Improved description of ligand polarization enhances transferability of ion-ligand interactions

Vered Wineman-Fisher,[†] Yasmine Al-Hamdani,[‡] Péter R. Nagy,[¶] Alexandre Tkatchenko,[‡] and Sameer Varma[*,†]

†*Department of Cell Biology, Microbiology and Molecular Biology, University of South Florida, Tampa, FL 33620*

‡*Physics and Materials Science Research Unit, University of Luxembourg, 162a avenue de la Fïancerie, Luxembourg City, L-1511, Luxembourg*

¶ *Department of Physical Chemistry and Materials Science, Budapest University of Technology and Economics, H-1521 Budapest, P.O.Box 91, Hungary*

E-mail: svarma@usf.edu

Abstract

Reliability of molecular mechanics (MM) simulations in describing biomolecular ion-driven processes depends on their ability to accurately model interactions of ions simultaneously with water and other biochemical groups. In these models, ion descriptors are calibrated against reference data on ion-water interactions, and it is then assumed that these descriptors will also satisfactorily describe interactions of ions with other biochemical ligands. Comparison against experiment and high-level quantum mechanical data show that this transferability assumption can break down severely. One approach to improve transferability is to assign cross-terms or separate sets of non-bonded descriptors for every distinct pair of ion type and its coordinating ligand. Here we propose an alternative solution that targets an error-source directly and corrects misrepresented physics. In standard model development, ligand descriptors are never calibrated or benchmarked in the high electric fields present near ions. We demonstrate for a representative MM model that when the polarization descriptors of its ligands are improved to respond to both low and high fields, ligand interactions with ions also improve, and transferability errors reduce substantially. In our case, the overall transferability error reduces from 3.3 to 1.8 kcal/mol. These improvements are observed without compromising on accuracy of low-field interactions of ligands in gas and condensed phases. Reference data for calibration and performance evaluation is taken from experiment and also obtained systematically from "gold-standard" CCSD(T) in the complete basis set limit, followed by benchmarked vdW-inclusive DFT.

Introduction

Ions are vital to all biological processes.[1] They participate by either interacting directly with biomolecules and modulating their activities, or serving as charge carriers in electrical responses of cells and tissues. Mechanistic understanding of these processes requires molecular details of how ions bind and dissociate from biomolecules. Consequently, understanding of such processes requires an understanding of the differences between an ion's hydrated and

biomolecule-bound states.

Molecular mechanics (MM) simulations can potentially provide such detailed atomistic insight. This has prompted systematic improvements to force field models of ionic interactions.[2–20] However, the majority of the effort has been directed toward improving interactions of ions with water, which does not, by itself, guarantee meaningful predictions of interactions of ions with other biochemical groups. In fact, a compilation of recent studies shows that it is this transferability assumption that breaks down for many fundamental test cases.[10,13,17,19,21–24] This is not surprising for non-polarizable models that do not utilize explicit functions for describing induced effects, and rely on the assumption that mean field approximations of induced effects in water are transferable. Certainly, inclusion of explicit polarization improves performance,[13,18–20] even in water,[2–9] however, large transferability errors still remain.[13,17–19,21–24]

One approach to improve transferability in MM models is to define cross-terms or separate sets of non-bonded (NB) descriptors for every distinct pair of ion and its coordinating chemical group (ligand).[11,14–20] This "NB-fix" approach is straightforward to implement and does not sacrifice computational efficiency. However, in most applications,[11,14–19] all error corrections are assigned to the Lennard-Jones (LJ) term, although there is no supporting information of this term being the source of error.

In a recent study,[25] we analyzed a polarizable MM model,[4,26] and reported that its polarization term was a source for transferability errors. Specifically, we noted that its polarization contribution was erroneous at the kind of high electric fields present near ions, which resulted in underestimated ion-ligand binding energies. At the same time, it did perform well in low dipolar electric fields where all MM models are calibrated and benchmarked. We proposed a solution in which different polarization cross-terms could be assigned to each distinct ion-ligand pair. Although this was also a NB-fix style approach, error corrections were not assigned to the LJ term, but directly to the error-source of transferability. This approach improved transferability, however, the question of whether a ligand's polarization

model could itself be recalibrated such that it performs well at both low and high fields remained unexplored. Here we explore this general approach, and test specifically whether the functional form of the polarization model is sufficiently versatile to perform well in both low and high fields. Additionally, we examine if its recalibration improves transferability while, at the same time, retain the model's existing accuracy in describing dipolar ligand-water and ligand-ligand interactions in the gas and condensed phases.

We focus on a set of six small polar molecules, including aldehydes (formaldehyde), alcohols (methanol and ethanol) and amides (acetamide, formamide and N-methylacetamide). These are representative of key chemical groups in proteins that interact with monovalent cations,[27,28] and so getting transferability right across these ligands is important for studying ion-driven processes in proteins. We continue to use the polarizable AMOEBA model[4,29] as our representative MM model, and, as we note in the results section, this representative MM model yields moderate transferability errors for Na^+ and K^+ ions – the RMS error is 3.8 kcal/mol and the maximum error exceeds 10 kcal/mol. Similar errors in water→ethanol and water→formamide transferability have also been reported for another widely used polarizable model,[18] even after NB-fix corrections. For calibration and performance evaluation, we use experimental data and also obtain additional reference data from coupled cluster theory with single, double, and perturbative triple excitations (CCSD(T)),[30] and systematically benchmarked vdW-corrected density functional theory (DFT).

