HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2023 October 10.

Published in final edited form as:

J Chem Inf Model. 2022 October 10; 62(19): 4713–4726. doi:10.1021/acs.jcim.2c00758.

Inclusion of high-field target data in AMOEBA's calibration improves predictions of protein-ion interactions

Julian A. Melendez[†], **Vered Wineman-Fisher**[†], **Sagar Pandit**[‡], **Sameer Varma**[†,‡]

[†]Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, 4202 E. Fowler Ave., Tampa, FL-33620, USA

[‡]Department of Physics, University of South Florida, 4202 E. Fowler Ave., Tampa, FL-33620, USA

Abstract

The reliability of molecular mechanics simulations to predict effects of ion binding to proteins depends on their ability to simultaneously describe ion-protein, ion-water and protein-water interactions. Force fields (FFs) to describe protein-water and ion-water interactions have been constructed carefully, and have also been refined routinely to improve accuracy. Descriptions for ion-protein interactions have also been refined, although in an *a posteriori* manner through use of 'non-bonded-fix (NB-fix)' approaches in which parameters from default Lennard-Jones mixing rules are replaced with those optimized against some reference data. However, even after NB-fix corrections, there remains a significant need for improvement. This is also true for polarizable FFs that include self-consistent inducible moments. Our recent studies on the polarizable AMOEBA FF suggested that the problem associated with modeling ion-protein interactions could be alleviated by recalibrating polarization models of cation-coordinating functional groups so that they respond better to the high electric fields present near ions. Here we present such a recalibration of carbonyls, carboxylates and hydroxyls in the AMOEBA protein FF, and report that it does improve predictions substantially – mean absolute errors in Na^+-protein and K^+-protein interaction energies decrease from 8.7 → 5.3 and 9.6 → 6.3 kcal/mol, respectively. Errors are computed with respect to estimates from vdW-inclusive density functional theory benchmarked against high-level quantum mechanical calculations and experiments. While recalibration does improve ion-protein interaction energies, they still remain underestimated, suggesting that further improvements can be made in a systematic manner through modifications in classical formalism. Nevertheless, we show that by applying our many-body NB-fix correction to Lennard-Jones components, these errors are further reduced to 2.7 and 2.6 kcal/mol, respectively, for Na^+ and K^+ ions. Finally, we show that the recalibrated AMOEBA protein FF retains its intrinsic reliability in predicting protein structure and dynamics in the condensed phase.

svarma@usf.edu .

Supporting Information Available

The supporting information includes 6 tables and 7 figures. These provide NMA and acetate vdW parameters, vdW cross-terms for D/E-R/K interactions, protein-ion vdW cross-terms, effects of simulation protocols on NMA density and dielectric constant, effects of vdW cross-terms on D/E-R/K interactions, list of PDB structures employed in this study, chemical environments of ions in PDB structures used in this study and replica exchange acceptance ratios and dihedral angle probabilities.

Introduction

Na^+ and K^+ ions are important to many biological process.[1] Their cellular concentrations are tightly regulated through membrane channels and transporters to enable various tasks, such as nerve conduction, cell volume homeostasis, cardiac rhythms and cellular uptake of nutrients. Na^+ and K^+ ions are also shown to serve as cofactors and allosteric modulators of several enzymes, including pyruvate kinases, dehydratases, bisphosphatases, deacetylases, and chaperons.[2,3] These ions have also been shown to bind to nucleotides, aiding RNA folding and stabilizing DNA duplexes and quadruplexes.[4,5]

Molecular mechanics (MM) simulations can, in principle, provide detailed insights into the roles of ions in biological processes at atomic resolution. For this, MM simulations must capture the correct balance between an ion's hydrated and biomolecule-bound states. Achieving this balance requires (a) effective sampling of conformational space, and (b) accurate descriptions of how ions interact with both water and biomolecular chemical groups. The former issue is being addressed through developments of statistical methods, and improvements in algorithms and computer hardware.[6] Here we address the latter, where despite many efforts,[7–27] there still remains need for improved accuracy.

MM descriptors for proteins, or protein force fields (FFs), have been constructed carefully using reference data from gas and condensed phase experiments as well as high level quantum mechanics calculations, and are refined routinely to improve accuracy.[28–36] However, ions are not included during parameterization of protein FF. Ion parameters are determined separately from reference data on ion-water interactions. In simulations consisting of both proteins and ions, ion-protein interaction energies are estimated using some predefined set of mixing rules for Lennard-Jones (LJ) terms. This, however, does not guarantee the reliability of predicted ion-protein interactions. In fact, large errors of the order of several kcal/mol have been reported in many studies.[15,19,23–25,37–39]

In recent years, errors in ion-protein interactions have been reduced to some extent in an *a posteriori* manner. 'Non-bonded-fix (NB-fix)' type approaches have been used in which parameters for ion-protein interactions obtained from mixing rules are substituted with those obtained from reference data on interactions of ions with small molecules representative of protein chemical groups.[17,20–25,27] Such a strategy is elegant in that it does not require recalibration of interactions of the coordinating chemical groups with other chemical groups. In addition, implementation is straightforward and does not increase computational cost. However, in almost all applications,[17,20–25] error corrections are assigned to the Lennard-Jones (LJ) term. LJ parameters are treated as free tunable parameters without any associated physical meaning. Thus, while such a procedure may improve overall accuracies, they may not be the main source of error.[40]

Since electronic polarization contributes substantially to ion-ligand interactions,[41] reliability was also expected to improve when FFs transitioned from fixed charge approaches for electrostatic interactions to more compute intensive formalisms that included self-consistent inducible electrostatic moments.[30–32,34,36,42,43] Indeed, performance improved.[19,24,26,27] However, there is still much need for improvement. In one of the most detailed

investigations carried out on the performance of FFs in predicting ion-protein interactions,[24] it was reported that while switching from non-polarizable to a Drude-based polarizable FFs did improve performance, root-mean-square (RMS) errors in ion-protein interaction energies in several test cases exceeded 10 kcal/mol, and in the worst case, errors were greater than 20 kcal/mol. Additionally, for interactions of Na^+ and K^+ ions with ethanol and N-methylacetamide (NMA), RMS errors were reported to be greater than 5 kcal/mol, and the maximum error exceeded 10 kcal/mol.[24] These errors were determined after including 2-body NB-fix corrections, which implies that to further improve accuracies, we need to look beyond NB-fix corrections.

To address the general problem of improving ion-protein interactions in MM FFs, we analyzed a different polarizable FF, AMOEBA,[44–49] developed originally by Ponder, Ren and others. We first recalibrated the vdW parameters of Na^+ and K^+ ions to improve their interactions with water in gas and aqueous phases.[50] Expectedly, the new vdW parameters also altered their interactions with other ligands.[50,51] However, the difference between ion-ligand and ion-water interaction energies remained unaffected.[50,51] For example, while recalibration of Na^+ vdW parameters strengthen Na^+-water interactions and improved their predictions, it also strengthened Na^+-methanol interactions by about the same amount. Consequently, this left the difference between ion-water and ion-methanol interaction energies unaltered and with errors similar to those present prior to recalibration. Getting this relative balance right is critical to properly simulating ion binding to proteins in which ions shed their inner-shell waters (partially or fully) in order to coordinate with protein chemical groups.

Our studies above also supported the idea[37] that the problem associated with predicting ion-protein interactions could perhaps be ligand-centric. Along these lines, we found that the ligand's polarization term was indeed an error source.[50–52] Specifically, we found that AMOEBA's polarization response was weak at the high electric fields present near ions, and we postulated that this is what produced weak ion-ligand binding. At the same time, AMOEBA's polarization response was good at lower dipolar electric fields. We note here that the original AMOEBA protein FF was not tested for their ability to predict their interactions with ions. Back then, reference data on interactions of ions with protein chemical groups was scarce. Additionally, the focus back then was to examine the viability and the broader impacts of including polarization effects in MM force fields.[44–46,48]

To remedy this, we first used high level quantum mechanical (QM) methods, including quantum Monte Carlo and CCSD(T), which were made accessible only recently, to generate the necessary benchmarking data.[39,50–54] Then, we explored two approaches to improve the polarization model. Both yielded comparable improvements. One solution was based on the NB-fix style approach in which we assigned different polarization cross terms to each distinct ion-ligand pair.[50] In the second approach, we showed that we could recalibrate FFs of representative small molecules so that they performed well at both dipolar electric fields and the kind of high electric fields present near monovalent cations.[51,52] Our recalibration also slightly improved low-field response, which was expected because the original polarization parameters were developed to be transferable across ligands, rather than ligand specific.[49] This incorporation of high field target data during recalibration

substantially improved ion-ligand interactions, while at the same time retaining the model's original accuracy in describing dipolar ligand-water and ligand-ligand interactions in the gas and condensed phases.[51,52]

Building on this earlier work[50–52] and preceding foundational developments on AMOEBA FF,[44–49] here we carry out the following studies:

- **i.** We extend our earlier work[51] on recalibrating AMOEBA FF for small molecules to respond better to high electric fields to two additional small molecules – NMA and acetate. At each stage of model development, we optimize all parameters simultaneously to capture their interdependent correlations with target data. This multi-dimensional optimization is carried out using our ParOpt software.[40,55] Reference data for parameter optimization and model validation are taken from published experiments and also new benchmarked QM calculations.
- **ii.** We use recalibrated NMA and acetate parameters, as well as recalibrated methanol and ethanol parameters from our previous work,[51] to describe carbonyl, carboxylate and hydroxyl groups in proteins. Carbonyls, carboxylates and hydroxyls are the three primary chemical groups in proteins that are known to interact with Na^+ and K^+ ions.[2,56] We then evaluate performance of recalibrated protein FF in predicting ion-protein interaction energies in ion clusters extracted from the protein data bank (PDB). Reference values for ion-protein interaction energies are obtained from vdW-inclusive DFT that we have, over the past years, systematically benchmarked for ion-ligand interactions against experiments and "gold-standard" QM methods, including quantum Monte Carlo and CCSD(T).[39,50–54]
- **iii.** We examine whether the recalibrated AMOEBA FF retains its intrinsic reliability in reproducing reference experimental data on protein structure and dynamics in the condensed phase.

We demonstrate that recalibration of AMOEBA protein FF parameters against both dipolar and high electric field reference data substantially reduces errors in ion-protein interaction energies – mean absolute errors (MAE) in interaction energies of Na^+ and K^+ ions in PDB clusters reduce from 8.7 → 5.3 and 9.6 → 6.3 kcal/mol, respectively. We note that even after recalibration the predicted ion-protein interaction energies continue to remain systematically underestimated. This suggests that there is still important physics missing from AMOEBA's functional forms describing ion-protein interactions. Nevertheless, by applying our many body NB-fix correction to Lennard-Jones components in ion-ligand interaction energies,[40] we report that these errors are further reduced to 2.7 and 2.6 kcal/mol, respectively, for Na^+ and K^+ ions.

Methods

Quantum mechanical reference data

Reference energies are obtained using the PBE0[57,58] exchange-correlation density functional supplemented by Tkatchenko-Scheffler[59] corrections for dispersion (vdW). The 25% exact exchange in PBE0 alleviates the delocalization error in DFT-based approximations, which is

particularly important for hydrogen-bonded and charge transfer systems.[59] All PBE0+vdW calculations are performed using the FHI-aims package[60] with 'really tight' basis sets. Total energies and electron densities are converged to within 10^{-6} eV and 10^{-3} eV/Å respectively.

