THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

High-Throughput Aqueous Electrolyte Structure Prediction Using IonSolvR and Equivariant Graph Neural Network Potentials

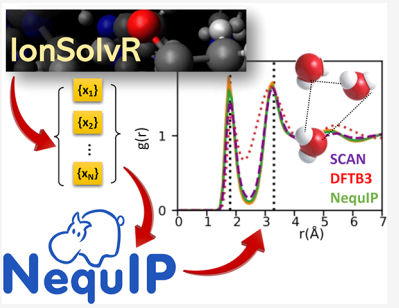
Sophie Baker, Joshua Pagotto, Timothy T. Duignan, and Alister J. Page*

Cite This: *J. Phys. Chem. Lett.* 2023, 14, 9508−9515

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Neural network potentials have recently emerged as an efficient and accurate tool for accelerating *ab initio* molecular dynamics (AIMD) in order to simulate complex condensed phases such as electrolyte solutions. Their principal limitation, however, is their requirement for sufficiently large and accurate training sets, which are often composed of Kohn−Sham density functional theory (DFT) calculations. Here we examine the feasibility of using existing density functional tight-binding (DFTB) molecular dynamics trajectory data available in the IonSolvR database in order to accelerate the training of E(3)-equivariant graph neural network potentials. We show that the solvation structure of Na^+ and Cl^- in aqueous NaCl solutions can be accurately reproduced with remarkably small amounts of data (i.e., 100 MD frames). We further show that these predictions can be systematically improved further via an embarrassingly parallel resampling approach.

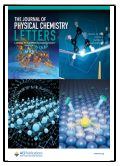


Water and aqueous electrolytes play a key role in a range of geological, biological, and chemical systems,[1] and hence, water is often referred to as the universal solvent.[2] Aqueous electrolytes such as NaCl(aq) are particularly important in the human body, which contains ∼0.8 mol salt.[3] As a result, the structures of water and aqueous electrolytes have been extensively studied over the last few decades[4] using experimental techniques such as X-ray[5,6] and neutron diffraction[5,7−9] as well as infrared spectroscopy.[6] Computational techniques, particularly molecular dynamics (MD), are also widely used.[10−14]

MD describes how the structural and dynamic properties of a system emerge and evolve over time by discretely propagating Newton's classical equations of motion for individual nuclei.[15] While classical MD force fields have been used extensively for the simulation of electrolyte systems, they rely on many parameters which invariably must be fitted to reproduce experimental data, limiting the utility of this approach. The addition of extra terms to describe more complex effects, such as polarizability, compounds this problem. *Ab initio* MD (AIMD) is an alternative whereby nuclear dynamics is propagated on the Born−Oppenheimer potential energy surface.[16] AIMD enables polarization effects to be described as well as bond making/breaking without fitted parameters (in principle). Furthermore, it has been applied to a variety of electrolyte solutions, yielding good agreement with experimental results.[10,17,18] However, the primary limitation of AIMD is its large computational cost, considering it almost exclusively relies on density functional theory (∼2−3 orders of magnitude slower than classical MD, depending on the exchange−correlation functional used[19]). This limits the simulations that can be computed with AIMD in terms of both system size and time scale; typically AIMD simulations are currently restricted to only hundreds of atoms and picosecond time scales.[20] Nevertheless, reliable ion−ion radial distribution functions in electrolyte solutions can require more than 1 ns of simulation time due to the low number of ions in typical simulation model systems and excessive configurational sampling necessitated by low ion diffusion rates.[12] Tight-binding approaches, such as density functional tight-binding (DFTB), offer a compromise in this respect but do require some parametrization.[13]

One solution that has recently emerged to overcome this impasse is the use of machine learning techniques, such as neural network potentials (NNPs),[20−24] which are highly flexible functions that map atomic coordinates to electronic energies and nuclear forces. A promising NNP method is the deep potential molecular dynamics (DeepMD) scheme,[22] which can reproduce the structure and properties of several systems, including molten[25−27] and aqueous[28−32] electrolytes. While DeepMD reduces the computation cost compared to AIMD trajectories by several orders of magnitude and scales linearly with system size,[22] it still requires considerable amounts of AIMD data for training (e.g., the non-smooth version of DeepMD required upward of 133,000 individual MD frames to train potentials for liquid water and ice[22]). This has been improved with the development of the DeepMD-smooth edition (DeepPot-SE),[33] which used ∼31,000 frames

Received: June 30, 2023
Accepted: October 12, 2023
Published: October 16, 2023



https://doi.org/10.1021/acs.jpclett.3c01783
J. Phys. Chem. Lett. 2023, 14, 9508−9515

