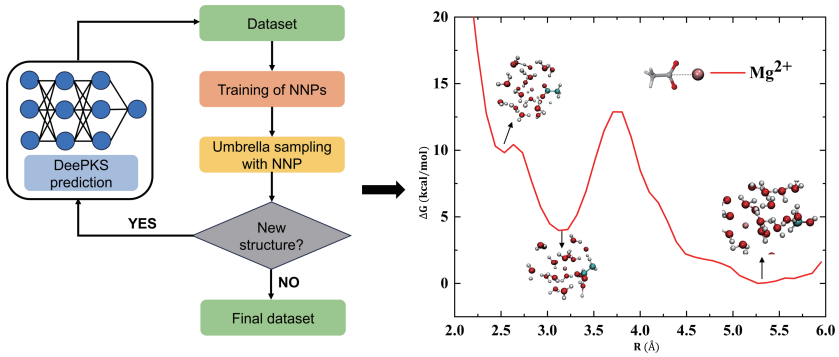
ARTICLE

Interaction of Magnesium Ion and Acetate Anion in Bulk Water: Toward High-Level Machine Learning Potential

Jiaying Gu[a], Jin Xiao[b], Xingyu Wu[a], Xi Zhu[a], Huimin Chen[a], John Z. H. Zhang[b,c,d,e,f], Tong Zhu[b,c,d*], Ya Gao[a,g*], Zhixiang Yin[a,g*]

a. School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China
b. Shanghai Engineering Research Center of Molecular Therapeutics & New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062, China
c. CAS Key Laboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
d. NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai 200062, China
e. Department of Chemistry, New York University, New York 10003, USA
f. Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China
g. Shanghai Frontiers Science Research Center for Druggability of Cardiovascular Noncoding RNA, Shanghai 201620, China

(Dated: Received on March 5, 2024; Accepted on April 25, 2024)



Metal ions play crucial roles in various biological functions, including maintaining homeostasis, regulating muscle contraction, and facilitating enzyme catalysis. However, accurately simulating the interaction between metal ions and amino acid side chain analogs using high-level wave function theories remains challenging due to the significant computational costs involved. In this study, deep potential molecular dynamics (DeePMD) simulation was employed to investigate the solvation structure of the Mg^{2+}-Ac^{-} ion pair in aqueous solution. To address the computational bottleneck associated with expensive quantum mechanics (QM) methods, the Deep Kohn-Sham (DeePKS) approach was utilized, which allows us to generate highly accurate self-consistent energy functionals while significantly reducing computational costs. The root mean square error and mean absolute error of energies and atomic forces indicate close agreement between DeePKS predictions and QM strongly constrained and appropriately normed (SCAN) calculations. Moreover, the neural network potential (NNP) generated using the SCAN-level dataset predicted by DeePKS exhibits higher accuracy compared to previous work, which employed a moderate BLYP functional. The potential of mean force for the Mg^{2+}-Ac^{-} system was further examined, revealing a preference for monodentate coordination of Mg^{2+} with a ~5.8 kcal/mol energy barrier between bidentate and monodentate geometries. Overall, this work provides a comprehensive, precise,

* Authors to whom correspondence should be addressed. E-mail: tzhu@lps.ecnu.edu.cn, gaoya@sues.edu.cn, 21190006@sues.edu.cn

and reliable methodology for investigating metal ions' properties in aqueous solutions.

Key words: Molecular dynamics simulation, Umbrella sampling, Neural network potential, Machine learning

I. INTRODUCTION

Ion-selective interactions have long been recognized as fundamental phenomena in numerous chemical, environmental, and biological processes that occur in aqueous solution. Metal ions play a central role in various biological functions, including maintaining homeostasis, regulating muscle contraction, and facilitating enzyme catalysis [1–3]. In proteins, metal ions often coordinate with the carboxylate of the side chain of residues such as ASP, GLU, CYS, and HIS, which significantly impact protein association and enzymatic activity [4]. Mg^{2+}, one of the most abundant elements in the human body, plays critical roles in various biomolecular processes, including binding with ATP to activate it biologically [5] and stabilizing the tertiary structure of DNA and RNA [2]. Therefore, it is of paramount importance to thoroughly investigate the energetic interaction mechanisms between metal ions and amino acid side chain analogs, such as acetate anions, to comprehend the chemical and physical behaviors underlying cellular processes.

