathrm{B}}T}{6\pi\eta r_{\mathrm{eff}}}, \tag{4.5}$$

where r_{eff} is the Stokes (or effective hydrodynamic) radius of a spherical solute diffusing in a continuum with viscosity η, assuming "stick" boundary conditions.[310] The SE equation can be substituted into the NE equation (shown in Equation 4.1, but lacking the deviation term) to yield the following form of the Walden rule:[252,297]

$$\Lambda\eta = k\left(\frac{1}{r_+} + \frac{1}{r_-}\right), \tag{4.6}$$

where r_+ and r_- are the cation and anion Stokes radii, respectively. Consequently, an adjusted Walden plot (a $\log_{10}$-$\log_{10}$ plot of Λ and $\eta^{-1}\left(\frac{1}{r_+} + \frac{1}{r_-}\right)$, with radii in Å) is more appropriate when comparing ILs with ions of varying size,[297] although still neglecting ion-ion coupling in the non-ideal liquid.

The calculated ionic self-diffusion coefficients, corrected for finite size effects and the calculated average viscosities, shown in Table 4.3, were used for the calculation of MD-based Stokes radii. The experimental Stokes radii were calculated using the experimental data shown in Table 4.3. It should be pointed out that the Stokes radii are determined from self-diffusion coefficients and viscosity. Therefore, these cannot be interpreted as the average size of the molecules. This is clear as even when experimental viscosities and diffusion coefficients are used, the Stokes radii are still substantially less than the actual average sizes. Thus, the Stokes radii should not be probed in this way, as no true physical meaning can be attached to it, and it should be used as a comparative measure of

sizes, rather than absolute values. For [C_4C_1pyr][TFSI], the average of the reported experimental ionic self-diffusion coefficients was used. The Stokes radii of the cation and anion in each IL were calculated from their ionic experimental self-diffusion coefficients and liquid viscosity using Equation 4.5 and the resulting values are reported in Table 4.8, where it is worth noting that the MD-calculated Stokes radii of these ions are greater than expected, compared to the other ILs and the experimental Stokes radii. Figure 4.9 (right)

Table 4.8: Calculated (MD) and experimental Stokes radii of the cations and anions in the six ionic liquids, at 298.15 K. All values are given in units of Å.

IL	r_+ (MD)	r_- (MD)	r_+ (exp.)	r_- (exp.)
[C_4C_1pyr][TFSI]	4.15	4.23	1.75	2.33
[C_4C_1pyr][FTFSI]	2.95	3.12		
[C_4C_1pyr][FSI]	3.16	2.89		
[C_4C_1im][TFSI]	2.52	2.80	1.53	1.95
[C_4C_1im][FTFSI]	2.94	3.24		
[C_4C_1im][FSI]	2.64	2.76		

shows the adjusted Walden plot, which gives only a moderate correlation with $r^2 = 0.77$. However, the calculated radii of the cation and anion in [C_4C_1im][FSI] ($r_+ = 2.62$ Å and $r_- = 2.86$ Å) are much smaller than in the analogous [C_4C_1pyr][FSI] IL ($r_+ = 5.71$ Å and $r_- = 5.01$ Å), which is not reflected a similar extent when comparing the other ILs. This suggests that the calculated self-diffusion coefficients are higher than required to be consistent with the calculated viscosity, which is further verified by an increase to $r^2 = 0.99$ if this IL is excluded from the correlation.

4.5.6 Enthalpy of Vaporisation

Enthalpies of vaporisation were computed using Equation 3.12, where the L1 simulations are used to represent the liquid phase. The statistical uncertainties associated with the calculated enthalpies of vaporisation are small, with the highest uncertainty of 2.5 % for [C_4C_1pyr][TFSI], which is also the IL showing the largest error, underestimating the measured value by ∼14 %. For both cations, the experimental vaporisation enthalpies are

greater with the smaller $[FSI]^-$ anion than with $[TFSI]^-$, and the measured $\Delta_{vap}H$ values are consistently greater for the ILs based on $[C_4C_1pyr]^+$ than for the ILs based on $[C_4C_1im]^+$. The calculated values follow the former trend for both cations as a function of anion size. Whereas the measured values differ by less than 4 kJ/mol within each cation family, the differences in the calculated values are much more exaggerated. Interestingly, for both cations, the $[FTFSI]^-$-based ILs, for which measured values are not available, have $\Delta_{vap}H$ values intermediate between the $[TFSI]^-$ and $[FSI]^-$-based ILs. The overall MAPE with respect to the available experimental values is 7 %.

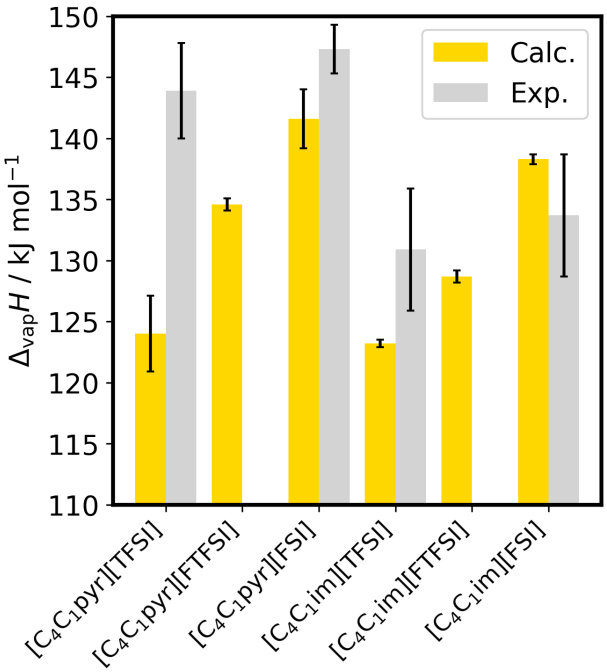


Fig. 4.10. Calculated and available experimental[277] enthalpies of vaporisation. The error bars show the 95 % CI for the mean when calculating the enthalpy of vaporisation using the L1 simulations.

Borodin[311] investigated the relationship between vaporisation enthalpy, ion transport, and molar volume for ILs, with properties calculated using the APPLE&P polarisable FF,[248] and found a significant correlation between the average ionic self-diffusion coefficient ($D_{s,avg} = (D_{s,+} + D_{s,-})/2$) and $\Delta_{vap}H$, which improved further by including the molecular volume of the ion pair (V_{pair}) and a fraction of the interaction energy.[311] For the

ILs simulated here, a simple linear correlation between $D_{s,avg}$ and $\Delta_{vap}H$ has a coefficient of determination of $r^2 = 0.73$, which improves to $r^2 = 0.86$ if the $[C_4C_1im][FSI]$ ILs is excluded. Furthermore, including V_{pair}, calculated from the molar mass of the ion pair and the liquid density, using a relationship similar to that used by Borodin[311] as $-\log(D_{s,avg} \times V_{pair})$ vs $\Delta_{vap}H$, improves the coefficient of determination to $r^2 = 0.86$, as shown in Figure 4.11. In addition, this correlation provides further evidence that $\Delta_{vap}H$ for $[C_4C_1im][FSI]$ is underestimated in our simulations, in that a higher calculated value would have led to a higher coefficient of determination (viz., $r^2 = 0.93$ if $\Delta_{vap}H([C_4C_1im][FSI])$ is removed).

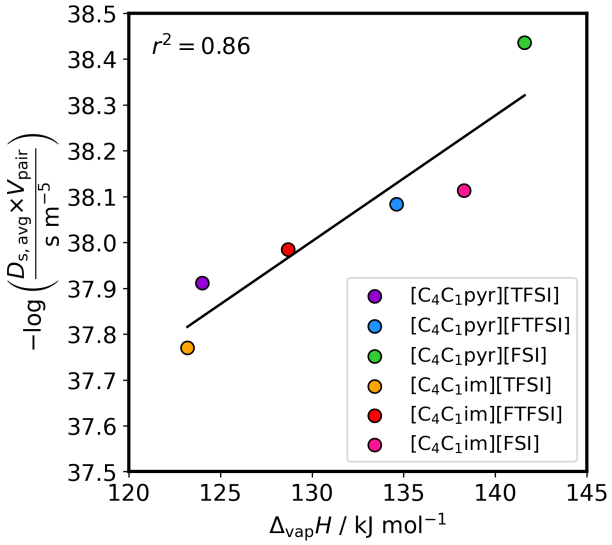


Fig. 4.11. Correlation between the average ionic self-diffusion coefficient multiplied by the volume of the ion pair ($D_{s,avg} \times V_{pair}$), and the enthalpy of vaporisation ($\Delta_{vap}H$).[311] r^2 is the coefficient of determination.

4.5.7 Ion-pair lifetimes

Ion-pair lifetimes are calculated as described in Section 3.2.6. The resulting values of the calculated ion-pair lifetimes, τ_{IP}, are shown in Table 4.9. The trend in relaxation times with respect to the nature of the anion is dissimilar within the two cation families; however, τ_{IP} is consistently greater for the $[C_4C_1pyr]^+$-based ILs than for the $[C_4C_1im]^+$-based ILs, sug-

gesting a looser association in the liquid for the latter family of ILs. The longer IP lifetimes of the $[C_4C_1pyr]^+$-based ILs, despite its weaker interaction strength, can be attributed to the higher viscosities of these ILs. This affects ion mobility, leading to slower ionic rearrangements and extended association times between ions. This suggests that the observed IP lifetimes are governed more by the liquid's bulk dynamic properties, such as viscosity and restricted motion, rather than direct ion-pair interaction strength. Interestingly, anions typically show a weaker interaction strength in the ion pair with the pyrrolidinium cation than with the imidazolium cation,[312] which might be interpreted as resulting in a shorter IP lifetime in the liquid; yet, the opposite is seen here. Using data generated from non-polarisable MD simulations of a wide variety of cation and anion classes, Zhang and Maginn[252] found a direct relationship between inverse IP or IC lifetimes and transport properties, in particular self-diffusivities and ionic conductivities, although based on the NE equation. Figure 4.12 (left) shows the correlation between the average ionic self-diffusion coefficient, $D_{\text{s,avg}}$, and the inverse of the IP lifetime, τ_{IP}^{-1}, for the six ILs studied in this work, which gives a remarkably high coefficient of determination of $r^2 = 0.998$! The same group conducted a follow-up study investigating the relationship between experimental viscosities and IP lifetimes.[253] Figure 4.12 (right) shows this correlation for the calculated properties, which produces a weaker, but still clear correlation with $r^2 = 0.78$.

Table 4.9: Ion pair lifetimes (τ_{IP}) with uncertainties expressed as the 95 % CI for the mean using the L1 simulations.

IL	$\tau_{\text{IP}}/\text{ns}$
$[C_4C_1pyr][TFSI]$	1.2 ± 0.6
$[C_4C_1pyr][FTFSI]$	1.4 ± 0.1
$[C_4C_1pyr][FSI]$	2.0 ± 0.2
$[C_4C_1im][TFSI]$	0.9 ± 0.1
$[C_4C_1im][FTFSI]$	1.2 ± 0.1
$[C_4C_1im][FSI]$	1.3 ± 0.1

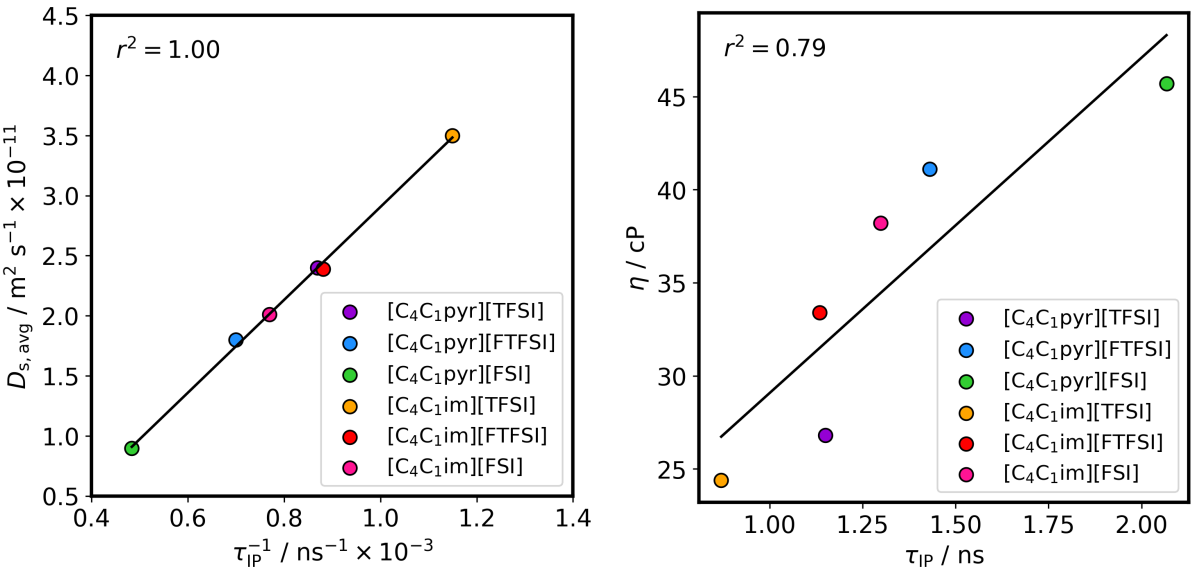


Fig. 4.12. The left panel shows the correlation between the average ionic self-diffusion coefficient ($D_{\mathrm{s,avg}}$) and the inverse of the IP lifetime (τ_{IP}^{-1}). The right panel shows the correlation between viscosity (η) and IP lifetime (τ_{IP}). r^2 is the coefficient of determination.

4.6 Conclusions

The work carried out in this chapter has confirmed that the CL&Pol polarisable FF is capable of predicting the thermodynamic and transport properties of ILs with reasonable accuracy.[IV] The empirical parameterization protocol developed by Goloviznina et al.[148] was used to extend the FF and obtain parameters for [C$_4$C$_1$im][FSI], [C$_4$C$_1$im][FTFSI], [C$_4$C$_1$pyr][FSI] and [C$_4$C$_1$pyr][FTFSI]. This protocol transfers parameters from the non-polarisable CL&P FF and requires only the geometry of optimised ion pair fragments, fragment dipole moments and polarisabilities to adjust the non-bonded interactions for the effects of polarisation.

The usefulness of any property prediction also depends on the associated uncertainty, and MD is no different. Therefore, all simulation results reported here included an estimate of the statistical uncertainty. This was achieved by determining the properties as an average over five or ten independent simulations and calculating the 95 % confidence

IV The accuracy is reasonable in the sense that it produces properties of the same order as experimental values, whereas previous (mostly non-polarisable) FFs calculate these values with errors of entire orders.

interval for the mean. The uncertainty in thermodynamic properties rivals that of experimental measurement; however, while self-diffusivity and viscosity showed uncertainties less than 25 %, this estimate for ionic conductivities was as high as 50 %. This highlights one of the challenges of using MD for IL transport property prediction which is independent of assessing the quality of the FF parameters.

The trends in the computed self-diffusion coefficients, viscosities, and ionic conductivities is opposite to that of the available experimental trends. One critical factor at play here is the k_{ij} scaling factor. Considering the $[C_4C_1im][TFSI]$ IL: the k_{ij} values used for the $[C_2C_1im]^+\cdots[TFSI]^-$ and $[C_4H_{10}]\cdots[FTFSI]$ interactions were 0.55 and 0.69, respectively, resulting in calculated properties indicating faster dynamics than that of experiment. In contrast, Philippi et al.[53] also simulated $[C_4C_1im][TFSI]$, using k_{ij} values of 0.65 and 0.77 for $[C_2C_1im]^+\cdots[TFSI]^-$ and $[C_4H_{10}]\cdots[FTFSI]$, respectively, observing diffusion behavior that was underpredicted relative to experimental values, indicating a less fluid system. This comparison highlights the sensitivity of the model dynamics to the k_{ij} scaling factor. While this does not establish which scaling factors are inherently correct, it underscores the significant impact of k_{ij} values on simulated properties.

Using the Walden rule combined with the Stokes-Einstein equation, the internal consistency of the calculation procedures used to obtain the self-diffusivity, viscosity, and ionic conductivity was validated. The calculated properties, determined independently, show the expected linear correlation suggested by these relationships. Additionally, this allowed us to interpret the reliability of the predictions and identify potential outliers. Taken together with reporting uncertainties, this is an important analysis that improves the reliability of the prediction methodology.

A remarkable correlation has been shown here between the MD-calculated self-diffusivity and viscosity with the ion pair lifetimes. Although this has been pointed out before,[252,313] including polarisation appears to improve this correlation significantly. Ex-

ploiting this relationship highlights a potential avenue for obtaining transport properties from more easily calculated, statistically sound properties.

As highlighted by Zhang and Maginn, discrete IPs do not exist in a liquid, and one should rather consider the lifetime of an ion cage (IC) as a time-based measure of the cation-anion interaction.[252] The lifetime of an IC expresses the persistence of a static solvation shell of anions around a central cation, or vice versa, and can be calculated similarly to the IP lifetime using an adapted definition of $p(t)$.[251,252] However, τ_{IP} was shown to correlate linearly with IC lifetimes[252] and consequently, while the concept of an IP in the liquid should not be taken too literally, τ_{IP} can be used equally well to gain insight into the strength of ion-ion interaction. Average IP lifetimes were calculated from the NVT trajectories using the PyLAT software package,[253] in which the first picosecond of the NVT trajectory was excluded when fitting to $C(t)$. Eqn 3.16 was used to calculate the relaxation time with the sum expanded to three exponential functions.

Chapter 5

Investigating the impact of polarisation and the performance of the CL&Pol FF for IL-based electrolytes

5.1 Introduction

Chapter 4 validated the ability of the CL&Pol FF to simulate pure ILs; this chapter extends the validation study to IL-based electrolytes. A comprehensive set of experimental data for a set of alkali metal cations in an IL has recently become available: Yamamoto et al.[95] published measured thermodynamic (density) and transport properties (ionic conductivity and shear viscosity) for the IL-electrolytes N-butyl-N-methylpyrrolidinium (fluorosulfonyl)(trifluorosulfonyl)imide ([C_4C_1pyr][FTFSI]) containing [M][FTFSI] (M = Li^+, Na^+, K^+, Rb^+ and Cs^+) salts. The [C_4C_1pyr][FTFSI] IL has been reported to have the desired properties envisaged for next-generation batteries (see Section 1.2.1). Furthermore, the

experimental data are for a comprehensive set of alkali metal-containing electrolytes, thus allowing for a systematic study of the factors considered in this chapter.

Regarding the treatment of interactions involving the alkali metal cations, it is worthwhile noting how the previous applications of CL&Pol FF to IL-based electrolytes (see Section 1.3) approached the scaling of the Lennard-Jones (LJ) well-depth, ε_{ij}. In the work of Goloviznina et al.,[174] CL&Pol was used for the simulation of 1-ethyl-3-methylimidazolium bis(trifluoromethylsolfonly)imide containing a lithium bis(trifluoromethylsolfonly)imide salt ([C_2C_1im][TFSI]/[Li][TFSI]). Here, no scaling of ε_{ij} is applied to interactions involving Li^+. This is based on the low polarisability of Li^+, which can be assumed to give negligible induction, essentially setting $k_{ij} = 1$. Massaro et al.[176] simulated the N-methyl-N-propylpyrrolidinium bis(fluorosulfonyl)imide IL mixed with varying concentrations of sodium bis(fluorosulfonyl)imide salts ([C_3C_1pyr][FSI]/[Na][FSI]), where once again, no scaling of interactions involving the alkali metal was carried out. While the argument of low polarisability is physically reasonable in justifying no scaling of ε_{ij} for interactions involving Li^+ (and possibly Na^+), the same argument cannot be made for the heavier alkali metals with larger polarisabilities. Thus, this chapter aims to investigate the transferability of the CL&Pol approach to previously unsimulated IL-based electrolyte systems. Furthermore, while comparisons involving polarisable and non-polarisable FFs based on (and including) CL&P(ol) have been carried out for simulation of pure ILs,[172] no such study has been carried out specifically for alkali metal cations in ILs.

This chapter uses molecular dynamics (MD) simulations of the [M][FTFSI]/[C_4C_1pyr][FTFSI] (M = Li^+, Na^+, K^+, Rb^+,Cs^+) electrolyte systems, with the alkali metal salts having a mole fractions $x_{\mathrm{M}} = 0.2$, to study 1) the effect of scaling of ε_{ij} for interactions involving metal cations, and 2) the relative performance of the fixed-charge CL&P, scaled fixed-charge sCL&P and polarisable CL&Pol FFs for the calculation of thermodynamic and transport properties of IL-based electrolytes.

5.2 Force field details

The CL&P and sCL&P FFs were implemented as outlined in Section 2.1.5, where the latter scales all atomic charges used in CL&P[133] by a factor of 0.8.[I] For the implementation of CL&Pol, LJ parameters are once again carried over from the CL&P FF (OPLS[185] parameters)and ε_{ij} values are scaled as prescribed by Goloviznina et al.[148] While all parameters were readily available for the $[C_4C_1pyr]^+$ and $[FTFSI]^-$ ions, only the Li^+ and Na^+ LJ parameters and atomic polarisabilities were included in the available parameterization files as of when this work was carried out. The list of LJ parameters and atomic polarisabilities of the alkali-earth metals required for this work were taken from the references used by Goloviznina et al.[148] during the development of CL&Pol: the LJ parameters are those of Jorgensen and Tirado-Rives[185] and atomic polarisabilities, α, were taken from Yu et al.[177] All metal cation non-bonded parameters are given in Table 5.1, showing that the values of σ_{ij} indicate an increase in atom size going down the alkali metal group, as would be expected given that this is a measure of the ion radius. The ε_{ij} values show the opposite trends to that of σ_{ij}, as well as more significant variation, although no physical interpretation can be made from this parameter. The studies of Fennell et al.[314] and Joung and Cheatham[315] highlight the variance in LJ parameters used by different FFs, and the latter showed that different combinations of σ_{ij} and ε_{ij} can reproduce the same hydration free energies. As such, as long as both parameters are taken from the same reference, they should, in theory, reproduce correct results.

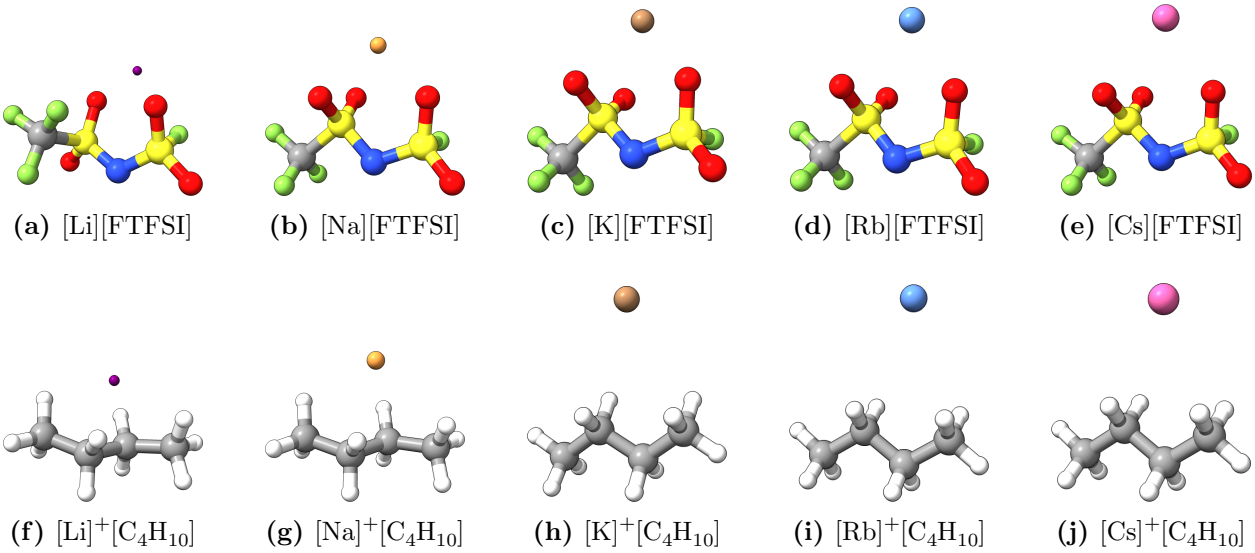
Regarding the scaling factor k_{ij}, the SAPT-based definition is used to compute k_{ij}^{SAPT} values in this chapter. In addition to the k_{ij}^{SAPT} scaling factors presented in Section 4.2 for $[C_1C_1pyr]^+$, $[C_4H_{10}]$ and $[FTFSI]^-$, k_{ij}^{SAPT} values needed to be computed for the pair interactions involving the metal cations and the fragments forming the cation and anion. The k_{ij}^{SAPT} values were computed for the fragments previously defined in Section 4.2

I Goloviznina et al.[148] used 0.8 to compare the performance of the CL&Pol FF against sCL&P during development of the former.

Table 5.1: Lennard-Jones parameters and atomic polarisabilities (α) for the metal cations (M) used in this work for simulation with the CL&Pol FF. The values of σ_{ii} and ε_{ii} are taken from Jorgensen and Tirado-Rives[185] and the α values from Yu et al.[177] The combining rules given in Equations 2.13 and 2.14 to obtain σ_{ij} and ε_{ij}, respectively

M	σ_{ii} / Å	ε_{ii} / kJ/mol	α / Å^3
Li^+	2.1265	0.0765	0.032
Na^+	3.3305	0.0116	0.157
K^+	4.9346	0.0014	0.83
Rb^+	5.6212	0.0007	1.37
Cs^+	6.716	0.0003	2.36

($[C_1C_1pyr]^+$, $[C_4H_{10}]$ and $[FTFSI]^-$) paired with each of the metal cations, respectively. These geometries were optimised at the B97-D3(BJ)/cc-pVDZ level, as done in Section 4.3.1, and the optimised ion-pair structures are illustrated in Figure 5.1. The k_{ij}^{SAPT} values were calculated as prescribed by Goloviznina et al.[148] in the development of the CL&Pol FF. A SAPT2+/aug-cc-pVDZ level energy decomposition was carried out, where the pseudopotentials of Lim et al.[316] were used for complexes containing K^+, Rb^+, Cs^+. The values calculated in this chapter are summarised in Table 5.2 and all k_{ij}^{SAPT} values for the other fragment-fragment interactions are taken from Table 4.2 in Chapter 4.



(a) [Li][FTFSI] **(b)** [Na][FTFSI] **(c)** [K][FTFSI] **(d)** [Rb][FTFSI] **(e)** [Cs][FTFSI]

(f) $[Li]^+[C_4H_{10}]$ **(g)** $[Na]^+[C_4H_{10}]$ **(h)** $[K]^+[C_4H_{10}]$ **(i)** $[Rb]^+[C_4H_{10}]$ **(j)** $[Cs]^+[C_4H_{10}]$

Fig. 5.1. Optimised geometries of the fragment pairs used for SAPT computations required for the calculation of thek_{ij}^{SAPT} values, given in Table 5.2.

Table 5.2: Values of k_{ij}^{SAPT} obtained from the induction energy, E_{ind}, and dispersion energy, E_{disp}, computed with the SAPT2+/aug-cc-pVDZ method. All energy values are in kJ/mol.

i	j	E_{ind}	E_{disp}	k_{ij}
Li^+	C_4H_{10}	-148.31	-3.21	0.02
	$[FTFSI]^-$	-110.10	-2.30	0.02
Na^+	C_4H_{10}	-89.72	-2.68	0.03
	$[FTFSI]^-$	-54.46	-1.45	0.03
K^+	C_4H_{10}	-67.56	-15.9	0.19
	$[FTFSI]^-$	-18.14	-2.17	0.11
Rb^+	C_4H_{10}	-63.07	-19.47	0.24
	$[FTFSI]^-$	-14.90	-2.47	0.14
Cs^+	C_4H_{10}	-59.99	-23.86	0.28
	$[FTFSI]^-$	-20.37	-7.31	0.26

In Table 5.2, it is seen that the magnitude of the induction energy, E_{ind}, decreases with an increase in atomic radius, while the dispersion energy, E_{disp}, initially decreases from Li^+ to Na^+ and subsequently increases for larger cations. The initial decrease in E_{disp} from Li^+ to Na^++ can be attributed to the smaller polarizability of the Li^+ ion. Due to its high charge density and tightly held electron cloud, Li^+ exhibits limited Van der Waals interactions with neighbouring species, resulting in relatively low dispersion contributions. Moving to Na^+, the ion's larger size and more diffused electron cloud lead to a slight reduction in charge density, which reduces dispersion interactions. Beyond Li^+, the trend reverses, and E_{disp} increases with atomic radius. This is explained by the further increase in cation size and electron cloud diffusivity, which enhances the cation's ability to induce temporary dipoles in surrounding molecules. Consequently, dispersion interactions become stronger, contributing more significantly to the overall energy. The result of these trends is that k_{ij}^{SAPT} increases with metal cation radius. This shows that although the polarisability of these ions increases with an increase in radius, the ratio of induction to dispersion decreases for larger ions. As highlighted above, the previous parameterization by Goloviznina et al.[174] of $[C_2C_1im][TFSI]/[Li][TFSI]$ did not scale any Li^+ pair interactions due to the low polarisability of the metal. While this is reasonable, the k_{ij}^{SAPT} values, as shown in Table 5.2, indicate that the Van der Waals interactions involving cations with larger polarisabilities

(K^+,Rb^+, Cs^+) are reduced by less than those involving the significantly less polarisable Li^+ and Na^+. Furthermore, the $[M]^+\cdots[C_4H_{10}]$ interactions are consistently reduced by less than the corresponding $[M]^+\cdots[FTFSI]^-$ interactions, where the latter would be expected to contain a greater degree of polarisation in the LJ potential and require a smaller scaling factor. This expected trend is seen in Chapter 4 where the charged-neutral fragment interactions are scaled down less than the charge-charge interactions, as would be expected. This suggests a potential breakdown in the scaling approach when metal cations are involved.

One of the sources of the so-called polarisation catastrophe is the correlation between neighbouring Drude particle (DP) and Drude core (DC) pairs, which is addressed with the addition of the Thole damping function. However, this function was not sufficient in reducing the correlation in strongly interacting hydrogen-bonded systems. As such, it was suggested that interactions involving induced dipoles and small, highly charged ions (charge-induced dipole interactions) should also be damped using the Tang-Toennies (TT) damping function, detailed in Section 2.1.5. As the systems studied in this chapter consist of highly-charged metal cations, the TT damping function is applied here to all short-range DP-DC pair interactions where the DC is of the metal cations. It should be noted that although suggested to correct a documented systematic error in the calculated density,[148] the σ_{ij} parameters remain unscaled here, as the purpose of this study is to assess only the effect of polarisation and the effect of introducing explicit polarisation into the FF.

In summary, four FF variations are used in this chapter:

1. **CL&P**: Implemented as detailed in Section 2.1.5.

2. **sCL&P**: The CL&P FF, with all atomic partial charges scaled by 0.8.

3. **CL&Pol(1)**: CL&Pol where all ε_{ij} values are scaled using k_{ij}^{SAPT}.

4. **CL&Pol(2)**: CL&Pol where all ε_{ij} values *excluding those obtained from interactions involving the metal cations* are scaled using k_{ij}^{SAPT}.

5.3 Simulation details

All MD simulations were carried out using the LAMMPS software package.[263] Periodic cubic boxes with side lengths of $\sim$40 Å containing 30 metal cations (M^+), 120 $[C_4C_1pyr]^+$ cations and 150 $[FTFSI]^-$ anions were generated with Packmol,[264] and input and data files were generated with the `fftool` utility.[265] The protocol used to carry out MD simulations using the fixed-charge CL&P and scaled fixed-charge sCL&P FFs differs from that of simulations using the two CL&Pol variations. The CL&Pol simulations were carried out as outlined in Section 4.4, resulting in five replicate 10 ns NVT trajectories from which properties could be calculated. The CL&P and sCL&P simulations were carried out for comparison. Due to the known slower dynamics produced by fixed-charge FFs, one longer trajectory was used in place of five shorter ones, to ensure that the simulations reach a regime where dynamic properties could be determined. Therefore, for the CL&P and sCL&P simulations, a 50 ns NVT trajectory was generated.

Given the difference in simulation protocol, it is necessary to adress the statistical validity of using one longer trajectory versus several shorter, independent simulations needs to be addressed. The use of a single long simulation, as applied for the CL&P and sCL&P FFs, provides a continuous representation of the system dynamics over time, which can capture slower, long-timescale effects. However, this approach has limitations in assessing the reproducibility and variability of computed properties. Independent shorter simulations, as implemented for the polarizable FFs, offer a statistically robust alternative by enabling the calculation of uncertainties and confidence intervals through replicate sampling. This ensures that the observed behaviour is not an artifact of a single trajectory and provides a more reliable measure of the system's equilibrium properties. By contrast, the longer, singular simulations may yield lower computational costs and are often preferred for exploratory studies or when computational resources are constrained. However, for studies requiring high precision and reproducibility, multiple shorter simulations are recommended to ensure statistical validity.

For the CL&P and sCL&P FFs, the simulations were carried out as follows: Hydrogen-containing bonds were constrained using the SHAKE algorithm with a maximum of 20 iterations and a tolerance of 1.0×10^{-4}. A 12 Å cutoff was applied to the LJ and real-space Coulombic pair potentials and a long-range tail correction was included for total energy and pressure.[266] The particle-particle particle-mesh (PPPM) Ewald summation was used to evaluate long-range electrostatics, with an accuracy of 1×10^{-5} and neighbour lists were built with a skin distance of 2.0 Å. A timestep of 1 fs was used and the equations of motion were integrated using the Verlet integrator. For the *NPT* simulations, the Nosé–Hoover barostat was applied with pressure set to 1 bar, relaxed every 1 ps and the volume of the box was saved every 10 ps to calculate the average density. For *NVT* simulations, configurations were saved every 1 ps for the calculation of self-diffusion coefficients and ionic conductivities. The simulations using the CL&P and sCL&P FFs were obtained as follows:

1. For each electrolyte system, a single independent initial configuration was generated from Packmol;

2. For each initial configuration, an energy minimization with a maximum of 1000 iterations of force/energy evaluations was carried out. The minimization criteria included an energy change between iterations less than 1.0×10^{-4} (unitless) or a maximum force on any atom less than 1.0×10^{-6} newton;

3. Each system was heated for 500 ps, from 100 K to 358 K, in the *NVT* ensemble;

4. The systems were then equilibrated for 10 ns in the *NPT* ensemble to obtain the average system density for each trajectory;

5. Finally, the 50 ns production trajectory was run in the *NVT* ensemble, using the average density determined in step 4 for each system.

The simulations run using the CL&Pol FF follow the same protocol as that in Section 4.4, reiterated here for convenience: Hydrogen-containing bonds were constrained using the SHAKE algorithm with a maximum of 20 iterations and a tolerance of 1.0×10^{-4}. A 12 Å cutoff was applied to the LJ and real-space Coulombic pair potentials and a long-range tail correction was included for total energy and pressure.[266] For all pair interactions between DPs within the 12 Å cutoff, the Thole damping function was used with a damping parameter of 2.6.[157,204] The particle-particle particle-mesh (PPPM) Ewald summation was used to evaluate long-range electrostatics, with an accuracy of 1×10^{-5} and neighbour lists were built with a skin distance of 3.0 Å. A timestep of 1 fs was used and the equations of motion were integrated using the Verlet integrator. A temperature-grouped Nosé–Hoover thermostat (tgNH) was used to allow better equipartitioning of the kinetic energy.[174,267] Molecular temperature was set at 358 K with a relaxation time of 100 fs and the Drude oscillator temperature was set to 1 K with a relaxation time of 20 fs. For *NPT* simulations, the Nosé–Hoover barostat was applied with pressure set to 1 bar, relaxed every 1 ps and the volume of the box was saved every 10 ps to calculate the average density. Configurations of the NVT simulations were saved every 1 ps for the calculation of self-diffusion coefficients and ionic conductivities. The details of the specific simulation types used in this work are outlined below:

1. For each system, 5 independent initial configurations were generated;

2. For each initial configuration, an energy minimization was performed with a maximum of 1000 iterations of force/energy evaluations. The minimization criteria included an energy change between iterations less than 1.0×10^{-4} (unitless) or a maximum force on any atom less than 1.0×10^{-6} newton;

3. Each system was heated for 500 ps, from 100 K to 358 K, in the *NVT* ensemble;

4. Equilibration of the systems was carried out for 5 ns in the *NPT* ensemble to obtain the average system density for each trajectory;

5. Finally, 5×10 ns production trajectories were generated in the *NVT* ensemble, using the average density determined in step 4 for each trajectory.

5.4 Liquid structure analysis

The solvation structure around the metal cation has been pointed out to be a significant factor affecting the mobility of the ions in the electrolyte.[42,52,56,245] Therefore, in this section, structural properties including radial distribution functions, coordination numbers and spatial distribution functions are computed and discussed.

While one cannot describe the liquid structure in the same way as for a regular, predictable arrangement of particles (solid), the radial distribution function (RDF) is the most useful measure of structure in a liquid. The RDF is a pair correlation function which gives the probability of finding an observed atom a at a distance r from a reference atom b, calculated as

$$g(r) = \frac{V}{N_a \cdot N_b} \sum_{i=1}^{N_a} \sum_{j>i}^{N_b} \Big\langle \delta\big(r - (\mathbf{r}_i(t) - \mathbf{r}_j(t)\big) \Big\rangle_t, \tag{5.1}$$

where V is the system volume, N_a and N_b are the total number of atoms of type a and b, respectively, $\mathbf{r}_i$ is the position vector of atom i, $\mathbf{r}_j$ is the position vector of j, and t is each timestep considered. RDFs between the metal cations, M^+ (M^+ = Li^+, Na^+,K^+,Rb^+, Cs^+), and the $[FTFSI]^-$ anion were obtained from each *NVT* trajectory with the TRAVIS analysis software,[317,318] where an average RDF over all replicate trajectories are reported for the simulations using Cl&Pol. The M^+ ion is taken as the reference atom and the anion center-of-mass (COM) is used as the observed point. The RDFs are illustrated in Figure 5.2, from which subsequent structural information was computed. This includes the coordination number (CN), calculated using equation 5.4, the first minimum, located at r_{min}, and the first maximum, located at r_{max}; the latter two of which are used to quantify the size of the first solvation shell. These three properties are reported for each system, modelled with each FF in Table 5.3. All FFs show an increase in r_{min} as the metal cation increases in

Table 5.3: First maximum (r_{max}), first minimum (r_{min}) and coordination number (CN) determined from the $M^+ \cdots [FTFSI]^-$ ($M^+ = Li^+$, Na^+,K^+,Rb^+, Cs^+) radial distribution functions (RDFs) computed using each indicated force field. All values given for both variations of the CL&Pol FFS are the averages over all five replicate simulations with the standard deviation (SD) indicated in parentheses.