Our choice of this vdW-inclusive DFT comes from three observations. Firstly, it is shown that PBE0+vdW yields an accuracy of 0.3 kcal/mol in comparison to "gold standard" quantum chemical reference data for a wide range of intermolecular interactions in molecular dimers.[61] Secondly, we have reported previously[50–54] that compared to quantum Monte Carlo calculations and CCSD(T) calculations in the complete basis set (CBS) limit, PBE0+vdW yields a MAE of 0.93 kcal/mol for interaction energies of monovalent cations (Na^+, K^+ and NH_4^+) with homogenous clusters of various small molecules (waters, alcohols, amides, aromatics and acetates). At the same time we do note that error in PBE0+vdW is systematic with respect to quantum Monte Carlo simulations and CCSD(T)/CBS in that PBE0+vdW generally overbinds. In the cases we examined, the extent of overbinding was found to be of the order of 1–2 kcal/mol. Finally, we have reported that, within Harmonic approximation, PBE0+vdW yields gas phase ion-ligand binding free energies with MAE of less than 1 kcal/mol with respect to experiments.[39,52]

Electron densities and polarizability tensors are obtained from second-order Møller-Plesset perturbation (MP2) theory[62] as implemented in Gaussian09.[63] We use Dunning's correlation consistent basis sets[64] augmented with diffuse functions, aug-cc-pVTZ, which was also used to construct the original AMOEBA model.

Molecular dynamics

All molecular dynamics (MD) simulations are carried out using Tinker9.[65] Temperature is regulated using an extended ensemble approach[66] and with a coupling constant of 0.1 ps. Pressure (under NPT conditions) is regulated using a Monte Carlo approach.[67,68] Equations of motions are integrated using the RESPA algorithm with an outer time step of 1 fs.[69] Electrostatic interactions outside the direct space, defined with a cutoff of 9 Å, are computed using the particle mesh Ewald (PME). vdW interactions are computed for inter-atomic distances smaller than 9 Å, with isotropic long range corrections applied beyond vdW cutoff. The convergence cutoff for induced dipoles is set at 0.01 D. Other control functions and parameters are set to their default values.

Multiparameter optimization

FF parameters are optimized using the Nelder-Mead simplex-based algorithm implement in our ParOpt software.[40,55] All parameters at a given FF development stage are optimized simultaneously. This is done to capture inter-parameter dependencies on target data. For each optimization run, we use original simplex coefficients for reflection ($\alpha = 1.0$), contraction ($\beta = 0.5$), expansion ($\gamma = 2.0$) and size ($\delta = 0.5$).[70,71] We use a convergence criteria of 10^{-5} and optimization is run using many different initial values of parameters selected randomly within supplied parameter ranges.[47] The number of times we run optimization to find parameter sets, and the parameter ranges we set depend on the explored phase space. Both these choices are context dependent and so are discussed at appropriate places in the results section.

Results and Discussion

First we recalibrate NMA and acetate molecules. Then we transfer their parameters, as well as those of methanol and ethanol molecules, which we recalibrated previously,[51] to proteins and evaluate performance of recalibrated protein FF in predicting ion-protein interaction energies. Finally, we examine whether the updated FF retains its intrinsic reliability in reproducing reference experimental data on protein structure and dynamics.

Note that when we refer to the original AMOEBA model, we are referring to the version distributed as part of the Tinker8 package. This FF was calibrated against the 2003 water model[44] and released in 2013.[49] This FF also includes the updates made in 2018 where the acetate molecule was recalibrated to improve Glu/Asp carboxylate parameters,[27] and explicit cross terms were introduced to improve carboxylate-amine and carboxylate-guanidinium interactions. We also note that there are currently two sets of AMOEBA parameters distributed for Na^+ and K^+ ions, one developed in 2003[45] and other refined recently in 2018.[72] We find that both sets of parameters perform similarly in predicting ion-protein interactions, and therefore, we include results from only the original parameters.[45] Recently, we had also refined Na^+ and K^+ ion parameters to improve their interactions with water in the gas and condensed phases.[50] In the developments we carry out here, we use our ion parameters, but the improvements we see are not due to our recalibration of ion parameters. We have shown previously that improving ion-water interactions has a negligible effect on predictions of relative interactions of ions with other chemical groups.[50]

Recalibration of NMA and acetate

We first recalibrate polarization models to improve their responses at high electric fields. Electronic polarization within AMOEBA is handled using an iterative atomic dipole induction scheme.[44] The anisotropic molecular polarizability is accounted for by assigning isotropic polarizabilities (α) to each atom. Additionally, to prevent a polarization catastrophe, polarization at short range is damped using the Thole scheme.[73] Damping is applied to one of the two polarization sites using the smearing function $\rho = (3a/4\pi)e^{-ar_{ij}^3/\sqrt{\alpha_i\alpha_j}}$ were r_{ij} is the distance between the two sites i and j, and 'a' is the dimensionless width parameter of the smeared charge distribution. In cases where interacting atoms have different 'a', the smaller value is chosen for applying damping. The original AMOEBA model uses the same atomic polarizabilities for each chemical element as the ones suggested by Thole, except for aromatic groups. The universal width factor of 0.57 suggested by Thole was modified to 0.39 in the original model as it better reproduced energies of water clusters.[44] The original atomic polarizabilities, although highly transferable, are known to systematically underestimate molecular polarizabilities compared to experiment, as also noted the original developers.[48]

We recalibrate the polarization models of NMA and acetate separately. Polarizabilities of all non-hydrogen atoms are optimized simultaneously using the Nelder-Mead algorithm (see methods) with upper bounds set at 2.5 $Å^3$. The target function for optimization is the root-mean-square error (RMSE) of molecular polarizability tensors computed with respect to values obtained from MP2 theory. Optimization is started using a random seed that assigns

random initial values of a to non-hydrogen atoms, and so each optimization can produce a parameter set that, in principle, is unique. We carry out 50 optimizations that generates 50 parameter sets. We also examine three different Thole damping width coefficients: (i) a = 0.39, which was used in the original AMOEBA model, (ii) a = 0.5, which is closer to the one proposed by Thole, and (iii) we also set the width coefficient to be a free parameter during optimization.

We find many parameter sets for polarization models that reproduce molecular polarizability tensors, and at the same time excel at simultaneously predicting both low and high field responses. Field responses are determined by calculating induced dipole moments along different directions generated by a point charge (+1) placed at different distances and along different orientational vectors about the small molecules. To examine how our choices for polarization parameter sets affect the overall model, we select five sets for NMA and five sets for acetate. These sets are quite different from each other (Table 1). For acetate, we did not find any parameter set with Thole damping coefficients of a = 0.39 or a = 0.50 for which polarizability RMSE was less than 0.2 $Å^3$ and $\alpha < 2.5$. Nevertheless, many parameter sets were available with a = 0.61, and to get consistency between acetate and NMA, we set a = 0.61 to select all five sets. We also note that this value of 'a' is actually much closer to that of 0.57 suggested originally by Thole.[73]

The polarizability tensors produced by the selected five sets are provided in Table 2. Figure 1 shows the performance of these sets in predicting field responses along selected directions. Compared to the original model, we note improvements in both polarizability tensors and in prediction of field responses. We also note that the performance gain is much greater in the high field region that is relevant to account for a response in the presence of monovalent cations. However, the performances of the five selected models are almost indistinguishable from each other.

Next, for each these five sets we determine vdW interaction parameters. AMOEBA uses a buffered 14–7 functional form to describe pairwise additive vdW interactions,[47]

$$U_{vdw} = \epsilon_{ij}\left(\frac{1.07}{\rho_{ij} + 0.07}\right)^7\left(\frac{1.12}{\rho_{ij}^7 + 0.12} - 2\right). \quad (1)$$

Here ϵ_{ij} is the potential well depth, and $\rho_{ij} = R_{ij}/R_{ij}^0$, where R_{ij} is the distance between sites i and j, and R_{ij}^0 is the minimum energy distance. For heterogeneous atom pairs, R_{ij}^0 and ϵ_{ij} are determined from the following mixing rules:

$$R_{ij}^0 = \frac{(R_{ii}^0)^3 + (R_{jj}^0)^3}{(R_{ii}^0)^2 + (R_{jj}^0)^2} \text{ and } \epsilon_{ij} = \frac{4\epsilon_{ii}\epsilon_{jj}}{(\epsilon_{ii}^{1/2} + \epsilon_{jj}^{1/2})^2} \quad (2)$$

We optimize vdW parameters, R_{ii}^0 and ϵ_{ii}, of all non-hydrogen atoms simultaneously using the Nelder-Mead algorithm. Target data comes from ligand-ligand clusters, similar to those used in the development of the original model, as well as ion-ligand clusters. Target data

is obtained from our reference vdW-inclusive DFT method (see methods), and all clusters are energy minimized. Specifically, for calibrating NMA, we use NMA-water and NMA-NMA dimers, and also Na^+-$(NMA)_{1–4}$ and K^+-$(NMA)_{1–4}$ clusters containing 1–4 NMA molecules. To determine acetate parameters, we use acetate-water, Na^+-$(acetate)_{1–2}$ and K^+-$(acetate)_{1–2}$ clusters containing 1–2 acetate molecules. We use both energy and distance data from these clusters. From ion-ligand clusters, we take distances between ions and all non-hydrogen (heavy) atoms of ligands, and from ligand-ligand clusters, we take distances between all non-bonded heavy atoms. To use both energy and distance data, we define target functions as sums of mean absolute percent errors (MAPE) in energies and distances.

For each of the five polarization models, we carry out twenty five cycles of optimization, which generates twenty five sets of vdW parameters. From these twenty five sets, we select the set that yields the lowest error in target function. The vdW parameters that we select for each of the five NMA and five acetate polarization models are listed in tables S1 and S2, respectively, in the supporting information.

Tables 3 and 4 compare the performances of the five NMA and five acetate models in yielding target distances and energies, respectively. In the recalibrated NMA models, RMSEs in ion-NMA distances range between 0.23 and 0.28 Å which are comparable to the RMSE in the original model (Table 3). In terms of target energies, we note that while recalibration substantially improves RMSEs in ion-NMA binding energies, it does come with a little compromise on accuracies of NMA-NMA and NMA-water dimerization energies. Additionally, we note that Sets 3, 4 and 5 perform better that sets 1 and 2 in yielding ion-NMA binding energies. The scenario with acetate recalibration is a bit different. We note that all five sets are comparable to each other in terms of errors with respect to target data. We see that acetate recalibration improves both acetate-water and acetate-ion binding energies (Table 4), but here there is a little compromise of ~ 0.05 Å in ion-acetate distances (Table 3).

To examine how these re-calibrations affect the balance of interactions of ions with water, we consider the following substitution reaction

$$\mathrm{AW}_n + n\mathrm{X} \rightleftharpoons \mathrm{AX}_n + n\mathrm{W}, \tag{3}$$

and determine the associated substitution energy as

$$\Delta E_{sub} = E_{\mathrm{AX}_n} - nE_{\mathrm{X}} - E_{\mathrm{AW}_n} + nE_{\mathrm{W}}\,. \tag{4}$$

Here ‘A’ refers to either a Na^+ or K^+ ion, ‘W’ refers to water and ‘X’ refers to NMA or acetate. Figure 2 shows the results obtained for all models. The original model underestimates the preference for NMA for clusters containing 1–3 molecules, and while it correctly predicts ΔE_{sub} for K^+ binding to 4 NMA, it overestimates the favorability of Na^+ for the cluster consisting for 4 NMA molecules. For acetate, we see that the original model again underestimates the preference of sodium for acetate and the error is largely corrected in the new sets with a slight overestimation of the preferences for $n = 1$ that further reduces going to $n = 2$. For K^+ we see no major changes compared to the original model which

correctly predicts the QM reference data. In general, recalibration improves ΔE_{sub} for all models, but the best improvements for NMA are seen for sets 3, 4 and 5, and for acetates, best improvements are seen for sets 2 and 5.