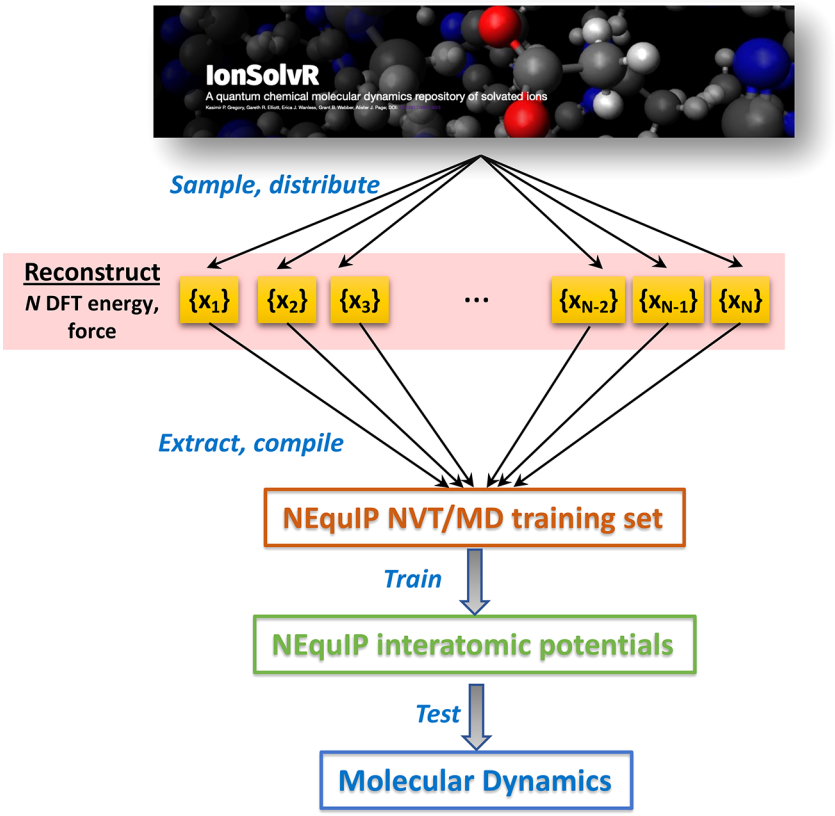


Figure 1. IonSolvR resampling and the NNP generation algorithm. Randomly chosen MD frames $\{x_i\}$ are first extracted from the full MD trajectory in IonSolvR. Energies and forces are then reconstructed using m-GGA DFT; as each frame is independent, this step is embarrassingly parallel. Energies and forces are then extracted and collated for training using NequIP. The IonSolvR logo is reproduced under CC BY 4.0.

to train a model to accurately predict the phase diagram of water and has been shown to reproduce other systems with small training sets, such as the GaP(110)−water interface, which required only 300 initial frames and 3257 frames after active learning.[34] Several other approaches to solving this problem have been reported in the literature. For instance, the DeePKS (Deep Kohn−Sham) method[35] is a neural network DFT model whereby a neural network energy functional, trained on a small number of AIMD frames, corrects a lower level of theory to one that approximates Kohn−Sham DFT. DeePKS accurately reproduces the structure of water and NaCl(aq) at various concentrations and pressures obtained using AIMD.[23] Promising approaches integrating DeepMD with concurrent learning strategies, such as DP-Gen,[36] have also been reported.

An alternative approach is the Neural Equivariant Interatomic Potentials (NequIP) method,[24] which is an E(3)-equivariant graph neural network approach that is trained on AIMD trajectories to construct interatomic potentials. Equivariance means that the features used in the model can be transformed under the symmetry operations of rotation, translation, and reflection; as a result, NequIP requires several orders of magnitude less training data compared to older NNP models for condensed-phase systems. For example, for ice and liquid water, NequIP yielded forces that were at least as good as those obtained using DeepMD despite being trained on only 133 AIMD frames.[24] Additionally, Duignan et al.[37] demonstrated that NequIP accurately predicts structure and ion diffusion in aqueous sodium chloride solutions compared to AIMD and experiment, respectively, with small training data requirements.

Nevertheless, NequIP, along with all NNP models, still requires the prior generation of training data through various techniques, such as random displacement of atoms, farthest-point sampling, active learning, and AIMD trajectory data. Each of these alternatives, ironically, takes a considerable amount of computational resources and time due to their reliance on density functional and/or wave function theory calculations. Circumventing this time constraint was one of the motivations behind the generation of the IonSolvR database by Gregory et al.,[13] a collection of more than 3000 AIMD nanosecond time scale simulations of aqueous and nonaqueous electrolyte solutions at variable concentrations, constructed from fast DFTB simulations.[38] Despite the recent emergence of concurrent learning strategies for the generation of machine learning potentials, using precomputed quantum-chemical MD trajectories as training datasets remains a useful strategy considering the abundance of trajectory data that have no-doubt accumulated over the last several decades. Herein we examine the feasibility of using the IonSolvR database for the generation of NNPs for aqueous electrolyte solutions. Using NaCl(aq) as an exemplar, we compare NNPs constructed using raw IonSolvR trajectory data and energy/force data calculated using both DFTB and meta-GGA (SCAN[39]) density functional theory, calculated using a high-throughput resampling approach (Figure 1).[40,41]

To assess whether NequIP NNPs can be generated reliably using raw IonSolvR trajectory data, we begin our discussion by assessing the accuracy of NNPs trained directly on the

https://doi.org/10.1021/acs.jpclett.3c01783
J. Phys. Chem. Lett. 2023, 14, 9508−9515

DFTB3/3ob energies and nuclear forces for structures contained in the IonSolvR database. Figure 2 compares radial distribution functions (RDFs) for the principal atomic interactions in NaCl(aq) obtained by using NNP-MD data generated in this work and those in the IonSolvR trajectory. In general, the DFTB-based NNPs (i.e., $NNP_{DFTB\text{-}500,1000,2000}$) accurately reproduce the key interactions observed in NaCl(aq) using DFTB/3ob. For instance, the principal O−H interactions in the first and second solvation shells found at ∼1.8 and ∼3.3 Å in the IonSolvR training data are in excellent agreement with those obtained using the NNPs generated here in both intensity and peak distance (Figure 2a). Notably, this agreement is obtained with only 500 random frames from the IonSolvR trajectory, and increasing the size of the dataset (i.e., $NNP_{DFTB\text{-}1000,2000}$) does not improve this agreement appreciably. Similarly, Figure 2b shows excellent agreement in the O−O interactions (at ∼2.8, 5.5, and 7 Å) predicted using these NNPs and those observed in the IonSolvR trajectory. The solvation environments of the Na^+ and Cl^- ions are considered in Figure 2c,d, respectively. There is excellent agreement between the NNP predictions and DFTB/3ob data for the Na^+−O interaction, which corresponds to two peaks in the Na^+−O $g(r)$ at ∼2.5 and 4.1 Å, respectively. These are closely reproduced by all NNPs in terms of both peak distance and intensity. As observed for the surrounding water structure, the size of the training set here does little to influence this agreement. On the other hand, the DFTB/3ob RDF for the Cl^-−H interaction, shown in Figure 2d, consists of two predominant peaks at ∼2.3 Å and 3.6 Å and a more diffuse long-range interaction at ∼5.6 Å. There is notable deviation between the predictions obtained using the NNPs reported here and the IonSolvR dataset for this RDF. $NNP_{DFTB\text{-}500}$ fails to accurately describe the 2.3 and 3.6 Å interactions, although the longer-range interaction at ∼5.6 Å is reproduced. This indicates that the solvation environment of the Cl^- anion observed using these NNPs is noticeably less structured than that observed in the original DFTB3/3ob dataset. Increasing the NNP training set here does not yield a consistent improvement in this respect.

