Polarizable AMOEBA Model for Simulating Mg^{2+}·Protein·Nucleotide Complexes

Julian M. Delgado,[†] Péter R. Nagy,[‡,¶,§] and Sameer Varma[*,†,‖]

†*Department of Molecular Biosciences, University of South Florida, 4202 E. Fowler Ave., Tampa, FL-33620, USA*

‡*Department of Physical Chemistry and Materials Science, Faculty of Chemical Technology and Biotechnology, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

¶*HUN-REN-BME Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

§*MTA-BME Lendület Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

‖*Department of Physics, University of South Florida, 4202 E. Fowler Ave., Tampa, FL-33620, USA*

E-mail: svarma@usf.edu

Abstract

Molecular mechanics (MM) simulations have the potential to provide detailed insights into the mechanisms of enzymes that utilize nucleotides as cofactors. In most cases, the activities of these enzymes also require the binding of divalent cations to catalytic sites. However, modeling divalent cations in MM simulations has been challenging. The inclusion of explicit polarization was considered promising, but despite improvements over non-polarizable force fields and despite the inclusion of 'Nonbonded-fix (NB-fix)' corrections, errors in interaction energies of divalent cations with proteins

remain large. Importantly, the application of these models fails to reproduce experimental structural data on Mg^{2+}·Protein·ATP complexes. Focusing on these complexes, here we provide a systematic assessment of the polarizable AMOEBA model and recommend critical changes that substantially improve its predictive performance. Our key results are as follows. We first show that our recent revision of the AMOEBA protein model (AMOEBABIO18-HFC), which contains high field corrections (HFC) to induced dipoles, dramatically improves Mg^{2+}-protein interaction energies, reducing mean absolute errors (MAE) from 17 to 10 kcal/mol. This further supports the general applicability of AMOEBABIO18-HFC. The inclusion of many-body NB-fix corrections further reduces MAE to 6 kcal/mol, which amounts to less than 2% error. The errors are estimated with respect to vdW-inclusive density functional theory that we benchmark against CCSD(T) calculations and experiments. We also present a new model of ATP with revised polarization parameters to better capture its high field response, as well as new vdW and dihedral parameters. The ATP model accurately predicts experimental Mg^{2+}-ATP binding free energy in the aqueous phase and provides new insights into how Mg^{2+} associates with ATP. Finally, we show that molecular dynamics (MD) simulations of Mg^{2+}·Kinase·ATP complexes carried out with these improvements lead to a better agreement in global and local catalytic site structures between MD and X-ray crystallography.

Introduction

Mg^{2+} is the most abundant divalent cation in the intracellular fluid.[1–3] As a hard metal, it interacts directly with water, carboxylates, phosphates, hydroxyls, and carbonyls present in proteins, nucleotides, nucleic acids, and lipids.[4–7] It is shown to have metabolic, structural as well as regulatory functions in many diverse physiological processes.[1,3,4,6,8–10] For example, it modulates membrane structure, binds to dissolved nucleotides, such as ATP, stabilizes DNA and RNA structures, and serves as a metal cofactor in various enzymes, including kinases,

adenylyl cyclases, and helicases.

Molecular mechanics (MM) simulations can, in principle, provide detailed insights into such Mg^{2+}-dependent processes. Here we focus on MM simulations of Mg^{2+}·Protein·Nucleotide complexes. To simulate such complexes reliably, the employed MM model must accurately describe a wide range of interactions, including Mg^{2+}-water, Mg^{2+}-protein, Mg^{2+}-nucleotide, protein-water, nucleotide-water, and protein-nucleotide interactions.

MM models to describe protein-water interactions have been constructed carefully using reference data from gas- and condensed-phase experiments as well as high-level quantum mechanics (QM) calculations and are refined regularly to improve accuracy.[11–19] However, Mg^{2+} ions are not included during the parametrization of protein MM models. Mg^{2+} parameters are determined separately from reference data on Mg^{2+}-water interactions,[20–24] where they have been shown to perform reasonably well. In simulations consisting of both proteins and Mg^{2+} ions, Mg^{2+}-protein interactions are estimated using a predefined set of mixing rules for Lennard-Jones (LJ) terms. This, however, does not guarantee the reliability of predicted Mg^{2+}-protein interactions. In fact, previous studies,[25–36] including ours, show that the use of mixing rules can produce large errors in interactions of monovalent cations with proteins. Not surprisingly, for divalents like Ca^{2+}, these errors are even larger[31] and can be $>$ 50 kcal/mol. Similar errors can be expected for Mg^{2+}-protein interactions. In recent years, errors in ion-protein interactions have been reduced to some extent in an *a posteriori* manner through so-called "Nonbonded-fix (NB-fix)"-type approaches 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.[30–33,37–41] However, even after including NB-fix corrections, and in MM models that describe dipole polarization explicitly (polarizable MM models), errors in interactions of proteins with monovalent and divalent cations can exceed 10 and 20 kcal/mol, respectively.[31,36] This implies that to further improve accuracies, we need to look beyond NB-fix corrections.

Assessing the accuracy of Mg^{2+}-nucleotide interactions poses an additional challenge related to limited experimental data. The clearest evidence from experiments is that Mg^{2+} binds strongly to nucleotides in solution. Compilation of data from 34 different experiments,[42] including NMR, calorimetry, spectrophotometry, and pH titration, which we summarize in **Table S8** in the supporting information, shows that in the aqueous phase Mg^{2+}-ATP binding free energy lies between -7.2 and -9.9 kcal/mol with an average of -8.5 ± 0.7 kcal/mol. Not surprisingly, most enzymes that utilize nucleotides as cofactors or substrates are known to interact not just with the nucleotides but also with bound divalent cations, especially Mg^{2+} ions, that coordinate with both nucleotide and protein.[43]

While Mg^{2+}-nucleotide binding free energy is known from experiment, the binding modes of Mg^{2+} to nucleotides remain ambiguous. Early NMR experiments in the aqueous phase suggest that Mg^{2+} coordinates with oxygens from only one of the three ATP phosphates.[44,45] More recent ^{31}P NMR experiments and ^{18}O vibrational studies suggest that the predominant binding mode is bidentate, that is, Mg^{2+} coordinates with oxygens from two phosphates.[46,47] Finally, reinterpretation of the original ^{31}P data and newer ^{31}P NMR experiments suggest that Mg^{2+}-ATP coordination is predominantly tridentate.[47–49] In relation to these experiments, QM geometry optimizations conducted in implicit solvent suggest that the most stable binding mode is one in which oxygens from all three phosphates coordinate with Mg^{2+}, followed by a bidentate binding mode.[50]

This ambiguity is further complicated by experimental structural data on proteins complexed with Mg^{2+} and nucleotides. Previous statistical analysis of structures from the protein data bank (PDB),[51,52] as well as our own analysis (**Figure S1** in supporting information), reveal three predominant Mg^{2+}-ATP binding modes, two modes in which Mg^{2+} coordinates with oxygens of two phosphates ($\beta\gamma$ or $\alpha\beta$) and the third in which Mg^{2+} coordinates with oxygens of all three phosphates ($\alpha\beta\gamma$). None of the non-polarizable MM models, including CHARMM and AMBER, have been shown to reproduce this structural data.[52] The polarizable AMOEBA model does reproduce Mg^{2+}-ATP binding free energy in the aqueous

phase,[53] but as shown in the results section, when it is employed to conduct molecular dynamics (MD) simulations of Mg^{2+}·Kinase·ATP complexes, it fails to reproduce experimental structural data on Mg^{2+}-ATP binding modes, ATP-protein binding modes as well as ATP conformations within kinases. This happens despite the inclusion of NB-fix cross-terms for Mg^{2+}-protein interactions. Perhaps this is because large errors are still present in Mg^{2+}-protein interactions, or perhaps this results from a combination of errors in Mg^{2+}-protein, Mg^{2+}-nucleotide, and protein-nucleotide interactions. A further assessment is required.

There are two general approaches to addressing this problem. In one approach, the MM parameter space can be explored to directly target experimental structures of Mg^{2+}·Protein·Nucleotide complexes. Alternatively, Mg^{2+}-protein, Mg^{2+}-nucleotide, and protein-nucleotide interactions can be assessed separately, and then the improved models can be validated against experimental structures of Mg^{2+}·Protein·Nucleotide complexes. Here we chose the latter approach. The advantage of the latter approach is that we are not relying on experimental structural data for potentially fixing the underlying physics of intermolecular interactions. Instead, we aim to assess, add/fix the fundamental physics of intermolecular interactions and then examine the extent to which such improvements modify the correspondence between simulated and experimental structures.

We focus on the assessment and improvement of these interactions in the polarizable AMOEBA model. We build on our earlier work[33–36] and preceding foundational developments of AMOEBA.[54–59] In earlier work, we developed new Mg^{2+} LJ parameters that reproduce CCSD(T) data on Mg^{2+}-water clusters and accurately predict aqueous phase properties of Mg^{2+} [34] (**Table S1** in supporting information). In earlier work, we also recalibrated the polarization and LJ interaction terms of carbonyls, hydroxyls and carboxylate groups in proteins so that they responded equally well to both the high electric fields present near cations and lower electric fields. We had shown that the revised AMOEBA protein model, AMOEBABIO18-HFC22 where HFC stands for high field correction, combined with many-body NB-fix (MBNB-fix) corrections to interactions between monovalent cations and

proteins reduced errors in ion-protein interactions from 9.2 to 2.7 kcal/mol.[36] The errors were estimated with respect to van der Waals-inclusive density functional theory (DFT) that was benchmarked against CCSD(T)/CBS and quantum Monte Carlo calculations as well as experiments.[29,33–35,60,61] We had also shown that the revised model retains its intrinsic reliability in predicting protein structure and dynamics in the condensed phase.[36] with respect to the reference energies In this work, we first show that AMOEBABIO18-HFC22 dramatically improves Mg^{2+}-protein interaction energies, reducing mean absolute errors (MAE) from 17 to 10 kcal/mol. Since divalent cations were not considered during the calibration of AMOEBABIO18-HFC22, this result demonstrates the general applicability of our high-field corrections to AMOEBA. We then explore two different approaches to develop AMOEBA MBNB-fix cross-terms for Mg^{2+}-protein interactions and show that our chosen approach further reduces the MAE to 6 kcal/mol. Next, we recalibrate the triphosphate parameters in ATP and evaluate its performance and predictions in describing Mg^{2+}-nucleotide interactions in the aqueous phase. We show that the revised model continues to predict the experimental Mg^{2+}-ATP binding free energy accurately. The revised model is also more flexible and shows high probabilities for Mg^{2+} to form mono-, bi-, and tridentate coordination with ATP, as seen in NMR experiments discussed above. Then we recalibrate protein-ATP interactions. Finally, we conduct MD simulations of two different Mg^{2+}·Kinase·ATP complexes and evaluate the performance of the improved model against experimental structures. We show that our proposed improvements lead to a better agreement in local and global structures between MD and X-ray crystallography.

QM reference data

Reference energies and structures of ion-ligand and ligand-ligand clusters are obtained from Perdew-Burke-Ernzerhof (PBE0)[62,63] exchange-correlation density functional supplemented by Tkatchenko-Scheffler[64] self-consistent corrections for dispersion. We refer to this dispersion-correct DFT as PBE0+vdW. This hybrid functional mixes 25% exact exchange to alleviate

the local or semilocal approximation of DFT-based methods. This is particularly important when significant intramolecular charge transfer is involved, for example for hydrogen-bonded systems or systems involving ions. All single-point calculations as well as geometry optimizations employing PBE0+vdW are performed with FHI-AIMS[65] and with 'really tight' basis sets. Total energies and electron densities are converged to 10^{-6} eV and 10^{-5} electrons, respectively, and geometries are optimized with a force tolerance of 10^{-3} eV/Å.

