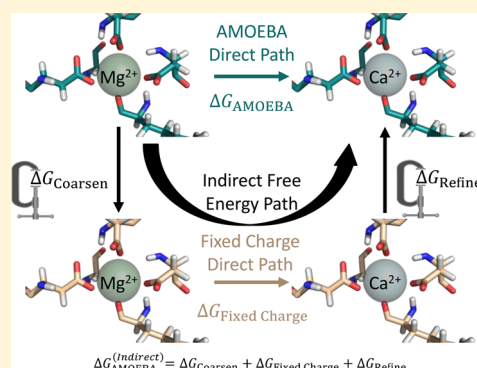


Scalable Indirect Free Energy Method Applied to Divalent Cation-Metalloprotein Binding

Jacob Litman,[†] Andrew C. Thiel,[‡] and Michael J. Schnieders^{*,†,‡}[†]University of Iowa, Department of Biochemistry, 51 Newton Road, 4-403 Bowen Science Building, Iowa City, Iowa 52242, United States[‡]University of Iowa, Department of Biomedical Engineering, 103 South Capitol Street, 5601 Seamans Center for the Engineering Arts and Sciences, Iowa City, Iowa 52242, United States

Supporting Information

ABSTRACT: Many biological processes are based on molecular recognition between highly charged molecules such as nucleic acids, inorganic ions, charged amino acids, etc. For such cases, it has been demonstrated that molecular simulations with fixed partial charges often fail to achieve experimental accuracy. Although incorporation of more advanced electrostatic models (such as multipoles, mutual polarization, etc.) can significantly improve simulation accuracy, it increases computational expense by a factor of 5–20×. Indirect free energy (IFE) methods can mitigate this cost by modeling intermediate states at fixed-charge resolution. For example, an efficient “reference” model such as a pairwise Amber, CHARMM, or OPLS-AA force field can be used to derive an initial estimate, followed by thermodynamic corrections to a more advanced “target” potential such as the polarizable AMOEBA model. Unfortunately, all currently described IFE methods encounter difficulties reweighting more than ~50 atoms between resolutions due to extensive scaling of both the magnitude of the thermodynamic corrections and their statistical uncertainty. We present an approach called “simultaneous bookending” (SB) that is fundamentally different from existing IFE methods based on a tunable sampling approximation, which permits scaling to thousands of atoms. SB is demonstrated on the relative binding affinity of $\text{Mg}^{2+}/\text{Ca}^{2+}$ to a set of metalloproteins with up to 2972 atoms, finding no statistically significant difference between direct AMOEBA results and those from correcting Amber to AMOEBA. The ability to change the resolution of thousands of atoms during reweighting suggests the approach may be applicable in the future to protein–protein binding affinities or nucleic acid thermodynamics.



INTRODUCTION

Cellular processes rely on a variety of charged compounds for their functions, including phosphorylated proteins,¹ phospholipids,² metal cations,³ and nucleic acids.⁴ The electric fields from these charged compounds can induce large polarization responses in nearby atoms. The mathematical models (i.e., molecular mechanics force fields) used to represent these systems in simulations must account for this polarization response. Although well-established fixed partial charge force fields are often designed to implicitly include a polarization response consistent with solvation in water, this approximation can produce significant errors for inhomogeneous electrostatic environments with strong fields.^{5–7} For example, metal ion-binding pockets coordinate compact, highly charged cations that produce large electrostatic fields within low-dielectric amino acid pockets. Recent simulations with an explicitly polarizable force field have accurately predicted the relative Ca^{2+} vs Mg^{2+} binding for a set of six divalent cation-binding metalloproteins, overcoming demonstrated limitations in standard fixed charge models.^{8,9} While alterations to fixed-

charge force fields have seen some success,^{10,11} these pairwise approaches have been shown to break down in some cases.^{12,13}

To address limitations in fixed partial charge force fields, more advanced models are emerging that include many-body polarization and/or atomic multipoles, including the AMOEBA model,^{7,14,15} which improves transferability across chemical environments and accuracy.^{6,16,17} However, current implementations of AMOEBA are 5–20× slower than standard fixed partial charge force fields.^{18,19} In the context of biological macromolecules, this can be a significant limitation, as even simulations based on pairwise fixed partial charge force fields can struggle to reach time scales key to protein functions.²⁰ For this reason, direct free energy simulations (i.e., a thermodynamic path defined by a single force field resolution) have often needed to sacrifice either accuracy or precision to achieve an affordable simulation protocol. To overcome this trade-off, so-called “indirect” free energy (IFE) methods were described in 1992 by Gao²¹ and

Received: February 25, 2019

Published: June 20, 2019

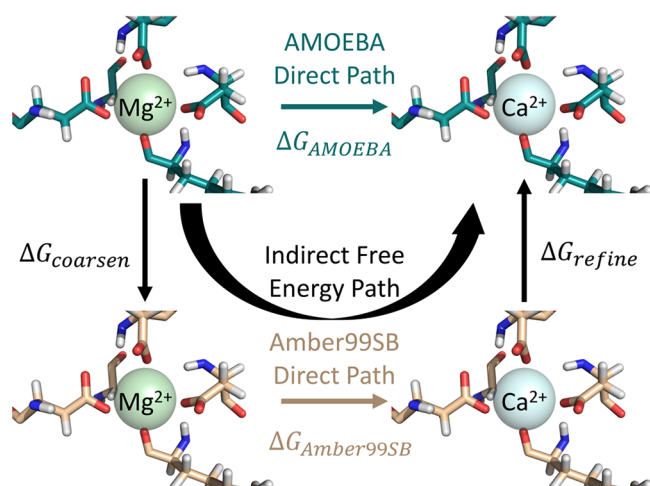


Figure 1. Thermodynamic path for indirect free energy calculations. The direct paths (horizontal legs) replace Mg^{2+} with Ca^{2+} in the binding pocket of 1B8L (triple mutant carp parvalbumin), and the indirect path corrects the Amber99SB (yellow-brown) direct path to the same result as the AMOEBA (green-blue) direct path. The focus of this work is on convergence of the indirect path free energy differences (i.e., the sum of the coarsen and refine vertical legs).

the Warshel group²² in an effort to combine the computational efficiency of fixed partial charge force fields with the accuracy of more advanced potential functions. Typically, these have been either a purely quantum mechanical (QM) potential or a hybrid quantum mechanics/molecular mechanics (QM/MM) target, though we have used a polarizable MM target potential in the past.⁶ This is illustrated by the thermodynamic cycle shown in Figure 1 for the specific case of the relative binding of $\text{Ca}^{2+}/\text{Mg}^{2+}$ to a metalloprotein. The relative binding affinity computed using an efficient potential $\Delta G_{\text{Amber99SB}}$ can be corrected to that of the more accurate potential ΔG_{AMOEBA} if the $\Delta G_{\text{Coarsen}}$ and ΔG_{Refine} indirect legs can be rigorously sampled (eq 1).

$$\Delta G_{\text{AMOEBA}}^{(\text{Indirect})} = \Delta G_{\text{Coarsen}} + \Delta G_{\text{Amber99SB}} + \Delta G_{\text{Refine}} \quad (1)$$

Ideally, IFE methods permit the bulk of the thermodynamic sampling to be performed using the efficient potential, followed by sampling of the correction legs to the more accurate target potential. For IFE methods to be superior to a simpler, direct usage of the target potential, the sum of the reference path (Figure 1, bottom leg) and indirect corrections (Figure 1, left/right legs) must be substantially faster than the direct target path (Figure 1, top leg) while achieving a similar uncertainty.

Unfortunately, changing the resolution of the potential for all atoms in a fully unconstrained fashion becomes intractable for large systems. Thus, prior IFE methods have either focused on small systems or refined the potential for only a subset of the system (e.g., fixed-charge MM to a hybrid QM/MM potential with a small QM region). Over the last three decades, progressively more atoms have been included in the coarsen and refine legs by exploring the use of a variety of advanced sampling methods (Figure 2 and Table 1). Early approaches kept the number of evaluations under the target potential very low^{21–23} by using well-separated snapshots from the reference simulation and single-step free energy perturbation (FEP).²⁴ Although this facilitates use of sophisticated target potentials (e.g., QM), its requirement for significant phase space overlap

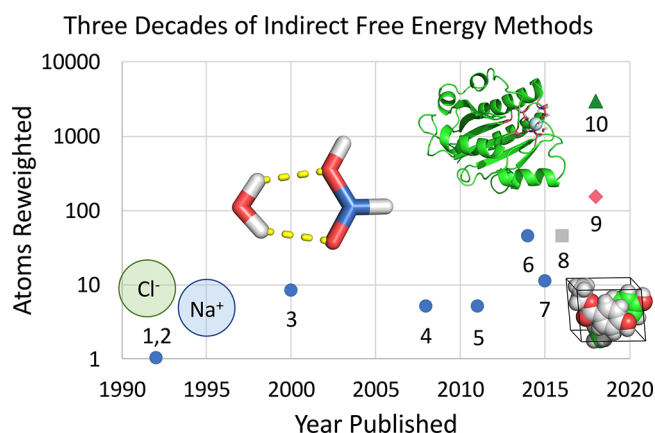


Figure 2. Scale of systems studied by indirect free energy methods and its growth over time. More details on these nine studies are given in Table 1. Blue circles are others' work (#1–7), and the gray square (#8) is our prior work.¹⁶ The red diamond (#9, unconstrained subset of atoms) and green triangle (#10, all atoms whose force field resolution is changed) are this work. A version of this figure with the Y-axis on a linear scale is available in Figure S2.

between reference and target potentials prevents reweighting more than a few dozen atoms.²⁵ To mitigate this limitation to some degree, more sophisticated statistical estimators such as multistate Bennett acceptance ratio (MBAR)²⁶ and non-Boltzmann Bennett (NBB)²⁷ have been employed to transform up to 44 atoms to a QM target.^{25,28,29} There is also an approach based on hybrid Monte Carlo (MC), wherein molecular dynamics (MD) under the reference potential is used to propose trial moves for Metropolis Monte Carlo³⁰ evaluated using the target potential. This approach has been applied to systems of at least 11 atoms^{31,32} but still suffers due to limited phase space overlap.