Over the past few decades, researchers have extensively utilized both experimental and theoretical approaches to explore the metal carboxylate binding constant and the hydration shell properties of Mg^{2+} in aqueous solution. X-ray diffraction [6–10] and neutron diffraction [11, 12] have provided valuable insights into the ionic coordination number and the solvation structure of Mg^{2+} in aqueous solution [9]. The sequential bond energies of $[Mg(H_2O)_n]^{2+}$ ($n=2$–10) were measured using the threshold collision-induced dissociation [13]. Vibrational spectroscopic methods have also been employed to determine the metal acetate binding constants, although inconsistencies in results have been reported [14, 15]. Denilson *et al.* introduced a novel Raman multivariate curve resolution approach to obtain the binding free energies for $Ac^{-}\cdots M^{2+}$ in aqueous solution (where M=Zn, Mg, and Ca) [16]. Theoretically, molecular dynamics (MD) simulations that effectively model electronic polarization via charge scaling have been utilized to calculate the binding free energies of different cations (Ca^{2+}, Zn^{2+}, and Mg^{2+}) to an acetate ion [16]. Studies combining *ab initio* methods with MD simulations have provided insights into the bidental nature of the -COO^{-} and X^{+} (X=Li, Na, K, and NH_4) interactions in aqueous solution [17]. Jing *et al.* demonstrated the importance of the many-body effects on Ca^{2+} selectivity compared to Mg^{2+} through *ab initio* calculations on model compounds of the binding pockets [18]. Although *ab initio* molecular dynamics (AIMD) studies have predicted coordination number of Ca^{2+} ranges from 2 to 10 and Mg^{2+} coordinates with 6 water molecules [19–21], the computational cost limits its applicability to relatively small systems and short time scales [22–24].

In recent years, machine learning (ML) methods have emerged as promising solutions to the computational challenges associated with AIMD, offering the potential to construct potential energy surfaces with high accuracy and efficiency [25–29]. Zhang *et al.* introduced the deep potential molecular dynamics (DeePMD) scheme, which utilizes the deep neural networks to fit potential energy surfaces based on expensive *ab initio* data. This approach enables the simulation of large-size systems over long-time scales with first-principles accuracy, significantly reducing computational cost [29–30]. Researchers have explored various systems using DeePMD, including aqueous solutions of NaCl, KCl, and NaBr, revealing insights into how salt ions affect the structure of water under different conditions [31]. Liu *et al.* employed DeePMD to investigate the solvation structures and properties of ion pairs such as Ca^{2+}-OH^{-} and Mg^{2+}-OH^{-} in water, shedding light on their interactions in aqueous environment [32]. However, a significant challenge in utilizing ML approaches is the reliance on reference datasets obtained from high-level quantum mechanics (QM) methods, which can be computationally expensive. The accuracy of the trained neural network potentials (NNPs) is thus limited by the quality of the reference data. To address this limitation, Liu *et al.* developed a deep potential based on fragment-based theory (DP-MP2) for water, offering an alternative approach for training NNPs [33].

In this work, DeePMD simulation was utilized to explore the solvation structure of the Mg^{2+}-Ac^{-} ion pair

DOI: 10.1063/1674-0068/cjcp2402023

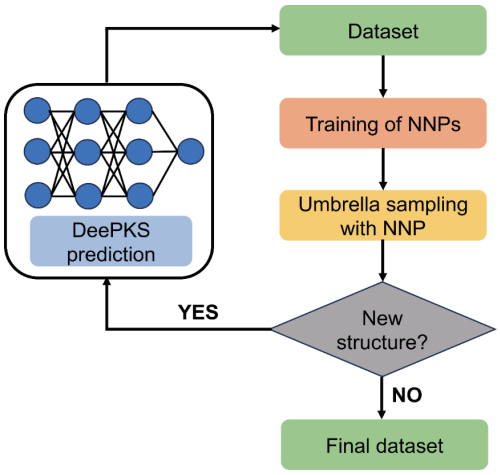


FIG. 1 The procedure for the construction of the neutral network potential.

in aqueous solution. To address the computational bottleneck associated with expensive QM methods, the deep Kohn-Sham (DeePKS) approach [34, 35] was used for the generation of highly accurate self-consistent energy functionals while significantly reducing computational costs. The procedure for constructing the neutral network potential is illustrated in FIG. 1. The accuracy of DeePKS model and the trained neutral network potential were examined to ensure reliability in our simulations. Overall, this work provides a comprehensive, precise, and reliable methodology for investigating metal ions' properties in aqueous solutions.