M	**FF**	r_{max}/**pm**	r_{min}/**pm**	**CN**
Li^+	Cl&P	332	598	4.1
	sCL&P	338	622	3.9
	CL&Pol(1)	425 (23)	550 (40)	4.0 (0.04)
	CL&Pol(2)	445 (18)	578 (30)	4.2 (0.04)
		—		
Na^+	Cl&P	368	602	4.3
	sCL&P	378	668	4.3
	CL&Pol(1)	435 (12)	572 (18)	4.1 (0.05)
	CL&Pol(2)	478 (20)	632 (10)	4.6 (0.1)
		—		
K^+	Cl&P	402	625	4.5
	sCL&P	408	715	4.7
	CL&Pol(1)	452 (20)	601 (15)	4.2 (0.2)
	CL&Pol(2)	418 (23)	685 (21)	4.9 (0.08)
		—		
Rb^+	Cl&P	415	688	4.6
	sCL&P	425	742	4.9
	CL&Pol(1)	482 (26)	641 (15)	4.7 0.3)
	CL&Pol(2)	425 (13)	698 (22)	5.1 (0.1)
		—		
Cs^+	Cl&P	432	728	4.8
	sCL&P	448	755	5.2
	CL&Pol(1)	428 (29)	694 (28)	5.1 (0.2)
	CL&Pol(2)	448 (19)	738 (11)	5.5 (0.09)

size, indicating the diameter within which ions in the first solvation shell fall is a direct result of the cation size, given that this will affect the probable distance between the metal ion and anion COM. For all systems, the RDFs obtained from the simulations using the CL&P and sCL&P FFs show similar features in shape, with a sharp r_{max} at ∼300–500 pm and extending to ∼600–750 pm. For these simulations, there is also wide shouldering present at distances between r_{max} and r_{min}, which becomes less pronounced with an increase in metal-cation radius. The shouldering of the first maximum peak (at $r_{\mathbf{max}}$) is less pronounced with the sCL&P FF and the r_{min} values are larger, leading to larger first solva-

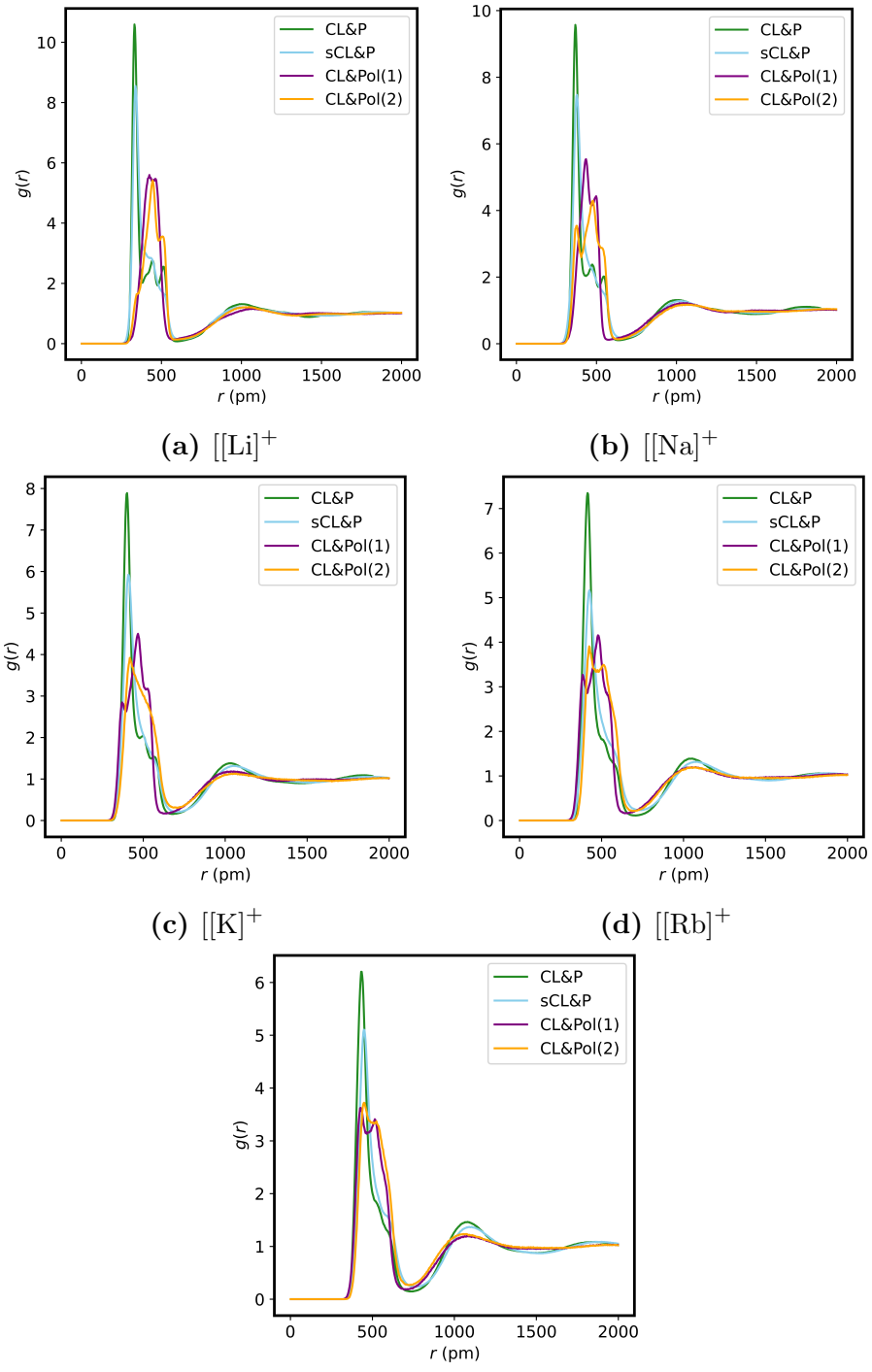
tion shells. While the CL&P and sCL&P FFs result in similar distances of r_{max}, those of the sCL&P simulations are larger than the CL&P equivalent, as would be expected due to the reduced interaction strength. The RDFs obtained from the polarisable simulations with the CL&Pol FF do not have the same shouldering as seen with the fixed charge FFs, resulting in larger r_{max} positions, but narrower first peaks and lower r_{min} values. Regarding the polarisable simulations, it is interesting that when k_{ij}^{SAPT} scaling is applied to M^+ interactions (CL&Pol(1)), the value of r_{min} decreases, indicating a decrease in the radius of the first solvation shell despite the reduction in the strength of the LJ interactions.

Considering the second maximum, $r_{\text{max,2}}$, the peaks of the second solvation shell, $g(r_{\text{max,2}})$, becomes more pronounced as the metal cation radius increases across all FFs; however, if one compares $g(r_{\text{max,2}})$ across all FFs for a single species, the CL&Pol FFs produces the least pronounced peaks. This indicates a softening of long-range interaction with polarisability, which shows one of the benefits of polarisable FFs versus scaled-charge FFs: the former has a range-dependent effect, whereas the latter applies a constant scaling of interactions, regardless of separation. Furthermore, CL&Pol(1) and CL&Pol(2) produce identical peaks at $r_{\text{max,2}}$, indicating that adjusting LJ parameters only significantly affects the interactions within the first solvation shell, ie., not long-range interactions.

Beyond the positioning and shape of the peaks, the peak height can also be used to compute the potential of mean force (PMF) to quantify the strength of interactions. The PMF provides a measure of the change in free energy (formally, the Helmholtz energy in the *NVT* ensemble) required to separate two atoms.[319] The PMF, $\omega(r)$, can be calculated from $g(r)$ as

$$\omega(r) = -k_{\text{B}}T \ln g(r), \tag{5.2}$$

where k_{B} is the Boltzmann constant and T is the temperature. Calculating accurate PMFs using Equation 5.2 requires sampling of all possible configurations, which is challenging with finite-length unbiased (equilibrium) simulations as those used here. However, assuming that the simulations are long enough to sample a representative set of configurations,



(a) $[[Li]^+$ **(b)** $[[Na]^+$

(c) $[[K]^+$ **(d)** $[[Rb]^+$

(e) $[[Cs]^+$

Fig. 5.2. Radial distribution functions (RDFs) for metal cations (Li^+, Na^+, K^+, Rb^+, Cs^+) with the centre of mass (COM) of the $[FTFSI]^-$ anions at 358 K obtained from the *NVT* trajectories using the CL&P, sCL&P, and CL&Pol(1) and CL&Pol(2) force fields.

it is suitable to take Equation 5.2 as

$$\omega(r) \approx -k_{\mathrm{B}}T \ln g(r). \tag{5.3}$$

In doing so, $\omega(r_{\mathrm{max}})$ can be estimated to provide an interpretation of the relative peak heights at r_{max} for the FFs tested. The values of $\omega(r_{\mathrm{max}})$ obtained for each system using each FF are shown in Figure 5.3. The absolute values of $\omega(r_{\mathrm{max}})$ tends to decrease

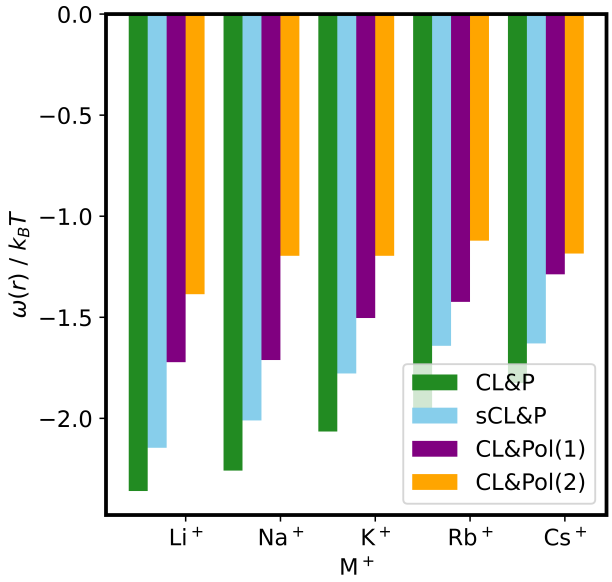


Fig. 5.3. Values of the potential of mean force, $\omega(r_{\mathrm{max}})$, calculated at r_{max} for all radial distribution functions of the $[\mathrm{FTFSI}]^-$ anion around the M^+ (Li^+, Na^+, K^+, Rb^+, Cs^+) cation in all CL&P, sCL&P, CL&Pol(1) and CL&Pol(2) simulations.

with an increase in the radius of the metal cations, suggesting weaker interactions as the metal ion size increases, which is expected. For all species, $\omega(r_{\mathrm{max}})$ is smaller in absolute value when using sCL&P, compared to values obtained with CL&P, which can be expected from weaker interactions,resulting from charge scaling. The values of $\omega(r_{\mathrm{max}})$ calculated for the CL&Pol(1 and 2) simulations are the smallest in the absolute value of all FFs, and the relative magnitude of the PMFs produced by the CL&Pol(1) and CL&Pol(2) simulations do agree with the corresponding values of r_{min}. While differences are apparent between the structuring obtained with CL&P and sCL&P, and likewise between CL&Pol(1)

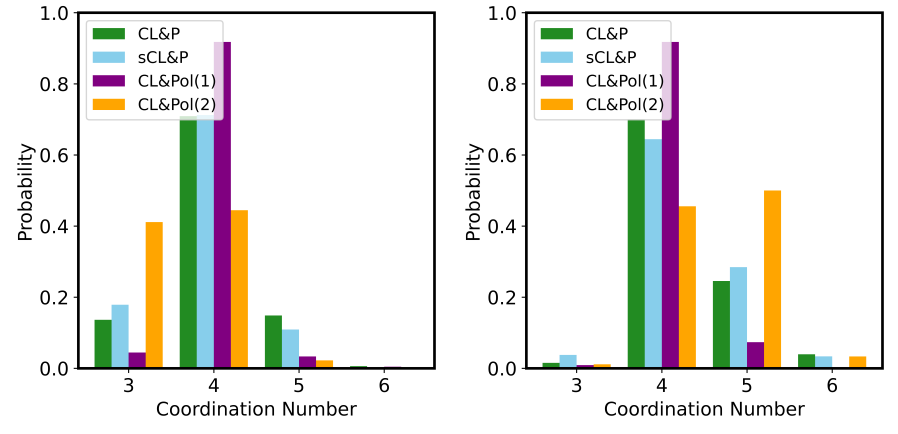
and CL&Pol(2), Figure 5.3 shows that the most significant difference results from whether Drude polarisation is present or not.

To explore the solvation structuring of the metal cations, the CNs of the $[FTFSI]^-$ anions in the first solvation shell around the metal cations were obtained from the corresponding integral of the RDF up to r_{min}:

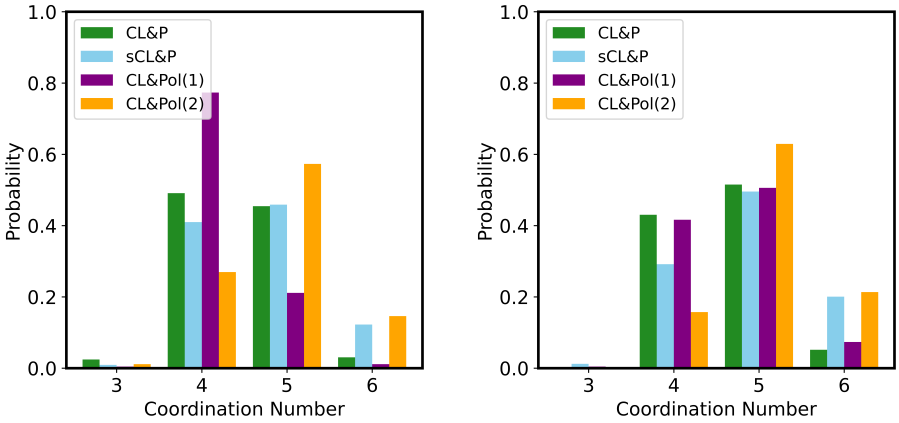
$$\text{CN}(r_{\text{min}}) = 4\pi\rho \int_0^{r_{\text{min}}} r^2 g(r)\, dr, \tag{5.4}$$

where ρ is the system density. The positions of r_{max} and r_{min} in the RDFs for each metal differ between the four FFs considered; however, the CNs of $[FTFSI]^-$ around each metal cation differ by at most 0.5. Furthermore, the average CNs overall systems fall within $\sim$4–6 (with larger values for the larger metal cations, as can be expected). To provide a more insightful interpretation of coordination, the CN distribution was calculated for each system, as a result of each FF. This was done by determining the CN of each individual metal cation in each trajectory, and binning the values to provide a distribution over the 30 ions. These distributions are illustrated in Figure 5.4. Here, it can be seen that, except for Cs^+, the polarisable FFs leads to a narrower distribution of CNs. Interestingly, the CN distribution varies significantly between the CL&Pol(1) and CL&Pol(2) simulations, even though the atomic sites which coordinate with the metal likely differ. As such, it is important to identify which of the approaches, CL&Pol(1) and CL&Pol(2), best represents IL-based electrolytes, given the significant structural differences produced by each.

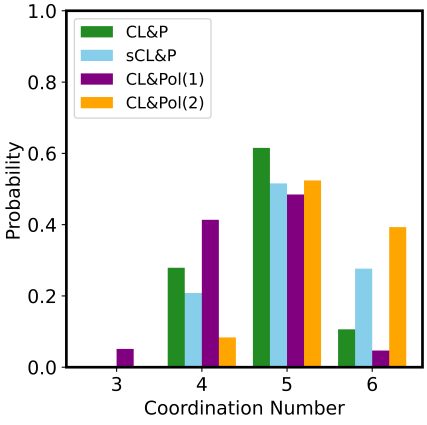
The spatial distribution function (SDF) provides the probability that an observed atom is at a certain position around a reference atom or molecule. To visualise the SDF, an iso-surface is defined, that being a surface which passes through all points with the same probability density. The isosurface can then be mapped onto the reference molecule. SDFs of the metal cations around the anions in each simulation were calculated, with the S$-$N$-$S atoms of the anion as the orientation reference vector, at an isosurface density value of 10



(a) $[Li]^+$ **(b)** $[Na]^+$



(c) $[K]^+$ **(d)** $[Rb]^+$



(e) $[Cs]^+$

Fig. 5.4. Coordination number distributions of the $[FTFSI]^-$ centre-of-mass around the metal cations (M^+) for the simulations obtained using the CL&P, sCL&P and CL&Pol force fields.

atoms/nm^3 using TRAVIS to provide a three-dimensional description of the metal coordination to the anion (See Figure 5.5). Across the CL&P and sCL&P FF simulations, there is significant coordination seen with the oxygen atoms that sit in the S−N−S plane, which less of is seen in the polarisable simulations. The SDFs also indicate that the coordination of the metals is predominantly with the oxygen atoms given that metal coordination is localised at these sites. The Cl&Pol(1) simulations show greater coordination with the arms of the anion, indicating a stronger interaction between the fluorine-containing peripheries and the metal cation when explicit polarisation is added. The CL&Pol(2) simulations produce SDFs that differ from CL&Pol(1), where the former has reduced metal ion interaction with the $-CF_3$ group. This indicates that the reduction of metal-anion interaction strength favours interactions with the $-CF_3$ group. Whether the difference in coordination distribution is due to an increase in the strength of interaction or a difference in geometry of the coordinating anions.

To begin, the anion conformer distribution was analysed across the *NVT* trajectories obtained with each FF (using the average distribution over all configurations for the CL&Pol generated trajectories). Using TRAVIS, the distribution of the dihedral angles defined in Figure 5.6 was computed, producing combined distribution functions (CDFs) of the two defined dihedrals, illustrated in Figure 5.7. These figures provide insight into the relative distribution by providing probability maps of the anion conformers present in the *NVT* trajectories. These maps are symmetrical about the lines $y = -x$ and $y = x$. Therefore, the following discussion refers to the cisoid conformer, where $-180^\circ < \tau$(F-S-N-S) $< 0^\circ$ and $0^\circ < \tau$(S-N-S-C) $< 180^\circ$ (top left), and transoid conformer where $-180^\circ < \tau$(F-S-N-S) $< 0^\circ$ and $-180^\circ < \tau$(S-N-S-C) $< 0^\circ$ (bottom left). The CDFs indicate a largely even distribution of cisoid and transoid conformers with the CL&P FF for the metal-doped systems. However, when charge scaling is introduced with sCL&P, the cisoid conformer is more dominant. Interestingly, the inclusion of Drude polarisation results in a similar distribution to that of CL&P and there is no notable difference between the CDFs produced by the CL&Pol(1)

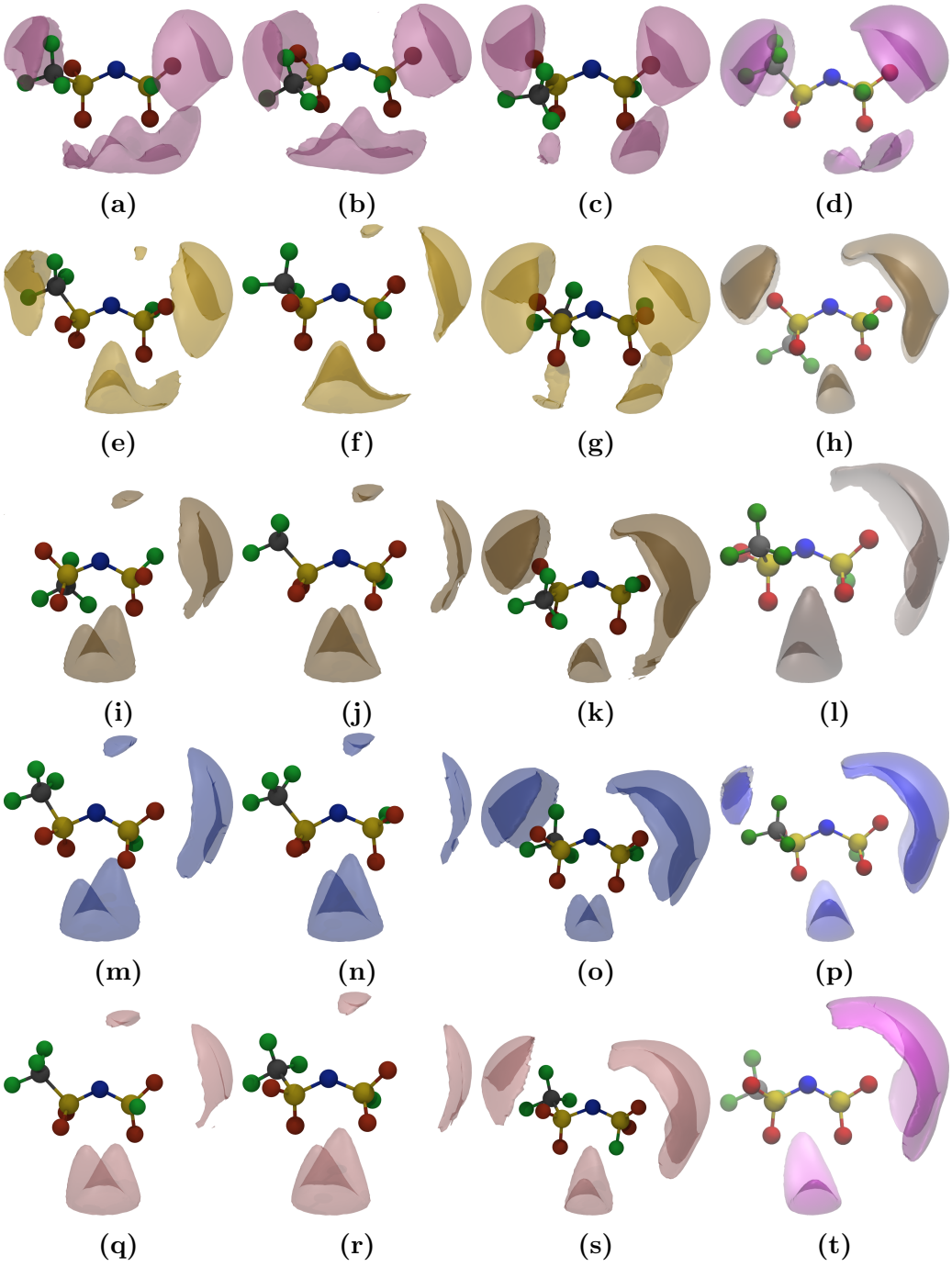


Fig. 5.5. Spatial distribution functions (SDFs) for Li^+ (a-d), Na^+ (e-h), K^+ (i-l), Rb^+ (m-p) and Cs^+ (q-t) around the $[FTFSI]^-$ anions at 358 K obtained using the CL&P (column 1), sCL&P (column 2), CL&Pol(1) (column 3), and CL&Pol(2) (column 4) force fields.

and CL&Pol(2) FFs. The CDFs therefore show that the difference in SDFs is not a result of conformer distribution, as there is no apparent correlation between the interaction regions and the dominating conformer. What the CDFs do elude to is that although scaling charges

is a common remedy to speed up dynamics, this can have a marked effect on conformer distribution!

While Figure 5.7 indicates that the anion conformer distribution is largely consistent between the pure and metal-doped system, the occurrence of the cisoid conformer is less prevalent in the pure system across all FFs, indicating that the metal cation does affect the conformer distribution. However, the effect of the metal cation is less significant than that of the FF, indicating that anion coordination is a direct consequence of the FF and it is the anion conformer that largely determines the coordination site of the metal cation, rather than the metal cation "pulling" the anion into a preferred conformer. Therefore, the distribution of metal coordination is a result of the anion conformer distribution and as such, the RDFs and solvation structures become highly dependent on the accurate parameterization of the anion torsion angles.

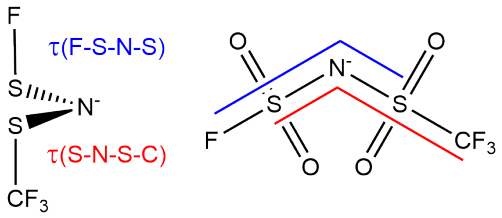
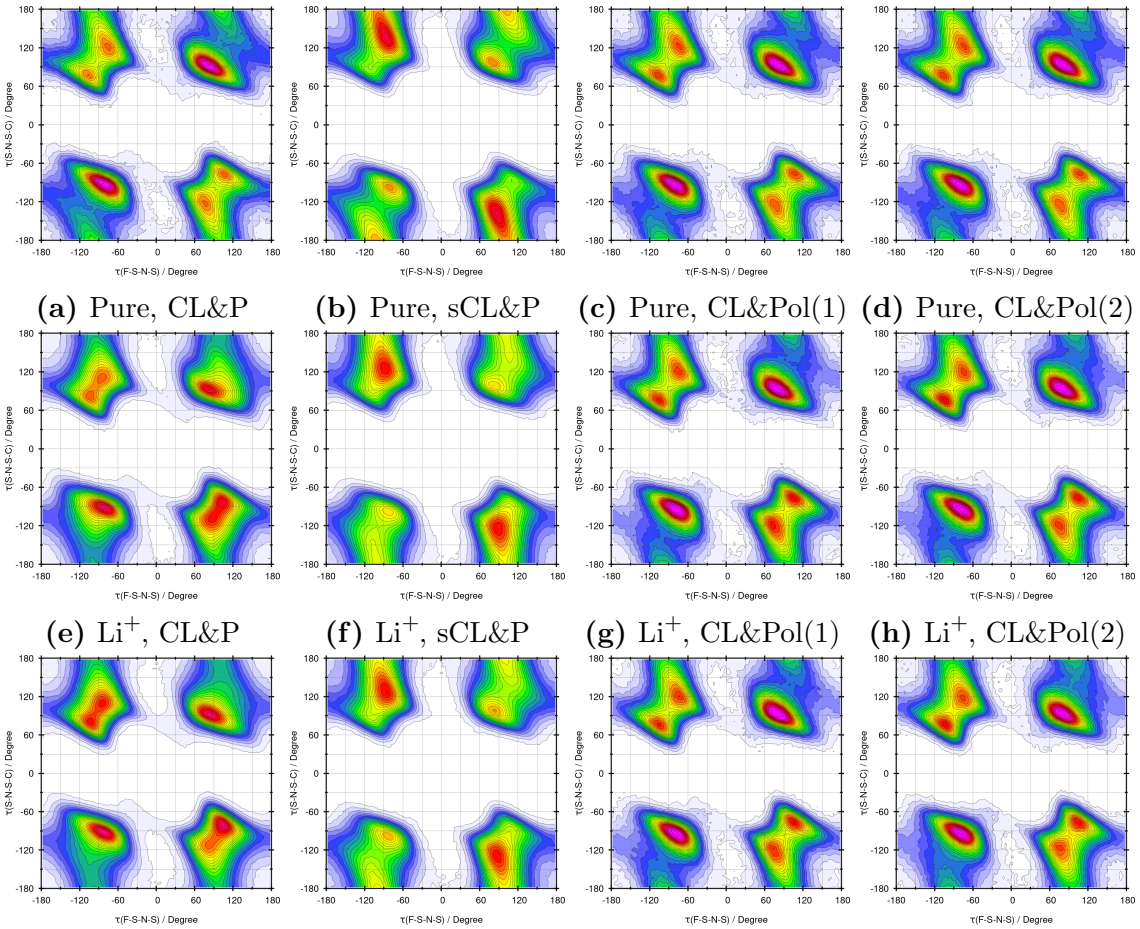


Fig. 5.6. Illustration of dihedrals used for dihedral angle distributions (See Figure 5.7).

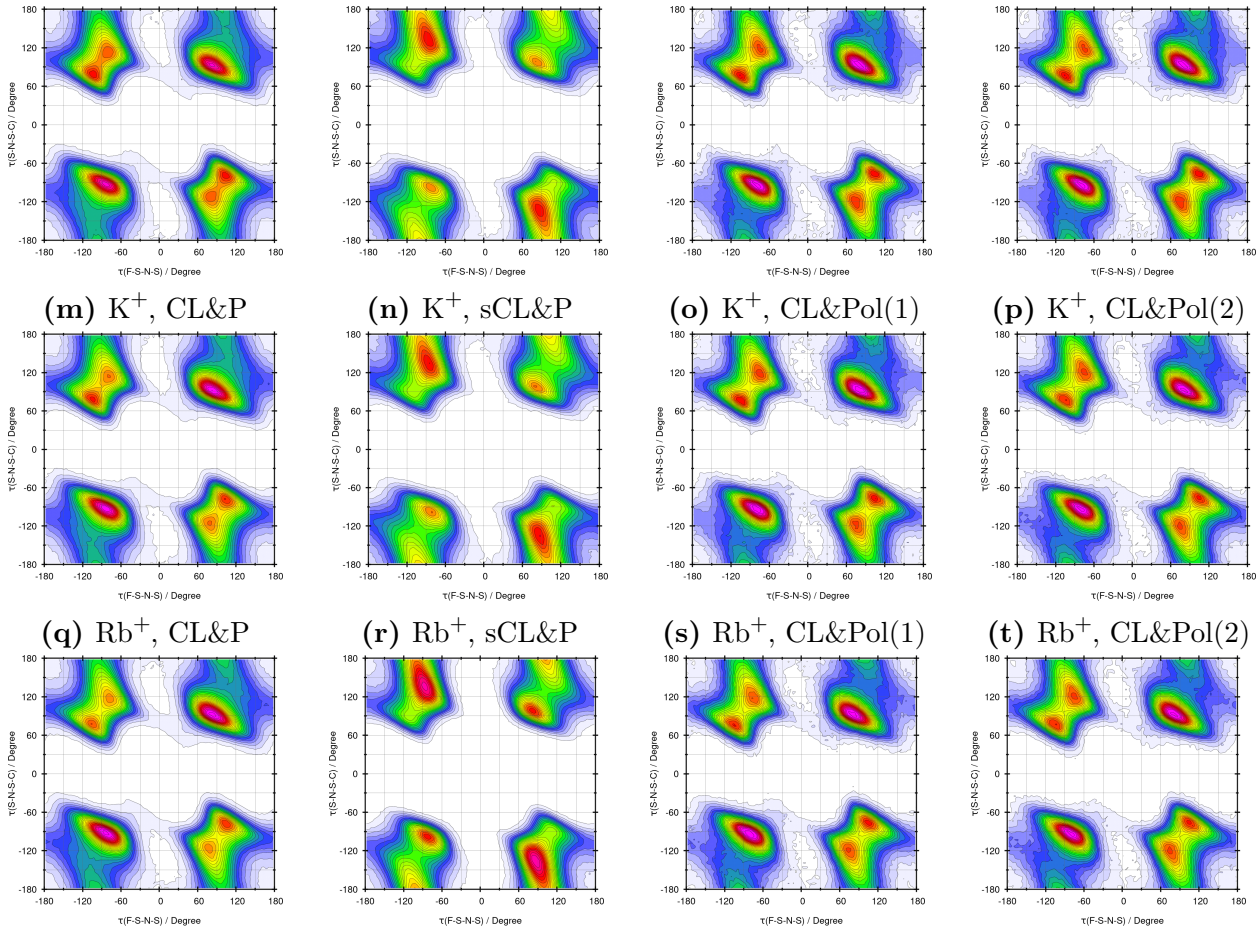


(a) Pure, CL&P **(b)** Pure, sCL&P **(c)** Pure, CL&Pol(1) **(d)** Pure, CL&Pol(2)

(e) Li^+, CL&P **(f)** Li^+, sCL&P **(g)** Li^+, CL&Pol(1) **(h)** Li^+, CL&Pol(2)

(i) Na^+, CL&P **(j)** Na^+, sCL&P **(k)** Na^+, CL&Pol(1) **(l)** Na^+, CL&Pol(2)

Fig. 5.7. Combined distribution functions (CDFs) showing the distribution of the dihedral angles, defined in Figure 5.6, in the *NVT* ensemble at 358 K. The CDFs are given for each [X][FTFSI]/[C_4C_1pyr][FTFSI] system considered, simulated using the CL&P, sCL&P, CL&Pol(1) and CL&Pol(2) force fields, showing the relative distribution of cisoid ($-180° < \tau$(F-S-N-S) $< 0°$ and $0° < \tau$(S-N-S-C) $< 180°$) and transoid ($-180° < \tau$(F-S-N-S) $< 0°$ and $-180° < \tau$(S-N-S-C) $< 0°$) anion conformers for each. CDFs corresponding to the CL&Pol force field variations are the average CDFs over all five *NVT* trajectories.



(m) K^+, CL&P **(n)** K^+, sCL&P **(o)** K^+, CL&Pol(1) **(p)** K^+, CL&Pol(2)

(q) Rb^+, CL&P **(r)** Rb^+, sCL&P **(s)** Rb^+, CL&Pol(1) **(t)** Rb^+, CL&Pol(2)

(u) Cs^+, CL&P **(v)** Cs^+, sCL&P **(w)** Cs^+, CL&Pol(1) **(x)** Cs^+, CL&Pol(2)

Fig. 5.7. Combined distribution functions (CDFs) showing the distribution of the dihedral angles, defined in Figure 5.6, in the *NVT* ensemble at 358 K. The CDFs are given for each [X][FTFSI]/[C_4C_1pyr][FTFSI] system considered, simulated using the CL&P, sCL&P, CL&Pol(1) and CL&Pol(2) force fields, showing the relative distribution of cisoid ($-180° < \tau$(F-S-N-S) $< 0°$ and $0° < \tau$(S-N-S-C) $< 180°$) and transoid ($-180° < \tau$(F-S-N-S) $< 0°$ and $-180° < \tau$(S-N-S-C) $< 0°$) anion conformers for each. CDFs corresponding to the CL&Pol force field variations are the average CDFs over all five *NVT* trajectories (cont.).

5.5 Calculated thermodynamic and transport properties

5.5.1 Density

The average density, ρ, was calculated from each NPT trajectory as explained in Section 3.2.1. The percentage deviation from experimental values for each calculated value is illustrated in Figure 5.6. Across all simulations, an increase in metal cation mass is consistent with an increase in the calculated values of ρ, matching the experimental trends,[95] as well as an increase in the absolute value of the error, with respect to experiment. The CL&P FF produces the lowest overall deviation from experiment, with an mean absolute percentage error (MAPE) of 0.6 %. While the scaling included with sCL&P serves to improve transport properties, it has been shown that the decrease in electrostatic interaction strength decreases the density,[145] as seen here with a significant decrease in the absolute value of densities calculated, and an increased MAPE of 2.18 %. While the sCL&P FF produces a greater MAPE, both CL&P and sCL&P errors are systematic in that they become more negative with an increase in metal cation radius. The CL&Pol(1) FF results in an MAPE of 2.16 %, comparable to that of sCL&P, but lacking the same systematic trend in errors. The CL&Pol(2) simulations produced the largest deviation, underpredicting densities with an MAPE of 3.11 %. This is a significant result indicating that not scaling metal interactions results in larger MAPEs, with the deviations being more marked with larger metal cations. Attention should be given to this result as, based on previous CL&Pol application to IL-based electrolytes, it is expected that for Li^+ interactions, k_{ij}^{SAPT} should equal 1.[174] However, the densities calculated here show that even for the relatively non-polarisable Li^+, the LJ parameters need reconsideration when incorporating Drude polarisability. It is worthwhile to reiterate that it has been suggested, when using the CL&Pol FF, the σ_{ij} values should be scaled to correct for the underprediction of ρ, but as the scope of this

chapter is a comparative study, σ_{ij} values were not scaled with the implementation of CL&Pol. For comparison, ρ was also computed using each FF for the pure system (ie., replacing the metal cation M^+ with $[C_4C_1pyr]^+$ and running the four simulation types as with the M-doped systems) and the percentage errors relative to experiment are also shown in Figure 5.6. The deviations seen using each FF for the pure IL, for each FF, are of a similar order to those of the electrolyte systems, suggesting that errors are systematic of the FF implemented, and not a consequence of the addition of the metal cations.

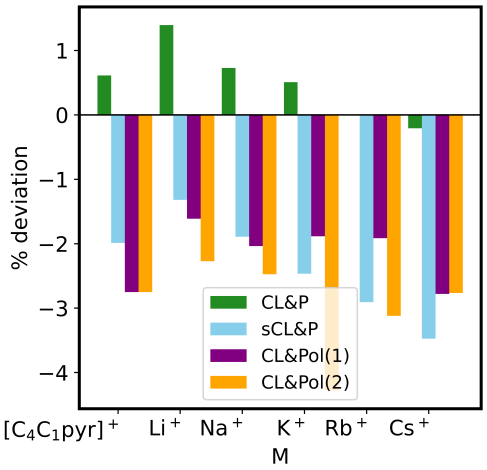


Fig. 5.6. Percentage deviation of calculated densities from experimental densities[95] for the [M][FTFSI]/[C_4C_1pyr][FTFSI] systems investigated.

5.6 Conductivity

Einstein-Helfand (EH) ionic conductivities, σ_{EH}, are calculated as detailed in Section 3.2.4. The calculated values of σ_{EH} are given in Table 5.4, and the values of σ_{EH} obtained with each FF, relative to experiment, are illustrated in Figure 5.7. The simulations run with the CL&P FF produce significantly underpredicted σ_{EH} values, with an MAPE of 97 %. The scaling of charges is expected to speed up ion motion and in turn, improve the transport properties. Using sCL&P does result in larger values of σ_{EH}, producing an MAPE of 72 %; however, these values are once again underpredicted, with the largest deviation with

the [Cs][FTFSI]/[C_4C_1pyr][FTFSI] system. If this system is excluded from analysis, the expected behaviour of charge scaling is seen, such that the MAPE would be 66 %. The calculated values of σ_{EH} from the CL&Pol(1) simulations produce an MAPE of 16 %. Thus, except for [Cs][FTFSI]/[C_4C_1pyr][FTFSI], values are still underpredicted. Even so, the large CI associated with $\sigma_{EH}(Cs^+)$ suggests that a significantly lower value could be expected for this system: Across the five simulations, all values of $\sigma_{EH}(Cs^+)$ are below 1.5 S/m except for one with a value of 4.03 S/m. If the outlier is removed, the 95 % CI is reduced and the average value of $\sigma_{EH}(Cs^+)$ drops to 1.12 S/m. Thus, the 95 % CI allows for a more refined analysis. A very satisfying result seen with the CL&Pol(1) simulations, not seen with the others, is that the experimental trend in $\sigma_{EH}(M^+)$ for [M][FTFSI]/[C_4C_1pyr][FTFSI] is reproduced: $\sigma_{EH}(Na^+) < \sigma_{EH}(Li^+) < \sigma_{EH}(K^+) < \sigma_{EH}(Rb^+) < \sigma_{EH}(Cs^+)$. The CL&Pol(2) simulations (with no k_{ij} scaling of M^+ pair interactions) produce an MAPE of 72 %, and do not reproduce the experimental trends as seen with CL&P(1). However, the large 95 % CIs do indicate that caution should be taken when interpreting any apparent trends. Given that the 95 % CIs indicate the range of values in which the conductivity is expected to fall, and these ranges overlap for the systems using both CL&Pol(1) and CL&Pol(2). The consequence is that, if a new set of configurations were considered for the calculation of σ_{EH}, it is equally probable that the average values obtained with CL&Pol(1) would not reproduce the experimental trend or that CL&Pol(2) would.