Our choice for PBE0+vdW stems from the three main reasons, discussed in more detail in our recent work.[36] Firstly, 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.[66,67] Secondly, we have reported previously[33–35,60,61] 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 homogeneous clusters of various small molecules (waters, alcohols, amides, carboxylates and aromatics). Finally, we have shown that with reference to LNO-CCSD(T), Mg^{2+} slightly overbinds water clusters by 1-2 kcal/mol and within the Harmonic approximation, it ywith respect to the reference energiesields gas-phase Mg^{2+}-water binding free energies with a MAE of less than 1 kcal/mol with respect to experiments.[35] Here, we further benchmark this method for interactions of Mg^{2+} ions with various neutral and negatively charged small molecules.

Reference CCSD(T) information to benchmark DFT+vdW is obtained using the local natural orbital (LNO) method[68–71] as implemented in the Mrcc package.[72,73] Optimized geometries from PBE0+vdW are used for single-point energy LNO-CCSD(T) calculations. Extrapolation of LNO-CCSD(T) results toward the approximation free CCSD(T) value and the corresponding local error estimates are performed using the Tight and very Tight LNO-CCSD(T) threshold sets[69,70] according to the extrapolation scheme of Ref. 70. For the LNO-CCSD(T) computations we employ Dunning's correlation-consistent basis sets augmented with diffuse functions (aug-cc-pVXZ, X=T, Q, 5) for first row elements, and the

corresponding core-valence basis sets[74] for Mg^{2+}. While deep-core electrons of all atoms are kept frozen, the sub-valence electrons of Mg^{2+} have to be correlated for reliable results. Extrapolation toward the complete basis set (CBS) limit of CCSD(T) results is performed via standard formulae,[75,76] yielding CBS(X,$X+1$) values from the aug-cc-pVXZ results. The remaining basis set incompleteness error (BSIE) of the final LNO-CCSD(T)/CBS(Q,5) interaction energies are estimated as the difference between the CBS(T,Q) and CBS(Q,5) results. The accuracy of the final Tight–very Tight LNO-CCSD(T)/CBS(Q,5) interaction energies can be conservatively characterized using a cumulative BSIE and local error estimate, which indicates that our reference results are within ±0.5 kcal/mol of the approximation-free CCSD(T)/CBS ones for all studied complexes.

Induced dipoles and polarizability tensors are computed using second-order Møller-Plesset perturbation (MP2) theory[77] implemented in Gaussian09.[78] We use Dunning's[79] triple-zeta correlation consistent polarize valence only basis set augmented with diffuse functions (aug-cc-pVTZ), which has been shown to have comparable accuracy to CCSD(T) for calculation of molecular dipole moments and polarizabilities.[80]

MD parameters

All molecular dynamics (MD) simulations are carried out using Tinker9.[81] Temperature is regulated using an extended ensemble approach[82] and with a coupling constant of 0.1 ps. In isobaric simulations, pressure is regulated using a Monte Carlo approach,[83,84] and with a coupling constant of 0.1 ps. Equations of motions are integrated using the RESPA algorithm with an outer time step of 3 fs and an inner time step of 0.5 fs.[85] Electrostatic interactions are computed using particle mesh Ewald (PME), with a direct space cutoff of 9 Å. Van der Waals interactions are computed for inter-atomic distances smaller than 9 Å. Both energy and virial are corrected with an isotropic long-range correction term.[86] The convergence cutoff for induced dipoles is set at a value of 10^{-2} Debye in condensed phase and 10^{-5} Debye in gas phase. Other control functions and parameters are set to their default values.

Energy optimizations employing MM models are performed using the optimize program in TINKER8[87] using an RMS potential gradient cutoff of 10^{-2} kcal mol^{-1} Å^{-1}.

MD of Mg^{2+}·Kinase·ATP complexes

The coordinates for the starting structure of glycogen synthase kinase-3 beta (GSK3β) are taken from PDB ID 1PYX, which was solved at a resolution of 2.4 Å.[88] The coordinates for the starting structure of cyclin-dependent kinase 2 (CDK2) are taken from PDB ID 1B38 whose structure was solved at a resolution of 2 Å.[89] In both cases, we simulate the monomeric forms of the enzyme, and in neither case, the enzyme are bound to substrates. Missing residues, all of which were $>$ 10 Å away from the ATP binding pocket, are added using Modeller.[90] The longest contiguous stretch of residues that we add for CDK2 is from residue numbers 36-43, which belongs to a loop. In the case of GSK3β, the longest contiguous stretch of residues we add is from residue numbers 86-90, and that also belongs to a loop. Protein termini are capped using ACE and NME when needed and the protonation states of titratable residues are assigned their standard protonation states under physiological conditions. Missing hydrogens are added by optimizing hydrogen bond networks using PDB2PQR.[91] Crystal waters are retained, and additional water molecules are introduced to fill cubic boxes of edge sizes 98.8 and 90.7 Å for GSK3β and CDK2, respectively. The GSK3β box contains 29,698 water molecules, and the CDK2 box contains 22,465 water molecules. A salt concentration of 20mM is set by randomly replacing water molecules with appropriate numbers of Na^+ and Cl^- ions. Additional 2 and 7 ions in CDK2 and GSK3β boxes are added to make the systems charge neutral. Following energy minimization, the boxes are simulated under NPT conditions (1 atm, 298 K) with harmonic restraints applied to all protein backbone atoms, ATP, and Mg^{2+} ions. Initially, a spring constant of 30 kcal/mol/Å^2 is used, which is decreased to 0 in steps of 6 kcal/mol/Å^2 per 900 picoseconds. After this, we perform unrestrained NPT simulations.

Double-decoupling method

The standard free energy change (ΔF_i) associated with the binding of Mg^{2+} to a specific site (i) on ATP in solution is computed as[92,93]

$$\Delta F_i = \Delta F_{rstr}^{\text{ATP-Mg(i)}}(aq.) - \Delta F^{\text{Mg}}(aq.) - \Delta F_{rstr}^{i}. \quad (1)$$

Here, $\Delta F_{rstr}^{\text{ATP-Mg(i)}}(aq.)$ is the free energy change associated with decoupling the vdW and electrostatics interactions of Mg^{2+} from ATP and the solvent, while it is held restrained at binding site i. $\Delta F^{\text{Mg}}(aq.)$ is the solvation energy of Mg^{2+} in the absence of ATP, and ΔF_{rstr}^{i} is the contribution from the restraining potential that is subtracted from the solvation energy of Mg^{2+}.

The electrostatic interactions of Mg^{2+} are decoupled from the rest of the system by linearly scaling the multipole and polarization parameters with a scaling factor (λ) that is changed from $\lambda = 1$ to $\lambda = 0$ with a constant step size $\Delta\lambda = 0.1$. As electrostatic interactions of Mg^{2+} are decoupled, we also turn on restraints that keep the Mg^{2+} bound to site i. vdW interactions of Mg^{2+} are decoupled using the following set of scaling values $\lambda = \{1, 0.9, 0.8, 0.75, 0.7, 0.65, 0.62, 0.6, 0.55, 0.5, 0.4, 0\}$. Therefore, a total of 22 independent simulations with different values of λ for electrostatics and vdW are run for each system. We use Bennet acceptance ratio (BAR)[94] to calculate the free energy difference between neighboring λ. Since we are using Ewald summation to compute the electrostatics with periodic boundary conditions, we ensure that the charge of the system remains neutral by simultaneously decoupling 2 Cl^- ions present in solution.[35] Therefore, equation 1 is rewritten as

$$\Delta F_i = \Delta F_{rstr}^{\text{ATP-Mg(i)Cl}_2}(aq.) - \Delta F^{\text{MgCl}_2}(aq.) - \Delta F_{rstr}^{i}. \quad (2)$$

In all cases, we apply restraints using a single harmonic potential defined as $u(r) = k(r - r_0)^2$, where k is spring constant and r_0 is the equilibrium distance that we chose separately for each binding mode as the average distance observed in 30 ns of unrestrained

MD in the $\lambda = \{1, 1\}$ state. For monodentate binding modes, the restraint is applied between Mg^{2+} and the coordinating phosphate oxygen. For multi-dentate binding modes, the restraint is applied between Mg^{2+} and the centers of masses of the coordinating phosphate oxygens. The contribution from the restraining potential is estimated analytically as

$$\Delta F^i_{rstr} = k_b T \ln(4\pi c^\circ \int_0^\infty r^2 e^{-\beta u(r)} dr) \tag{3}$$

where, c° is the unit concentration. To assess the effect of the choice of the spring constant, we compute ΔF_i for the original model using two different spring constants, $k = \{5, 10\}$ kcal/mol/Å^2. We note that ΔF_i differences by less than 0.2 kcal/mol. The values we present in the results section are computed using $k = 5$ kcal/mol/Å^2.

The error associated with the binding free energy (ΔF_i^{bind}) of each mode is computed as $\delta\Delta F_i^{bind} = \sqrt{\sum_j (\delta^j \Delta F_i^{bind})^2}$, where $\delta^j \Delta F_i^{bind}$ is the Montecarlo error for each window j for each mode i. We keep track of both the free energy and associated error as a function of simulation time, and based on this, we conduct 15 ns sampling for each λ value to get errors below 0.5 kcal/mol.

Multiparameter optimization

MM parameters are optimized using the Nelder-Mead simplex-based algorithm as implemented in the ParOpt software.[95,96] The Nelder-Mead coefficients are set to their standard values for reflection ($\alpha = 1.0$), contraction ($\beta = 0.5$), expansion ($\gamma = 2.0$) and size ($\delta = 0.5$). For each optimization, a random point within the domain of the parameter space is selected and the algorithm is iterated until the simplex is converged, that is when the root mean square distance between the simplex vertices and the centroid is below 10^{-5} or until 1000 steps are completed. The parameters with the lowest value for the target function are selected after the algorithm is run starting at many different initial values. The parameter domain that we explore as well as the number of optimizations we perform are context-dependent,

and, therefore, are discussed at appropriate places in the results section.

Results and Discussion

This section is organized as follows. We first assess and propose improvements to Mg^{2+}-protein interactions. To account for high electric fields response of phosphates, we present a recalibration of nucleotide triphosphate parameters. Then we propose a new set of Mg^{2+}-phosphate interaction cross-terms and assess the model's performance in predicting Mg^{2+}-ATP binding in the aqueous phase. Finally, we propose a new set of nucleotide-protein interaction cross-terms and assess the performance of the revised model in simulating Mg^{2+}·Kinase·ATP complexes.

Mg^{2+}-protein interactions

To assess the performance of AMOEBA in predicting Mg^{2+}-protein interaction energies, we extract Mg^{2+} clusters from PDB structures and examine how AMOEBA performs in predicting interaction energies between Mg^{2+} ions and the rest of the cluster. We extract clusters from PDB in an unbiased manner as follows. First, we download all X-ray structures from the PDB that contain Mg^{2+} ions and have resolutions better than 1.5 Å.[4] From these structures, we extract all atoms that are within 6 Å from Mg^{2+}. Even if a single atom of an amino acid is within this cutoff, the entire amino acid is included as part of the cluster. Then we discard all clusters that contain non-protein or non-water atoms or more than one Mg^{2+} ion. At this stage, we are left with 29 clusters. These clusters contain a single Mg^{2+} ion, peptide fragments and water. 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 water, are added. The clusters are then energy minimized using the

original AMOEBA model with position restraints applied to atoms that were resolved in PDB structures.