Recently we presented an alternative approach that samples a *dual force field* (DFF) potential defined via interpolation between reference and target potentials.¹⁶ Although this overcomes lack of phase space overlap by defining a smooth path between resolutions, it also requires a significant number of energy evaluations under the more expensive target potential. This may limit the choice of target potential to advanced classical models¹⁶ or approximate QM methods.^{33,34} Nevertheless, prior DFF corrections in the context of absolute crystal thermodynamics resulted in a 95% reduction in condensed-phase sampling of the target polarizable AMOEBA potential by using the fixed charge OPLS-AA model to sample vacuum to solid phase transitions.¹⁶

This work builds on the DFF approach to describe an IFE method called *simultaneous bookending* (SB). Whereas the DFF method remains limited by extensive scaling of the uncertainty in $\Delta G_{\text{coarsen}}$ and ΔG_{refine} with system size, SB is shown to dramatically reduce statistical uncertainty by converging the sum $\Delta G_{\text{coarsen}} + \Delta G_{\text{refine}}$ *simultaneously* during a single simulation. By introducing a tunable sampling approximation, SB represents a fundamentally different approach to tackling the inherent challenge of IFE methods by improving convergence of the sum $\Delta G_{\text{coarsen}} + \Delta G_{\text{refine}}$ while transforming large systems in their entirety to the target potential. The method is demonstrated on a series of six divalent cation-binding metalloproteins^{9,35–40} of up to 2972 atoms (Figure 3, Table 2) by recapitulating a direct path for relative Mg^{2+} versus Ca^{2+} binding affinity.

Table 1. Prior Indirect Free Energy Methods^a

ID	Largest System Studied	Atoms Reweighted	Method
1	Cl [−] Solvation	1	Monte Carlo Free Energy Perturbation ²¹
2	Na ⁺ Solvation	1	Free Energy Perturbation ²²
3	Water-Formic Acid Dimer	8	Molecular Mechanics Based Importance Function ³¹
4	Water-Methane Relative Solvation	5	Replica exchange Monte Carlo Molecular Dynamics with Thermodynamic Integration ³⁴
5	Water-Methane Relative Solvation	5	Free Energy Perturbation ²³
6	Amitriptyline Solvation	44	Non-Boltzmann Bennett ²⁸
7	Ethane Solvation	11	Hybrid Monte Carlo and Free Energy Perturbation ³²
8	Ethylparaben $\Delta G_{\text{deposition}}$	44	Dual Force Field ¹⁶
9	Calbindin D _{9k} Ca ²⁺ vs Mg ²⁺ selectivity	154	Simultaneous Bookending: Number of atoms whose force field resolution changed, and were not constrained between IFE legs (This work)
10	CD11a I-domain Ca ²⁺ vs Mg ²⁺ selectivity	2972	Simultaneous Bookending: Total number of atoms whose force field resolution changed, some of which were constrained between IFE legs (This work)

^aIndirect free energy (*i.e.* bookending) methods from Figure 2, with a description of the largest system studied and method used. For ID 8, the asymmetric unit of ethylparaben contains two 22-atom molecules. ID 9 refers to just the “unconstrained” atoms (see Theory), whereas ID 10 refers to full system size. For both IDs 9 and 10, all atoms are reweighted to the target potential, using a novel sampling-based approximation that constrains atoms between the coarsen and refine legs. IDs 1, 2, and 4–7 use potential-based approximations. IDs 3 and 8 have no IFE approximation.

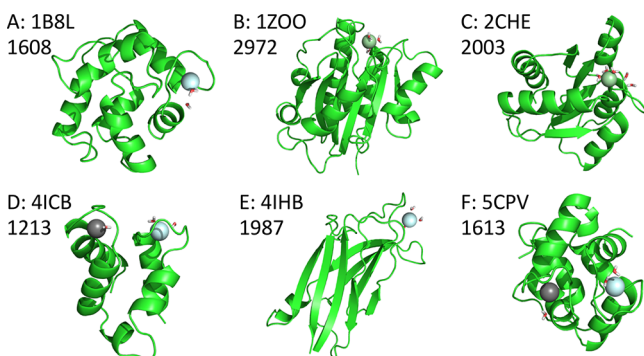


Figure 3. All structures studied in this work (Table 2). Background (nonmutated) Ca²⁺ ions are in gray, alchemical Ca²⁺ ions are in pale cyan, and alchemical Mg²⁺ ions are in pale green. Each structure is listed with the number of atoms used in simulation. A: 1B8L. B: 1ZOO. C: 2CHE. D: 4ICB. E: 4IHB. F: 5CPV.

Table 2. Metalloproteins Studied^a

PDB ID (reference)	Protein	Binding Domain	Specificity (kcal/mol)
1B8L ³⁵	Parvalbumin β Triple Mutant	EF Hand	−1.6
1ZOO ³⁶	CD11A	I Domain	1.7
2CHE ³⁷	CheY	CheY	−0.5
4ICB ³⁸	Calbindin D _{9k}	EF Hand	−6.2
4IHB ³⁹	Dysferlin C2A	C2	−1.7
5CPV ^{35,40}	Parvalbumin β	EF Hand	−5.6

^aThe metalloproteins studied in this work are described, as well as Ca²⁺ specificity (defined as $\Delta G_{\text{binding, Ca}^{2+}} - \Delta G_{\text{binding, Mg}^{2+}}$). For 5CPV, the Ca²⁺ value was taken from Cates *et al.* 1999³⁵ and rounded to −11 kcal/mol, while the Mg²⁺ value was taken from Henzl *et al.* 2003,⁴⁰ consistent with prior simulation work.⁹

THEORY

Direct Free Energy Difference. Alchemical sampling of the free energy difference between bound Ca²⁺ and bound Mg²⁺, with potential energy defined by $U_{\text{Ca}}(\mathbf{x})$ and $U_{\text{Mg}}(\mathbf{x})$, respectively, employed the following target function

$$U_{\text{Mg} \rightarrow \text{Ca}}(\mathbf{x}, \lambda) = f(\lambda) \cdot U_{\text{Ca}}(\mathbf{x}) + f(1 - \lambda) \cdot U_{\text{Mg}}(\mathbf{x}) \quad (2)$$

This relies on a state variable λ and a switching function $f(\lambda)$ to interpolate between the Ca²⁺ bound system and the Mg²⁺ bound system. Free energy differences are computed by integrating $\langle \partial U / \partial \lambda \rangle$ over the λ path between resolutions based on an enhanced sampling method^{41,42} (see the sampling strategy in the Methods for more details).

Use of a soft core treatment of van der Waals interaction is unnecessary in this case, which simplified use of GPU acceleration via Force Field X-OpenMM (FFX-OpenMM). The AMOEBA transformations used a simple linear switch $f(\lambda) = \lambda$, while the Amber99SB transformation used a trigonometric switch $f(\lambda) = \sin^2(\pi/2 \cdot \lambda)$. This was due to a large change in Åqvist van der Waals radii⁴³ between Mg²⁺ and Ca²⁺, with no compensatory change in polarizability (Table S1). This isolates most of the cost of enlarging the Amber99SB/Åqvist ion-binding pocket to the first segment of the path, which is softened by the trigonometric switch.

Indirect Free Energy Correction from Two Independent Bookending Simulations. In principle, correction of fixed charge (FC) relative binding thermodynamics to quantities consistent with the AMOEBA force field can be accomplished using two DFF simulations with potential energy functions defined by

$$U_{\text{DFF}}^{(\text{Coarsen})}(\mathbf{x}_C, \lambda) = f(1 - \lambda) \cdot U_{\text{Mg, AMOEBA}}(\mathbf{x}_C) + f(\lambda) \cdot U_{\text{Mg, FC}}(\mathbf{x}_C) \quad (3)$$

$$U_{\text{DFF}}^{(\text{Refine})}(\mathbf{x}_R, \lambda) = f(1 - \lambda) \cdot U_{\text{Ca, FC}}(\mathbf{x}_R) + f(\lambda) \cdot U_{\text{Ca, AMOEBA}}(\mathbf{x}_R) \quad (4)$$

This work uses the same trigonometric $f(\lambda) = \sin^2(\pi/2 \cdot \lambda)$ switch as above, which helps to soften the large $\partial U / \partial \lambda$ partial derivative exhibited at the ends of the DFF path (*i.e.*, near $\lambda = 0$ and $\lambda = 1$). However, the Results section demonstrates the extensive scaling of DFF statistical uncertainty, which becomes increasingly intractable as system size grows. This observation motivated the *simultaneous bookending* method described below.

Indirect Free Energy Correction from a Single Simultaneous Bookending Simulation. For many scenarios, the chemical systems for the “coarsen” and “refine” steps are composed almost entirely of chemically equivalent atoms, differing only due to alchemical modifications (such as Ca^{2+} to Mg^{2+} , an amino acid mutation, change of ligand, etc.). This observation can be leveraged to massively reduce the sampling difficulty of both DFF steps by combining them into a single “simultaneous bookending” (SB) simulation with total potential energy

$$U_{\text{SB}}(\mathbf{x}_C, \mathbf{x}_R, \mathbf{x}_S, \lambda) = U_{\text{DFF}}^{(\text{Coarsen})}(\mathbf{x}_C, \mathbf{x}_S, \lambda) + U_{\text{DFF}}^{(\text{Refine})}(\mathbf{x}_R, \mathbf{x}_S, \lambda) \quad (5)$$

where a tunable set of atomic coordinates distal from the alchemical site are shared ($\mathbf{x}_S$) throughout the simulation and the independent “coarsen” ($\mathbf{x}_C$) and “refine” ($\mathbf{x}_R$) coordinates are reduced to a neighborhood of atoms centered on the alchemical modification. We refer to atoms assigned to the shared coordinate set $\mathbf{x}_S$ as atoms with “constrained” or “shared” coordinates.