Overall, we see that improving NMA and acetate polarizability models does improve their gas phase interactions with ions, and without compromise on their gas phase interactions with uncharged polar small molecules. Next we examine the performance of these recalibrated models in predicting experimental condensed phase properties. Specifically, we simulate liquid NMA and determine its density (ρ), heat of vaporization (ΔH_v), self-diffusion constant (D_{self}) and the dielectric constant (ϵ). For acetate and also for NMA, we determine hydration free energies ($\Delta G(aq.)$).

To determine liquid NMA properties, we first carry out MD simulations of cubic boxes containing 512 NMA molecules for 1 ns under NPT simulations, with P=1 atm and T=308 K. We note that for all five NMA models, densities equilibrate within the first 500 ps. We, therefore, estimate average densities from the final 500 ns. From these MD simulations, we select snapshots closest to the average density and initiate 30 ns long MD simulations under NVT conditions. We use the final 29 ns of this trajectory to determine the remaining properties.

Heats of vaporization (ΔH_v) are computed as

$$\Delta H_v = \left(\left\langle U_{gas}(T)\right\rangle - \left\langle U_{liquid})T\right\rangle\right) + RT \tag{5}$$

where $U_{gas}(T)$ and $U_{liquid}(T)$ are the total potential energies of the molecules in the gas and liquid phases, and R is the gas constant. $U_{liquid}(T)$ are computed from each snapshot of the NVT runs and then normalized by the number of molecules (N = 512). $U_{gas}(T)$ are computed from separate MD simulations of a single NMA molecule under isothermal conditions with a number density set at 0.024 nm^{-3}.

Self-diffusion constants are computed using the Einstein equation, and corrected for system-size effects using a correction term[75] obtained from the thermodynamic theory of diffusion,[76,77]

$$D_{self} = \lim_{\Delta t \to \infty} \left\langle r^2(\Delta t)\right\rangle / 6\Delta t + k_b T\alpha / 6\pi\eta L \tag{6}$$

The second term in the equation above is the correction term, where L is the unit length of the box, η = 0.0344 is the viscosity of the solvent taken from its experimental value at 308 K[78] and α = 2.837 is a constant of the model. In the first term, the mean square displacement ($r^2(\Delta t)$) is calculated for the center of mass of each molecule and then averaged over all molecules. Limiting value of ratio in the first term $\left\langle r^2(\Delta t)\right\rangle / \Delta t$ is calculated from slope of the line from Δt = 2 ns to Δt = 27 ns. The errors provided are the difference of the diffusion coefficients obtained from doing a fit up from Δt = 2 to Δt = 14.5 ns and from Δt = 14.5 ns to Δt = 27 ns.

Finally, dielectric constants are determined from dipole moment fluctuations[79]

$$\epsilon = \epsilon_\infty + \frac{1}{3V\epsilon_0 k_B T}(\langle \mathbf{M}^2 \rangle - \langle \mathbf{M} \rangle^2) . \quad (7)$$

Here, $\mathbf{M} = \sum \vec{\mu}_i$ is the vectorial sum of dipole moments $\vec{\mu}_i$ of all NMA molecules in the box. corresponds to the individual dipole moment of the molecules in the box, and ϵ_∞) is determined from the Clausius-Mossotti approximation[80]

$$\frac{\epsilon_\infty - 1}{\epsilon_\infty + 2} = \frac{4\pi\alpha}{3\langle v \rangle} . \quad (8)$$

In the relationship above, V is the box volume, k_b is Boltzmann-constant, ϵ_0 is the vacuum permittivity, $\langle v \rangle$ is the average molecular volume, and α is the molecular polarizability. To monitor convergence of dielectric constant, we calculate it in increments of 200 ps (see Figure S2 in Supporting Information). Based on this criteria, we report values determined from the asymptote calculated from the final 10 ns of the simulations. Additionally, we understand that in the thermodynamic limit, $\langle M \rangle = 0$ for isotropic systems, and we do find that the contribution of the second term in equation 7 is very small and does not affect the results qualitatively.

Table 5 lists the predictions of liquid NMA from the five recalibrated parameter sets, and also compares them to experimental values and those obtained from the original model. The first thing we note is that none of the sets, including the original set, reproduce all the experimental liquid NMA properties. Among these, Set_5 performs the best, and its overall predictions are comparable to that of the original model.

Set_5, however, like all other sets and the original model, does not perform satisfactorily in reproducing the NMA's dielectric constant. Note that the experimental value we list in Table 5 is taken from a polynomial fit of values measured in the range 303 – 393 K, as also used by others.[81,82] To examine the dependence of dielectric constant predictions on our protocol, we carry out the following additional tests. First, we use an identical protocol to compute the dielectric constants of liquid methanol and ethanol, whose parameters we recalibrated previously.[51] At T = 298 K, we get values of 37±1 and 25±1 for methanol and ethanol, which closely match their respective values of 33 and 24 obtained from experiment.[83,84] Next, we examine the effect of the chosen box volume. Using Set_5, we carry out two additional 30 ns long NVT simulations. Instead of starting from the coordinates with box size that matches the average volume in NPT, we start runs from coordinates in the smallest and the largest boxes observed in the NPT run and carry out 30 ns long MD simulations. Additionally, we extend our NPT run from 1 ns to 30 ns, and estimate dielectric constant from NPT ensemble by substituting V with average volume $\langle V \rangle$ in Equation 8. We find that ϵ estimated from the NPT ensemble lies within the values determined from the NVT ensembles initiated using minimum and maximum box volumes (Figure S3 in supporting information).

Next, we estimate hydration free energies ($\Delta G(aq.)$) of NMA and acetate for the different parameter sets. We estimate them under NVT conditions and using Bennett's acceptance ratio (BAR).[88] The conformational ensembles needed for BAR are obtained from MD

of cubic boxes containing 1500 water molecules, and using a protocol we described previously,[52] but with the following two exceptions. Firstly, vdW interactions are decoupled using a different set of λ = {1.0, 0.9, 0.8, 0.75, 0.7, 0.65, 0.6, 0.5, 0.4, 0.2, 0.0}.[89] Secondly, each window is run for 500 ps and the last 400 ps are used to determine the hydration free energy. Note that we also subtract out the contribution of discharging the solute in vacuum, for which we use the same protocol. Following BAR calculations, the correction term $-RT\ln C_l/C_g$ is added to adjust for solute concentration differences between gas (C_g = 0.041 M) and condensed (C_l) phases.[90] For acetate, we use C_l = 1 M,[91] and for NMA, we use C_l = 0.01 M.[92]

The results of these calculations are listed in Table 6 and compared to experimental estimates. Note that the experimental values for acetate[91] are determined using the TATB scheme, which is not expected to include air–water interface potential effects, and so can be compared directly to results of our BAR calculations with concentration term corrections.[52,93,94] For acetate, we find that all five sets perform similarly, and we note marked improvement of recalibration on predicted hydration free energy. In the case of NMA, we see that all the recalibrated models except for Set_3 perform similar to the original model, with hydration free energies underestimated by 1.4 – 1.9 kcal/mol.

Taken together, we note that improving polarizability descriptors in NMA and acetate improve their gas phase interactions with ions. Additionally, they perform just as well as the original parameters in predicting their interactions with uncharged polar small molecules in both gas and condensed phases. For NMA, we find that Set_5 outperforms all other parameter sets, and so we select its parameters to describe chemical groups in proteins. In the case of acetate, we do not find any standout criteria to select between its five sets. But we do note that for acetate Set_3 and Set_5 slightly outperform the others in predicting ΔE_{sub} in Figure 2, and Set_5 does slightly better than Set_3 in predicting $\Delta G(aq.)$. We, therefore, select Set_5 to describe carboxylate groups in proteins.

Modifications to AMOEBA protein FF

We use the recalibrated parameter sets of NMA and acetate, as well as the recalibrated parameters of methanol and ethanol from previous work,[51] to describe their corresponding chemical groups in proteins. We use two schemes. In the first scheme, we retain the original mapping between small molecules and protein chemical groups. In this scheme, we use NMA to describe peptide backbones, peptide termini (ACE/NME) and side chains of Asn and Gln. Acetate provides parameters for side chain carboxylates in Glu and Asp, and methanol is used to describe hydroxyls in Tyr, Ser and Thr side chains. In the second scheme, we make three modifications to the original scheme. Firstly, while we continue to use methanol to describe Tyr's hydroxyl group, we use ethanol to describe side chains of Ser and Thr. Secondly, in addition to describing backbone N,H,C and O atoms with those of NMA, we also use NMA's methyl carbon to describe C_α of backbone, methyl carbons of NME and ACE caps, C_β of Asn and and C_γ of Gln. We reason that since polarization of NMA was calibrated as a whole, this approach more readily captures the chemical environment of NMA. Finally, in line with the same idea, we use ethanol's CH_2 to describe Ser/Thr C_β and ethanol's CH_3 to describe Thr's methyl group. As we will see

later, these modifications discernibly improve ion-protein interactions. In both the original and modified schemes, we are assuming that the polarization and vdW parameters of small molecules appropriately describe their corresponding chemical groups in proteins, although we do note that the atomic multipoles in small molecules are a bit different from those in protein chemical groups.

Since we are updating Glu and Asp parameters, we also need to examine how this affects their salt bridge interactions with Arg and Lys. We note that the original FF consists of explicit LJ-based NB-fix type cross terms for carboxylate-amine and carboxylate-guanidinium interactions. We generate new cross-terms for the updated carboxylate parameters, which are listed in Table S3 of the Supporting Information. Figure S4 of the Supporting Information shows that the new carboxylate parameters along with the new cross terms perform well against our reference QM method.

Protein-ion interactions

To examine the performance of the updated protein FF in predicting ion-protein interactions, we use it to determine Na^+ and K^+ ion interaction energies in clusters extracted from the PDB. Clusters are extract from PDB in an unbiased manner, as follows. We consider X-ray structures with resolutions better than 2.5 Å. From these X-ray structures, and for each Na^+ or K^+ ion in them, we extract amino acids and waters that are within 6 Å from the ion. Even if a single atom of an amino acid is within this cutoff, the entire amino acid is included as part of the cluster. At this stage the clusters contain an ion, peptide fragments and waters. Peptide fragments in each cluster are then analyzed and if their ends are separated by less than two amino acids, the connecting amino acids are also included in the cluster, even if the connecting amino acids are outside the 6 Å cutoff. The peptide fragment ends are then capped with ACE and NME, and all missing hydrogens, including those on waters are added. We then select the thirty smallest (non-redundant) Na^+ and thirty smallest K^+ clusters. The largest Na^+ and K^+ clusters that we consider consist of 136 and 145 atoms, respectively. These clusters differ from each other in terms of the numbers of hydroxyls, carboxylates and carbonyls as well as numbers of waters in the ion's first coordination shell. Details of each cluster are included in Table S4 of the supporting information.

The selected clusters are then energy minimized using the original AMOEBA FF using an RMS potential gradient cutoff of 0.01 kcal mol^{-1} $Å^{-1}$. Note that during optimization, we applied position restraints on all the atoms that were present in X-ray structures. Consequently, optimization resulted in changes in positions of only those atoms (mainly hydrogens) that were absent from X-ray structures. Our choice for optimizing hydrogen positions using only one FF was to get consistency in structure when comparing different force fields. Additionally, examination of clusters indicates that the ion-coordinator distances span across a range of 1.2 Å (0.3 Å smaller than the QM minima and 0.9 Å larger than the QM minima), which gives us a good distribution of geometries around energy minima.