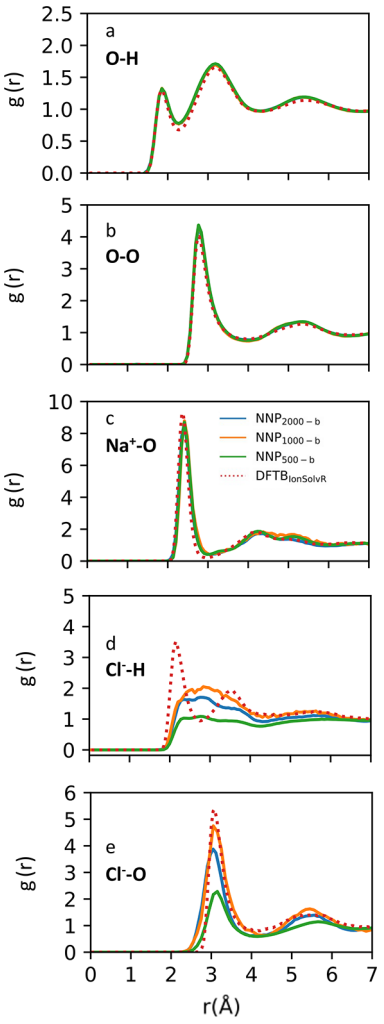


Figure 2. Comparison of partial radial distribution functions, $g(r)$, for the principal interactions in 0.8 M NaCl(aq) solution obtained using $NNP_{DFTB\text{-}500}$, $NNP_{DFTB\text{-}1000}$, and $NNP_{DFTB\text{-}2000}$ and those observed in the IonSolvR DFTB/3ob trajectory: (a) O−H, (b) O−O, (c) Na^+−O, (d) Cl^--H, and (e) Cl^-−O. All NNPs were trained using a training:test ratio of 80%:20%.

The origins of the deviations in NNP predictions for Cl^- interactions is considered further in Figure S1, which shows that each NNP generated here reliably reproduces DFTB/3ob forces for all atoms in the randomly sampled dataset. Interestingly, the Cl^-−O oxygen interaction (Figure 2e) predicted using NequIP NNPs and DFTB3/3ob are in good agreement, suggesting that these issues are limited to predominantly short-range interactions involving Cl^- (such as electrostatic interactions and hydrogen bonding).

The preceding discussion has demonstrated that generating NNPs using NequIP trained on IonSolvR trajectory data yields an accurate description of the Na^+ solvation structure in an aqueous NaCl solution. The issues associated with the solvation of the Cl^- anion seen in Figure 2d are not a reflection of NequIP's ability to recreate forces but instead a result of the DFTB3/3ob forces themselves pertaining to the Cl−H interaction (Figure S1 clearly shows that $NNP_{DFTB\text{-}500}$ is able to reliably recreate the DFTB forces it was trained from for each atom type). This result provides the motivation for examining whether the forces provided in the IonSolvR database can be systematically improved by using a more accurate description of the Born−Oppenheimer potential energy surface. SCAN[39] is chosen here for this purpose, as it provides the necessary level of accuracy without incurring excessive computational cost and has been shown to outperform other GGA and meta-GGA functionals for bulk water,[42] Na^+(aq),[17] and NaCl(aq).[42]

Figure 3 compares the structural predictions of NaCl(aq) solution obtained using NNPs trained on IonSolvR+mGGA datasets with those obtained using AIMD with SCAN/DZVP[17] and neutron diffraction.[8] The remainder of our discussion will be based on NNPs trained using an optimal training:test ratio of 80%:20% (Figure S2 presents further details on the impact of the training:test ratio). The height of this peak can vary significantly (e.g., between 5 and 25) depending on solution conditions; for instance, changing concentration causes an increase in the first peak and a decrease in the second and third,[11,12] while increasing temperature generally causes the peak height to decrease. Thus, recreating experimental peak heights is less important than recreating the peak distances. We therefore gauge the accuracy of our results here based on the length scale of this interaction rather than the peak heights.

https://doi.org/10.1021/acs.jpclett.3c01783
J. Phys. Chem. Lett. 2023, 14, 9508−9515

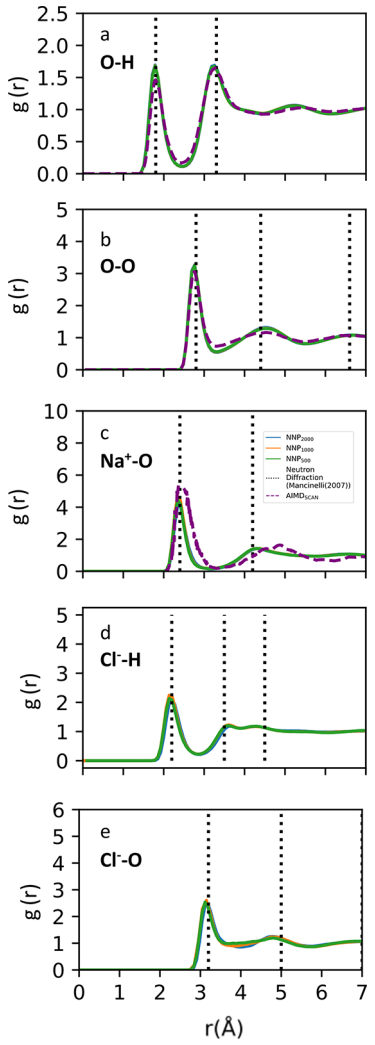


Figure 3. Partial radial distribution functions, $g(r)$, for the principal interactions in 0.8 M NaCl(aq) solution obtained from NNP-MD simulations trained using IonSolvR+mGGA datasets: (a) O−H, (b) O−O, (c) Na^+−O, (d) Cl^-−H and (e) Cl^-−O. All NNPs were trained using a training:test ratio of 80%:20%. $AIMD_{SCAN}$ data are SCAN/DZVP data from ref 17.

