PERSPECTIVE

Check for updates

Towards predictive design of electrolyte solutions by accelerating *ab initio* simulation with neural networks

Junji Zhang, [a] Joshua Pagotto[a] and Timothy T. Duignan [*ab]

Electrolyte solutions play a vital role in a vast range of important materials chemistry applications. For example, they are a crucial component in batteries, fuel cells, supercapacitors, electrolysis and carbon dioxide conversion/capture. Unfortunately, the determination of even their most basic properties from first principles remains an unsolved problem. As a result, the discovery and optimisation of electrolyte solutions for these applications largely relies on chemical intuition, experimental trial and error or empirical models. The challenge is that the dynamic nature of liquid electrolyte solutions require long simulation times to generate trajectories that sufficiently sample the configuration space; the long range electrostatic interactions require large system sizes; while the short range quantum mechanical (QM) interactions require an accurate level of theory. Fortunately, recent advances in the field of deep learning, specifically equivariant neural network potentials (NNPs), can enable significant accelerations in sampling the configuration space of electrolyte solutions. Here, we outline the implications of these recent advances for the field of materials chemistry and identify outstanding challenges and potential solutions.

1 Introduction

Electrolyte solutions are crucially important for materials chemistry, particularly for energy and sustainability applications. For example, the choice of electrolyte solution impacts the efficiency and safety of battery storage, supercapacitor and carbon capture/conversion technologies, in which they are deployed. They are also fundamental to the water splitting/treatment processes. In all of these applications the specific ions that are present can have a dramatic impact on the overall behaviour of the system.[1] For example, sodium ions are more weakly solvated than lithium ions. This can result in lower desolvation energy barriers and hence improve the kinetics of battery charging.[2] Additionally, replacing potassium with sodium has been shown to substantially accelerate the

[a]*School of Chemical Engineering, The University of Queensland, St Lucia, Brisbane 4072, Australia. E-mail: t.duignan@uq.edu.au*

[b]*School of Environment and Science, Griffith University, Nathan, Brisbane 4111, Australia. E-mail: t.duignan@griffith.edu.au*
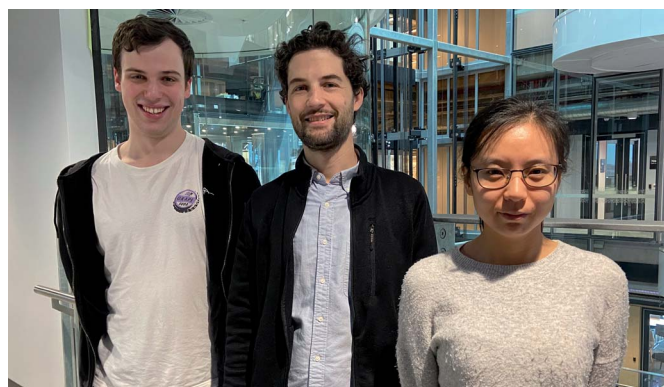
Tim T. Duignan holds a Discovery Early Career Researcher Award in the School of Chemical Engineering at the University of Queensland and has just commenced as a Lecturer in Applied Maths and Physics at Griffith University. He is working on modelling and simulation of electrolyte solutions combining continuum solvent models, ab initio molecular dynamics and deep learning. He is focussed on developing these techniques for application to improving electrochemical energy storage and CO₂ capture and conversion. He completed his PhD in the Applied Mathematics Department at the Australian National University in Canberra before carrying out postdoctoral research at Pacific Northwest National Laboratory in Washington State, USA. Junji Zhang is completing her PhD at the University of Queensland in Chemical Engineering. She completed a master's in chemical engineering at the University of Queensland and a bachelor's of biochemical engineering at Hebei University of Technology. Joshua Pagotto is completing his undergraduate studies in Chemistry and Chemical Engineering at the University of Queensland.

extraction of CO_2 from air using hydroxide solutions.[3,4] Finally, the cation can have a significant effect on the products produced by electrocatalytic CO_2 reduction[5] and water splitting.[6]

To address the energy and sustainability challenges faced by modern society, significant advances in the performance of all of these applications are necessary. Towards these advances, scientists and engineers need the ability to design electrolyte solutions based on a quantitatively accurate molecular understanding of their properties. This ambition is depicted in Fig. 1.

Quantum density functional theory (DFT) is a crucial and well established tool in understanding, designing and optimising new solid state materials. Although, there are still some limitations and challenges associated with the accuracy and efficiency of these methods.[7] In contrast, the use of DFT for understanding, designing and optimising electrolyte solutions has been more limited. In fact, even the most basic properties of electrolyte solutions such as the solubility of sodium chloride in water still cannot be predicted accurately from first principles.[8] Additionally, explaining specific ion effects remains a significant challenge.[1,9] Knowing the properties of the electrolyte

Fig. 1 Electrolyte solutions composed of ions dissolved in a solvent play a fundamental role in a vast range of important materials science applications, such as electrochemical energy storage and CO_2 capture/conversion, which are key to combating climate change.

solution such as conductivities and solubilities is critical for large scale models of important systems. Additionally, many key phenomena occur at the electrode–electrolyte interface.[5,10] Therefore, an accurate theoretical treatment of the solid state alone is often of only limited use. For example, intercalation in batteries and electrocatalytic reduction are both interfacial processes.

Currently, researchers rely heavily on experimental databases of electrolyte solution properties. Unfortunately, these often contain unreliable data or are missing key pieces of information.[11,12] For instance, the activity coefficients of many fundamental electrolytes such as lithium bicarbonate and rubidium hydroxide in water have never been reported in the experimental literature to the best of our knowledge.[13]

In recent years, researchers have focused on developing predictive approaches to the design of electrolyte solutions for battery applications using large datasets[14,15] such as the electrolyte genome project.[16] This work could be accelerated by improving the availability and reliability of the underlying data they rely on.

Our understanding of electrolyte solutions lags behind that of solid state electrodes because electrolytes are inherently more dynamic. This means that any reliable treatment of electrolyte solutions requires a well converged statistical description to determine the relative prevalence of various configurations, which significantly increases the computational cost of studying electrolyte systems. Additionally, long range electrostatic interactions play a key role in the behaviour of electrolyte solutions requiring large system sizes. Finally, small inaccuracies in the predicted energies of different structures can bias the statistical sampling leading to significant errors in predictions. This means that high levels of quantum mechanical (QM) theory are required. In contrast, solid materials, normally occupy a well defined minimum energy configuration, removing the need for statistical sampling and reducing the sensitivity to minor errors in the predicted energies.

Given the challenging nature of this problem, significant approximations are almost universally used. One is to abandon the expensive quantum mechanical description, and instead employ classical molecular dynamics simulations (CMD). Alternatively, the complicated statistical description is abandoned and calculations using fixed ion–solvent clusters (ISC) of molecules are performed. Both of these methods can be useful in certain circumstances. However, despite their wide use, these methods have never been demonstrated to predict various basic benchmark properties of electrolyte solutions and cannot therefore be expected to provide quantitatively accurate predictions of realistic systems. This is a fundamental limitation of these approaches as they neglect key aspects of the system.