Methods

Molecular dynamics

All MD simulations are carried out using TINKER version 7.1.[26] The following control functions and parameters are chosen to be different from defaults. Integration is carried out using the RESPA algorithm with an outer time step of 1 fs.[31] Temperature is regulated using an extended ensemble approach[32] and with a coupling constant of 0.1 ps, and pressure is

regulated using a Monte Carlo approach[33,34] with a coupling constant of 0.1 ps. Electrostatic interactions are computed using particle mesh Ewald with a direct space cutoff of 9 Å. van der Waals interactions are computed explicitly for inter-atomic distances smaller than 9 Å. The convergence cutoff for induced dipoles is set at 0.01 Debye.

Reference energies

For selected ion-ligand combinations, we first compute interaction energies using complete basis set (CBS) extrapolated[35,36] and counterpoise corrected (CP)[37] CCSD(T) energies. Dunning's correlation-consistent basis sets augmented with diffuse functions (aug-cc-pVXZ, X=Q, 5) are employed for first row elements, while the corresponding weighted core-valence basis sets[38,39] are used for the alkali metal ions. Sub-valence electrons of Na^+ and K^+ are correlated in the CCSD(T) calculations, while deep-core electrons of all atoms are kept frozen. The basis set incompleteness error (BSIE) of the CBS(Q,5) interaction energies is estimated as the difference of the CP corrected and uncorrected CCSD(T) energies. The local natural orbital (LNO) scheme[40,41] is employed to accelerate the CCSD(T) calculations as implemented in the Mrcc package.[42,43] Approximation-free CCSD(T) energy and corresponding local error estimates are evaluated using the Tight and very Tight LNO-CCSD(T) threshold sets[41,44] according to the extrapolation scheme of Ref. 44. The cumulative BSIE and local error estimates indicate that the LNO-CCSD(T)/CBS(Q,5) interaction energies are within ±0.2 kcal/mol of the approximation-free CCSD(T)/CBS ones for all studied complexes.

Since CCSD(T) is significantly more expensive than DFT, we use the reference information from CCSD(T) and DMC to benchmark a vdW-corrected DFT exchange-correlation functional, namely PBE0+vdW.[45,46] The PBE0 hybrid functional contains 25% exact exchange and is supplemented by Tkatchenko-Scheffler corrections for dispersion (vdW). Exact exchange is particularly important in hydrogen bonded and charge transfer systems since it alleviates the delocalization error in DFT based approximations. All PBE0+vdW calculations are performed using the FHI-AIMS package[47] with 'really tight' basis sets. Total

energies are converged to within 10^{-6} eV and electron densities are converged to within 10^{-5} electrons. Geometry optimizations are carried out with force criterion of 10^{-3} ev/Å and the PBE0+vdW functional. The starting configurations for optimizations are taken from our previous studies,[24,48] where they were optimized using the B3LYP density functional. The ion-ligand cluster geometries used in CCSD(T) are those obtained from PBE0+vdW optimizations.

Results

We first recalibrate ligand descriptors to satisfy reference data for local interactions and then evaluate the effects of these changes on predicting their electric field responses, condensed phase properties and interactions with ions. We note that the recommended strategy to calibrate force fields is to include certain condensed phase properties as optimization targets.[49,50] Here, we are not including them as targets because we want to examine how improving local interactions affects predictions of condensed phase properties.

Recalibrating dipole polarizabilities

In the original AMOEBA model,[29] each atom is assigned an isotropic polarizability (α), and apart from atoms belonging to aromatic groups, their values are similar to those proposed by Thole[51] (Table S1 of the supporting information). However, as also noted by the authors of the original model,[29] these α produce molecular polarizabilities that are generally smaller than reference values obtained from experiment (Table 1). Perhaps that is why the induced dipole moments computed using the original model are underestimated (see next subsection).

Atomic polarizabilities can be recalibrated against experimental values, but experimental tensor components are not available for all molecules. We obtain these from Møller-Plesset second order perturbation (MP2) theory[52] implemented in Gaussian09.[53] These values are provided in Table 1. We use Dunning's correlation-consistent basis sets augmented with

diffuse functions, and note that differences between values computed using aug-cc-pVTZ and aug-cc-pVQZ basis sets are marginal. We note that the computed molecular polarizabilities quantitatively agree with experiment, expect those of Formaldehyde and Acetamide that are overestimated by a little over 5%. To maintain a consistent parameterization protocol, we chose to recalibrate polarizabilities against MP2 values.

Table 1: Comparison of original (Orig) and recalibrated (Pol) molecular polarizabilities (in Å^3) against reference values taken from experiment and computed using MP2 theory. [a]taken from Ref. 54; [b]taken from Ref. 29; [c]taken from Ref. 55.