RDFs for the O−H interaction obtained from neutron diffraction experiments[9] (Figure 3a) show a peak at ∼1.8 Å, which corresponds to the principal hydrogen-bonding interaction in water, and a second peak at ∼3.3 Å, which corresponds to the second solvation shell. SCAN/DZVP accurately reproduces the length scales of both interactions. $NNP_{SCAN\text{-}500,1000,2000}$ predictions are all in excellent agreement with SCAN/DZVP, aside from marginally overstructuring the principal hydrogen-bonding interaction at 1.8 Å, a result that is obtained irrespective of the dataset size. However, for the longer-range O−H interaction at 3.3 Å, the NNPs reported here make essentially indistinguishable predictions compared to those made using SCAN/DZVP. This is also the case regarding the longer-range weak interaction at ∼5.2 Å.

Similar agreement is observed in Figure 3b for the O−O RDF in these simulations. Experimental RDFs for the O−O interaction obtained from neutron diffraction[9] show three peaks at ∼2.8, 4.4, and 6.6 Å. Figure 3b shows that SCAN/DZVP can reliably reproduce the structures of each of these peaks. The DFTB3/3ob trajectory (Figure 3b) also accurately reproduces the principal 2.8 Å interaction, despite slightly overstructuring it compared to SCAN/DZVP. However, the 4.4 and 6.6 Å interactions are overestimated slightly at ∼5.5 and 7 Å. Interestingly, Figure 3b shows that replacing DFTB3/3ob with SCAN/DZVP energy and force data markedly improves the structure of the O−O RDF obtained using NequIP NNPs, particularly at longer range. For instance, there is good agreement between SCAN/DZVP and $NNP_{SCAN\text{-}500,1000,2000}$ predictions both in length scale and in $g(r)$ peak height.

Figure 3c compares NNP predictions with SCAN/DZVP and experimental data for the Na^+−O RDF. Neutron diffraction data[8] feature two peaks corresponding to the first and second solvation shells of the Na^+ cation at ∼2.4 and 4.5 Å, respectively.[9] The DFTB3/3ob IonSolvR trajectory reproduces the length scale of the first solvation shell closely, while slightly underestimating the length scale of the second solvation shell at ∼4.1 Å. SCAN/DZVP also accurately predicts these length scales,[17] and the peak height for the first solvation shell is also in excellent agreement with XRD data.[43] As for the H−O and O−O interactions discussed above, Figure 3c shows excellent agreement between $NNP_{SCAN\text{-}500,1000,2000}$ data for the length of the Na^+−O interaction and both SCAN/DZP[17] and experimental results.[8] This indicates that the NNPs generated here reliably describe the solvation environment of the Na^+ cation. We note, however, that compared to SCAN/DZP data, this interaction is slightly understructured using the $NNP_{SCAN\text{-}500,1000,2000}$ potentials.

Figure 3d shows the RDFs for the Cl^-−H interaction in an aqueous NaCl solution. Neutron diffraction experiments[8] yield peaks at ∼2.2, 3.5, and 4.5 Å, corresponding to the first, second, and third hydration shells of the Cl^- anion. The DFTB3/3ob RDF reproduces these features fairly accurately, showing peaks at ∼2.3, 3.6, and 5.6 Å. While SCAN/DZVP MD data are unavailable for a direct comparison, Figure 3d shows that $NNP_{SCAN\text{-}500,1000,2000}$ data generated here closely reproduce the experimental peak distances in each case, which is a considerable improvement compared to the Cl^-−H solvation structure shown in Figure 2d (i.e., without SCAN/DZVP energies and forces). This is consistent with the agreement observed in the longer-range Cl^-−O interaction, as shown in Figure 3e. Importantly, the use of SCAN/DZVP energies and forces in the datasets here results in the contraction of the long-range interaction to ∼4.3 Å, which is substantially better performance than that observed using DFTB3/3ob. This contraction is consistent with that observed for the O−O interaction as discussed above. These NNP RDFs predict a weaker solvation structure around the Cl^- anion compared to DFTB3/3ob, but we note that solvation structure in water is often overpronounced using DFTB3/3ob.[44]

The preceding discussion has demonstrated that the IonSolvR database can be used as a resource for creating high-quality mGGA-level datasets for training NNPs using NequIP. We note that all results presented in Figures 2 and 3 were generated using simple random sampling of the raw trajectory data. More sophisticated sampling techniques, such as farthest-point sampling, which provide more uniformly distributed training data do not markedly improve these results for the largest training sets employed here (see Figure S3).