The smallest cluster in our set consists of 65 atoms and the largest cluster consists of 285 atoms. We note that in 18 of these 29 clusters, Mg^{2+} is 5-fold or 6-fold coordinated. In the remaining clusters, the smaller coordination is perhaps due to missing water molecules, although, we understand that coordination number is not just a property of the ion, but is also influenced by the environment.[97] These clusters also differ from each other in terms of the numbers of hydroxyls, carboxylates, and carbonyls as well as the number of waters in the Mg^{2+} ion's first coordination shell (see **Table S3** in supporting information for details). We note, however, that the hydroxyl oxygen is underrepresented in this cluster set, but this is expected for Mg^{2+} based on a previous PDB survey.[98] Additionally, we note that these clusters also provide a good distribution of geometries around energy minima. The peak of the distribution of distances between Mg^{2+} and their first shell coordinators is close to the optimal distance in Mg^{2+}-water clusters[34,99] and both the repulsive hard-wall and first shell outer boundary distances are represented (see **Figure S2** in supporting information).

Using each of these clusters, we determine interaction energies (ΔE) between Mg^{2+} and everything else in the clusters. **Figure 1** compares the interaction energies obtained from the original AMOEBA model against those obtained from our benchmarked QM method (PBE0+vdW). We note that predictions from the original model correlate excellently with reference QM in the entire range of interaction energies [-210, -820] kcal/mol. However, both the MAE of 17 kcal/mol and the maximum error of 53 kcal/mol are very high. The magnitudes of these errors are similar to that noted for the CHARMM DRUDE MM model in the case of Ca^{2+} ions, despite the inclusion of NB-fix corrections.[31] When considering only those clusters in which Mg^{2+} is 6-fold coordinated, the MAE of 22 kcal/mol is even larger. We provide an assessment of 6-fold coordinated clusters separately because 6-fold coordination is known to be the preferred coordination of Mg^{2+} ions when they bind proteins.[5]

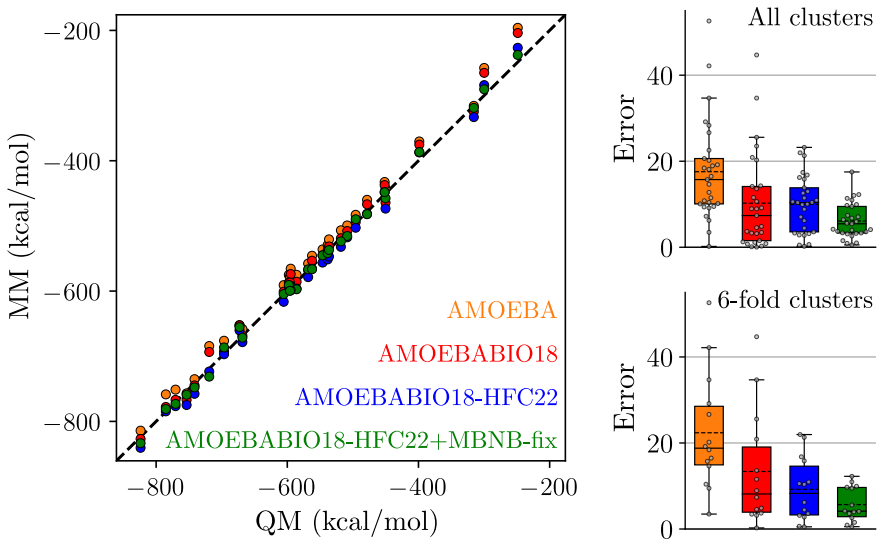

Figure 1: Performance of AMOEBA models in predicting Mg^{2+}-protein interaction energies. The scatter plot compares the interaction energies computed using benchmarked QM (PBE0+vdW), against four AMOEBA models: AMOEBA,[59] AMOEBABIO18,[100] AMOEBABIO18-HFC22[36] and AMOEBABIO18-HFC22+MBNB-fix that includes Mg^{2+}-protein MBNB-fix corrections. See related text for explanations of these models. Next to the scatter plot, we show the distribution of absolute errors in the predictions of AMOEBA models computed with respect to PBE0+vdW. The mean and median are represented by dashed and solid lines, respectively.

Figure 1 also shows the performance of another AMOEBA model, which we refer to as the 'AMOEBABIO18' model. The modifications that this model has concerning Mg^{2+} protein interactions are revised Glu/Asp carboxylate parameters, modified Mg^{2+} polarizability, and LJ parameters and explicit NB-fix cross-terms for interactions of carboxylates with Mg^{2+} and amines.[101] The AMOEBABIO18 model performs better than the original model, with a smaller MAE of 10 kcal/mol, and also a smaller MAE of 13 kcal/mol if only 6-fold clusters are considered. The maximum error of 45 kcal/mol is also lower compared to the original model.

Next, we determine the performance of our revised AMOEBABIO18-HFC22 model.[36] This model has new Mg^{2+}-water parameters that reproduce CCSD(T) cluster data and accurately predict aqueous phase properties of Mg^{2+}[35] (**Table S1** in supporting information). Additionally, the revised model has new parameters for hydroxyls, carbonyls, and carboxylates in proteins that were shown to perform well at the high electric fields present near

cations.[36] However, the revised model does not have any NB-fix corrections for Mg^{2+} interactions with proteins. We note that even without NB-fix corrections it performs better than the original and AMOEBABIO18 models. The MAE reduces to 10 kcal/mol for all clusters and 9 kcal/mol for 6-fold clusters. Additionally, the maximum error now reduces dramatically to 22 kcal/mol.

To further reduce this error, we develop ligand-specific NB-fix type LJ cross-terms for Mg^{2+}. The ligands are carbonyls (Asn/Gln/backbone/ACE/NME), carboxylates (Asp/Glu) and hydroxyls (Tyr/Ser/Thr). The typical strategy to develop these cross-terms is to use reference energies from two-body interactions.[30–32,37–41] For example, reference energies between Mg^{2+} and NMA can be computed at different distances and used as target data to optimize Mg^{2+}-carbonyl LJ cross-terms. This strategy, however, does not account for many-body cooperativity effects. To account for them, we have shown that LJ cross-terms can also be determined from reference data on many-body clusters.[96]

To examine whether we should use the traditional two-body cluster approach or use higher-order clusters, we carry out the following exercise. We first determine Mg^{2+}-carbonyl and Mg^{2+}-carboxylate LJ cross-terms using reference data on Mg^{2+}-NMA and Mg^{2+}-acetate dimers, respectively, and use them to predict interaction energies in 6-fold clusters made up of waters, NMA, and acetate. Our set of 6-fold clusters consists of 100 geometries belonging to 10 unique chemical compositions shown in **Figure S3** of the supporting information. Alternatively, we determine cross-terms using the same 6-fold clusters we used for testing the results of the former strategy and use them to predict dimer energies. In all cases, the reference energies are determined from PBE0+vdW. For Mg^{2+}-acetate and Mg^{2+}-NMA dimers, we also benchmark PBE0+vdW against LNO-CCSD(T)/CBS (**Figure 2 (a)**) and find that it performs excellently. The NMA and acetate parameters are taken from our earlier work that perform better at reproducing high-field reference data.[36] In both strategies, we use the Nelder-Mead algorithm for optimizing LJ cross-terms using a parameter range of [3.0, 4.2] Å and [0.1, 0.9] kcal/mol for R^0 and ϵ, respectively. We carry out 100 different

optimizations starting with different seeds, which produce 100 different sets of parameters. We chose the parameter set that yields the smallest error with respect to target energies.

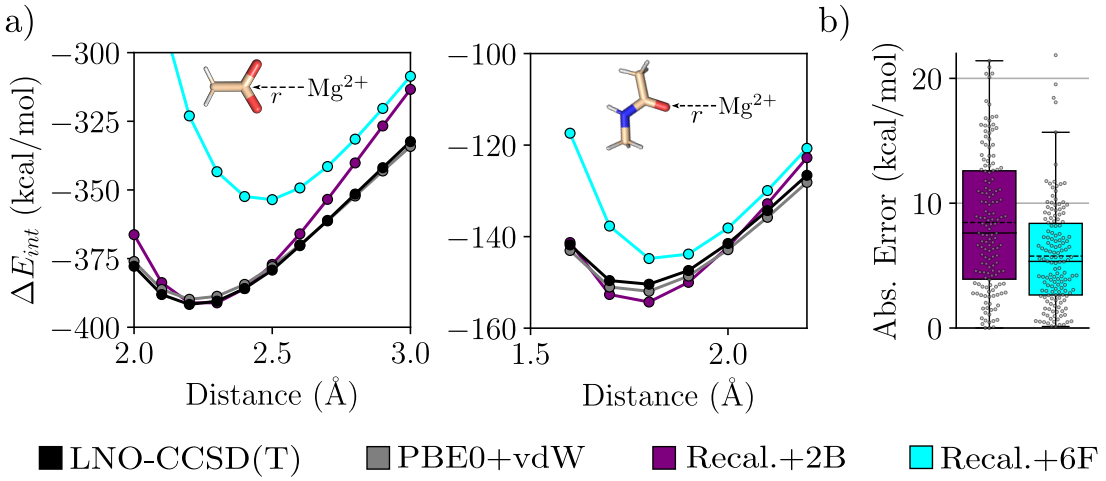

Figure 2: (a) Distance-dependent interaction energies of Mg^{2+} with acetate and NMA. 'Revised+2B' refers to our revised AMOEBA model for small molecules (AMOEBA09-HFC22)[36] along with Mg^{2+}-carbonyl and Mg^{2+}-carboxylate LJ cross-terms determined using Mg^{2+}-NMA and Mg^{2+}-acetate dimers. 'Revised+6F' refers to our revised AMOEBA model for small molecules along with Mg^{2+}-carbonyl and Mg^{2+}-carboxylate LJ cross-terms determined using 6-fold clusters of NMA, acetate and water. (b) Distributions of the absolute error in interaction energies of Mg^{2+} ions with 6-fold clusters computed using the 'Revised+2B' and 'Revised+6F' models.

The results of this exercise are shown in **Figure 2**. We make three key observations. Firstly, LJ cross-terms determined using 2-body clusters do not perform as well compared to those determined using 6-fold clusters in estimating interaction energies of 6-fold clusters. Secondly, the LJ cross-terms determined using 6-fold clusters perform poorly in predicting interaction energies of Mg^{2+}-NMA/acetate dimers. This means that cross-terms determined using higher-order clusters should be used with caution to compute energies of lower-order clusters and vice versa. Thirdly, even when considering only 6-fold clusters, we are unable to reduce the MAE below 5 kcal/mol, which means that we may have reached the predictive limit of AMOEBA's functional form. Together, these observations imply there is still essential physics missing from the AMOEBA model. Nevertheless, these results suggest that since we intend to use Mg^{2+} for simulating protein binding in 6-fold coordination geometries, we need to determine cross-terms using higher-order clusters.