5.6.1 Viscosity

The value of η is estimated from the plateau of the running integrals, determined using the same protocol as described in Section 3.2.3, for each NVT simulation run with each FF. The calculated values are presented in Table 5.4 and the running integrals for all simulations are given in Appendix B.1, indicating where η was estimated for each trajectory. The values of η determined from the CL&P simulations are exceptionally over-predicted, with an error of $\sim$50 times that of the experimental values due to the running integrals

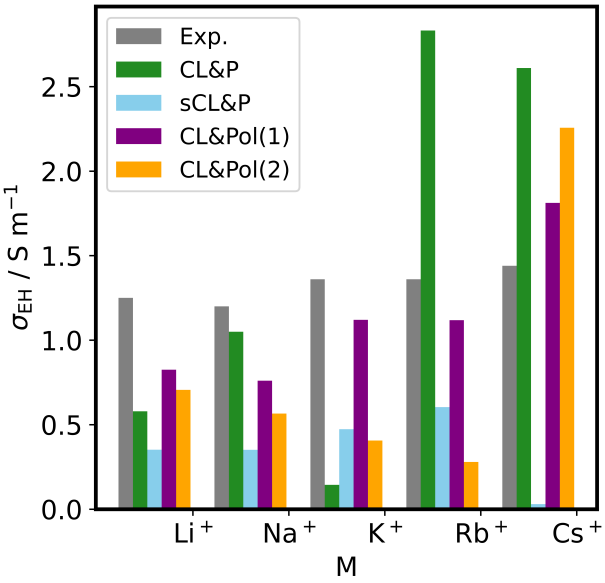


Fig. 5.7. Calculated ionic conductivities (σ_{EH}) for the [M][FTFSI]/[C_4C_1pyr][FTFSI] systems investigated, obtained from the NVT simulations run using the CL&P, sCL&P, CL&Pol(1) and CL&Pol(2) FFs and a comparison against the available experimental data of Yamamoto et al.[95] The error bars show the 95 % CI for the mean when calculating conductivity using the replicate simulations.

not showing a distinct plateau within the 50 ns of simulation time. While these simulations would likely reach a plateau after a significantly longer simulation run, this is not feasible from the point of computational resources. Furthermore, by the 50 ns point, the estimated values already sit much higher than the experiment, suggesting that even if the simulation were run long enough to result in a plateau in the running integral, this would be *at least* as high as the last point shown here, already significantly overpredicted. When scaling is applied this provides a significant improvement on the running integrals produced, but a large MAPE of 107 % persists. The CL&Pol(1) FF produces values of η with an MAPE of 185 % and the CL&Pol(2) simulations produce viscosities with an MAPE of 110 %. As with the calculated ionic conductivities, the large 95 % CIs of the calculated values produced with both CL&Pol implementations highlight the inconsistency associated with this calculation. Thus, care should be taken to conclude as to whether the scaling metal interactions do or do not provide more accurate values of η than CL&Pol. This is once again a prime

example of the importance of reporting CIs with calculated properties, as these must be taken into consideration when interpreting errors. Therefore, what is of importance to these results is that including charge scaling or polarisability significantly improves the mobility of the systems, compared to the fixed-charge equivalent, in turn improving the quality of the data used to calculate η.

Figure 5.8 shows the abilities of the FFs to reproduce trends in η across the metal cations. Regarding the experimental values, there is no clear trend when moving through the set of metal cations. Furthermore, the experimental values of η all fall within a range of $\sim$4 cP which indicates that η is largely independent of the metal cation and is rather a function of the IL itself, and trying to differentiate between the viscosities of the systems studied here with MD is beyond the capabilities of such simulations. The sCL&P, while producing a MAPE comparable to that of both CL&Pol implementations, fails to predict the similarity in values of η over the set. CL&Pol(2) does provide a more consistent range of η values, while still showing larger deviations. Once again, the large CIs produced also indicate the very clear need for multiple simulations, and show how averaging over more simulations is necessary to capture the behaviour of viscosity.

5.6.2 Self-diffusion coefficients

While no experimental self-diffusion coefficients were available for the systems studied here, it was shown in Chapter 4 that the CIs associated with the calculated self-diffusion coefficients for a species i, $D^{\infty}_{\mathrm{s},i}$, are typically much lower than those for the collective properties σ_{EH} and η. Therefore, even without experimental values, self-diffusion coefficients corrected for finite-size effects were computed as described in Section 3.2.2 to provide a measure of ion transport and mobility with lower uncertainty than the calculated collective properties. Based on data available for ILs with $[C_4C_1pyr]^+$ and metal cations[52,58,274] it can be expected that $[C_4C_1pyr]^+$ should have higher self-diffusion coefficients than the metal cations, which is seen across all FFs. While the CL&P simulations are capable of reproduc-

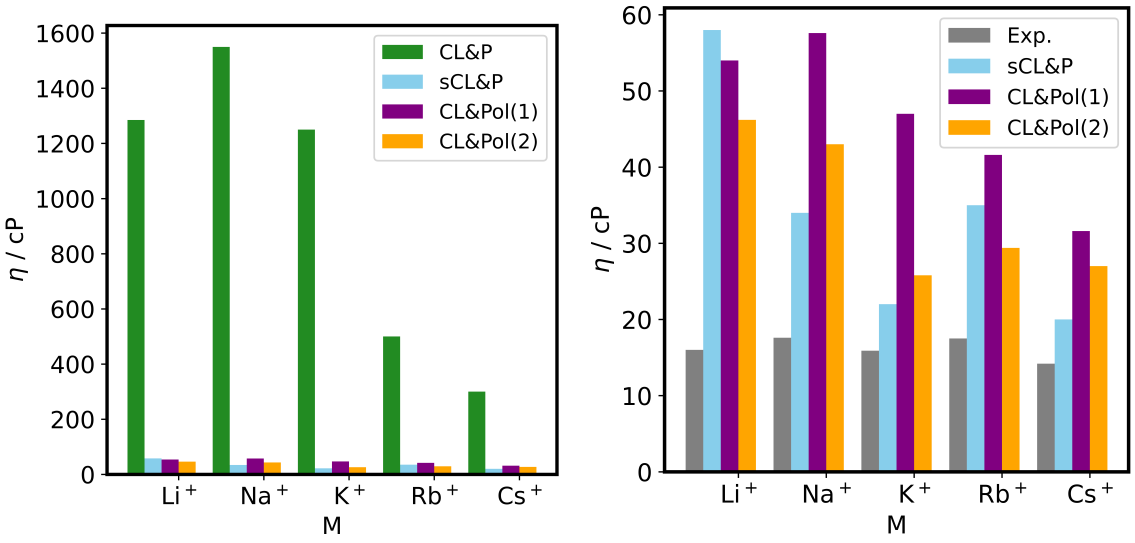


Fig. 5.8. (a) Calculated viscosities (η) for the [M][FTFSI]/[C_4C_1pyr][FTFSI] systems investigated, obtained from the simulations run using the CL&P, sCL&P and CL&Pol FFs and (b) a comparison of the latter two to the available experimental data of Yamamoto et al.[95]. The error bars show the 95 % CI for the mean when calculating viscosity using the replicate simulations.

ing the same trends in mobility as both CL&Pol FF versions implemented, the values of the CL&P diffusivities are an entire order of magnitude smaller than those produced with the other FFs. Using the β parameter to assess the linearity of the MSD (see Appendix B.2), it is apparent that the CL&P MSDs are not sufficiently linear indicating that the simulations do not reach a linear regime, resulting in lower than expected diffusivity. These results show that using the fixed charge FF will produce much lower diffusion coefficients, and while the sCL&P FF does raise these values in line with the CL&Pol(1) and CL&Pol(2) values, the linear regions are not as well attained as with the polarisable FFs, yielding caution when interpreting the sCL&P results.

The accuracy of self-diffusion coefficients depends significantly on the length of the molecular dynamics simulation, as longer trajectories enable better sampling of ion motion, reduce statistical noise, and capture long-timescale dynamics such as ion-pair dissociation or cluster rearrangements. While this study used a 50 ns trajectory to balance

computational cost and precision, longer simulations may be required for improved accuracy, particularly in systems with strong ion pairing or complex structural dynamics. The calculated self-diffusion coefficients reflect the influence of metal-anion structuring discussed earlier in the chapter, where tight metal-anion coordination, such as with Li+ restricts ionic mobility, resulting in lower diffusion coefficients. In contrast, weaker interactions or larger coordination environments, as seen for Na^+ and K^+, promote higher mobility and greater diffusion. These findings highlight how local ion organization directly impacts macroscopic transport properties, emphasizing the interplay between simulation protocol and the underlying structural dynamics.

A significant result of the CL&Pol(1) simulations is slightly faster diffusion of all the species in the [Li][FTFSI]/[C_4C_1pyr][FTFSI] system than in the [Na][FTFSI]/[C_4C_1pyr][FTFSI] systems. The same effect is seen experimentally from the calculated values of σ_{EH} and η for these systems,[91] thus suggesting that only this FF gives a description of diffusivity that is representative of the expected trends. It is interesting that when metal cation interactions are not reduced with k_{ij}^{SAPT}, as with CL&Pol(2), the diffusion coefficients are higher than with CL&Pol(1) — it would be expected that a reduction in interaction strength would increase the diffusion coefficients. Although CL&Pol(2) does produce overall higher diffusion coefficients than CL&Pol(1), the difference produced by the two FFs for i, $D_{\mathrm{s,M}}^{\infty}$ is less marked than for i, $D_{\mathrm{s},+}^{\infty}$ and i, $D_{\mathrm{s},-}^{\infty}$. This suggests that the scaling down of interactions with the metal cations causes the systems to be dominated by the interactions of $[C_4C_1\mathrm{pyr}]^+$ and $[\mathrm{FTFSI}]^-$, effectively increasing the relative strength of these interactions, decreasing the overall mobility of the cation and anion.

5.7 Conclusion

The study carried out in this section analysed two important factors regarding the simulation of alkali metal cations in [C_4C_1pyr][FTFSI]: The need for explicit polarisation, and

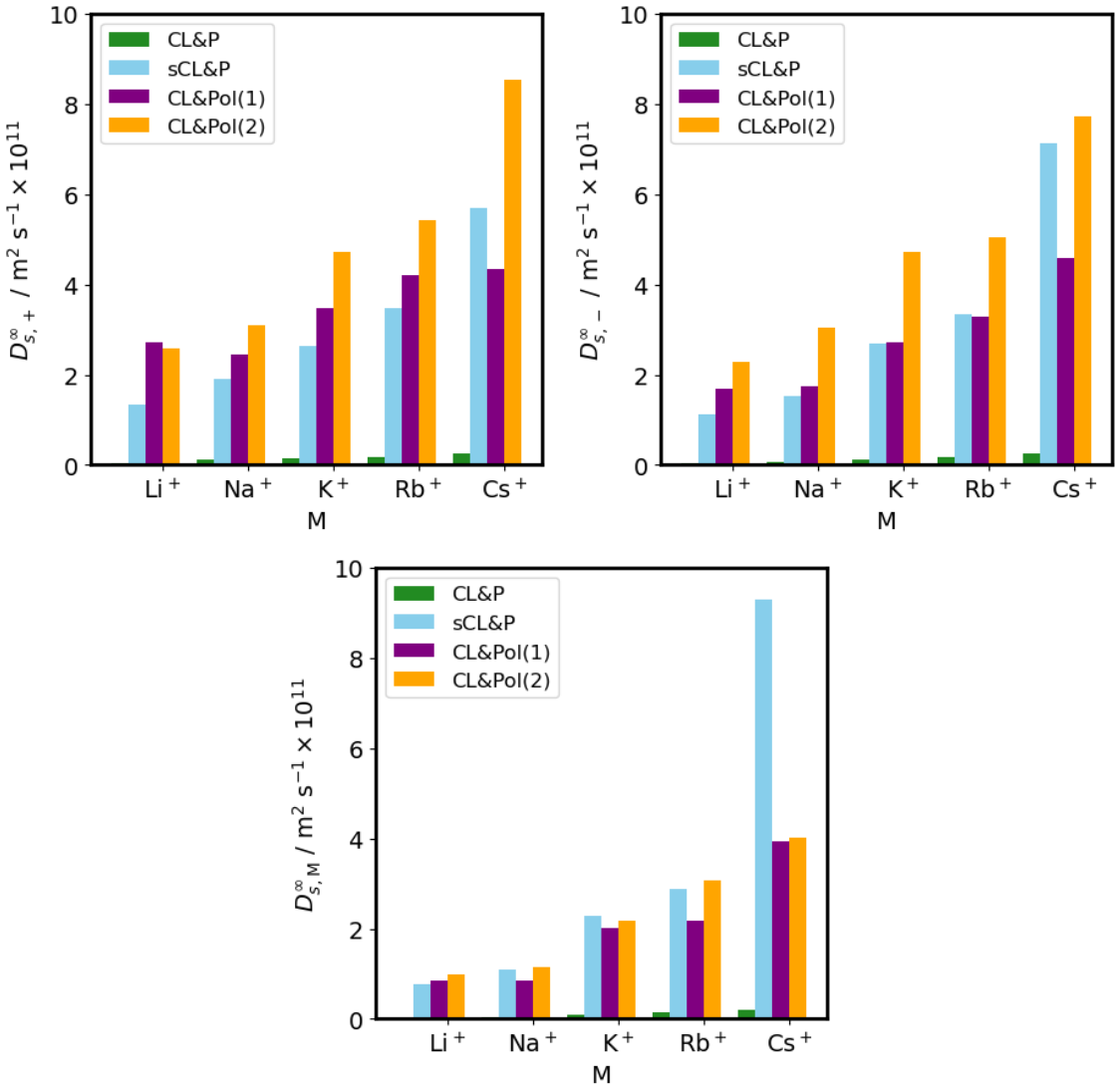


Fig. 5.9. Calculated self-diffusion coefficients for $[C_4C_1pyr]^+$ (left), $[FTFSI]^-$ (middle) and M (right) for each [M][FTFSI]/$[C_4C_1pyr]$[FTFSI] system.

Table 5.4: Calculated self-diffusion coefficients of the cation (left) and anion (middle) and metal cation (right). The error bars show the 95 % CI for the mean when calculating self-diffusivity using the replicate simulations.

	ρ/g/cm^3	$D^{\infty}_{s,+}/10^{-11}$ m^2/s	$D^{\infty}_{s,-}/10^{-11}$ m^2/s	$D^{\infty}_{s,M}/10^{-11}$ m^2/s	σ_{EH} S/m	η/cP
[Li][FTFSI]/[C$_4$C$_1$pyr][FTFSI]						
Exp.	1.364	-	-	-	1.25	16
CL&P	1.383	0.051	0.042	0.012	0.016	1285
sCL&P	1.346	1.349	1.127	0.761	0.352	58
CL&Pol(1)	1.342 (0.001)	2.718 (0.57)	1.678 (0.60)	0.849 (0.39)	0.825 (0.53)	54 (19.6)
CL&Pol(2)	1.333 (0.001)	2.577 (0.80)	2.287 (1.49)	0.982 (0.20)	0.439 (0.47)	46.2 (17.8)
[Na][FTFSI]/[C$_4$C$_1$pyr][FTFSI]						
Exp.	1.374	-	-	-	1.2	17.6
CL&P	1.384	0.109	0.068	0.027	0.029	1550
sCL&P	1.348	1.898	1.538	1.087	0.351	34
CL&Pol(1)	1.346 (0.002)	2.434 (0.52)	1.752 (0.20)	0.843 (0.14)	0.76 (0.227)	57.6 (38.4)
CL&Pol(2)	1.341 (0.002)	3.083 (1.28)	3.051 (1.28)	1.518 (0.62)	0.30 (0.56)	43.0 (23.9)
[K][FTFSI]/[C$_4$C$_1$pyr][FTFSI]						
Exp.	1.379	-	-	-	1.36	15.9
CL&P	1.386	0.156	0.121	0.089	0.004	1250
sCL&P	1.345	2.628	2.694	2.285	0.473	22
CL&Pol(1)	1.353 (0.002)	3.466 (0.55)	2.731 (0.55)	2.007 (0.78)	1.12 (0.606)	47 (38.2)
CL&Pol(2)	1.321 (0.080)	4.721 (1.7)	4.722 (1.6)	2.172 (1.88)	0.41 (0.22)	25.8 (13.4)
[Rb][FTFSI]/[C$_4$C$_1$pyr][FTFSI]						
Exp.	1.41	-	-	-	1.36	17.5
CL&P	1.410	0.166	0.177	0.136	0.077	500
sCL&P	1.369	3.482	3.329	2.890	0.604	35
CL&Pol(1)	1.383 (0.001)	4.195 (0.92)	3.286 (0.71)	2.188 (0.60)	1.12 (0.615)	41.6 (26.1)
CL&Pol(2)	1.366 (0.003)	5.42 (0.30)	5.05 (0.41)	3.07 (0.60)	0.28 (0.25)	29.4 (9.6)
[Cs][FTFSI]/[C$_4$C$_1$pyr][FTFSI]						
Exp.	1.439	-	-	-	1.44	14.2
CL&P	1.436	0.262	0.255	0.194	0.070	300
sCL&P	1.389	0.413	0.507	0.543	0.04	100
CL&Pol(1)	1.399 (0.001)	4.353 (0.55)	4.578 (0.92)	3.927 (1.23)	1.81 (1.604)	31.6 (31.9)
CL&Pol(2)	1.399 (0.002)	8.533 (0.62)	7.73 (0.92)	4.014 (1.08)	2.257 (0.74)	27.0 (8.3)

the effect of k_{ij}^{SAPT} scaling of interactions involving the metal cations. While charge scaling (sCL&P) provides transport properties of comparable accuracy to that of the CL&Pol simulations, RDFs and SDFs show that Drude polarisation produces significantly different structuring than both fixed-charge variations.

The comparison of the CL&Pol(1) and CL&Pol(2) simulations was sought after as two previous studies in which CL&Pol was applied used OPLS parameters for Li^+ without scaling, and Na^+ parameters from the CHARMM-Drude FF. Therefore, the parameterization for a series of alkali metals required investigation. The work carried out here showed that scaling of these interactions, for all metals considered, produced differing structural and transport properties, even for Li^+. This is significant as, based on the low polarisability of Li^+, it is often assumed that LJ interactions involving this cation should not require adjustment; however, the CL&Pol(1) simulations consistently performed the best for [Li][FTFSI]/[C_4C_1pyr][FTFSI]. This indicates that the introduction of Drude induction does require an adjustment (reduction) of alkali metal parameters, which is achieved with the use of k_{ij}^{SAPT}. Although providing an adjustment which improves accuracy, the fundamental intention of k_{ij}^{SAPT} is to remove induction contributions, which should show a correlation with the polarisability. As the low-polarisability atoms have the smallest values of k_{ij}^{SAPT} (large induction contributions) here, this calls into question whether the use of k_{ij}^{SAPT} to scale alkali-metal interactions is theoretically and physically sound.

Part II

Further Development of the CL&Pol Force Field for the Simulation of Metal Cations in Ionic Liquids

Chapter 6

Resolving Van der Waals energy components with Symmetry-Adapted Perturbation Theory

6.1 Introduction

As detailed in Section 2.3, Symmetry-Adapted Perturbation Theory (SAPT) describes the interaction energy of a dimer, E_{int}, as the sum of contributions of electrostatic, induction, dispersion and exchange energy components. SAPT is an appealing method as one can, in principle, expand each perturbation to the infinite order to produce the full configuration interaction limit of E_{int}.[235] In a large and systematic analysis carried out by Parker et al.,[233] E_{int} was obtained with various SAPT decomposition schemes at varying levels of complexity and benchmarked against values obtained with the "gold standard" Coupled-Cluster with Singles, Doubles and Perturbative Triples (CCSD(T)) method, extrapolated to the Complete Basis Set (CBS) limit. The gold, silver and bronze SAPT standards were subsequently determined, with average absolute relative deviations (AARDs) of 0.63, 1.23 and 2.05 kJ/mol, respectively, calculated over the S22, HBC6, NBC10 and

HSG datasets. It is significant to note that the datasets used in this benchmark consisted of neutral monomers, and later work by Lao et al.[320] showed that the recommended gold and silver SAPT levels of Parker et al. are outperformed by other methods when applied to systems containing charged monomers. When applied to the AHB21 dataset consisting of anion-neutral complexes, the gold, silver and bronze levels produced AARDs (relative to the CCSD(T)/CBS equivalents) of 1.88, 3.22 and 6.61 kJ/mol, respectively, while using the previously unranked SAPT2+(CCD)δMP2/aug-cc-pVTZ method produced an AARD of 1.17 kJ/mol. Across the CHB6 (cation-neutral complexes) dataset, the gold and silver levels produced AARDs of 3.05 and 6.86 kJ/mol, respectively, whereas the SAPT2+3δMP2/aug-cc-pVQZ method produced an AARD of 0.84 kJ/mol. Lastly, for the IL16 dataset (consisting of cation-anion pairs), the gold, silver and bronze levels produced AARDs of 3.85, 2.59 and 21.76 kJ/mol, respectively, and again the SAPT2+3δMP2/aug-cc-pVQZ method produced the lowest AARD of 1.38 kJ/mol. Therefore, the study by Lao et al.[320] indicated that higher levels of SAPT are required to achieve even silver standard accuracy for charged systems, and the robustness of these methods when applied to non-neutral classes of chemical systems deserves more attention.

Another factor to consider concerning the accuracy of SAPT energy components is the size of the basis set. As detailed in Section 2.2, Hartree-Fock methods ignore correlated motion and to provide truly accurate energies, correlation must be accounted for. The correlated methods discussed in Section 2.2 use Slater determinants to expand the wavefunctions and as such, show extremely slow convergence, as this expansion does not accurately represent the wavefunction behaviour at the cusp (where the interelectronic distances approach zero).[321] One approach to dealing with this deficiency, is to use a basis set which is dependent on the interelectronic distance, introducing explicit correlation, which is not practical for large systems. Another way to obtain accurate energies, without explicit correlation, is to make use of hierarchized correlation-consistent basis sets, which can be extrapolated towards the CBS limit. It is well known that the electronic energy converges

smoothly to the basis set limit[322,323] using Dunning's suite of systematically constructed correlation-consistent basis sets;[324] therefore, with these basis sets, SAPT components can be extrapolated toward the CBS limit. This provides an appealing approach to obtain SAPT interaction energies void of basis-set incompleteness errors. In addition, it has been suggested by Kodrycka et al.[325] that as the SAPT energy components (and the terms of which they consist) show different convergence behaviours, the convergence of each component can be assessed independently. One of the first studies in which the CBS limit for SAPT components was reported is that of Misquitta and Szalewicz.[326] Extrapolation of the SAPT E_{int} was performed for the He_2, Ne_2, $(H_2O)_2$ and $(CO_2)_2$ dimers using the aug-cc-pVQZ and aug-cc-pV5Z basis sets[324] and Helgaker's two-point extrapolation scheme that was proposed to estimate correlation energies at the CBS limit,[323,327]

$$E_{\text{CBS}}^{\text{est.}} = \frac{E_{X+1}(X+1)^p - E_X X^p}{(X+1)^p - X^p}. \tag{6.1}$$

Here E_X and E_{X+1} are the energies computed using correlation-consistent basis sets with cardinal numbers X and $X+1$, $E_{\text{CBS}}^{\text{est.}}$ is the estimated CBS limit, and $p = 3$; this will henceforth be referred to as a Helgaker extrapolation. Where other values of p are used, this will be referred to as an $X/X+1$ extrapolation, where a*X*Z is shorthand for aug-cc-pVXZ. Equation 6.1 follows from expressing the convergence of the correlation energy as follows:

$$E_X = E_{\text{CBS}} + AX^{-p}, \tag{6.2}$$

and equating the CBS limits written using consecutive E_X and E_{X+1} values to eliminate the constant A. While $p = 3$ for the Helgaker extrapolation, Truhlar[328] and Schwenke[329] suggested using a variable exponent that can be determined by fitting to numerical data.

Řezáč and Hobza[330] explored the convergence behaviour of the SAPT components for 10 small organic complexes using the series of aXZ basis sets with X = D, T, Q and 5. Two methods to obtain the DFT-SAPT dispersion energy at the CBS limit were proposed:

The first is an $X/X+1$ extrapolation using Equation 6.1 with $p = 2.868$. The value of p was optimized by fitting the calculated dispersion energies (the sum of the dispersion and exchange-dispersion terms) across the entire series of basis sets to Equation 6.2, with A and E_{CBS} as variables. The DFT-SAPT/CBS E_{int} was then obtained by adding the resulting extrapolated dispersion component to the non-dispersion components, the latter calculated using the large a5Z basis set. Any two consecutive basis sets considered produced values of E_{int} within close agreement to the CCSD(T)/CBS total interaction energies, with root-mean-square deviations (RMSDS) of 0.15 kJ/mol and 0.025 kJ/mol for aDZ/aTZ and aTZ/aQZ extrapolations, respectively. The second approach was to simply scale the dispersion component obtained with each basis set to fit the CBS limit. Scaling the aDZ dispersion component by 1.193 and adding it to the aTZ non-dispersion components gave an RMSD of 0.435 kJ/mol while scaling the aTZ dispersion by 1.051 and adding to the aTZ non-dispersion components gave an RMSD of 0.209 kJ/mol. The basis set dependence of DFT-SAPT interaction energies was also evaluated by Heßelmann and Korona,[331] using the S12L data set consisting of large organic complexes, and correlation-consistent basis sets up to X = Q. These authors showed that the aDZ basis set gave electrostatic and exchange energies that are well-converged with only a minimal underestimation of induction, compared to results calculated using the aQZ basis set. Furthermore, the non-augmented cc-pVTZ basis set was deemed sufficient to reliably estimate the aQZ results for non-dispersion terms. On the contrary, cc-pVTZ produced worse results for dispersion than those with aug-cc-pVDZ, illustrating the importance of diffuse functions for this contribution. Dispersion energies calculated with the cc-pVDZ, aug-cc-pVDZ and aug-cc-pVTZ basis sets deviated from the CBS limit, estimated using an aTZ/aQZ extrapolation with Equation 6.1 and $p = 3$, with AARDs of 35 %, 15 % and 7 %, respectively. However, since dispersion energies extrapolated using the cc-pVDZ and cc-pVTZ basis sets were found to systematically underestimate the CBS results by about 8 %, Heßelmann and Korona suggested that it would be reasonable to estimate DFT-SAPT/CBS energies

for large complexes by scaling the cc-pVDZ/cc-pVTZ extrapolated dispersion by 1.08 and adding this to the non-dispersion terms calculated using cc-pVTZ. SAPT interaction energies at the CBS limit based on DFT treatment of monomers have been reported in several other studies.[320,332–340] In all these works, the DFT-SAPT/CBS (sometimes referred to as SAPT(DFT)/CBS) interaction energies were obtained by extrapolating only the dispersion energy using Dunning's augmented correlation-consistent basis sets and Equation 6.1, with $p = 3$, and adding the extrapolated values to the non-dispersion (electrostatics, exchange-repulsion and induction) components computed with a sufficiently large basis set, usually the larger of the two used in the extrapolation. The majority of these references used an aTZ/aQZ extrapolation.

Although most of the previous work focused on DFT-SAPT, CBS extrapolations using WFT-based SAPT have also been reported. Przybytek[341] used the A24 data set of non-covalent complexes[220] and compared the uncoupled dispersion calculated using an explicitly correlated F12 approach with the corresponding aQZ/a5Z CBS limit estimated using Equation 6.1 and $p = 3$. Wang et al.[342] developed a database of SAPT2+ energy components for 101 small molecules at seven intermolecular distances (S101$\times$7). Examining a smaller subset of five molecules showed that the electrostatic and induction energies obtained with the aTZ basis set were reasonably close to the CBS limit, as approximated by the a5Z results. Exchange-repulsion and induction were then estimated using an aDZ/aTZ extrapolation in which Equation 6.1, with optimized p values of 3.0 and 4.3, respectively, were used. Furthermore, the extrapolated dispersion energies were scaled by a factor of 0.89 to account for missing higher-order terms in SAPT2+ (see Section 2.3). This scaling factor was determined by minimizing the absolute difference between the CCSD(T)/CBS and SAPT2+/CBS interaction energies across the S66 data set. It is therefore evident that the CBS limits of WFT-based SAPT have not been investigated to the same extent as that of DFT-SAPT.

This chapter aimed to develop a protocol to compute high-quality interaction energy components using SAPT, that could subsequently be used for the development of physically-motivated force field (FF) parameters. Thus, the results here present several systematically developed protocols and the careful consideration that is given to (1) the level of perturbation and (2) the basis set size, allowing one to produce highly accurate energy components. Consequently, this chapter first presents a systematic study of the basis set convergence of both non-correlated (electrostatics, exchange and induction) and correlated (dispersion) SAPT components using Dunning's correlation-consistent basis set.[324] The second section of the results then presents an approach to include high-order SAPT components, without lower-level computations.

6.2 Methodology

6.2.1 Dataset selection and extension

The AHB21 dataset was used in this work, consisting of 21 hydrogen-bonded dimers, each composed of a neutral molecule and one of the F^-, Cl^-, N_3^-, SH^-, or HCO_2^- anions, shown in Figure 6.1. This dataset was selected because Lao et al.[320] showed that to obtain accurate interaction energies, higher orders of perturbation and larger basis sets are needed for anion-neutral than for cation-neutral (or cation-anion) dimers; this observation is important as this chapter specifically explores the convergence of the SAPT interaction energy as the basis set size increases. The equilibrium geometries, optimised at the MP2/aug-cc-pVTZ level of theory, were taken from Lao et al.,[320] as well as CCSD(T)/CBS interaction energies. The dataset was then extended to include distance-dependent geometries, for which interaction energies and energy components could be obtained, as including the distance dependence of the interaction energy is particularly important in the context of FF development.[342] The distance-dependent points were generated by scaling the inter-monomer distance of the equilibrium geometries as follows (Illustrated in Figure 6.2):

- The intermonomer vector (IV) was determined as follows: If a hydrogen bond was present between the monomers, the IV was determined as the distance between the hydrogen-bond donor and acceptor atoms (See (a) in Figure 6.2). If there were multiple hydrogen bonds, the IV was taken as the average over all donor-acceptor vectors (See (b) in Figure 6.2). If there were no hydrogen bonds present, the IV was then defined by the centre-of-mass (COM) of each monomer (See (c) in Figure 6.2).

- The IV of each equilibrium geometry was then scaled, keeping the individual ion geometries fixed, by vector scaling factors (VSFs) of 0.9, 1, 1.25, 1.5 and 2, generating five geometries for each dimer at varying intermonomer distances, resulting in the AHB21$\times$5 dataset.

6.2.2 Computational Details

6.2.2.1 SAPT computations

SAPT energy decompositions were carried out for each geometry (each dimer complex at each VSF) in the AHB21$\times$5 dataset. Dunning's augmented correlation consistent aug-cc-pVXZ basis sets[324] were used, and are abbreviated to aXZ where X is the cardinal number, of which X = D, T, Q, 5, 6 were used. Both SAPT0 and SAPT2+ energy decompositions (see Section 2.3) were carried out for the AHB21$\times$5 dataset. The SAPT0 decompositions carried out here include the $\delta E^{(2)}_{\text{HF}}$ contribution, where this term was summed to the induction terms and as such was included in the induction energy component. Where SAPT2+ was used, the δE_{MP2} correction was summed to $\delta E^{(2)}_{\text{HF}}$. Additionally, the dispersion terms were computed with coupled-cluster doubles (CCD) amplitudes to allow for highly accurate dispersion calculations; this computation is henceforth abbreviated as SAPT2+(CCD)δMP2. Furthermore, the calculation of all second-order SAPT terms (including δMP2) implemented the frozen core approximation, and density fitting was used. Use of the single-exchange approximation (SEA)[236,237] has been shown to produce errors for

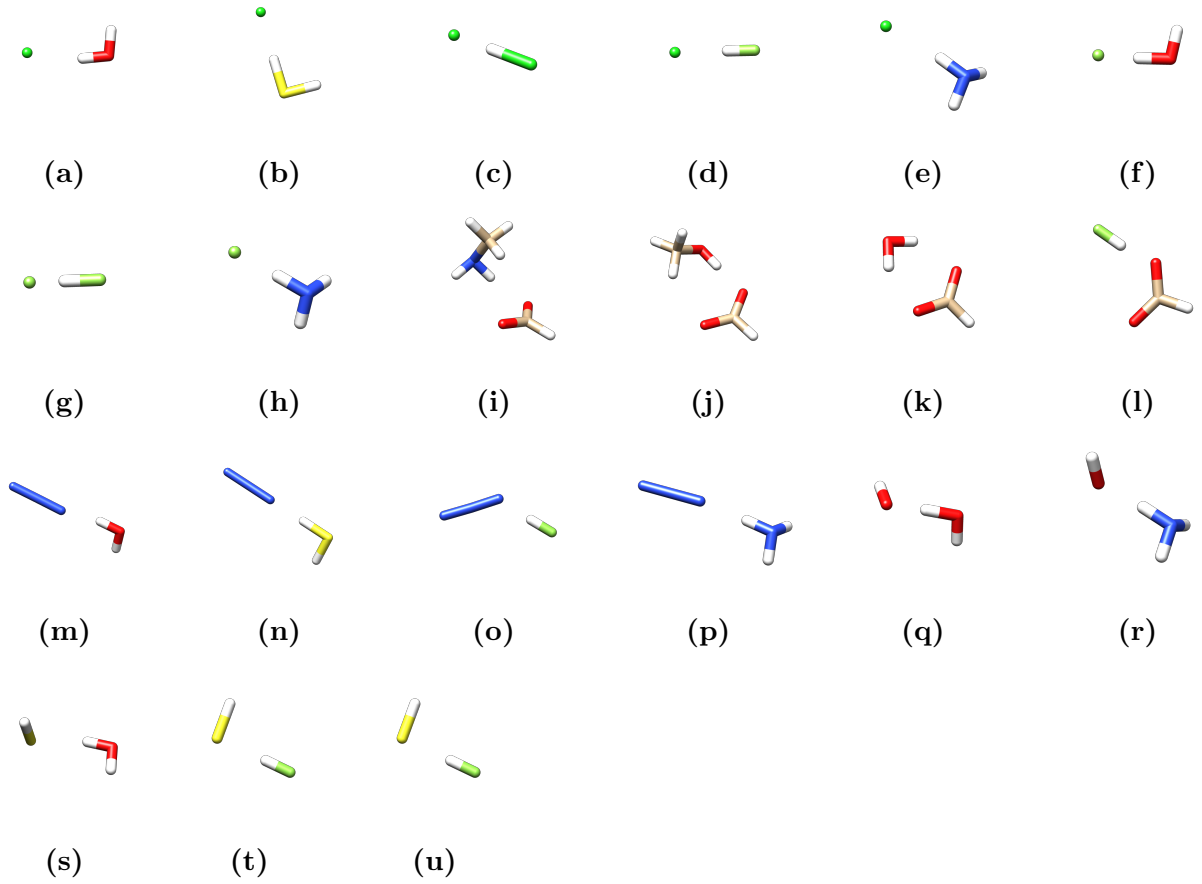


Fig. 6.1. Structures of the equilibrium geometries of the complexes in the AHB21×5 dataset. The structures correspond to the following equilibrium complexes: (a) $Cl^-(H_2O)$; (b) $Cl^-(H_2S)$; (c) $Cl^-(HCl)$; (d) $Cl^-(HF)$; (e) $Cl^-(NH_3)$; (f) $F^-(H_2O)$; (g) $F^-(HF)$; (h) $F^-(NH_3)$; (i) $HCO_2^-(MeNH_2)$; (j) $HCO_2^-(MeOH)$; (k) $HCO_2^-(H_2O)$; (l) $HCO_2^-(HF)$; (m) $N_3^-(H_2O)$; (n) $N_3^-(H_2S)$; (o) $N_3^-(HF)$; (p) $N_3^-(NH_3)$; (q) $OH^-(H_2O)$; (r) $OH^-(NH_3)$; (s) $SH^-(H_2O)$; (t) $SH^-(HF)$ and (u) $SH^-(NH_3)$.

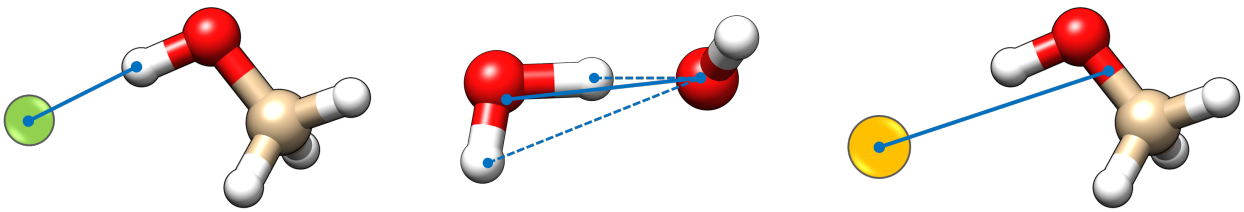


Fig. 6.2. Illustration of the determination of the intermonomer vector (IV) for: (a) F^-(MeOH), containing a single hydrogen bond donor-acceptor pair; (b) $OH^-(H_2O)$, containing multiple hydrogen bonds; and (c) F^-(MeOH), containing no hydrogen bonds. The solid blue lines represent the IVs, and the dashed lines represent the vectors determined by individual hydrogen bonds.

second- and third-order exchange terms at intermolecular distances shorter than the equilibrium.[343] The use of VSF=0.9 therefore warranted non-approximated computations. As such, second-order exchange-containing terms were computed without the SEA, and the analogous values with the SEA applied were also computed for comparison. Psi4 version 1.4[344] was used for all SAPT computations.

6.2.2.2 Reference interaction energies

The CBS protocol outlined below was used to compute reference interaction energies (used to compute AARDS) as was done by Lao et al.[320] to ensure consistency between the original AHB21 dataset and the additional geometries generated here. Therefore, the reference interaction energies were calculated as CCSD(T)/CBS interaction energies such that

$$E_{\text{int,CBS}}^{\text{CCSD(T)}} = E_{\text{HF}} + E_{\text{corr,CBS}}^{\text{CCSD(T)}}. \tag{6.3}$$

In Equation 6.3, E_{HF} is the Hartree-Fock (HF) energy calculated with the aQZ basis set, and $E_{\text{corr,CBS}}^{\text{CCSD(T)}}$ is the CCSD(T) correlation energy estimated at the CBS limit using the Helgaker scheme.[323,327] For the larger HCO_2^- containing complexes, $E_{\text{int,CBS}}^{\text{CCSD(T)}}$ was extrapolated using explicitly correlated CCSD(T)-F12 correlation energies computed with the cc-pVDZ-

F12/cc-pVTZ-F12 basis sets, and added to the HF/cc-pVTZ-F12 energy. The corresponding near-complete CABS auxiliary basis sets were used,[345,346] and the HF-F12 basis set correction[347,348] was included. The CCSD(T)-F12 implementation in the ORCA software,[349] used for these computations, uses the fixed diagonal amplitude ansätz C[350,351] including the F12-based non-iterative, perturbative treatment of triples (i.e., CCSD(T)$_{\overline{F12}}$), as implemented by Valeev and Daniel Crawford.[352] The frozen core approximation was used for all correlation energy calculations. All CCSD(T)-F12 calculations were carried out using ORCA 4.0[349] and all other calculations with Psi4.[344] The calculated reference interaction energies for the AHB21×5 dataset are given in Table 6.1.