Using each of these clusters, we determine interaction energies (ΔE) between ions and everything else in the clusters. Figure 3 compares the interaction energies obtained from the original and recalibrated protein FFs against those obtained from our benchmarked

QM method. We note first that predictions from the original FF correlate excellently with reference QM, even for cases in which interaction energies are large ($\Delta E < -200$ kcal/mol). However, the original FF systematically underestimates the strengths of the ion-protein interactions. The MAE is somewhat high − 8.7 and 9.6 kcal/mol, respectively, for Na^+ and K^+ ions – but comparable to the performance of another polarizable FF.[24] In some cases, errors exceed 15 kcal/mol, which is also expected based on performance of another polarizable FF.[24] We note, however, that some of the error in the AMOEBA FF may be artifactual due to the inherent bias in the reference QM method that we are using. We have reported previously that the employed PBE0+vdW DFT tends to overbind monovalent cations by 1–2 kcal/mol compared to CCSD(T)/CBS.[50–54]

Recalibration of parameters does reduce the MAE in ion-protein interaction energy from 8.7 → 5.3 and 9.6 → 6.3 kcal/mol, respectively, for Na^+ and K^+ ions, and the maximum errors are also reduced (Figure 3). These errors are larger than those that we got for model clusters used during calibration of small molecule parameters (Table 4). To examine the source of this residual error, we first examine the chemistries of PDB clusters (Table S4 of the supporting information). However, we find no clear correlations. Another possibility could be that residual errors are associated with the approximations that are typically made to transfer small molecules parameters to protein chemical groups. In fact, we do note that we get some improvement in predictions when we modify the scheme to map small molecule parameters to proteins. Specifically, representing Ser/Thr side chains using ethanol parameters instead of methanol parameters, and transferring NMA's methyl carbon parameters to protein backbone and Gln/Asn side chains does improve predictions by about a kcal/mol (see Figure S5 of Supporting Information). Additionally, it is possible that some of the residual error may be due to the differences between small molecule and protein chemical group multipoles. Finally, it is also possible that the residual errors could be due to the absence of some essential physics missing from the classical Hamiltonian. This is supported by the observation that the predicted ion-protein interaction energies are still systematically underestimated. For example, our previous work[37] reported on presence of non-negligible charge penetration effects in ion-ligand interactions, and based on recent studies, their explicit descriptions in AMOEBA's classical Hamiltonian does improve predictions in model ion-ligand clusters.[95,96]

We now examine if these issues can be circumvented and if the residual errors can be reduced by relaxing the need to adhere to the use of LJ mixing rules (Equation 2). In a recent study, we showed that errors in ion-ligand interactions, even in non-polarizable FFs, can be reduced substantially by applying many-body NF-fix corrections (MB-NB-fix).[40] The basic idea underlying MB-NB-fix is essentially the same as traditional NF-fix approach,[17,20–25,27] but here we also strive to capture (i) inter-parameter dependencies by optimizing all parameters simultaneously and (ii) many-body effects by including many-body clusters during parameter optimization.[40] We use interaction energies from the PDB clusters above as target data and determine R_{ij}^0 and ϵ_{ij} parameters for interactions between each ion type and protein carbonyls, hydroxyls, and carboxylates. To sample the extensive multi-dimensional phase space, we carry out 200 cycles of optimization, and start each cycle from a random assignment of initial parameter values. Figure 4 provides an example of

the explored parameter space for K^+ and the carbonyl oxygen over 200 cycles. Each cycle generates a vdW parameter set, and we select the one that yields the smallest error.

The final set of parameters we chose are listed in Table S5 of the Supporting Information. We note that the optimized R_{ij}^0 are close to the ones that would be derived from LJ mixing rules, and the changes are mainly in ϵ_{ij} (Figure S6 of supporting information). In almost all cases, ϵ_{ij} are larger. This perhaps produces deeper energy wells. However, note that these are mere observations and should not be taken as guidelines for future optimizations, as the relationship between ϵ_{ij} and R_{ij}^0 are complex, as also evident from Figure 4. This optimization of LJ interaction parameters reduces MAE in ion-protein interaction energies down to 2.7 and 2.6 kcal/mol for Na^+ and K^+ ions, respectively (Figure 3). The error is no longer systematic and the spread in error is also substantially reduced.

Protein structure and dynamics

While the FF modifications introduced above do improve ion-protein interactions, we note that we modified not only side chains, but also the protein backbone. Consequently, it is essential to examine how these modifications affect predictions of protein structure and dynamics.

We first examine how FF modifications affect the dihedral potential energy surface of alanine dipeptide (ACE-Ala-NME). We generate alanine dipeptide conformations regularly spaced about its backbone dihedrals (ψ, ϕ), and energy minimize them with constraints on ψ and ϕ. Geometry optimizations are carried out using a RMS potential gradient cutoff of 0.01 kcal mol^{-1} $Å^{-1}$. We compare the resulting minimum energy potential energy maps obtained from original and recalibrated FF against that we obtained previously using CCSD(T)/cc-pVTZ (frozen core).[97] This comparison in shown in Figure 5. The original model shows great agreement with the reference QM with a MAE of 1.04 kcal/mol. Recalibration has a minor effect on the overall topology of the potential energy surface, and slightly increases the MAE to 1.12 kcal/mol. Predictions of the three most stable minima (C_{7eq}, C_5 and C_{7ax}) are also maintained, and the recalibrated model also predicts a higher energy barrier between the [C_{7eq}, α_L] and [C_{7ax}, α_R] regions. This suggests that a recalibration of backbone dihedral parameters is perhaps not necessary.

To better gauge whether backbone dihedral parameters as well as those involving C_β atoms need to be recalibrated, we compute J-coupling for $(Ala)_5$ peptide and compare results against experiment. To mimic experimental conditions,[98] we protonate both the N- and C-termini of $(Ala)_5$. The peptide is placed in a cubic box whose size we determined by imposing a minimum distance of 8 Å between the peptide and the cell wall. The simulation box contains 630 waters and 1 chloride ion. To enhance conformational sampling, we use temperate replica exchange MD (t-REMD)[99] implemented using an in-house shell script. We use 32 replicas with temperatures distributed exponentially between 284 and 600 K. Each replica is run under NVT conditions, and the volume for each replica is determined from an equilibration run of 1 ns under NPT conditions, with T=298 K and P=1 atm. Replica exchange attempts are made every 10^3 steps. The resulting acceptance ratios of neighbor temperature exchanges were found between the recommended[99] values of 0.1 and

0.3. The specific values for each temperature pair can be found in Table S6 of the Supporting Information.

Each of the 32 replicas are simulated for 5×10^7 steps. We use the final 4/5th of the trajectory at T=298 K and the Karplus relationship[100] to compute J-coupling constants. Error estimates are determined by dividing the analyzed trajectory into two halves. We also note that the dihedral distributions in the two halves are similar (see Figure S7 of Supporting Information). The J-coupling constants computed for the recalibrated FF are listed in Table 7 and compared to experimental estimates[98] and also to those obtained from the original FF.[49] We find that both the original and the recalibrated FFs perform excellently against experiment with RMSEs of 0.33 and 0.68, respectively.

Finally, we evaluate the effect of recalibration on predicting protein structure and dynamics. We select ubiquitin as our model protein as it has a mixture of secondary structure motifs (β-sheets and α-helices) of varying lengths[101] and second order parameters (S^2) for each residue are also available from NMR relaxation experiments.[102] We solvate X-ray structure (1UBQ)[101] in a cubic box containing 7014 waters and 150 mM of NaCl. We perform a 200 ns long MD simulation under NPT conditions with T=298 K and P=1 atm.

Figure 6 shows the root-mean-square displacement (RMSD) of backbone atoms relative to X-ray structure. Just as the original FF, the recalibrated FF also maintains the overall backbone structure with an average RMSD of 1.07 Å. Figure 7 compares the backbone ^{15}N-^{1}H second order parameters for each residue against those obtained from NMR experiments.[102] The recalibrated FF reproduces correctly protein dynamics in both low flexibility regions (β-sheets and α-helices) and high flexibility regions (loops and termini). We compute order parameters using an in-house implementation of the isotropic reorientational eigenmode dynamics (iRED) formalism.[103] iRED does not rely on fitting structural snapshots in trajectory and instead extracts internal motions and global tumbling from a covariance matrix of the desired interactions. We construct the covariance matrix using the second-order Legendre polynomial of the angle between pairs of amide bond vectors and average over all frames of the trajectory. The reorientational modes and corresponding eigenvalues are then obtained by diagonalizing the covariance matrix. The modes are then sorted and the corresponding order parameters (S^2) for each residue are computed from the contribution of each mode excluding the largest five modes that correspond to global tumbling.

Overall, we find that AMOEBA's recalibration retains its intrinsic reliability in predicting protein structure and dynamics in the condensed phase.

Conclusion

We have presented an improved protein model for interactions with Na^+ and K^+. Our model builds upon the already very accurate AMOEBA model and improves the predictability of carbonyl, carboxylate and hydroxyl interactions with ions. We also show that the recalibrated model does not affect AMOEBA's intrinsic reliability in predicting protein structure and dynamics. We have shown that errors in transferability and protein-ion

interactions can be reduced substantially by first improving the polarization model of the small molecules and then recalibrating the vdW parameters using both high- and low-field target data. We also show that by incorporating a more inclusive description of the small molecules when the parameters are transferred to proteins, the error in protein-ion interactions is also reduced. Nevertheless, the error after transferring such parameters is still larger compared to the one obtained during optimization of the small molecules. We introduce a many-body NB-fix correction that reduces the error to nearly chemical standards. The full set of parameters can be downloaded from http://labs.cas.usf.edu/cbb/research.htm. Our method presented here is general and can be applied outside the AMOEBA model to improve protein-ion interactions in other force fields.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

The authors acknowledge the use of computer time from Research Computing at USF, and funding from the National Institute of Health (Grant No. R01 GM118697). The authors also thank Guy Dayhoff for help with software installation.

Data and Software Availability

Software packages, Tinker9, FHI-aims and Gaussian9 are available at https://github.com/TinkerTools/, https://fhi-aims.org and https://gaussian.com, respectively. New parameters for the AMOEBA force field are provided in the Supporting Information and the updated AMOEBA force field PRM file can be downloaded from http://labs.cas.usf.edu/cbb/research.htm. PDB structures used in this study are tabulated in the Supporting Information.

References

(1). Alberts B; Johnson AD; Lewis J; Morgan D; Raff M; Robert K; Walter P Molecular Biology of the Cell; W. W. Norton and Company, 2014; p 1464.

(2). Page MJ; Di Cera E Role of Na^+ and K^+ in Enzyme Function. Physiol. Rev. 2006, 86, 1049–1092. [PubMed: 17015484]

(3). Vašák M; Schnabl J In The Alkali Metal Ions: Their Role for Life; Sigel A, Sigel H, Sigel RKO, Eds.; Springer International Publishing: Cham, 2016; pp 259–290.

(4). Largy E; Mergny J-L; Gabelica V In The Alkali Metal Ions: Their Role for Life; Sigel A, Sigel H, Sigel RKO, Eds.; Springer International Publishing: Cham, 2016; pp 203–258.

(5). Auffinger P; D'Ascenzo L; Ennifar E In The Alkali Metal Ions: Their Role for Life; Sigel A, Sigel H, Sigel RKO, Eds.; Springer International Publishing: Cham, 2016; pp 167–201.

(6). Schlick T Molecular-dynamics Based Approaches for Enhanced Sampling of Long-time, Large-scale Conformational Changes in Biomolecules. F1000 Biol. Rep. 2009, 1, 51. [PubMed: 20948633]

(7). Lybrand TP; Kollman PA Water–water and Water–ion Potential Functions Including Terms for Many Body Effects. J. Chem. Phys. 1985, 83, 2923–2933.

(8). Perera L; Berkowitz ML Many Body Effects in Molecular Dynamics Simulations of Na+(H2O)n and Cl−(H2O)n Clusters. J. Chem. Phys. 1991, 95, 1954–1963.