Research into systems involving electrolyte solutions has been held back by this fundamental challenge for decades. We believe a fundamentally new and different approach is long overdue. In this perspective we outline how exciting recent developments in the field of deep learning, specifically equivariant neural network potentials (NNPs), have the potential to finally provide such an approach. These methods enable long
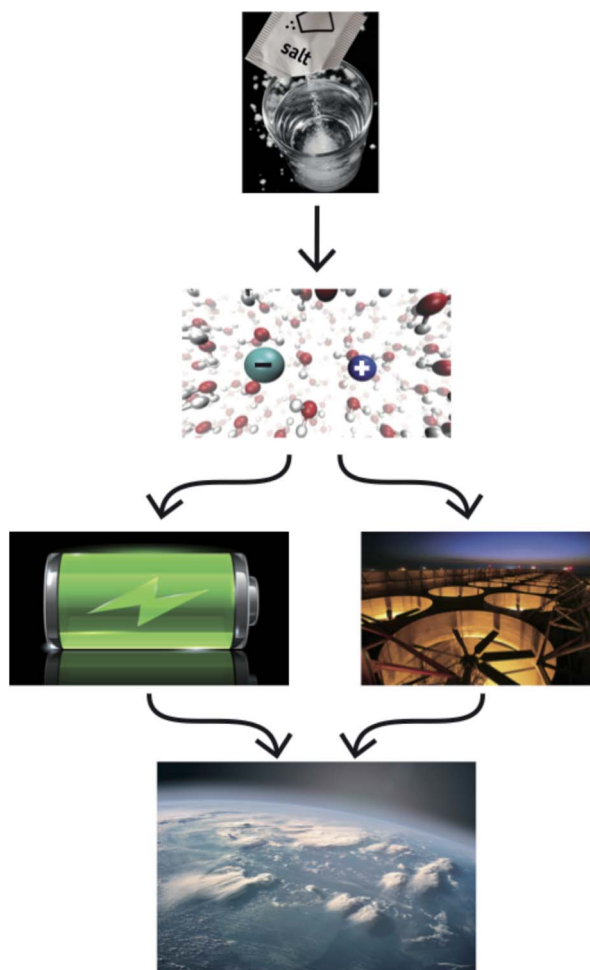
time and large spatial scale simulations of electrolyte solutions while maintaining a high level of quantum chemical accuracy. Before the utility of this method can be demonstrated with application to real systems of direct practical relevance, it is important to demonstrate its validity by reproducing key fundamental benchmark properties of electrolyte solutions. In this article, we provide such a demonstration, showing that the hydration structure of a sodium ion in water can be adequately reproduced. We also outline additional fundamental properties of electrolyte solutions that should also serve as key benchmarks. Specifically, we focus on predicting the radial distribution functions, chemical potentials, diffusivities and reaction rates. In addition to providing key benchmarks, these properties are also necessary inputs into large scale modelling of practically important systems. Their determination is therefore highly important in cases where these values are not experimentally available.

We believe this technique is approaching a level where it will soon have real predictive value for designing the next generation of materials for critical energy and sustainability applications.

2 Key properties

Firstly, we discuss in greater detail some of the key benchmark properties of electrolyte solutions that any approach should aim to predict quantitatively. Once these fundamental properties are reproduced using these methods, we can move on to more complex and practically relevant systems such as electrode–electrolyte interfaces.

2.1 Chemical potential

The chemical potential is one of the most fundamental properties of any electrolyte solution. The chemical potential of a given species is defined as the rate of free energy change with addition of that species to the system. In simplified form, it is given by:

$$\mu_i = \mu_i^0 + k_B T \ln c_i \gamma_i \qquad (1)$$

μ_i^0 is the chemical potential of a single ion in water at infinite dilution, or equivalently the solvation free energy, c_i is the concentration of the species i, and γ_i is the activity coefficient of the species i at the concentration c_i.[17] In essence, this quantity represents the stability of a particular species. It is the key thermodynamic property for determining chemical and phase equilibria, as it describes the tendency of a chemical species to change phase or undergo chemical reactions. For example, the chemical potential is required to predict salt solubilities and ion speciation.[18] Chemical potentials can be related to many other important properties of electrolyte solution systems such as redox potentials, standard electrode potentials, pK$_a$'s *etc.*[19]

2.2 Activity and osmotic coefficients

Activity coefficients describe the deviation of a given species from ideal (Raoultian) solution behaviour. They are closely

related to the osmotic coefficients, which capture the deviation from ideal behaviour of the osmotic pressure, another key thermodynamic property of electrolyte solutions. Osmotic coefficients can be converted from activity coefficients based on the Gibbs–Duhem equation.[18] Researchers have focused on developing theoretical methods to calculate thermodynamic properties such as activity/osmotic coefficients for over a century.[18] However, this problem remains largely unsolved. Essentially, all processes of practical interest occur at concentrations where the non-ideal behaviour captured by these coefficients can neither be neglected nor accurately estimated with classical theories. The only option currently to describe these effects is to use equations with parameters extensively fitted to experiment, such as the Pitzer equations.[20] However, this approach fails for any case that isn't well characterised, and in practice often these non-ideal effects are ignored entirely.

2.3 Radial distribution function

The radial distribution function (RDF) is another critical property that describes the structure of electrolyte solution. It corresponds to the normalised average density of atoms around a given reference atom. In general, it can be defined as the ratio of the averaged local density of particles at the distance r, to the bulk density of the particles, as follows:

$$g(r) = \rho(r)/\rho(\infty) \qquad (2)$$

where ρ is the density. The radial distribution function can be calculated directly from molecular dynamic simulations, or through integral equation approaches, such as the Ornstein–Zernike equation with a closure approximation. It can be used to connect microscopic information to macroscopic properties such as activity/osmotic coefficients and many other thermodynamic properties *via* Kirkwood–Buff theory.[21,22]

2.4 Potential of mean force

Changes in the free energy are critical for understanding the thermodynamics and kinetics of electrolyte solutions. The potential of mean force (PMF) describes the free energy of two species as a function of separation averaging over all other molecular configurations. For electrolyte solutions, the PMF can be calculated directly from the radial distribution function as:

$$w(r) = -k_B T \ln [g(r)] \qquad (3)$$

It can therefore be indirectly computed by Monte Carlo or molecular dynamic simulations. While in principle the PMF depends on concentration, its infinite dilution limit is a particularly important quantity that is useful for determining thermodynamic properties.[21,23]

2.5 Kinetic properties

While most equilibrium properties can be related to the chemical potential, kinetic properties can be more challenging to determine, particularly for cases involving rare events.

19562 | *J. Mater. Chem. A*, 2022, **10**, 19560–19571

This journal is © The Royal Society of Chemistry 2022

However, some important kinetic properties can be related to thermodynamic properties such as the chemical potential. For instance, activities can also be related to chemical reaction rates *via* the Brönstead–Bjerreum equation[24] and free energy barrier heights or even binding energies can often be a good predictor of kinetic processes.[2] Some kinetic properties, such as diffusion coefficients, can be determined directly from MD simulations when they are run sufficiently long. For rare-event problems, enhanced sampling techniques such as metadynamics or umbrella sampling can be employed.

3   Theoretical methods

Currently there is no well established method for quantitatively predicting any of the properties outlined above. It is important to understand the reason for this inadequacy of existing approaches in order to contextualise the usefulness for the latest deep learning advances.

A summary of the key points of this section is provided in Table 1.