II. METHODS

A. Models

For Mg^{2+}-Ac^{-} system, the initial configurations with different distances between Mg^{2+} and the carbon atom of carboxyl group were obtained from our previous work [36], generated through short AIMD simulations. In the system, Mg^{2+} and Ac^{-} were immersed in a cubic water box with the dimension 16 Å (128 water molecules in total), and 41 windows spanning the reaction coordinate from 2.0 Å to 6.0 Å, with an interval of 0.1 Å, were set to comprehensively examine the system's dynamic behavior.

B. DeePKS model generation

DeePKS is a machine learning-based DFT model that utilizes a computationally efficient neural network-based functional model to develop a correction term applied to an economical DFT model. The detailed principles of the DeePKS method have been extensively demonstrated in previous literatures [37, 38]. It has undergone through testing and shows that the DeePKS model reproduces the target energies and forces even when trained on a significantly small dataset. In this work, the DeePKS method serves as a bridging mechanism between cost-effective Perdew-Burke-Ernzerhof (PBE) [39] GGA functional and expensive strongly constrained and appropriately normed (SCAN) meta-GGA [40] calculations, facilitating the prediction of energy and atomic forces during the deep potential generation procedure. To get the DeePKS model from PBE to SCAN, a total of 410 frames, consisting of 10 frames from each configuration across 41 distance windows, were calculated using CP2K [41] to acquire their energies and atomic forces. The SCAN functional for the exchange and correlation terms, in conjunction with the Grimme's D3 correction, were employed ensuring accurate descriptions of molecular structures [42]. Core-valence interactions were described utilizing Goedecker–Teter–Hutter (GTH) pseudopotentials optimized for the SCAN functional. Double-zeta valence polarized basis set MOLOPT (DZVP-MOLOPT-SR-GTH) was employed to represent atomic species [43], with a plane wave kinetic energy cutoff (E_{cut}) set to 1000 Ry.

The energies and forces from 155 structures were allocated for the training set, while 255 structures were reserved for the validation set during the DeePKS training. In this work, a radius cutoff of 5 Bohr was employed, with a kinetic energy cutoff set at 100 Ry. Initially, the DeePHF model was trained and utilized as the starting point for DeePKS training. During each training round of DeePKS, SCF calculations were performed and only converged structures were retained for subsequent training iterations. The DeePKS model comprised a neural network with three hidden layers, each containing 108 neurons. Training utilized the Adam optimizer with a batch size of 16 and a learning rate of 1×10^{-4}. The learning rate exponentially decayed by a factor of 0.5 every 1000 epochs. The total number of epochs was capped at 5,000. Iterative training of DeePKS ceased once the mean absolute error (MAE) of the training and validation sets converged.

C. Deep potential generation

The deep potential generator (DP-GEN) [44] was utilized for generating deep potential (DP) models. A dataset comprising 2027 data points, with 50 data points for each configuration, was derived from DeePKS

DOI: 10.1063/1674-0068/cjcp2402023

calculations and subsequently trained using the DeePMD-kit package [45], integrated with the DP-GEN package. During each training iteration, the embedding network consisted of three hidden layers with the sizes of (25, 50, 100), while the fitting network comprised three hidden layers with the sizes of (240, 240, 240). An exponentially decaying learning rate, ranging from 1.0×10^{3} to 3.5×10^{-8}, was applied. The prefactors of the energy and the force terms in the loss function were dynamically adjusted, transitioning from 0.02 to 1 and from 1000 to 1, respectively. The DP model underwent training for 2×10^{5} steps. To ensure robustness, four DP models were generated for each training session, and each initialized with different weights but trained on the same reference dataset. Concurrent learning algorithm was employed to ensure comprehensive coverage of the target chemical space. Subsequently, the initial four NNP models were subjected to restrained MD simulations in each window using LAMMPS software [46] combined with the umbrella sampling method [47, 48]. Periodic boundary conditions were applied and PLUMED plugin [48] was utilized to constrain the reaction coordinates. Within DP-GEN, the model deviation, denoted as ε, represented the maximum standard deviation of the predictions for the atomic forces, serving as a key metric to assess differences between models,