The other issue concerns whether LJ cross-terms should be determined using small molecule representatives of proteins or directly from protein clusters. Previously, we had chosen to do this using protein clusters, as the multipoles in small molecules slightly differ from their representative chemical groups in proteins.[36] We use the interaction energies from the 6-fold PDB clusters as target data and determine R^0_{ij} and ϵ_{ij} such that they minimize the root mean square error (RMSE) to this target. The best parameter set obtained after the exhaustive search is provided in **Table S4** of the supporting information. We note that the biggest difference to the terms calculated from the mixing rules relates to making the energy well (ϵ_{ij}) deeper. **Figure 1** shows that for 6-fold clusters these new LJ cross-terms further reduce MAE to 6 kcal/mol and the maximum error to 12 kcal/mol. Even when all the clusters are taken into account the MAE remains around 6 kcal/mol and the overall distribution of errors also shrinks, except for a single case whose error lies around 17 kcal/mol.

Overall, our revised parameters dramatically reduce MAE in Mg^{2+}-protein interaction energies from 17 to 6 kcal/mol and maximum error from 53 to 17 kcal/mol. For 6-fold clusters that are expected to be common in Mg^{2+}·Protein·Nucleotide complexes,[98] MAE reduces from 22 to 6 kcal/mol and the maximum error reduces from 53 to 12 kcal/mol.

Polarization, LJ and torsional parameters of the triphosphate

In earlier work, we showed that the accuracy of an MM model to describe cation-ligand interactions depends strongly on the ability of ligands to respond to the high electric field present near cations.[33–36] To assess this in AMOEBA's ATP model,[53] we compute the induced dipole moment of dimethyl phosphate (DMP) as a function of its distance from a unit positive point charge. DMP is also used in the parameterization of other MM models of nucleotides, including CHARMM and AMBER.[102–104] We do this calculation in two ways. In the first approach, we compute DMP's field response using AMOEBA's DMP model.[105] We note that the atomic polarizabilities of AMOEBA's DMP reproduce the molecular polarizability tensor obtained from MP2 theory. Our calibration of atomic polarizabilities against

MP2 yields very similar values compared to the original DMP's model. In the second approach, we transfer the polarizabilities of AMOEBA's triphosphate group to AMOEBA's DMP model. The results of these calculations are shown in **Figure 3** and compared to reference values obtained from MP2 theory. We find that DMP's polarization model outperforms the triphosphate's polarization model in predicting both low and high electric field responses. We, therefore, transfer atomic polarizabilities of DMP to ATP's triphosphate group, as listed in **Table S5** of the supporting information.

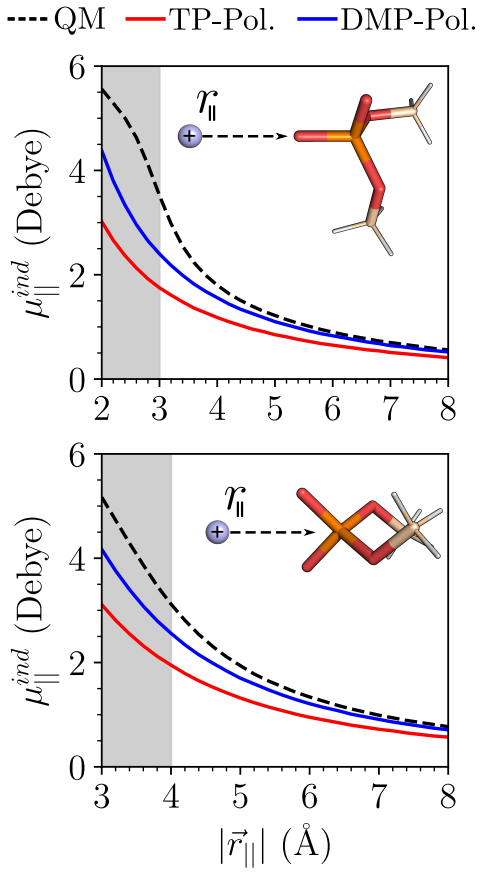

Figure 3: Induced dipole moment of DMP as a function of distance ($|\vec{r}|$) from a unit positive point charge. The primary component of the induced dipole, $\mu_{||}^{ind}$, which is parallel to the distance vector, is shown. DMP-pol refers to the set of atomic polarizabilities in AMOEBA's DMP model,[105] and TP-pol refers to the set of atomic polarizabilities in ATP/GTP's triphosphate group.[53]

Since we modify triphosphate's polarization term, we will need to recalibrate triphosphate's interaction with water. To do this, we extract ten random snapshots from an MD simulation of ATP in water. Then we discard all water, except those that coordinate directly

with ATP's triphosphate group. From these ATP-water clusters, we compute interaction energies between ATP and all water molecules. We use the interaction energies from the original model as reference and recalibrate the LJ parameters of the triphosphate oxygens to minimize the RMS error with respect to the reference energies. After recalibration, the RMS error is 0.04 kcal/mol/water. We also note that the new parameters are only slightly different from the original ones, and are provided in **Table S5** of the supporting information. To further validate the new parameters, we compute interaction energies of triphosphate with waters placed at varying distances from the triphosphate oxygens and compare them against PBE0+vdW. To avoid protonation of triphosphate in QM calculations, we perform these calculations in the presence of all waters within triphosphate's first coordination shell. The results are shown in **Figure S4** of the supporting information. We find that the recalibrated model performs well against QM in describing triphosphate-water interaction energies.

Since polarization and LJ parameters are modified, we examine whether dihedral parameters also need to be revised. Since the triphosphate is defined using two dihedrals, we generate a 2-dimensional potential energy surface (2D-PES). We rotate the dihedral angles Oα-Pβ-Oβ-Pγ (ψ) and Pα-Oα-Pβ-Oβ (ϕ) in increments of 15° and generate 576 geometries. We energy minimize these geometries with dihedral restraints using the original AMOEBA model. We then compute the total potential energy of each geometry and scale the PES by subtracting out the energy of the most stable geometry. **Figure 4** compares the PES computed from the original model against that obtained from PBE0+vdW. We note that the original model deviates from QM by an average of 6.3 kcal/mol. Furthermore, the energetic barrier around $\psi = \phi = 0°$ is substantially weaker in the original model. Additionally, the relatively more stable geometries are more localized in ϕ around $\psi = 0°$, which perhaps enhances the torsional rigidity of the original model.

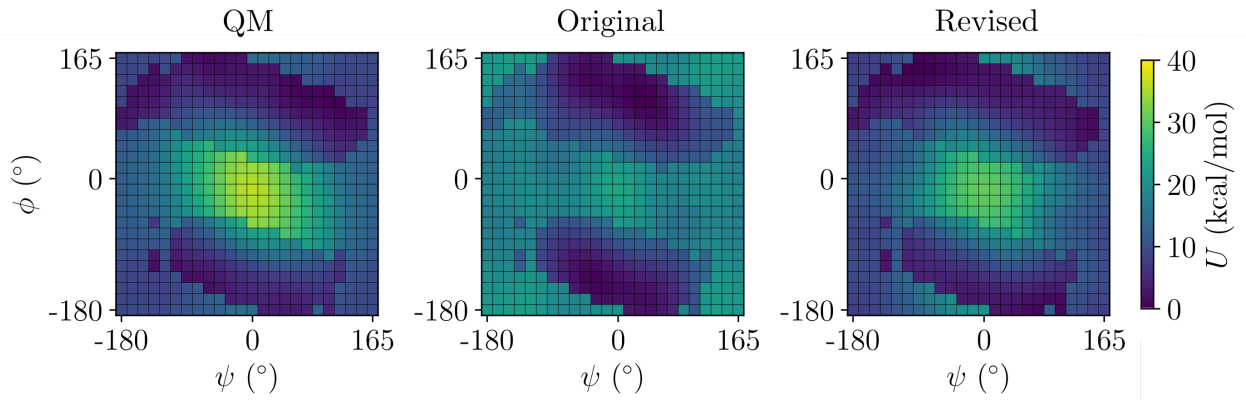

Figure 4: 2D-PES of tripohsophate dihedrals Oα-Pβ-Oβ-Pγ (ψ) and Pα-Oα-Pβ-Oβ (ϕ). The QM 2D-PES is computed using PBE0+vdW. 'Original' refers to the ATP parameters developed by Walker et al.[101] and 'Revised' refers to the modifications of the triphosphate parameters in this work.

To recalibrate the dihedrals, we construct an error function that includes not only the 2D-PES determined above but also a set of four optimized structures, one each in the local minima of the cis-cis, cis-trans, trans-cis, and trans-trans conformations, as shown in **Figure S5** of the supporting information. We carry out a total of 100 Nelder-Mead optimizations starting from randomly selected simplex and explore a parameter range of [-20, 20] kcal/mol for the coefficients of the dihedral term. The overall parameter space explored is shown in **Figure S6** of the supporting information. The parameter set that minimizes the error function reduces the PES MAE from 6.3 to 1.4 kcal/mol and the largest RMSD in structure drops from 0.9 to 0.5 Å. Both the original and revised parameters for the triphosphate are provided in **Tables S5 and S6** of the supporting information. **Figure 4** shows the 2D-PES generated from the final set of polarization, LJ and dihedral parameters. We note that the main features of the 2D-PES that were concerning in the original model are now resolved.

Mg^{2+}-triphosphate interactions

To determine Mg^{2+}-triphosphate LJ cross-terms, we make a minor adjustment to the protocol used above for determining Mg^{2+}-protein interactions. Instead of using triphosphates, we will use its small molecule representative DMP. The main reason we do not use triphosphates is because it carries a high negative charge of -4 eu. This high negative charge will produce

long electron density tails in gas phase QM calculations, that are likely to be much shorter in the condensed phase due to inter-molecular electron repulsion. Since our goal is to get simulations right in the condensed phase, we decided not to capture the electrostatic effect of the long tails in the gas phase. Therefore, we avoid using triphosphates to obtain LJ cross-terms, and instead use DMP.

We obtain LJ cross-terms with DMP using reference energies and structures associated with the following substitution reaction

$$[\mathrm{MgW}_6]^{2+} + n(\mathrm{DMP}) \rightleftharpoons [\mathrm{Mg(DMP)}_n\mathrm{W}_{6-n}]^{2+} + n\mathrm{W} \tag{4}$$

where $n = \{1, 2\}$. **Table 1** lists the substitution energies computed using different methods. We first note excellent agreement between PBE0+vdW and LNO-CCSD(T). The large discrepancies that we find in the original model are partially mediated by the use of our new Mg^{2+} parameters and are further reduced by introducing LJ cross-terms between Mg^{2+} and the phosphate oxygens. We also note that all models correctly capture the optimized geometry obtained from QM (**Figure S7** in supporting information). The LJ cross-terms are provided in **Table S7** of the supporting information. We note that transferring Mg^{2+}-DMP LJ cross-terms to Mg^{2+}-triphosphate LJ interactions can lead to errors due to differences between triphosphate and DMP multipoles. However, the differences in multipoles are small, and we expect the associated error to be of the order of 1-2 kcal/mol.

Table 1: Substitution energies in kcal/mol computed for Equation 4 using different methods.

Method	$n = 1$	$n = 2$	MAE
LNO-CCSD(T)	−202.0	−315.9	
PBE0+vdW	−201.9	−316.8	0.5
AMOEBABIO18	−214.3	−330.5	13.5
AMOEBABIO18-HFC22	−209.8	−323.1	7.5
AMOEBABIO18-HFC22+NBfix	−206.5	−315.5	2.4

Mg^{2+}-ATP binding in bulk water

After introducing modifications to the triphosphate group and its interaction with Mg^{2+}, we now predict the structure and thermodynamics of Mg^{2+}-ATP binding in the aqueous phase.