3.1   Continuum solvent models (CSM)

Continuum solvent models (CSMs) or implicit solvent models are widely applied in modelling aqueous electrolyte solutions, especially for thermodynamic property prediction. While there are a diverse range of these models, they are defined by the fact that they treat the solvent as a continuous medium rather than being composed of explicit molecules. Compared with explicit solvent models, continuum solvent models have far fewer degrees of freedom, and hence much lower computational requirements. Continuum solvent models applied to electrolytes originated over a century ago with the work of several Nobel prize winning scientists.[18,25] Subsequently, computational methods were developed that enabled continuum solvent models to be combined with quantum mechanical calculations.[26] In continuum solvent models, a cavity is constructed with an appropriate shape and size that contains the solute. The key parameters needed for these models are the ionic cavity size and solvent dielectric constant. Born developed a model calculating the solvation free energy of ions with this approach,[25] which shows good qualitative agreement although it requires some improvements.[27] Debye and Hückel also developed a continuum solvent model for activity coefficients computation which works quantitatively at low concentrations.[18]

The use of these techniques is extremely common in materials science applications where particular implementations such as the Poisson–Boltzmann equation, PCM, SMD or COSMO[26] are most often used. Sophisticated implementations of these methods have enabled the study of electrolyte solutions at electrode interfaces.[28,29] The essential problem with these models is that they do not include realistic short range interactions of ions with other ions and solvent molecules. This limits their rigorous range of validity to very low concentrations. Some success has been made with building modified continuum solvent models with more realistic short range interactions to extend the theory to higher concentrations but some reliance on adjustable parameters invariably remains.[28,30,31]

3.2   Classical molecular dynamics (CMD)

Classical molecular dynamics (CMD) simulates molecules as explicit particles with charges to account for the electrostatic interactions; Lennard-Jones potential to characterise short range Pauli repulsion and medium range dispersion interactions; and harmonic potentials to capture chemical bonds. More sophisticated models have been developed to describe effects such as polarisation and charge transfer. The potential energy/forces are calculated and then Newton's second law is used to compute a trajectory of all the atoms as a function of time. These methods are now highly developed and applied in many scientific fields, such as material design and drug discovery. CMD simulations can be run for very large systems with relatively moderate computational cost without significant technical difficulty. However, their key limitation is that they require many parameters which can be very difficult to determine. These parameters have to be adjusted to reproduce properties such as solvation free energies and activity coefficients.[22,32] This limits the generalisability and usefulness of these models as they are only applicable to electrolyte solutions that are already well characterised experimentally. Furthermore, significant effort needs to be invested to optimise and adjust their parameters and generalising these parameters to new systems can lead to new issues. For example, parameters optimised for bulk may fail when used to simulate an interface.[33]

CMD can reproduce experimental properties such as activity and osmotic coefficients in some cases, if the parameters are carefully adjusted. However, almost invariably significant

Table 1   A summary of the different approaches to theoretical study of electrolyte solutions and a subjective assessment of their properties in the criteria of accuracy efficiency, simplicity and interpretability

Approach	Accuracy	Efficiency	Simplicity	Interpretability
Continuum solvent models (CSM)	Low	High	High	High
Classical molecular dynamics (CMD)	Moderate	Moderate	Moderate	Moderate
Ion–solvent clusters (ISC)	Moderate	Low	Moderate	Moderate
Ab initio molecular dynamics (AIMD)	High	Low	Moderate	Low
ISC/CSM	Moderate	Low	Moderate	Moderate
AIMD/CMD (a.k.a. QM/MM)	Moderate	Low	Low	Low
Neural network potential molecular dynamics (NNP-MD)	High	Moderate	Moderate	Low

This journal is © The Royal Society of Chemistry 2022

J. Mater. Chem. A, 2022, **10**, 19560–19571 | **19563**

deviations and breakdowns occur compared with experiments due to difficulties associated with the accurate determination of the parameters.[21–23]

Sophisticated many-body inter atomic potentials extensively fitted to high level quantum chemistry calculations on small clusters is a promising pathway currently being explored to overcome the challenge of parameter determination of classical methods.[34–36]

3.3 Ion–solvent clusters (ISCs)

In cases where reliable parameters do not exist, and where CSMs fail to provide sufficient accuracy, researchers often use quantum chemistry calculations on small ion–solvent clusters (ISCs).[37,38] These clusters are often embedded in a continuum solvent model. However, it is difficult to make definitive conclusions from this approach because the range of configurations the clusters can occupy is extremely large. This makes it very difficult to obtain a representative sample. As a result it is often unclear whether the conclusions drawn are reliable or an artefact of the particular cluster structures used. The additional effort associated with identifying the correct representative minimum energy structures and limited improvements in accuracy also limit the usefulness of this approach.

3.4 *Ab initio* molecular dynamics (AIMD)

Ab initio molecular dynamics (AIMD) calculates the energies and forces on the nuclei using quantum chemistry rather than with a classical force field. This means it bypasses the parameter determination problem of CMD. The key challenge for AIMD is finding approximations capable of solving the Schrödinger equation with sufficient accuracy and efficiency. DFT is the most common level of theory used. It usually relies on the Born–Oppenheimer approximation, which assumes that electronic motions and nuclear motions are separable. Although AIMD is much more expensive computationally than CMD, there is still a significant and growing amount of research into using AIMD to research structural properties of electrolyte solutions[39] and even a few cases where it has been used to study thermodynamic properties.[31,40,41] For example, AIMD has been used to demonstrate the iodide anion is not strongly adsorbed to the air–water interface, contradicting the predictions of CMD simulations.[33] AIMD has also been used to compute ion solvation energies accurately[41] and even activities in a moderate concentration range where like ion interactions can be neglected.[31] However, the high computational demand of this approach limits its practical use. Specifically, it is only available to researchers with access to large computational resources and can only be applied to small scale, simple properties. Additionally, even at a quite high level of theory some inaccuracies can remain that need to be corrected.[42] We are therefore stuck in a trade off between efficiency and accuracy reminiscent of the well known trade off between energy and power density observed for electrochemical energy storage systems as shown in Fig. 2.

These approaches can be combined in various ways such as QM/MM simulation where AIMD and CMD are combined[43] or QM/CM where AIMD and ISC models are combined.[44] While these methods are promising, the challenges associated with interfacing these different methods introduces significant additional complexity, requiring significant careful and detailed assessment.

3.5 Neural network potential molecular dynamics (NNP-MD)

The use of neural network potential molecular dynamics (NNP-MD) has the potential to allow large spatial and long temporal scale molecular simulations while maintaining a DFT level of accuracy. The underlying theoretical details of these methods are beyond the scope of this review and we refer the reader to several excellent reviews of these methods that have recently been published for more detail.[45–51] At a high level these methods work by defining a mapping between atomic coordinates to energies and forces (occasionally virial tensors also). This mapping contains a large number of parameters (weights and biases) that can be systematically adjusted to minimise the error on a set of training data, combined with an algorithm to systematically optimise the parameters (Backpropagation). The training data consists of a set of coordinates and their respective energies/forces. This training data can be obtained from short AIMD runs or from resampling of classical MD runs with DFT calculations. A standard approximation is to treat the total energy as the sum of individual atomic energies.