However, there are two noteworthy exceptions here. The first is improved prediction of the ion solvation structure at low sampling densities. Our main discussion so far has employed

https://doi.org/10.1021/acs.jpclett.3c01783
J. Phys. Chem. Lett. 2023, 14, 9508−9515

datasets comprising between 500 and 2000 frames from a single IonSolvR trajectory (i.e., sampling densities between 0.5 and 2%, respectively). As large-scale mGGA calculations carry considerable computational expense, we have additionally considered datasets comprising as few as 100 frames to ascertain the "lower limit" on IonSolvR sampling density. The performance of NNPs trained on farthest-point sampling datasets is considered in Figure S3, which shows accurate prediction of ion−solvent and solvent−solvent interactions with as few as 100 frames. Notably, the predicted solvent−solvent interactions (i.e., the H−O and O−O interactions) are essentially invariant to the size of the dataset. Farthest-point sampling provides a distinct advantage over random sampling in this respect, with NNP-MD simulations using NNPs trained on datasets of this size frequently showing numerical instabilities.

The second exception is the predicted Na^+-Cl^- interaction in NaCl(aq). We have not focused on this interaction here intentionally, as it is considerably undersampled in any IonSolvR-based dataset since these trajectories include only a single ion pair. Consequently, it is unreasonable to expect the Na^+-Cl^- interaction to be reliably predicted using NNPs trained solely on IonSolvR data. The main feature of the DFTB/3ob Na^+-Cl^- interaction is a so-called solvent-shared ion pair at ∼4.6 Å. By contrast, the contact ion pair observed observed in neutron diffraction at ∼2.75 Å[8] is significantly under-represented using DFTB3/3ob (Figure S4), and our 2000-frame IonSolvR dataset constructed via random sampling excluded this structure entirely. NNPs trained upon this dataset were therefore unable to reproduce the intensity of the Na^+-Cl^- contact ion pair interaction whatsoever (we note, however, that, unsurprisingly, there is strong agreement between NNP and IonSolvR descriptions of the longer-range Na^+-Cl^- contact structure; Figure S4). Conversely, these Na^+-Cl^- contact ion pair structures can be captured via farthest-point sampling, even for the smallest datasets considered here (100 frames). Nevertheless, they are still so infrequent that the NNP description of the Na^+-Cl^- interaction frequency at short range remained unreliable (Figure S4). We note, however, that the length scales of both the contact ion pair and solvent-shared ion pair are in excellent agreement with neutron diffraction experiments, indicating that SCAN itself provides an accurate description of the Na^+-Cl^- interaction potential.

Aqueous electrolytes are essential to the biological, chemical, and geological systems that support life, but predicting their structure and properties remains difficult, largely due to the large computational cost associated with running *ab initio* molecular dynamics. Neural network potentials have been suggested as a solution to this problem but are limited due to the requirement of generating *ab initio* trajectories for training. Here we have demonstrated that NNPs can be generated rapidly using the recently reported IonSolvR database and that predicting the solvation structure in aqueous NaCl solutions can be systematically improved further by sampling the Born−Oppenheimer PES at a lower level of theory and training using energies and forces sampled at a higher level of theory. This approach enables the use of embarrassingly parallel algorithms, which significantly decrease the computing time required for generating training data. Our results demonstrate that E(3)-equivariant NNPs constructed using NequIP maintain and in some cases improve upon the accuracy of the DFTB3/3ob for predicting ion solvation structure in NaCl(aq) and indeed match the accuracy of SCAN/DZVP AIMD data where sufficient data are available. Importantly, this agreement was achieved in some cases by using as few as 100 points on the DFTB3/3ob potential energy surface, depending on the sampling method employed. The apparent limitation of the IonSolvR trajectory used as the basis of this study is that the single-ion-pair simulation leads to undersampling of the Na^+-Cl^- interaction. Improved NNPs would therefore be obtained given training simulations comprising additional ion pairs. Nevertheless, SCAN-based NNPs were capable of reproducing the approximate length scale of these interactions, despite the paucity of training data. We also showed that an accurate description of the Na^+-Cl^- interaction length scale can be recovered by incorporating a minimal amount of training data. These results clearly demonstrate that E(3)-equivariant NNPs can be optimized to reproduce specific interactions in condensed-phase MD simulations in a highly efficient manner.

METHODS

IonSolvR Dataset Generation. IonSolvR datasets consisted of 2000, 1000, and 500 randomly selected frames (consisting of nuclear Cartesian coordinates, total energy, and nuclear forces) extracted from a 1 ns DFTB-MD trajectory (100,000 frames) of ∼0.8 M NaCl(aq) from IonSolvR.[13] Frames consisted of a periodic unit cell containing 100 water molecules and 1 NaCl ion pair. Energies and nuclear coordinates were sampled from the original IonSolvR trajectory, while nuclear forces (not originally included in the IonSolvR database) were calculated for each dataset element using third-order DFTB[45] in conjunction with the 3ob parameter set.[46]

IonSolvR+mGGA Dataset Generation. SCAN datasets consisted of the same 2000, 1000, and 500 randomly selected frames as the IonSolvR datasets detailed in the preceding section. The nuclear Cartesian coordinates of each frame therefore corresponded to those obtained using DFTB3/3ob, but the total energies and forces were calculated using the strongly constrained and appropriately normed (SCAN)[39] exchange−correlation functional, respectively, as implemented in the CP2K program, in conjunction with the Γ-point approximation.[47] All SCAN calculations employed DZVP basis sets for Na, Cl, O, and H[48] and energy convergence thresholds of $1 \times 10^{-7} E_h$.[48] The fact that these energy and force calculations are independent of each other enables them to be calculated via an embarrassingly parallel approach, thereby enabling a significant decrease in the real time required for dataset construction.