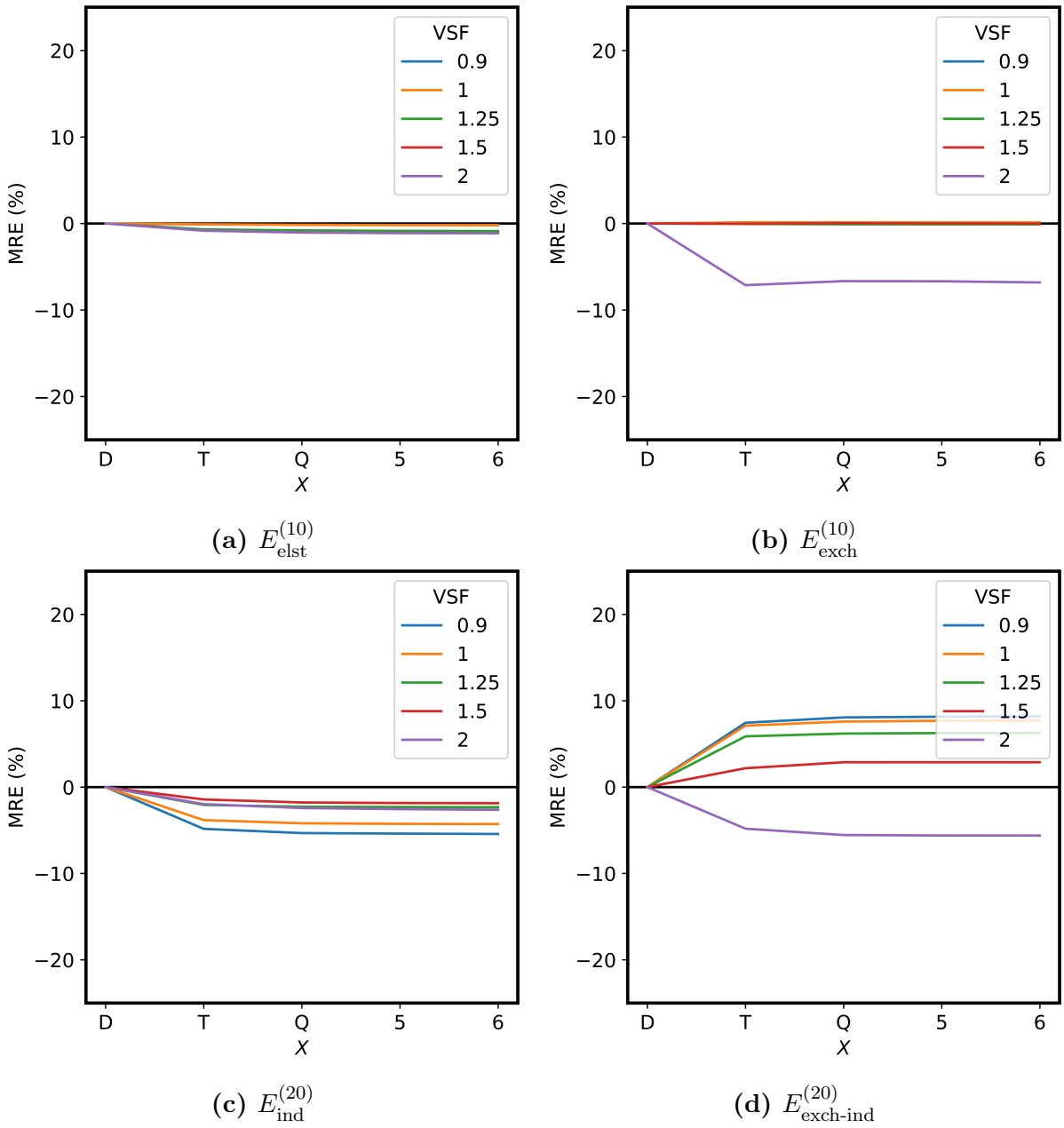
6.3 Basis set convergence of SAPT0 terms

The SAPT0 interaction energy, $E_{\text{int}}^{\text{SAPT0}}$, is defined in this work as given in Equation 2.104, repeated here for convenience:

$$E_{\text{int}}^{\text{SAPT0}} = E_{\text{elst}}^{(10)} + E_{\text{exch}}^{(10)} + E_{\text{ind,resp}}^{(20)} + E_{\text{exch-ind,resp}}^{(20)} + E_{\text{disp}}^{(20)} + E_{\text{exch-disp}}^{(20)} + \delta E_{\text{HF}}^{(2)}. \tag{2.104}$$

Given that no intramonomer correlation is present in each of the above terms, they might have been expected to show a quick convergence with an increase in X; however, the dispersion terms ($E_{\text{disp}}^{(20)}$ and $E_{\text{exch-disp}}^{(20)}$) are known to show a slower convergence due to the presence of *intermonomer* correlation effects.[330,331,339] The value of each term computed with each basis set is expressed as a percentage relative to the value computed with the aDZ basis set, and this difference is henceforth referred to as the mean relative error (MRE). The resulting MREs over the full dataset are illustrated for each term in Figure 6.3, highlighting the slower convergence of the correlated terms ($E_{\text{disp}}^{(20)}$ and $E_{\text{exch-disp}}^{(20)}$) as the basis set size increases. Despite this, all terms do show a systematic convergence, suggesting that in general the CBS limit can be extrapolated where it cannot be estimated with a SAPT decomposition using a sufficiently large basis set. The $E_{\text{exch}}^{(10)}$ and $E_{\text{exch-ind}}^{(20)}$ terms at VSF $= 2$

do show a difference in convergence behaviour than at other VSFs, exhibiting a decrease in value with an increase in basis-set. While Figure 6.3 indicates that increasing the basis set size results in a ∼5–10 % decrease in the values of $E^{(10)}_{\text{exch}}$ and $E^{(20)}_{\text{exch-ind}}$ at VSF $= 2$, the absolute values are always less than 0.1 kJ/mol at this separation, consistently contributing less than 1 % to the overall energy. As exchange is a close-range effect it can be expected that the values will approach zero quickly with an increase in monomer separation; therefore, while the figures do indicate a breakdown in the convergence behaviour of these terms at long distances, the values are low enough such that they become independent of basis set. The MRE of the $\delta E^{(2)}_{\text{HF}}$ term indicates that the VSF $= 1.5$ geometries initially show a decrease in value when increasing the basis set from aDZ to aTZ, after which the MRE becomes more positive. However, the MREs at VSF $= 1.5$ all fall within 0.5 % of each other, and the term is considered to show convergence with an increase in basis set size at all separation distances.



(a) $E^{(10)}_{\text{elst}}$ **(b)** $E^{(10)}_{\text{exch}}$

(c) $E^{(20)}_{\text{ind}}$ **(d)** $E^{(20)}_{\text{exch-ind}}$

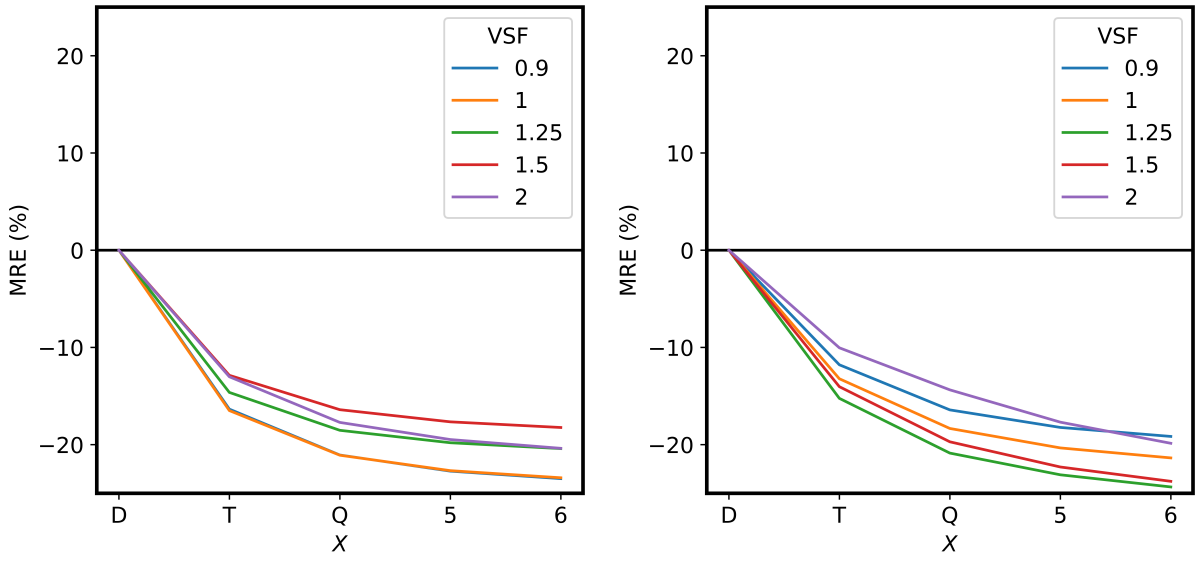
Fig. 6.3. Mean relative energy (MRE) for each of the SAPT0 terms, at each basis set, relative to the aDZ basis set, averaged over all geometries at each VSF in the AHB21×5 dataset.

The absolute percentage deviation (AD) between energy values determined with consecutive basis sets was calculated for each component as the difference between the [X+1]Z and aXZ basis sets. Ideally, if the AD is zero across all complexes at all distances,

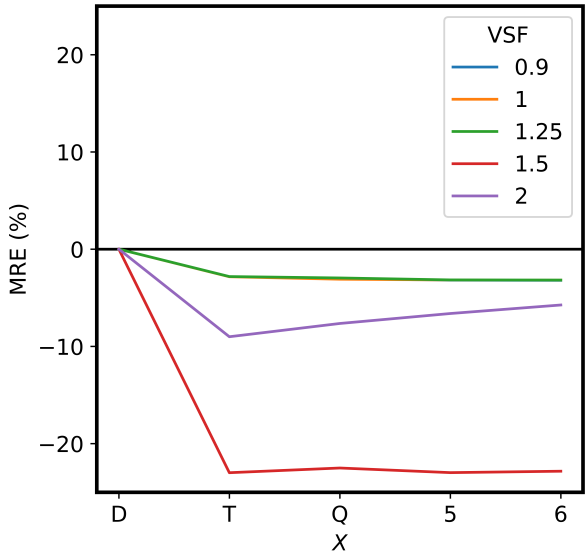
that term could in general be considered converged at the smaller of the two consecutive basis sets (aXZ).

Table 6.1: CCSD(T)/CBS reference interaction energies, $E_{\text{int,CBS}}^{\text{CCSD(T)}}$, calculated for the AHB21×5 dataset, given in units of kJ/mol (kcal/mol given in parentheses). The VSF indicates the factor by which the equilibrium intermonomer distance was scaled. The $E_{\text{int,CBS}}^{\text{CCSD(T)}}$ values calculated by Lao et al.[320] for the original AHB21 dataset are given, in units of kcal/mol for comparison.

	VSF					
Complex	0.9	1	1[320]	1.25	1.5	2
$Cl^-(H_2O)$	-50.7 (-12.1)	-65.3 (-15.6)	-15.6	-46.6 (-11.1)	-29.1 (-7.0)	-14.0 (-3.3)
$Cl^-(H_2S)$	-38.4 (-9.2)	-60.0 (-14.4)	-14.4	-39.6 (-9.5)	-21.7 (-5.2)	-8.7 (-2.1)
$Cl^-(HCl)$	-102.9 (-24.6)	-171.5 (-41.0)	-41.8	-105.4 (-25.2)	-55.2 (-13.2)	-21.5 (-5.1)
$Cl^-(HF)$	-84.8 (-20.3)	-106.8 (-25.5)	-25.5	-76.5 (-18.3)	-48.0 (-11.5)	-23.0 (-5.5)
$Cl^-(NH_3)$	-27.3 (-6.5)	-37.6 (-9.0)	-9.0	-25.9 (-6.2)	-15.6 (-3.7)	-7.2 (-1.7)
$F^-(H_2O)$	-108.8 (-26.0)	-136.0 (-32.5)	-32.5	-95.52 (-22.8)	-58.0 (-13.9)	-25.7 (-6.2)
$F^-(HF)$	-210.6 (-50.3)	-274.8 (-65.7)	-65.7	-187.2 (-44.8)	-111.8 (-26.7)	-49.0 (-11.7)
$F^-(NH_3)$	-58.3 (-13.9)	-74.4 (-17.8)	-17.8	-51.0 (-12.2)	-29.9 (-7.1)	-12.5 (-3.0)
$(HCO_2)^-(CH_3NH_2)$	-41.0 (-9.8)	-53.6 (-12.8)	-12.8	-35.9 (-8.6)	-21.6 (-5.2)	-10.2 (-2.4)
$(HCO_2)^-(CH_3OH)$	-59.9 (-14.3)	-86.4 (-20.7)	-20.75	-54.1 (-12.9)	-30.6 (-7.3)	-13.5 (-3.2)
$(HCO_2)^-(H_2O)$	-73.0 (-17.5)	-88.0 (-21.0)	-21.0	-61.0 (-14.7)	-38.4 (-9.2)	-19.0 (-4.5)
$(HCO_2)^-(HF)$	-89.7 (-21.4)	-131.4 (-31.4)	-31.4	-79.0 (-18.9)	-43.9 (-10.5)	-19.2 (-4.6)
$N_3^-(H_2O)$	-27.1 (-6.5)	-59.1 (-14.1)	-14.1	-35.0 (-8.4)	-19.3 (-4.6)	-8.6 (-2.0)
$N_3^-(H_2S)$	-10.2 (-2.4)	-46.3 (-11.1)	-11.1	-25.5 (-6.1)	-12.5 (-3.0)	-4.9(-1.2)
$N_3^-(HF)$	-70.4 (-16.8)	-108.8 (-26.0)	-26.0	-66.9 (-16.0)	-37.3 (-8.9)	-16.4 (-3.9)
$N_3^-(NH_3)$	-11.4 (-2.7)	-33.3 (-8.0)	-8.0	-19.0 (-4.6)	-10.1 (-2.4)	-4.4 (-1.0)
$(OH)^-(H_2O)$	-119.7 (-28.6)	-156.0 (-37.3)	-37.3	-105.8 (-25.3)	-61.5 (-14.7)	-25.3 (-6.1)
$(OH)^-(NH_3)$	-54.0 (-12.9)	-71.3 (-17.0)	-17.0	-47.2 (-11.3)	-26.4 (-6.3)	-10.4 (-2.5)
$(SH)^-(H_2O)$	-51.2 (-12.3)	-65.8 (-15.7)	-15.7	-46.9 (-11.2)	-28.8 (-6.9)	-13.3 (-3.2)
$(SH)^-(HF)$	-88.0 (-21.0)	-109.8 (-26.2)	-26.0	-78.1 (-18.7)	-47.9 (-11.5)	-21.8 (-5.2)
$(SH)^-(NH_3)$	-25.4 (-6.1)	-36.1 (-8.6)	-8.6	-24.52 (-5.9)	-14.1 (-3.4)	-6.0 (-1.4)

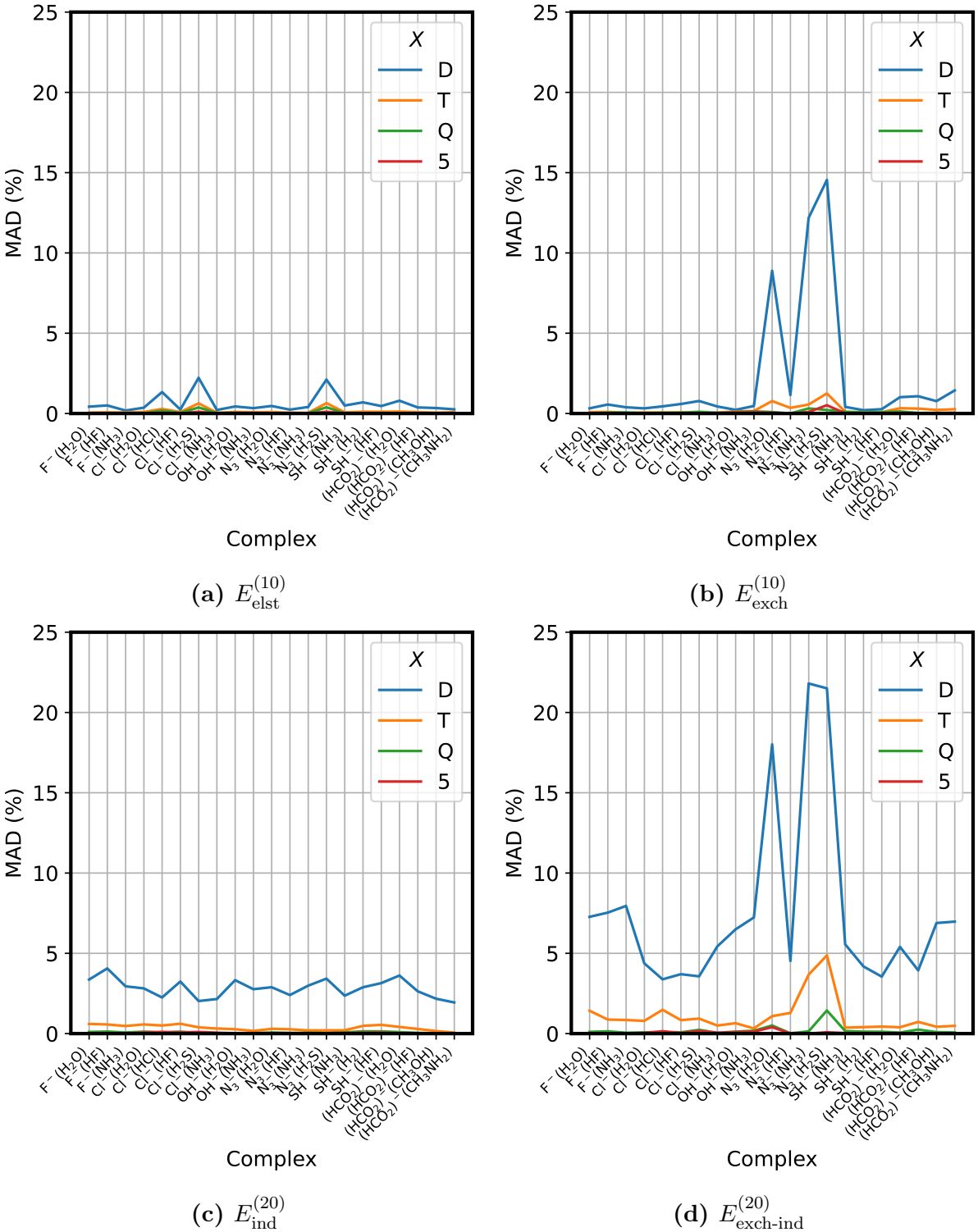


(e) $E^{(20)}_{\text{disp}}$ **(f)** $E^{(20)}_{\text{exch-disp}}$



(g) $\delta E^{(2)}_{\text{HF}}$

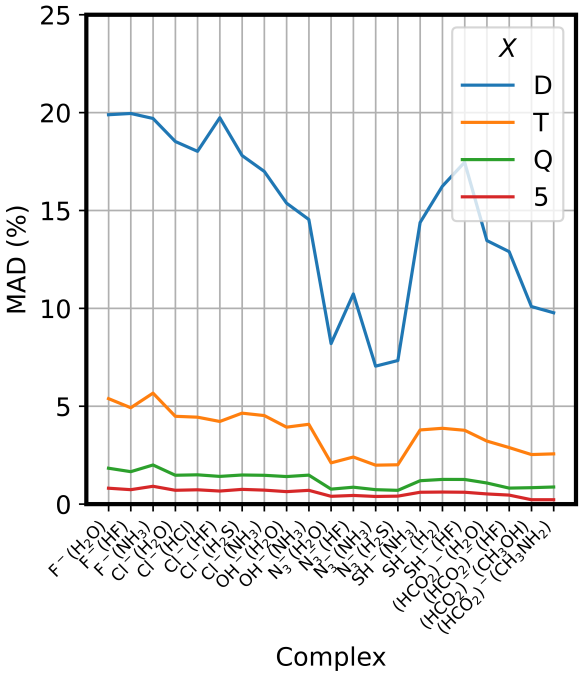
Fig. 6.2. Mean relative energy (MRE) for each of the SAPT0 terms, at each basis set, relative to the aDZ basis set, averaged over all geometries at each VSF in the AHB21×5 dataset (cont.).



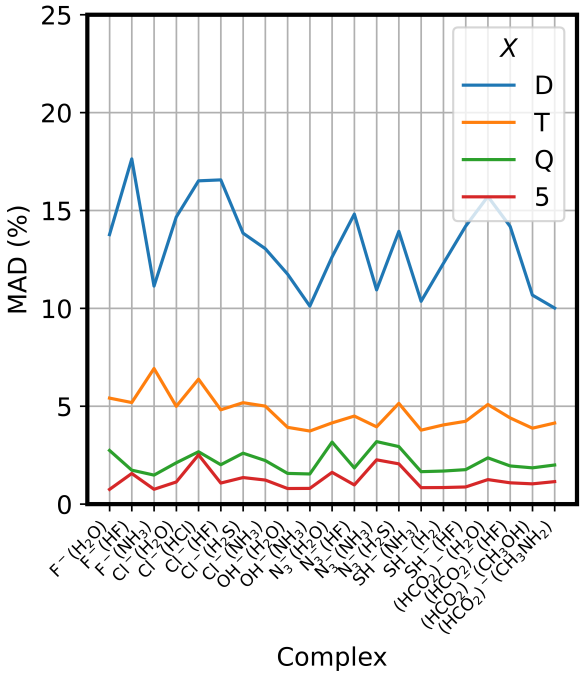
(a) $E^{(10)}_{\text{elst}}$ **(b)** $E^{(10)}_{\text{exch}}$

(c) $E^{(20)}_{\text{ind}}$ **(d)** $E^{(20)}_{\text{exch-ind}}$

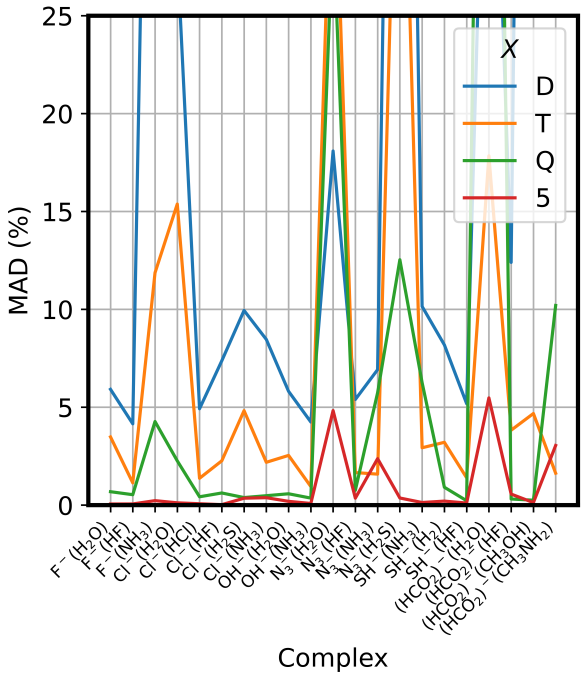
Fig. 6.2. Percentage mean absolute deviation (MAD %) for each complex over all distances for each SAPT0 energy component. The absolute deviation is determined for each AHB21×5 complex as the percentage difference between the energy at the $[X+1]$Z and aXZ basis sets.



(e) $E^{(20)}_{\text{disp}}$



(f) $E^{(20)}_{\text{exch-disp}}$



(g) $\delta E^{(2)}_{\text{HF}}$

Fig. 6.2. Percentage mean absolute deviation (MAD %) for each complex over all distances for each SAPT0 energy component. The absolute deviation is determined for each AHB21×5 complex as the percentage difference between the energy at the $[X+1]$Z and aXZ basis sets (cont.).

Figure 6.2 shows the mean AD (MAD) for each complex in the dataset — the ADs averaged over all VSFs for a complex — showing that the ADs of the non-correlated terms ($E^{(10)}_{\text{elst}}$, $E^{(10)}_{\text{exch}}$, $E^{(20)}_{\text{ind}}$, $E^{(20)}_{\text{exch-ind,resp}}$) approach zero with an increase in basis set, whereas the correlated terms do not. The $E^{(10)}_{\text{elst}}$ and $E^{(10)}_{\text{exch}}$ and terms can be considered sufficiently converged with the aTZ basis set, and the $E^{(20)}_{\text{ind}}$ and $E^{(20)}_{\text{exch-ind,resp}}$ terms with the aQZ basis set. In addition, since the δ_{HF} term is typically (as done so here) grouped with induction terms, which can be considered converged with the aQZ basis set, this term should be calculated similarly, although the aTZ level would be sufficient. Last, if only electrostatics and/or exchange components are needed, computation using the aTZ basis set would be sufficient. On the other hand, Figure 6.2 shows that the correlated terms do not meet the same level of convergence, compared to the non-correlated terms, using the aQZ basis set, or any of the basis sets used here, confirming the need for extrapolation of these terms. Therefore, an estimate of the CBS limit of $E^{(20)}_{\text{disp}}$ and $E^{(20)}_{\text{exch-disp}}$ required extrapolation.

The two-point extrapolation scheme, given in Equation 6.1, can be used to estimate the CBS limit of each non-correlated term using the values calculated with consecutive basis sets with cardinal numbers X and $X+1$. While this scheme was originally proposed with a value of $p = 3$,[323,327] Řezáč and Hobza[330] showed that optimising the value of p can improve the estimated CBS limit of the dispersion energy. Therefore, the scheme was optimised for each dispersion term through the determination of the value of the exponent p for each possible set of consecutive basis sets, and assessing the AARDs for each combination, relative to the theoretical CBS values, E_{CBS}. Equation 6.2 was used to find the theoretical CBS limit of $E^{(20)}_{\text{disp}}$ and $E^{(20)}_{\text{exch-disp}}$ for all 105 geometries in the AHB21×5 dataset, producing E_{CBS} for each. The values of $E^{(20)}_{\text{disp,CBS}}$ and $E^{(20)}_{\text{exch-disp,CBS}}$ were then used to optimise the value of p in Equation 6.1 for each term. All possible pairs of consecutive basis sets were considered, for each complex at each distance, for the $X/X+1$ extrapolations. Values of p ranging from 0.1 to 5 in increments of 0.01 were tested and $E^{\text{est.}}_{\text{CBS}}$ assessed, relative to the corresponding E_{CBS}, taking the error as the difference between the two.

Column A of Table 6.4 presents the average optimal p value for each $X/X+1$ two-point extrapolation, with the standard deviation (SD) over all geometries of all complexes to provide a measure of how consistent the p value is over all complexes at all distances, for which the distance-dependent errors are shown in Table 6.3. The average p values for all extrapolations of $E_{\text{disp}}^{(20)}$ are similar, varying by at most 0.3, but the error drops as the basis sets used to get larger, even though the SD, over all distance-dependent geometries, increases. The T/Q extrapolation does produce a notable improvement in AARD over that of the D/T extrapolation, and the improvement produced by extrapolation of larger basis sets (Q/5 and 5/6 extrapolations) does not offer significantly lower AARDs than that of the T/Q extrapolation.

Considering the extrapolation of $E_{\text{exch-disp}}^{(20)}$, the deviation in p values and AARDs are much larger than for $E_{\text{disp}}^{(20)}$. It is expected that the value of p should vary, given that each complex will have different convergence behaviour, but as to why the SD for $E_{\text{exch-disp}}^{(20)}$ is almost twice as large as that of $E_{\text{disp}}^{(20)}$ required further analysis. Through the determination of the theoretical E_{CBS} for the Cl^-(HCl) complex at VSF $= 0.9$, using Equation 6.2, the basis-set dependency was best fit to an exponential function with a "theoretical" p value of ~ 35. The theoretical p values were also assessed for all other geometries at all VSFs where the values ranged between ∼1–3. The implication here is that the convergence behaviour of Cl^-(HCl) at VSF $= 0.9$ is significantly different to the other complexes at all separations, and cannot be described using the determined average p value. Furthermore, as this complex requires a much larger estimate of p than the other geometries when fitting to Equation 6.1, this geometry increases the average p value as well as the SD. Lao et al.[320] highlighted that R^-(HR) systems (Cl^-(HCl), F^-(HF), $OH^-(H_2O)$) essentially share the proton between the two R^- atoms at equilibrium and at closer distances (VSF≤ 1) the "intermonomer" distances fall well within the range of a covalent bond and so that the fragments are no longer non-covalently bound and describing the interaction with perturbation theory is unphysical, and the SAPT formalism fails. The percentage error for

the extrapolations of each dimer at VSF $= 0.9$ is presented in Table 6.2. The percentage

Table 6.2: Percentage error, relative to CBS limits determined with Equation 6.2, for each complex, for the indicated extrapolations, using the optimal p value given in parenthesis, of $E^{(20)}_{\text{exch-disp}}$ at VSF $= 0.9$

Complex	**D/T (1.72)**	**T/Q (1.83)**	**Q/5 (1.67)**	**5/6 (1.63)**
$F^-(H_2O)$	12.81	13.42	12.32	12.01
$F^-(HF)$	30.45	30.90	29.70	29.20
$F^-(NH_3)$	6.51	7.51	6.25	6.11
$Cl^-(H_2O)$	1.65	2.37	1.05	0.85
$Cl^-(HCl)$	120.58	123.25	121.06	121.19
$Cl^-(HF)$	4.20	4.89	3.35	3.24
$Cl^-(H_2S)$	2.64	3.41	2.20	2.07
$Cl^-(NH_3)$	0.55	0.96	0.16	0.44
$OH^-(H_2O)$	25.00	24.85	24.06	23.77
$OH^-(NH_3)$	7.06	7.57	6.57	6.46
$N_3^-(H_2O)$	2.49	2.92	1.89	1.71
$N_3^-(HF)$	8.59	8.53	7.51	7.16
$N_3^-(NH_3)$	0.21	1.08	0.03	0.17
$N_3^-(H_2S)$	4.50	4.70	3.78	3.52
$SH^-(NH_3)$	1.21	1.26	0.37	0.06
$SH^-(H_2O)$	2.67	2.91	1.88	1.63
$SH^-(HF)$	5.48	5.89	4.66	4.46
$(HCO_2)^-(H_2O)$	3.11	4.04	3.06	2.98
$(HCO_2)^-(HF)$	11.20	11.47	9.09	9.62
$(HCO_2)^-(MeOH)$	4.54	5.10	4.19	2.73
$(HCO_2)^-(MeNH_2)$	0.97	1.89	1.11	0.05

errors are ∼31 % for $F^-(HF)$, ∼25 % for $OH^-(H_2O)$ and ∼120 % for $Cl^-(HCl)$, with the next highest error being approximately 20 % lower than that of $OH^-(H_2O)$. This confirms that extrapolation of $E^{(20)}_{\text{exch-disp}}$ at closer distances using the optimised p value therefore results in unphysical (or unrealistic) estimates of $E^{(20)}_{\text{exch-disp}}$ at the CBS limit. If these points are removed from the assessment of the error, the AARD at VSF $= 0.9$ is ∼2 % for all $X/X+1$ extrapolations considered.

It is also seen in Table 6.3 that the AARDs for the VSF $= 2$ complexes are ∼5–8 times greater than the VSF $= 1.5$ complexes, and ∼10–20 times greater than the VSF ≤ 1.25 complexes. At longer intermonomer distances the exchange effects approach zero, and by VSF $= 2$ this term is virtually zero, and a significant basis-set dependence cannot be

Table 6.3: AARD % at each vector scaling factor (VSF) for each two-point extrapolation considered. Values of p used for each extrapolation are those of column (A) in Table 6.4.

Term	**VSF**	**D/T**	**T/Q**	**Q/5**	**5/6**
$E^{(20)}_{\mathbf{disp}}$	0.9	1.41	1.28	1.22	1.16
	1.0	1.32	1.06	0.92	0.84
	1.25	1.09	0.63	0.38	0.37
	1.5	1.23	0.43	0.31	0.29
	2.0	2.76	1.18	1.21	1.05
$E^{(20)}_{\mathbf{exch\text{-}disp}}$	0.9	13.49	13.68	12.29	12.03
	1.0	3.15	3.10	1.94	1.82
	1.25	2.29	1.49	0.55	0.42
	1.5	5.71	3.84	2.98	3.07
	2.0	27.23	26.26	25.02	24.71

extracted. Consequently, fitting this data to an exponential function is not realistic and the CBS limit cannot be reasonably estimated using Equation 6.1 for these complexes. Therefore, the AARDs for $E^{(20)}_{\text{exch-disp}}$ at VSF $= 2$ are not truly representative of the quality of the estimation schemes used. Following this, the average values of p were determined excluding the VSF $= 2$ and R^{-}(HR) VSF $= 0.9$ points, and are presented in column B of Table 6.4 with the corresponding errors. The average value of p does not change significantly, but the SD drops by $\sim$ 50 %, and the AARDs are much lower. Once again, it is seen that the error drops with an increase in the basis sets used for extrapolation; however, the decrease in AARD is largest between the D/T and T/Q extrapolations, after which the further decrease in error is largely insignificant.

While the distance behaviour of the convergence is important for the intended purpose of FF development and parameterization, SAPT components at the CBS limit might well be needed in applications where only equilibrium geometries are analyzed. Consequently, column C in Table 6.4 shows the results when using only the equilibrium (VSF $= 1$) geometries. As can be expected, AARDs are much lower over this smaller set but the optimized values of p remain within the same range as fitting over the full data set for which the values of p given in column C of Table 6.4. Using the scheme with these values of p would

be more suitable for the analysis of only equilibrium complexes, and the performance of these schemes for such an application is indicated by the consistently low AARDs.

Table 6.4: Optimised values of p for the $X/X+1$ (X = D, T, Q, 5, 6) extrapolations of $E^{(20)}_{\text{disp}}$ and $E^{(20)}_{\text{exch-disp}}$ determined using: (A) all AHB21×5 complexes; (B) all AHB21×5 complexes excluding the VSF = 2 and R⁻(HR) VSF = 0.9 complexes and (C) only VSF = 1 complexes.

		A		**B**		**C**	
Term	Extrapolation	p	AARD	p	AARD	p	AARD
$E^{(20)}_{\mathbf{disp}}$	D/T	2.65 ±0.39	3.46	2.75 ±0.33	2.48	2.71 ±0.16	1.89
	T/Q	2.77 ±0.36	2.79	2.88 ±0.25	1.95	2.86 ±0.11	1.49
	Q/5	2.50 ±0.39	2.70	2.61 ±0.33	1.91	2.59 ±0.23	1.51
	5/6	2.64 ±0.69	2.65	2.71 ±0.63	1.88	2.57 ±0.26	1.46
$E^{(20)}_{\mathbf{exch\text{-}disp}}$	D/T	1.72 ±0.60	7.67	1.86 ±0.35	2.04	1.90 ±0.12	0.66
	T/Q	1.83 ±0.68	6.64	2.01 ±0.35	1.26	2.09 ±0.13	0.32
	Q/5	1.67 ±0.64	5.96	1.80 ±0.33	0.98	1.85 ±0.14	0.25
	5 6	1.63 ±0.74	5.84	1.71 ±0.43	0.95	1.67 ±0.12	0.17

As stated, the most notable improvement in error is between the D/T and T/Q extrapolations for both $E^{(20)}_{\mathbf{disp}}$ and $E^{(20)}_{\mathbf{exch\text{-}disp}}$. As such, the latter extrapolation is considered sufficient and will be used exclusively in the remainder of this chapter. The optimal values of p used are those determined without the VSF = 2 and R⁻(XR) VSF = 0.9 complexes, for which the p values used for extrapolation of $E^{(20)}_{\mathbf{disp}}$ and $E^{(20)}_{\mathbf{exch\text{-}disp}}$ are 2.88 and 2.01, respectively. It is worth noting that the errors associated with the D/T extrapolation are still reasonable and if the computations for the T/Q extrapolation were not feasible, one could use the D/T extrapolation as shown here, without a major increase in error.

6.4 Effect of the single-exchange approximation

Up to this point, only the effect of the basis sets applied in the extrapolation was considered when assessing the accuracy. As a further assessment, the accuracy of the SAPT implementation was considered, giving attention to the effect of the single-exchange approximation (SEA).[168] The second-order exchange terms ($E^{(20)}_{\text{exch-ind,resp}}$ and $E^{(20)}_{\text{exch-disp}}$) are traditionally calculated with the SEA; however, the closed form solution has been extended to these

terms.[236,237] Lao and Herbert[343] showed that the SEA leads to significant errors at intermolecular distances shorter than that of the equilibrium geometry for ionic species, and while the SEA provides reasonable accuracy for neutral systems, exchange terms for charged systems should be calculated *without* the SEA. Furthermore, Naseem-Khan et al.[353] recently highlighted that the SEA results in a large underestimation of exchange-induction terms for strongly bound systems, and the effect becomes more detrimental with the inclusion of third-order SAPT terms. The deviation of $E^{(20)}_{\text{exch-ind,resp}}$ and $E^{(20)}_{\text{exch-disp}}$ as a result of the SEA were assessed, and the errors are presented in Table 6.5, where, the MAD and maximum AD of the approximated (S^2), relative to the exact, non-approximated (S^∞) exchange terms, at each scale distance using the aQZ basis set are given. The complexes producing the largest deviations are those with the strongest overall interaction energies (see Table 6.1) indicating complexes with stronger interactions are more greatly affected by the SEA. That being, these complexes produce the largest absolute, as well as relative, deviations of the S^2 values from S^∞ equivalents. This supports the recommendation of Lao and Herbert[343] that the exchange terms of charged, strongly interacting systems should be calculated without the SEA. As expected, the largest deviations occur at the shorter intermonomer distance, as exchange effects are larger at shorter ranges. As the separation increases, the deviation due to the SEA approaches zero; however, the results presented here clearly emphasize that the SEA should not be used at short intermonomer distances.

While the SEA does have a significant effect on $E^{(20)}_{\text{exch-ind,resp}}$, the inclusion of the $\delta E^{(2)}_{\text{HF}}$ term in $E^{\text{SAPT0}}_{\text{ind}}$, as described in Section 2.3, corrects this error and the overall SAPT induction energy is unaffected by the use of the SEA. The SEA is only reflected in the $E^{(20)}_{\text{exch-disp}}$ term and therefore, using the closed-form (S^∞) solution rather than the SEA (S^2) solution is crucial for complexes where the dispersion energy makes a significant contribution to the total interaction energy. The final recommendation is thus that if individual components are required, the SEA should not be employed; however, if the contribution of $E^{\text{SAPT0}}_{\text{disp}}$

Table 6.5: Mean absolute deviation (MAD) and maximum absolute deviation (AD_{max}) between the S^2 and S^∞ values of $E^{(10)}_{exch}$, $E^{(20)}_{exch\text{-}ind,resp}$ and $E^{(20)}_{exch\text{-}disp}$ for the AHB21×5 dataset at each scaled distance. MAD and AD_{max} values are given in kJ/mol.

Component	**VSF**	**MAD**	**AD_{max}** (Value)
$E^{(10)}_{exch}$	0.9	21.4	Cl^-(HCl) (110)
	1	4.33	Cl^-(HCl) (25.7)
	1.25	8.78×10^{-2}	Cl^-(HCl) (4.97×10^{-1})
	1.5	2.21×10^{-3}	F^-(HF) (1.38×10^{-2})
	2	2.84×10^{-6}	$OH^-(H_2O)$ (2.26×10^{-5})
$E^{(20)}_{exch\text{-}ind,resp}$	0.9	26.1	Cl^-(HCl) (198)
	1	4.37	Cl^-(HCl) (34.2)
	1.25	7.58×10^{-2}	Cl^-(HCl) (4.68×10^{-1})
	1.5	2.22×10^{-3}	$OH^-(H_2O)$ (1.27×10^{-2})
	2	6.66×10^{-6}	$F^-(H_2O)$ (1.58×10^{-4})
$E^{(20)}_{exch\text{-}disp}$	0.9	1.81	Cl^-(HCl) (10.89)
	1	2.66×10^{-1}	F^-(HF) (1.40)
	1.25	5.67×10^{-3}	Cl^-(HCl) (4.89×10^{-2})
	1.5	4.76×10^{-4}	Cl^-(HCl) (3.67×10^{-3})
	2	6.17×10^{-5}	$F^-(NH_3)$ (4.91×10^{-4})

(the sum of $E^{(20)}_{disp}$ and $E^{(20)}_{exch\text{-}disp}$) is small, total interaction energies can be computed with the SEA, without resulting in a significant error.