(9). Grossfield A; Ren P; Ponder JW Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. J.Amer. Chem. Soc. 2003, 125, 15671–15682. [PubMed: 14664617]

(10). Carrillo-Tripp M; Saint-Martin H; Ortega-Blake I A Comparative Study of the Hydration of Na^+ and K^+ with Refined Polarizable Model Potentials. J. Chem. Phys. 2003, 118, 7062–7073.

(11). Spångberg D; Hermansson K Many-body Potentials for Aqueous Li^+, Na^+, Mg^{2+}, and Al^{3+}: Comparison of Effective Three-body Potentials and Polarizable Models. J. Chem. Phys. 2004, 120, 4829–4843. [PubMed: 15267343]

(12). Lamoureux G; Roux B Absolute Hydration Free Energy Scale for Alkali and Halide Ions Established from Simulations with a Polarizable Force Field. J. Phys. Chem. B 2006, 110, 3308–3322. [PubMed: 16494345]

(13). Whitfield TW; Varma S; Harder E; Lamoureux G; Rempe SB; Roux B Theoretical Study of Aqueous Solvation of K+ Comparing ab Initio, Polarizable, and Fixed-Charge Models. J. Chem. Theory Comput. 2007, 3, 2068–2082.

(14). Lee Warren G; Patel S Hydration Free Energies of Monovalent Ions in Transferable Intermolecular Potential Four Point Fluctuating Charge Water: An Assessment of Simulation Methodology and Force Field Performance and Transferability. J. Chem. Phys. 2007, 127, 064509(01)–064509(19). [PubMed: 17705614]

(15). Warren GL; Patel S Comparison of the Solvation Structure of Polarizable and Nonpolarizable Ions in Bulk Water and Near the Aqueous Liquid-Vapor Interface. J. Phys. Chem. C 2008, 112, 7455–7467.

(16). Joung IS; Cheatham TE Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. J. Phys. Chem. B 2008, 112, 9020–9041. [PubMed: 18593145]

(17). Baker CM; Lopes PEM; Zhu X; Roux B; MacKerell AD Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. J. Chem. Theory Comput. 2010, 6, 1181–1198.

(18). Luo Y; Roux B Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. J. Phys. Chem. Lett. 2010, 1, 183–189.

(19). Varma S; Rogers DM; Pratt LR; Rempe SB Design Principles for K^+ Selectivity in Membrane Transport. J. Gen. Physiol. 2011, 138, 279–279.

(20). Yoo J; Aksimentiev A Improved Parametrization of Li^+, Na^+, K^+, and Mg^{2+} Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. J. Phys. Chem. Lett. 2012, 3, 45–50.

(21). Fyta M; Netz RR Ionic Force Field Optimization Based on Single-ion and Ion-pair Solvation Properties: Going Beyond Standard Mixing Rules. J. Chem. Phys. 2012, 136, 124103(01)–124103(11). [PubMed: 22462831]

(22). Mamatkulov S; Fyta M; Netz RR Force Fields for Divalent Cations Based on Single-ion and Ion-pair Properties. J. Chem. Phys. 2013, 138, 024505(1)–024505(12). [PubMed: 23320702]

(23). Savelyev A; MacKerell AD Balancing the Interactions of Ions, Water, and DNA in the Drude Polarizable Force Field. J. Phys. Chem. B 2014, 118, 6742–6757. [PubMed: 24874104]

(24). Li H; Ngo V; Da Silva MC; Salahub DR; Callahan K; Roux B; Noskov SY Representation of Ion-Protein Interactions Using the Drude Polarizable Force-Field. J. Phys. Chem. B 2015, 119, 9401–9416. [PubMed: 25578354]

(25). Savelyev A; MacKerell AD Competition among Li^+, Na^+, K^+, and Rb^+ Monovalent Ions for DNA in Molecular Dynamics Simulations Using the Additive CHARMM36 and Drude Polarizable Force Fields. J. Phys. Chem. B 2015, 119, 4428–4440. [PubMed: 25751286]

(26). Soniat M; Pool G; Franklin L; Rick SW Ion association in aqueous solution. Fluid Phase Equilibria 2016, 407, 31–38, Aqueous Solutions.

(27). Jing Z; Qi R; Liu C; Ren P Study of Interactions Between Metal Ions and Protein Model Compounds by Energy Decomposition Analyses and the AMOEBA Force Field. J. Chem. Phys. 2017, 147, 161733(01)–161733(15). [PubMed: 29096462]

(28). Ponder JW; Case DA Protein Simulations; Advances in Protein Chemistry; Academic Press, 2003; Vol. 66; pp 27–85. [PubMed: 14631816]

(29). Hornak V; Abel R; Okur A; Strockbine B; Roitberg A; Simmerling C Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins 2006, 65, 712–725. [PubMed: 16981200]

(30). Cieplak P; Dupradeau F-Y; Duan Y; Wang J Polarization Effects in Molecular Mechanical Force Fields. J Phys. Condens. Matter 2009, 21, 333102–333102. [PubMed: 21828594]

(31). Cisneros GA; Karttunen M; Ren P; Sagui C Classical Electrostatics for Biomolecular Simulations. Chem. Rev. 2014, 114, 779–814. [PubMed: 23981057]

(32). Baker CM Polarizable Force Fields for Molecular Dynamics Simulations of Biomolecules. Wiley Interdiscip. Rev. Comput. Mol. Sci. 2015, 5, 241–254.

(33). Lopes PEM; Guvench O; MacKerell J, Alexander D Current Status of Protein Force Fields for Molecular Dynamics Simulations. Methods Mol. Biol. 2015, 1215, 47–71. [PubMed: 25330958]

(34). Lemkul JA; Huang J; Roux B; MacKerell J, Alexander D An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. Chem. Rev. 2016, 116, 4983–5013. [PubMed: 26815602]

(35). Nerenberg PS; Head-Gordon T New Developments in Force Fields for Biomolecular Simulations. Curr. Opin. Struct. Biol. 2018, 49, 129–138. [PubMed: 29477047]

(36). Jing Z; Liu C; Cheng SY; Qi R; Walker BD; Piquemal J-P; Ren P Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. Annu. Rev. Biophys. 2019, 48, 371–394. [PubMed: 30916997]

(37). Varma S; Rempe SB Multibody Effects in Ion Binding and Selectivity. Biophys. J. 2010, 99, 3394–3401. [PubMed: 21081088]

(38). Rogers DM; Beck TL Quasichemical and Structural Analysis of Polarizable Anion Hydration. J. Chem. Phys. 2010, 132, 014505(01)–014505(12). [PubMed: 20078170]

(39). Rossi M; Tkatchenko A; Rempe SB; Varma S Role of Methyl-Induced Polarization in Ion Binding. Proc. Natl. Acad. Sci. U.S.A. 2013, 110, 12978–12983. [PubMed: 23878238]

(40). Saunders M; Wineman-Fisher V; Jakobsson E; Varma S; Pandit SA High-Dimensional Parameter Search Method to Determine Force Field Mixing Terms in Molecular Simulations. Langmuir 2022, 38, 2840–2851, [PubMed: 35192365]

(41). Varma S; Rempe SB Coordination Numbers of Alkali Metal Ions in Aqueous Solutions. Biophys. Chem. 2006, 124, 192–199, Ion Hydration Special Issue. [PubMed: 16875774]

(42). Halgren TA; Damm W Polarizable force fields. Curr. Opin. Struct. Biol. 2001, 11, 236–242. [PubMed: 11297934]

(43). Warshel A; Kato M; Pisliakov AV Polarizable Force Fields: History, Test Cases, and Prospects. J. Chem. Theory Comput. 2007, 3, 2034–2045, [PubMed: 26636199]

(44). Ren P; Ponder JW Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation. J. Phys. Chem. B 2003, 107, 5933–5947.

(45). Grossfield A; Ren P; Ponder JW Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. J. Amer. Chem. Soc. 2003, 125, 15671–15682. [PubMed: 14664617]

(46). Ren P; Ponder JW Temperature and Pressure Dependence of the AMOEBA Water Model. J. Phys. Chem. B 2004, 108, 13427–13437.

(47). Ponder JW; Wu C; Ren P; Pande VS; Chodera JD; Schnieders MJ; Haque I; Mobley DL; Lambrecht DS; DiStasio RA Jr Current Status of the AMOEBA Polarizable Force Field. J. Phys. Chem. B 2010, 114, 2549–2564. [PubMed: 20136072]

(48). Ren P; Wu C; Ponder JW Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. J. Chem. Theory Comput. 2011, 7, 3143–3161. [PubMed: 22022236]

(49). Shi Y; Xia Z; Zhang J; Best R; Wu C; Ponder JW; Ren P Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. J. Chem. Theory Comput. 2013, 9, 4046–4063. [PubMed: 24163642]

(50). Wineman-Fisher V; Al-Hamdani Y; Addou I; Tkatchenko A; Varma S Ion-Hydroxyl Interactions: From High-Level Quantum Benchmarks to Transferable Polarizable Force Fields. J. Chem. Theory Comput. 2019, 15, 2444–2453, [PubMed: 30830778]

(51). Wineman-Fisher V; Al-Hamdani Y; Nagy PR; Tkatchenko A; Varma S Improved description of ligand polarization enhances transferability of ion–ligand interactions. J. Chem. Phys. 2020, 153, 094115. [PubMed: 32891085]

(52). Wineman-Fisher V; Delgado JM; Nagy PR; Jakobsson E; Pandit SA; Varma S Transferable interactions of Li+ and Mg2+ ions in polarizable models. J. Chem. Phys. 2020, 153, 104113. [PubMed: 32933310]

(53). Rahman S; Wineman-Fisher V; Al-Hamdani Y; Tkatchenko A; Varma S Predictive QM/MM Modeling of Modulations in Protein–Protein Binding by Lysine Methylation. J. Mol. Biol. 2021, 433, 166745. [PubMed: 33307090]

(54). Rahman S; Wineman-Fisher V; Nagy P; Al-Hamdani Y; Tkatchenko A; Varma S Methyl-induced Polarization Destabilizes Non-covalent Interactions of N-methylated Lysines. Chem. Euro. J. 2021,

(55). Fogarty JC; Chiu S-W; Kirby P; Jakobsson E; Pandit SA Automated Optimization of Water-water Interaction Parameters for a Coarse-grained Model. J. Phys. Chem. B 2014, 118, 1603–1611. [PubMed: 24460506]

(56). Glusker JP In Metalloproteins: Structural Aspects; Anfinsen C, Edsall JT, Richards FM, Eisenberg DS, Eds.; Adv. Protein Chem.; Academic Press, 1991; Vol. 42; pp 1–76.

(57). Perdew JP; Burke K; Ernzerhof M Gradient Approximation Made Simple. Phys. Rev. Lett. 1996, 77, 3865–3868. [PubMed: 10062328]

(58). Adamo C; Barone V Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 model. J. Chem. Phys. 1999, 110, 6158–6170.

(59). Tkatchenko A; Scheffler M Accurate Molecular van der Waals Interactions from Ground-state Electron Density and Free-atom Reference Data. Phys. Rev. Lett. 2009, 102, 073005. [PubMed: 19257665]

(60). Blum V; Gehrke R; Hanke F; Havu P; Havu V; Ren X; Reuter K; Scheffler M Ab initio Molecular Simulations with Numeric Atom-centered Orbitals. Comput. Phys. Commun. 2009, 180, 2175–2196.

(61). Marom N; Tkatchenko A; Rossi M; Gobre VV; Hod O; Scheffler M; Kronik L Dispersion Interactions with Density-Functional Theory: Benchmarking Semiempirical and Interatomic Pairwise Corrected Density Functionals. J. Chem. Theory Comput. 2011, 7, 3944–3951, [PubMed: 26598340]

(62). Møller C; Plesset MS Note on an Approximation Treatment for Many-electron Systems. Phys. Rev. 1934, 46, 618.