$$\varepsilon = \max_i \sqrt{\left(\|f_i - \bar{f}_i\|\right)^2}, \quad \bar{f} = \langle f \rangle$$

where f_i denotes the predicted force on the atom i, $\bar{f}_i$ denotes the mean of the forces predicted by the four NNP models. During the MD simulations, atomic structures with force deviations falling within the range of $0.15\ \text{eV/Å} < \varepsilon < 0.3\ \text{eV/Å}$ were identified as candidate configurations. In each iteration, up to 50 candidates from each window were selected, which were incorporated into the training set for the next iteration after predicting their energies and atomic forces using the DeePKS model. Additionally, the simulation time was gradually increased in each iteration, until the target simulation time of 4 ns was reached.

III. RESULTS AND DISCUSSION

A. Verification of the DeePKS model

To validate the robustness of the CP2K calculation, energies and atomic forces for 7 randomly selected structures were calculated employing SCAN functional with the GTH pseudopotentials optimized for the PBE functional using both CP2K and VASP. The root mean square error (RMSE) for energies and forces are 4.9 kcal/mol and 1.4 $\text{kcal·mol}^{-1}\text{·Å}^{-1}$, respectively, affirming the accuracy of the CP2K approach. Subsequently, SCAN functional with the GTH pseudopotentials optimized for SCAN were employed to calculate the energies and forces of 410 frames, yielding a broad range of energy (−1411718.40 kcal/mol to −1411523.87 kcal/mol) and force (−121.50 to 118.59 $\text{kcal·mol}^{-1}\text{·Å}^{-1}$) respectively, indicating comprehensive exploration of the conformational space. The 410 frames with the SCAN energy and force labels were used to train the PBE-based DeePKS model. The energy and force RMSEs for the 155-train set and 255-validation set were shown in Table I. For the train set, the RMSE of the energy and the force were $1.13\times10^{-3}\ \text{kcal·mol}^{-1}\text{·atom}^{-1}$ and 1.27 $\text{kcal·mol}^{-1}\text{·Å}^{-1}$, and for the validation set, the RMSE of the energy and the force were $2.44\times10^{-3}\ \text{kcal·mol}^{-1}\text{·atom}^{-1}$ and 1.31 $\text{kcal·mol}^{-1}\text{·Å}^{-1}$. FIG. 2 depicts the correlation between energies and atomic forces as predicted by the DeePKS model and those calculated by CP2K across the overall 410 data set. It can be seen that with a few frames, the DeePKS model yields highly accurate predictions. To further validate the DeePKS model's performance, 41 new frames were generated as a test set. Compared to CP2K calculations, the DeePKS model exhibited RMSE values of $2.30\times10^{-3}\ \text{kcal·mol}^{-1}\text{·atom}^{-1}$ for energy and 1.20 $\text{kcal·mol}^{-1}\text{·Å}^{-1}$ for force, demonstrating comparable accuracy to the train and validation sets. Notably, while a single SCAN calculation by CP2K takes approximately 2 h on the 36-core processor to complete, the DeePKS model can accurately reproduce these results within 35 min with the same core processor, a remarkable computational time-saving of over 3 orders of magnitude. Thus, the DeePKS model can serve as a reliable bridge between computationally expensive *ab initio* calculations such as SCAN and cost-ef-

TABLE I The performance of DeePKS prediction for the Mg^{2+}-Ac^{-} system.