In the limit of sharing no coordinates ($\mathbf{x}_S$ is empty, $\mathbf{x}_C = \mathbf{x}_R$ and $\mathbf{x}_R = \mathbf{x}_R$), the SB simulation is unconstrained and equivalent to the two step DFF approach. On the other hand, if the number of degrees of freedom contained in $\mathbf{x}_C$ and $\mathbf{x}_R$ includes only alchemical atoms, then the SB simulation enforces equivalent sampling of the environment (i.e., all nonalchemical atoms) for both chemical end states. This latter limit is clearly an approximation, which can be appreciated by considering that the coordination of Ca^{2+} and Mg^{2+} ions is distinct (i.e., the distance from Mg^{2+} to coordinating oxygen atoms is shorter than when coordinating Ca^{2+}). This “constrained phase space approximation” is discussed in more detail in the SI text. Thus, the SB approach should only constrain coordinates distal from the region of the system undergoing chemical modifications. From an alchemical perspective, all energy terms that depend only on shared coordinates $\mathbf{x}_S$ do not contribute to the partial derivative of the SB potential energy with respect to λ or the ensemble average thermodynamic force $\langle \partial U_{\text{SB}} / \partial \lambda \rangle$ (see SI text for a detailed derivation).

$$\begin{aligned} \partial U_{\text{SB}}(\mathbf{x}_C, \mathbf{x}_R, \mathbf{x}_S, \lambda) / \partial \lambda \\ = \partial U_{\text{DFF}}^{(\text{Coarsen})}(\mathbf{x}_C, \mathbf{x}_S, \lambda) / \partial \lambda + \partial U_{\text{DFF}}^{(\text{Refine})}(\mathbf{x}_R, \mathbf{x}_S, \lambda) / \partial \lambda \end{aligned} \quad (6)$$

For example, the partial derivative with respect to λ of bonded forces fluctuates significantly in DFF simulations because two arbitrary force fields (e.g., Amber99SB and AMOEBA) differ in their functional forms, equilibrium values, and/or force constants. However, in a SB simulation, all bonded (and nonbonded) terms defined only by shared degrees of freedom ($\mathbf{x}_S$) are included in both the coarsening and refining transformations and thereby have equal magnitude, have opposite sign, and cancel. Without constrained coordinates (i.e., direct use of two DFF simulations), the instantaneous contributions of these terms to $\partial U_{\text{SB}} / \partial \lambda$ fluctuate and must be sampled in both the “coarsen” and “refine” directions. The key insight motivating the SB approach is that the ensemble-average contribution to free energy differences of energy terms distal to an alchemical transformation approaches zero. By constraining coordinates, these instantaneous contributions to the thermodynamic force are zero at every time step, and thus no sampling is needed to

remove their contribution. To minimize the approximation caused by constraining coordinates, parts of the system that do not have (nearly) identical ensembles between end-states must be assigned unique degrees of freedom. This work adopts the heuristic of only constraining coordinates for residues and molecules that do not have an atom within 5 Å of the ion. Due to the mobility of solvent molecules (i.e., the freedom to diffuse toward or away from the alchemical region), implicit solvent is particularly attractive for initial demonstrations of the SB approach.

This is illustrated with a simple SB simulation of a constrained divalent cation ($\text{Ca}^{2+}/\text{Mg}^{2+}$) surrounded by three constrained water molecules and three unconstrained pairs of water molecules (Figure 4, Movie S1). The three

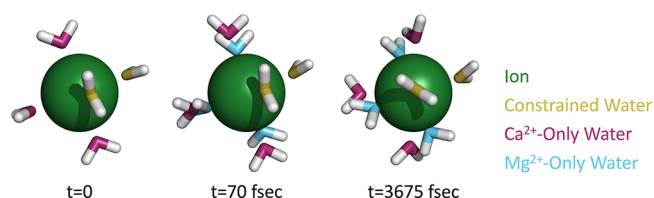


Figure 4. Illustration of simultaneous bookending (SB) using a divalent metal cation (Ca^{2+} or Mg^{2+}) surrounded by six water molecules, half of which are constrained. The ion is in green, constrained waters are in yellow, unconstrained Ca^{2+} waters are in purple, and unconstrained Mg^{2+} waters are in cyan. In the first frame, the unconstrained Mg^{2+} waters are not visible, as both topologies are initialized from identical coordinates (i.e., the Mg^{2+} waters are “hidden” inside the Ca^{2+} waters). The constrained water molecules feel both ions, while the unconstrained water molecules only (directly) feel one ion. The time step was 0.35 fs, the tempering parameter was $2.0k_B T$, and the bias magnitude was 0.001 kcal/mol. This is animated in Movie S1.

constrained water molecules (in yellow) maintain a consensus distance around the ion, intermediate between a Ca^{2+} distance and an Mg^{2+} distance. Despite being initialized with identical coordinates, the unconstrained water molecule pairs immediately separate (and would eventually exchange with each other and the constrained water molecules), with the Ca^{2+} -specific copies (purple) maintaining a greater distance from the ion than the Mg^{2+} -specific copies (cyan). The bonded terms in the constrained water molecules do not directly contribute to any free energy difference; while this was not a significant advantage for this model system, it is a major factor in reducing SB uncertainty for larger protein systems. Of particular note, the constraints are between “identical” atoms in the two systems and are not simply freezing the atoms or limiting accessible coordinates.

In this system, the Ca^{2+} and Mg^{2+} ions are effectively a single particle with a single position and momentum. There are six water– Ca^{2+} interactions and six water– Mg^{2+} interactions. At $\lambda = 0.3$ (assuming a linear switch), the water– Ca^{2+} interactions will have 30% fixed-charge character and 70% AMOEBA character, while the water– Mg^{2+} interactions will have 70% fixed-charge character and 30% AMOEBA character. These 12 water–ion interactions are provided by nine water positions: three constrained waters which provide three water– Ca^{2+} and three water– Mg^{2+} interactions, three Ca^{2+} -only waters that provide water– Ca^{2+} interactions, and three Mg^{2+} -only waters that provide the final three water– Mg^{2+} interactions.

METHODS

Materials and Data Availability. The Force Field X software package is freely available from Github (<https://github.com/mjschnie/ffx>), with documentation at a laboratory Web site (<https://ffx.biochem.uiowa.edu/>). Files not included in the SI (e.g., simulation script files) are available upon request.

Sampling Strategy: Transition-Tempered Orthogonal Space Random Walk. The simulations in this work used the transition-tempered orthogonal space random walk (TT-OSRW) algorithm. This method combines transition-tempered metadynamics⁴⁴ with orthogonal space random walk,⁴¹ using thermodynamic integration to estimate free energy differences. Four TT-OSRW walkers per simulation were managed via the Parallel Java library.⁴⁵ Uncertainties are reported as the standard deviation of five replicate simulations. Zero-sum tests (Figures 5A and 6) are reported as population standard deviation with a mean of 0 (eq 12), while other simulations (Figures 5B and 7) report sample standard deviation (eq 13).

The total TT-OSRW potential energy $U_{\text{TT-OSRW}}(\lambda, \mathbf{x})$ is the sum of the potential of interest (i.e. $U(\lambda, \mathbf{x})$ below is either eq 2 or eq 5) and a time-dependent bias defined by $f_m(\lambda) + g_m(\lambda, F_\lambda)$ to yield

$$U_{\text{TT-OSRW}}(\lambda, \mathbf{x}) = U(\lambda, \mathbf{x}) + f_m(\lambda) + g_m(\lambda, F_\lambda) \quad (7)$$

where $g_m(\lambda, F_\lambda)$ is a sum of repulsive potentials given by

$$g_m(\lambda, F_\lambda) = \sum_i h(t_i) \exp \left[\frac{|\lambda - \lambda(t_i)|^2}{2w_1^2} \cdot \frac{|F_\lambda - F_\lambda(t_i)|^2}{2w_2^2} \right] \quad (8)$$

that is used to generate the 1D bias $f_m(\lambda)$ based on the relationship

$$\begin{aligned} f_m(\lambda_j) &= \left. \frac{\partial G}{\partial \lambda} \right|_{\lambda_j} = \langle \partial U / \partial \lambda \rangle_{\lambda_j} \\ &= \frac{\int_{\partial U / \partial \lambda} \partial U / \partial \lambda \exp[k_B T \cdot g_m(\lambda, F_\lambda)] \delta(\lambda - \lambda_j)}{\int_{\partial U / \partial \lambda} \exp[k_B T \cdot g_m(\lambda, F_\lambda)] \delta(\lambda - \lambda_j)} \end{aligned} \quad (9)$$

With the exception of the time dependence of contributed biases $h(t_i)$, this transition tempered formulation is identical to the original nontempered OSRW implementation for AMOEBA.^{41,42} The transition tempering defines a non-constant bias height $h(t_i)$ that decays as the simulation progresses. As the bias progressively flattens the path, $h(t_i)$ decreases asymptotically to 0 based on

$$h(t_i) = h(t_0) \exp \left[\frac{\min(0, V_{\text{th}} - V^*(t_i))}{\Delta T} \right] \quad (10)$$

The tempering threshold V_{th} requires the entire path be at least lightly covered before the transition to tempering begins ($V_{\text{th}} = 2$ kcal/mol in this work). When $V_{\text{th}} = 0$, TT-OSRW is not thresholded. Exponential decay is controlled by a parameter ΔT (in this work $16^*k_B T$ was chosen) and

$$V^*(t_i) = \min_{\lambda} [\max_{F_\lambda} g_m(\lambda, F_\lambda)] \quad (11)$$

where the max operation, for each fixed λ , is over the range of F_λ values (i.e., the 2D g_m histogram is reduced to a 1D function of λ), followed by the min operation over λ . In this work, the two-dimensional bias histogram was defined by λ bins of width

0.005 and F_λ bins of width 2 kcal/mol. The initial Gaussian bias height $h(t_0)$ was set to 0.05 kcal/mol, with standard deviations equal to one bin in either dimension and truncated after five bins during evaluation of the 2D bias (eq 8).