Once these parameters have been optimised the network can be used to predict the energies and forces, for any set of coordinates at a much lower computational cost than the original method. This enables much longer temporal and larger spatial scale simulations. Additionally, because the energy is a sum of individual atomic energies, these methods can be applied to simulate much larger scale systems than the training data.
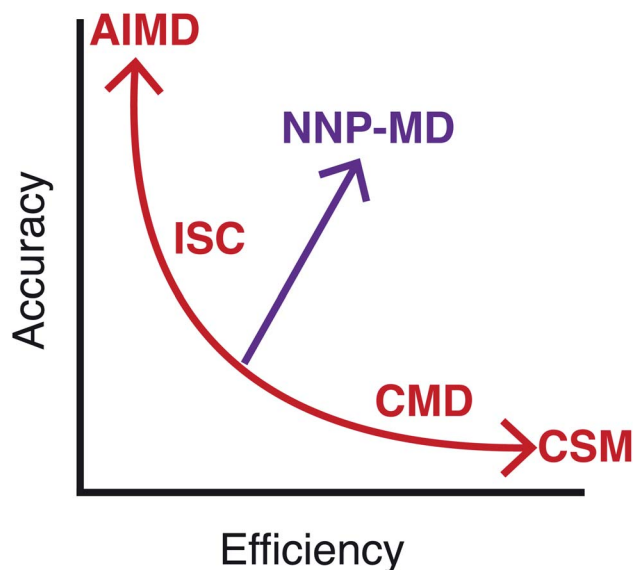

Fig. 2 A depiction of the fundamental accuracy/efficiency tradeoff between the different methods currently used to model electrolyte solutions and the potential for neural network potential molecular dynamics (NNP-MD) to breakthrough the accuracy/efficiency tradeoff.

19564 | *J. Mater. Chem. A*, 2022, **10**, 19560–19571

This journal is © The Royal Society of Chemistry 2022

These methods have already been scaled up to simulate simple systems containing billions of atoms.[52] They have also very recently been incorporated into an automated workflow to enable calculation of redox potentials, acidity constants, and solvation free energies.[53]

An additional key advantage of these approaches is that thermodynamic integration can be rigorously implemented relying only on a neural network trained on the initial and final states.[53,54] This is because the intermediate states can be created by a linear combination of the two force fields. Thermodynamic integration is a key tool for the calculation of the chemical potential and related properties so this is an important advantage.

3.5.1 Architectures. A standard (NNP) architecture converts the atomic coordinate information of some local cluster of atoms into a descriptor vector, these descriptors are then fed into a standard neural network that predicts a single energy value for each atom. These can be summed to compute the energy of the whole system, and differentiated to compute the force. DeePMD-kit[55] is an example of a software package, which implements this architecture. While this type of approach has shown promise in various applications[52,56–59,59–63] it can have large training data requirements, which limits its usefulness as the generation of sufficiently large training data is still computationally very demanding.

Therefore, a significant amount of research effort is currently focused on developing improved, more sophisticated architectures to describe the potential energy surface, which in principle should be able to reduce the amount of required training data.[64–74] Many of these approaches incorporate new and exciting architectures known as graph neural networks (GNNs) or message passing neural networks (MPNNs). These approaches normally represent each atom as a multidimensional feature vector, which is a function of the atomic number and is iteratively updated using information about the distances and feature vectors of surrounding atoms. These features are combined with more standard neural networks to determine atomic energies.

A crucial feature of these architectures is that they are carefully designed to maintain invariance to molecular rotation, translation and permutation. Recently, there has been an increased focus on building models that are additionally equivariant to rotations.[48,75,76] In essence, this assures that rotating the molecular configuration input into the neural network, results in equivalent rotations of the tensors associated with the atoms. An example of this is the recently developed Neural Equivariant Interatomic Potentials (NequIP) approach.[69] NequIP has been shown to perform with high accuracy based on a much smaller data set compared with DeepMD, *i.e.*, using three orders of magnitude less AIMD training data. Equivariance was demonstrated to be key to this improvement.

The specific speed up possible with these methods depends significantly on the particular methods and implementation used. In the application we demonstrate below, NequIP enables a speed up of approximately three orders of magnitude compared to the original AIMD method. While this is still slower than standard CMD, it still enables qualitatively new phenomena to be studied using only a single GPU. This is a rapidly developing field with many new approaches from the field of deep learning being incorporated and tested with promising results,[77] indicating that further improvement is likely possible.

These methods are now mature enough to be applied to compute structural and kinetic properties of many systems, particularly electrolyte solutions, at an *ab initio* level of accuracy. Below, we provide a simple demonstration of these methods by applying them to simulate the structure and diffusivity of a sodium ion in water. These properties can then be fed into well-established theories such as Kirkwood–Buff theory,[78] the modified Poisson–Boltzmann equation[28,79] and kinetic models to predict important practical properties. We demonstrated an example of such an approach for computing the activities/osmotic coefficients of aqueous sodium and chloride ions in a recent preprint.[80]

Once the reliability of this approach is firmly established it should be scaled up to many cases to contribute to a database of key properties of electrolyte solutions such as IonSolvR[81] to eventually improve and supplement the experimental databases relied on today. This crucial information can then be used to design and optimise electrolyte solutions for the many important materials science applications where they play a role. Additionally, it should be possible to use these techniques to directly simulate more complex phenomena such those that occur at the electrode–electrolyte interfaces.[82] For example, charge transfer of an ion with a battery electrode material. Additionally, the linear scaling characteristic of these methods allows very large scale systems to be studied.[74,83] Although, additional effort will be needed to incorporate long range electrostatic and Casimir–Lifshitz forces into these large scale simulations. These methods hold particular promise for electrolyte solutions with slower dynamics where the long equilibration time makes direct AIMD particularly challenging, such as the organic electrolytes used in batteries, as well as water in salt electrolytes and ionic liquids.

The scaling of deep learning methods using much larger models and training data sets has recently been demonstrated to be surprisingly impressive and effective in the fields of natural language processing and image recognition. The potential for similar gains in this field is also exciting and should be further explored.[84]

4 Example application

To demonstrate the promise of this approach, we outline the use of NequIP to compute the hydration structure and diffusivity of a sodium ion. As outlined above reproducing these basic features is the key first step toward predicting many more practically important properties of electrolyte solutions.

4.1 Workflow

Fig. 3 outlines a workflow for computing important properties of electrolyte solutions. First, high quality short AIMD

This journal is © The Royal Society of Chemistry 2022

J. Mater. Chem. A, 2022, **10**, 19560–19571 | **19565**

simulations are performed using software such as CP2K and the highest level of DFT functional feasible. In this case we use the strongly constrained and appropriate normalised functional (SCAN).[85] Secondly, calculations are performed on small clusters extracted from the AIMD simulation at a more accurate level of theory, in this case MP2, in order to confirm the reliability and accuracy of the DFT level of theory. Correction terms can potentially be derived if necessary. We have previously shown that SCAN can very reliably describe the sodium ion–water interaction so no correction for this term is required.[42] Next NequIP is used to train a NNP. This is then used to run much longer time scale NNP-MD simulations by using the NequIP interface with LAMMPS. Computational details are provided below. We train on a 2.4 M simulation of sodium chloride using the SCAN functional.