Following the optimisation of the scheme in Section 6.3 and the considerations outlined above, the SAPT0/CBS protocol was implemented: The non-correlated terms ($E^{(10)}_{elst}$, $E^{(10)}_{exch}$, $E^{(20)}_{ind}$, $E^{(20)}_{exch\text{-}ind,resp}(S^\infty)$ and $\delta E^{(2)}_{HF}$) were obtained with the aQZ basis set and the correlated terms ($E^{(20)}_{disp}$ and $E^{(20)}_{exch\text{-}disp}(S^\infty)$ with a T/Q extrapolation, using optimised p values of 2.88 and 2.01, respectively. Finally, all exchange-containing terms are calculated without the SEA. The AARD of the overall interaction energy, given in Table 6.6, is 17.9 % (relative to the CCSD(T)/CBS interaction energies). A component-wise AARD cannot be assessed relative to the reference energies, and rather a measure of "completeness" is given. As the SAPT2+δMP2(CCD) method is the most complete (and obtainable) description of each component, this is used to determine the component-wise AARDs as follows: the AARDs of each non-correlated SAPT0/CBS component is given relative to the corresponding SAPT2+δMP2(CCD)/a6Z equivalent, and the AARD of the correlated

(dispersion) component is given relative to the theoretical SAPT2+δMP2(CCD)/CBS dispersion, determined by fitting the basis-set dependent SAPT2+δMP2(CCD) dispersion to Equation 6.2. These AARDs are presented in the top block of Table 6.6. The AARDs of E_{elst} and E_{ind} are consistently the lowest over all VSFs, showing only a slight increase with the increase in dimer separation. The intramonomer contribution to E_{elst} and E_{ind} (obtained at the SAPT2+δMP2(CCD)/a6Z level) are relatively small, accounting for at most 10 % of the components over all VSFs. Therefore, the SAPT0 terms provide most of the contribution to the E_{elst} and E_{ind} and obtaining these components with the SAPT0/CBS provides a largely complete description at all dimer separations. When considering E_{exch} and E_{disp}, the AARDs are much larger due to the higher-order terms accounting for $\sim$10–50 % of these components. These components also show a greater sensitivity to VSF, with the AARD of E_{exch} increasing significantly at longer distances. The SAPT0/CBS method does not account for the intramonomer contribution to E_{exch}, which only becomes significant (as a percentage of the total exchange energy) at greater separation where the intermonomer exchange falls away. Therefore, the exclusion of the higher-order exchange terms will produce a greater AARD at larger distances. The E_{disp} component shows the same trend such that the error increases at larger VSFs, indicating that the higher-order dispersion contributions have a greater significance at longer separations. This can be explained as the SAPT0/CBS E_{disp} does not include intramonomer effects, which, at long distance, are the dominating contributions to the dispersion energy, rather than the intermonomer contributions, and the deficit of intermonomer contributions will as such, be felt more heavily at larger VSFs. Therefore, the AARDs presented here can be considered to represent the basis set-independent error of the SAPT0/CBS approach, that is, if higher accuracy is required, intramonomer effects must be considered through higher-level SAPT components.

Table 6.6: AARD (%) over all geometries at each VSF obtained using the SAPT0/CBS and f_{2+}SAPT0/CBS approaches. AARDs of non-correlated components are given with respect to the SAPT2+δMP2(CCD)/a6Z equivalents, AARDs of correlated components are given relative to the SAPT2+δMP2(CCD)/CBS equivalents and AARDs of overall energies with respect to the CCSD(T)/CBS interaction energy. The p values used for extrapolation of $E^{(20)}_{\text{disp}}$ and $E^{(20)}_{\text{exch-disp}}$ are 2.88 and 2.01, respectively. The f_i scaling factors used to obtain the f_{2+}SAPT0/CBS components are those presented in Table 6.7.

Protocol	**VSF**	$E_{\mathbf{elst}}$	$E_{\mathbf{exch}}$	$E_{\mathbf{ind}}$	$E_{\mathbf{disp}}$	$E_{\mathbf{int}}$
SAPT0/CBS	0.9	2.46	6.23	5.48	7.80	58.2
	1	2.49	9.55	5.72	11.1	14.1
	1.25	3.00	17.5	6.53	15.4	4.73
	1.5	3.96	24.8	6.65	15.9	4.16
	2	4.88	34.9	6.32	14.6	5.26
	Overall	3.36	18.6	6.14	13.0	17.9
f_{2+}SAPT0/CBS	0.9	3.30	6.99	2.67	6.61	12.8
	1	2.90	5.05	2.80	7.75	6.40
	1.25	1.98	8.76	2.85	8.49	9.22
	1.5	2.51	15.0	2.69	10.1	10.6
	2	3.03	26.0	2.54	13.0	11.6
	Overall	2.74	12.4	2.71	8.88	6.22

6.5 Extension of the SAPT0/CBS protocol

To improve the accuracy of the overall interaction energy, higher-order SAPT components need to be included; however, doing so increases the computational resources and time required. In the work of Řezáč and Hobza,[330] E_{disp} (consisting of the DFT-SAPT computed E^2_{disp} and E^2_{disp}) was scaled to obtain the CBS limit. In a later study by Heßelmann and Korona,[331] the two DFT-SAPT dispersion components were scaled independently to their respective CBS limits. The goal of these two studies was to obtain scaling factors to account for the finite basis set effect, which has been addressed in Section 6.3. However, the scaling of SAPT terms to compensate for deficiencies, as done in these works, inspired the work carried out here. To compensate for the lack of higher-order terms, a scaling factor for each component (whereas previous work only focused on scaling of dispersion components[330,331]) was calculated to approximate higher-quality components from the computed

SAPT0/CBS components. The most complete components (as described in the section above) were used to determine scaling factors. Consequently, the f_{2+}SAPT0/CBS interaction energy is obtained by scaling each component i of the SAPT0/CBS interaction energy by a factor f_i to fit the equivalent SAPT2+(CCD)δMP2/a6Z and SAPT2+(CCD)δMP2/CBS terms for the non-correlated and correlated components, respectively. The scaling factors were determined by minimizing S_i for each component i over the full AHB21$\times$5 dataset,

$$S_i = \left| E_i^{\text{SAPT2+(CCD)}\delta\text{MP2/X}} - E_i^{\text{SAPT0/CBS}} \times f_{2+,i} \right|, \tag{6.4}$$

where $X =$ a6Z or CBS, depending on the component considered. The resulting $f_{2+,i}$ values are summarised in Table 6.7 and the AARDs resulting from the f_{2+}SAPT0/CBS protocol are presented in the bottom block of Table 6.6. The AARD of the total interaction energy across the AHB21$\times$5 dataset drops from 17.9 % with the SAPT0/CBS approach to 6.22 % with f_{2+}SAPT0/CBS. The overall AARD for each component also drops significantly, and the range of AARDs over the VSFs for each component is narrower, indicating that scaling basis-set corrected SAPT0 terms as such improves the quality and dimer separation behaviour of each component. The inclusion of higher-order effects makes the most significant improvement to E_{int} at shorter separations. Furthermore, E_{ind} and E_{disp} show the most marked improvement of all components at short distances, highlighting the significance of the contribution of these components to the overall energy. Last, while the f_{2+}SAPT0/CBS approach shows a large improvement for E_{exch}, the effect on the overall energy is minimal due to the small contribution of this component for long-range interactions.

6.6 Conclusion

The approaches developed here provide a way to obtain SAPT2+(CCD)δMP2 level energy components from SAPT0 computations that, when summed, produce interaction energy with an AARD of 6.22 % compared to CCSD(T)/CBS reference interaction energies. The

Table 6.7: Recommended scaling factors, $f_{2+,i}$, used to approximate SAPT2+(CCD)δMP2/a6Z non-correlated and SAPT2+(CCD)δMP2/CBS correlated components i from the equivalent SAPT0/CBS components.

Term	$f_{2+,i}$
E_{elst}	0.982
E_{exch}	1.141
E_{ind}	1.106
E_{disp}	1.12

recommendations made here are as follows: Non-correlated energy components should be obtained at the SAPT0/aQZ level of theory, and correlated energy components require extrapolation to the CBS limit. The dispersion terms should be extrapolated using the Helgaker Scheme shown in Equation 6.1, using the aTZ and aQZ basis sets. The values of p used in these extrapolations were optimised for $E^{(20)}_{\text{disp}}$ and $E^{(20)}_{\text{exch- disp}}$ basis set to have values of 2.88 and 2.01, respectively. The SDs calculated for these optimised values indicate that the value of p is species-dependent, and consideration should be given to this when performing such extrapolations for species of different nature to those tested here. The SAPT0/CBS protocol produces SAPT0 energy components void of basis set incompleteness errors; however, to improve the overall accuracy of the components, and therefore overall interaction energy, higher-order SAPT terms are required. The f_{2+}SAPT0/CBS approach showed that the quality of all SAPT0/CBS components can be improved by applying the scaling factors reported in Table 6.7 Last if individual energy components are required, the SEA should not be employed to provide a complete description of affected terms.

The accuracy is affected by the VSF, and the percentage error is greater at large distances, but the impact is greatest at closer distances where the absolute value of the interaction energy is greatest. This shows that the SAPT0 components are largely inaccurate at short distances and the error persists even with an increase in basis set size, and emphasises the need for higher-order SAPT components. This work shows that higher-order terms can be included for using the scaling factors determined. In addition, an assessment of the SEA shows that once again the effects are much greater at shorter distances, and

this approximation should not be employed for intermonomer distances shorter than that of the equilibrium geometry.

While the methodology developed here is shown to be effective for studying small complexes, the redistribution of electron density is limited due to the restricted size of the systems involved. This is a significant consideration, especially in the context of ILs, where cations and anions are large, delocalized, and capable of extensive electron density reorganization.[53,241] Such shifts in electron density are critical for accurately modelling the solvation effects and ion pairing in ILs. The use of small complexes may not fully capture these dynamics, which could affect the accuracy of the model when applied to ILs.

importparts/part7b.tex

Chapter 7

Summary of conclusions and future work

This thesis presents a comprehensive study on the application and development of the CL&Pol force field (FF) for Molecular Dynamics (MD) simulation of pure ionic liquids (ILs) and their mixtures with alkali-metal salts, forming IL-based electrolytes. The CL&Pol FF represents a significant advancement in the accurate simulation of complex molecular systems by incorporating explicit polarization effects through the use of Drude oscillators. The focus of this research spanned from the validation of the FF against experimental data to the extension of the developed calculation methodology for various IL-based systems, and finally the refinement of the description of van der Waals interactions within these models.

In Chapter 4, the CL&Pol FF was applied to a selected set of ILs composed of imidazolium and pyrrolidinium-based cations paired with imide-type anions. Validation was performed by comparing simulated properties against experimental data. A significant outcome of this study was the ability to quantify the reliability of the calculated properties with the 95 % confidence interval, achieved with the use of replicate trajectories. Being able to quantify the uncertainty of these properties allowed for a more reliable assessment

of the accuracy with respect to experimentally measured values. The simulation and calculation methodology used here provided a robust foundation for subsequent applications.

Chapter 5 extended the methodology employed in Chapter 4 to IL-based electrolytes consisting of the N-butyl-N-methylpyrrolidinium (fluorosulfonyl)(trifluorosulfonyl)imide and a 0.2 mole fraction of alkali metal cation salts. MD simulations of these were carried out and structural and transport properties were computed using the established protocols. The findings of this chapter emphasised the importance of explicit polarisation for the simulation of these systems, with a significant difference in structuring around the metal cations as a result of FF used. Concerning transport properties, the calculation of viscosity and ionic conductivity highlighted the significant role of polarisation in describing the mobility in IL-based electrolyte systems. Furthermore, this chapter highlighted the importance of considering the adjustment of metal cation parameters using the k_{ij} methodology, showing that the treatment of the metal cation parameters required reconsideration. Thus, fueling the work carried out in Chapters 6 and 7.

In Chapter 6, the novel f_{2+}SAPT0/CBS methodology was developed to compute highly accurate Symmetry-Adapted Perturbation Theory (SAPT) energy components devoid of basis-set incompleteness. Here, electrostatic, induction, exchange and dispersion, energies were obtained at a level equivalent to SAPT2+(CCD)δMP2 from SAPT0 computations extrapolated to the complete basis set. The approach emphasizes generating physically meaningful parameters through robust, cost-effective energy component computation. The energy components were treated independently, summing to give an AARD of 6.22 % relative to the "gold standard" CCSD(T)/CBS interaction energies, indicating the overall reliability of the energy components computed. These energy components were instrumental in parameterizing an exchange-dispersion potential compatible with the CL&Pol Drude framework.

Chapter 7 then used the f_{2+}SAPT0/CBS methodology developed in Chapter 6 to obtain exchange and dispersion energy components, for a set of homodimers and heterodimers

containing alkali metal cations. The SAPT components were then used to parameterize a dispersion-exchange potential. The buffered Lennard-Jones potential proved most accurate in replicating the SAPT dispersion and exchange energy profiles, parameterized using homodimer systems, referred to as the bLJ (HOM) potential. Initial applications were focused on metal cations, and the parameterization was extended to include simple IL components.

Overall, this thesis was successful in providing several contributions towards the simulation of novel IL-based electrolyte systems using the CL&Pol Drude polarisable FF. This included refining the simulation protocol and calculation methodology, as well as developing a framework in which Cl&Pol compatible parameters can be obtained. Future work will be aimed at the bLJ (HOM) potential for MD simulation of the systems studied here, as well as more realistic IL-based electrolyte systems. This will allow for an assessment of the accuracy produced by these parameters, as well as a measure of the transferability. Additionally, further development is envisioned for the f_{2+}SAPT0/CBS method. This involves extending the dataset to incorporate a larger variety of dimer systems, as well as including higher-order SAPT computations.

References

1. D. R. MacFarlane, M. Kar and J. M. Pringle, *Fundamentals of Ionic Liquids*, Wiley-VCH, Weinheim, 2017.

2. P. Wasserscheid and W. Keim, *Angew. Chem., Int. Ed.*, 2000, **39**, 3773–3789.

3. S. Chowdhury, R. S. Mohan and J. L. Scott, *Tetrahedron*, 2007, **63**, 2363–2389.

4. K. N. Marsh, J. A. Boxall and R. Lichtenthaler, *Fluid Ph. Equilibria*, 2004, **219**, 93–98.

5. T. Welton, *Chem. Rev.*, 1999, **99**, 2071–2083.

6. U. P. R. M. Preiss, J. M. Slattery and I. Krossing, *Ind. Eng. Chem. Res.*, 2009, **48**, 2290–2296.

7. J. O. Valderrama, A. Toro and R. E. Rojas, *J. Chem. Thermodyn.*, 2011, **43**, 1068–1073.

8. J. O. Valderrama, G. Martinez and C. A. Faúndez, *Int. J. Thermophys.*, 2011, **32**, 942–956.

9. P. Walden, *Bull. Acad. Imper. Sci.*, 1914, **1800**, 405–422.

10. F. H. Hurley and T. P. Wier Jr., *J. Electrochem. Soc.*, 1951, **98**, 207–212.

11. W. T. Ford, R. J. Hauri and S. G. Smith, *J. Am. Chem. Soc.*, 1974, **96**, 4316–4318.

12. J. C. Nardi, C. L. Hussey and L. A. King, *US Pat.*, 4 122 245, 1978.

13. J. Robinson and R. A. Osteryoung, *J. Am. Chem. Soc.*, 1979, **101**, 323–327.

14. J. S. Wilkes, J. A. Levisky, R. A. Wilson and C. L. Hussey, *Inorg. Chem.*, 1982, **21**, 1263–1264.

15. E. I. Cooper and C. A. Angell, *Solid State Ion.*, 1983, **9**, 617–622.

16. M. G. Freire, C. M. S. S. Neves, I. M. Marrucho, J. A. P. Coutinho and A. M. Fernandes, *J. Phys. Chem. A*, 2010, **114**, 3744–3749.

17. J. S. Wilkes and M. J. Zaworotko, *J. Chem. Soc. Chem. Commun.*, 1982, **13**, 965–966.

18. J. P. Hallet and T. Welton, *Chem. Rev.*, 2011, **111**, 3508–3576.

19. A. George, A. Brandt, K. Tran, S. M. S. N. S. Zahari, D. Klein-Marcuschamer, N. Sun, N. Sathitsuksanoh, J. Shi, V. Stavila, R. Parthasarathi, S. Singh, B. M. Holmes, T. Welton, B. A. Simmons and J. P. Hallett, *Green Chem.*, 2015, **17**, 1728–1734.

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

21. K. Wang, H. Adidharma, M. Radosz, P. Wan, X. Xu, C. K. Russell, H. Tian, M. Fan and J. Yu, *Green Chem.*, 2017, **19**, 4469–4493.

22. X. Li, Y. Zheng and C. Y. Li, *Energy Storage Mater.*, 2020, **29**, 273–280.

23. K. K. Maniam and S. Paul, *Appl. Sci.*, 2020, **10**, year.

24. S. Rahali, R. Zarrougui, M. Marzouki and O. Ghodbane, *J. Electroanal. Chem*, 2020, **871**, 114289.

25. E. B. Molodkina, M. R. Ehrenburg, P. Broekmann and A. V. Rudnev, *J. Electroanal. Chem*, 2020, **860**, 113892.

26. T. C. Lourenço, L. G. Dias and J. L. F. Da Silva, *ACS Appl. Energy Mater.*, 2021, **4**, 4444–4458.

27. A. Tkacheva, J. Zhang, B. Sun, D. Zhou, G. Wang and A. M. McDonagh, *J. Phys. Chem. C*, 2020, **124**, 5087–5092.

28. M. Angell, G. Zhu, M.-C. Lin, Y. Rong and H. Dai, *Adv. Funct. Mater.*, 2020, **30**, 1901928.

29. V. M. Ortiz-Martínez, L. Gómez-Coma, G. Pérez, A. Ortiz and I. Ortiz, *Sep. Purif. Technol.*, 2020, **252**, 117436.

30. L. G. da Trindade, L. Zanchet, J. C. Souza, E. R. Leite, E. M. A. Martini and E. C. Pereira, *Ionics*, 2020, **26**, 5661–5672.

31. L. Zanchet, L. G. da Trindade, W. Bariviera, K. M. Nobre Borba, R. D. M. Santos, V. A. Paganin, C. P. de Oliveira, E. A. Ticianelli, E. M. A. Martini and M. O. de Souza, *J. Mater. Sci.*, 2020, **55**, 6928–6941.

32. A. Al-Othman, P. Nancarrow, M. Tawalbeh, A. Ka'ki, K. El-Ahwal, B. El Taher and M. Alkasrawi, *Int. J. Hydrog. Energy*, 2021, **46**, 6100–6109.

33. A. M. Sampaio, G. F. L. Pereira, M. Salanne and L. J. A. Siqueira, *Electrochim. Acta*, 2020, **364**, 137181.

34. G. Santos Junior, V. Fortunato, G. Silva, P. Ortega and R. Lavall, *Electrochim. Acta*, 2019, **325**, 134900.

35. H. Zhang, Y. Ling, Y. Peng, J. Zhang and S. Guan, *Inorg. Chem. Commun.*, 2020, **115**, 107856.

36. Z. Li, Q. Xu, L. Zhang, X. Wang, F. He, J. Cheng and H. Xie, *Sustain. Energy Fuels*, 2020, **4**, 3418–3427.

37. Q. Xu, X. Wang, J. Cheng, L. Zhang, F. He and H. Xie, *RSC Adv*, 2020, **10**, 36504–36513.

38. F. Poli, D. Momodu, G. E. Spina, A. Terella, B. K. Mutuma, M. L. Focarete, N. Manyala and F. Soavi, *Electrochim. Acta*, 2020, **338**, 135872.

39. I. López-Martin, E. Burello, P. N. Davey, K. R. Seddon and G. Rothenberg, *ChemPhysChem*, 2007, **8**, 690–695.

40. J. N. A. Canongia Lopes and A. A. H. Pádua, *J. Phys. Chem. B*, 2006, **110**, 3330–3335.

41. G. Kaur, H. Kumar and M. Singla, *J. Mol. Liq.*, 2022, **351**, 118556.

42. Q. Huang, T. C. Lourenco, L. T. Costa, Y. Zhang, E. J. Maginn and B. Gurkan, *J. Phys. Chem. B*, 2019, **123**, 516–527.

43. A. Bard, L. Faulkner and H. White, *Electrochemical Methods: Fundamentals and Applications*, Wiley, 2022.

44. H. Tokuda, S. Tsuzuki, M. A. B. H. Susan, K. Hayamizu and M. Watanabe, *J. Phys. Chem. B*, 2006, **110**, 19593–19600.

45. L. M. Diamante and T. Lan, *J. Food Process. Preserv.*, 2014, **2014**, 1–6.

46. M. Ngadi and L. Yu, *Can. Biosyst. Eng.*, 2004, **46**, 315–318.

47. P. Wasserscheid and T. Welton, *Ionic Liquids in Synthesis*, Wiley-VCH, Weinheim, 2008.

48. P. Stilbs, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1987, **19**, 1–45.

49. A. Noda, K. Hayamizu and M. Watanabe, *J. Phys. Chem. B*, 2001, **105**, 4603–4610.

50. H. Every, A. G. Bishop, M. Forsyth and D. R. MacFarlane, *Electrochim. Acta*, 2000, **45**, 1279–1284.

51. A. Wettstein, D. Diddens and A. Heuer, *Phys. Chem. Chem. Phys.*, 2022, **24**, 6072–6086.

52. O. Borodin, G. A. Giffin, A. Moretti, J. B. Haskins, J. W. Lawson, W. A. Henderson and S. Passerini, *J. Phys. Chem. C*, 2018, **122**, 20108–20121.

53. F. Philippi, K. Goloviznina, Z. Gong, S. Gehrke, B. Kirchner, A. A. H. Pádua and P. A. Hunt, *Phys. Chem. Chem. Phys.*, 2022, **24**, 3144–3162.

54. F. Castiglione, A. Famulari, G. Raos, S. V. Meille, A. Mele, G. B. Appetecchi and S. Passerini, *J. Phys. Chem. B*, 2014, **118**, 13679–13688.

55. P. Ray, A. Balducci and B. Kirchner, *J Phys Chem B*, 2018, **122**, 10535–10547.

56. R. Gupta, T. R. Kartha and B. S. Mallik, *ACS Omega*, 2019, **4**, 19556–19564.

57. M. Gouverneur, F. Schmidt and M. Schonhoff, *Phys Chem Chem Phys*, 2018, **20**, 7470–7478.

58. C. J. Solano, S. Jeremias, E. Paillard, D. Beljonne and R. Lazzaroni, *J. Chem. Phys.*, 2013, **139**, 034502.

59. P. Johansson, L. E. Fast, A. Matic, G. B. Appetecchi and S. Passerini, *J. Power Sources*, 2010, **195**, 2074–2076.

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

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

62. S. P. M. Ventura, F. A. e Silva, M. V. Quental, D. Mondal, M. G. Freire and J. A. P. Coutinho, *Chem. Rev.*, 2017, **117**, 6984–7052.

63. V. I. Parvulescu and C. Hardacre, *Chem. Rev.*, 2007, **107**, 2615–2665.

64. J. E. Bara, T. K. Carlisle, C. J. Gabriel, D. Camper, A. Finotello, D. L. Gin and R. D. Noble, *Ind. Eng. Chem. Res.*, 2009, **48**, 2739–2751.

65. G. A. Giffin, *J. Mater. Chem. A*, 2016, **4**, 13378–13389.

66. D. R. MacFarlane, N. Tachikawa, M. Forsyth, J. M. Pringle, P. C. Howlett, G. D. Elliott, J. H. Davis, M. Watanabe, P. Simon and C. A. Angell, *Energy Environ. Sci.*, 2014, **7**, 232–250.

67. X. Kang, X. Sun and B. Han, *Adv. Mater.*, 2016, **28**, 1011–1030.

68. S. H. Strauss, *Chem. Rev.*, 1993, **93**, 927–942.

69. A. B. Rupp and I. Krossing, *Acc. Chem. Res.*, 2015, **48**, 2537–2546.

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

71. T. Rüther, A. I. Bhatt, A. S. Best, K. R. Harris and A. F. Hollenkamp, *Batter. Supercaps*, 2020, **3**, 793–827.

72. D. R. MacFarlane, P. Meakin, J. Sun, N. Amini and M. Forsyth, *J. Phys. Chem. B*, 1999, **103**, 4164–4170.

73. G. B. Appetecchi, M. Montanino, D. Zane, M. Carewska, F. Alessandrini and S. Passerini, *Electrochim. Acta*, 2009, **54**, 1325–1332.

74. H. Sakaebe and H. Matsumoto, *Electrochem. commun.*, 2003, **5**, 594–598.

75. K. R. Harris, L. A. Woolf, M. Kanakubo and T. Rüther, *J. Chem. Eng. Data*, 2011, **56**, 4672–4685.

76. Y. Katayama, T. Morita, M. Yamagata and T. Miura, *Electrochem.*, 2003, **71**, 1033–1035.

77. H. Matsumoto, H. Kageyama and Y. Miyazaki, *Electrochem.*, 2003, **71**, 1058–1060.

78. P. C. Howlett, D. R. MacFarlane and A. F. Hollenkamp, *Electrochem. Solid-State Lett.*, 2004, **7**, A97.

79. H. Yoon, P. C. Howlett, A. S. Best, M. Forsyth and D. R. MacFarlane, *J. Electrochem. Soc.*, 2013, **160**, A1629.

80. C. Michot, M. Armand, M. Gautier and N. Ravet, U.S. Patent 174465, Apr. 2002.

81. H.-B. Han, S.-S. Zhou, D.-J. Zhang, S.-W. Feng, L.-F. Li, K. Liu, W.-F. Feng, J. Nie, H. Li, X.-J. Huang, M. Armand and Z.-B. Zhou, *J. Power Sources*, 2011, **196**, 3623–3632.

82. H. Matsumoto, H. Sakaebe, K. Tatsumi, M. Kikuta, E. Ishiko and M. Kono, *J. Power Sources*, 2006, **160**, 1308–1313.

83. H. Yoon, A. S. Best, M. Forsyth, D. R. MacFarlane and P. C. Howlett, *Phys. Chem. Chem. Phys.*, 2015, **17**, 4656–4663.

84. H. Matsumoto, N. Terasawa, T. Umecky, S. Tsuzuki, H. Sakaebe, K. Asaka and K. Tatsumi, *Chem. Lett.*, 2008, **37**, 1020–1021.

85. J. Reiter, S. Jeremias, E. Paillard, M. Winter and S. Passerini, *Phys. Chem. Chem. Phys.*, 2013, **15**, 2565–2571.

86. A. S. Shaplov, E. I. Lozinskaya, P. S. Vlasov, S. M. Morozova, D. Y. Antonov, P.-H. Aubert, M. Armand and Y. S. Vygodskii, *Electrochim. Acta*, 2015, **175**, 254–260.

87. H. Matsumoto, H. Kageyama and Y. Miyazaki, *ChemComm*, 2002, 1726–1727.

88. J.-P. Hoffknecht, M. Drews, X. He and E. Paillard, *Electrochim. Acta*, 2017, **250**, 25–34.

89. C. Maton, N. De Vos and C. V. Stevens, *Chem. Soc. Rev.*, 2013, **42**, 5963–5977.

90. A. Lewandowski and A. Świderska Mocek, *J. Power Sources*, 2009, **194**, 601–609.

91. T. Yamamoto, K. Matsumoto, R. Hagiwara and T. Nohira, *J. Phys. Chem. C*, 2017, **121**, 18450–18458.

92. R. Hagiwara, K. Matsumoto, J. Hwang and T. Nohira, *Chem Rec*, 2019, **19**, 758–770.

93. H. Yang, X.-F. Luo, K. Matsumoto, J.-K. Chang and R. Hagiwara, *J. Power Sources*, 2020, **470**, 228406.

94. T. Yamamoto and T. Nohira, *ChemComm*, 2020, **56**, 2538–2541.

95. T. Yamamoto, S. Nishijima and T. Nohira, *J. Phys. Chem. B*, 2020, **124**, 8380–8387.

96. J. Liu, J.-G. Zhang, Z. Yang, J. P. Lemmon, C. Imhoff, G. L. Graff, L. Li, J. Hu, C. Wang, J. Xiao, G. Xia, V. V. Viswanathan, S. Baskaran, V. Sprenkle, X. Li, Y. Shao and B. Schwenzer, *Adv. Funct. Mater.*, 2013, **23**, 929–946.

97. T. Placke, R. Kloepsch, S. Dühnen and M. Winter, *J. Solid State Electrochem.*, 2017, **21**, 1939–1964.

98. J. Xu, X. Cai, S. Cai, Y. Shao, C. Hu, S. Lu and S. Ding, *Energy Environ. Mater.*, 2023, e12450.

99. D. Larcher and J. M. Tarascon, *Nat. Chem.*, 2015, **7**, 19–29.

100. P. K. Nayak, L. Yang, W. Brehm and P. Adelhelm, *Angew. Chem. Int. Ed.*, 2018, **57**, 102–120.

101. L. Zhang, X. Li, M. Yang and W. Chen, *Energy Stor. Mater.*, 2021, **41**, 522–545.

102. A. Eftekhari, Z. Jian and X. Ji, *ACS Appl. Mater. Interfaces*, 2017, **9**, 4404–4419.

103. W. Zhang, Y. Liu and Z. Guo, *Sci. Adv.*, 2019, **5**, 7412.

104. N. Yabuuchi, K. Kubota, M. Dahbi and S. Komaba, *Chem. Rev.*, 2014, **114**, 11636–82.

105. K. Kubota, M. Dahbi, T. Hosaka, S. Kumakura and S. Komaba, *Chem. Rec.*, 2018, **18**, 459–479.

106. I. Hasa, S. Passerini and J. Hassoun, *J. Power Sources*, 2016, **303**, 203–207.

107. A. Fukunaga, T. Nohira, R. Hagiwara, K. Numata, E. Itani, S. Sakai and K. Nitta, *J. Appl. Electrochem.*, 2016, **46**, 487–496.

108. C.-Y. Chen, K. Matsumoto, T. Nohira and R. Hagiwara, *J. Electrochem. Soc*, 2015, **162**, A2093–A2098.

109. N. Wongittharom, C.-H. Wang, Y.-C. Wang, C.-H. Yang and J.-K. Chang, *ACS Appl. Mater. Interfaces*, 2014, **6**, 17564–17570.

110. H. Kim, J. C. Kim, M. Bianchini, D.-H. Seo, J. Rodriguez-Garcia and G. Ceder, *Adv. Energy Mater*, 2018, **8**, 1702384.

111. X. Wu, D. P. Leonard and X. Ji, *Chem. Mater.*, 2017, **29**, 5031–5042.

112. Y. Liu, Z. Tai, J. Zhang, W. K. Pang, Q. Zhang, H. Feng, K. Konstantinov, Z. Guo and H. K. Liu, *Nat. Commun*, 2018, **9**, 3645.

113. R. A. Adams, A. Varma and V. G. Pol, *J. Power Sources*, 2018, **375**, 131–137.

114. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **45**, 11225–11236.

115. S. V. Sambasivarao and O. Acevedo, *J. Chem. Theory Comput*, 2009, **5**, 1038–1050.

116. B. Doherty, X. Zhong, S. Gathiaka, B. Li and O. Acevedo, *J. Chem. Theory Comput*, 2017, **13**, 6131–6145.

117. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

118. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell Jr, *J. Comput. Chem.*, 2010, **31**, 671–690.

119. C. G. Hanke, S. L. Price and R. M. Lynden-Bell, *Mol. Phys.*, 2001, **99**, 801–809.

120. D. E. Williams and S. R. Cox, *Acta Crystallogr. B.*, 1984, **40**, 404–417.

121. L.-Y. Hsu and D. E. Williams, *Acta Crystallogr. A.*, 1980, **36**, 277–281.

122. D. E. Williams and D. J. Houpt, *Acta Crystallogr. B.*, 1986, **42**, 286–295.

123. T. I. Morrow and E. J. Maginn, *J. Phys. Chem. B*, 2002, **106**, 12807–12813.

124. Z. Liu, S. Huang and W. Wang, *J. Chem. Phys. B*, 2004, **108**, 12978–12989.

125. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc*, 1995, **117**, 5179–5197.

126. V. V. Chaban, I. V. Voroshylova and O. N. Kalugin, *Phys. Chem. Chem. Phys.*, 2011, **13**, 7910–7920.

127. J. N. Canongia Lopes, J. Deschamps and A. A. H. Pádua, *J. Phys. Chem. B*, 2004, **108**, 2038–2047.

128. J. N. A. Canongia Lopes and A. A. H. Pádua, *J. Phys. Chem. B*, 2004, **108**, 16893–16898.

129. J. N. A. Canongia Lopes, A. A. H. Pádua and K. Shimizu, *J. Phys. Chem. B*, 2008, **112**, 5039–5046.

130. B. L. Bhargava and S. Balasubramanian, *Chem. Phys.*, 2007, **127**, 114510.

131. W. Zhao, H. Eslami, W. L. Cavalcanti and F. Müller-Plathe, *Z. Phys. Chem.*, 2007, **221**, 1647–1662.

132. V. V. Chaban and I. V. Voroshylova, *J. Phys. Chem. B*, 2015, **119**, 6242–6249.

133. J. N. Canongia Lopes and A. A. H. Pádua, *Theor. Chem. Acc.*, 2012, **131**, 1–11.

134. L. Hakim, Y. Ishii, K. Matsumoto, R. Hagiwara, K. Ohara, Y. Umebayashi and N. Matubayasi, *J. Phys. Chem. B*, 2020, **124**, 7291–7305.

135. J. N. A. Canongia Lopes and A. A. H. Pádua, *J. Phys. Chem. B*, 2006, **110**, 7485–7489.

136. J. N. A. Canongia Lopes, M. F. Costa Gomes and A. A. H. Pádua, *J. Phys. Chem. B*, 2006, **110**, 16816–16818.

137. M. Deetlefs, C. Hardacre, M. Nieuwenhuyzen, A. A. H. Pádua, O. Sheppard and A. K. Soper, *J. Phys. Chem. B*, 2006, **110**, 12055–12061.

138. C. Cadena, Q. Zhao, R. Q. Snurr and E. J. Maginn, *J. Phys. Chem. B*, 2006, **110**, 2821–2832.

139. S. Tsuzuki, W. Shinoda, H. Saito, M. Mikami, H. Tokuda and M. Watanabe, *J. Phys. Chem. B*, 2009, **113**, 10641–10649.

140. J. Schmidt, C. Krekeler, F. Dommert, Y. Zhao, R. Berger, L. D. Site and C. Holm, *J. Phys. Chem. B*, 2010, **114**, 6150–6155.

141. S. Zahn, B. Kirchner and D. Mollenhauer, *ChemPhysChem*, 2016, **17**, 3354–3358.

142. Y. Zhang and E. J. Maginn, *J. Phys. Chem. B*, 2012, **116**, 10036–10048.

143. A. Mondal and S. Balasubramanian, *J. Phys. Chem. B*, 2014, **118**, 3409–3422.

144. F. Dommert, K. Wendler, R. Berger, L. Delle Site and C. Holm, *ChemPhysChem*, 2012, **13**, 1625–1637.

145. V. Chaban, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16055–16062.

146. C. Schröder, *Phys. Chem. Chem. Phys.*, 2012, **14**, 3089–3102.

147. T. G. A. Youngs and C. Hardacre, *ChemPhysChem*, 2008, **9**, 1548–1558.

148. K. Goloviznina, J. N. Canongia Lopes, M. Costa Gomes and A. A. H. Pádua, *J. Chem. Theory Comput.*, 2019, **15**, 5858–5871.

149. A. d. O. Cavalcante, M. C. C. Ribeiro and M. S. Skaf, *J. Chem. Phys*, 2014, **140**, 144108.

150. Y. Wu, Y. Li, N. Hu and M. Hong, *Phys. Chem. Chem. Phys*, 2014, **16**, 2674–2685.

151. T. Yan, C. J. Burnham, M. G. Del Pópolo and G. A. Voth, *J. Phys. Chem. B*, 2004, **108**, 11877–11881.

152. M. Salanne, L. J. A. Siqueira, A. P. Seitsonen, P. A. Madden and B. Kirchner, *Faraday Discussions*, 2012, **154**, 171–188.

153. P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.

154. P. Ren and J. W. Ponder, *J. Comput. Chem.*, 2002, **23**, 1497–1506.

155. J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, J. DiStasio, Robert A., M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.

156. P. Ren, C. Wu and J. W. Ponder, *J. Chem. Theory Comput.*, 2011, **7**, 3143–3161.

157. B. Thole, *Chem. Phys.*, 1981, **59**, 341–350.

158. E. A. Vázquez-Montelongo, J. E. Vazquez-Cervantes and G. A. Cisneros, *Int. J. Mol. Sci.*, 2020, **21**, 697.

159. E. A. Vázquez-Montelongo, G. A. Cisneros and H. M. Flores-Ruiz, *J. Mol. Liq.*, 2019, **296**, 111846.

160. H. Torabifard, L. Reed, M. T. Berry, J. E. Hein, E. Menke and G. A. Cisneros, *J. Chem. Phys*, 2017, **147**, 161731.

161. J. E. Vázquez-Cervantes and G. A. Cisneros, *J. Phys. Chem. B*, 2023, **127**, 5481–5493.

162. O. Borodin, *J. Phys. Chem. B*, 2009, **113**, year.

163. V. Lesch, Z. Li, D. Bedrov, O. Borodin and A. Heuer, *Phys. Chem. Chem. Phys.*, 2016, **18**, 382–92.

164. L. Suo, O. Borodin, T. Gao, M. Olguin, J. Ho, X. Fan, C. Luo, C. Wang and K. Xu, *Science*, 2015, **350**, 938–943.

165. F. Wang, O. Borodin, T. Gao, X. Fan, W. Sun, F. Han, A. Faraone, J. A. Dura, K. Xu and C. Wang, *Nat. Mater.*, 2018, **17**, 543–549.

166. C. Schröder and O. Steinhauser, *J Chem Phys*, 2010, **133**, 154511.

167. J. G. McDaniel and J. R. Schmidt, *J. Phys. Chem. A*, 2013, **117**, 2053–66.

168. E. Clementi, C. Cagliari and S. Université L. Pasteur, *Methods and Techniques in Computational Chemistry: METECC-94 Vol. A, B Og C*, STEF, 1993.

169. J. G. McDaniel, E. Choi, C. Y. Son, J. R. Schmidt and A. Yethiraj, *J. Phys. Chem. B*, 2016, **120**, 7024–36.

170. J. G. McDaniel, *J. Phys. Chem. B*, 2018, **122**, 4345–4355.

171. F. Philippi, D. Rauber, O. Palumbo, K. Goloviznina, J. McDaniel, D. Pugh, S. Suarez, C. C. Fraenza, A. Pádua, C. W. M. Kay and T. Welton, *Chem. Sci.*, 2022, **13**, 9176–9190.

172. Z. Zhang, E. Zofchak, J. Krajniak and V. Ganesan, *J. Phys. Chem. B*, 2022, **126**, 2583–2592.

173. N. Wang and E. J. Maginn, *J. Phys. Chem. B*, 2024, **128**, 871–881.

174. K. Goloviznina, Z. Gong, M. F. Costa Gomes and A. A. H. Pádua, *J. Chem. Theory Comput.*, 2021, **17**, 1606–1617.

175. K. Goloviznina, Z. Gong and A. A. H. Pádua, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2021, e1572.

176. A. Massaro, J. Avila, K. Goloviznina, I. Rivalta, C. Gerbaldi, M. Pavone, M. F. Costa Gomes and A. A. H. Pádua, *Phys. Chem. Chem. Phys.*, 2020, **22**, 20114–20122.

177. H. Yu, T. W. Whitfield, E. Harder, G. Lamoureux, I. Vorobyov, V. M. Anisimov, A. D. J. MacKerell and B. Roux, *J. Chem. Theory Comput.*, 2010, **6**, 774–786.

178. P. Kubisiak, P. Wróbel and A. Eilmes, *J. Phys. Chem. B*, 2019, **124**, 413–421.

179. F. Jensen, *Introduction to Computational Chemistry*, John Wiley and Sons Ltd, Sussex, 2006.

180. E. J. Maginn, R. A. Messerly, D. J. Carlson, D. R. Roe and J. R. Elliott, *Living J. Comp. Mol. Sci.*, 2019, **1**, 6324.

181. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, 2017.

182. A. Leach, *Molecular Modelling: Principles and Applications*, Prentice Hall, Essex, 2nd edn, 2001.

183. S. W. Rick and S. J. Stuart, in *Potentials and Algorithms for Incorporating Polarizability in Computer Simulations*, 2002, pp. 89–146.

184. W. C. Swope, H. C. Andersen, P. H. Berens and K. R. Wilson, *J. Chem. Phys*, 1982, **76**, 637–649.

185. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.

186. A. J. Stone, *The Theory of Intermolecular Forces*, Oxford University Press, Oxford, UK, 2nd edn, 2013.

187. M. Born and J. E. Mayer, *Z. Phys.*, 1932, **75**, 1–18.

188. J. N. Israelachvili, *Intermolecular and Surface Forces*, Academic Press, San Diego, US, 3rd edn, 2011.

189. P. P. Ewald, *Ann. Phys.*, 1921, **369**, 253–287.

190. V. Natoli and D. M. Ceperley, *J. Comput. Phys.*, 1995, **117**, 171–178.

191. Z.-M. Chen, T. Çağin and W. A. Goddard Iii, *J. Comput. Chem.*, 1997, **18**, 1365–1370.

192. S. W. de Leeuw, J. W. Perram and E. R. Smith, *Proc. R. Soc. A: Math.*, 1980, **373**, 27–56.

193. A. K. Rappe and I. Goddard, William A., *J. Phys. Chem.*, 1991, **95**, 3358–3363.

194. S. W. Rick, S. J. Stuart and B. J. Berne, *Chem. Phys.*, 1994, **101**, 6141–6156.

195. B. A. Bauer and S. Patel, *Theor. Chem. Acc.*, 2012, **131**, 1153.

196. P. Drude, C. Mann and R. Millikan, *The Theory of Optics*, Longmans, Green, and Company, 1901.

197. G. Lamoureux and B. Roux, *J. Chem. Phys.*, 2003, **119**, 3025–3039.

198. G. Lamoureux, J. MacKerell, Alexander D. and B. Roux, *J. Phys. Chem.*, 2003, **119**, 5185–5197.

199. P. J. van Maaren and D. van der Spoel, *J. Phys. Chem. B*, 2001, **105**, 2618–2626.

200. R. Parr, R. Donnelly, M. Levy and W. Palke, *J. Chem. Phys*, 1978, **68**, 3801–3807.

201. R. Parr and R. Pearson, *J. Am. Chem. Soc.*, 1983, **105**, 7512–7516.

202. M. Sprik, *J. Phys. Chem.*, 1991, **95**, 2283–2291.

203. E. Heid, M. Heindl, P. Dienstl and C. Schröder, *J. Chem. Phys.*, 2018, **149**, 044302.

204. A. Dequidt, J. Devemy and A. A. Pádua, *J. Chem. Inf. Model.*, 2016, **56**, 260–268.

205. K. T. Tang and J. P. Toennies, *J. Chem. Phys.*, 1984, **80**, 3726–3741.

206. A. A. H. Pádua, *J. Chem. Phys.*, 2017, **146**, 204501.

207. B. Jeziorski, R. Moszynski and K. Szalewicz, *Chem. Rev.*, 1994, **94**, 1887–1930.

208. P.-O. Löwdin, *J. Mol. Spectrosc.*, 1959, **3**, 46–66.

209. M. Head-Gordon, J. A. Pople and M. J. Frisch, *Chem. Phys. Lett.*, 1988, **153**, 503–506.

210. J. A. Pople, R. Seeger and R. Krishnan, *Int. J. Quantum Chem*, 1977, **12**, 149–163.

211. R. Krishnan and J. A. Pople, *Int. J. Quantum Chem*, 1978, **14**, 91–100.

212. P.C.Hohenberg, W. Kohn and L.J.Sham, *Adv. Quantum Chem.*, 1990, **21**, 1–26.

213. P. A. M. Dirac, *Proc. Cambridge Phil. Soc.*, 1930, **26**, 376.

214. L. H. Thomas, *Proc. Camb. Phil. Soc.*, 1927, **23**, 542.

215. W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.

216. J. Řezáč and P. Hobza, *Chem. Rev.*, 2016, **116**, 5038–5071.

217. K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479–483.

218. M. Kodrycka and K. Patkowski, *J. Chem. Phys.*, 2019, **151**, 070901.

219. K. E. Riley, M. Pitonak, P. Jurecka and P. Hobza, *Chem. Rev.*, 2010, **110**, 5023–5063.

220. J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2013, **9**, 2151–2155.

221. P. Bagus, K. Hermann and C. W. Bauschlicher, *J. Chem. Phys.*, 1984, **80**, 4378–4386.

222. W. J. Stevens and W. Fink, *Chem. Phys. Lett.*, 1987, **139**, 15–22.

223. Y. Mo, J. Gao and S. Peyerimhoff, *J. Chem. Phys.*, 2000, **112**, 5530–5538.

224. Y. Mo, L. Song and Y. Lin, *J. Phys. Chem. A*, 2007, **111 34**, 8291–8301.

225. Y. Mo, P. Bao and J. Gao, *Phys. Chem. Chem. Phys.*, 2011, **13 15**, 6760–6775.

226. R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. Bell and M. Head-Gordon, *J. Phys. Chem. A*, 2007, **111 36**, 8753–65.

227. M. J. S. Phipps, T. Fox, C. S. Tautermann and C.-K. Skylaris, *Chem. Soc. Rev.*, 2015, **44**, 3177–3211.

228. L. Zhao, M. von Hopffgarten, D. M. Andrada and G. Frenking, *WIREs Comput. Mol. Sci.*, 2018, **8**, e1345.

229. K. Patkowski, *WIREs Comput. Mol. Sci.*, 2019, **10**, e1452.

230. G. Jansen, *WIREs Comput. Mol. Sci.*, 2014, **4**, 127–144.

231. R. Moszynski, B. Jeziorski and K. Szalewicz, *Chem. Phys.*, 1994, **100**, 1312–1325.

232. E. G. Hohenstein and C. D. Sherrill, *Chem. Phys.*, 2010, **132**, 184111.

233. T. M. Parker, L. A. Burns, R. M. Parrish, A. G. Ryno and C. D. Sherrill, *J. Chem. Phys.*, 2014, **140**, 094106.

234. H. L. Williams, K. Szalewicz, R. Moszynski and B. Jeziorski, *Chem. Phys.*, 1995, **103**, 4586–4599.

235. R. M. Parrish, E. G. Hohenstein and C. D. Sherrill, *J. Chem. Phys.*, 2013, **139**, 174102.

236. R. Schäffer and G. Jansen, *Theor. Chem. Acc.*, 2012, **131**, 1235.

237. R. Schäffer and G. Jansen, *Mol. Phys.*, 2013, **111**, 2570–2584.

238. D. H. Zaitsau, A. V. Yermalayeu, S. P. Verevkin, J. E. Bara and A. D. Stanton, *Ind. Eng. Chem. Res.*, 2013, **52**, 16615–16621.

239. T. L. Greaves and C. J. Drummond, *Chem. Rev.*, 2015, **115**, 11379–11448.

240. W. Silva, M. Zanatta, A. S. Ferreira, M. C. Corvo and E. J. Cabrita, *Int. J. Mol. Sci.*, 2020, **21**, 7745.

241. F. Philippi, D. Pugh, D. Rauber, T. Welton and P. A. Hunt, *Chem. Sci.*, 2020, **11**, 6405–6422.

242. E. Braun, J. Gilmer, H. B. Mayes, D. L. Mobley, J. I. Monroe, S. Prasad and D. M. Zuckerman, *Living J Comput Mol Sci*, 2019, **1**, year.

243. I.-C. Yeh and G. Hummer, *J. Phys. Chem. B*, 2004, **108**, 15873–15879.

244. B. Dünweg and K. Kremer, *J. Chem. Phys.*, 1993, **99**, 6983–6997.

245. H. Wang, R. S. DeFever, Y. Zhang, F. Wu, S. Roy, V. S. Bryantsev, C. J. Margulis and E. J. Maginn, *J. Chem. Phys.*, 2020, **153**, 214502.

246. B. L. Holian and D. J. Evans, *J. Chem. Phys.*, 1983, **78**, 5147–5150.

247. Y. Zhang and E. J. Maginn, *J. Chem. Theory Comput.*, 2015, **11**, 3537–3546.

248. O. Borodin, *J. Phys. Chem. B*, 2009, **113**, 11463–11478.

249. F. Müller-Plathe, *Acta Polym.*, 1994, **45**, 259–293.

250. W. Zhao, F. Leroy, B. Heggen, S. Zahn, B. Kirchner, S. Balasubramanian and F. Müller-Plathe, *J. Am. Chem. Soc.*, 2009, **131**, 15825–15833.

251. M. Kohagen, M. Brehm, J. Thar, W. Zhao, F. Müller-Plathe and B. Kirchner, *J. Phys. Chem. B*, 2011, **115**, 693–702.

252. Y. Zhang and E. J. Maginn, *J. Phys. Chem. Lett.*, 2015, **6**, 700–705.

253. M. T. Humbert, Y. Zhang and E. J. Maginn, *J. Chem. Inf. Model.*, 2019, **59**, 1301–1305.

254. J. Zhang and M. Dolg, *Phys. Chem. Chem. Phys.*, 2015, **17**, 24173–81.

255. J. Zhang and M. Dolg, *Phys. Chem. Chem. Phys.*, 2016, **18**, 3003–3010.

256. J. Zhang and V. Glezakou, *Int J Quantum Chem*, 2020, **121**, e26553.

257. S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671.

258. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Rev. C.01*, 2016.

259. S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.

260. S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

261. R. M. Parrish, L. A. Burns, D. G. A. Smith, A. C. Simmonetta, E. A. Eugene DePrince, G. Hohenstein, U. Bozkaya, A. Y. Sokolov, R. D. Remigio, R. M. Richard,

J. F. Gonthier, A. M. James, H. R. McAlexander, A. Kumar, X. Masaaki Saitow, P. V. Wang, Benjamin P. Pritchard, H. F. Schaefer, K. Patkowski, R. A. King, E. F. Valeev, F. A. E. J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Theory Comput.*, 2017, **13**, 3185–3197.

262. J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill and T. D. Crawford, *Wiley Interdiscip. Rev. Comput. Mol. Sci*, 2012, **2**, 556–565.

263. A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, *Comp. Phys. Comm.*, 2022, **271**, 108171.

264. L. Martinez, R. Andrade, E. G. Birgin and J. M. Martinez, *J. Comput. Chem.*, 2009, **30**, 2157–2164.

265. A. A. H. Pádua, *fftool*, 2013, `github.com/agiliopadua/fftool`.

266. H. Sun, *J. Phys. Chem. B*, 1998, **102**, 7338–7364.

267. C. Y. Son, J. G. McDaniel, Q. Cui and A. Yethiraj, *J. Phys. Chem. Lett.*, 2019, **10**, 7523–7530.

268. B. A. D. Neto, E. C. Meurer, R. Galaverna, B. J. Bythell, J. Dupont, R. G. Cooks and M. N. Eberlin, *J. Phys. Chem. Lett.*, 2012, **3**, 3435–3441.

269. V. V. Chaban and O. V. Prezhdo, *J. Phys. Chem. Lett.*, 2012, **3**, 1657–1662.

270. J. P. Leal, J. M. S. S. Esperança, M. E. Minas da Piedade, J. N. Canongia Lopes, L. P. N. Rebelo and K. R. Seddon, *J. Phys. Chem. A*, 2007, **111**, 6176–6182.

271. M. S. Kelkar and E. J. Maginn, *J. Phys. Chem. B*, 2007, **111**, 9424–9427.

272. A. Kazakov, J. Magee, R. Chirico, E. Paulechka, V. Diky, C. Muzny, K. Kroenlein and M. Frenkel, *NIST Standard Reference Database 147: NIST Ionic Liquids Database - (ILThermo)*, 2024, `http://ilthermo.boulder.nist.gov`.

273. N. Sánchez-Ramírez, B. D. Assresahegn, D. Bélanger and R. M. Torresi, *J. Chem. Eng. Data*, 2017, **62**, 3437–3444.

274. J. B. Haskins, W. R. Bennett, J. J. Wu, D. M. Hernandez, O. Borodin, J. D. Monk, J. Bauschlicher, C. W. and J. W. Lawson, *J. Phys. Chem. B*, 2014, **118**, 11295–309.

275. H. Tokuda, K. Ishii, M. A. B. H. Susan, S. Tsuzuki, K. Hayamizu and M. Watanabe, *J. Phys. Chem. B*, 2006, **110**, 2833–2839.

276. F. M. Gaciño, T. Regueira, L. Lugo, M. J. P. Comuñas and J. Fernández, *J. Chem. Eng. Data*, 2011, **56**, 4984–4999.

277. D. H. Zaitsau and S. P. Verevkin, *Int. J. Thermophys.*, 2022, **43**, 24.

278. T. Makino, M. Kanakubo, T. Umecky, A. Suzuki, T. Nishida and J. Takano, *J. Chem. Eng. Data*, 2012, **57**, 751–755.

279. C. A. Nieto de Castro, E. Langa, A. L. Morais, M. L. M. Lopes, M. J. V. Lourenço, F. J. V. Santos, M. S. C. S. Santos, J. N. C. Lopes, H. I. M. Veiga, M. Macatrão, J. M. S. S. Esperança, C. S. Marques, L. P. N. Rebelo and C. A. M. Afonso, *Fluid Phase Equilib.*, 2010, **294**, 157–179.

280. H. Tokuda, K. Hayamizu, K. Ishii, M. A. B. H. Susan and M. Watanabe, *J. Phys. Chem. B*, 2004, **108**, 16593–16600.

281. S. Jiang, Y. Hu, Y. Wang and X. Wang, *J. Phys. Chem. Ref. Data*, 2019, **48**, 033101.

282. A. Nazet, S. Sokolov, T. Sonnleitner, T. Makino, M. Kanakubo and R. Buchner, *J. Chem. Eng. Data*, 2015, **60**, 2400–2411.

283. Y. Guan, R. Clark, F. Philippi, X. Zhang and T. Welton, *J. Chem. Phys.*, 2022, **156**, 204312.

284. A. S. L. Gouveia, C. E. S. Bernardes, L. C. Tomé, E. I. Lozinskaya, Y. S. Vygodskii, A. S. Shaplov, J. N. C. Lopes and I. M. Marrucho, *Phys. Chem. Chem. Phys.*, 2017, **19**, 29617–29624.

285. M. Klajmon and C. Červinka, *J. Chem. Theory Comput.*, 2021, **17**, 6225–6239.

286. M. Klajmon and C. Červinka, *J. Phys. Chem. B*, 2022, **126**, 2005–2013.

287. H. Tokuda, K. Hayamizu, K. Ishii, M. A. B. H. Susan and M. Watanabe, *J. Phys. Chem. B*, 2005, **109**, 6103–6110.

288. S. M. Urahata and M. C. Ribeiro, *J Chem Phys*, 2005, **122**, 024511.

289. H. Liu and E. Maginn, *J. Chem. Phys.*, 2011, **135**, 124507.

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

291. M. Casalegno, G. Raos, G. B. Appetecchi, S. Passerini, F. Castiglione and A. Mele, *J. Phys. Chem. Lett.*, 2017, **8**, 5196–5202.

292. P. Kubisiak and A. Eilmes, *J. Phys. Chem. B*, 2020, **124**, 9680–9689.

293. Y. Shao, K. Shigenobu, M. Watanabe and C. Zhang, *J. Phys. Chem. B*, 2020, **124**, 4774–4780.

294. J. O. Bockris and A. K. N. Reddy, *Volume 1: Modern Electrochemistry: Ionics.*, Springer, New York, NY, 2nd edn, 1998.

295. M. Kanakubo, K. R. Harris, N. Tsuchihashi, K. Ibuki and M. Ueno, *J. Phys. Chem. B*, 2007, **111**, 2062–2069.

296. K. R. Harris, *J. Phys. Chem. B*, 2010, **114**, 9572–9577.

297. D. R. MacFarlane, M. Forsyth, E. I. Izgorodina, A. P. Abbott, G. Annat and K. Fraser, *Phys. Chem. Chem. Phys.*, 2009, **11**, 4962–4967.

298. K. Ueno, H. Tokuda and M. Watanabe, *Phys. Chem. Chem. Phys.*, 2010, **12**, 1649–1658.

299. K. R. Harris and M. Kanakubo, *J. Chem. Eng. Data*, 2016, **61**, 2399–2411.

300. H. K. Kashyap, H. V. R. Annapureddy, F. O. Raineri and C. J. Margulis, *J. Phys. Chem. B*, 2011, **115**, 13212–13221.

301. F. Philippi, D. Rauber, M. Springborg and R. Hempelmann, *J. Phys. Chem. A*, 2019, **123**, 851–861.

302. O. Hollóczki, F. Malberg, T. Welton and B. Kirchner, *Phys. Chem. Chem. Phys.*, 2014, **16**, 16880–16890.

303. B. Kirchner, F. Malberg, D. S. Firaha and O. Hollóczki, *J. Phys.: Condens. Matter*, 2015, **27**, 463002.

304. P. Walden, *Z. Phys. Chem.*, 1906, **55**, 207–249.

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

306. W. Xu and C. A. Angell, *Science*, 2003, **302**, 422–425.

307. M. Yoshizawa, W. Xu and C. A. Angell, *J. Am. Chem. Soc.*, 2003, **125**, 15411–15419.

308. C. A. Angell, C. T. Imrie and M. D. Ingram, *Polym. Int.*, 1998, **47**, 9–15.

309. C. Schreiner, S. Zugmann, R. Hartl and H. J. Gores, *J. Chem. Eng. Data*, 2010, **55**, 1784–1788.

310. H. J. V. Tyrrell and K. R. Harris, *Diffusion in Liquids*, Butterworths, London, 1984, pp. 258–310.

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

312. K. Karu, M. Mišin, H. Ers, J. Sun and V. Ivaništšev, *Int. J. Quantum Chem.*, 2018, **118**, e25582.

313. M. T. Humbert, Y. Zhang and E. J. Maginn, *Mol. Syst. Des. Eng.*, 2017, **2**, 293–300.

314. C. J. Fennell, A. Bizjak, V. Vlachy and K. A. Dill, *J. Phys. Chem. B*, 2009, **113**, 6782–6791.

315. I. S. Joung and I. Cheatham, Thomas E., *J. Phys. Chem. B*, 2008, **112**, 9020–9041.

316. I. S. Lim, P. Schwerdtfeger, B. Metz and H. Stoll, *J. Chem. Phys.*, 2005, **122**, 104103.

317. M. Brehm and B. Kirchner, *J. Chem. Inf. Model*, 2011, **51**, 2007–2023.

318. M. Brehm, M. Thomas, S. Gehrke and B. Kirchner, *J. Chem. Phys.*, 2020, **152**, 164105.

319. D. McQuarrie, *Statistical Mechanics*, University Science Books, 2000.

320. K. U. Lao, R. Schaffer, G. Jansen and J. M. Herbert, *J. Chem. Theory Comput.*, 2015, **11**, 2473–2486.

321. C. Schwartz, *Phys. Rev.*, 1962, **126**, 1015–1019.

322. D. Feller, *J. Chem. Phys.*, 1992, **96**, 6104–6114.

323. T. Helgaker, W. Klopper, H. Koch and J. Noga, *J. Chem. Phys.*, 1997, **106**, 9639–9646.

324. T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

325. M. Kodrycka, C. Holzer, W. Klopper and K. Patkowski, *J. Chem. Theory Comput.*, 2019, **15**, 5965–5986.

326. A. J. Misquitta and K. Szalewicz, *J. Chem. Phys.*, 2005, **122**, 214109.

327. A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen and A. K. Wilson, *Chem. Phys. Lett.*, 1998, 243–252.

328. D. G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45–48.

329. D. W. Schwenke, *J. Chem. Phys.*, 2005, **122**, 14107.

330. J. Řezáč and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**, 685–689.

331. A. Heßelmann and T. Korona, *J. Chem. Phys.*, 2014, **141**, 094107.

332. A. Heßelmann, G. Jansen and M. Schütz, *J. Chem. Phys.*, 2005, **122**, 014103.

333. A. Tekin and G. Jansen, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1680–1687.

334. R. Sedlák, J. Fanfrlík, D. Hnyk, P. Hobza and M. Lepšík, *J. Phys. Chem. A*, 2010, **114**, 11304–11311.

335. C. Leforestier, A. Tekin, G. Jansen and M. Herman, *J. Chem. Phys.*, 2011, **135**, 234306.

336. A. Heßelmann, *J. Phys. Chem. A*, 2011, **115**, 11321–11330.

337. S. Yourdkhani, T. Korona and N. L. Hadipour, *J. Comput. Chem.*, 2015, **36**, 2412–2428.

338. S. Yourdkhani, M. Chojecki, M. Hapka and T. Korona, *J. Phys. Chem. A*, 2016, **120**, 6287–6302.

339. R. Sedlak and J. Řezáč, *J. Chem. Theory Comput.*, 2017, **13**, 1638–1646.

340. A. Heßelmann, *J. Chem. Theory Comput.*, 2018, **14**, 1943–1959.

341. M. Przybytek, *J. Chem. Theory Comput.*, 2018, **14**, 5105–5117.

342. Q. Wang, J. A. Rackers, C. He, R. Qi, C. Narth, L. Lagardere, N. Gresh, J. W. Ponder, J. P. Piquemal and P. Ren, *J. Chem. Theory Comput.*, 2015, **11**, 2609–2618.

343. K. U. Lao and J. M. Herbert, *J. Phys. Chem. A*, 2012, **116**, 3042–3047.

344. D. G. A. Smith, L. A. Burns, A. C. Simmonett, R. M. Parrish, M. C. Schieber, R. Galvelis, P. Kraus, H. Kruse, R. D. Remigio, A. Alenaizan, A. M. James, S. Lehtola, J. P. Misiewicz, M. Scheurer, R. A. Shaw, J. B. Schriber, Y. Xie, Z. L. Glick, D. A. Sirianni, J. S. O'Brien, J. M. Waldrop, A. Kumar, E. G. Hohenstein, B. P. Pritchard, B. R. Brooks, H. F. SchaeferIII, A. Y. Sokolov, K. Patkowski, A. E. DePrinceIII, U. Bozkaya, R. A. King, F. A. Evangelista, J. M. Turney, T. D. Crawford and C. D. Sherrill, *J. Chem. Phys.*, 2020, **152**, 184108.

345. E. F. Valeev, *Chem. Phys. Lett.*, 2004, **395**, 190–195.

346. F. A. Bischoff, S. Wolfsegger, D. P. Tew and W. Klopper, *Mol. Phys.*, 2009, **107**, 963–975.

347. T. B. Adler, G. Knizia and H.-J. Werner, *J. Chem. Phys.*, 2007, **127**, 221106.

348. J. Noga and J. Šimunek, *Chem. Phys.*, 2009, **356**, 1–6.

349. F. Neese, *WIRES Comput. Mol. Sci.*, 2018, **8**, e1327.

350. S. Ten-no, *Chem. Phys. Lett.*, 2004, **398**, 56–61.

351. S. Ten-no, *Chem. Phys.*, 2004, **121**, 117–129.

352. E. F. Valeev and T. Daniel Crawford, *Chem. Phys.*, 2008, **128**, 244113.

353. S. Naseem-Khan, N. Gresh, A. J. Misquitta and J. P. Piquemal, *J. Chem. Theory Comput.*, 2021, **17**, 2759–2774.

Appendix A

Chapter 4

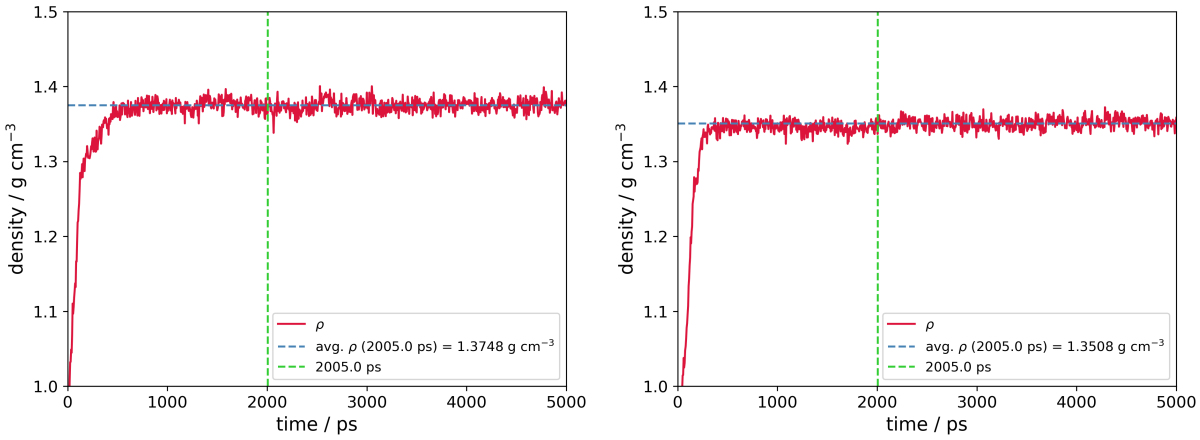
A.1 Final simulation box dimensions

Table A.1: Final box dimensions used for NVT simulations in Chapter 4.

IL	Box dimensions ($\AA$)
[C_4C_1pyr][TFSI]	42.45706
[C_4C_1pyr][FTFSI]	40.96032
[C_4C_1pyr][FSI]	39.4123
[C_4C_1im][TFSI]	41.74013
[C_4C_1im][FTFSI]	40.21818
[C_4C_1im][FSI]	38.56756

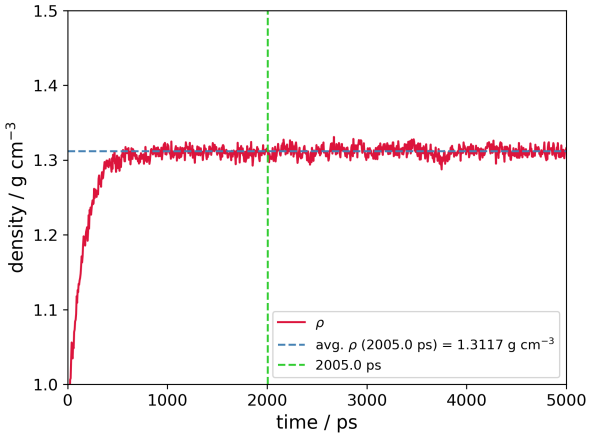
A.2 Density plots

A.3 Beta plots

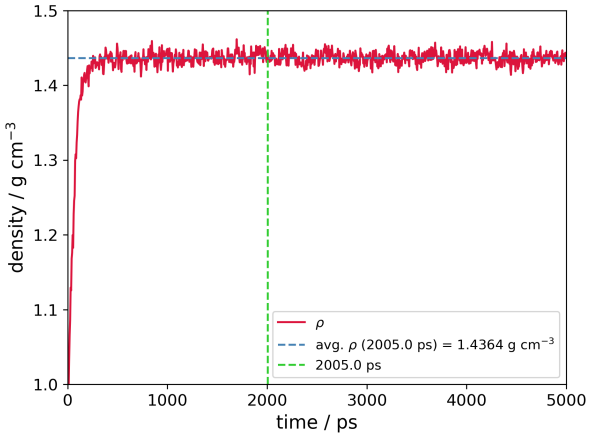


(a) $[C_4C_1pyr][TFSI]$

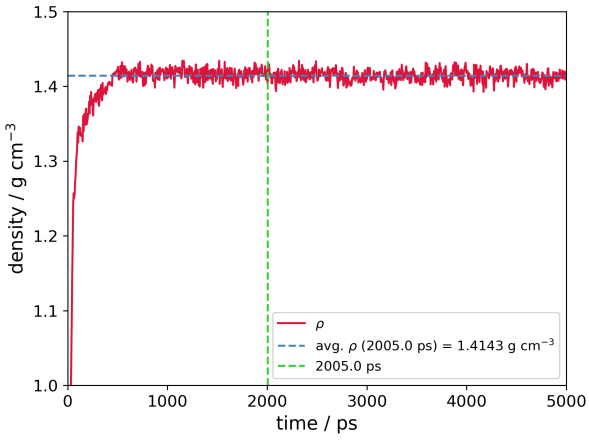
(b) $[C_4C_1pyr][FTFSI]$



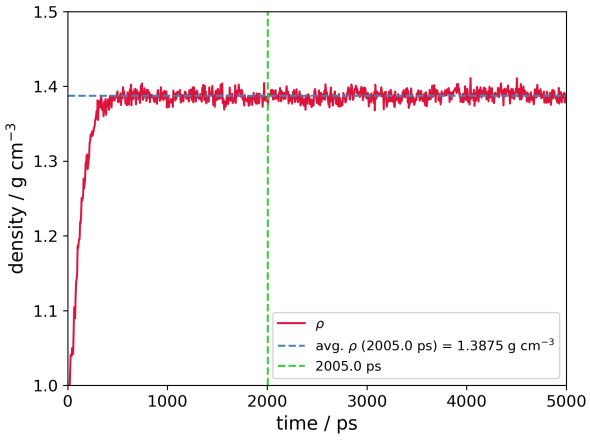
(c) $[C_4C_1pyr][FSI]$



(d) $[C_4C_1im][TFSI]$

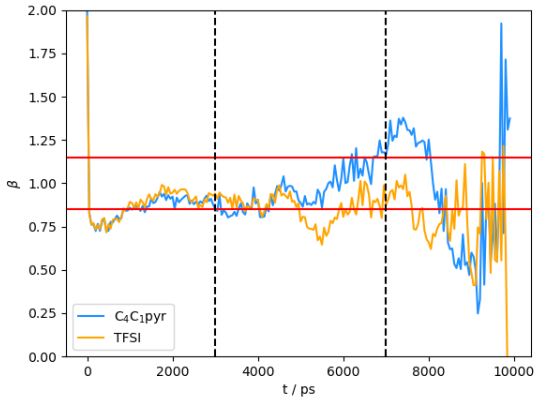


(e) $[C_4C_1im][FTFSI]$

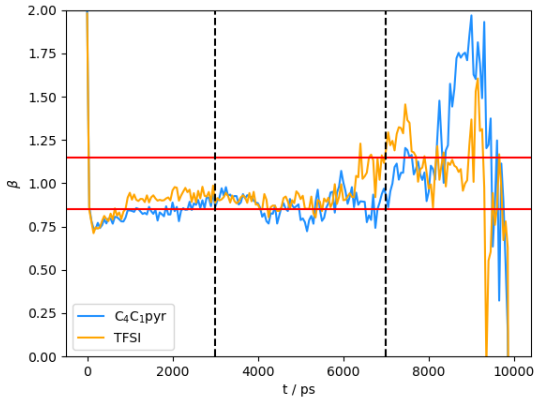


(f) $[C_4C_1im][FSI]$

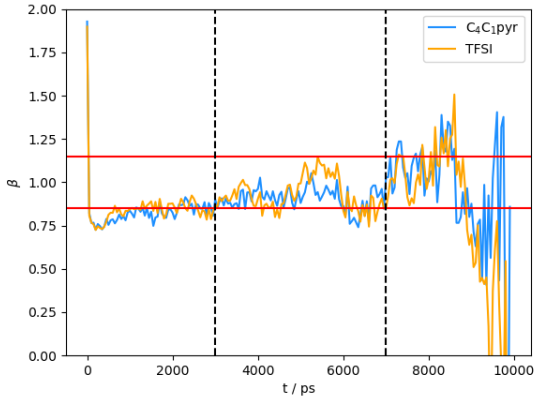
Fig. A.1. Exemplar plots of density over the duration of the *NPT* simulations for each IL.



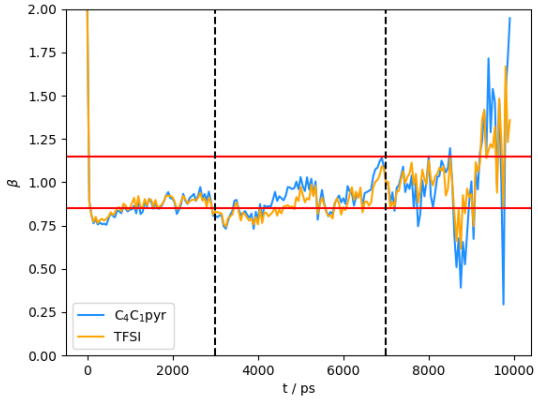
(a) Configuration 1



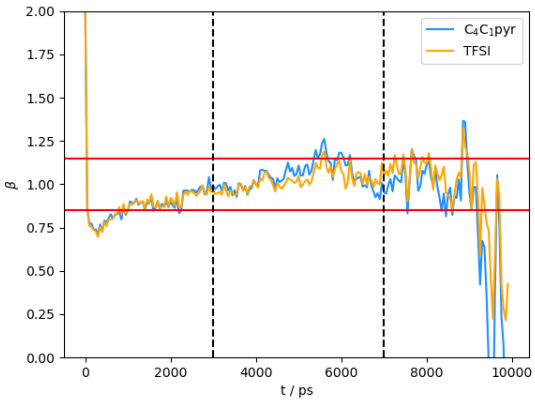
(b) Configuration 2



(c) Configuration 3

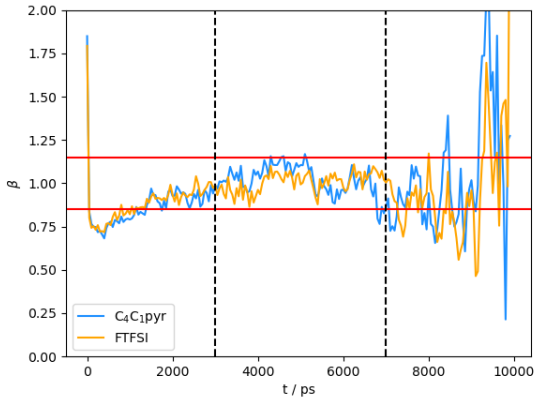


(d) Configuration 4

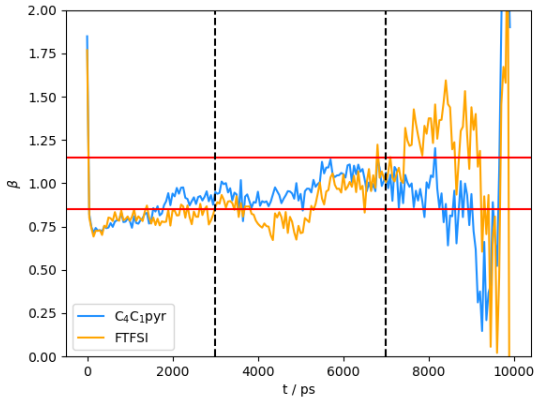


(e) Configuration 5

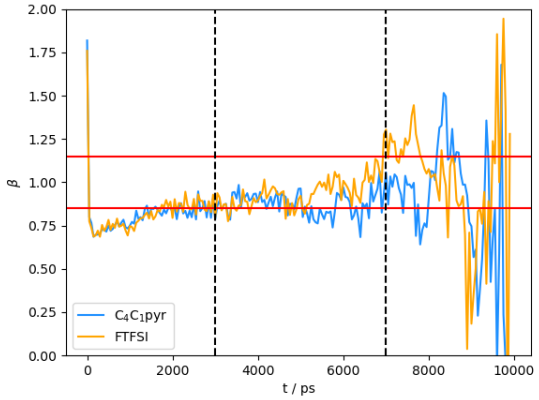
Fig. A.2. Beta plots obtained for each configuration of [C_4C_1pyr][TFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



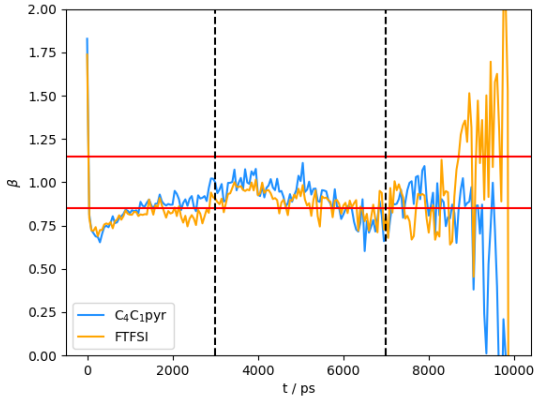
(a) Configuration 1



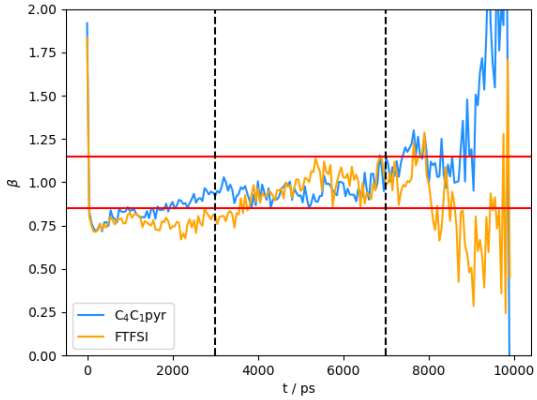
(b) Configuration 2



(c) Configuration 3

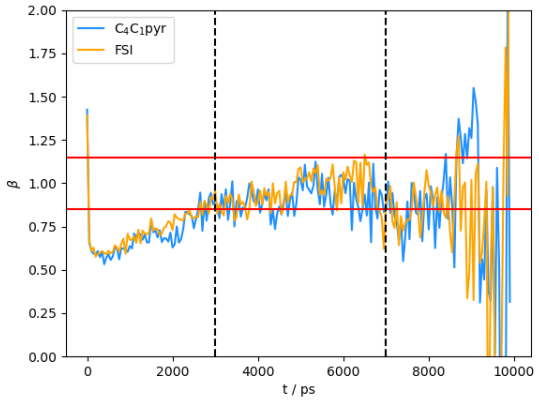


(d) Configuration 4

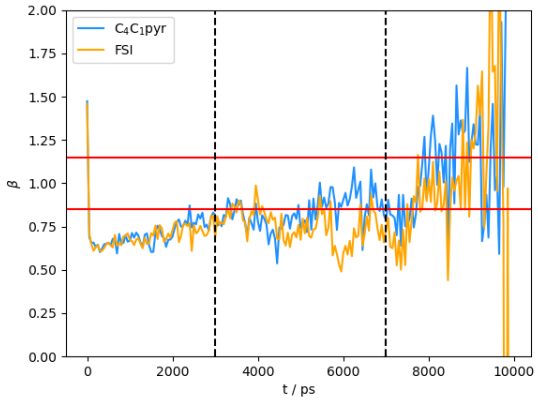


(e) Configuration 5

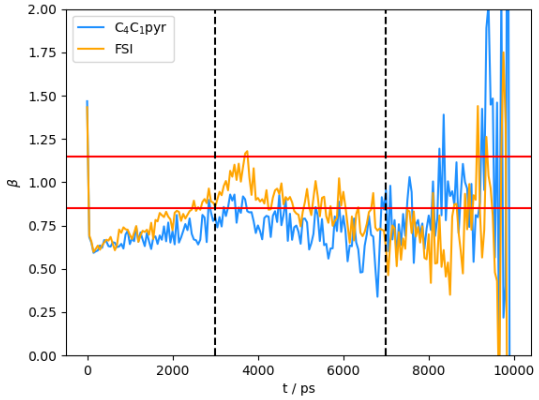
Fig. A.3. Beta plots obtained for each configuration of [C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