The OSRW sampling strategy was chosen for consistency with prior work,¹⁶ its single-simulation nature, and its pre-existing implementation in Force Field X. Its advantages and disadvantages for use with SB and other IFE methods are a potential avenue of investigation. However, the ideas of SB are consistent with any AFE technique that permits coupled systems, and this work is focused on proving the validity of the SB theory.

Structure Preparation. The starting structures for the metalloproteins studied (Table 2) were based on prior work by the Ren lab.⁹ All bulk ions and almost all waters were deleted, replaced by Generalized Kirkwood (GK)⁴⁶ or Generalized Born (GB)⁴⁷ implicit solvent. However, 0–5 restrained water molecules per ion (Table S2) were kept to mitigate inaccuracies of pure implicit-solvent simulations near the divalent cations. Directly coordinating waters (oxygen within 3 Å of each ion) were considered “bound” waters, whereas waters within 4.5 Å with clear analogues in the crystal structure were considered “loose” waters. All waters were restrained by their oxygen to the ion using flat-bottom distance restraints with a force constant of 5.0 kcal/mol/Å. Loose waters had a flat bottom from 1.0 to 5.0 Å, while bound waters had a flat bottom from 1.69 to 2.49 Å for magnesium and 2.02 to 2.82 Å for calcium (effectively 0.4 Å around the M–O distance reported by Marcus).⁴⁸ Additionally, the ion was restrained to a pair of coordinating carboxyl groups (specifically the carboxyl carbon) with a 5.0 kcal/mol/Å force constant and a 1.0 to 5.0 Å flat bottom. These restraints had the effect of keeping the ion and solvent molecules in the binding site, while still being loose enough to permit local flexibility. The “bound” waters were assumed to be directly coordinating the ion, providing the directionality and explicit hydrogen bonding GK/GB implicit solvents lack, while “loose” waters were added to fill cavities that the bound waters could occasionally swing into. For 4IHB, under Amber99SB/GB only, an 80 kcal/mol/Å flat-bottom restraint was added between the ion and the carboxyl carbon of Glu 74, with a flat bottom from 6.5 to 1000 Å, so as to eliminate a slow time-scale and likely unphysical movement toward the ion. Prior to simulations, the structures were optimized to an RMS gradient criterion of 0.01 kcal/mol/Å using the L-BFGS algorithm^{49,50} and the reference FFX Java (CPU) implementations of AMOEBA/GK and Amber99SB/GB. For direct simulations, minimizations were with respect to the calcium-bound structure, while for indirect SB simulations, λ was set to 0.5 and all coordinates were constrained.

Simulation Details. All simulations used Langevin dynamics⁵¹ at 298.15 K either in vacuum (Figure 4), in implicit solvent (Figures 5 and 7), or in the NVT ensemble (Figure 6). Direct-path condensed-phase simulations utilized a 1.0 fs time step; however, SB simulations used a 0.7 fs time step. This shorter time step is due to constrained atoms being propagated under the influence of two potential energy functions simultaneously (i.e., the coarsen and refine dual force field legs) and thereby twice the normal amount of force. The highest-frequency motions are harmonic bond stretches whose frequency is proportional to the square root of the force constant. If the force constant is doubled, the time step should be reduced by a factor of approximately $2^{-1/2}$ to maintain integration accuracy.^{52,53}

The target force field was the 2018 AMOEBA protein force field, which includes updates to the 2013 AMOEBA force field¹⁴ for calcium and magnesium,⁸ plus other modifications described by Jing et al.⁹ The reference force field was Amber99SB⁵⁴ with its port of the Åqvist parameters⁴³ and a flexible 3-point water model hereafter referred to as Dang–Pettitt water.⁵⁵ It is important to note that Amber99SB/GB was chosen for convenience (being already present in Force Field X) rather than because of any expectation of accurate modeling of divalent cation interactions; indeed, *less* accurate reference results are a stronger demonstration of indirect methods. Nevertheless, based on results from the Ren group,⁹ it is clear that some sort of polarization model is necessary for accurate modeling of divalent cations, even if one were to more carefully choose a fixed-charge model. In the rare case that a pair of constrained atoms differed in their mass (e.g., the fully pinned transformation in Figure 5B, the ion in Figure 4), the mass of the second topology's atom (i.e., magnesium) was used. Alternatively, the convention of using the heavier mass could be adopted, although both choices sample identical thermodynamic ensembles.

Implicit solvent calculations used the Generalized Born (GB) or Generalized Kirkwood (GK) algorithms, a pairwise descreening overlap scale factor of 0.69,^{46,56} a solvent relative permittivity of 78.3, and a base radius scaling factor of 1.06.⁵⁷ For the divalent cations, custom GB/GK radii were fit to match hydration free energies estimated by Asthagiri et al.⁵⁸ (Table S1). The apolar term was based on estimating atomic solvent-exposed surface area from effective Born radii, as described more fully by Schaefer et al.,⁵⁹ and a surface tension of 0.0049 kcal/mol/Å. No cutoffs were applied for any nonbonded term. GPU-accelerated calculations were performed at mixed precision, while CPU-only calculations (e.g., minimizations) were performed at pure double precision.

For AMOEBA, the self-consistent field convergence criterion was 10^{−5} RMS Debye. GPU-accelerated FFX-OpenMM simulations used a DIIS solver,^{19,60} while native Java calculations such as minimizations used a preconditioned conjugate gradient solver.⁶¹ As the SB energy function did not require a van der Waals softcore treatment (λ dependence is applied outside force field energy evaluations), the GPU-accelerated OpenMM package could be used to evaluate the energy and gradient for all simulations. Dynamics and the TT-OSRW algorithm were both propagated in FFX on the CPU, but evaluations of the AMOEBA and Amber99SB force fields were performed on a mix of Nvidia GPUs. Simulation lengths are reported as the aggregate of four TT-OSRW walkers, where each walker contributes to the same histogram but with independent trajectories. Five independent replicates of each simulation were completed to define standard deviation as an uncertainty estimate. After simulations completed 16 ns of aggregate sampling (4 ns per walker), the standard deviation of the free energy was computed. In some cases, extra sampling was required to further reduce uncertainty; dynamics was restarted in 16 ns increments from saved velocities, bias histogram, λ values, *etc.* (Table S3). Most simulations achieved our goal of standard deviations of less than 1.0 kcal/mol despite unphysically large free energy differences arising from inaccuracies in Amber99SB.

Explicit solvent calculations (Figure 6) used periodic boundary conditions with particle-mesh Ewald (PME) for long-range electrostatics.^{62,63} PME used a real-space cutoff of 8.0 Å, a mesh density of 1.2 grid points per Å, an Ewald

parameter of 0.545, and fifth-order B-splines. The van der Waals cutoff was set to 12.0 Å, with a multiplicative switch applied starting at 10.8 Å. The time step was 1.0 fs, and uncertainty was calculated as $\sqrt{2}$ times the standard deviation of five replicate dual force field (DFF) simulations switching between AMOEBA water and Dang–Pettitt water.

Statistical Analysis. Comparison of simulation groups generally used heteroscedastic two-tailed Student's *t* tests; homoscedastic tests were performed when F-tests indicated identical variances at $p > 0.95$ (i.e., our null hypothesis was that means were equal, but variances were unequal). For zero-sum tests, uncertainty was calculated via population standard deviation with known true mean of zero (eq 12), while non-zero-sum uncertainties were calculated with sample standard deviation (eq 13).

$$\sigma_{\text{zero-sum}} = \sqrt{\frac{\sum_{i=1}^n (y_i - 0)^2}{n}} \quad (12)$$

$$\sigma_{\text{sample}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}} \quad (13)$$

Protocol Deviations. Two deviations from the typical protocol were necessary to converge results in an adequate time frame. The first was 2CHE, where Phe 15 had to be constrained to eliminate a very large source of uncertainty, despite being inside the 5 Å sphere. This is a result of Phe 15 being relatively solvent-exposed and mobile, thus causing significant interpotential differences solely from sampling its partial solvation/desolvation. The second was 4IHB, where Glu 74 had to be restrained away from the ion under Amber99SB (both direct-path and SB correction). This is because under Amber99SB/GB (but never under AMOEBA/GK), Glu 74 could swing about by several Ångströms to coordinate the ion, a long-time scale motion that was difficult to converge. While the Amber ion mutation for 1ZOO also had atypically high uncertainty, no singular motion could be identified as being a strong contributor to uncertainty; it appears to be a very flexible pocket, a problem likely aggravated by limitations in the Amber99SB/GB potential.

RESULTS

To obtain an initial estimate of how the SB sampling strategy affected estimates of relative free energy differences, we applied the method to SCPV while systematically varying the number of constrained atoms (Figures 3 and 5 and Table 3). In addition to totally constrained SB and totally unconstrained SB, we chose spheres of constraint removal where all residues with an atom inside a given radius of the ion in the starting structure were left unconstrained.

The first test utilized a zero-sum transformation (Figure 5A) of Ca²⁺ back to Ca²⁺. In this case, the true answer is zero, and any nonzero result is due to sampling uncertainty. When all degrees of freedom were constrained, there was zero uncertainty (Figure 5A, leftmost point), a special case that only applies for fully constrained zero-sum SB. In general, there was a trend of increasing uncertainty as fewer atomic degrees of freedom were constrained, with a free energy difference of zero always within the margins of statistical uncertainty.

The second test transformed Mg²⁺ to Ca²⁺ (Figure 5B). In this case, the true free energy difference can be estimated *via* the difference of direct-path AMOEBA and Amber simulations.