DeePKS	Force/ ($\text{kcal·mol}^{-1}\text{·Å}^{-1}$)		Energy/ ($10^{-3}\ \text{kcal·mol}^{-1}\text{·atom}^{-1}$)	
	MAE	RMSE	MAE	RMSE
Train set	0.899	1.267	0.899	1.129
Validation set	0.899	1.313	1.728	2.442
Test set	0.876	1.198	1.728	2.304

DOI: 10.1063/1674-0068/cjcp2402023

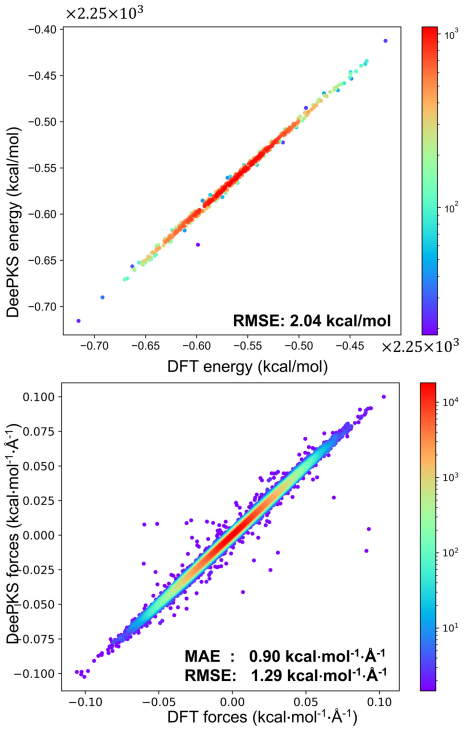


FIG. 2 The correlation of energies (upper) and atomic forces (lower) predicted by the DeePKS model and those calculated by CP2K across the overall 410 data set. The color bar indicates the density of data points.

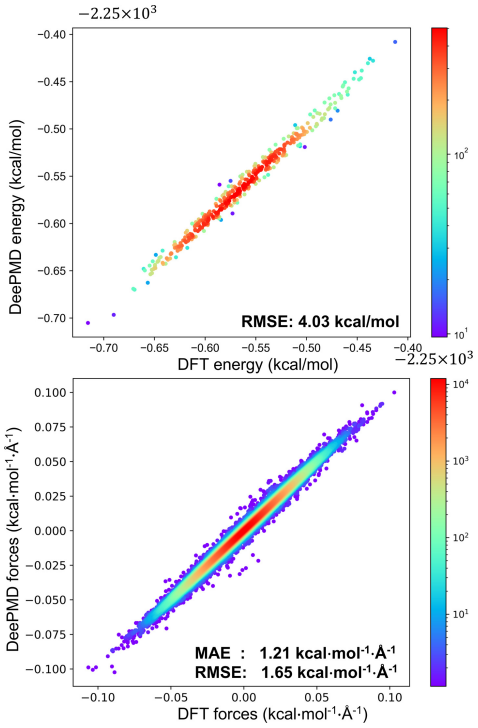


FIG. 3 The correlation of energies (upper) and atomic forces (lower) predicted by the NNP model and those calculated by CP2K across the 410 data set. The color bar indicates the density of data points.

fective DFT calculations, offering an efficient and precise approach to computational chemistry investigations.

B. Performance of the NNP

For atomic forces, the MAE between the NNP prediction and the DeePKS calculation is 1.22 kcal·mol^{-1}·Å^{-1}, with the RMSE of 1.66 kcal·mol^{-1}·Å^{-1}. For the energy, it is 8.13 ×10^{-3} kcal·mol^{-1}·atom^{-1} with RMSE of 9.52×10^{-3} kcal·mol^{-1}·atom^{-1}. FIG. 3 displays the correlation between energies and atomic forces as predicted by the NNP model and those calculated by CP2K for the 410 data set. The MAE and RMSE indicate that the atomic forces and energies obtained from NNP and DeePKS are in close agreement. In our previous work, DFT calculations were conducted using the BLYP functional with a double-ξ valence polarized (DZVP) basis set along with GTH pseudopotentials due to computational affords. The MAE and RMSE between NNP predictions and DFT calculations for atomic forces were 1.31 and 2.00 kcal·mol^{-1}·Å^{-1}, respectively, which are larger than those obtained in the current work. Additionally, for energy, the RMSE was 36.87×10^{-3} kcal·mol^{-1}·atom^{-1}, significantly higher than 9.52×10^{-3} kcal·mol^{-1}·atom^{-1} observed in the present work. These differences highlight the importance of the accuracy of the label dataset, which should be obtained through high-level QM method calculations, determining the performance of the NNP.