(63). Frisch MJ; Trucks GW; Schlegel HB; Scuseria GE and Robb MA; Cheeseman JR; Scalmani G; Barone V; Mennucci B; Petersson GA; Nakatsuji H; Caricato M; Li X; Hratchian HP; Izmaylov AF; Bloino J; Zheng G; Sonnenberg JL; Hada M; Ehara M; Toyota K; Fukuda R; Hasegawa J; Ishida M; Nakajima T; Honda Y; Kitao O; Nakai H; Vreven T; Montgomery JA Jr; Peralta JE; Ogliaro F; Bearpark M; Heyd JJ; Brothers E; Kudin KN; Staroverov VN; Kobayashi R; Normand J; Raghavachari K; Rendell A; Burant JC; Iyengar SS; Tomasi J; Cossi M; Rega N; Millam JM; Klene M; Knox JE; Cross JB; Bakken V; Adamo C; Jaramillo J; Gomperts R; Stratmann RE; Yazyev O; Austin AJ; Cammi R; Pomelli C; Ochterski JW; Martin RL; Morokuma K; Zakrzewski VG; Voth GA; Salvador P; Dannenberg JJ; Dapprich S; Daniels AD; Farkas N; Foresman JB; Ortiz JV; Cioslowski J; Fox DJ. Gaussian 09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

(64). Dunning TH Jr Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen. J. Chem. Phys. 1989, 90, 1007–1023.

(65). Wang Z; Ponder JW Tinker9: Next Generation of Tinker with GPU Support. 2021; https://github.com/TinkerTools/tinker9.

(66). Bussi G; Donadio D; Parrinello M Canonical Sampling Through Velocity Rescaling. J. Chem. Phys. 2007, 126, 014101. [PubMed: 17212484]

(67). Åqvist J; Wennerström P; Nervall M; Bjelic S; Brandsdal BO Molecular Dynamics Simulations of Water and Biomolecules with a Monte Carlo Constant Pressure Algorithm. Chem. Phys. Lett. 2004, 384, 288–294.

(68). Chow K-H; Ferguson DM Isothermal-isobaric Molecular Dynamics Simulations with Monte Carlo Volume Sampling. Comput. Phys. Commun. 1995, 91, 283–289.

(69). Tuckerman M; Berne BJ; Martyna GJ Reversible Multiple Time Scale Molecular Dynamics. J Chem. Phys. 1992, 97, 1990–2001.

(70). Fan S-KS; Zahara E A Hybrid Simplex Search and Particle Swarm Optimization for Unconstrained Optimization. Eur. J. Oper. Res. 2007, 181, 527–548.

(71). Wang PC; Shoup TE Parameter Sensitivity Study of the Nelder–Mead Simplex Method. Adv. Eng. Softw. 2011, 42, 529–533.

(72). Wang Z Polarizable Force Field Development, and Applications to Conformational Sampling and Free Energy Calculation. Ph.D. thesis, Washington University, 2018.

(73). Thole BT Molecular Polarizabilities Calculated with a Modified Dipole Interaction. Chem. Phys. 1981, 59, 341–350.

(74). Bosque R; Sales J Polarizabilities of Solvents from the Chemical Composition. J. Chem. Inf. Comput. Sci. 2002, 42, 1154–1163. [PubMed: 12377003]

(75). Dünweg B; Kremer K Molecular Dynamics Simulation of a Polymer Chain in Solution. J. Chem. Phys. 1993, 99, 6983–6997.

(76). Yeh I-C; Hummer G System-size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. J. Phys. Chem. B 2004, 108, 15873–15879.

(77). Dutta P; Botlani M; Varma S Water Dynamics at Protein–Protein Interfaces: Molecular Dynamics Study of Virus–Host Receptor Complexes. J. Phys. Chem. B 2014, 118, 14795–14807. [PubMed: 25420132]

(78). Williams WD; Ellard JA; Dawson LR Solvents Having High Dielectric Constants. VI. Diffusion in N-methylacetamide1, 2. J. Amer. Chem. Soc. 1957, 79, 4652–4654.

(79). De Leeuw S; Perram JW; Smith E Computer Simulation of the Static Dielectric Constant of Systems with Permanent Electric Dipoles. Annu. Rev. Phys. Chem. 1986, 37, 245–270. [PubMed: 21819241]

(80). Neumann M; Steinhauser O Computer Simulation and the Dielectric Constant of Polarizable Polar Systems. Chem. Phys. Lett. 1984, 106, 563–569.

(81). Harder E; Anisimov VM; Whitfield T; MacKerell AD; Roux B Understanding the Dielectric Properties of Liquid Amides from a Polarizable Force Field. J. Phys. Chem. B 2008, 112, 3509–3521. [PubMed: 18302362]

(82). Haynes W; Lide D; Bruno T CRC Handbook of Chemistry and Physics, 97th edn, Vol. 2016–2017. Elements in the Earth's Crust and in the Sea 2017, 14–17.

(83). Lide DR CRC Handbook of Chemistry and Physics; CRC press, 2004; Vol. 85.

(84). Kindt J; Schmuttenmaer C Far-infrared dielectric properties of polar liquids probed by femtosecond terahertz pulse spectroscopy. The Journal of Physical Chemistry 1996, 100, 10373–10379.

(85). Victor PJ; Hazra DK Excess Molar Volumes, Viscosity Deviations, and Isentropic Compressibility Changes in Binary Mixtures of N-methylacetamide+2-methoxyethanol and N-methylacetamide+water at (308.15, 313.15, and 318.15) K. J. Chem. Eng. Data 2002, 47, 79–82.

(86). MacKerell AD Jr; Shim JH; Anisimov VM Re-evaluation of the Reported Experimental Values of the Heat of Vaporization of N-methylacetamide. J Chem. Theory Comput. 2008, 4, 1307–1312. [PubMed: 20445813]

(87). Swiergiel J; Jadzyn J Static Dielectric Permittivity and Electric Conductivity of N-methylacetamide + N, N-dimethylacetamide Mixtures. J. Chem. Eng. Data 2009, 54, 2296–2300.

(88). Bennett CH Efficient Estimation of Free Energy Differences from Monte Carlo Data. J. Comput. Phys. 1976, 22, 245–268.

(89). Shi Y; Wu C; Ponder JW; Ren P Multipole Electrostatics in Hydration Free Energy Calculations. J. Comp. Chem. 2011, 32, 967–977. [PubMed: 20925089]

(90). Varma S; Rempe SB Structural Transitions in Ion Coordination Driven by Changes in Competition for Ligand Binding. J. Amer. Chem. Soc. 2008, 130, 15405–15419. [PubMed: 18954053]

(91). Marcus Y Thermodynamics of Solvation of Ions. Part 5. Gibbs Free Energy of Hydration at 298.15 K. J. Chem. Soc., Faraday Trans. 1991, 87, 2995–2999.

(92). Wolfenden R Interaction of the Peptide Bond with Solvent Water: A Vapor Phase Analysis. Biochemistry 1978, 17, 201–204. [PubMed: 618544]

(93). Pliego JR; Miguel ELM Absolute Single-Ion Solvation Free Energy Scale in Methanol Determined by the Lithium Cluster-Continuum Approach. J. Phys. Chem. B 2013, 117, 5129–5135, [PubMed: 23570440]

(94). Pollard TP; Beck TL Re-examining the Tetraphenyl-arsonium/tetraphenyl-borate (TATB) Hypothesis for Single-ion Solvation Free Energies. J. Chem. Phys. 2018, 148, 222830. [PubMed: 29907029]

(95). Mao Y; Demerdash O; Head-Gordon M; Head-Gordon T Assessing Ion–Water Interactions in the AMOEBA Force Field Using Energy Decomposition Analysis of Electronic Structure Calculations. Journal of Chemical Theory and Computation 2016, 12, 5422–5437, [PubMed: 27709939]

(96). Rackers JA; Wang Q; Liu C; Piquemal J-P; Ren P; Ponder JW An Optimized Charge Penetration Model for Use with the AMOEBA Force Field. Physical Chemistry, Chemical Physics 2017, 19, 276–291.

(97). Varma S; Botlani M; Hammond JR; Scott HL; Orgel JP; Schieber JD Effect of Intrinsic and Extrinsic Factors on the Simulated D-band Length of Type I Collagen. Proteins 2015, 83, 1800–1812. [PubMed: 26214145]

(98). Graf J; Nguyen PH; Stock G; Schwalbe H Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. J. Amer. Chem. Soc. 2007, 129, 1179–1189. [PubMed: 17263399]

(99). Sugita Y; Okamoto Y Replica-exchange Molecular Dynamics Method for Protein Folding. Chem. Phys. Lett. 1999, 314, 141–151.

(100). Karplus M Contact Electron-spin Coupling of Nuclear Magnetic Moments. J. Chem. Phys. 1959, 30, 11–15.

(101). Vijay-Kumar S; Bugg CE; Cook WJ Structure of Ubiquitin Refined at 1.8 Å Resolution. J. Mol. Biol. 1987, 194, 531–544. [PubMed: 3041007]

(102). Tjandra N; Feller SE; Pastor RW; Bax A Rotational Diffusion Anisotropy of Human Ubiquitin from ^{15}N NMR Relaxation. J. Amer. Chem. Soc. 2002, 117, 12562–12566.

(103). Prompers JJ; Brüschweiler R General Framework for Studying the Dynamics of Folded and nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. J. Amer. Chem. Soc. 2002, 124, 4522–4534. [PubMed: 11960483]

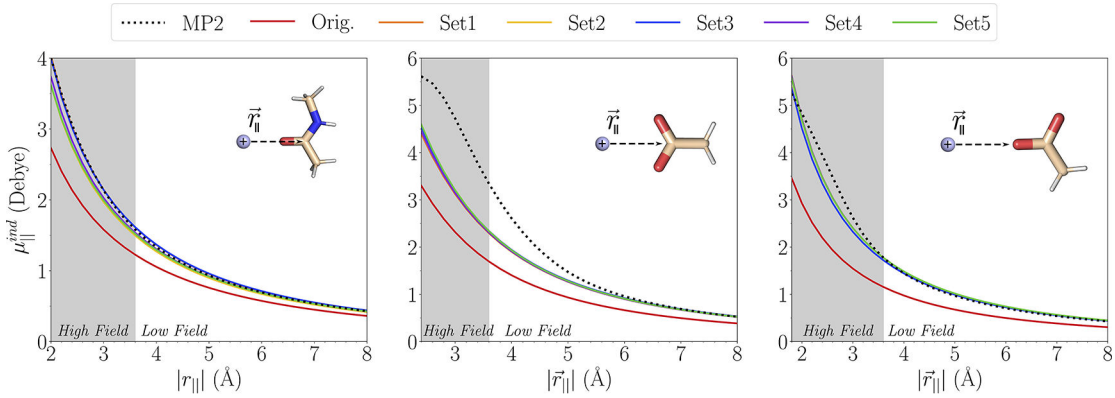

Figure 1:

Effect of recalibrating NMA's and acetate's atomic polarizabilities on their predicted induced dipole moment components $\mu_{\|}^{ind}$ parallel to the distance vector. Induced dipoles are computed for different distances from a unit positive point charge (+1) placed at different distances ($|\vec{r}|$), and compared against values determined from MP2/aug-cc-pVTZ theory.