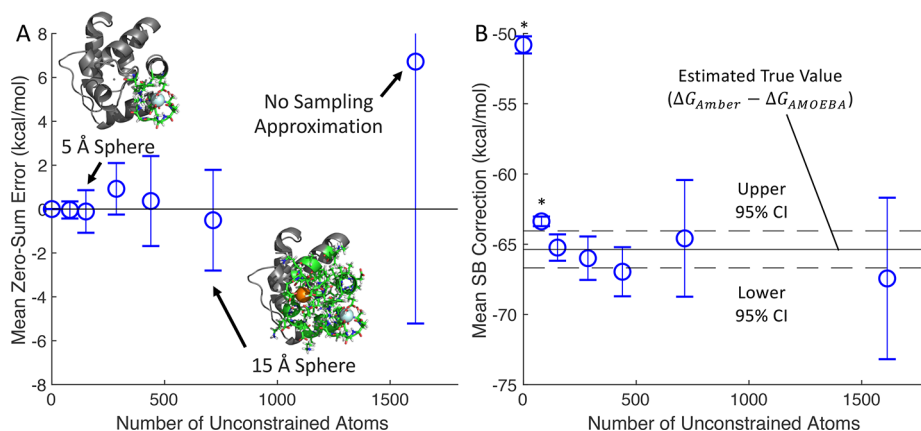


Figure 5. “Simultaneous bookending” indirect free energy method applied to SCPV (Table 2), with AMOEBA/GK as the target potential and Amber99SB/GB as the reference potential. An increasing number of atoms are left unconstrained (X-axis), in “spheres” around the ion. Spheres of size 0, 3 Å, 5 Å, 9 Å, 12 Å, and 15 Å, and all atoms were used. (A) A zero-sum transformation of Ca^{2+} back to Ca^{2+} , with a known true value of zero. Also shown are pictorial representations of the constrained/unconstrained regions for 5 and 15 Å spheres, with constrained atoms in gray cartoon representation and unconstrained atoms as sticks (with the background ion as an orange sphere). Uncertainties are population standard deviation (eq 12) of 5 replicates of 16 ns of TT-OSRW sampling. (B) A Mg^{2+} to Ca^{2+} transformation, with an estimated true mean (solid horizontal line) calculated as the difference between the Amber and AMOEBA direct paths (Table 4). Uncertainties are calculated with sample standard deviation (eq 13). The dashed lines indicate uncertainty in this estimated true mean (95% confidence interval). Asterisks on the 0 and 3 Å points indicate statistically significant difference from this estimated true mean at $p > 0.95$. All values for these SB simulations are provided in Table 3.

With the exception of the fully constrained point, this recapitulated the monotonic increase in uncertainty found in the zero-sum simulations and additionally showed the effect of the sampling approximation. Constraining inside a 5 Å sphere produced incorrect results (as compared to the Amber-AMOEBA direct-path difference) at $p > 0.95$ (Figure 5B). There was a drop in uncertainty between the fully constrained and 3 Å sphere. This indicates that for this system, the nonbonded terms can vary more strongly if there is no local flexibility to accommodate the difference in size between Mg^{2+} and Ca^{2+} .

To supplement this, we also performed dual force field calculations, similar in nature to unconstrained zero-sum SB on explicit solvent boxes of increasing size (Figure 6). The uncertainty scaled approximately as the square root of the number of waters, which is consistent with an increasing number of identical and independent sources of variance. This uncertainty scaling is not so cleanly demonstrated by the protein system (Figure 5), due to the heterogeneity found in a more complicated macromolecular system.

To examine the effectiveness of SB for a challenging biological application, a set of six metalloproteins with experimentally established affinities to both calcium and magnesium was examined.⁹ While not quite as good as the original explicit solvent results,⁹ AMOEBA/GK was highly correlated with experiment (Figure 7A, Tables 4 and 5). Although its mean unsigned error of 1.3 kcal/mol was slightly better than AMOEBA in explicit solvent at 1.6 kcal/mol (Table 4), its Pearson correlation was reduced (Table 5). Meanwhile, Amber99SB/GB had an extraordinarily poor mean unsigned error (MUE) of 57.5 kcal/mol (Figure 7A, Table 4), displayed an incorrect trend, and had poor correlation with experiment (Table 5). Remarkably, the SB method was able to correct Amber99SB/GB to the direct AMOEBA/GK path (Figures 7B and 7C and Tables 4 and 5). While the 2.2 kcal/mol MUE to experiment for SB corrected-Amber99SB/GB is comparable to that for explicit solvent AMOEBA, the 1.0 kcal/mol MUE to the AMOEBA/GK direct path is strong evidence

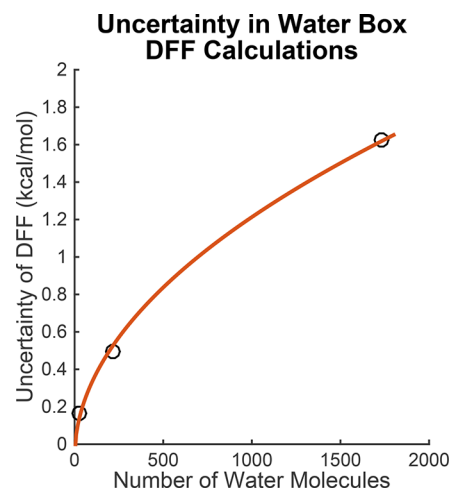


Figure 6. Uncertainty in dual force field simulations between AMOEBA water¹⁵ and Dang–Pettitt water⁵⁵ after 16 ns of transition-tempered orthogonal space random walk. All uncertainties are ± 1 standard deviation from five replicate simulations. Points at 27, 216, and 1728 waters. Square root trend line (in orange): $y = 0.0406 \cdot \sqrt{x} - 0.0703$.

for the success of SB. Student’s t tests between each direct AMOEBA/GK and indirect SB pair showed only one statistically significant difference at $\alpha = 0.05$ (for 4ICB at $p = 0.016$). This significance, however, disappears upon use of a Bonferroni multiple test correction,^{64,65} suggesting it may simply be a coincidence caused by six independent tests.

In particular, we extended the SCPV simulations to assess long-time scale convergence of our TT-OSRW based simulation protocol as applied to direct-path and the standard 5 Å sphere SB, using one of our “best-behaved” systems. For this system, we reached 64 ns of AMOEBA and SB sampling per replicate and 96 ns of Amber sampling per replicate. Neither the mean nor the uncertainty values drifted significantly between our initial stopping points (16 ns direct-path, 32 ns SB) and our final stopping points (Figure

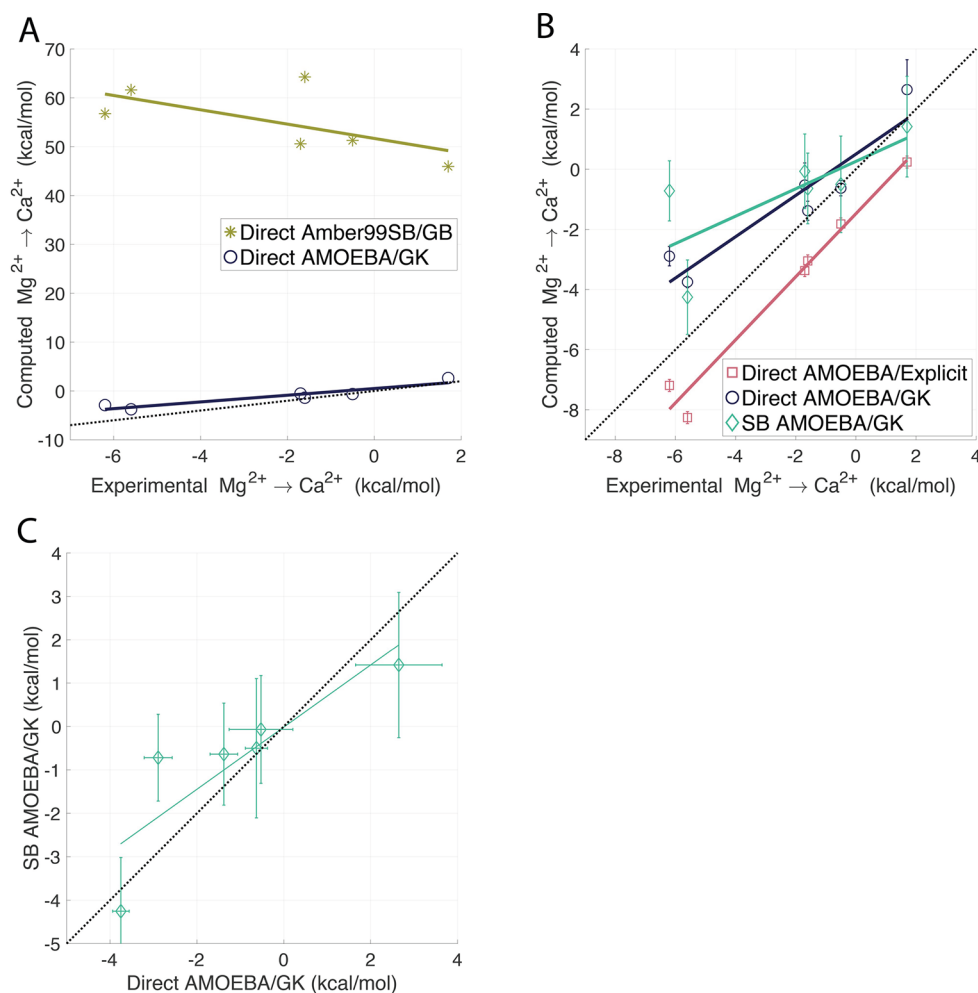


Figure 7. Various computational estimates (Tables 4 and 5) of relative $\text{Ca}^{2+}/\text{Mg}^{2+}$ binding (versus experimental values) to six metalloproteins (Table 2), with dotted $y = x$ reference lines. The values are for $\Delta G_{\text{binding}}(\text{Ca}^{2+}) - \Delta G_{\text{binding}}(\text{Mg}^{2+})$. (A) Direct free energy estimates in implicit solvent. (B) AMOEBA free energy estimates given from a direct path in implicit solvent, a direct path in explicit solvent, and an indirect (simultaneous bookending) path in implicit solvent. (C) Simultaneous bookending versus direct free energy estimates for AMOEBA in implicit solvent. All uncertainties are ± 1 sample standard deviation from five independent simulations (eq 13).