Ligand	**Method**	α_{avg}	α_{xx}	α_{yy}	α_{zz}
Formaldehyde	Expt.[a]	2.45	2.76	2.76	1.83
	MP2	2.64	3.31	2.67	1.94
	Orig[b]	2.45	2.78	2.56	2.01
	Pol	2.66	3.14	2.71	2.14
Formamide	Expt.[a]	4.08 (4.22[c])	5.24	$\alpha_{yy} + \alpha_{zz} = 7.01$	
	MP2	4.22	5.58	4.09	3.00
	Orig[b]	3.65	4.32	3.87	2.74
	Pol	4.29	5.16	4.43	3.27
Acetamide	Expt.[a]	5.67	6.70	$\alpha_{yy} + \alpha_{zz} = 10.3$	
	MP2	6.06	7.09	6.45	4.62
	Orig[b]	5.43	6.27	5.71	4.30
	Pol	6.12	7.04	6.50	4.81
NMA	Expt.[c]	7.85			
	MP2	7.81	9.25	8.11	6.08
	Orig[b]	7.28	8.84	7.14	5.85
	Pol	7.74	9.41	7.69	6.12
Methanol	Expt.[a]	3.32 (3.26[c])	4.09	3.23	2.65
	MP2	3.22	3.52	3.09	3.05
	Orig[b]	3.20	3.62	3.03	2.93
	Pol	3.21	3.58	3.08	2.96
Ethanol	Expt.[a]	5.26 (5.13[c])	6.39	4.82	4.55
	MP2	5.07	5.52	4.98	4.72
	Orig[b]	4.95	5.38	4.94	4.53
	Pol	5.08	5.52	5.01	4.71

We use enumeration to optimize α, and also optimize atomic α of each molecular chemistry, that is, aldehyde, alcohol and amide, separately. The latter is to implicitly incorporate bonding chemistries into α. The new set of atomic α are provided in Table S1 of the supporting information, and the new set of molecular polarizabilties are listed in Table 1.

Optimization does improve polarizabilties of all molecules, however, that of formamide still remains somewhat lower compared to reference values. In fact, we find no combinations of atomic α that reproduce reference values for formamide (Fig. S1 of supporting information). At the same time, we do note that Thole damping coefficients in the polarizable model can be modified to potentially further improve correspondence with reference data.

Electric field response

To evaluate the field response of a molecule, we determine its induced dipole in the presence of a unit point charge (+1) placed at incrementally increasing distances ($|\mathbf{r}|$) from its coordinating oxygen. Fig. 1a shows the results of these calculations for two representative molecules, NMA and ethanol, for which polarizability recalibration produces one of the largest and smallest improvements, respectively. In the original model, we note that the induced dipoles of all molecules are consistently underestimated at short distances from the point charge, but the absolute error decreases with increasing distance from the point charge. In other words, while the original model performs well at low electric fields ($|\mathbf{r}| > 3.5$ Å), its error increases at stronger fields that are present in an ion's first coordination shell ($|\mathbf{r}| < 3.5$ Å). Recalibration of polarizability improves their field responses in both the low and field regions, although the performance gain is observed to be much greater in the high field region, as shown in Fig. 1b. Indeed, errors do remain at very high fields, except for formalydehyde. This may perhaps be due to limitations in the Thole polarization model, and variations to the model used in the original force field have, in fact, shown improvements in describing many-body interactions.[56]

Condensed phase properties

Since we modify atomic α, the Lennard-Jones (LJ) parameters of small molecules also need to be recalibrated. Following protocol of the original model, we recalibrate them to reproduce homo- and hetero-dimer binding energies and geometries. Here, instead of using

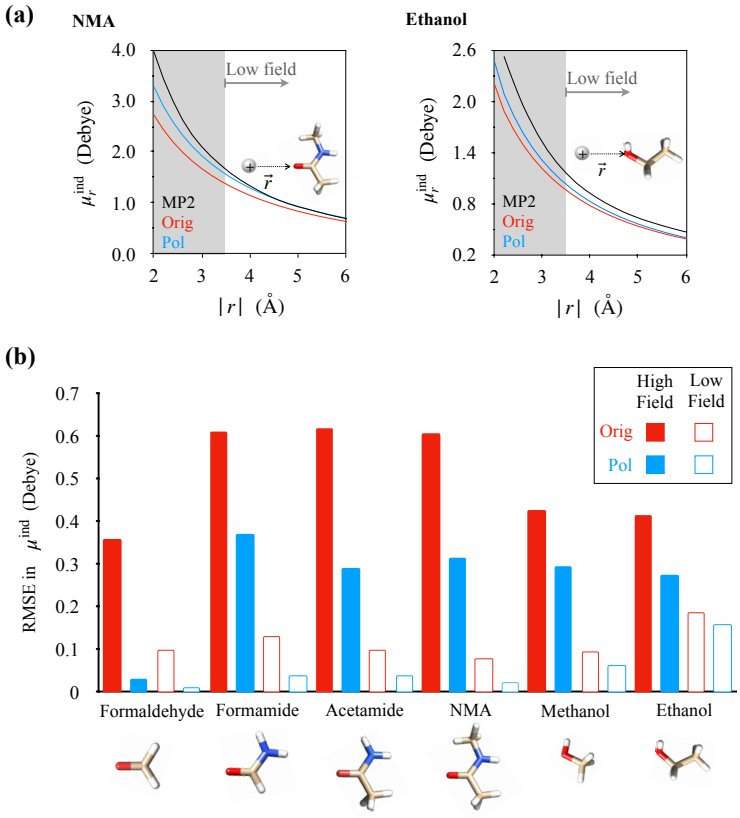

Figure 1: Effect of recalibrating ligand polarizabilities on their predicted induced dipole moments μ^{ind}. (a) Induced dipoles of NMA and ethanol are estimated for different distances ($|\mathbf{r}|$) from a positive point charge, and compared against corresponding values from MP2/aug-cc-pVTZ theory. The subscript r in μ_r^{ind} is the component of the induced dipole along the vector $\mathbf{r}$ which is parallel to an interaction axis. For Formamide, Acetamide, and NMA, μ^{ind} is estimated along multiple axes, and their full sets of calculations are provided in Figure S2 of the Supporting Information. (b) Root mean square errors (RMSE) are determined with respect to MP2 values, but separately for molecule-charge distances less than and greater than 3.5 Å, which we refer to as high and low field regions, respectively.