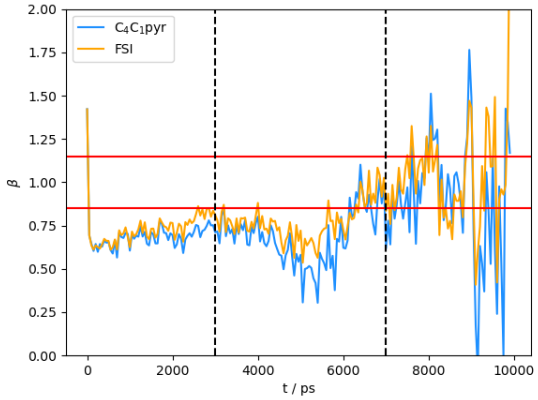
(a) Configuration 1



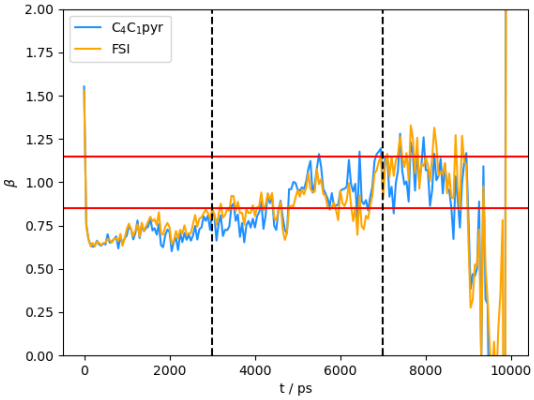
(b) Configuration 2



(c) Configuration 3

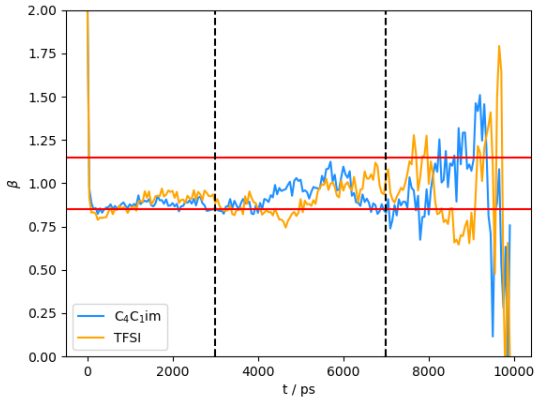


(d) Configuration 4

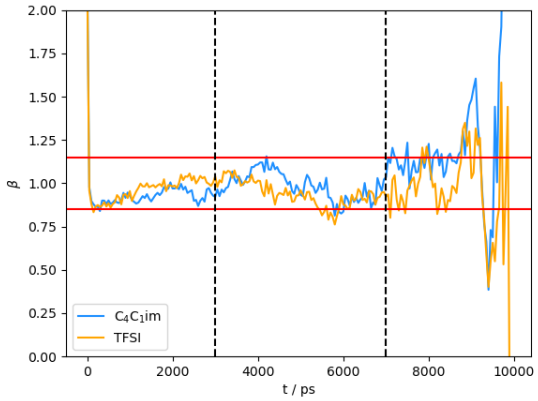


(e) Configuration 5

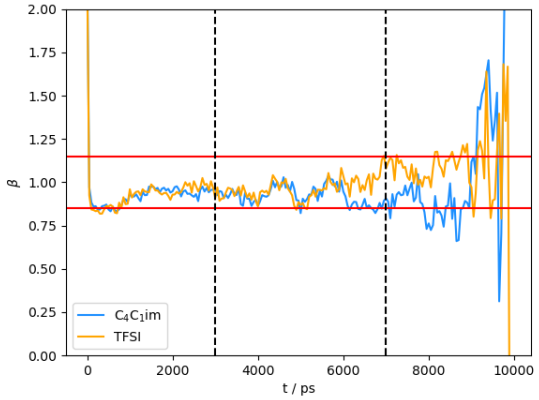
Fig. A.4. Beta plots obtained for each configuration of [C_4C_1pyr][FSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



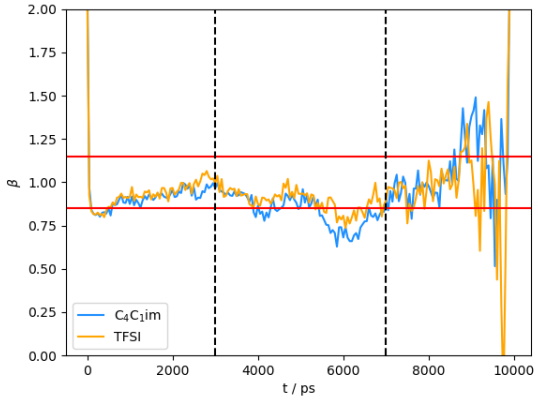
(a) Configuration 1



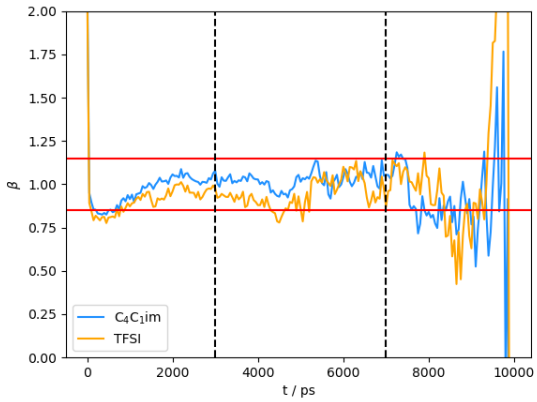
(b) Configuration 2



(c) Configuration 3

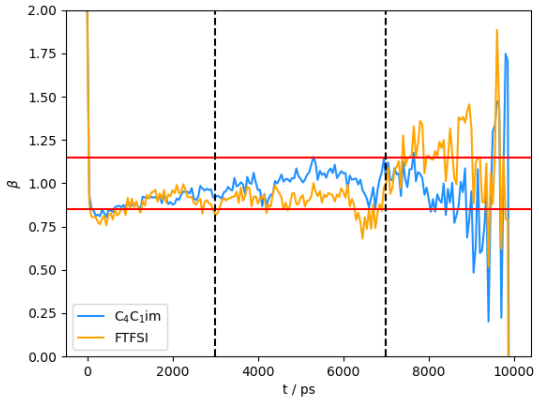


(d) Configuration 4

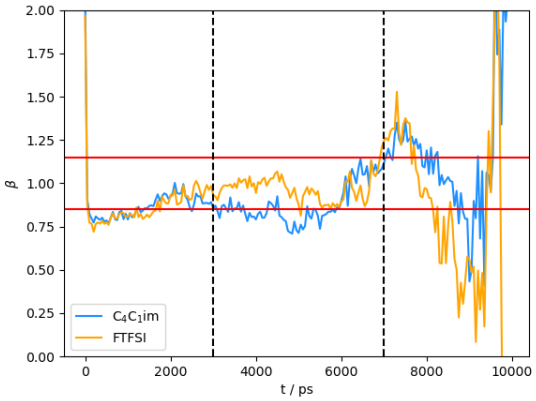


(e) Configuration 5

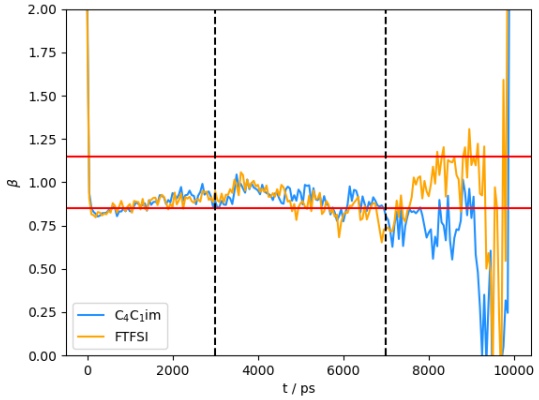
Fig. A.5. Beta plots obtained for each configuration of $[C_4C_1im][TFSI]$. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



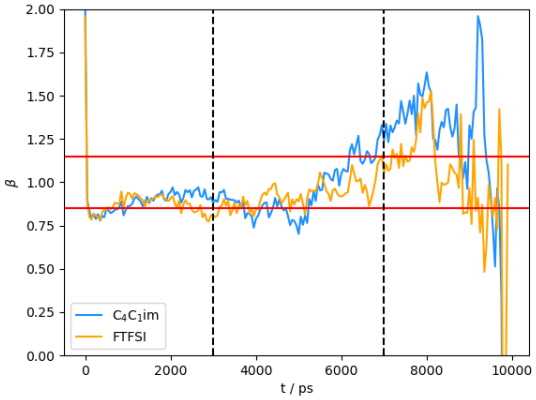
(a) Configuration 1



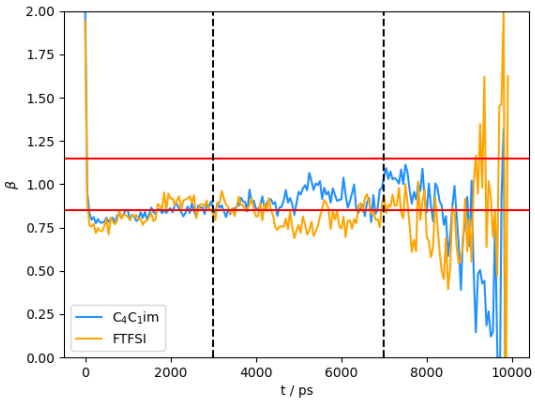
(b) Configuration 2



(c) Configuration 3

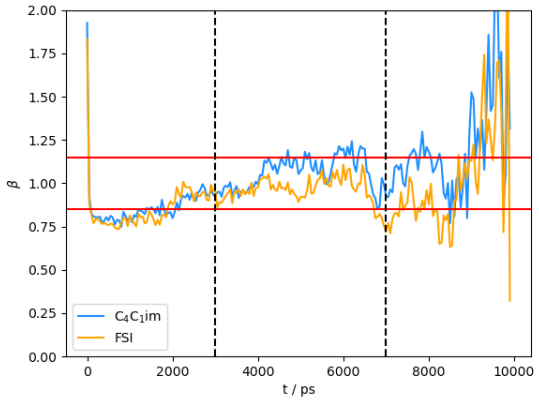


(d) Configuration 4

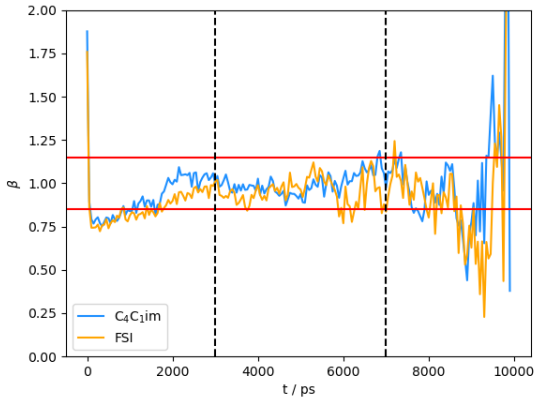


(e) Configuration 5

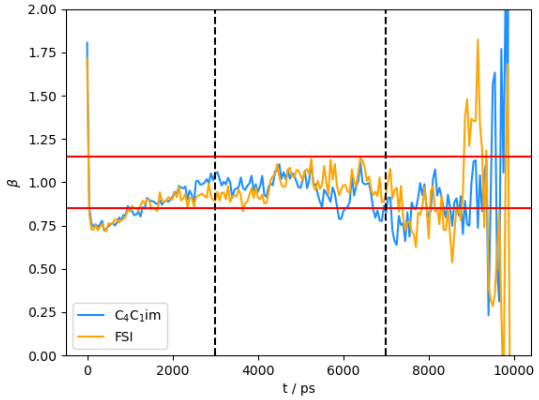
Fig. A.6. Beta plots obtained for each configuration of $[C_4C_1im][FTFSI]$. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



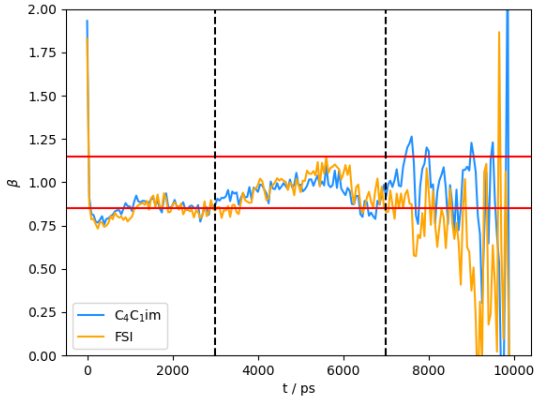
(a) Configuration 1



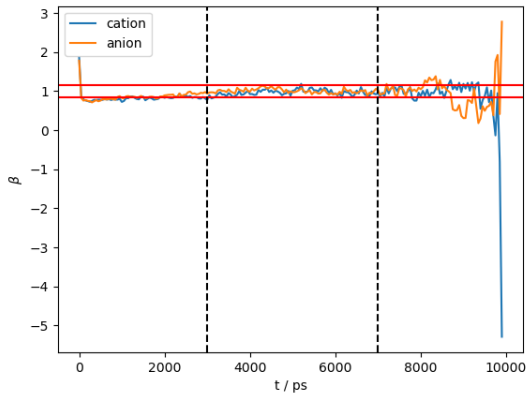
(b) Configuration 2



(c) Configuration 3



(d) Configuration 4



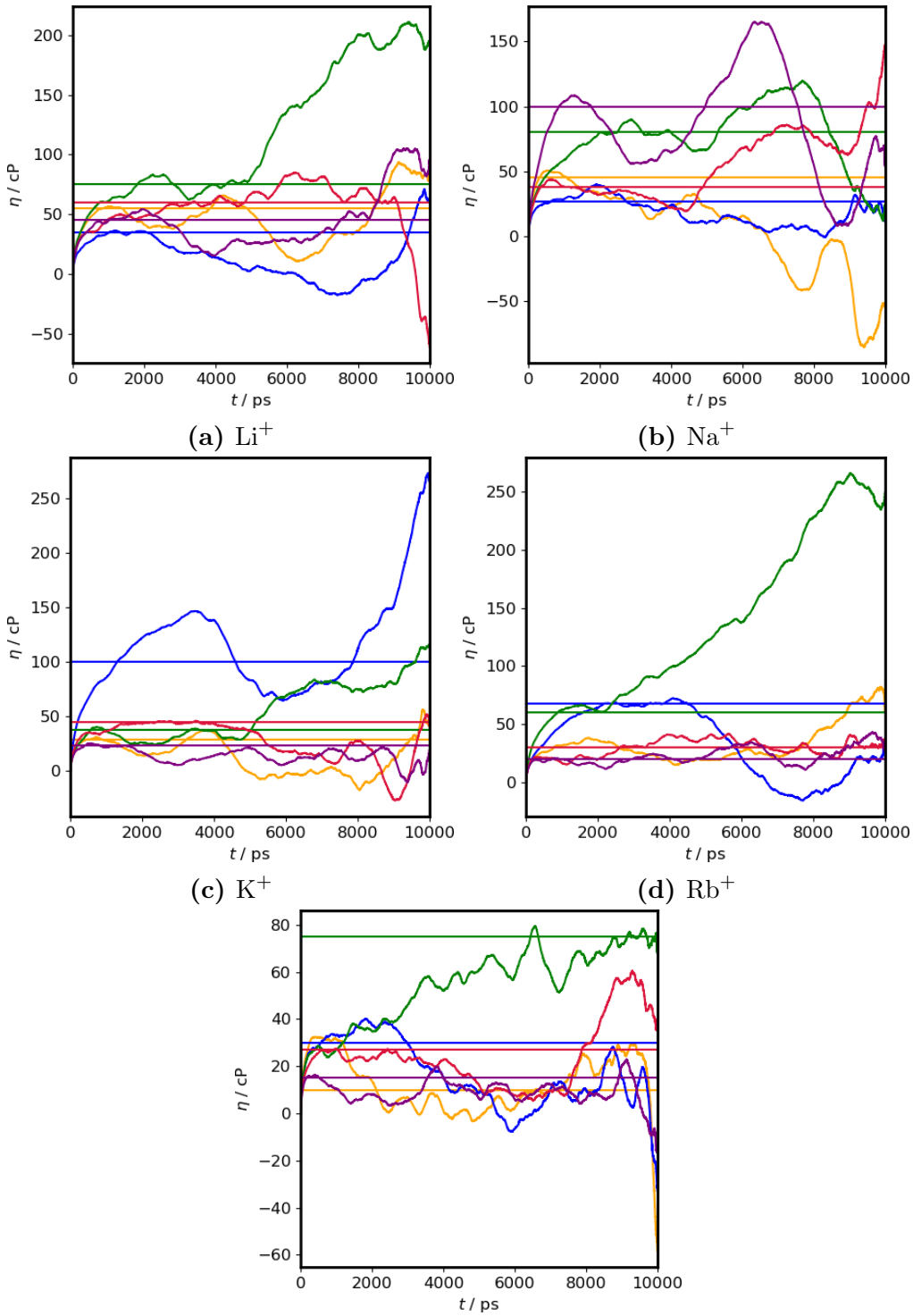
(e) Configuration 5

Fig. A.7. Beta plots obtained for each configuration of $[C_4C_1im][FSI]$. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.

Appendix B

Chapter 5

B.1 Viscosity fitting regions

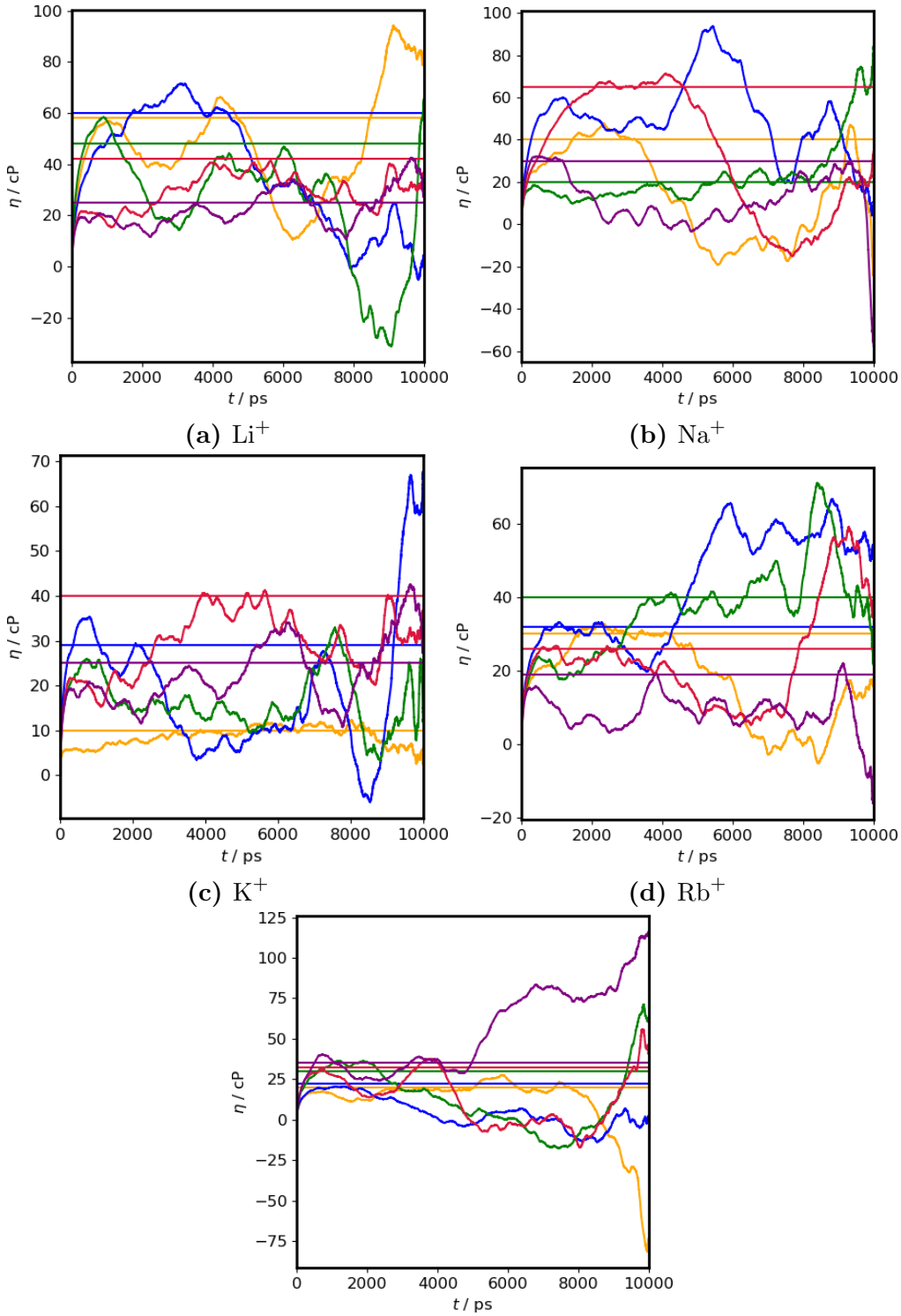


(a) Li^+ **(b)** Na^+

(c) K^+ **(d)** Rb^+

(e) Cs^+

Fig. B.3. Running integrals determined for each trajectory obtained using the CL&P(1) FF for each [M][FTFSI]/[C_4C_1pyr][FTFSI] system indicated for each of the five trajectories, where the integral and corresponding plateau are illustrated in the same colour.



(a) Li^+ **(b)** Na^+

(c) K^+ **(d)** Rb^+

(e) Cs^+

Fig. B.4. Running integrals determined for each trajectory obtained using the CL&P (2) FF for each [M][FTFSI]/[C_4C_1pyr][FTFSI] system indicated for each of the five trajectories, where the integral and corresponding plateau are illustrated in the same colour.

B.2 Beta plots

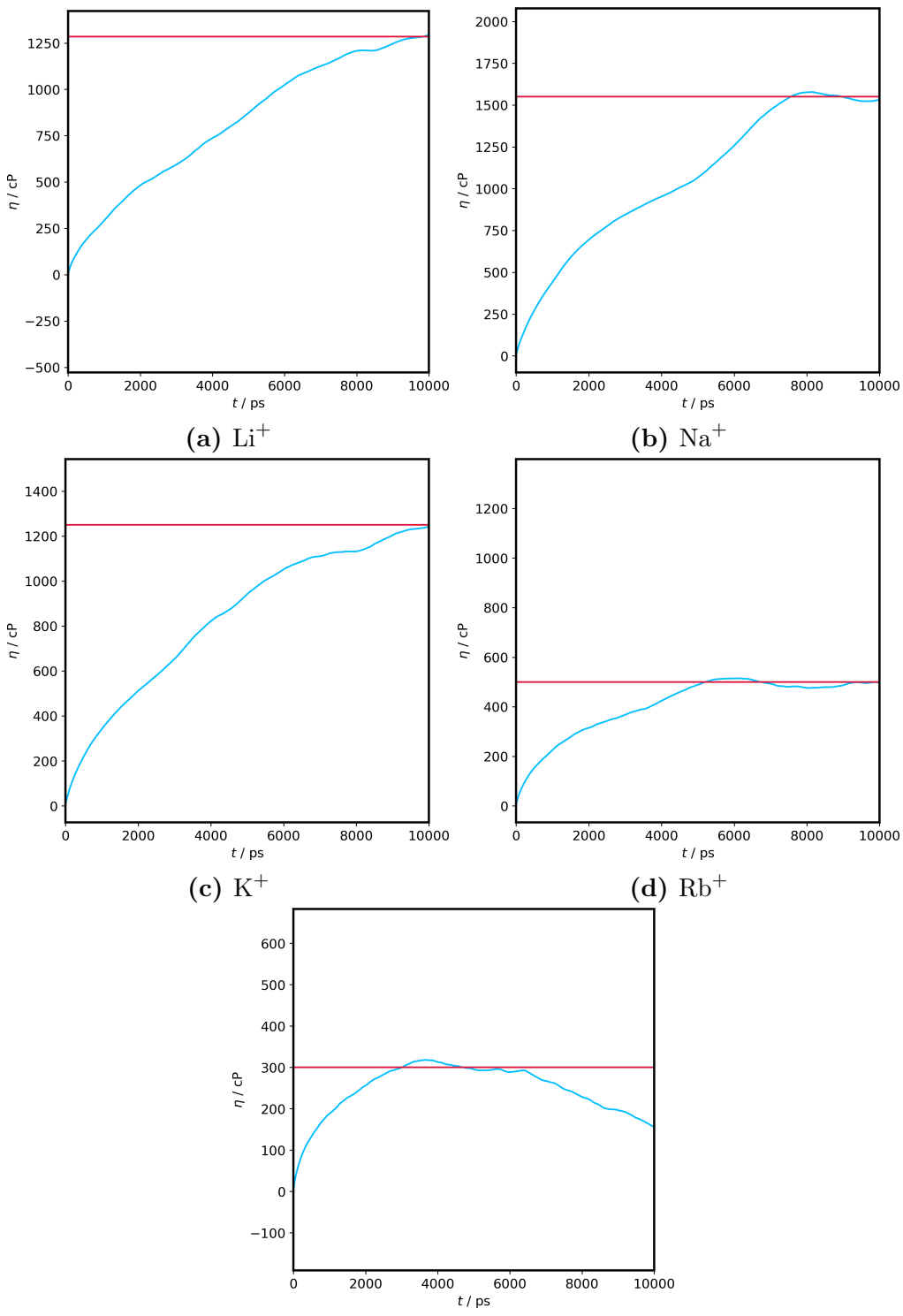
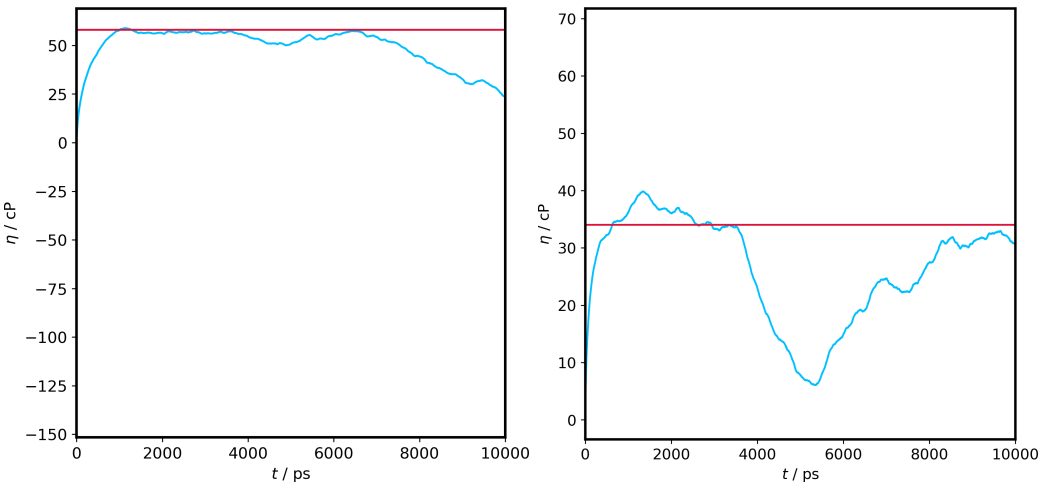
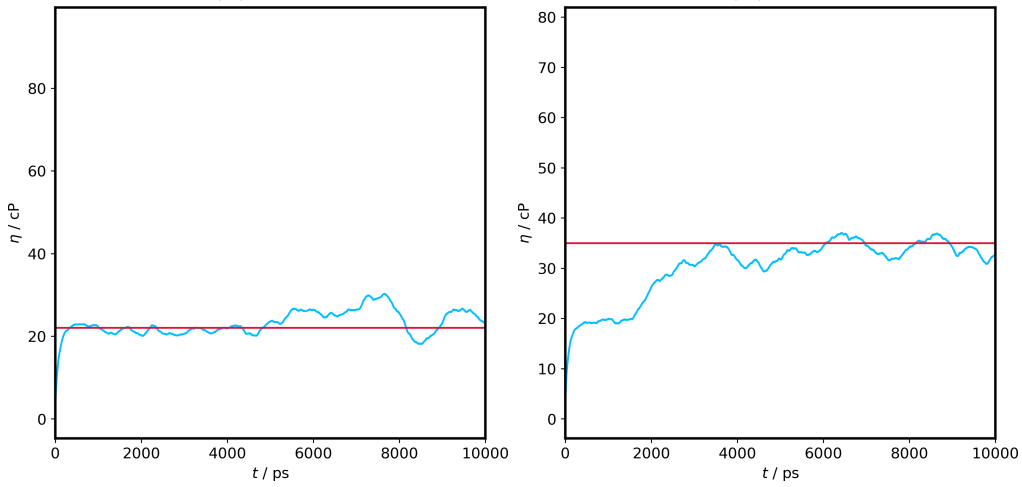


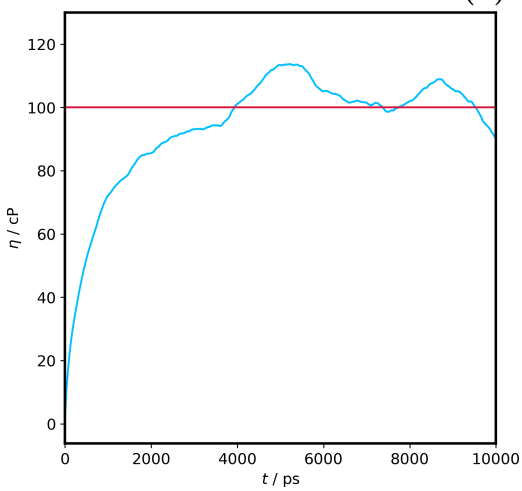
Fig. B.1. Running integrals determined for each trajectory obtained using the CL&P FF for each [M][FTFSI]/[C_4C_1pyr][FTFSI] system indicated.



(a) Li^+ **(b)** Na^+

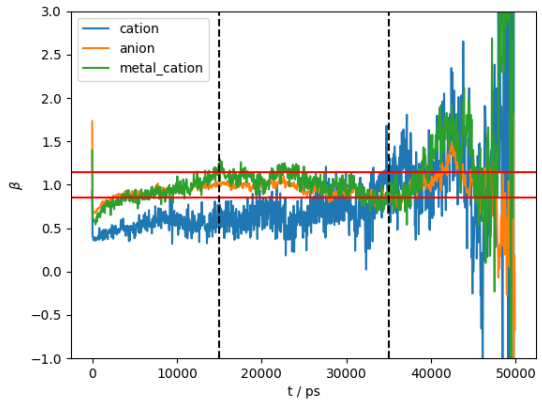


(c) K^+ **(d)** Rb^+

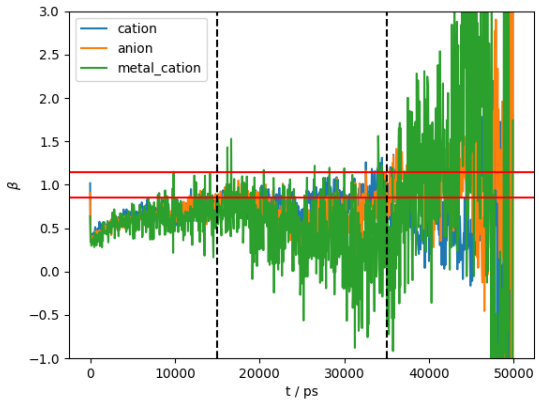


(e) Cs^+

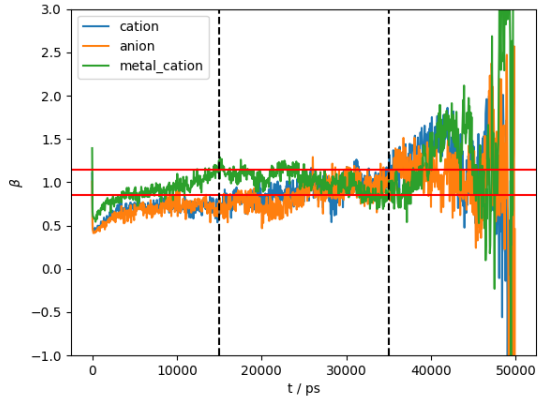
Fig. B.2. Running integrals determined for each trajectory obtained using the sCL&P FF for each [M][FTFSI]/[C_4C_1pyr][FTFSI] system indicated.



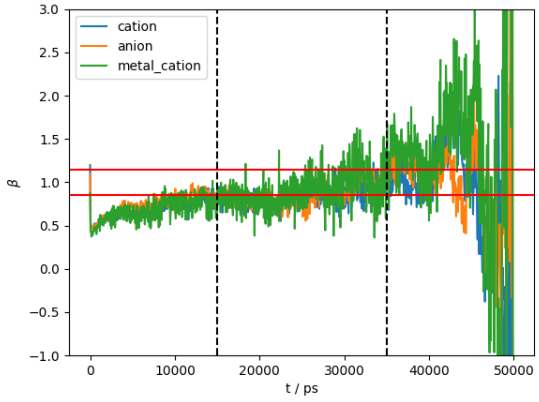
(a) Li^+



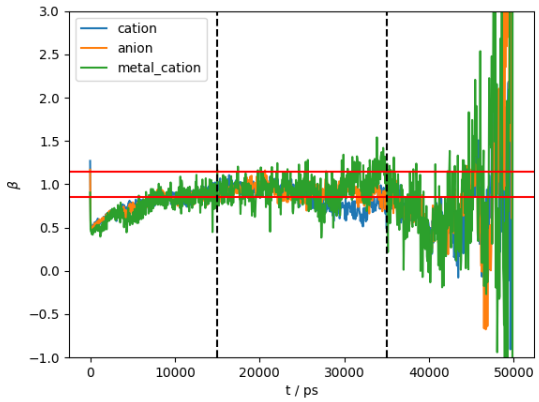
(b) Na^+



(c) K^+

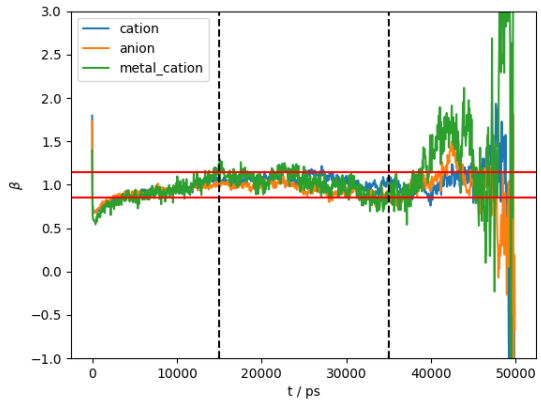


(d) Rb^+

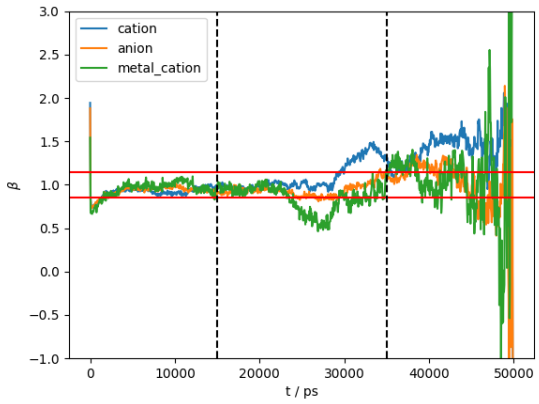


(e) Cs^+

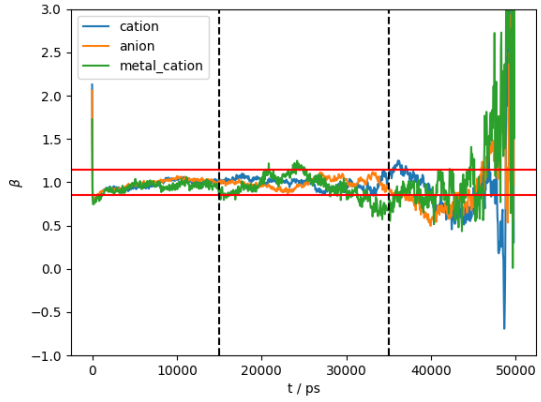
Fig. B.5. Beta plots obtained using the CL&P FF for each [M][FTFSI]/[C_4C_1pyr][FTFSI] system indicated. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



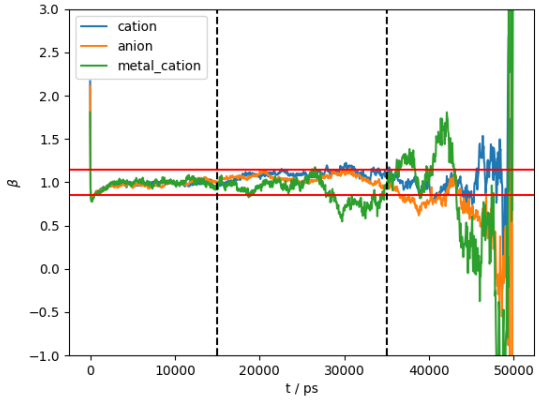
(a) Li^+



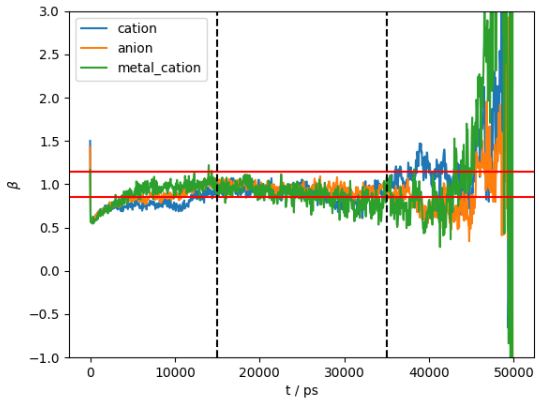
(b) Na^+



(c) K^+

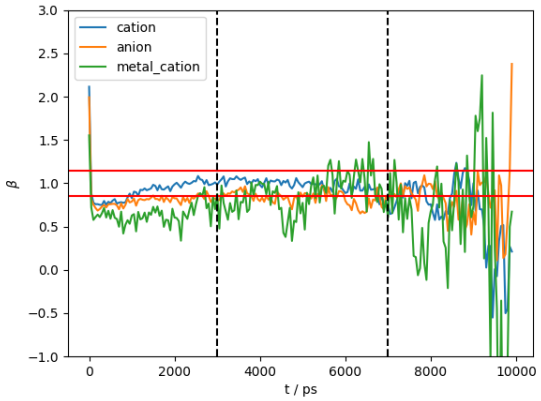


(d) Rb^+

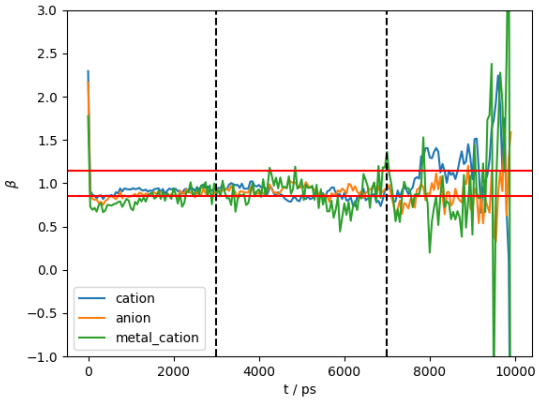


(e) Cs^+

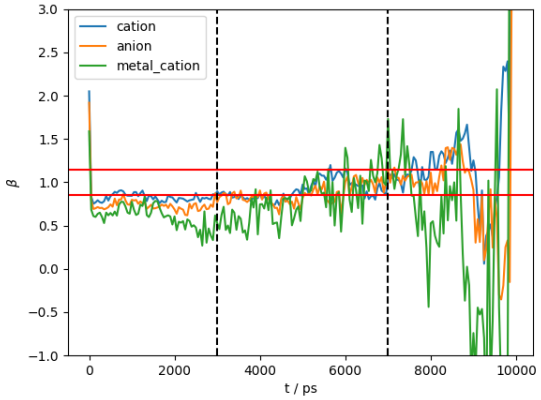
Fig. B.6. Beta plots obtained using the sCL&P FF for each [M][FTFSI]/[C_4C_1pyr][FTFSI] system indicated. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