S1), suggesting that our protocol reaches initial estimates quickly but did not further reduce uncertainty past that. It may be possible to improve upon the TT-OSRW parameters used here (e.g., we have found in other recent studies that an initial Gaussian bias height $h(t_0)$ of 0.002 kcal/mol improves convergence relative to the value of 0.05 kcal/mol used here), but our aim was to demonstrate the validity of the SB algorithm.

Despite the relatively short per-replicate simulation times (16–96 ns), this work represents a significant amount of protein-scale AMOEBA sampling (Table S3). Between the metalloproteins (Figures 5, 7), we ran a total of 800 ns of direct-path AMOEBA sampling, 1040 ns of direct-path Amber sampling, and 2480 ns of SB sampling (of which 1120 ns was solely for Figure 5). After normalizing for the 0.7 fs time step of SB, the two force fields per direct-path mutation, and the four force fields per SB simulation, this amounts to about 8.7 trillion AMOEBA force field evaluations and 9.2 trillion Amber force field evaluations (i.e., the equivalent of 8.7 μs of AMOEBA and 9.2 μs of Amber sampling with 1 fs time steps). The efficiency of this type of simulation can be significantly improved by reducing CPU-GPU communication (as our forces were calculated on GPU, but Newton's laws of motion

integrated on CPU), utilizing methods such as hydrogen mass repartitioning to increase time step or devising a single-topology parameter interpolation rather than interpolating two separate force fields. Although it is tempting to consider reducing bonded force constants for constrained atoms by a factor of 2 to permit restoring the SB time step back to 1 fs, this is a non-negligible approximation (see the Supporting Information for a detailed demonstration).

CONCLUSIONS

Previous indirect free energy methods are characterized by their use of tunable *potential energy-based approximations*, wherein a limited subset of atoms are coarsened and refined between potential energy resolutions. Here the “Simultaneous Bookending” IFE method was introduced, which opens the door to a unique class of indirect free energy methods characterized by the use of tunable *sampling-based approximations*. For SB, all atoms are coarsened and refined between potential energy resolutions; however, some atomic degrees of freedom are constrained during *simultaneous* sampling of the coarsen and refine free energy legs. This allows SB to reweight dramatically more atoms between resolutions than prior methods, which was demonstrated by correcting relative

Table 3. Zero-Sum Ca^{2+} Simultaneous Bookending Results^a

Constraints	Atoms Unconstrained	$\Delta\Delta G$ (kcal/mol)	Uncertainty (kcal/mol)
A: Zero-Sum			
All Constrained	0	0	0
3 Å Sphere	80	−0.25	0.32
3.8 Å Sphere	106	−0.04	0.39
5 Å Sphere	151	−0.11	0.97
9 Å Sphere	286	0.92	1.17
12 Å Sphere	439	0.37	2.05
15 Å Sphere	715	−0.51	2.29
Unconstrained	1613	6.72	11.94
B: Mg^{2+} to Ca^{2+}			
AMOEBA-Amber*	N/A	−65.37	1.32
All Constrained	0	−50.81	0.61
3 Å Sphere	80	−63.75	0.45
3.8 Å Sphere	106	−63.37	0.34
5 Å Sphere	151	−65.24	0.94
9 Å Sphere	286	−66.00	1.55
12 Å Sphere	439	−66.96	1.75
15 Å Sphere	715	−64.58	4.15
Unconstrained	1613	−67.44	5.76

^aSimultaneous bookending (SB) was applied to SCPV with varying numbers of atoms constrained (Figure 5). (A) Zero-sum calculations, of Ca^{2+} back to Ca^{2+} , with a known true value of zero. (B) Corrections for Mg^{2+} to Ca^{2+} , with a true value estimated from direct-path simulations (see Figure 7 and Table 4). Uncertainties are the standard deviation of five replicate SB simulations at 16 nanoseconds, with the direct-path estimate being from five replicates run to full length (Table S3). The asterisk indicates the direct-path estimate, which is not from SB.

divalent cation binding thermodynamics from fixed-charge Amber99SB/GB to polarizable AMOEBA/GK. Increasing the extent of the SB approximation systematically reduced uncertainty, given a fixed amount of sampling, from more than 6 kcal/mol down to less than 0.5 kcal/mol (Figure 5).

The SB method represents a substantial jump in the number of atoms reweighted by an indirect free energy method (Figure 2). Whereas other IFE methods make the approximation that only a relatively small (~50 atoms at most) region can be converted to the more sophisticated potential, SB converted all atoms to AMOEBA by constraining the sampling of all atoms outside a 5 Å sphere from the location of the divalent cation. Despite the reference model producing large, poorly correlated errors (Figure 7A, Table 4), SB simulations corrected Amber99SB/GB free energy differences to such an extent

Table 5. Computed Calcium Specificity Trend Lines^a

Result	Correlation To	Slope	Y-Intercept	R ²
AMOEBA/Explicit	Experiment	1.05	−1.48	0.97
Direct Amber99SB/GB	Experiment	−1.47	51.69	0.40
Direct AMOEBA/GK	Experiment	0.69	0.50	0.87
SB AMOEBA/GK	Direct AMOEBA/GK	0.72	−0.02	0.73
SB AMOEBA/GK	Experiment	0.46	0.26	0.55

^aLinear regression lines for correlating computed measures of Ca^{2+} vs Mg^{2+} specificity to other measures (Figure 7, Table 4). The final two rows correlate simultaneous bookending (SB) to direct AMOEBA/GK and experiment, respectively.

that they became statistically indistinguishable from direct AMOEBA/GK results. It is important to note that not only did SB correct large errors, it corrected a large range of errors; SB free energy corrections were not simply a constant offset added to the Amber99SB/GB results. Not surprisingly, indirect AMOEBA/GK results based on SB corrections were better correlated to the direct AMOEBA/GK path than to either Amber99SB/GB or experiment (Table 5).

The SB approach, and potentially follow-on approaches utilizing a similar sampling-based approximation, have advantages and disadvantages versus potential-based approximations. The primary advantage is outright elimination of the potential-based approximation; long-range effects are included at full target potential detail. There are no “cross-potential” terms (i.e., there is no polarizable MM boundary with fixed charge MM) as the entire system is smoothly and uniformly interpolated between target and reference potentials. This results in a scalable algorithm that can theoretically address very large systems, scaling only by the size of the unconstrained region.

Five issues appear with SB, of which the first three are general to large-scale IFE methods and the latter two are, to our knowledge, specific to the nature of SB. The first issue is inaccuracy in the conformational ensemble of distal parts of the system. With potential-based approximation, these distal regions are represented by a less detailed model that may not properly capture key dynamics. With the sampling-based approximation, the distal region is forced to sample a hybrid topology that is not quite equivalent to either of the physical end states. With neither approximation, large systems are prohibitively difficult to converge.

A second issue is the need for an identical number of background degrees of freedom between resolutions (i.e., both

Table 4. Computed Calcium Specificities^{a9}

System	Direct Amber99SB/GB	SB Correction	SB AMOEBA/GK	Direct AMOEBA/GK	AMOEBA Explicit	Experiment
1B8L	64.27 ± 0.52	−65.16 ± 1.09	−0.89 ± 1.21	−1.38 ± 0.31	−3.05 ± 0.2	−1.6
4ICB	56.76 ± 0.60	−59.04 ± 1.98	−2.27 ± 2.07	−2.89 ± 0.32	−7.19 ± 0.2	−6.2
1ZOO	45.96 ± 1.46	−43.10 ± 1.70	2.85 ± 2.24	3.00 ± 1.29	0.24 ± 0.2	1.7
SCPV	61.62 ± 1.04	−65.87 ± 0.67	−4.25 ± 1.24	−3.75 ± 0.19	−8.26 ± 0.2	−5.6
4IHB	50.50 ± 1.09	−50.12 ± 2.03	0.39 ± 2.30	−0.53 ± 0.73	−3.37 ± 0.2	−1.7
2CHE	51.30 ± 0.90	−49.68 ± 1.47	1.61 ± 1.73	−0.63 ± 0.26	−1.82 ± 0.2	−0.5
MUE	57.40		1.62	1.27	1.59	
MSE	57.40		1.52	1.23	−1.59	

^aThe experimental and computed $\Delta G_{\text{binding, Ca}^{2+}} - \Delta G_{\text{binding, Mg}^{2+}}$ values for six metalloproteins (Figure 7), plus mean unsigned error (MUE) and mean signed error (MSE) with respect to experimental values. All values are in kcal/mol. The values for AMOEBA in explicit solvent were taken from prior work.

resolutions must be based on either flexible models or rigid models with identical equilibrium distances and angles). Here this was addressed by use of purely flexible water models to accommodate the AMOEBA model, despite the wide use of rigid definitions for many fixed-charge water models.^{66–68} Another possibility is to use paired potentials with identical, rigid bonded geometry, but this depends on use of a reference potential specifically designed for compatibility with the target potential. As before, this issue arises for all IFE methods and is not specific to SB.

A third consideration is the handling of solvent or, more generally, the issue of highly mobile atoms that can approach and then leave an alchemical site. In this work, we chose to use implicit solvent coupled with a few restrained solvent molecules proximal to the divalent cations. Others have also observed that deficiencies in implicit solvent models can be ameliorated by inclusion of a limited amount of explicit solvent, effectively pushing the explicit–implicit boundary approximation further away from key atoms.^{69–72} In the future, we will explore the possibility of using a small number of restrained solvent molecules without SB constraints (for topology-specific interactions), with bulk solvent represented by solvent molecules with SB constraints.