MP2 theory to obtain reference data, we use the vdW-corrected PBE0 density functional (PBE0+vdW)[45,46] that has been demonstrated to produce a wide range of intermolecular interactions in molecular dimers with an accuracy of 0.3 kcal/mol against S22 and S66 datasets.[57] Nevertheless, we further benchmark its performance by comparing reference data for formamide dimers against CCSD(T) (Figure S3 of Supporting Information). All of the dimer reference data used for LJ recalibration is provided in Figure S3 of the Supporting Information.

To evaluate the performance of the recalibrated model, we determine four condensed phase properties:[49,50] density (ρ), heat of vaporization (ΔH_v), lattice energy (ΔE_l) and self diffusion constant (D_{self}).

We compute densities and heats of vaporization from the final 1 ns of 5 ns long MD trajectories of $N = 512$ solvent molecules contained in cubic boxes and simulated under isothermal (T=298 K) and isobaric (P=1 bar) conditions. Statistical errors are obtained from block averaging using progressively smaller time windows of 0.8, 0.6, 0.4 and 0.2 ns. The heat of vaporization is computed as

$$\Delta H_v = (\langle U_{\text{gas}} \rangle - \langle U_{\text{liquid}} \rangle)/N + RT \tag{1}$$

where U_{gas} and U_{liquid} are the total potential energies of molecules in the gas and liquid phases, and R is the gas constant. U_{liquid} are computed from the same trajectory data from which densities are calculated above. U_{gas} are computed from separate MD trajectories of isolated molecules under isochoric and isothermal conditions with a ligand number density of 0.024 nm^{-3}. Lattice energies are determined as $\Delta E_l = U_l/n$, where U_l is the potential energy a single unit cell under periodic conditions computed after energy minimization, and n is the number molecules in the unit cell. Coordinates of the formamide unit cell are taken from Ref. 58 and those of the remaining molecules are taken from the Crystallography Open Database.[59]

Finally, self diffusion constants are computed using Einstein's equation, and also corrected for periodic cell size using the relationship,

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

obtained from the thermodynamic theory of diffusion.[60–62] In the expression above, $r(\Delta t)$ is the center of mass displacement, L is the unit length of the cubic box, η is the viscosity, and $\alpha = 2.837$ is a constant. Data for computing the first term is taken from separate 5.5 ns long MD simulations conducted under NVT conditions and at volumes fixed at their average values found in NPT simulations. For the average value of D_{self}, statistics are obtained from the final 5 ns of each trajectory, and the slope $\langle r^2(\Delta t) \rangle / \Delta t$ is determined from $\Delta t = 0.5$ to $\Delta t = 4.5$ ns. Statistical error is computed by block averaging where progressively smaller amounts of simulation data are taken and slopes are re-computed with correspondingly smaller Δt windows.[62]

Table 2 shows the results of these calculations, and also compares them to reference data. We note that the values predicted from the recalibrated model are very similar to those obtained using the original model, with the exception of perhaps ethanol where the percentage change is higher. Note that in the calibration of the original model, condensed phase properties, like density, were included as part of the optimization target, which we did not include in our recalibration. Since recalibration improves ligand induced dipoles even at low dipolar fields, it also improves the relative balance between contributions from polarization and LJ forces.

Transferability of ionic interactions

We evaluate transferability of ionic interactions by determining substitution energies

$$\Delta E = E_{\text{AX}_n} - nE_{\text{X}} - E_{\text{AW}_n} + nE_{\text{W}} \tag{3}$$

Table 2: Effect of recalibrating ligands on predictions of their condensed phase properties. Note that the statistical errors for ρ are not listed, but for all systems, they are smaller than 0.005 g/cc. [a]taken from Ref. 63; [b]taken from Ref. 64; [c]taken from Ref. 58; [d]taken from Ref. 65;[e]taken from Ref. 66; [f]taken from Ref. 67; [g]taken from Ref. 29; [h]taken from Ref. 68; [i]taken from Ref. 69; [j]taken from Ref. 70;[k]taken from Ref. 71; [l]taken from Ref. 72; [m]taken from Ref. 73; [n]taken from Ref. 74.

Ligand	Method	ρ (g/cc)	ΔH_v (kcal/mol)	ΔE_l (kcal/mol)	D_{self} (10^{-5} cm^2/s)
Formamide	Expt.	1.13[a]	14.3[b]	-18.9[c]	0.55[d]
	CCSD(T)	-	-	-21.5[e]	
	PBE0+vdW	-	-	-20.1[f]	
	Orig[g]	1.12	14.1 ± 0.4	-18.2	0.53 ± 0.03
	Pol	1.10	13.8 ± 0.3	-17.6	0.67 ± 0.03
NMA	Expt.	0.95[a]	13.3–14.3[h,i]		0.41[j]
	Orig	0.95	14.2 ± 0.2		0.34 ± 0.01
	Pol	0.92	13.7 ± 0.2		0.35 ± 0.02
Methanol	Expt.	0.78[a]	9.0[i]	-11.8[k]	2.41[d]
	CCSD(T)	-	-	-12.9[d]	
	Orig	0.77	9.1 ± 0.3	-12.9	2.12 ± 0.02
	Pol	0.73	9.6 ± 0.2	-13.5	2.93 ± 0.09
Ethanol	Expt.	0.79[l]	10.1[m]	-12.5[n]	1.07[d]
	CCSD(T)	-	-	-9.3[e]	
	Orig	0.77	10.4 ± 0.30	-14.0	0.91 ± 0.04
	Pol	0.82	12.2 ± 0.33	-16.2	0.53 ± 0.01

for the reactions below

$$\mathrm{AW}_n + n\mathrm{X} \rightleftharpoons \mathrm{AX}_n + n\mathrm{W}, \tag{4}$$

where A refers to either a Na^+ or K^+ ion, W refers to water and X refers to a small molecule other than water.