We first identify 'stable' binding modes of Mg^{2+}-ATP complexes in the aqueous phase. Overall, there are 15 possible coordination binding modes between Mg^{2+} and the different oxygen atoms of ATP (see **Figure S8** in the supporting information). We solvate optimized geometries of each of these 15 6-fold coordinated binding modes in separate 70 Å cubic boxes of water with 100 mM of KCl. The choice of the specific ionic strength is to emulate experimental conditions in which KCl is used as a buffer for determining Mg^{2+}-ATP binding free energies.[42] Note that there are two additional K^+ in each box to balance the net charge. We first conduct 3 ns long NPT simulations (1 atm, 298 K) with harmonic distance restraints (25 kcal/mol/Å^2) between Mg^{2+} and its six coordinators. We then reduce these harmonic restraints over the next 1.2 ns of NPT simulations. Then we carry out an additional 300 ps of NPT simulations to determine average box volumes. Next, we select snapshots from the final 300 ps trajectories that have volumes closest to the average volumes and use them to initiate 15 unrestrained NVT runs, one for each binding mode. We carry out NVT simulations for 30 ns.

Analysis of these trajectories reveals the following seven stable binding modes in the revised ATP model: three monodentate modes (α, β, and γ), three bidentate modes ($\alpha\beta$, $\beta\gamma$, and $\alpha\gamma$) and one tridentate mode ($\alpha\gamma\gamma$). These binding modes are shown in **Figure 5(a)**. Note that $\alpha\gamma$ and $\alpha\gamma\gamma$ are shown together as we find them to switch frequently between each other. Mg^{2+} ions in the remaining eight modes switch coordinations to form one of these seven binding modes (see **Figure S9** in the supporting information). In comparison to the revised model, the original model shows only four binding modes, α, β, γ, and $\alpha\gamma$ consistent with a previous study.[53] These four binding modes are a subset of the seven binding modes observed with the revised model. This perhaps means that our revised model is more flexible in that it can adopt more conformations in the presence of Mg^{2+}.

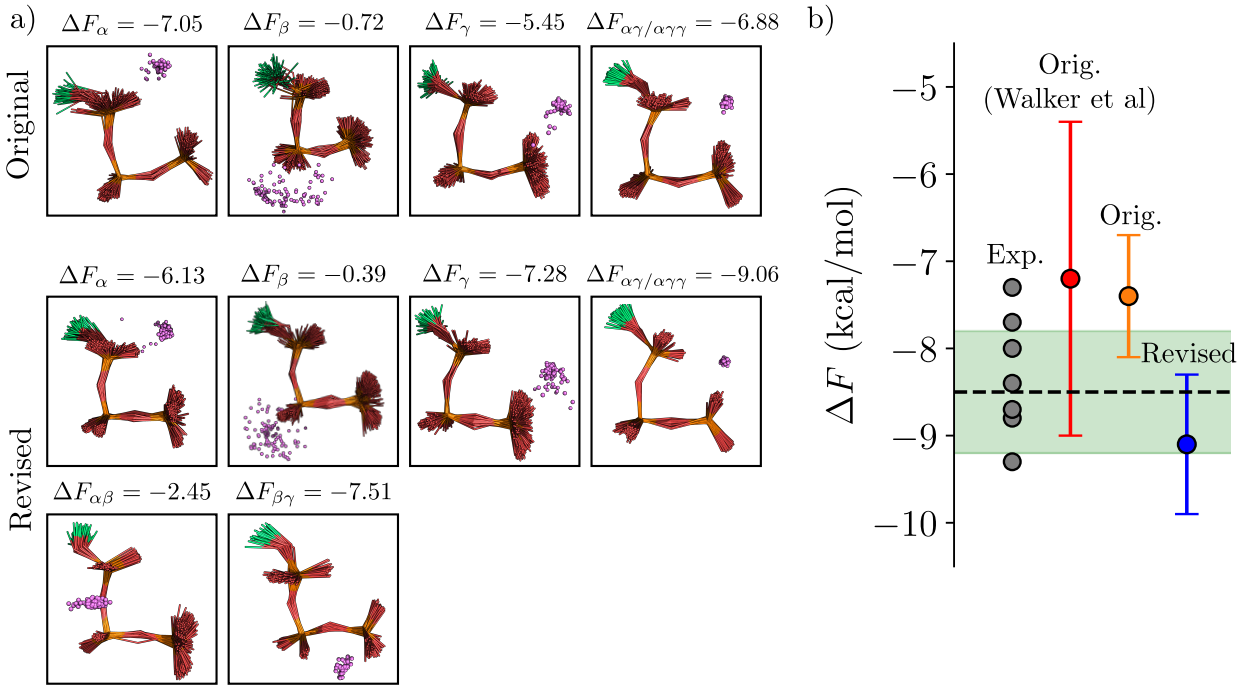

Figure 5: (a) Predicted binding modes of Mg^{2+} to ATP in solution. Mg^{2+} is shown as pink spheres and the truncated adenosine groups are colored green. The binding free energies of Mg^{2+} to each mode ΔF_i are provided in units of kcal/mol. 'Original' refers to the ATP parameters developed by Walker et al.[101] and the Mg^{2+} parameters used in that development. 'Revised' refers to the modifications to triphosphate and Mg^{2+}-triphosphate parameters introduced in this work and the Mg^{2+} parameters we developed previously.[35] (b) Mg^{2+}-ATP binding free energies from different models are compared to experiment. Note that estimates from the revised model are predictions as they are a direct outcome of the model that we revised solely based on QM targets consisting of local interactions. This is not the case with the original model. Note also that two estimates are provided for the original model, one taken from Walker et al[53] and the other determined in this work. The grey circles are estimates from seven experimental methods, compiled by others,[42] and summarized in **Table S8** in the supporting information. The dashed black line is the average of all experiments and the green shaded region represents the associated standard deviation.

To quantify Mg^{2+}-ATP affinity in each of these binding modes, we compute standard binding free energies using the double decoupling method[92,93] (details provided in the methods section). The results are shown in **Figure 5(a)**. We make two main observations. Firstly, revision of the model not only yields new binding modes but also changes the relative strengths of binding modes seen in the original model. Specifically, in the original model, the energetically most favored binding mode is α, followed by $\alpha\gamma$ and γ. However, in the revised model, the $\alpha\gamma/\alpha\gamma\gamma$ is most favorable followed by $\beta\gamma$ and γ. In the revised model, we find that any interaction involving the γ oxygen leads to stronger binding. Secondly, while the original model did not produce any of the three binding modes ($\alpha\beta$, $\beta\gamma$, and $\alpha\beta\gamma$)

preferentially seen in PDB structures of Mg^{2+}·Protein·ATP complexes, the revised model suggests that the $\beta\gamma$-mode does exist in the aqueous phase with high probability. The other prominent bidentate binding mode, the $\alpha\beta$-mode also exists in the aqueous phase, but with lower probability compared to the $\beta\gamma$-mode. We do not observe the third binding mode observed in PDB structures, the $\alpha\beta\gamma$-mode, which suggests that it perhaps results due to complexation with protein.

Figure 5(b) shows the total Mg^{2+}-ATP binding free energy, estimated by Boltzmann averaging the standard binding free energies of all modes $\Delta F = -k_b T \ln \sum_i e^{-\beta \Delta F_i}$. The standard error, as shown by Walker et al.[53] is $\delta\Delta F = \sqrt{\sum_i (\delta \Delta F_i)^2}$, where $\delta\Delta F_i$ is the error of each mode. We first note excellent agreement between the previously calculated ΔF of -7.2 ± 1.8 kcal/mol[53] and the one we obtained with the same original model (-7.4 ± 0.7 kcal/mol), but following a systematic protocol to identify binding modes. The revised model yields $\Delta F = -9.1 \pm 0.8$ kcal/mol. All of these values are within the experimental range compiled by Goldberg and Tewari[42] and summarized in **Table S8** in the supporting information. Note that the predictions from the revised model are a direct outcome of the revisions we made to the model using QM targets consisting only of local interactions.

Protein-triphosphate interactions

To model nucleotide binding to proteins, we also need to assess the accuracies of protein-triphosphate interactions. **Figure 6** shows the interaction energies between DMP and Lys/Arg for different distances between the charged groups. For DMP-Arg interactions, we find an overall good agreement between PBE0+vdW and estimates from both AMOEBABIO18 (MAE= 1.7 kcal/mol) and revised AMOEBABIO18-HFC22 (MAE= 1.9 kcal/mol) models. The inclusion of NB-fix cross-terms between phosphate oxygen and Arg guanidinium nitrogen has only a little impact on MAE. Therefore, we decided to stick to the LJ parameters computed from mixing rules. In the case of DMP-Lys interactions, the MAE is 3.2 kcal/mol for the AMOEBABIO18 model and 3.4 kcal/mol for the AMOEBABIO18-HFC22

model. The inclusion of LJ cross-terms reduces the MAE to 0.5 kcal/mol. These cross-terms are listed in **Table S7** of the supporting information.

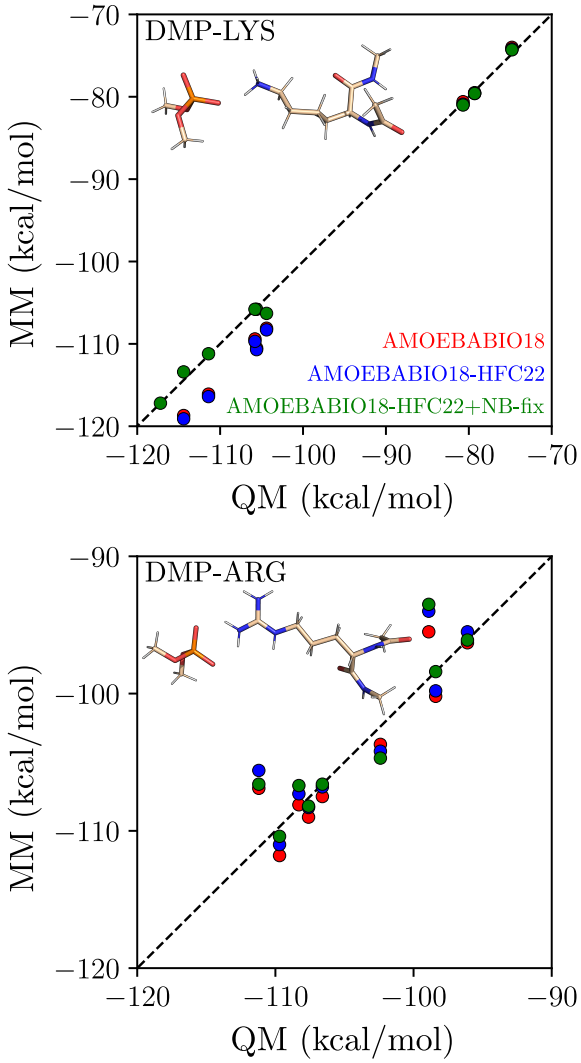

Figure 6: DMP-lysine and DMP-arginine interaction energies for 9 different dimer geometries. The dimer geometries differ from each other mainly in terms of their distances ranging between 3.4 and 5 Å.