A fourth issue, the first specific to SB, is obligate pairing of IFE legs. For example, consider mutating alanine to all other 19 amino acids. While just 19 SB simulations are needed, versus 20 DFF simulations (19 mutations plus alanine itself), each SB time step requires four potential energy function evaluations (two systems at target and reference potentials), while DFF and potential-based approximations require just two evaluations (one system at target and reference potential). While not an issue for single changes (e.g., only considering Mg^{2+} to Ca^{2+}), one-to-many efforts approach a doubling of effort. The advantage is, of course, that IFE corrections for large systems can be practically converged without the potential-based approximation.

Finally, a reduced time step is necessitated by the doubling of forces (e.g., especially bonded terms) in the distal constrained atom region. Approaches to mitigate the reduced time step include multiple time step algorithms, hydrogen mass repartitioning, or simply increasing the masses of atoms in the distal region. Despite altered kinetics, the latter approach permits the unconstrained (and assumed more-important) region to sample at full speed. On the other hand, simply halving bonded term force constants in the background region is not a desirable approach, as this results in a further approximation the conformational ensemble and thus incorrect thermodynamics (as discussed further in the [Supporting Information](#)).

Despite these limitations, we obtained statistically significant agreement between SB and direct-path AMOEBA for at least five of six metalloproteins ([Figure 7](#), [Table 4](#)). The possible outlier, 4ICB, was significantly different at $\alpha = 0.05$, but this significance disappears after a Bonferroni multiple test correction^{64,65} to reduce α to 0.0083. From this, we conclude that SB is a valid IFE approach, based on a novel sampling-based approximation.

Use of the Amber99SB/GB/Åqvist potential for the reference path simulations had both advantages and disadvantages. This potential was convenient due to its broad adoption in software platforms (e.g., in Amber, Force Field X, OpenMM, TINKER, etc.), and the magnitude/range of relative binding affinity errors served to provide a challenging

application of the SB algorithm. However, this potential was not designed for protein-ion binding thermodynamics. Beyond the large errors obtained for Amber99SB/GB/Åqvist relative binding free energies, these simulations were consistently harder to converge than those under AMOEBA/GK ([Tables 4](#) and [S3](#)). While limitations in Amber99SB/GB were helpful in demonstrating the power of the SB algorithm, future applications will benefit from choosing more accurate fixed charge reference potentials with better phase space overlap with the target potential. As an example of more accurate fixed-charge results, Jing et al.⁹ obtained results with the more recent ff14SB Amber force field,⁷³ HFE ion parameters (optimized for hydration free energy),⁷⁴ and explicit solvent. However, their work was still negatively correlated with experiment, producing a mean signed error of 11.7 kcal/mol.⁹ Meanwhile, AMOEBA was well correlated with experiment under either explicit solvent⁹ or implicit solvent ([Figure 7B](#), [Tables 4](#) and [5](#)). Our implicit-solvent results had similar mean unsigned error as the Ren lab explicit solvent results,⁹ though with a different sign on the mean signed error and a weaker correlation ([Figure 7B](#), [Table 5](#)).

While we have demonstrated that SB is a *valid* IFE method, application to the relative binding of divalent cations was not amenable to demonstrating increased efficiency relative to using a direct free energy difference path. Since the direct path under AMOEBA was well-behaved and converged rapidly, SB corrections actually required more simulation time than the direct AMOEBA/GK path ([Table S3](#)). Efficiency advantages will be explored in future work for free energy differences that depend on longer time scale motions or possibly target a more expensive target potential such as the Gaussian Electrostatics Model (GEM)⁷⁵ or the Sum of Interactions Between Fragments *Ab initio* computed (SIBFA).⁷⁶ It may also be possible to reweight coarse-grain simulations to all-atom resolution, although this introduces an additional challenge due to the number of degrees of freedom changing between resolutions. Finally, we note that it should be possible to combine the SB sampling approximation with more conventional potential energy approximations into hybrid IFE schemes.

■ ASSOCIATED CONTENT

§ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.9b00147.

Supplementary text: Detailed mathematical derivation for simultaneous bookending and a discussion of why halving background bonded terms produces an additional layer of approximation. [Figure S1](#): Mean and standard deviation over time of replicate simulations on SCPV. [Figure S2](#): Version of [Figure 2](#) with a linear-scale Y-axis. [Table S1](#): Force field parameters used for divalent metal cations. [Table S2](#): Number of explicit solvent molecules used in each system. [Table S3](#): Amount of sampling per replicate simulation ([Figure 7](#)). [Table S4](#): Results for simultaneous bookending applied to an ion–water cluster (see supplementary text). [Table S5](#): Statistical tests applied to ion–water cluster SB results ([PDF](#))

[Movie S1](#): Short simultaneous bookending trajectory for an ion–water cluster, from which still frames were extracted for [Figure 4](#) ([MPG](#))

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: michael-schnieders@uiowa.edu.

ORCID

Jacob Litman: 0000-0002-3256-8747

Andrew C. Thiel: 0000-0002-8652-3042

Michael J. Schnieders: 0000-0003-1260-4592

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Pengyu Ren and Zhifeng Jing for providing data and the starting structures. We thank James Dama for help with the initial implementation of transition-tempering for OSRW. This research was supported in part through computational resources provided by The University of Iowa, Iowa City, Iowa. In particular, we would like to acknowledge support and guidance from Joe Hetrick, Glenn Johnson, Brenna Miller, Ben Rogers, and John Sexton. We are grateful for support from the following grants: JML was supported by NIH/CBB Award T32 GM0008365 and the University of Iowa Presidential Graduate Research Fellowship. MJS was supported by NIH RO1DK110023, NIH RO1DC012049, and NSF CHE-1751688.

■ REFERENCES

- (1) Lemmon, M. A.; Schlessinger, J. Cell Signaling by Receptor Tyrosine Kinases. *Cell* **2010**, *141* (7), 1117–1134.
- (2) Lemmon, M. A. Membrane recognition by phospholipid-binding domains. *Nat. Rev. Mol. Cell Biol.* **2008**, *9*, 99.
- (3) Dudev, T.; Lim, C. Principles Governing Mg, Ca, and Zn Binding and Selectivity in Proteins. *Chem. Rev.* **2003**, *103* (3), 773–788.
- (4) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **2000**, *289* (5481), 920–930.
- (5) Rocklin, G. J.; Boyce, S. E.; Fischer, M.; Fish, I.; Mobley, D. L.; Shoichet, B. K.; Dill, K. A. Blind Prediction of Charged Ligand Binding Affinities in a Model Binding Site. *J. Mol. Biol.* **2013**, *425* (22), 4569–4583.
- (6) Jiao, D.; Zhang, J. J.; Duke, R. E.; Li, G. H.; Schnieders, M. J.; Ren, P. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J. Comput. Chem.* **2009**, *30* (11), 1701–1711.
- (7) 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, R. A., Jr.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114* (8), 2549–2564.
- (8) Jing, Z.; Qi, R.; Liu, C.; Ren, P. Study of interactions between metal ions and protein model compounds by energy decomposition analyses and the AMOEBA force field. *J. Chem. Phys.* **2017**, *147* (16), 161733.
- (9) Jing, Z.; Liu, C.; Qi, R.; Ren, P. Many-body effect determines the selectivity for Ca²⁺ and Mg²⁺ in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (32), E7495–E7501.
- (10) Kohagen, M.; Lepšík, M.; Jungwirth, P. Calcium Binding to Calmodulin by Molecular Dynamics with Effective Polarization. *J. Phys. Chem. Lett.* **2014**, *5* (22), 3964–3969.
- (11) Li, P. F.; Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10* (1), 289–297.
- (12) Pokorna, P.; Krepl, M.; Kruse, H.; Sponer, J. MD and QM/MM Study of the Quaternary HutP Homohexamer Complex with mRNA, L-Histidine Ligand, and Mg²⁺. *J. Chem. Theory Comput.* **2017**, *13* (11), 5658–5670.
- (13) Hitznerberger, M.; Hofer, T. S. The influence of metal-ion binding on the structure and surface composition of Sonic Hedgehog: a combined classical and hybrid QM/MM MD study. *Phys. Chem. Chem. Phys.* **2016**, *18* (32), 22254–22265.
- (14) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9* (9), 4046–4063.
- (15) Ren, P.; Ponder, J. W. Temperature and Pressure Dependence of the AMOEBA Water Model. *J. Phys. Chem. B* **2004**, *108* (35), 13427–13437.
- (16) Nessler, I. J.; Litman, J. M.; Schnieders, M. J. Toward polarizable AMOEBA thermodynamics at fixed charge efficiency using a dual force field approach: application to organic crystals. *Phys. Chem. Chem. Phys.* **2016**, *18*, 30313.
- (17) Park, J.; Nessler, I.; McClain, B.; Macikenas, D.; Baltrusaitis, J.; Schnieders, M. J. Absolute Organic Crystal Thermodynamics: Growth of the Asymmetric Unit into a Crystal via Alchemy. *J. Chem. Theory Comput.* **2014**, *10* (7), 2781–2791.
- (18) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chemical Science* **2018**, *9* (4), 956–972.
- (19) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y. T.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *Plos Computational Biology* **2017**, *13* (7), 17.
- (20) Zuckerman, D. M. Equilibrium Sampling in Biomolecular Simulations. In *Annual Review of Biophysics*; Rees, D. C., Dill, K. A., Williamson, J. R., Eds.; Annual Reviews: Palo Alto, 2011; Vol. 40, pp 41–62.
- (21) Gao, J. L. Absolute Free-Energy of Solvation from Monte-Carlo Simulations Using Combined Quantum and Molecular Mechanical Potentials. *J. Phys. Chem.* **1992**, *96* (2), 537–540.
- (22) Vaidehi, N.; Wesolowski, T. A.; Warshel, A. Quantum-mechanical calculations of solvation free energies. A combined *ab initio* pseudopotential free-energy perturbation approach. *J. Chem. Phys.* **1992**, *97* (6), 4264–4271.
- (23) Beierlein, F. R.; Michel, J.; Essex, J. W. A Simple QM/MM Approach for Capturing Polarization Effects in Protein-Ligand Binding Free Energy Calculations. *J. Phys. Chem. B* **2011**, *115* (17), 4911–4926.
- (24) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method 0.1. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22* (8), 1420–1426.
- (25) Dybeck, E. C.; König, G.; Brooks, B. R.; Shirts, M. R. Comparison of Methods To Reweight from Classical Molecular Simulations to QM/MM Potentials. *J. Chem. Theory Comput.* **2016**, *12* (4), 1466–1480.
- (26) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129* (12), 124105.
- (27) König, G.; Boresch, S. Non-Boltzmann sampling and Bennett's acceptance ratio method: How to profit from bending the rules. *J. Comput. Chem.* **2011**, *32* (6), 1082–1090.
- (28) König, G.; Pickard, F. C.; Mei, Y.; Brooks, B. R. Predicting hydration free energies with a hybrid QM/MM approach: an evaluation of implicit and explicit solvation models in SAMPL4. *J. Comput.-Aided Mol. Des.* **2014**, *28* (3), 245–257.
- (29) Genheden, S.; Cabedo Martinez, A. I.; Criddle, M. P.; Essex, J. W. Extensive all-atom Monte Carlo sampling and QM/MM corrections in the SAMPL4 hydration free energy challenge. *J. Comput.-Aided Mol. Des.* **2014**, *28* (3), 187–200.
- (30) Metropolis, N.; Ulam, S. The Monte Carlo Method. *J. Am. Stat. Assoc.* **1949**, *44* (247), 335–341.