C. Modeling Mg^{2+}-Ac^{-} system in water solution

Following the MD simulations conducted with NNPs, the potential of mean force (PMF) curve for the Mg^{2+}-Ac^{-} system was computed with respect to the reaction coordinates using the weighted histogram analysis method (WHAM) [49]. As depicted in FIG. 4, the PMF curve reveals three distinct minima. The first minimum, located at a reaction coordinate of 2.5 Å, corresponds to the formation of the contact bidentate structure. In this configuration, both oxygen atoms on the carboxyl group are coordinated with Mg^{2+}. The second minimum, observed at a reaction coordinate of 3.2 Å, represents a monodentate structure where Mg^{2+} interacts with only one oxygen atom in the carboxyl group. The third minimum, situated around the reaction coordinate of 5.2 Å, corresponds to a solvent-shared structure. In this configuration, Mg^{2+} coordinates with 6 wa-

DOI: 10.1063/1674-0068/cjcp2402023

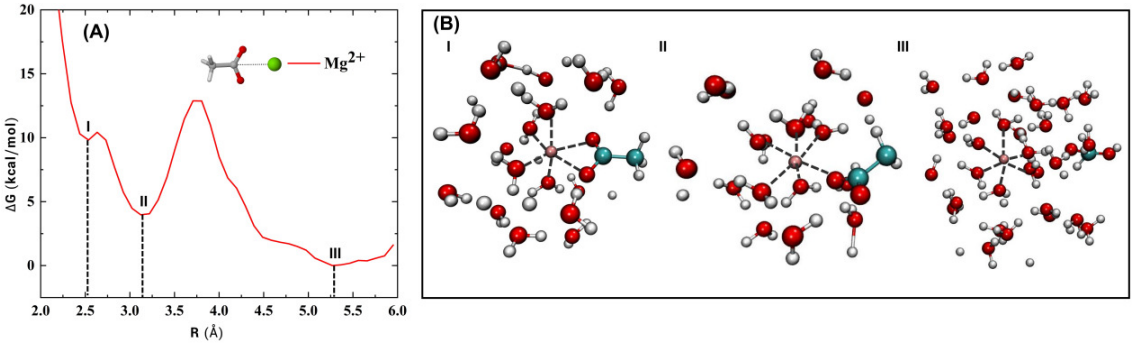


FIG. 4 (A) Potential of mean force curve of Mg^{2+}-Ac^{-} ion pairs in aqueous solution. (B) Representative structures corresponding to the minima points of the PMF curve in the MD simulation. I, II, and III refer to contact bidentate structure, monodentate structure and solvent-shared structure, respectively. Red: oxygen atom, white: hydrogen atom, green: carbon atom, pink: magnesium atom.

ter molecules, forming an octahedral solvation structure. The PMF results are consistent with previous study and demonstrate the reliability of the trained NNP model. Furthermore, the preference of Mg^{2+} for the monodentate structure over the contact bidentate structure is observed from the PMF curve, with the energy barrier of ~5.8 kcal/mol, which is higher than our previous work. In addition, the energy barrier is 3.9 kcal/mol between the monodentate structure and the solvent-share structure when the Mg^{2+} and Ac^{-} ions are separated. The observation suggests that the smaller radius of Mg^{2+} (86 pm) compared to Ca^{2+} (114 pm) results in stronger electrostatic repulsion with the ligand. Consequently, Mg^{2+} favors an octahedral coordination geometry with 6 ligands and is less inclined to adopt a bidentate coordination arrangement.