4.2 Sodium–oxygen RDF prediction

Fig. 4 shows the sodium oxygen RDF computed with the NequIP NNP, showing excellent agreement with pure AIMD results and rescaled XRD data previously reported.[42,86]

Most impressively only 2000 frames from a short AIMD simulation were used in the training of the NNP demonstrating the remarkably low training data requirements of this approach. There is some disagreement in the second solvation
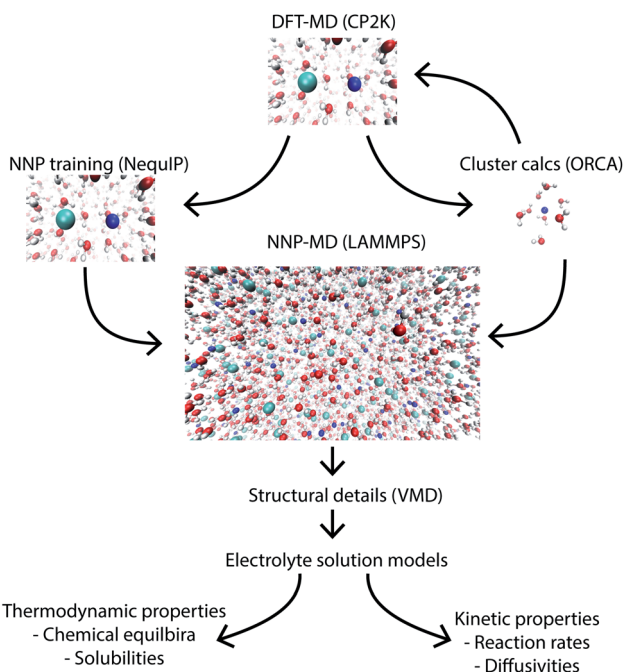

Fig. 3 A depiction of the proposed workflow to compute electrolyte solution properties. Starting with a short AIMD simulation a neural network potential is trained using NequIP to accurately and efficiently predict the energies and forces. Correction terms are determined using higher level quantum mechanical calculations on small clusters if necessary. These are combined with the neural network potential to perform large scale/long time scale molecular dynamics simulations. Structural and kinetic information can then be extracted from these simulations and be used in statistical mechanical theories to predict important practical properties such as chemical potentials.

layer compared with the original AIMD SCAN simulation. This is attributable to the fact that the SCAN RDF is not fully converged due to the significant computational expense associated with using this functional. This is demonstrated by the fact that the second peak agrees well with the better converged previously published simulation of Na–O RDF using the corrected revised Perdew–Becke–Ernzerhof with Grimme dispersion correction DFT functional (revPBE-D3-corr.)[87–89] shown in red.[39]

The longer time scale accessible with this method enables the calculation of properties that would otherwise be too difficult to converge with direct AIMD simulation such as diffusivities as shown in Fig. 5.

5 Challenges

There are several challenging problems which remain to be solved with this approach.

5.1 *Ab initio* accuracy

The first and perhaps most significant problem is that the DFT functionals used to generate the training data can have significant inaccuracies associated with them.[90–92] One potential solution to this problem is to add correction potentials adjusted to minimise this error using either higher level calculations on small clusters[39,42,93] or experimental information. Using machine learning to determine the corrections is also a possibility.[94] Alternatively, training a full NNP purely on data from small cluster calculations, which can be performed at higher levels, may also be feasible.[95] Another potential solution is employing a substantially higher level of theory such as the random phase approximation (RPA), second order Møller–Plesset (MP2) or double hybrids DFT functionals, which are now becoming feasible for periodic systems.[96–98] However, these
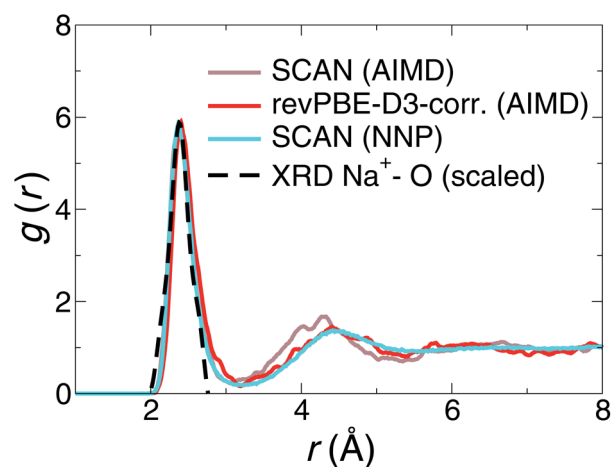

Fig. 4 Sodium oxygen radial distribution function predicted using a neural network potential showing excellent agreement with scaled X-ray diffraction data (black dashed), direct AIMD simulations using SCAN (brown) and with a corrected version of the revPBE-D3 functional (red).[87–89]
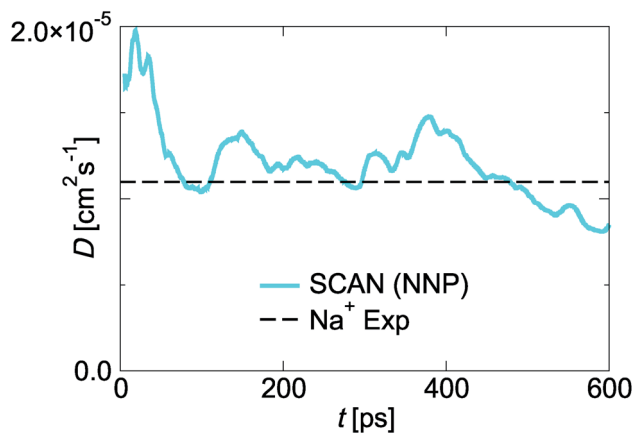
19566 | *J. Mater. Chem. A*, 2022, **10**, 19560–19571

This journal is © The Royal Society of Chemistry 2022

Fig. 5 Sodium ion diffusivity predicted using a neural network potential showing good experimental agreement.

levels of theory also require significantly larger basis sets, which can be a challenge.

5.2 Long range interactions

An additional problem is that most NNPs currently only use local structural information to determine energies. However, often long range interactions can be important.[99,100] In particular, the electrostatic interactions are long ranged.[101] Such long range interactions are likely to be particularly important for asymmetric systems or low concentrations. A lot of sophisticated work is focused on addressing this problem for complex systems.[102–106] But for electrolyte solutions these interactions are already known analytically, *i.e.*, a dielectric screened Coulomb interaction. It is therefore not clear that it is essential to capture them with a NNP. Delta learning can be applied instead, where everything except the long range interaction is learnt with the NNP, which is then added to the explicit long range interactions for full MD simulations.[80,107]

5.3 Stability

Finally, these methods can have stability issues which do not necessarily correlate with the error observed on a validation data set.[108] This can result in simulations crashing or making nonphysical or inaccurate predictions. This is normally attributable to insufficient training data. Using better architectures with smaller training data requirements is one way to address this problem. Alternatively, a larger and more diverse training data set can be built by using high temperature MD simulations to make sure it covers a larger range of configuration space. It is common to use randomly sampled data from an AIMD simulation to generate the training data. A better approach could be to bias this sampling towards low probability/high energy structures to counteract the thermodynamic bias to lower energy regions. Another option is the use of active learning which involves identifying structures where the NNP cannot provide reliable predictions and adding these to the training data set.[53,82,109–112] This approach can be used to significantly

lower training data requirements. A downside of this approach is that it requires repeating the NNP training process several times. It is also important to identify any underlying noise/errors in the training data itself which can lead the NNP to learn nonphysical behaviours.

5.4 Interpretability

Molecular dynamics simulation is already generally less interpretable than more traditional methods such as continuum solvent models, as the key explanatory information must be extracted from the large trajectories. NNP-MD and AIMD are generally even less interpretable than CMD as they cannot provide a partitioning into different intermolecular interactions. However, in our view it is more important to obtain accurate simulations from which interpretable information can be extracted from additional calculations such as energy decomposition analysis, rather than relying on directly interpretable but less accurate methods.