Reference energies needed for evaluating performance are obtained from PBE0+vdW.[45,46] Table 3 shows that predictions from PBE0+vdW for six different types of ion-ligand clusters agree with higher-level quantum methods, including Monte Carlo (QMC)[25] and LNO-CCSD(T). Additionally, the LNO scheme[40,41] employed to accelerate the CCSD(T) does not compromise accuracy in relation to values obtained using QMC,[25] and also previous estimates of ion-water interaction energies obtained without the LNO scheme.[75] We also note agreement of PBE0+vdW with QMC and LNO-CCSD(T) in terms of both interaction energies per ligand, and also the trend with respect to cluster-size. We had also noted in our earlier study[24] that under a harmonic approximation, PBE0+vdW also predicts gas phase ion-water cluster enthalpies and free energies consistent with experiment.

Table 3: Cluster binding energies (in kcal/mol), normalized by the number of ligands in clusters, from first principles methods: QMC, LNO-CCSD(T) and PBE0+TS. [a]Taken from our earlier work.[25]

	H_2O			*CH_3OH*			*NH_2CHO*	
Na^+/#ligands	QMC[a]	LNO-CCSD(T)	PBE0+vdW[a]	QMC[a]	LNO-CCSD(T)	PBE0+vdW[a]	LNO-CCSD(T)	PBE0+vdW
1	-24.5 ± 0.2	-24.4	-24.7	-26.5 ± 0.3	-26.1	-26.3	-36.9	-37.2
2	-23.0 ± 0.3	-23.1	-23.5	-24.4 ± 0.6	-24.5	-25.0	-33.3	-34.1
3	-21.6 ± 0.4	-21.6	-22.0	-22.9 ± 0.6	-22.7	-23.4	-29.4	-30.2
4	–	-20.0	-20.4	-21.1 ± 1.2	-20.9	-21.7	-26.0	-26.7
K^+/#ligands	QMC[a]	LNO-CCSD(T)	PBE0+vdW[a]	QMC[a]	LNO-CCSD(T)	PBE0+vdW[a]	LNO-CCSD(T)	PBE0+vdW
1	-17.9 ± 0.3	-18.2	-18.2	-19.0 ± 0.3	-19.4	-19.3	-28.3	-28.6
2	-17.1 ± 0.3	-17.1	-17.2	-18.1 ± 0.4	-18.2	-18.1	-25.8	-26.0
3	-15.9 ± 0.5	-16.2	-16.3	-16.8 ± 0.5	-17.1	-17.2	-23.3	-23.5
4	-15.3 ± 0.4	-15.3	-15.4	-16.6 ± 0.6	-15.9	-16.3	-20.5	-21.1

Fig. 2 shows the effect of ligand parameter recalibration on substitution energies. In calculations using the original model, we employ the original LJ descriptors of Na^+ or K^+ ions,[4] and in the recalibrated model, we use our new ion LJ descriptors.[25] The original vdW descriptors of Na^+ and K^+ were (ϵ = 0.26 kcal/mol, r_0 = 3.02 Å) and (0.35, 3.71), respectively, and our new descriptors are (0.48, 2.50) for Na^+ and (0.59, 3.51) for K^+.

Overall, we find that the RMSE with respect to reference data reduces from 3.3 kcal/mol to 1.8 kcal/mol, and the maximum error drops from 9.8 kcal/mol to 6.3 kcal/mol. The extent of improvement in water→alcohol substitution energies is similar to what we have noted previously when we had employed a NB-fix style approach to modify the polarization term.[25] Note that the improvement in transferability is not due to recalibration of ion LJ parameters, as we demonstrated previously.[25]

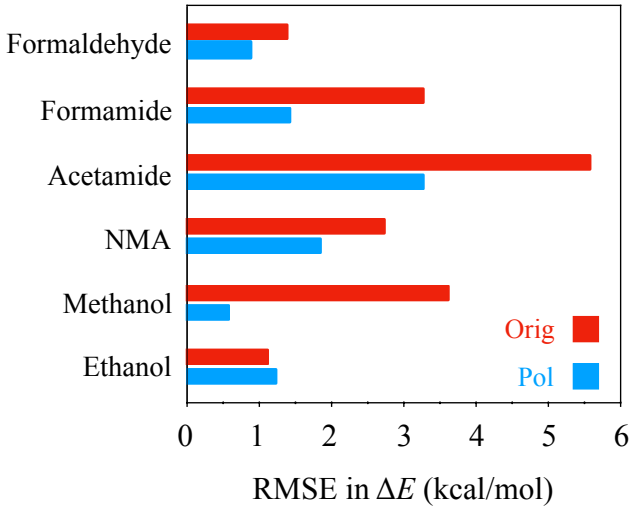

Figure 2: Effect of recalibrating ligands on ion-ligand substitution energies, ΔE. RMSE is obtained with respect to PBE0+vdW values, and all of the data used for computing RMSE's is shown in Figure S4 of the Supporting Information.