MD simulations of Mg^{2+}·Kinase·ATP complexes

To assess the performance of our revised model, we perform MD simulations of two protein kinases, one complexed with ATP and two Mg^{2+} ions, and the other complexed with ATP and one Mg^{2+} ion. The kinase structure complexed with two Mg^{2+} ions is that of glycogen synthase kinase-3 beta (GSK3β), and the kinase structure complexed with one Mg^{2+}

ion is that of cyclin-dependent kinase 2 (CDK2). We perform 300 ns long MD simulations using both the 'original' and the 'revised' models. The 'original' model refers to the AMOEBABIO18 model with ATP parameters developed in 2021.[53] The revised model refers to the AMOEBABIO18-HFC22 model along with all the revisions made in this work.

Figure 7 shows the root-mean-square deviation (RMSD) of backbone atoms determined with respect to X-ray structures. Overall, we find that the revised model exhibits a smaller deviation relative to X-ray structures. The difference in RMSD between the two models emerges primarily from the structures of loops, which remain closer to the X-ray structure in the revised model. This is deduced quantitatively from examining RMSDs computed without including residues that belong to loops, and in these calculations, the RMSDs in both models are similar. This is also clear from visual inspection of superimposed structures shown in **Figure 7**. This is consistent with what we noted previously[36] in the case of another protein where the original model produced more flexible loops than the revised model. We had also noted previously that the revised model performed better than the original model at reproducing second-order NMR parameters.[36] Of particular interest is the Gly-rich loop in GSK3β that interacts directly with bound ATP and is known to modulate its catalytic activity.[106] We find that in both the original and the revised models, the Gly-rich loop is flexible, but in the original model, we see large deviations from the X-ray structure. In the revised model, the fluctuations are around the X-ray structure. This difference is not noted in the case of CDK2, which suggests that the differences noted in the case of GSK3β's Gly-rich loop may be due to the enhanced electrostatics from the second Mg^{2+} ion, although further investigation is necessary. Overall, the revised model performs better at reproducing the global structures of the complexes.

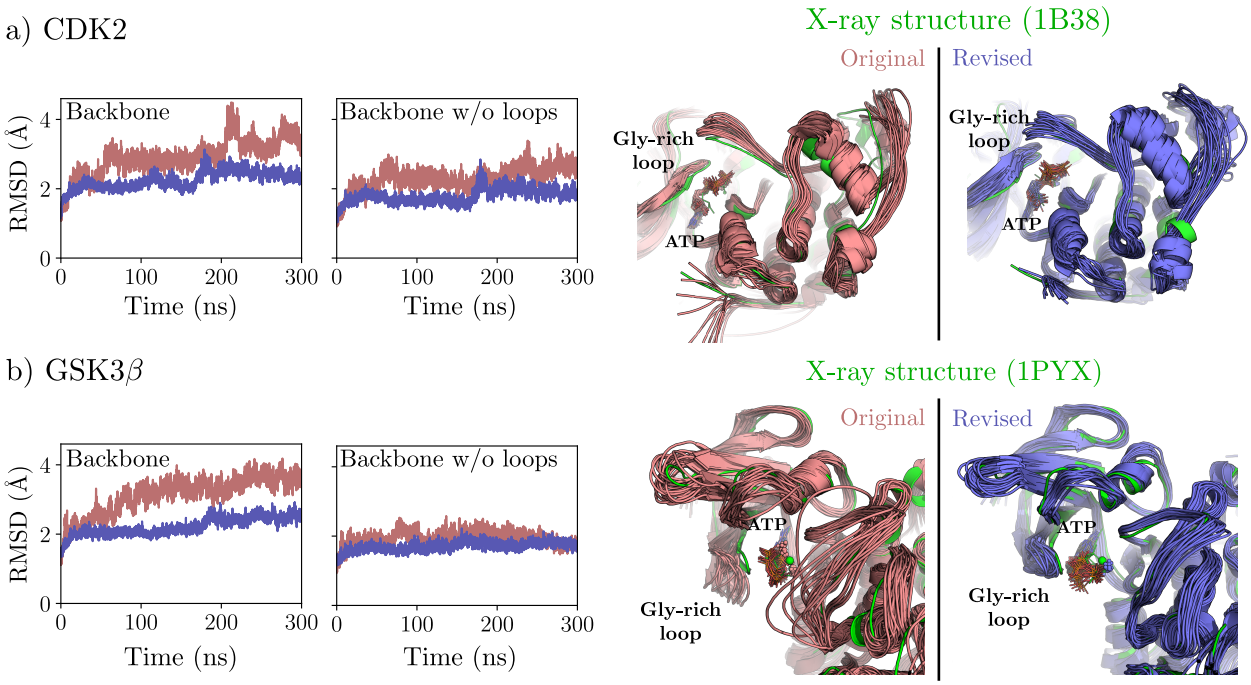

Figure 7: Effect of model revision on the global structure of (a) CDK2 and (b) GSK3β complexes. The figures on the left show backbone RMSDs with and without the inclusion of residues that belong to loops. The figures on the right show twenty equally spaced snapshots from simulations superimposed over X-ray structures (green) used for starting simulations. Mg^{2+} ions are shown as spheres.

Next, we evaluate the coordination environments of Mg^{2+} ions. **Figure 8** shows the inner shell (< 3 Å) coordination numbers of Mg^{2+}. In CDK2's X-ray structure, Mg^{2+} is resolved in the $\alpha\beta\gamma$ binding mode to ATP. Mg^{2+} is 6-fold coordinated and the remaining three coordinators are N132, D145, and one water. In the original model, Mg^{2+} continues to coordinate with N132, D145, and one water, but its binding to ATP switches to a $\alpha\gamma$ mode, dropping the coordination number from 6 to 5. In the revised model, Mg^{2+} also continues to coordinate with N132, D145, and one water. The binding to ATP also switches, but to a $\alpha\gamma\gamma/\alpha\alpha\gamma$ mode, leaving the coordination number around 6.

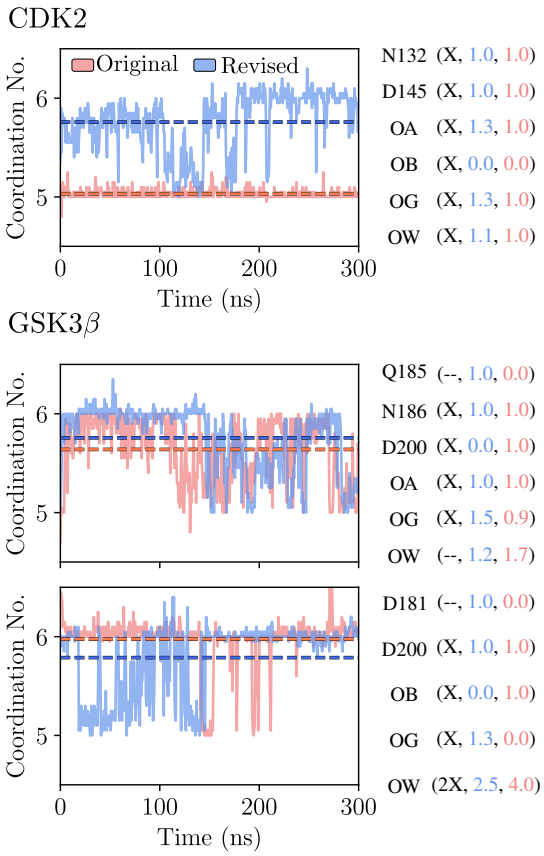

Figure 8: Coordination numbers of Mg^{2+} ions in simulations compared to X-ray structures. The figures on the left show total coordination numbers as functions of simulation time. The dashed lines show averages over the entire lengths of the simulations. The text on the right shows probabilities of Mg^{2+} coordination with different ligands. 'X' and '–' imply whether the coordination is present or absent in X-ray structures.

In GSK3β's X-ray structure, neither of the two Mg^{2+} ions are resolved in 6-fold coordinated states: one is 4-fold coordinated, and the other is 5-fold coordinated. One Mg^{2+} coordinates ATP in the $\alpha\gamma$ mode and the other Mg^{2+} coordinates ATP in the $\beta\gamma$ mode. The original model retains the $\alpha\gamma$ binding mode, but the $\beta\gamma$ switches to β mode. Both Mg^{2+} ions have average coordination numbers close to 6, and in both cases, the new coordinators that Mg^{2+} ions pick up to fill their inner shells are waters. The revised model also retains the $\alpha\gamma$ binding mode of Mg^{2+} to ATP, but the $\beta\gamma$ binding mode switches to the γ mode. Both Mg^{2+} ions also have average coordination numbers close to 6, but in contrast to the original model, the new coordinators that Mg^{2+} ions pick up to fill their inner shells are not waters, but protein groups. One Mg^{2+} picks up Q185, and the other picks up D181.

To further access the geometry of Mg^{2+} ions, we calculate the distance between the two Mg^{2+} ions. This applies only to GSK3β. In the X-ray structure, this distance is 4.1 Å. In both models, the average distance is a bit larger and around 5 Å (see **Figure S10** in the supporting information). Nevertheless, we do note that the revised model does sample short distances in the interval of [3.6-4.3] Å, which the original model does not. Overall, while both models only partially retain Mg^{2+} coordination geometries seen in X-ray structures, the revised model performs better at retaining coordination numbers.

Next, we assess polar contacts between ATP and protein. These are summarized in **Figure 9** and are computed using a protocol described in our earlier work.[107] We focus our attention on the three main parts of ATP: the nitrogen atoms from the base (N1 and N6), the hydroxyl oxygens from the sugar (O2′ and O3′), and the triphosphate oxygens (OA, OB, and OG). Overall, we note that while both models retain the contacts of adenosine base, the revised model performs discernibly better at retaining the contacts of sugar hydroxyls and the triphosphate oxygens.

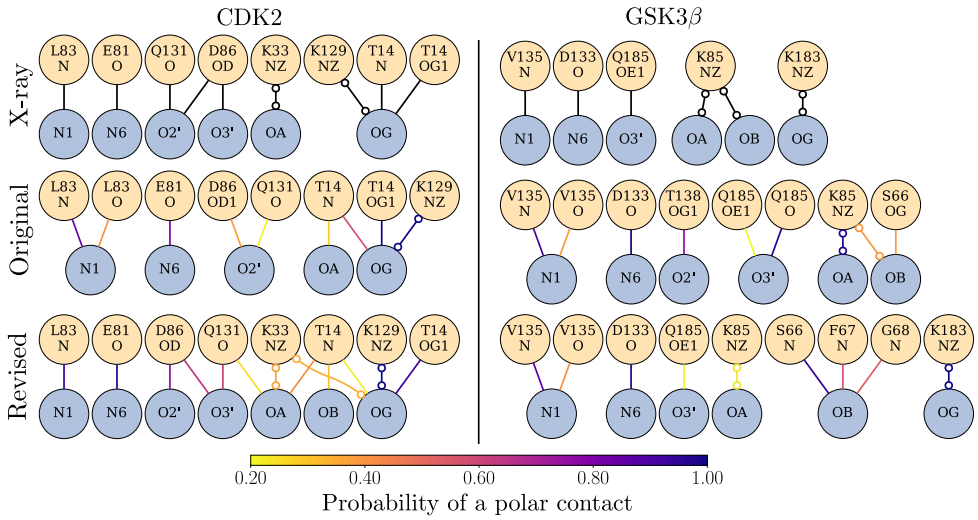

Figure 9: Probabilities of polar contacts between ATP and protein in simulations compared to those observed in X-ray structures. Lines denote contacts between donors and acceptors and the symbol 'o-o' denotes salt bridges. The ATP atoms are shown as circles filled in blue, and protein atoms are shown as circles filled in orange. The probability of a given contact is color-coded on the lines connecting the circles. Contacts with a probability lower than 0.2 are not shown.