6 Other uses of deep learning

There are various other exciting potential uses of deep learning in this area. While these are less developed at this stage, they are also extremely promising. This could result in the near future with the entire simulation process outlined above being replaced with neural networks. For example, one promising approach is to replace the full QM calculations with a deep learning architecture to predict the energies. DM21 (ref. 113) and Orbnet[114] are two examples of this. DM21 has recently been used for this purpose to enable the simulation of water, although some issues remain.[115] The other method is to use deep learning to directly predict the molecular trajectories themselves rather than relying on direct solution of Newton's second law to predict the motion of the atoms.[116–118] Other exciting developments are more generalisable neural networks able to predict forces and energies for general classes of molecules such as ANI.[119]

7 Conclusion

The use of neural network potentials offers a pathway to break through the fundamental efficiency/accuracy trade off that has plagued the field of electrolyte solution modelling for decades. This approach, in the near term, should enable the prediction of many important properties of electrolyte solutions from first principles. The inability to do this has been a major factor preventing the use of first principles methods in the design and optimisation of electrolyte solutions for many important materials science applications where they play a fundamental role. Here, we have demonstrated a pathway to use these techniques to determine important fundamental properties of electrolyte solutions. We have also addressed some challenges this approach faces and outlined promising pathways to overcome these. This is a rapidly developing and exciting area of research that holds the promise to lead to significant advances in the field of materials science.

This journal is © The Royal Society of Chemistry 2022

J. Mater. Chem. A, 2022, **10**, 19560–19571 | **19567**

8 Computational details

The computational details for the SCAN and corrected revPBE-D3 functional AIMD simulations of sodium in water are provided in a previous publication.[39] We used Born–Oppenheimer *ab initio* molecular dynamics simulations within the constant volume NVT (at 300 K) ensemble using periodic boundary conditions, which are performed within the CP2K simulation suite (**https://www.cp2k.org**) containing the *QuickStep* module for the DFT calculations.[120] The D3 dispersion correction due to Grimme[89] was used for revPBE. A 0.5 fs time step was used. We used a double ζ basis set that has been optimized for the condensed phase[121] in conjunction with GTH pseudopotentials[122] using a 400 Ry cutoff for the auxiliary plane wave basis for the revPBE-D3 simulations and a 1200 Ry cutoff for the SCAN[85] simulations.[123,124] A Nosé–Hoover thermostat was attached to every degree of freedom to ensure equilibration.[125] The energies were accumulated for ≈ 12 ps after 3 ps of equilibration. The sodium and potassium simulations for revPBE-D3 consisted of one sodium ion in a box of 96 water molecules of fixed dimensions of 14.3 (ref. 3) Å^3 giving a density of 1 g cm^{-3}. The same settings were used for the corrected revPBE-D3 simulations with the exception that the multiple force evaluation option was used to combine the DFT forces with the pairwise forces computed using the FIST method.

8.1 NequIP

To train the NNP, we ran 2.4 M NaCl SCAN simulations at 300 K and 400 K in CP2K using the above details for 12 and 8 ps respectively. Forces and energies from 2000 frames extracted from these simulations were used to train the NNP with NequIP. 500 frames were held out as a validation set. An equal weighting on forces and energies was used in the default loss function.[69] We decrease the initial learning rate of 0.01 by a decay factor of 0.5 whenever the validation RMSE in the forces has not seen an improvement for three epochs. A radial cutoff distance of 5 Å was used. Two interaction blocks were used with the maximum l set to two each with 8 even scalars, vectors and tensors. Only even parity was used. All the other parameters were set to the defaults. The NequIP plugin for LAMMPS[126] was used to perform NVT simulations at 300 K for 600 ps. A Nosé–Hoover thermostat was attached to every degree of freedom to ensure equilibration.[125]

8.2 Diffusion coefficients calculation

Diffusion coefficients were computed from the mean squared displacements (MSD) of sodium ions in our NNP MD trajectories. This conversion was carried out using the following diffusion coefficient-MSD relationship:

$$D = \frac{\text{MSD}}{6t} \tag{4}$$

The results were finally adjusted by finite size corrections.[127] We have used the experimental value for the viscosity of pure water when determining the finite size correction. The experimental values compared with are for the sodium ion diffusivity in a 2.4 M NaCl solution.[128]

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 K. P. Gregory, G. R. Elliott, H. Robertson, A. Kumar, E. J. Wanless, G. B. Webber, V. S. J. Craig, G. G. Andersson and A. J. Page, *Phys. Chem. Chem. Phys.*, 2022, **24**, 12682–12718.

2 G. Åvall, J. Mindemark, D. Brandell and P. Johansson, *Adv. Energy Mater.*, 2018, **8**, 1703036.

3 R. Pohorecki and W. Moniuk, *Chem. Eng. Sci.*, 1988, **43**, 1677–1684.

4 D. W. Keith, G. Holmes, D. S. Angelo and K. Heidel, *Joule*, 2018, **2**, 1573–1594.

5 J. Resasco, L. D. Chen, E. Clark, C. Tsai, C. Hahn, T. F. Jaramillo, K. Chan and A. T. Bell, *J. Am. Chem. Soc.*, 2017, **139**, 11277–11287.

6 C. Ding, X. Zhou, J. Shi, P. Yan, Z. Wang, G. Liu and C. Li, *J. Phys. Chem. B*, 2015, **119**, 3560–3566.

7 A. Urban, D.-H. Seo and G. Ceder, *npj Comput. Mater.*, 2016, **2**, 16002.

8 S. Boothroyd, A. Kerridge, A. Broo, D. Buttar, J. Anwar, P. Chemistry and C. Physics, *Phys. Chem. Chem. Phys.*, 2018, **20**, 20981–20987.

9 K. P. Gregory, E. J. Wanless, G. B. Webber, V. S. J. Craig and A. J. Page, *Chem. Sci.*, 2021, **12**, 15007–15015.

10 Y. Huang, L. Zhao, L. Li, M. Xie, F. Wu and R. Chen, *Adv. Mater.*, 2019, **31**, 1808393.

11 P. M. May and D. Rowland, *J. Chem. Eng. Data*, 2017, **62**, 2481–2495.

12 S. Vaque Aura, J.-S. Roa Pinto, N. Ferrando, J.-C. de Hemptinne, A. ten Kate, S. Kuitunen, N. Diamantonis, T. Gerlach, M. Heilig, G. Becker and M. Brehelin, *J. Chem. Eng. Data*, 2021, **66**, 2976–2990.

13 T. T. Duignan and X. S. Zhao, *Ind. Eng. Chem. Res.*, 2021, **60**, 14948–14954.

19568 | *J. Mater. Chem. A*, 2022, **10**, 19560–19571

This journal is © The Royal Society of Chemistry 2022

14 M. A. Makeev and N. N. Rajput, *Curr. Opin. Chem. Eng.*, 2019, **23**, 58–69.

15 A. D. Sendek, B. Ransom, E. D. Cubuk, L. A. Pellouchoud, J. Nanda and E. J. Reed, *Adv. Energy Mater.*, 2022, 2200553.

16 X. Qu, A. Jain, N. N. Rajput, L. Cheng, Y. Zhang, S. P. Ong, M. Brafman, E. Maginn, L. A. Curtiss and K. A. Persson, *Comput. Mater. Sci.*, 2015, **103**, 56–67.

17 A. Ben-Naim, *J. Phys. Chem.*, 1978, **82**, 792–803.

18 R. A. Robinson and R. H. Stokes, *Electrolyte solutions*, Butterworth & Co., Devon, 1959.

19 P. Hünenberger and M. Reif, *Single-ion solvation: experimental and theoretical approaches to elusive thermodynamic quantities*, The Royal Society of Chemistry, 2011.