Finally, we examine how ligand recalibration affects their structures around ions in the condensed phase. To examine this, we simulate both ions in all four solvents under NPT conditions (P=1 atm. and T=298 K) for 5 ns, and use the final nanosecond of each trajectory to compute the radial distribution functions (RDFs) of solvent oxygens around ions. We find that ligand recalibration has little effect on RDFs (Fig. S5 of the Supporting Information), suggesting that their parameter-sensitivity is less compared to substitution energy. Note that, as expected,[76] the coordination structures of solvents around ions vary with solvent chemistry.

Overall, we find that for all, but one, small molecules recalibration of their polarizabilities substantially improves their interactions with ions, and with minimal affects on their

condensed phase properties. The only exception is ethanol whose recalibration leads to slightly larger errors in its predicted condensed phase properties (Table 2), with almost no effect on its interactions with ions (Fig. 2). In fact, recalibration of its polarizability also had little effect on its predicted field response (Fig. 1a). Therefore, we consider the original ethanol parameters, which were tuned numerically to simultaneously reproduce a subset of condensed phase properties, to be superior to those obtained here that were not calibrated specifically to reproduce condensed phase properties.

Conclusions

In standard MM model development, ligand descriptors are calibrated against low electric field reference data, which does not guarantee performance at the much higher electric fields present near ions. In fact, even in our representative model, ligands perform well at low fields, but errors get progressively larger with increasing field strengths. Here we demonstrate that when the polarization descriptors of ligands are calibrated and benchmarked to satisfy reference data at not only low, but also high fields, their interactions also improve with ions. Performance gain at high fields does not have to be the expense of accuracy at low fields, as long as the underlying functional form is sufficiently flexible. Therefore, as an alternative to patching ion-ligand interactions in *a posteriori* manner,[11,14–20,25] this work recommends future development of MM models to also consider ligand calibration and performance evaluation at high fields. This would make MM models intrinsically more compatible with modeling ionic interactions.

Acknowledgements

Authors acknowledge the use of computer time from Research Computing at USF and UL. VW, YA, AT and SV ackwoledge funding from NIH grant number R01GM118697. PRN is grateful for financial support of NKFIH, Grant No. KKP126451, ÚNKP-19-4-BME-418 New

National Excellence Program of the Ministry for Innovation and Technology and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Supplementary Material

Contains one table and five figures.

Availability of Data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

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

(2) Lybrand, T. P.; Kollman, P. A. Water–water and water–ion potential functions including terms for many body effects. *The Journal of Chemical Physics* **1985**, *83*, 2923–2933.

(3) Perera, L.; Berkowitz, M. L. Many body effects in molecular dynamics simulations of $Na^+(H_2O)_n$ and $Cl^-(H_2O)_n$ clusters. *The Journal of Chemical Physics* **1991**, *95*, 1954–1963.

(4) Grossfield, A.; Ren, P.; Ponder, J. W. Ion Solvation Thermodynamics from Simulation with a Polarizable Force Field. *Journal of the American Chemical Society* **2003**, *125*, 15671–15682.

(5) Carrillo-Tripp, M.; Saint-Martin, H.; Ortega-Blake, I. A comparative study of the hydration of Na+ and K+ with refined polarizable model potentials. *The Journal of Chemical Physics* **2003**, *118*, 7062–7073.

(6) Spångberg, D.; Hermansson, K. Many-body potentials for aqueous Li+, Na+, Mg2+, and Al3+: Comparison of effective three-body potentials and polarizable models. *The Journal of Chemical Physics* **2004**, *120*, 4829–4843.

(7) Lamoureux, G.; Roux, B. Absolute Hydration Free Energy Scale for Alkali and Halide Ions Established from Simulations with a Polarizable Force Field. *The Journal of Physical Chemistry B* **2006**, *110*, 3308–3322.

(8) Whitfield, T. W.; Varma, S.; Harder, E.; Lamoureux, G.; Rempe, S. B.; Roux, B. Theoretical Study of Aqueous Solvation of K+ Comparing ab Initio, Polarizable, and Fixed-Charge Models. *Journal of Chemical Theory and Computation* **2007**, *3*, 2068–2082.

(9) 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. *The Journal of Chemical Physics* **2007**, *127*, 064509(01)–064509(19).

(10) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *The Journal of Physical Chemistry B* **2008**, *112*, 9020–9041.

(11) Baker, C. M.; Lopes, P. E. M.; Zhu, X.; Roux, B.; MacKerell, A. D. Accurate Calculation of Hydration Free Energies using Pair-Specific Lennard-Jones Parameters in the CHARMM Drude Polarizable Force Field. *Journal of Chemical Theory and Computation* **2010**, *6*, 1181–1198.

(12) Luo, Y.; Roux, B. Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *The Journal of Physical Chemistry Letters* **2010**, *1*, 183–189.

(13) Varma, S.; Rogers, D. M.; Pratt, L. R.; Rempe, S. B. Design principles for K+ selectivity in membrane transport. *The Journal of General Physiology* **2011**, *138*, 279–279.

(14) Yoo, J.; Aksimentiev, A. Improved Parametrization of Li+, Na+, K+, and Mg2+ Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *The Journal of Physical Chemistry Letters* **2012**, *3*, 45–50.

(15) Fyta, M.; Netz, R. R. Ionic force field optimization based on single-ion and ion-pair solvation properties: Going beyond standard mixing rules. *The Journal of Chemical Physics* **2012**, *136*, 124103(01)–124103(11).