IV. CONCLUSION

In this work, we used DeePMD simulation to investigate the solvation structure of the Mg^{2+}-Ac^{-} ion pair in aqueous solution. The accuracy of NNP within DeePMD simulations hinges on the precision of the label dataset, specifically the accuracy of the QM method employed. To circumvent the bottleneck posted by costly QM methods in DeePMD simulations, DeePKS method was used as an intermediary between inexpensive PBE and expensive SCAN calculation to predict the energy and atomic forces during the generation of the deep potential. Results demonstrated that the DeePKS model accurately predicts the correlation between energies and atomic forces compared to CP2K for the 410 data set and indicated that with significantly a few frames, the DeePKS model achieves accurate predictions. Moreover, the accuracy of the NNP generated using the SCAN-level data set predicted by DeePKS surpassed that of previous work, which is obtained by a moderate DFT method considering the computational effort. The MAE and RMSE analyses suggested close agreement between the atomic forces and energies obtained from NNP and DeePKS. The solvation structure and structural properties of the Mg^{2+}-Ac^{-} ion pair in aqueous solutions were then investigated using DeePMD simulation. The potential of mean force for the Mg^{2+}-Ac^{-} system indicated three representative structures, consistent with previous research. Furthermore, it is observed that Mg^{2+} tends to form monodentate coordination with the acetate ion in octahedral geometry with 5 water molecules indicated by a ~5.8 kcal/mol energy barrier between the monodentate and bidentate geometries. In the work, the significant challenge in utilizing machine learning approaches, that is the reliance on reference datasets obtained from high-level QM methods, is alleivated by the DeePKS methodology. In addition, DeePMD was used to generate the NNP based on the data with the accuracy of high-level wave function theory. In summary, the overall approach outlined provides a comprehensive, precise, and reliable framework for quantitative predictions of the properties of metal ions' systems or other condensed-phase systems.

V. ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (No.22373065, No.62072296, No.22222303, No.22173032, No.21933010), the National Key R&D Program of China (No.2023YFF1204903), NYU-ECNU Center for Computational Chemistry at NYU Shanghai, the Opening Project of Shanghai Frontiers Science Research Center for Druggability of Cardiovascular noncoding RNA. We gratefully acknowl-

DOI: 10.1063/1674-0068/cjcp2402023

edge the computational support from the Supercomputer Center of East China Normal University (ECNU Public Platform for Innovation 001) and Xiaogan 3D Scientific Computing Center.