20 K. S. Pitzer, *J. Phys. Chem.*, 1973, **77**, 268–277.

21 I. Kalcher and J. Dzubiella, *J. Chem. Phys.*, 2009, **130**, 134507.

22 M. Fyta and R. R. Netz, *J. Chem. Phys.*, 2012, **136**, 124103.

23 L. Vrbka, M. Lund, I. Kalcher, J. Dzubiella, R. R. Netz and W. Kunz, *J. Chem. Phys.*, 2009, **131**, 154109.

24 S. R. Logan, *Trans. Faraday Soc.*, 1967, **63**, 3004–3008.

25 V. M. Born, *Z. Phys.*, 1920, **1**, 45–48.

26 J. M. Herbert, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1519.

27 T. T. Duignan, D. F. Parsons and B. W. Ninham, *J. Phys. Chem. B*, 2013, **117**, 9421–9429.

28 K. Schwarz and R. Sundararaman, *Surf. Sci. Rep.*, 2020, **75**, 100492.

29 M. M. Melander, *Curr. Opin. Electrochem.*, 2021, **29**, 100749.

30 T. T. Duignan, D. F. Parsons and B. W. Ninham, *Phys. Chem. Chem. Phys.*, 2014, **16**, 22014–22027.

31 T. T. Duignan, M. D. Baer and C. J. Mundy, *Curr. Opin. Colloid Interface Sci.*, 2016, **23**, 58–65.

32 D. Horinek, S. I. Mamatkulov and R. R. Netz, *J. Chem. Phys.*, 2009, **130**, 124507.

33 M. D. Baer and C. J. Mundy, *J. Phys. Chem. Lett.*, 2011, **2**, 1088–1093.

34 G. R. Medders, V. Babin and F. Paesani, *J. Chem. Theory Comput.*, 2014, **10**, 2906–2910.

35 D. Zhuang, M. Riera, G. K. Schenter, J. L. Fulton and F. Paesani, *J. Phys. Chem. Lett.*, 2019, **10**, 406–412.

36 A. Caruso and F. Paesani, *J. Chem. Phys.*, 2021, **155**, 064502.

37 V. S. Bryantsev, M. S. Diallo and W. A. Goddard, *J. Phys. Chem. B*, 2008, **112**, 9709–9719.

38 L. Tomanik, E. Muchova and P. Slavicek, *Phys. Chem. Chem. Phys.*, 2020, **22**, 22357–22368.

39 T. T. Duignan, G. Schenter, C. J. Mundy and X. S. Zhao, *J. Chem. Theory Comput.*, 2020, **16**, 5401–5409.

40 K. Leung, S. B. Rempe and O. A. von Lilienfeld, *J. Chem. Phys.*, 2009, **130**, 204507.

41 T. T. Duignan, M. D. Baer, G. K. Schenter and C. J. Mundy, *Chem. Sci.*, 2017, **8**, 6131–6140.

42 T. Duignan, G. K. Schenter, J. Fulton, T. Huthwelker, M. Balasubramanian, M. Galib, M. D. Baer, J. Wilhelm, J. Hutter, M. D. Ben, X. S. Zhao and C. J. Mundy, *Phys. Chem. Chem. Phys.*, 2020, **22**, 10641–10652.

43 J. Chen, J. Kato, J. B. Harper, Y. Shao and J. Ho, *J. Phys. Chem. B*, 2021, **125**, 9304–9316.

44 A. Muralidharan, L. Pratt, M. Chaudhari and S. Rempe, *Chem. Phys. Lett.: X*, 2019, **4**, 100037.

45 A. C. Mater and M. L. Coote, *J. Chem. Inf. Model.*, 2019, **59**, 2545–2559.

46 F. Noé, A. Tkatchenko, K.-R. Müller and C. Clementi, *Annu. Rev. Phys. Chem.*, 2020, **71**, 361–390.

47 J. Behler, *Chem. Rev.*, 2021, **121**, 10037–10072.

48 A. D. White, *Living J. Comp. Mol. Sci.*, 2022, **3**, 1499.

49 E. Kocer, T. W. Ko and J. Behler, *Annu. Rev. Phys. Chem.*, 2022, **73**, 163–186.

50 T. Wen, L. Zhang, H. Wang, W. E and D. J. Srolovitz, *Mater. Futur.*, 2022, **1**, 022601.

51 A. Kolluru, M. Shuaibi, A. Palizhati, N. Shoghi, A. Das, B. Wood, C. L. Zitnick, J. R. Kitchin and Z. W. Ulissi, 2022, arXiv:2206.02005.

52 Z. Guo, D. Lu, Y. Yan, S. Hu, R. Liu, G. Tan, N. Sun, W. Jiang, L. Liu, Y. Chen, L. Zhang, M. Chen, H. Wang and W. Jia, 2022, arXiv:2201.01446.

53 F. Wang and J. Cheng, *J. Chem. Phys.*, 2022, **157**, 024103.

54 R. Jinnouchi, F. Karsai and G. Kresse, *Phys. Rev. B*, 2020, **101**, 060201.

55 H. Wang, L. Zhang, J. Han and W. E, *Comput. Phys. Commun.*, 2018, **228**, 178–184.

56 M. Hellström and J. Behler, *Phys. Chem. Chem. Phys.*, 2017, **19**, 82–96.

57 B. Onat, E. D. Cubuk, B. D. Malone and E. Kaxiras, *Phys. Rev. B*, 2018, **97**, 1–9.

58 M. Galib and D. T. Limmer, *Science*, 2021, **371**, 921–925.

59 L. Zhang, H. Wang, R. Car and E. Weinan, *Phys. Rev. Lett.*, 2021, **126**, 236001.

60 Y. Shi, C. C. Doyle and T. L. Beck, *J. Phys. Chem. Lett.*, 2021, **12**, 10310–10317.

61 C. Zhang, S. Yue, A. Z. Panagiotopoulos, M. L. Klein and X. Wu, *Nat. Commun.*, 2022, **13**, 822.

62 C. Malosso, L. Zhang, R. Car, S. Baroni and D. Tisi, *npj Comput. Mater.*, 2022, **8**, 139.

63 Y. Shi, S. T. Lam and T. L. Beck, *Chem. Sci.*, 2022, **13**, 8265–8273.

64 J. Klicpera, F. Becker and S. Günnemann, 2021, arXiv:2106.08903.

65 K. Atz, F. Grisoni and G. Schneider, *Nat. Mach. Intell.*, 2021, **3**, 1023–1032.

66 M. Haghighatlari, J. Li, X. Guan, O. Zhang, A. Das, C. J. Stein, F. Heidar-Zadeh, M. Liu, M. Head-Gordon, L. Bertels, H. Hao, I. Leven and T. Head-Gordon, *Digit. Discov.*, 2022, **1**, 333–343.

67 Z. Li, K. Meidani, P. Yadav and A. B. Farimani, *J. Chem. Phys.*, 2022, **156**, 144103.

68 C. W. Park, M. Kornbluth, J. Vandermause, C. Wolverton, B. Kozinsky and J. P. Mailoa, *npj Comput. Mater.*, 2021, **7**, 73.

69 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nat. Commun.*, 2022, **13**, 2453.

70 P. Tholke and G. D. Fabritiis, 2022, arXiv:2202.02541.

This journal is © The Royal Society of Chemistry 2022

J. Mater. Chem. A, 2022, **10**, 19560–19571 | **19569**

71 I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner and G. Csányi, 2022, arxiv:2206.07697.

72 A. Musaelian, S. Batzner, A. Johansson, L. Sun, C. J. Owen, M. Kornbluth and B. Kozinsky, 2022, arXiv:2204.05249.

73 H. E. Sauceda, L. E. Gálvez-González, S. Chmiela, L. O. Paz-Borbón, K.-R. Müller and A. Tkatchenko, *Nat. Commun.*, 2022, **13**, 3733.