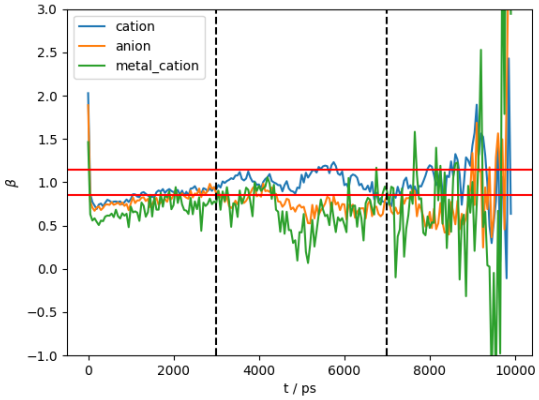
(a) Configuration 1



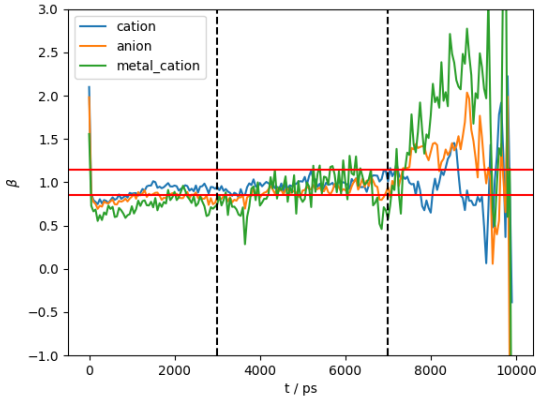
(b) Configuration 2



(c) Configuration 3

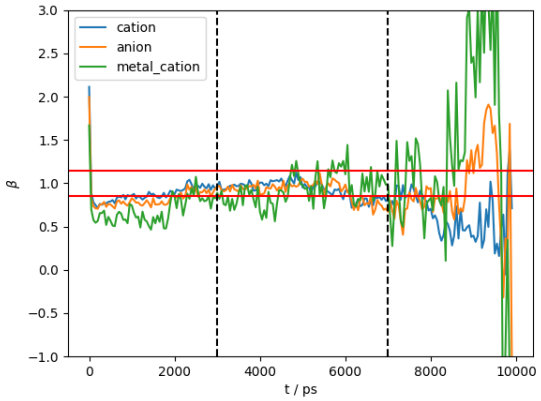


(d) Configuration 4

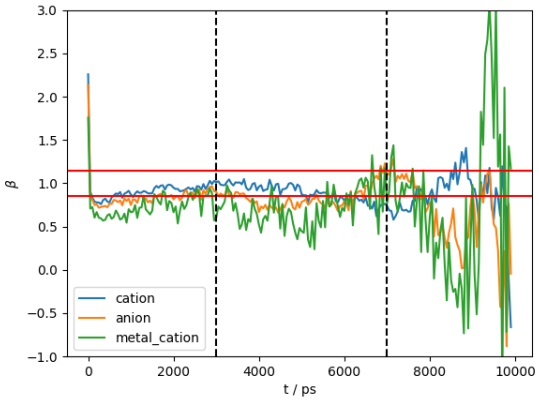


(e) Configuration 5

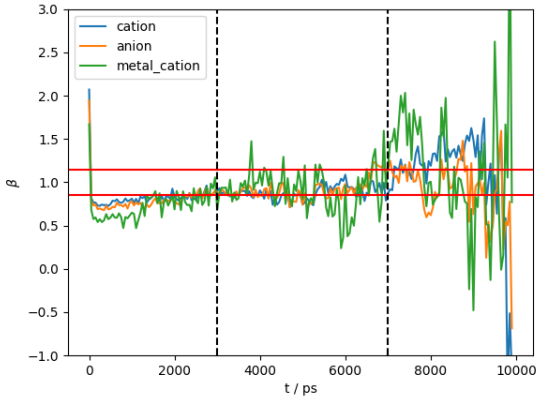
Fig. B.7. Beta plots obtained using the CL&Pol(1) for each configuration of [Li][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



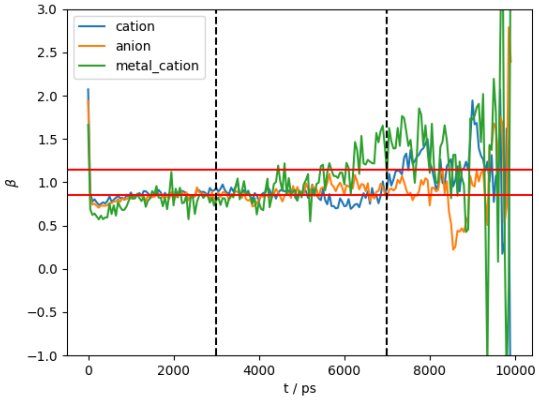
(a) Configuration 1



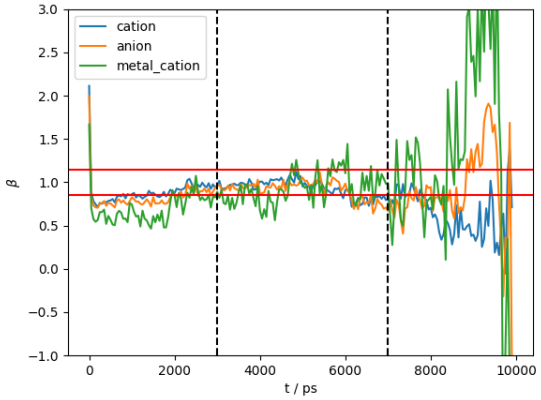
(b) Configuration 2



(c) Configuration 3

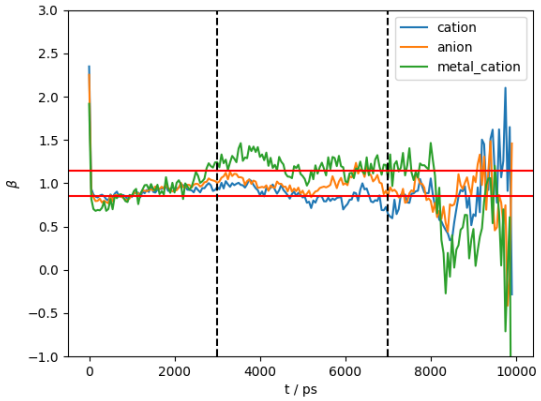


(d) Configuration 4

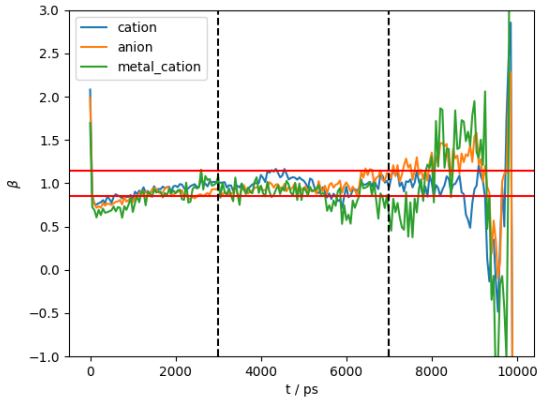


(e) Configuration 5

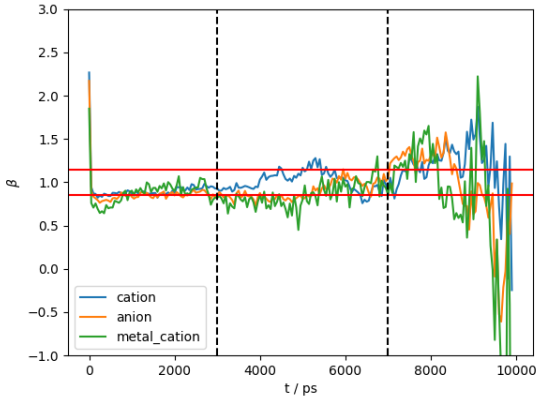
Fig. B.8. Beta plots obtained using the CL&Pol(1) for each configuration of [Na][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



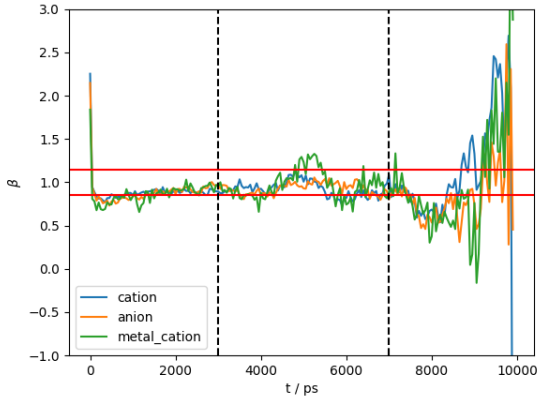
(a) Configuration 1



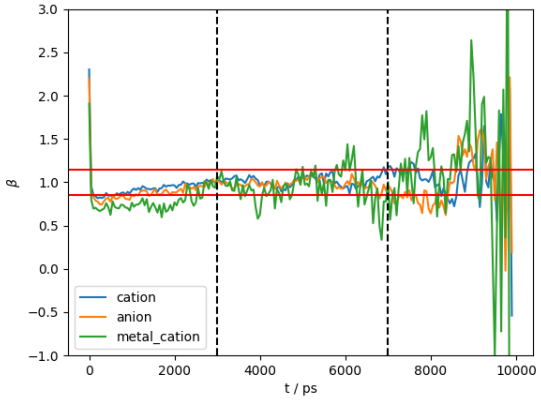
(b) Configuration 2



(c) Configuration 3

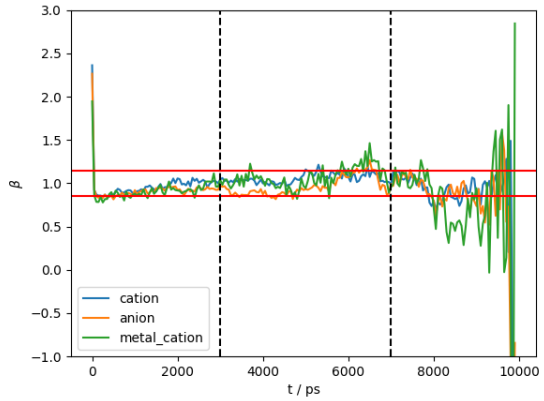


(d) Configuration 4

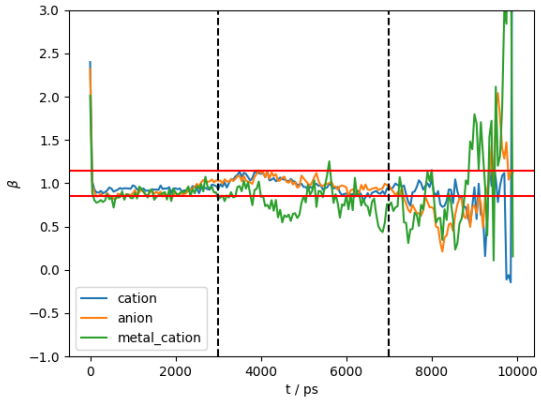


(e) Configuration 5

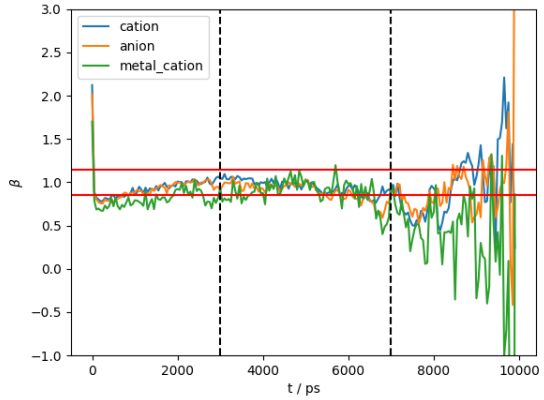
Fig. B.9. Beta plots obtained using the CL&Pol(1) for each configuration of [K][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



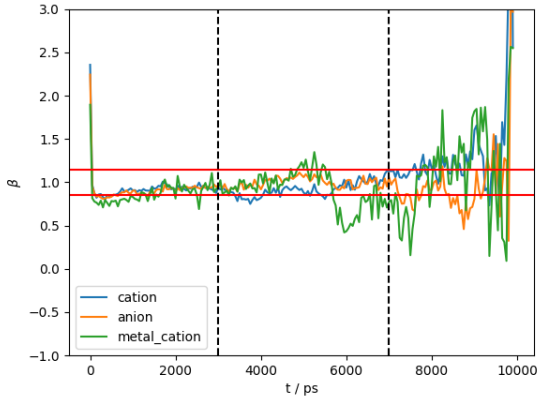
(a) Configuration 1



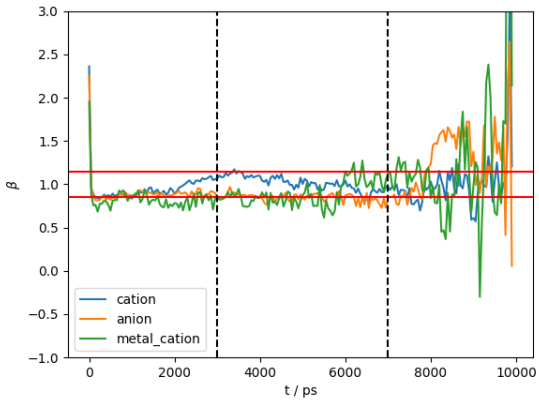
(b) Configuration 2



(c) Configuration 3

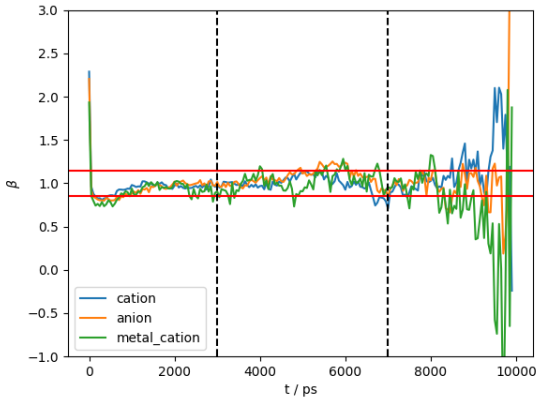


(d) Configuration 4

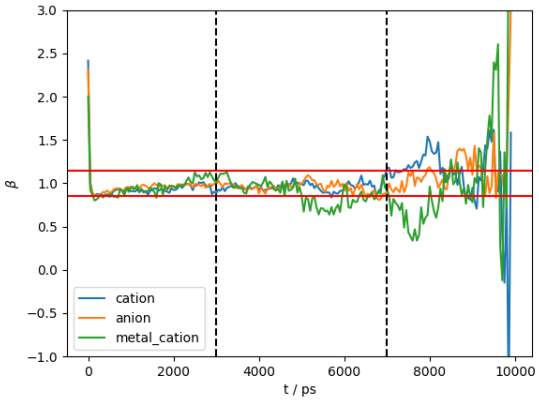


(e) Configuration 5

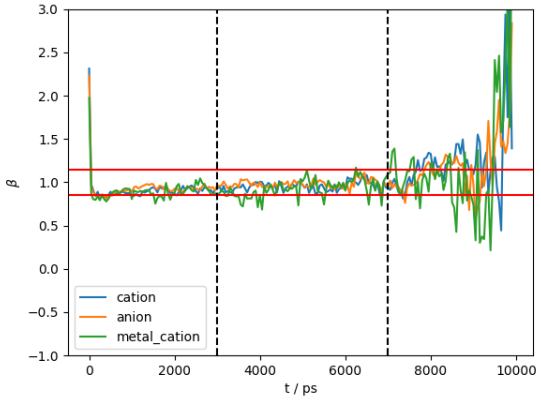
Fig. B.10. Beta plots obtained using the CL&Pol(1) for each configuration of [Rb][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



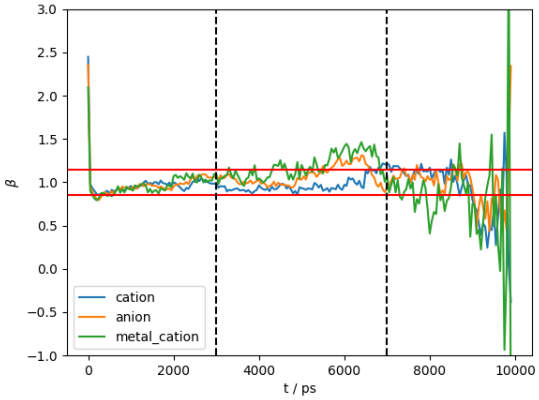
(a) Configuration 1



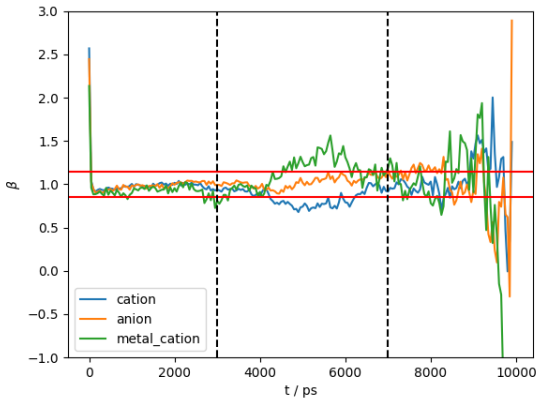
(b) Configuration 2



(c) Configuration 3

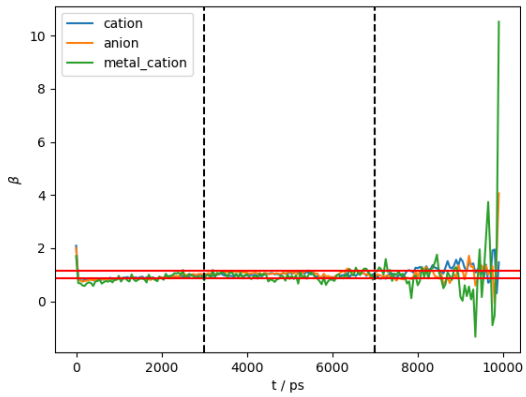


(d) Configuration 4

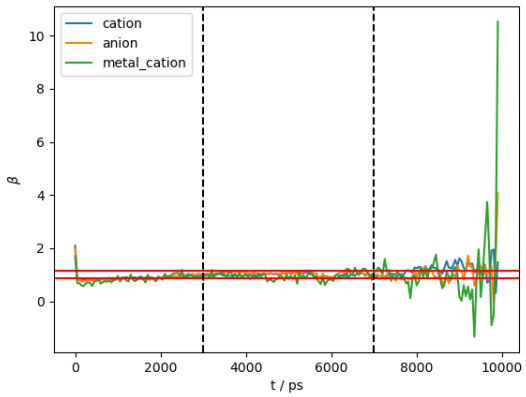


(e) Configuration 5

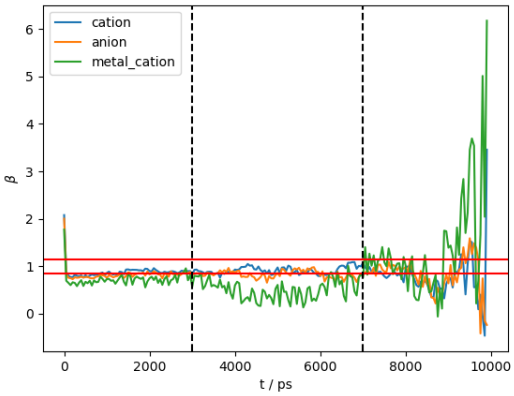
Fig. B.11. Beta plots obtained using the CL&Pol(1) for each configuration of [Cs][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



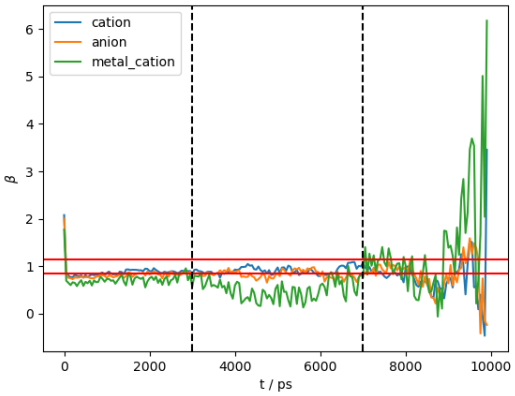
(a) Configuration 1



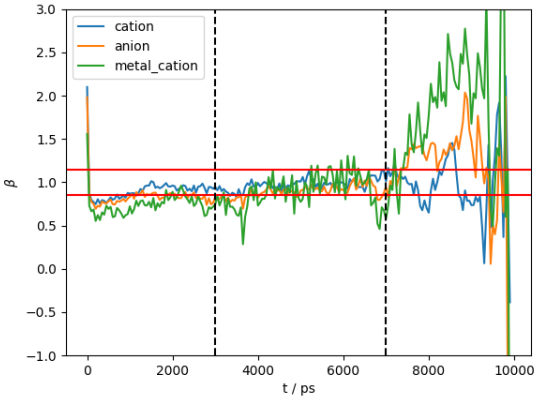
(b) Configuration 2



(c) Configuration 3

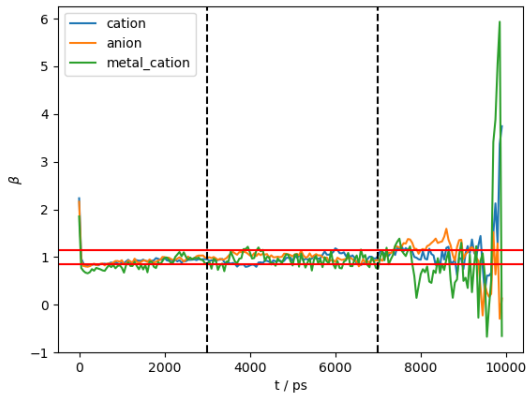


(d) Configuration 4

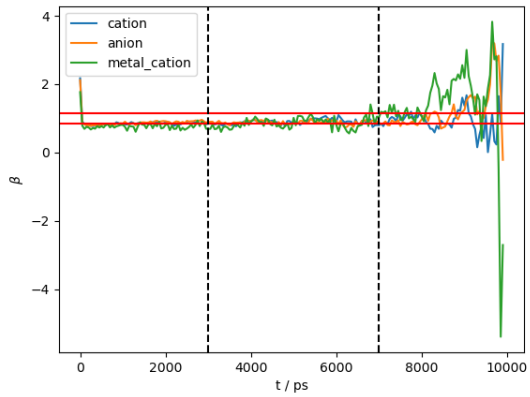


(e) Configuration 5

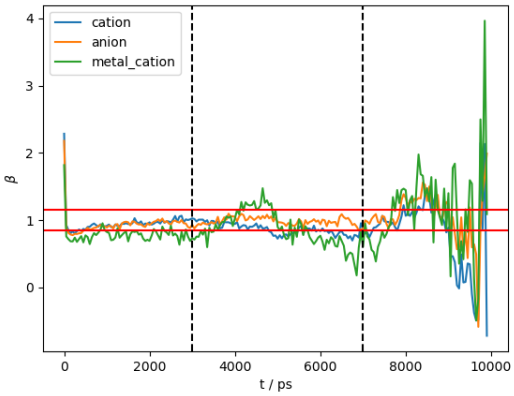
Fig. B.12. Beta plots obtained using the CL&Pol(2) for each configuration of [Li][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



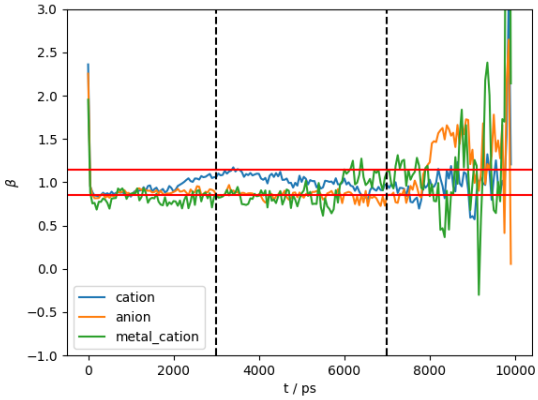
(a) Configuration 1



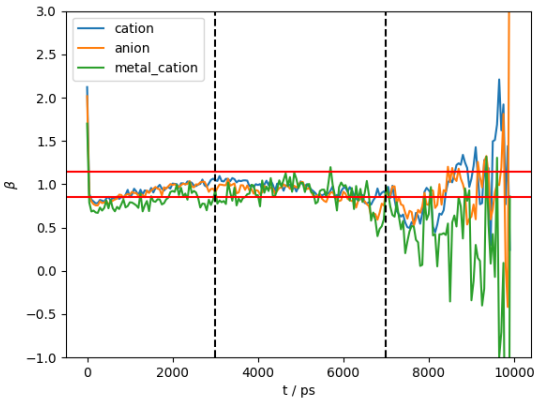
(b) Configuration 2



(c) Configuration 3

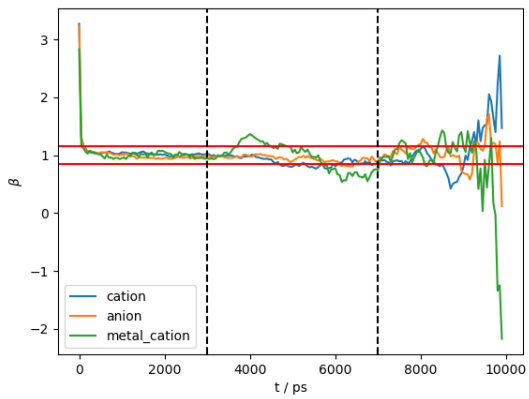


(d) Configuration 4

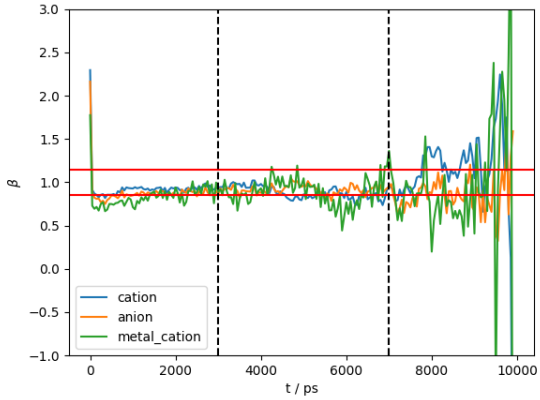


(e) Configuration 5

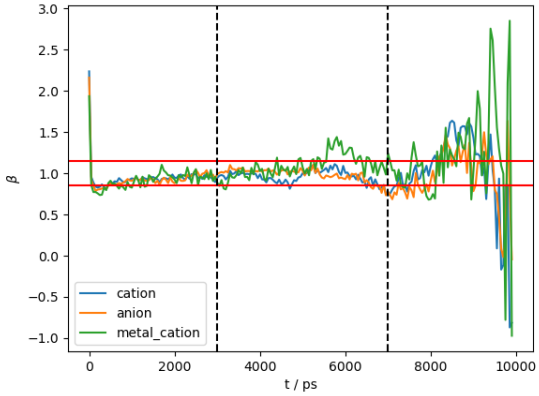
Fig. B.13. Beta plots obtained using the CL&Pol(2) for each configuration of [Na][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



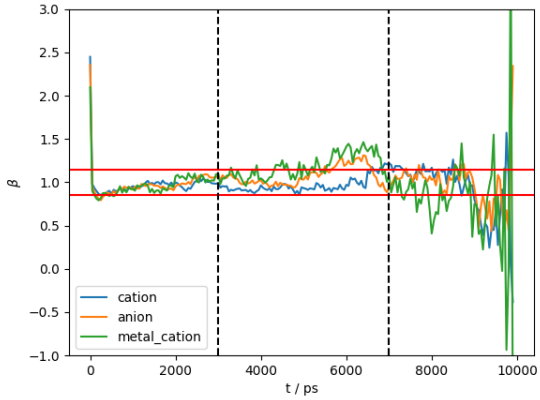
(a) Configuration 1



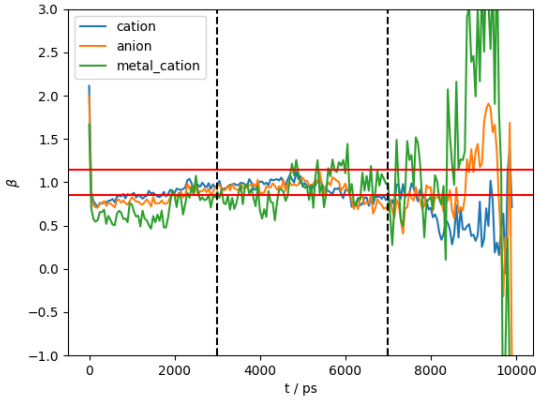
(b) Configuration 2



(c) Configuration 3

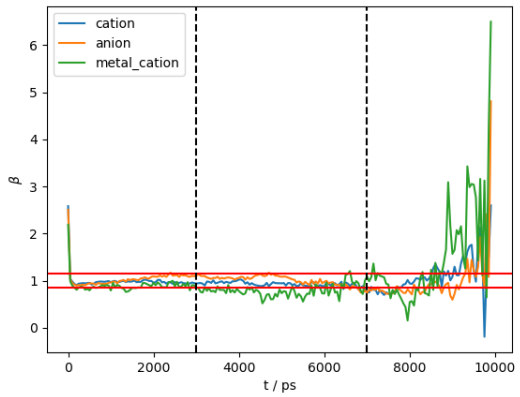


(d) Configuration 4

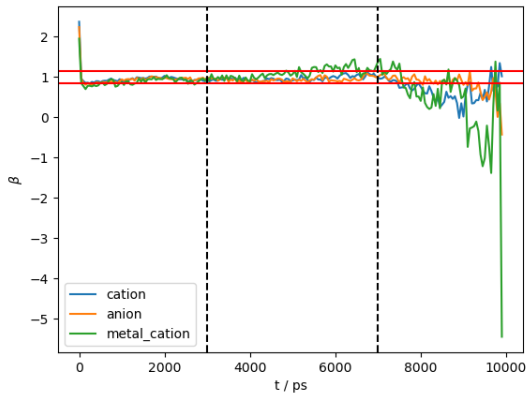


(e) Configuration 5

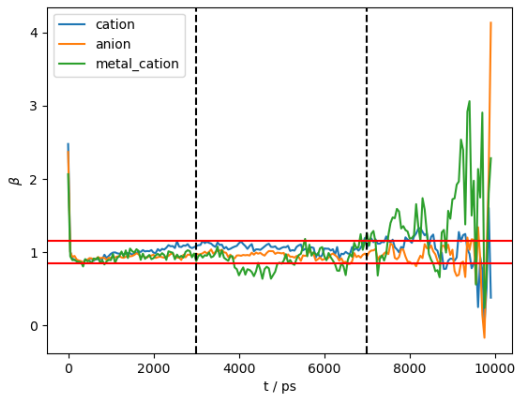
Fig. B.14. Beta plots obtained using the CL&Pol(2) for each configuration of [K][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



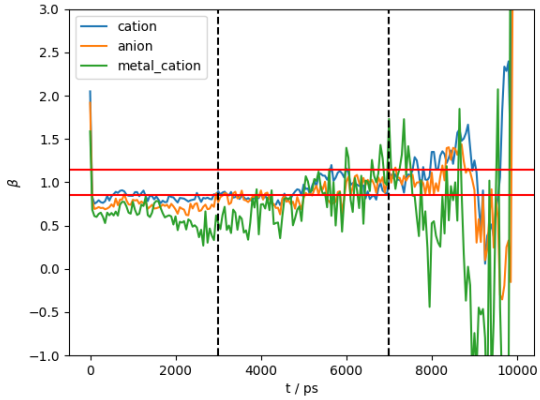
(a) Configuration 1



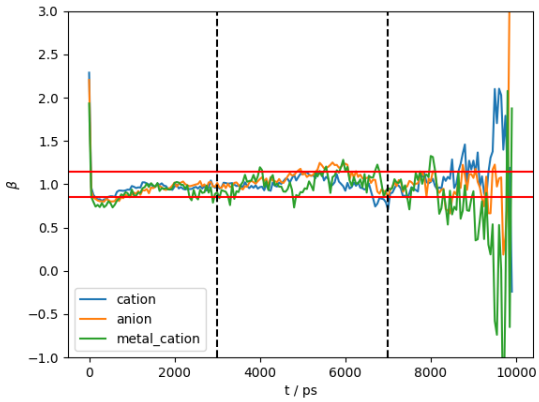
(b) Configuration 2



(c) Configuration 3

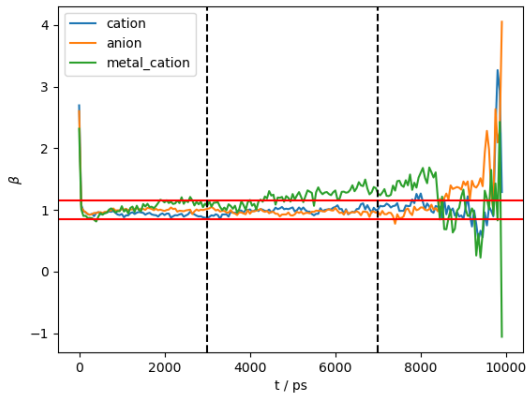


(d) Configuration 4

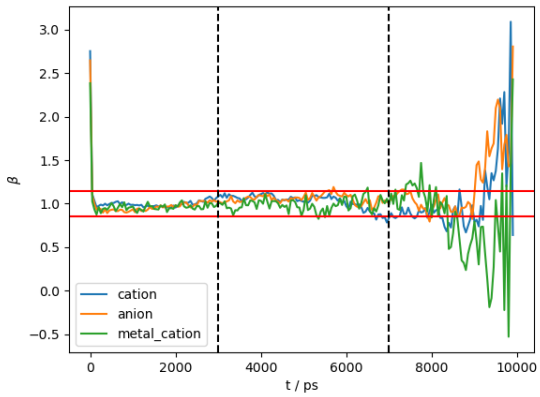


(e) Configuration 5

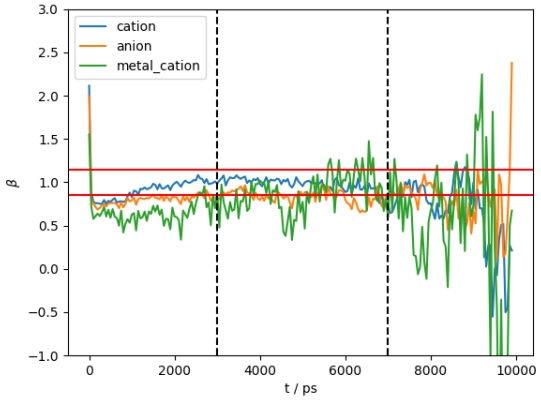
Fig. B.15. Beta plots obtained using the CL&Pol(2) for each configuration of [Rb][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.



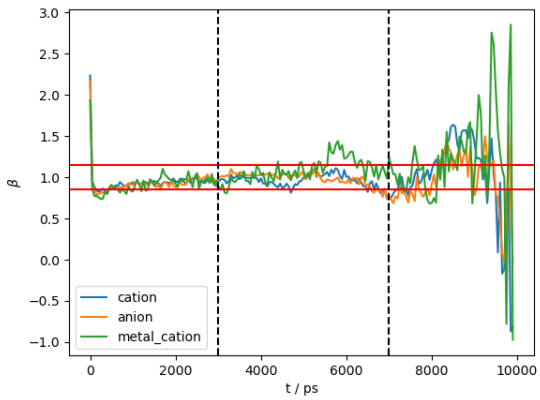
(a) Configuration 1



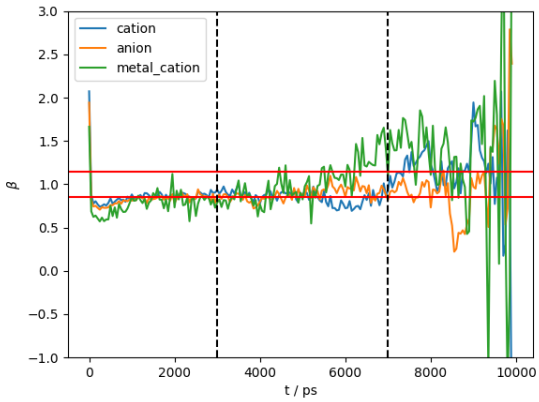
(b) Configuration 2



(c) Configuration 3



(d) Configuration 4



(e) Configuration 5

Fig. B.16. Beta plots obtained using the CL&Pol(2) for each configuration of [Cs][FTFSI]/[C_4C_1pyr][FTFSI]. The area where $\beta = 1 \pm 0.15$ is indicated by the horizontal red lines and the region used for fitting is indicated with the black dashed vertical lines.

Appendix C

Chapter 7

C.1 Genetic Algorithm code

```
import numpy as np
np.random.seed(19950612)
import pandas as pd
from sklearn import preprocessing
from sklearn.model_selection import GroupShuffleSplit
from tensorflow.python.keras import models
from tensorflow.python.keras import layers
from tensorflow.python.keras.callbacks import EarlyStopping,
    ModelCheckpoint
from sklearn.metrics import mean_absolute_error
from statistics import stdev

mae = pd.Series([])
aard = pd.Series([])
bias0 = pd.Series([])
bias1 = pd.Series([])

# Import stratified dataset
```

```
data = pd.read_csv("stratified_esp3.csv")

# Split into test and train set
train = data[data["Split"] == "train"]
test = data[data["Split"] == "test"]
val = data[data["Split"]=="val"]
x_train = train.iloc[:,4:]
x_test = test.iloc[:,4:]
x_val = val.iloc[:,4:]
target_train = train["Heat Capacity"].values
target_test = test["Heat Capacity"].values
target_val = val["Heat Capacity"].values

# Scale data
scaler = preprocessing.StandardScaler()
features_train = scaler.fit_transform(x_train)
features_test = scaler.transform(x_test)
features_val = scaler.transform(x_val)

# Specify parameters
h_nodes = 7 # set number of hidden nodes
early_stopping = True # early stopping switched on
batch_size = 10 # set batch size
activation = "relu" # set activation function
optimizer = "rmsprop" # set optimiser

#Fit neural network 30 times
for i in range(30):

    # Start neural network
    network = models.Sequential()

    # Add first hidden layer
```

```
    network.add(layers.Dense(units=h_nodes,
                             activation=activation,
                             input_shape=(features_train.shape[1],)))

    # Add output layer
    network.add(layers.Dense(units=1))

    # Compile neural network
    network.compile(loss="mse",
                    optimizer=optimizer,
                    metrics=["mse"])

    # Call back used for early stopping
    if early_stopping == True:
        callbacks = [EarlyStopping(monitor="val_loss", patience=2),
                     ModelCheckpoint(filepath="best_model.h5",monitor="
    val_loss",save_best_only=True)]
    else:
        callbacks = None

    # Train neural network
    history = network.fit(features_train, # training features
                          target_train, # target features
                          epochs=500, # number of training epochs
                          callbacks=callbacks,
                          verbose=0,
                          batch_size=batch_size,
                          validation_data=(features_val, target_val))

    # Predicted values of test set
    predicted_target = network.predict(features_test)
    mae[i] = mean_absolute_error(target_test,predicted_target)
```

```
82      aard[i] = (mean_absolute_error(target_test,predicted_target)/
    target_test.mean())*100
83      print("Iteration: ", i+1 ,"; Epochs ran: ",len(history.history['loss'
    ]),"; MAE: ","%.2f" % mae[i],"; AARD%: ","%.2f" % aard[i])
84
85  print("AVERAGE ERROR = ","%.2f" % mae.mean(),"+-","%.2f" % stdev(mae))
86  print("AARD% = ","%.2f" % aard.mean(),"+-","%.2f" % stdev(aard))
```