Sampling Method. NNP training datasets consisting of between 100 and 2000 IonSolvR frames were constructed via random sampling or farthest-point sampling. To enable direct comparison between random datasets of different sizes, random datasets were derived as subsets of a single 2000-frame dataset obtained from a 1 ns DFTB-MD trajectory (100,000 frames) of ∼0.8 M NaCl(aq) from IonSolvR. For the farthest-point sampling dataset, each frame in the IonSolvR trajectory was represented using the smooth overlap of atomic positions (SOAP) descriptor,[49−51] as implemented in DScribe.[52] Dataset elements were then chosen based on Euclidean distances. The SOAP descriptor employed parameters r_{cut} = 8.0 Å, n_{max} = 8, and l_{max} = 8 (and we note that varying these parameters gave negligible difference to dataset composition). The SOAP descriptor only encoded the environment around the Na^+ and Cl^- ions in the IonSolvR trajectory to optimize the elements of this dataset.

https://doi.org/10.1021/acs.jpclett.3c01783
J. Phys. Chem. Lett. 2023, 14, 9508−9515

Neural Network Potential Generation. NNPs were trained on atomic coordinates, total energies, and forces in each respective dataset using NequIP (see Table S1 for NNP hyperparameters employed).[24] Following the approach of Duignan et al.,[37] to simplify the NNP optimization, long-range dielectrically screened ion−ion Coulomb interactions were removed from the DFT energies and forces during NNP optimization, calculated separately in LAMMPS,[53] and subsequently added back in manually once the NNP had been trained. NNPs trained using the IonSolvR and IonSolvR+mGGA datasets are referred to using the nomenclature "$NNP_{DFTB\text{-}X}$" and "$NNP_{SCAN\text{-}X}$", respectively, where X is the size of the dataset. For each dataset, five independent NNPs were generated to ensure consistent results. The impact of the training:test set ratio on NNPs trained using IonSolvR+mGGA datasets is considered in Figure S2.

Molecular Dynamics Simulations. Trained NNPs were verified using 2 ns MD simulations of 0.8 M NaCl(aq) (1 NaCl ion pair and 100 molecules) in an *NVT* ensemble. While this system size is small, it was chosen to enable a direct comparison with the original IonSolvR trajectory. We note also that the dimensions of the periodic unit cell in these NNP-MD simulations, and hence the density of the simulation, were held fixed at the values used in IonSolvR for the same reason. Temperature was enforced via a Nosé−Hoover thermostat[54] at 300 K and a time step of 0.5 fs; we note that this was consistent with the temperature and thermostat used in IonSolvR. All MD simulations employed LAMMPS[53] and were analyzed using MDAnalysis.[55] All radial distribution functions presented are the averages of those obtained from five independent 2 ns *NVT* MD trajectories. Unless stated otherwise, all MD results were generated using NNPs trained using a training:test set ratio of 80%:20% (see Figure S2).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.3c01783.

NNP hyperparameters, comparison of NNP and IonSolvR nuclear forces, NNP validation set average force RMSE data, and comparison of NaCl(aq) radial distribution functions using NNPs with random and farthest-point sampling datasets (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Alister J. Page − *Discipline of Chemistry, College of Engineering, Science and Environment, University of Newcastle, Newcastle, NSW 2308, Australia;* orcid.org/0000-0002-8444-2775; Email: alister.page@newcastle.edu.au

Authors

Sophie Baker − *Discipline of Chemistry, College of Engineering, Science and Environment, University of Newcastle, Newcastle, NSW 2308, Australia*

Joshua Pagotto − *School of Chemical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia*

Timothy T. Duignan − *School of Chemical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia; Queensland Micro- and Nanotechnology Centre, Griffith University, Brisbane, QLD 4111, Australia;* orcid.org/0000-0003-3772-8057

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.3c01783

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.J.P. acknowledges Australian Research Council Discovery Project (DP190100788). T.T.D. acknowledges the Australian Research Council (ARC) funding via Projects DE200100794 and DP200102573. This research/project was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

■ REFERENCES

(1) Brodholt, J. P. Molecular Dynamics Simulations of Aqueous NaCl Solutions at High Pressures and Temperatures. *Chem. Geol.* **1998**, *151* (1−4), 11−19.

(2) Pohorille, A.; Pratt, L. R. Is Water the Universal Solvent for Life? *Origins Life Evol. Biospheres* **2012**, *42* (5), 405−409.

(3) Bouazizi, S.; Nasr, S. Self-Diffusion Coefficients and Orientational Correlation Times in Aqueous NaCl Solutions: Complementarity with Structural Investigations. *J. Mol. Liq.* **2011**, *162* (2), 78−83.

(4) Ohtaki, H.; Radnai, T. Structure and Dynamics of Hydrated Ions. *Chem. Rev.* **1993**, *93* (3), 1157−1204.

(5) Skipper, N. T.; Neilson, G. W. X-Ray and Neutron Diffraction Studies on Concentrated Aqueous Solutions of Sodium Nitrate and Silver Nitrate. *J. Phys: Condens. Matter* **1989**, *1* (26), 4141.

(6) Mähler, J.; Persson, I. A Study of the Hydration of the Alkali Metal Ions in Aqueous Solution. *Inorg. Chem.* **2012**, *51* (1), 425−438.

(7) Neilson, G. W. The Application of Neutron Scattering Methods to Aqueous Electrolyte Solutions. *Physica B+C* **1983**, *120* (1−3), 325−334.

(8) Mancinelli, R.; Botti, A.; Bruni, F.; Ricci, M. A.; Soper, A. K. Hydration of Sodium, Potassium, and Chloride Ions in Solution and the Concept of Structure Maker/Breaker. *J. Phys. Chem. B* **2007**, *111* (48), 13570−13577.

(9) Mancinelli, R.; Botti, A.; Bruni, F.; Ricci, M. A.; Soper, A. K. Perturbation of Water Structure Due to Monovalent Ions in Solution. *Phys. Chem. Chem. Phys.* **2007**, *9* (23), 2959−2967.

(10) Bankura, A.; Carnevale, V.; Klein, M. L. Hydration Structure of Salt Solutions from Ab Initio Molecular Dynamics. *J. Chem. Phys.* **2013**, *138* (1), 014501.