74 A. Johansson, Y. Xie, C. J. Owen, J. S. Lim, L. Sun, J. Vandermause and B. Kozinsky, arXiv:2204.12573, 2022.

75 N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff and P. Riley, 2018, arxiv:1802.08219.

76 V. G. Satorras, E. Hoogeboom and M. Welling, 2021, arXiv:2102.09844.

77 Y.-l. Liao and T. Smidt, 2022, arXiv:2206.11990.

78 D. M. Rogers, *J. Chem. Phys.*, 2018, **148**, 054102.

79 T. T. Duignan, *J. Colloid Interface Sci.*, 2021, **600**, 338–343.

80 J. Pagotto, J. Zhang and T. T. Duignan, *ChemRxiv*, 2022, DOI: **10.26434/chemrxiv-2022-jndlx**.

81 K. P. Gregory, G. R. Elliott, E. J. Wanless, G. B. Webber and A. J. Page, *Sci. Data*, 2022, **9**, 430.

82 C. Schran, F. L. Thiemann, P. Rowe, E. A. Müller, O. Marsalek and A. Michaelides, *Proc. Natl. Acad. Sci. U. S. A.*, 2021, **118**, e2110077118.

83 D. Lu, H. Wang, M. Chen, L. Lin, R. Car, W. E, W. Jia and L. Zhang, *Comput. Phys. Commun.*, 2021, **259**, 107624.

84 N. C. Frey, R. Soklaski, S. Axelrod, S. Samsi, R. Gómez-Bombarelli, C. W. Coley and V. Gadepally, *ChemRxiv*, 2022, DOI: **10.26434/chemrxiv-2022-3s512**.

85 J. Sun, A. Ruzsinszky and J. P. Perdew, *Phys. Rev. Lett.*, 2015, **115**, 036402.

86 M. Galib, M. D. Baer, L. B. Skinner, C. J. Mundy, T. Huthwelker, G. K. Schenter, C. J. Benmore, N. Govind and J. L. Fulton, *J. Chem. Phys.*, 2017, **146**, 084504.

87 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

88 Y. Zhang and W. Yang, *Phys. Rev. Lett.*, 1998, **80**, 890.

89 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

90 M. Riera, A. W. Götz and F. Paesani, *Phys. Chem. Chem. Phys.*, 2016, **18**, 30334–30343.

91 M. J. Gillan, D. Alfè and A. Michaelides, *J. Chem. Phys.*, 2016, **144**, 130901.

92 M. Galib, G. K. Schenter, C. J. Mundy, N. Govind and J. L. Fulton, *J. Chem. Phys.*, 2018, **149**, 124503.

93 T. T. Duignan, *ChemRxiv*, 2021, DOI: **10.26434/chemrxiv-2021-7jxpq-v2**.

94 R. T. McGibbon, A. G. Taube, A. G. Donchev, K. Siva, F. Hernández, C. Hargus, K. H. Law, J. L. Klepeis and D. E. Shaw, *J. Chem. Phys.*, 2017, **147**, 161725.

95 V. Zaverkin, D. Holzmüller, R. Schuldt and J. Kästner, *J. Chem. Phys.*, 2022, **156**, 114103.

96 M. Del Ben, M. Schönherr, J. Hutter and J. VandeVondele, *J. Phys. Chem. Lett.*, 2013, **4**, 3753–3759.

97 Y. Yao and Y. Kanai, *J. Phys. Chem. Lett.*, 2021, **12**, 6354–6362.

98 F. Stein and J. Hutter, *J. Chem. Phys.*, 2022, **156**, 074107.

99 S. Yue, M. C. Muniz, M. F. Andrade, L. Zhang, R. Car and A. Z. Panagiotopoulos, *J. Chem. Phys.*, 2021, **154**, 034111.

100 C. G. Staacke, H. H. Heenen, C. Scheurer, G. Csányi, K. Reuter and J. T. Margraf, *ACS Appl. Energy Mater.*, 2021, **4**, 12562–12569.

101 A. Gao, R. C. Remsing and J. D. Weeks, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 1293–1302.

102 S. A. Ghasemi, A. Hofstetter, S. Saha and S. Goedecker, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2015, **92**, 045131.

103 S. P. Niblett, M. Galib and D. T. Limmer, *J. Chem. Phys.*, 2021, **155**, 164101.

104 T. W. Ko, J. A. Finkler, S. Goedecker and J. Behler, *Nat. Commun.*, 2021, **12**, 398.

105 A. Gao and R. C. Remsing, *Nat. Commun.*, 2021, **13**, 1572.

106 L. Zhang, H. Wang, M. C. Muniz, A. Z. Panagiotopoulos, R. Car and W. E, *J. Chem. Phys.*, 2022, **156**, 124107.

107 G. P. Pun, R. Batra, R. Ramprasad and Y. Mishin, *Nat. Commun.*, 2019, **10**, 2339.

108 S. Stocker, J. Gasteiger, F. Becker, S. Günnemann and J. T. Margraf, *ChemRxiv*, 2022, DOI: **10.26434/chemrxiv-2022-mc4gb**.

109 L. Zhang, D. Y. Lin, H. Wang, R. Car and E. Weinan, *Phys. Rev. Mater.*, 2019, **3**, 023804.

110 Y. Zhang, H. Wang, W. Chen, J. Zeng, L. Zhang, H. Wang and W. E, *Comput. Phys. Commun.*, 2020, **253**, 107206.

111 Y. Zhai, A. Caruso, S. Gao and F. Paesani, *J. Chem. Phys.*, 2020, **152**, 144103.

112 N. Wilson, D. Willhelm, X. Qian, R. Arróyave and X. Qian, *Comput. Mater. Sci.*, 2022, **208**, 111330.

113 J. Kirkpatrick, B. Mcmorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. R. Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-sánchez, D. Hassabis and A. J. Cohen, *Science*, 2021, **374**, 1385–1389.

114 Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby and T. F. Miller, *J. Chem. Phys.*, 2020, **153**, 124111.

115 E. Palos, E. Lambros, S. Dasgupta and F. Paesani, *J. Chem. Phys.*, 2022, **156**, 161103.

116 P. R. Vlachas, J. Zavadlav, M. Praprotnik and P. Koumoutsakos, *J. Chem. Theory Comput.*, 2022, **18**, 538–549.

117 X. Fu, T. Xie, N. J. Rebello, B. D. Olsen and T. Jaakkola, 2022, arXiv:2204.10348.

118 L. Winkler, K.-R. Müller and H. E. Sauceda, *Mach. Learn. Sci. Technol.*, 2022, **3**, 025011.

119 J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192–3203.

120 J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.

121 J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, 114105.

122 S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 1703–1710.

123 G. Miceli, J. Hutter and A. Pasquarello, *J. Chem. Theory Comput.*, 2016, **12**, 3456–3462.

19570 | *J. Mater. Chem. A*, 2022, **10**, 19560–19571

This journal is © The Royal Society of Chemistry 2022

124 Y. Yao and Y. Kanai, *J. Chem. Theory Comput.*, 2018, **14**, 884–893.

125 G. J. Martyna, M. L. Klein and M. Tuckerman, *J. Chem. Phys.*, 1992, **97**, 2635–2643.

126 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

127 I.-C. Yeh and G. Hummer, *J. Phys. Chem. B*, 2004, **108**, 15873–15879.

128 V. Vitagliano and P. A. Lyons, *J. Am. Chem. Soc.*, 1956, **78**, 1549–1552.