Finally, we examine the conformation of the triphosphate group in the protein binding

site. **Figure 10** shows a 2D-histogram of the dihedral angles O3α-Pβ-O3β-Pγ and Pα-O3α-Pβ-O3β. In both complexes, our revisions to AMOEBA shift the distributions away from the centers ($\psi = \phi = 0°$) and toward larger angles. Specifically, in CDK2, the two density maxima in the original model shift from around $(\psi, \phi) = (30°, -30°), (-30°, 30°)$ to $(60°, -30°), (-90°, 30°)$. In GSK3β, a single density maximum around $(45°, -45°)$ splits into two density maxima around $(75°, -60°), (-90°, 45°)$. Comparing these 2D-histograms to **Figure 4** might suggest that the revisions to the triphosphate dihedral potentials contribute to this shift in dihedral distributions. However, note that substantial changes are also made to protein polar groups as well as to interactions between Mg^{2+}, proteins and nucleotides, and so we consider this shift in dihedral distributions to be a combined effect of all the changes introduced in the MM model. Importantly, in both cases, our revisions shift the distributions discernibly closer to the angles observed in X-ray structures. We also examine distributions in the triphosphate P_α-P_β-P_γ angle and find that both models perform equally well against the angles noted in the X-ray structures.

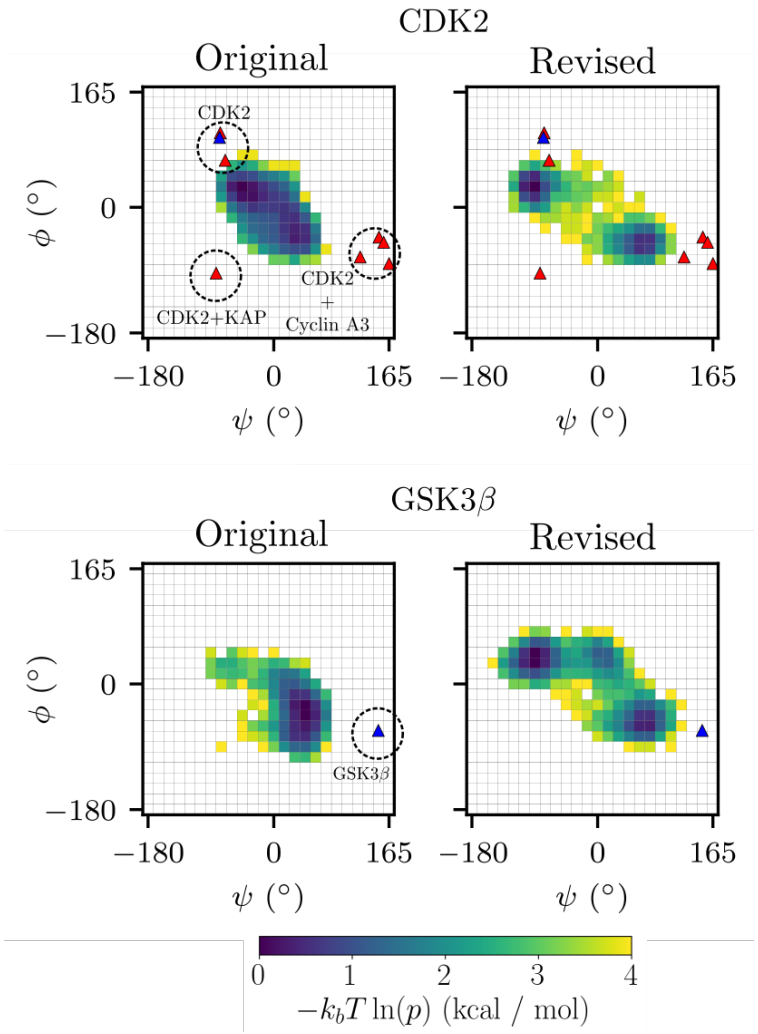

Figure 10: 2D-histogram of the triphosphate dihedral angles O3α-Pβ-O3β-Pγ (ψ) and Pα-O3α-Pβ-O3β (ϕ). The red triangles represent the dihedral angles from different X-ray structures from the PDB and the blue triangles represent the X-ray structure used for these simulations. Triangles labeled as 'CDK2+Cyclin A3' and 'CDK2+KAP' are in X-ray structures bound to Cyclin A3 and KAP, respectively.

Taken together, we note that the revised model performs better than the original model in reproducing the global X-ray structure, as well as the local geometries at catalytic sites, including Mg^{2+} coordinations, ATP coordinations, and ATP conformations. However, there are still differences between local geometries predicted by the revised model and the X-ray structure.

Conclusions

We present an improved polarizable AMOEBA model for simulating Mg^{2+}·Protein·Nucleotide complexes. We follow an approach in which we first make specific improvements to Mg^{2+}-protein, Mg^{2+}-nucleotide, and protein-nucleotide interactions, assessing them individually using benchmarked QM and experimental data. We then evaluate the performance of these improvements by carrying out MD simulations of Mg^{2+}·Kinase·ATP complexes and comparing MD structural data against X-ray structural data.

We first show that our revised AMOEBA polarizable model for proteins, AMOEBABIO18-HFC22[36] that responds better to high electric fields and was built on the strong foundations of the AMOEBABIO18 protein model, does better at predicting Mg^{2+}-protein interactions. Inclusion of many-body LJ cross-terms (MBNB-fix) further improves accuracy and brings the MAE down to 6 kcal/mol. Clearly, there is still room for improvement in Mg^{2+}-protein interactions to bring MAE down to benchmarked DFT accuracy, or to the almost DFT level accuracy of 2.7 kcal/mol that we achieved previously[36] for interactions of proteins with monovalent cations. Note that employing LJ cross-terms developed using reference data from 2-body interactions, which is a typical strategy employed in the development of polarizable and non-polarizable models, yields larger errors in many-body interactions. Conversely, employing LJ cross-terms developed using reference data from many-body interactions yields larger errors in 2-body interactions. Therefore, the many-body LJ cross-terms we provide should not be used for studying two-body Mg^{2+}-protein interactions. This also implies that there is still some essential physics that is missing in the description of these interactions.

We also present a new set of the molecular polarizability, LJ terms and torsional angles of the triphosphate group, as well as new Mg^{2+}-phosphate cross-terms that accurately predict Mg^{2+}-ATP binding free energies in the aqueous phase, without need for any a posteriori adjustments. How Mg^{2+} associates with ATP in the aqueous phase (Mg^{2+}-ATP binding modes) remains ambiguous from experiments, and there is supporting evidence from NMR that Mg^{2+} can form mono-, bi- and tri-dentate coordinations with ATP. Our revised model

predicts the existence of all these Mg^{2+}-ATP binding modes with high probabilities.

Analysis of X-ray structures of Mg^{2+}·Protein·ATP complexes reveals three prominent Mg^{2+}-ATP binding modes, $\beta\gamma$, $\alpha\beta$, and $\alpha\beta\gamma$. While the original model does not predict the existence of any of these binding modes in the aqueous phase, the revised model predicts both bidentate modes to exist in the aqueous phase, with the $\beta\gamma$-mode having a much higher probability compared to the $\alpha\beta$-mode. Since we do not observe the third $\alpha\beta\gamma$-mode in the aqueous phase, our model suggests that it perhaps results from complexation of Mg^{2+}·ATP with protein.

Finally, MD simulations of two different Mg^{2+}·Kinase·ATP complexes show that the revisions we introduce in Mg^{2+}-protein, Mg^{2+}-nucleotide, and protein-nucleotide interactions substantially improve the correspondence between MD and X-ray structures. We note improvements in both global and local catalytic site structures. In terms of local structures, we note improvements in predictions of Mg^{2+}-ATP binding modes, ATP-protein binding modes as well as ATP orientations within kinases.

However, there are still discernible differences in local catalytic site geometries between the revised model and X-ray structures. Most notable is the revised model's inability to retain $\beta\gamma$ or $\alpha\beta\gamma$ binding modes of Mg^{2+} to ATP, although it does maintain the $\alpha\gamma$ binding mode. These modes are also not retained by the original model. However, we note that these modes do not switch to binding modes that are more probable in the aqueous phase. The $\beta\gamma$ mode switches to the γ mode, although both $\beta\gamma$ and γ modes are equally likely in the aqueous phase. The tridentate $\alpha\beta\gamma$ mode remains tridentate, but fluctuates between $\alpha\gamma\gamma$ and $\alpha\alpha\gamma$ modes, although the $\alpha\alpha\gamma$ mode is not observed in the aqueous phase. This implies that the protein is strongly influencing the formation of these modes, and perhaps the errors that are still present in the Mg^{2+}-protein interactions are somehow influencing the switching of these Mg^{2+}-ATP binding modes in the catalytic site. It remains to be seen whether further improvements in Mg^{2+}-protein interactions close this gap between MD simulations and X-ray crystallography.

Acknowledgement

The authors acknowledge the use of computer time from Research Computing at USF and Hungarian HPC Infrastructure at NIIF Institute. Funding for this study was provided by the National Institute of Health (grant numbers R01 GM118697 and R01 GM147210), ERC (starting grant numbers 101076972), National Research, Development, and Innovation Office (grant numbers. FK142489 and KKP126451), and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

Supporting Information Available

The supporting information includes 8 tables and 10 figures.

References

(1) Wolf, F. I.; Torsello, A.; Fasanella, S.; Cittadini, A. Cell physiology of magnesium. *Molecular aspects of medicine* **2003**, *24*, 11–26.

(2) Romani, A. M. Cellular magnesium homeostasis. *Archives of biochemistry and biophysics* **2011**, *512*, 1–23.

(3) Romani, A. M. P. In *Encyclopedia of Metalloproteins*; Kretsinger, R. H., Uversky, V. N., Permyakov, E. A., Eds.; Springer New York: New York, NY, 2013; pp 1255–1264.

(4) Black, C.; Huang, H.-W.; Cowan, J. Biological coordination chemistry of magnesium, sodium, and potassium ions. Protein and nucleotide binding sites. *Coordination Chemistry Reviews* **1994**, *135*, 165–202.

(5) Tang, S.; Yang, J. J. In *Encyclopedia of Metalloproteins*; Kretsinger, R. H., Uver-

sky, V. N., Permyakov, E. A., Eds.; Springer New York: New York, NY, 2013; pp 1243–1250.

(6) Hartwig, A. Role of magnesium in genomic stability. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **2001**, *475*, 113–121.

(7) Le, C. T. M.; Houri, A.; Balage, N.; Smith, B. J.; Mechler, A. Interaction of small ionic species with phospholipid membranes: The role of metal coordination. *Frontiers in Materials* **2019**, *5*, 80.

(8) Dunn, M. J.; Grant, R. Red blood cell calcium and magnesium: effects upon sodium and potassium transport and cellular morphology. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1974**, *352*, 97–116.

(9) Romero, P. J. The role of membrane-bound magnesium in the permeability of ghosts to K+. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1974**, *339*, 116–125.

(10) Wolf, F. I.; Trapani, V. Cell (patho) physiology of magnesium. *Clinical science* **2008**, *114*, 27–35.

(11) Ponder, J. W.; Case, D. A. *Protein Simulations*; Advances in Protein Chemistry; Academic Press, 2003; Vol. 66; pp 27–85.