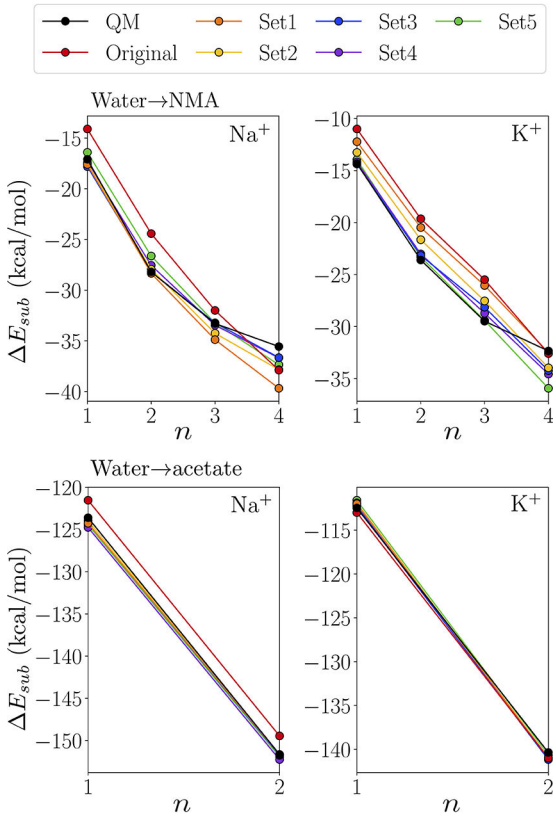

Figure 2:

Performance of recalibrated NMA and acetate models ($Sets_{1–5}$) in estimating Water→NMA and Water→acetate substitution energies in Na^+ and K^+ clusters against benchmarked QM.

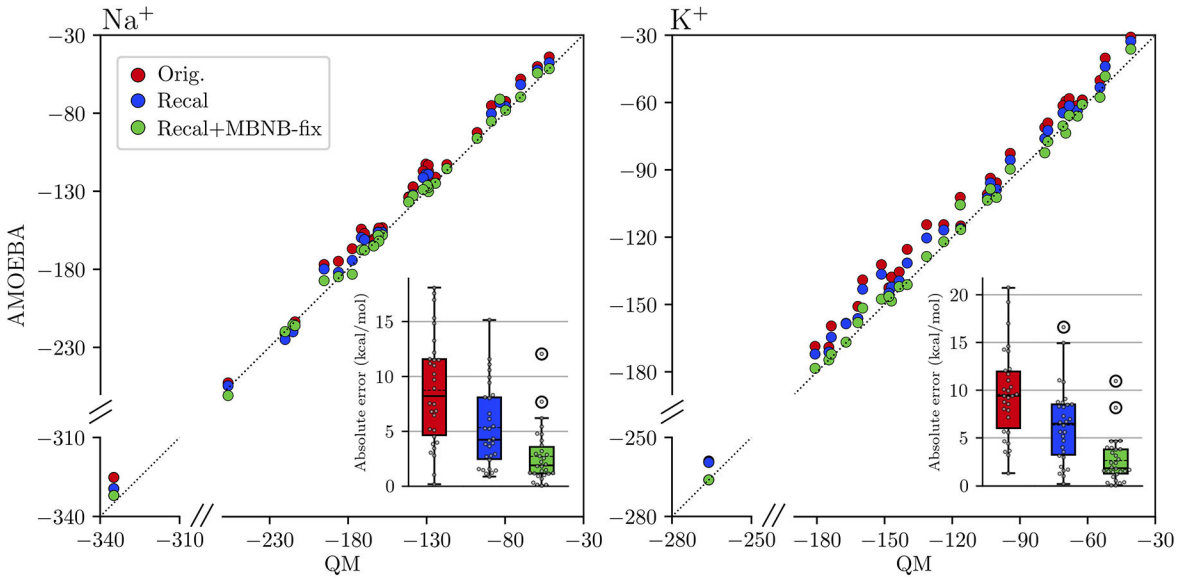

Figure 3:

Performance of recalibrated AMOEBA model in predicting ion-protein interaction energies (ΔE in kcal/mol) in thirty clusters of Na^+ and K^+ ions extracted from the PDB in an unbiased manner. QM refers to our benchmarked vdW-inclusive DFT, 'Orig.' refers to the original AMOEBA FF, 'Recal' refers to our recalibrated FF and 'Recal+MBNB-fix' refers to our recalibrated FF with ion-protein many body NB-fix corrections. The inset shows the spread in the error, as well as mean (dashed lines) and medians (solid lines).

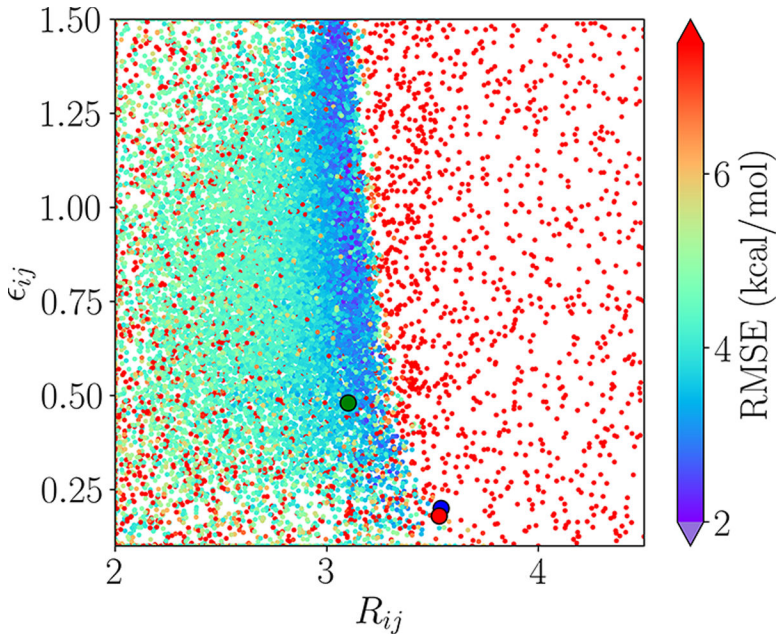

Figure 4:

Parameter space explored by 200 iterations of the Nelder-Mead algorithm for LJ interaction between K^+ and carbonyl oxygen. Cross terms determined using mixing rules are shown as dots for both the Original (red) and the recalibrated models (blue). The parameter set selected from cross term optimization is marked with a green dot.

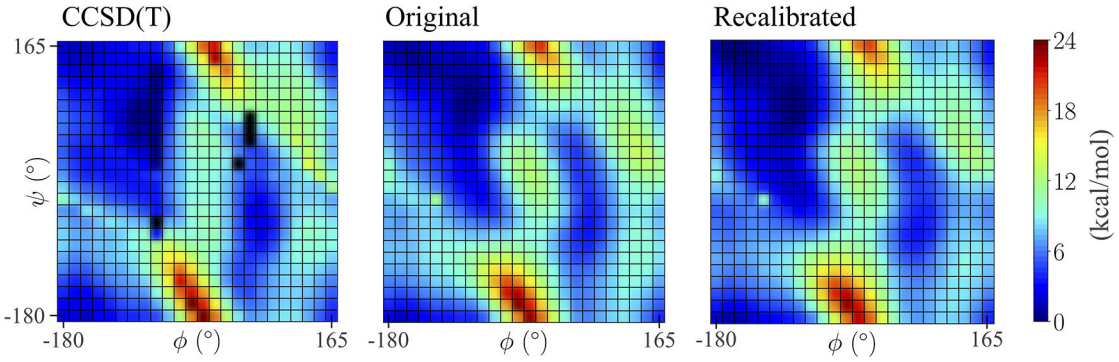

Figure 5:

Performance of recalibrated AMOEBA model in predicting alanine dipeptide's potential energy as a function of backbone dihedral angles (ψ,ϕ) in 15° increments. Predictions are compared against results from CCSD(T) calculations taken from our earlier work.[97] The black regions in the CCSD(T) plot indicate dihedral combinations for which CCSD(T) data could not be converged. Predictions from the original AMOEBA model are also included for comparison.

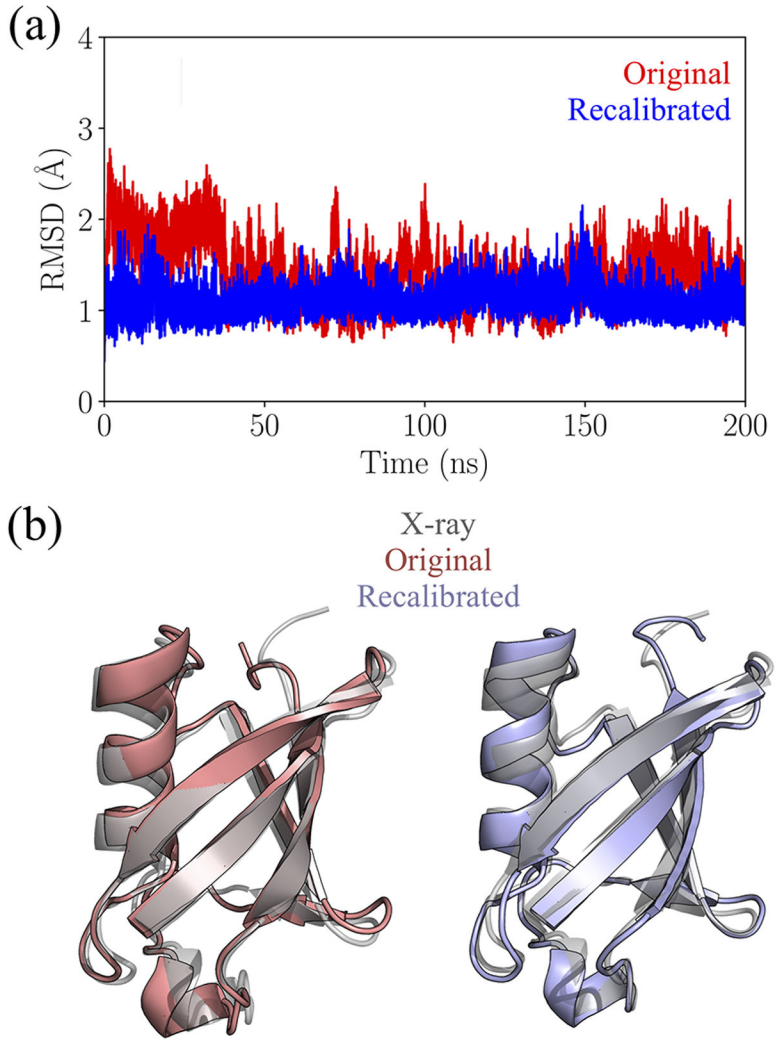

Figure 6:

Performance of recalibrated AMOEBA model in predicting protein structure. (a) Time evolution of ubiquitin's backbone RMSDs in a 200 ns long MD simulation. (b) Comparison of 200 ns snapshots of against X-ray structure.

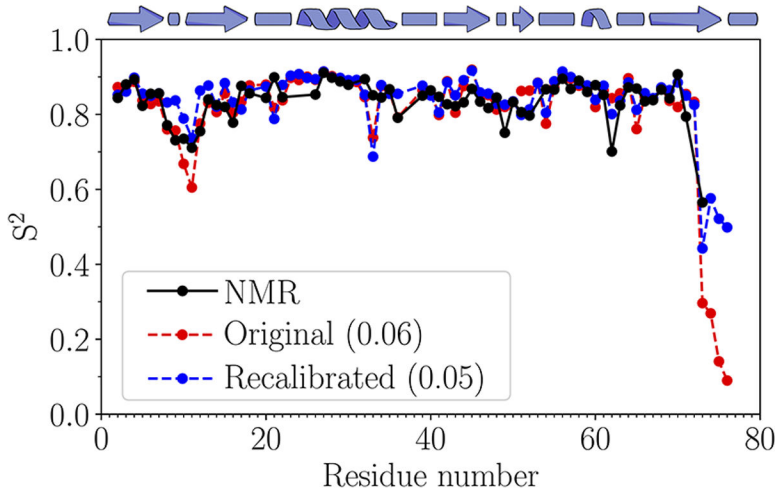

Figure 7:

Performance of the recalibrated AMOEBA model in predicting ubiquitin's backbone ^{15}N-^{1}H NMR second-order parameters (S^2).[102] Predictions from the original AMOEBA model are also included for comparison. The linear cartoon at the top indicates ubiquitin's secondary structure, where arrows and spirals represent β-sheets and and α-helices, respectively. RMSE errors for both models are shown in parentheses.