(16) Mamatkulov, S.; Fyta, M.; Netz, R. R. Force fields for divalent cations based on single-ion and ion-pair properties. *The Journal of Chemical Physics* **2013**, *138*, 024505(1)–024505(12).

(17) Savelyev, A.; MacKerell, A. D. Balancing the Interactions of Ions, Water, and DNA in the Drude Polarizable Force Field. *The Journal of Physical Chemistry B* **2014**, *118*, 6742–6757.

(18) Li, H.; Ngo, V.; Da Silva, M. C.; Salahub, D. R.; Callahan, K.; Roux, B.; Noskov, S. Y. Representation of Ion–Protein Interactions Using the Drude Polarizable Force-Field. *The Journal of Physical Chemistry B* **2015**, *119*, 9401–9416.

(19) Savelyev, A.; MacKerell, A. D. Competition among Li+, Na+, K+, and Rb+ Monovalent Ions for DNA in Molecular Dynamics Simulations Using the Additive CHARMM36 and Drude Polarizable Force Fields. *The Journal of Physical Chemistry B* **2015**, *119*, 4428–4440.

(20) 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. *The Journal of Chemical Physics* **2017**, *147*, 161733(01)–161733(15).

(21) Warren, G. L.; Patel, S. Comparison of the Solvation Structure of Polarizable and Nonpolarizable Ions in Bulk Water and Near the Aqueous Liquid, Vapor Interface. *The Journal of Physical Chemistry C* **2008**, *112*, 7455–7467.

(22) Varma, S.; Rempe, S. B. Multibody Effects in Ion Binding and Selectivity. *Biophysical Journal* **2010**, *99*, 3394–3401.

(23) Rogers, D. M.; Beck, T. L. Quasichemical and structural analysis of polarizable anion hydration. *The Journal of Chemical Physics* **2010**, *132*, 014505(01)–014505(12).

(24) Rossi, M.; Tkatchenko, A.; Rempe, S. B.; Varma, S. Role of methyl-induced polarization in ion binding. *Proceedings of the National Academy of Sciences of the USA* **2013**, *110*, 12978–12983.

(25) 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. *Journal of Chemical Theory and Computation* **2019**, *15*, 2444–2453, PMID: 30830778.

(26) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *Journal of Chemical Theory and Computation* **2013**, *9*, 4046–4063.

(27) Glusker, J. P. In *Metalloproteins: Structural Aspects*; Anfinsen, C., Edsall, J. T., Richards, F. M., Eisenberg, D. S., Eds.; Advances in Protein Chemistry; Academic Press, 1991; Vol. 42; pp 1 – 76.

(28) Page, M. J.; Di Cera, E. Role of Na+ and K+ in Enzyme Function. *Physiological Reviews* **2006**, *86*, 1049–1092.

(29) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. *Journal of Chemical Theory and Computation* **2011**, *7*, 3143–3161.

(30) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order per-

turbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479.

(31) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. *The Journal of Chemical Physics* **1992**, *97*, 1990–2001.

(32) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **2007**, *126*, 014101(1)–014101(8).

(33) Chow, K.-H.; Ferguson, D. M. Isothermal-isobaric molecular dynamics simulations with Monte Carlo volume sampling. *Computer physics communications* **1995**, *91*, 283–289.

(34) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical physics letters* **2004**, *384*, 288–294.

(35) Karton, A.; Martin, J. M. L. Comment on: "Estimating the Hartree–Fock limit from finite basis set calculations". *Theor. Chem. Acc.* **2006**, *115*, 330.

(36) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639.

(37) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553.

(38) Prascher, B. P.; Woon, D. E.; Peterson, K. A.; Dunning, T. H.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. VII. Valence, core-valence, and scalar relativistic basis sets for Li, Be, Na, and Mg. *Theor. Chem. Acc.* **2011**, *128*, 69.

(39) Hill, J. G.; Peterson, K. A. Gaussian basis sets for use in correlated molecular calculations. XI. Pseudopotential-based and all-electron relativistic basis sets for alkali metal (K-Fr) and alkaline earth (Ca-Ra) elements. *J. Chem. Phys.* **2017**, *147*, 244106.

(40) Nagy, P. R.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Redundancy-free triples correction using Laplace transform. *J. Chem. Phys.* **2017**, *146*, 214106.

(41) Nagy, P. R.; Samu, G.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications. *J. Chem. Theory Comput.* **2018**, *14*, 4193.

(42) Kállay, M. et al. The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **2020**, *152*, submitted.

(43) MRCC, a quantum chemical program suite written by M. Kállay, P. R. Nagy, Z. Rolik, D. Mester, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, I. Ladjánszki, L. Szegedy, B. Ladóczki, K. Petrov, M. Farkas, P. D. Mezei, and B. Hégely. See http://www.mrcc.hu/ (Accessed October 1, 2019).

(44) Nagy, P. R.; Kállay, M. Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods. *J. Chem. Theory Comput.* **2019**, *15*, 5275.

(45) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of chemical physics* **1999**, *110*, 6158–6170.

(46) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Physical Review Letters* **2009**, *102*, 073005(1)–073005(4).

(47) 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. *Computer Physics Communications* **2009**, *180*, 2175–2196.

(48) Varma, S.; Rempe, S. B. Structural transitions in ion coordination driven by changes in competition for ligand binding. *Journal of the American Chemical Society* **2008**, *130*, 15405–15419.

(49) Salas, F. J.; Mendez-Maldonado, G. A.; Nunez-Rojas, E.; Aguilar-Pineda, G. E.; Dominguez, H.; Alejandre, J. Systematic Procedure To Parametrize Force Fields for Molecular Fluids. *Journal of Chemical Theory and Computation* **2015**, *11*, 683–693, PMID: 26579602.