(12) 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.

(13) Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *Journal of physics. Condensed matter : an Institute of Physics journal* **2009**, *21*, 333102–333102.

(14) Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. Classical Electrostatics for Biomolecular Simulations. *Chemical Reviews* **2014**, *114*, 779–814.

(15) Baker, C. M. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2015**, *5*, 241–254.

(16) Lopes, P. E. M.; Guvench, O.; MacKerell, J., Alexander D Current status of protein force fields for molecular dynamics simulations. *Methods in molecular biology (Clifton, N.J.)* **2015**, *1215*, 47–71.

(17) Lemkul, J. A.; 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. *Chemical reviews* **2016**, *116*, 4983–5013.

(18) Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Current Opinion in Structural Biology* **2018**, *49*, 129–138.

(19) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual review of biophysics* **2019**, *48*, 371–394.

(20) Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P. Simulation of Ca2+ and Mg2+ solvation using polarizable atomic multipole potential. *The journal of physical chemistry B* **2006**, *110*, 18553–18559.

(21) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell Jr, A. D.; Roux, B. Simulating monovalent and divalent ions in aqueous solution using a Drude polarizable force field. *Journal of chemical theory and computation* **2010**, *6*, 774–786.

(22) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz Jr, K. M. Rational design of particle mesh Ewald compatible Lennard-Jones parameters for+ 2 metal cations in explicit solvent. *Journal of chemical theory and computation* **2013**, *9*, 2733–2748.

(23) Grotz, K. K.; Cruz-León, S.; Schwierz, N. Optimized magnesium force field parameters for biomolecular simulations with accurate solvation, ion-binding, and water-exchange properties. *Journal of chemical theory and computation* **2021**, *17*, 2530–2540.

(24) Allnér, O.; Nilsson, L.; Villa, A. Magnesium ion–water coordination and exchange in biomolecular simulations. *Journal of chemical theory and computation* **2012**, *8*, 1493–1502.

(25) 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.

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

(27) 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).

(28) 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.

(29) 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.

(30) 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.

(31) 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.

(32) 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.

(33) 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.

(34) Wineman-Fisher, V.; Al-Hamdani, Y.; Nagy, P. R.; Tkatchenko, A.; Varma, S. Improved description of ligand polarization enhances transferability of ion–ligand interactions. *The Journal of Chemical Physics* **2020**, *153*, 094115.

(35) Wineman-Fisher, V.; Delgado, J. M.; Nagy, P. R.; Jakobsson, E.; Pandit, S. A.; Varma, S. Transferable interactions of Li+ and Mg2+ ions in polarizable models. *The Journal of Chemical Physics* **2020**, *153*, 104113.

(36) Delgado, J. A.; Wineman-Fisher, V.; Pandit, S.; Varma, S. Inclusion of High-Field Target Data in AMOEBA's Calibration Improves Predictions of Protein–Ion Interactions. *Journal of Chemical Information and Modeling* **2022**, *62*, 4713–4726.

(37) 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.

(38) 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).

(39) 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).

(40) 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.

(41) 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).

(42) Goldberg, R. N.; Tewari, Y. B. Thermodynamics of the disproportionation of adenosine 5'-diphosphate to adenosine 5'-triphosphate and adenosine 5'-monophosphate: I. Equilibrium model. *Biophysical chemistry* **1991**, *40*, 241–261.

(43) Saris, N.-E. L.; Mervaala, E.; Karppanen, H.; Khawaja, J. A.; Lewenstam, A. Magnesium: an update on physiological, clinical and analytical aspects. *Clinica chimica acta* **2000**, *294*, 1–26.

(44) Cohn, M.; Hughes Jr, T. R. Nuclear magnetic resonance spectra of adenosine di-and triphosphate: II. Effect of complexing with divalent metal ions. *Journal of Biological Chemistry* **1962**, *237*, 176–181.

(45) Tran-Dinh, S.; Neumann, J. A 31P-NMR study of the interaction of Mg2+ ions with nucleoside diphosphates. *Nucleic Acids Research* **1977**, *4*, 397–403.

(46) Ramirez, F.; Marecek, J. F. Coordination of magnesium with adenosine 5'-diphosphate and triphosphate. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1980**, *589*, 21–29.

(47) Takeuchi, H.; Murata, H.; Harada, I. Interaction of adenosine 5'-triphosphate with Mg2+: vibrational study of coordination sites by use of 18O-labeled triphosphates. *Journal of the American Chemical Society* **1988**, *110*, 392–397.

(48) Kuntz, G.; Swift, T. Contrasting structures of magnesium and calcium adenosine-triphosphate complexes as studied by nuclear relaxation. FEDERATION PROCEEDINGS. 1973; p 546.

(49) Bishop, E. O.; Kimber, S. J.; Orchard, D.; Smith, B. E. A 31P-NMR study of mono-and dimagnesium complexes of adenosine 5'-triphosphate and model systems. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* **1981**, *635*, 63–72.

(50) Dudev, T.; Grauffel, C.; Lim, C. How native and alien metal cations bind ATP: Implications for lithium as a therapeutic agent. *Scientific reports* **2017**, *7*, 42377.

(51) Komuro, Y.; Re, S.; Kobayashi, C.; Muneyuki, E.; Sugita, Y. CHARMM force-fields with modified polyphosphate parameters allow stable simulation of the ATP-bound structure of Ca^{2+}-ATPase. *Journal of Chemical Theory and Computation* **2014**, *10*, 4133–4142.

(52) Buelens, F. P.; Leonov, H.; de Groot, B. L.; Grubmüller, H. ATP–Magnesium Coordination: Protein Structure-Based Force Field Evaluation and Corrections. *Journal of Chemical Theory and Computation* **2021**, *17*, 1922–1930.

(53) Walker, B.; Jing, Z.; Ren, P. Molecular dynamics free energy simulations of ATP: Mg2+ and ADP: Mg2+ using the polarisable force field AMOEBA. *Molecular simulation* **2021**, *47*, 439–448.

(54) Ren, P.; Ponder, J. W. Polarizable atomic multipole water model for molecular mechanics simulation. *The Journal of Physical Chemistry B* **2003**, *107*, 5933–5947.

(55) 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.

(56) Ren, P.; Ponder, J. W. Temperature and pressure dependence of the AMOEBA water model. *The Journal of Physical Chemistry B* **2004**, *108*, 13427–13437.

(57) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A. Current status of the AMOEBA polarizable force field. *The journal of physical chemistry B* **2010**, *114*, 2549–2564.

(58) 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.

(59) 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.

(60) 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. *Journal of Molecular Biology* **2021**, *433*, 166745.

(61) 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. *Chemistry–A European Journal* **2021**,

(62) Perdew, J. P.; Burke, K.; Ernzerhof, M. Gradient Approximation Made Simple. *Phys Rev Lett.* **1996**, *77*, 3865–3868.

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

(64) 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.

(65) 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.

(66) DiLabio, G. A.; Otero-de-la Roza, A. Noncovalent interactions in density functional theory. *Reviews in computational chemistry* **2016**, *29*, 1–97.

(67) 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, PMID: 26598340.

(68) 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.

(69) 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.

(70) 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.

(71) Gyevi-Nagy, L.; Kállay, M.; Nagy, P. R. Integral-direct and parallel implementation of the CCSD(T) method: Algorithmic developments and large-scale applications. *J. Chem. Theory Comput.* **2020**, *16*, 336.

(72) Kállay, M. et al. The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **2020**, *152*, 074107.

(73) Kállay, M. et al. MRCC, *a quantum chemical program suite. See https://www.mrcc.hu/* **Accessed August 1, 2023**,

(74) 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.

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

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

(77) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Physical review* **1934**, *46*, 618.

(78) Frisch, M. J. et al. Gaussian 09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

(79) Dunning Jr, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of chemical physics* **1989**, *90*, 1007–1023.

(80) Hickey, A. L.; Rowley, C. N. Benchmarking quantum chemical methods for the calculation of molecular dipole moments and polarizabilities. *The Journal of Physical Chemistry A* **2014**, *118*, 3678–3687.

(81) Wang, Z.; Ponder, J. W. Tinker9: Next Generation of Tinker with GPU Support. 2021; `https://github.com/TinkerTools/tinker9`.

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

(83) 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.

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

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

(86) Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids*; Oxford university press, 2017.

(87) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: software tools for molecular design. *Journal of chemical theory and computation* **2018**, *14*, 5273–5289.

(88) Bertrand, J.; Thieffine, S.; Vulpetti, A.; Cristiani, C.; Valsasina, B.; Knapp, S.; Kalisz, H.; Flocco, M. Structural characterization of the GSK-3β active site using selective and non-selective ATP-mimetic inhibitors. *Journal of molecular biology* **2003**, *333*, 393–407.

(89) Brown, N. R.; Noble, M. E.; Lawrie, A. M.; Morris, M. C.; Tunnah, P.; Divita, G.; Johnson, L. N.; Endicott, J. A. Effects of phosphorylation of threonine 160 on cyclin-

dependent kinase 2 structure and activity. *Journal of Biological Chemistry* **1999**, *274*, 8746–8756.

(90) Šali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology* **1993**, *234*, 779–815.

(91) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic acids research* **2004**, *32*, W665–W667.

(92) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophysical Journal* **1998**, *72*, 1047–1069.

(93) Hamelberg, D.; McCammon, J. A. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: double-decoupling method. *Journal of the American Chemical Society* **2004**, *126*, 7683–7689.

(94) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics* **1976**, *22*, 245–268.

(95) Fogarty, J. C.; Chiu, S.-W.; Kirby, P.; Jakobsson, E.; Pandit, S. A. Automated optimization of water-water interaction parameters for a coarse-grained model. *The Journal of Physical Chemistry B* **2014**, *118*, 1603–1611.

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

(97) Varma, S.; Rempe, S. Structural transitions in ion coordination driven by changes in competition for ligand binding. *Journal of American Chemical Society* **2010**, *130*, 14505–14519.

(98) Dudev, T.; Lin,; Dudev, M.; Lim, C. First- second shell interactions in metal binding sites in proteins: A PDB survey and DFT/CDM calculations. *Journal of the American Chemical Society* **2003**, *125*, 3168–3180.

(99) Ohtaki, H.; Radnai, T. Structure and dynamics of hydrated ions. *Chemical reviews* **1993**, *93*, 1157–1204.

(100) 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.

(101) 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.

(102) Foloppe, N.; MacKerell, A. D., Jr All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *Journal of computational chemistry* **2000**, *21*, 86–104.

(103) Pastor, R.; MacKerell Jr, A. Development of the CHARMM force field for lipids. *The journal of physical chemistry letters* **2011**, *2*, 1526–1532.

(104) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society* **1995**, *117*, 5179–5197.

(105) Zhang, C.; Lu, C.; Wang, Q.; Ponder, J. W.; Ren, P. Polarizable multipole-based force field for dimethyl and trimethyl phosphate. *Journal of chemical theory and computation* **2015**, *11*, 5326–5339.

(106) Bossemeyer, D. The glycine-rich sequence of protein kinases: a multifunctional element. *Trends in biochemical sciences* **1994**, *19*, 201–205.

(107) Delgado, J. A.; Duro, N.; Rogers, D. M.; Tkatchenko, A.; Pandit, S.; Varma, S. Molecular basis for higher affinity of SARS-CoV-2 spike RBD for human ACE2 receptor. *Proteins* **2021**, *89*, 1134–1144.