(11) Ghaffari, A.; Rahbar-Kelishami, A. MD Simulation and Evaluation of the Self-Diffusion Coefficients in Aqueous NaCl Solutions at Different Temperatures and Concentrations. *J. Mol. Liq.* **2013**, *187*, 238−245.

(12) Lyubartsev, A. P.; Laaksonen, A. Concentration Effects in Aqueous NaCl Solutions. A Molecular Dynamics Simulation. *J. Phys. Chem.* **1996**, *100* (40), 16410−16418.

(13) Gregory, K. P.; Elliott, G. R.; Wanless, E. J.; Webber, G. B.; Page, A. J. A Quantum Chemical Molecular Dynamics Repository of Solvated Ions. *Sci. Data* **2022**, *9* (1), 430.

(14) Fulton, J. L.; Schenter, G. K.; Baer, M. D.; Mundy, C. J.; Dang, L. X.; Balasubramanian, M. Probing the Hydration Structure of Polarizable Halides: A Multiedge XAFS and Molecular Dynamics Study of the Iodide Anion. *J. Phys. Chem. B* **2010**, *114* (40), 12926−12937.

(15) Alder, B. J.; Wainwright, T. E. Phase Transition for a Hard Sphere System. *J. Chem. Phys.* **1957**, *27* (5), 1208−1209.

(16) Iftimie, R.; Minary, P.; Tuckerman, M. E. Ab Initio Molecular Dynamics: Concepts, Recent Developments, and Future Trends. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6654−6659.