[1] P. F. Li and K. M. Merz Jr., Chem. Rev. **117**, 1564 (2017).
[2] J. H. F. de Baaij, J. G. J. Hoenderop, and R. J. M. Bindels, Physiol. Rev. **95**, 1 (2015).
[3] S. Tejpar, H. Piessevaux, K. Claes, P. Piront, J. G. Hoenderop, C. Verslype, and E. Van Cutsem, Lancet Oncol. **8**, 387 (2007).
[4] K. D. Collins, Biophys. Chem. **119**, 271 (2006).
[5] M. J. Laires, C. P. Monteiro, and M. Bicho, Front. Biosci. **9**, 262 (2004).
[6] M. M. Probst, T. Radnai, K. Heinzinger, P. Bopp, and B. M. Rode, J. Phys. Chem. **89**, 753 (1985).
[7] P. Smirnov, M. Yamagami, H. Wakita, and T. Yamaguchi, J. Mol. Liq. **73/74**, 305 (1997).
[8] T. Megyes, T. Grósz, T. Radnai, I. Bakó, and G. Pálinkás, J. Phys. Chem. A **108**, 7261 (2004).
[9] R. Caminiti, G. Licheri, G. Piccaluga, and G. Pinna, Chem. Phys. Lett. **47**, 275 (1977).
[10] T. Megyes, S. Bálint, T. Grósz, T. Radnai, I. Bakó, and P. Sipos, J. Chem. Phys. **128**, 044501 (2008).
[11] N. A. Hewish, G. W. Neilson, and J. E. Enderby, Nature **297**, 138 (1982).
[12] Y. S. Badyal, A. C. Barnes, G. J. Cuello, and J. M. Simonson, J. Phys. Chem. A **108**, 11819 (2004).
[13] D. R. Carl and P. B. Armentrout, ChemPhysChem **14**, 681 (2013).
[14] F. Quilès and A. Burneau, Vib. Spectrosc. **16**, 105 (1998).
[15] W. W. Rudolph and G. Irmer, J. Solution Chem. **46**, 190 (2017).
[16] D. M. de Oliveira, S. R. Zukowski, V. Palivec, J. Hénin, H. Martinez-Seara, D. Ben-Amotz, P. Jungwirth, and E. Duboué-Dijon, Phys. Chem. Chem. Phys. **22**, 24014 (2020).
[17] E. F. Aziz, N. Ottosson, S. Eisebitt, W. Eberhardt, B. Jagoda-Cwiklik, R. Vácha, P. Jungwirth, and B. Winter, J. Phys. Chem. B **112**, 12567 (2008).
[18] Z. F. Jing, C. W. Liu, R. Qi, and P. Y. Ren, Proc. Natl. Acad. Sci. USA **115**, E7495 (2018).
[19] C. F. Schwenk, H. H. Loeffler, and B. M. Rode, J. Chem. Phys. **115**, 10808 (2001).
[20] C. I. León-Pimentel, J. I. Amaro-Estrada, J. Hernández-Cobos, H. Saint-Martin, and A. Ramírez-Solís, J. Chem. Phys. **148**, 144307 (2018).
[21] F. C. Lightstone, E. Schwegler, R. Q. Hood, F. Gygi, and G. Galli, Chem. Phys. Lett. **343**, 549 (2001).
[22] A. Sengupta, A. Seitz, and K. M. Merz Jr., J. Am. Chem. Soc. **140**, 15166 (2018).
[23] M. Y. Yang, L. Bonati, D. Polino, and M. Parrinello, Catal. Today **387**, 143 (2022).
[24] A. Y. Mehandzhiyski, E. Riccardi, T. S. van Erp, T. T. Trinh, and B. A. Grimes, J. Phys. Chem. B **119**, 10710 (2015).
[25] J. Behler and M. Parrinello, Phys. Rev. Lett. **98**, 146401 (2007).
[26] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Phys. Rev. Lett. **104**, 136403 (2010).
[27] K. T. Schütt, P. J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K. R. Müller, *31st International Conference on Neural Information Processing Systems*, Long Beach, 992 (2017).
[28] J. S. Smith, O. Isayev, and A. E. Roitberg, Chem. Sci. **8**, 3192 (2017).
[29] L. F. Zhang, J. Q. Han, H. Wang, R. Car, and W. N. E, Phys. Rev. Lett. **120**, 143001 (2018).
[30] L. F. Zhang, J. Q. Han, H. Wang, W. A. Saidi, R. Car, and W. N. E, *32nd International Conference on Neural Information Processing Systems* (*NIPS*), Montreal, 4441 (2018).
[31] C. Y. Zhang, S. W. Yue, A. Z. Panagiotopoulos, M. L. Klein, and X. F. Wu, Nat. Commun. **13**, 822 (2022).
[32] J. C. Liu, R. X. Liu, Y. Cao, and M. H. Chen, Phys. Chem. Chem. Phys. **25**, 983 (2023).
[33] J. F. Liu, J. G. Lan, and X. He, J. Phys. Chem. A **126**, 3926 (2022).
[34] Y. X. Chen, L. F. Zhang, H. Wang, and W. N. E, J. Phys. Chem. A **124**, 7155 (2020).
[35] Y. X. Chen, L. F. Zhang, H. Wang, and W. N. E, J. Chem. Theory Comput. **17**, 170 (2021).
[36] Q. Zhang and T. Zhu, Chin. J. Chem. Phys. **36**, 162 (2023).
[37] Y. X. Chen, L. F. Zhang, H. Wang, and W. N. E, Comput. Phys. Commun. **282**, 108520 (2023).
[38] W. F. Li, Q. Ou, Y. X. Chen, Y. Cao, R. X. Liu, C. Y. Zhang, D. Y. Zheng, C. Cai, X. F. Wu, H. Wang, M. H. Chen, and L. F. Zhang, J. Phys. Chem. A **126**, 9154 (2022).
[39] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).
[40] J. W. Sun, A. Ruzsinszky, and J. P. Perdew, Phys. Rev. Lett. **115**, 036402 (2015).
[41] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, WIRES Comput. Mol. Sci. **4**, 15 (2014).
[42] J. Hostaš and J. Řezáč, J. Chem. Theory Comput. **13**, 3575 (2017).
[43] J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).
[44] Y. Z. Zhang, H. D. Wang, W. J. Chen, J. Z. Zeng, L. F. Zhang, H. Wang, and W. N. E, Comput. Phys. Commun. **253**, 107206 (2020).
[45] H. Wang, L. F. Zhang, J. Q. Han, and W. N. E, Comput. Phys. Commun. **228**, 178 (2018).
[46] B. Roux, Comput. Phys. Commun. **91**, 275 (1995).
[47] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[48] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Comput. Phys. Commun. **185**, 604 (2014).
[49] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **16**, 1339 (1995).

DOI: 10.1063/1674-0068/cjcp2402023