Table 1:

Atomic polarizabilities (α) and width coefficients (a) corresponding to the five different parameter sets of NMA and acetate explored in this study.

	NMA					Acetate			
	α_{CO}	α_{O}	α_{N}	α_{CH_3}	a	α_{COO}	α_{OOC}	α_{CH_3}	a
Orig.	1.33	0.84	1.07	1.33	0.39	1.33	1.20	1.33	0.39
Set_1	1.23	1.87	1.46	0.83	0.39	1.59	1.95	1.46	0.61
Set_2	1.00	1.50	1.49	1.24	0.39	1.97	1.82	1.72	0.61
Set_3	1.34	1.69	1.34	0.75	0.50	2.34	1.84	1.52	0.61
Set_4	1.13	1.50	1.05	1.20	0.50	1.84	1.97	1.33	0.61
Set_5	1.26	1.30	0.35	1.25	0.61	2.41	2.00	0.84	0.61

Table 2:

Molecular polarizability tensors of NMA and acetate (components in Å^3) determined from experiment,[74] MP2 theory, AMOEBA model (Orig.) and five different sets of recalibrated AMOEBA models ($Set_{1–5}$). RMSE (root-mean-square error) is computed with respect to MP2 theory.

NMA	Expt.	MP2	Orig.	Set_1	Set_2	Set_3	Set_4	Set_5
α_{xx}		9.25	8.84	9.18	9.33	9.24	9.29	9.25
α_{yy}		8.11	7.14	8.01	7.85	8.10	7.99	8.11
α_{zz}		6.08	5.85	6.22	6.19	6.09	6.08	6.08
α_{ave}	7.85	7.81	7.28	7.80	7.79	7.81	7.79	7.81
RMSE			0.62	0.11	0.17	0.01	0.07	0.00
Acetate								
α_{xx}		7.82	5.71	7.80	7.82	7.82	7.80	7.79
α_{yy}		7.59	5.44	7.59	7.60	7.59	7.59	7.56
α_{zz}		5.17	4.26	5.22	5.16	5.17	5.21	5.23
α_{ave}		6.86	5.13	6.87	6.86	6.86	6.87	6.87
RMSE			1.81	0.02	0.01	0.00	0.02	0.03

Table 3:

RMSE (in Å) in ion-ligand and ligand-ligand distances in recalibrated models. RMSE is estimated with respect to our benchmarked vdW-inclusive DFT.

	Orig.	Set_1	Set_2	Set_3	Set_4	Set_5
NMA-NMA	0.08	0.25	0.24	0.46	0.09	0.16
NMA-water	0.08	0.21	0.26	0.18	0.15	0.15
acetate-water	0.14	0.14	0.14	0.13	0.14	0.13
ion-NMA	0.29	0.28	0.25	0.23	0.26	0.25
ion-acetate	0.05	0.13	0.12	0.12	0.12	0.13

Table 4:

Performance of recalibrated NMA and acetate models in yielding target binding energies. Target binding energies are taken from a benchmarked vdW-inclusive DFT method. Target energies are divided into two categories. Low field target energies include NMA-NMA, NMA-water and acetate-water dimerization energies as well as binding energies of one acetate with two waters. High field target energies include binding energies of ions with 1–4 NMA and 1–2 acetate molecules. RMSE are estimated with respect to target energies. All energies are in kcal/mol.

	Target	Orig.	Set_1	Set_2	Set_3	Set_4	Set_5
Low field							
NMA-NMA dimer	−9.2	−8.6	−7.3	−8.1	−8.1	−8.8	−7.4
Water-NMA dimerl	−5.4	−5.2	−4.5	−4.4	−5.9	−5.9	−4.9
Water-NMA dimer2	−8.3	−8.2	−7.3	−7.4	−6.8	−6.9	−7.3
RMSE		0.4	1.3	1.0	1.1	0.9	1.2
Acetate-water	−19.7	−17.6	−19.0	−19.0	−19.0	−19.1	−19.0
Acetate-$(water)_2$	−36.5	−33.9	−36.5	−36.5	−36.5	36.5	−36.5
RMSE		2.4	0.5	0.5	0.5	0.4	0.5
High field							
$Na^+(NMA)_1$	−41.8	−37.3	−42.7	−42.6	−42.9	−42.4	−41.5
$Na^+(NMA)_2$	−75.3	−68.3	−75.6	−75.2	−75.4	−74.8	−73.9
$Na^+(NMA)_3$	−99.2	−93.1	−100.5	−99.9	−98.8	−99.1	−98.8
$Na^+(NMA)_4$	−117.2	−113.2	−120.3	−118.5	−117.3	−117.3	−118.0
RMSE		5.5	1.8	0.8	0.6	0.4	0.8
$K^+(NMA)_1$	−32.6	−28.3	−30.4	−31.5	−32.6	−32.4	−32.2
$k^+(nma)_2$	−58.0	−52.3	−54.8	−56.0	−57.5	−57.3	−57.6
$k^+(nma)_3$	−78.4	−71.9	−74.7	−76.1	−76.8	−77.3	−78.0
$k^+(nma)_4$	−93.8	−90.8	−93.4	−94.9	−95.2	−95.5	−96.9
RMSE		5.0	2.7	1.7	1.1	1.1	1.6
$Na^+(acetate)_1$	−148.3	−144.7	−149.3	−149.0	−149.3	−149.8	−149.5
$Na^+(acetate)_2$	−198.7	−193.3	−198.9	−198.9	−199.0	−199.5	−199.2
RMSE		4.6	0.7	0.5	0.8	1.1	0.9
$K^+(acetate)_1$	−130.7	−130.3	−130.2	−130.1	−130.5	−130.4	−129.8
$K^+(acetate)_2$	−174.8	−173.7	−175.3	−175.3	−175.5	−175.5	−175.0
RMSE		0.8	0.5	0.5	0.5	0.5	0.7

Table 5:

Performance of recalibrated models in predicting condensed phase properties of liquid NMA at 308 K. Statistical errors are estimated from block averaging.

	ρ (g/cc)	$D_{self} \times 10^5$ cm^2/s	ΔH_v (kcal/mol)	ϵ
Expt.	0.946[a]	0.411[b]	12.6 – 13.8[c]	172[d]
Orig.	0.948 ± 0.007	0.293 ± 0.013	14.7 ± 2.3	107 ± 2
Set_1	0.969 ± 0.007	0.744 ± 0.021	13.1 ± 2.3	87 ± 1
Set_2	0.970 ± 0.005	0.341 ± 0.016	15.6 ± 2.4	83 ± 1
Set_3	0.835 ± 0.010	0.184 ± 0.006	14.0 ± 2.3	118 ± 3
Set_4	0.970 ± 0.005	0.081 ± 0.005	16.8 ± 2.3	97 ± 3
Set_5	0.916 ± 0.006	0.490 ± 0.010	13.0 ± 2.2	123 ± 1

[a] taken from Ref. 85

[b] from Ref. 78

[c] from Ref. 86 and

[d] from Ref. 87.

Table 6:

Effect of recalibrating small molecule models on predicted hydration free energies. Error estimates are obtained using Monte Carlo bootstrapping.

	Δ*G(aq.)* (kcal/mol)	
	NMA.	**Acetate**
Expt.	−10.0[a]	−94.0[b]
Orig.	−8.6 ± 0.1	−83.9 ± 0.3
Set_1	−7.8 ± 0.1	−94.7 ± 0.3
Set_2	−8.6 ± 0.1	−94.6 ± 0.3
Set_3	−7.4 ± 0.1	−94.7 ± 0.3
Set_4	−8.8 ± 0.1	−94.6 ± 0.3
Set_5	−8.1 ± 0.1	−94.3 ± 0.3

[a] taken from reference 92 and

[b] is taken from reference 91.

Table 7:

Performance of recalibrated (Recal) AMOEBA model in predicting $(Ala)_5$'s Jcoupling constants measured in NMR experiments.[98] Conformational sampling is carried out using t-REMD simulations. Values for the original model (Orig) are taken from Ref. 49

Residue	J-coupling type	J_{Expt}	J_{Orig}	J_{Recal}
Ala-2	$^1J(\mathrm{N},C_\alpha)$	11.36	11.066	11.041 ± 0.021
Ala-3	$^1J(\mathrm{N},C_\alpha)$	11.26	10.923	10.873 ± 0.037
Ala-4	$^1J(\mathrm{N},C_\alpha)$	11.25	10.922	10.948 ± 0.003
Ala-2	$^2J(\mathrm{N},C_\alpha)$	9.20	8.448	8.122 ± 0.008
Ala-3	$^2J(\mathrm{N},C_\alpha)$	8.55	8.170	8.223 ± 0.027
Ala-4	$^2J(\mathrm{N},C_\alpha)$	8.40	8.232	8.015 ± 0.096
Ala-5	$^2J(\mathrm{N},C_\alpha)$	8.27	8.250	8.159 ± 0.006
Ala-2	$^3J(\mathrm{C},\mathrm{C})$	0.19	0.866	1.115 ± 0.006
Ala-2	$^3J(H_\alpha,\mathrm{C})$	1.85	1.729	1.974 ± 0.021
Ala-3	$^3J(H_\alpha,\mathrm{C})$	1.86	1.705	1.980 ± 0.008
Ala-4	$^3J(H_\alpha,\mathrm{C})$	1.89	1.713	2.014 ± 0.024
Ala-5	$^3J(H_\alpha,\mathrm{C})$	2.19	1.929	1.809 ± 0.058
Ala-2	$^3J(H_N,\mathrm{C})$	1.13	1.087	1.948 ± 0.015
Ala-4	$^3J(H_N,\mathrm{C})$	1.15	1.315	1.962 ± 0.037
Ala-5	$^3J(H_N,\mathrm{C})$	1.16	1.216	1.647 ± 0.011
Ala-2	$^3J(H_N,C_\beta)$	2.30	1.819	1.474 ± 0.003
Ala-3	$^3J(H_N,C_\beta)$	2.24	1.833	1.513 ± 0.027
Ala-4	$^3J(H_N,C_\beta)$	2.14	1.743	1.425 ± 0.012
Ala-5	$^3J(H_N,C_\beta)$	1.96	1.584	1.785 ± 0.033
Ala-2	$^3J(H_N,H_\alpha)$	5.59	6.269	7.181 ± 0.001
Ala-3	$^3J(H_N,H_\alpha)$	5.74	5.988	7.229 ± 0.046
Ala-4	$^3J(H_N,H_\alpha)$	5.98	6.079	7.309 ± 0.002
Ala-5	$^3J(H_N,H_\alpha)$	6.54	6.607	6.460 ± 0.071
Ala-2	$^3J(H_N,C_\alpha)$	0.67	0.421	0.660 ± 0.001
Ala-3	$^3J(H_N,C_\alpha)$	0.68	0.614	0.663 ± 0.008
Ala-4	$^3J(H_N,C_\alpha)$	0.69	0.648	0.648 ± 0.019
Ala-5	$^3J(H_N,C_\alpha)$	0.73	0.663	0.617 ± 0.002
RMSE			0.33	0.68