(17) Duignan, T. T.; Schenter, G. K.; Fulton, J. L.; Huthwelker, T.; Balasubramanian, M.; Galib, M.; Baer, M. D.; Wilhelm, J.; Hutter, J.; Del Ben, M.; Zhao, X. S.; Mundy, C. J. Quantifying the Hydration Structure of Sodium and Potassium Ions: Taking Additional Steps on Jacob's Ladder. *Phys. Chem. Chem. Phys.* **2020**, *22* (19), 10641−10652.
(18) Duignan, T. T.; Baer, M. D.; Schenter, G. K.; Mundy, C. J. Real Single Ion Solvation Free Energies with Quantum Mechanical Simulation. *Chem. Sci.* **2017**, *8* (9), 6131−6140.
(19) Rode, B. M.; Schwenk, C. F.; Hofer, T. S.; Randolf, B. R. Coordination and Ligand Exchange Dynamics of Solvated Metal Ions. *Coord. Chem. Rev.* **2005**, *249* (24), 2993−3006.
(20) Dajnowicz, S.; Agarwal, G.; Stevenson, J. M.; Jacobson, L. D.; Ramezanghorbani, F.; Leswing, K.; Friesner, R. A.; Halls, M. D.; Abel, R. High-Dimensional Neural Network Potential for Liquid Electrolyte Simulations. *J. Phys. Chem. B* **2022**, *126* (33), 6271−6280.
(21) Jacobson, L. D.; Stevenson, J. M.; Ramezanghorbani, F.; Ghoreishi, D.; Leswing, K.; Harder, E. D.; Abel, R. Transferable Neural Network Potential Energy Surfaces for Closed-Shell Organic Molecules: Extension to Ions. *J. Chem. Theory Comput.* **2022**, *18* (4), 2354−2366.
(22) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), 143001.
(23) Li, W.; Ou, Q.; Chen, Y.; Cao, Y.; Liu, R.; Zhang, C.; Zheng, D.; Cai, C.; Wu, X.; Wang, H.; Chen, M.; Zhang, L. DeePKS +ABACUS as a Bridge between Expensive Quantum Mechanical Models and Machine Learning Potentials. *J. Phys. Chem. A* **2022**, *126* (49), 9154−9164.
(24) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials. *Nat. Commun.* **2022**, *13* (1), 2453.
(25) Liang, W.; Lu, G.; Yu, J. Theoretical Prediction on the Local Structure and Transport Properties of Molten Alkali Chlorides by Deep Potentials. *J. Mater. Sci. Technol.* **2021**, *75*, 78−85.
(26) Shi, Y.; Lam, S. T.; Beck, T. L. Deep Neural Network Based Quantum Simulations and Quasichemical Theory for Accurate Modeling of Molten Salt Thermodynamics. *Chem. Sci.* **2022**, *13* (28), 8265−8273.
(27) Dajnowicz, S.; Agarwal, G.; Stevenson, J. M.; Jacobson, L. D.; Ramezanghorbani, F.; Leswing, K.; Friesner, R. A.; Halls, M. D.; Abel, R. High-Dimensional Neural Network Potential for Liquid Electrolyte Simulations. *J. Phys. Chem. B* **2022**, *126* (33), 6271−6280.
(28) Zhang, W.; Zhou, L.; Yang, B.; Yan, T. Molecular Dynamics Simulations of LiCl Ion Pairs in High Temperature Aqueous Solutions by Deep Learning Potential. *J. Mol. Liq.* **2022**, *367*, 120500.
(29) Zhang, C.; Yue, S.; Panagiotopoulos, A. Z.; Klein, M. L.; Wu, X. Dissolving Salt Is Not Equivalent to Applying a Pressure on Water. *Nat. Commun.*. **2022**, *13* (1), 822.
(30) O'Neill, N.; Schran, C.; Cox, S. J.; Michaelides, A. Crumbling Crystals: On the Dissolution Mechanism of NaCl in Water. *arXiv (Physics.Chemical Physics)*, November 8, 2022, 2211.04345, ver. 1. https://arxiv.org/abs/2211.04345 (accessed 2023-06-30).
(31) Hellström, M.; Behler, J. Structure of Aqueous NaOH Solutions: Insights from Neural-Network-Based Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2017**, *19* (1), 82−96.
(32) Jinnouchi, R.; Karsai, F.; Kresse, G. Making Free-Energy Calculations Routine: Combining First Principles with Machine Learning. *Phys. Rev. B* **2020**, *101* (6), 060201.
(33) Zhang, L.; Wang, H.; Car, R.; E, W. Phase Diagram of a Deep Potential Water Model. *Phys. Rev. Lett.* **2021**, *126* (23), 236001.
(34) Fan, X. T.; Wen, X. J.; Zhuang, Y. B.; Cheng, J. Molecular Insight into the GaP(110)-Water Interface Using Machine Learning Accelerated Molecular Dynamics. *J. Energy Chem.* **2023**, *82*, 239−247.
(35) Chen, Y.; Zhang, L.; Wang, H.; E, W. DeePKS: A Comprehensive Data-Driven Approach toward Chemically Accurate Density Functional Theory. *J. Chem. Theory Comput.* **2021**, *17* (1), 170−181.
(36) Zhang, Y.; Wang, H.; Chen, W.; Zeng, J.; Zhang, L.; Wang, H.; E, W. DP-GEN: A Concurrent Learning Platform for the Generation of Reliable Deep Learning Based Potential Energy Models. *Comput. Phys. Commun.* **2020**, *253*, 107206.
(37) Pagotto, J.; Zhang, J.; Duignan, T. T. Predicting Electrolyte Solution Properties by Combining Neural Network Accelerated Molecular Dynamics and Continuum Solvent Theory. Presented at NeurIPS 2022 AI for Science: Progress and Promises, New Orleans, LA, December 2, 2022.
(38) Hourahine, B.; Aradi, B.; Blum, V.; Bonafé, F.; Buccheri, A.; Camacho, C.; Cevallos, C.; Deshaye, M. Y.; Dumitrică, T.; Dominguez, A.; Ehlert, S.; Elstner, M.; van der Heide, T.; Hermann, J.; Irle, S.; Kranz, J. J.; Köhler, C.; Kowalczyk, T.; Kubař, T.; Lee, I. S.; Lutsker, V.; Maurer, R. J.; Min, S. K.; Mitchell, I.; Negre, C.; Niehaus, T. A.; Niklasson, A. M. N.; Page, A. J.; Pecchia, A.; Penazzi, G.; Persson, M. P.; Řezáč, J.; Sánchez, C. G.; Sternberg, M.; Stöhr, M.; Stuckenberg, F.; Tkatchenko, A.; Yu, V. W. -z.; Frauenheim, T. DFTB+, a Software Package for Efficient Approximate Density Functional Theory Based Atomistic Simulations. *J. Chem. Phys.* **2020**, *152* (12), 124101.
(39) Sun, J.; Ruzsinszky, A.; Perdew, J. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Phys. Rev. Lett.* **2015**, *115* (3), 036402.
(40) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11* (5), 2087−2096.
(41) Stöhr, M.; Medrano Sandonas, L.; Tkatchenko, A. Accurate Many-Body Repulsive Potentials for Density-Functional Tight Binding from Deep Tensor Neural Networks. *J. Phys. Chem. Lett.* **2020**, *11* (16), 6835−6843.
(42) Yao, Y.; Kanai, Y. Free Energy Profile of NaCl in Water: First-Principles Molecular Dynamics with SCAN and wB97X-V Exchange-Correlation Functionals. *J. Chem. Theory Comput.* **2018**, *14* (2), 884−893.
(43) Galib, M.; Baer, M. D.; Skinner, L. B.; Mundy, C. J.; Huthwelker, T.; Schenter, G. K.; Benmore, C. J.; Govind, N.; Fulton, J. L. Revisiting the Hydration Structure of Aqueous Na^+. *J. Chem. Phys.* **2017**, *146* (8), 084504.
(44) Goyal, P.; Qian, H. J.; Irle, S.; Lu, X.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q. Molecular Simulation of Water and Hydration Effects in Different Environments: Challenges and Developments for DFTB Based Models. *J. Phys. Chem. B* **2014**, *118* (38), 11007−11027.
(45) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7* (4), 931−948.
(46) Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parameterization of the DFTB3Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *J. Chem. Theory Comput.* **2015**, *11* (1), 332−342.
(47) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; Golze, D.; Wilhelm, J.; Chulkov, S.; Bani-Hashemian, M. H.; Weber, V.; Borštnik, U.; Taillefumier, M.; Jakobovits, A. S.; Lazzaro, A.; Pabst, H.; Müller, T.; Schade, R.; Guidon, M.; Andermatt, S.; Holmberg, N.; Schenter, G. K.; Hehn, A.; Bussy, A.; Belleflamme, F.; Tabacchi, G.; Glöß, A.; Lass, M.; Bethune, I.; Mundy, C. J.; Plessl, C.; Watkins, M.; VandeVondele, J.; Krack, M.; Hutter, J. CP2K: An Electronic Structure and Molecular Dynamics Software Package - Quickstep: Efficient and Accurate Electronic Structure Calculations. *J. Chem. Phys.* **2020**, *152* (19), 194103.
(48) VandeVondele, J.; Hutter, J. Gaussian Basis Sets for Accurate Calculations on Molecular Systems in Gas and Condensed Phases. *J. Chem. Phys.* **2007**, *127* (11), 114105.
(49) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* **2013**, *87* (18), 184115.
(50) Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Himanen, L.; Foster, A. S. Machine Learning Hydrogen Adsorption on

Nanoclusters through Structural Descriptors. *npj Comput. Mater.* **2018**, *4* (1), 37.
(51) De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016**, *18* (20), 13754−13769.
(52) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of Descriptors for Machine Learning in Materials Science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
(53) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1−19.
(54) Evans, D. J.; Holian, B. L. The Nose−Hoover Thermostat. *J. Chem. Phys.* **1985**, *83* (8), 4069−4074.
(55) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32* (10), 2319−2327.

https://doi.org/10.1021/acs.jpclett.3c01783
J. Phys. Chem. Lett. 2023, 14, 9508−9515