- (31) Iftimie, R.; Salahub, D.; Wei, D.; Schofield, J. Using a classical potential as an efficient importance function for sampling from an ab initio potential. *J. Chem. Phys.* **2000**, *113* (12), 4852–4862.
- (32) Sampson, C.; Fox, T.; Tautermann, C. S.; Woods, C.; Skylaris, C. K. A "Stepping Stone" Approach for Obtaining Quantum Free Energies of Hydration. *J. Phys. Chem. B* **2015**, *119* (23), 7030–7040.
- (33) Caldararu, O.; Olsson, M. A.; Misini Ignjatović, M.; Wang, M.; Ryde, U. Binding free energies in the SAMPL6 octa-acid host–guest challenge calculated with MM and QM methods. *J. Computer-Aided Molecular Design* **2018**, *32*, 1027.
- (34) Woods, C. J.; Manby, F. R.; Mulholland, A. J. An efficient method for the calculation of quantum mechanics/molecular mechanics free energies. *J. Chem. Phys.* **2008**, *128* (1), 8.
- (35) Cates, M. S.; Berry, M. B.; Ho, E. L.; Li, Q.; Potter, J. D.; Phillips, G. N. Metal-ion affinity and specificity in EF-hand proteins: coordination geometry and domain plasticity in parvalbumin. *Structure* **1999**, *7* (10), 1269–1278.
- (36) Vorup-Jensen, T.; Waldron, T. T.; Astrof, N.; Shimaoka, M.; Springer, T. A. The connection between metal ion affinity and ligand affinity in integrin I domains. *Biochim. Biophys. Acta, Proteins Proteomics* **2007**, *1774* (9), 1148–1155.
- (37) Needham, J. V.; Chen, T. Y.; Falke, J. J. Novel Ion Specificity of a Carboxylate Cluster Mg(II) Binding Site: Strong Charge Selectivity and Weak Size Selectivity. *Biochemistry* **1993**, *32* (13), 3363–3367.
- (38) Andersson, M.; Malmendal, A.; Linse, S.; Ivarsson, I.; Forsen, S.; Svensson, L. A. Structural basis for the negative allostery between Ca²⁺- and Mg²⁺-binding in the intracellular Ca²⁺-receptor calbindin D-9k. *Protein Sci.* **1997**, *6* (6), 1139–1147.
- (39) Abdullah, N.; Padmanarayana, M.; Marty, N. J.; Johnson, C. P. Quantitation of the Calcium and Membrane Binding Properties of the C2 Domains of Dysferlin. *Biophys. J.* **2014**, *106* (2), 382–389.
- (40) Henzl, M. T.; Larson, J. D.; Agah, S. Estimation of parvalbumin Ca²⁺- and Mg²⁺-binding constants by global least-squares analysis of isothermal titration calorimetry data. *Anal. Biochem.* **2003**, *319* (2), 216–233.
- (41) Zheng, L.; Chen, M.; Yang, W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (51), 20227–20232.
- (42) Schnieders, M. J.; Baltusaitis, J.; Shi, Y.; Chatterjee, G.; Zheng, L.; Yang, W.; Ren, P. The Structure, Thermodynamics and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. *J. Chem. Theory Comput.* **2012**, *8* (5), 1721–1736.
- (43) Aqvist, J. Ion-water interaction potentials derived from free energy perturbation simulations. *J. Phys. Chem.* **1990**, *94* (21), 8021–8024.
- (44) Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A. Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. *J. Chem. Theory Comput.* **2014**, *10* (9), 3626–3633.
- (45) Kaminsky, A. In *Parallel Java: A Unified API for Shared Memory and Cluster Parallel Programming in 100% Java*, 2007 IEEE International Parallel and Distributed Processing Symposium, 26–30 March 2007; 2007; pp 1–8.
- (46) Schnieders, M. J.; Ponder, J. W. Polarizable atomic multipole solutes in a generalized Kirkwood continuum. *J. Chem. Theory Comput.* **2007**, *3* (6), 2083–2097.
- (47) Bashford, D.; Case, D. A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129–152.
- (48) Marcus, Y. Ionic radii in aqueous solutions. *Chem. Rev.* **1988**, *88* (8), 1475–1498.
- (49) Nocedal, J. Updating quasi-Newton matrices with limited storage. *Math. Comput.* **1980**, *35* (151), 773–782.
- (50) Liu, D. C.; Nocedal, J. On the limited memory BFGS method for large-scale optimization. *Math. Program.* **1989**, *45* (3), 503–528.
- (51) Guarnieri, F.; Still, W. C. A rapidly convergent simulation method: Mixed Monte Carlo/stochastic dynamics. *J. Comput. Chem.* **1994**, *15* (11), 1302–1310.
- (52) Leach, A. R. *Molecular Modeling: Principles and Applications*, 2nd ed.; Prentice Hall: Harlow, England, 2001; p 744.
- (53) Zhang, G.; Schlick, T. LIN: A new algorithm to simulate the dynamics of biomolecules by combining implicit-integration and normal mode techniques. *J. Comput. Chem.* **1993**, *14* (10), 1212–1233.
- (54) 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: Struct., Funct., Genet.* **2006**, *65* (3), 712–725.
- (55) Dang, L. X.; Pettitt, B. M. Simple Intramolecular Model Potentials for Water. *J. Phys. Chem.* **1987**, *91* (12), 3349–3354.
- (56) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, *246* (1–2), 122–129.
- (57) Bondi, A. van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68* (3), 441–451.
- (58) Asthagiri, D.; Pratt, L. R.; Paulaitis, M. E.; Rempe, S. B. Hydration Structure and Free Energy of Biomolecularly Specific Aqueous Dications, Including Zn²⁺ and First Transition Row Metals. *J. Am. Chem. Soc.* **2004**, *126* (4), 1285–1289.
- (59) Schaefer, M.; Bartels, C.; Karplus, M. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.* **1998**, *284* (3), 835–848.
- (60) Pulay, P. Convergence acceleration of iterative sequences. the case of SCF iteration. *Chem. Phys. Lett.* **1980**, *73* (2), 393–398.
- (61) Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations. *J. Chem. Theory Comput.* **2014**, *10* (4), 1638–1651.
- (62) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N * log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.
- (63) Schnieders, M. J.; Fenn, T. D.; Pande, V. S. Polarizable Atomic Multipole X-Ray Refinement: Particle Mesh Ewald Electrostatics for Macromolecular Crystals. *J. Chem. Theory Comput.* **2011**, *7* (4), 1141–1156.
- (64) Bonferroni, C. Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze* **1936**, *8*, 3–62.
- (65) Shaffer, J. P. Multiple Hypothesis Testing. *Annu. Rev. Psychol.* **1995**, *46* (1), 561–584.
- (66) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
- (67) Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **2000**, *112* (20), 8910–8922.
- (68) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.
- (69) Gregersen, B. A.; Lopez, X.; York, D. M. Hybrid QM/MM Study of Thio Effects in Transphosphorylation Reactions: The Role of Solvation. *J. Am. Chem. Soc.* **2004**, *126* (24), 7504–7513.
- (70) Zhang, L.; Xie, D.; Xu, D.; Guo, H. Supermolecule density functional calculations suggest a key role for solvent in alkaline hydrolysis of p-nitrophenyl phosphate. *Chem. Commun.* **2007**, No. 16, 1638–1640.
- (71) Lin, J.-H.; Baker, N. A.; McCammon, J. A. Bridging Implicit and Explicit Solvent Approaches for Membrane Electrostatics. *Biophys. J.* **2002**, *83* (3), 1374–1379.
- (72) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. Improved Efficiency of Replica Exchange Simulations through Use of a Hybrid Explicit/Implicit Solvation Model. *J. Chem. Theory Comput.* **2006**, *2* (2), 420–433.

(73) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713.

(74) Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for + 2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9* (6), 2733–2748.

(75) Cisneros, G. A.; Piquemal, J. P.; Darden, T. A. Generalization of the Gaussian electrostatic model: Extension to arbitrary angular momentum, distributed multipoles, and speedup with reciprocal space methods. *J. Chem. Phys.* **2006**, *125* (18), 16.

(76) Gresh, N. Development, validation, and applications of anisotropic polarizable molecular mechanics to study ligand and drug-receptor interactions. *Curr. Pharm. Des.* **2006**, *12* (17), 2121–2158.