(50) Perez de la Luz, A.; Aguilar-Pineda, J. A.; Mendez-Bermúdez, J. G.; Alejandre, J. Force Field Parametrization from the Hirshfeld Molecular Electronic Density. *Journal of Chemical Theory and Computation* **2018**, *14*, 5949–5958, PMID: 30278120.

(51) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics* **1981**, *59*, 341–350.

(52) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **1934**, *46*, 618–622.

(53) Frisch, M.; et. Al, Gaussian09 Revision A.1. **2009**,

(54) Applequist, J.; Carl, J. R.; Fung, K.-K. Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities. *Journal of the American Chemical Society* **1972**, *94*, 2952–2960.

(55) Bosque, R.; Sales, J. Polarizabilities of solvents from the chemical composition. *Journal of chemical information and computer sciences* **2002**, *42*, 1154–1163.

(56) Liu, C.; Qi, R.; Wang, Q.; Piquemal, J.-P.; Ren, P. Capturing Many-Body Interactions with Classical Dipole Induction Models. *Journal of Chemical Theory and Computation* **2017**, *13*, 2751–2761, PMID: 28482664.

(57) Marom, N.; Tkatchenko, A.; Rossi, M.; Gobre, V. V.; Hod, O.; Scheffler, M.; Kronik, L. Dispersion interactions with density-functional theory: Benchmarking semiempirical and interatomic pairwise corrected density functionals. *Journal of Chemical Theory and Computation* **2011**, *7*, 3944–3951.

(58) Otero-De-La-Roza, A.; Johnson, E. R. A benchmark for non-covalent interactions in solids. *The Journal of chemical physics* **2012**, *137*, 054103.

(59) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database - an open-access collection of crystal structures. *Journal of applied crystallography* **2009**, *42*, 726–729.

(60) Dunweg, B.; Kremer, K. Molecular dynamics simulation of a polymer chain in solution. *The Journal of Chemical Physics* **1993**, *99*, 6983–6997.

(61) Yeh, I.-C.; Hummer, G. System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *The Journal of Physical Chemistry B* **2004**, *108*, 15873–15879.

(62) Dutta, P.; Botlani, M.; Varma, S. Water Dynamics at Protein–Protein Interfaces: Molecular Dynamics Study of Virus–Host Receptor Complexes. *The Journal of Physical Chemistry B* **2014**, *118*, 14795–14807, PMID: 25420132.

(63) Pacak, P. Refractivity and density of some organic solvents. *Chem. Papers* **1991**, *45*, 227–232.

(64) Barone, G.; Castronuovo, G.; Della Gatta, G.; Elia, V.; Iannone, A. Enthalpies of vaporization of seven alkylamides. *Fluid phase equilibria* **1985**, *21*, 157–164.

(65) Holz, M.; Mao, X.-a.; Seiferling, D.; Sacco, A. Experimental study of dynamic isotope

effects in molecular liquids: Detection of translation-rotation coupling. *The Journal of chemical physics* **1996**, *104*, 669–679.

(66) Červinka, C.; Fulem, M.; Růžička, K. CCSD(T)/CBS fragment-based calculations of lattice energy of molecular crystals. *The Journal of Chemical Physics* **2016**, *144*, 064505.

(67) Reilly, A. M.; Tkatchenko, A. Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *The Journal of Chemical Physics* **2013**, *139*, 024705.

(68) Kreis, R.; Wood, R. Enthalpy of fusion and cryoscopic constant of N-methylacetamide. *The Journal of Chemical Thermodynamics* **1969**, *1*, 523–526.

(69) Chickos, J. S.; Acree Jr, W. E. Enthalpies of vaporization of organic and organometallic compounds, 1880–2002. *Journal of physical and chemical reference data* **2003**, *32*, 519–878.

(70) Williams, W. D.; Ellard, J. A.; Dawson, L. R. Solvents Having High Dielectric Constants. VI. Diffusion in N-Methylacetamide1, 2. *Journal of the American Chemical Society* **1957**, *79*, 4652–4654.

(71) Brugmans, M. J.; Vos, W. L. Competition between vitrification and crystallization of methanol at high pressure. *The Journal of chemical physics* **1995**, *103*, 2661–2669.

(72) Khattab, I. S.; Bandarkar, F.; Fakhree, M. A. A.; Jouyban, A. Density, viscosity, and surface tension of water+ ethanol mixtures from 293 to 323K. *Korean Journal of Chemical Engineering* **2012**, *29*, 812–817.

(73) Diogo, H. P.; Santos, R. C.; Nunes, P. M.; da Piedade, M. E. M. Ebulliometric apparatus for the measurement of enthalpies of vaporization. *Thermochimica acta* **1995**, *249*, 113–120.

(74) Hunter, E.; Lias, S.; Mallard, W.; Linstrom, P. NIST Chemistry WebBook, NIST Standard Reference Database No. 69. 1998.

(75) Riera, M.; Götz, A. W.; Paesani, F. The i-TTM model for ab initio-based ion–water interaction potentials. II. Alkali metal ion–water potential energy functions. *Physical Chemistry Chemical Physics* **2016**, *18*, 30334–30343.

(76) Varma, S.; Rempe, S. B. Structural Transitions in Ion Coordination Driven by Changes in Competition for Ligand Binding. *Journal of the American Chemical Society* **2008**, *130*, 15405–15419.