

A Transferable Nonbonded Pairwise Force Field to Model Zinc Interactions in Metalloproteins

Ruibo Wu,^{†,‡} Zhenyu Lu,[†] Zexing Cao,[‡] and Yingkai Zhang^{*,†}

*Department of Chemistry, New York University, New York, New York 10003, United States
 and Department of Chemistry and State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China*

Received September 15, 2010

Abstract: Herein we introduce a novel practical strategy to overcome the well-known challenge of modeling the divalent zinc cation in metalloproteins. The main idea is to design short–long effective functions (SLEF) to describe charge interactions between the zinc ion and all other atoms. This SLEF approach has the following desired features: (1) It is pairwise, additive, and compatible with widely used atomic pairwise force fields for modeling biomolecules; (2) It only changes interactions between the zinc ion and other atoms and does not affect force field parameters that model other interactions in the system; (3) It is a nonbonded model that is inherently capable to describe different zinc ligands and coordination modes. By optimizing two SLEF parameters as well as zinc van der Waals parameters through force matching based on Born–Oppenheimer ab initio quantum mechanical/molecular mechanical (QM/MM) molecular dynamics (MD) simulations, we have successfully developed the first SLEF force field (SLEF1) to describe zinc interactions. Extensive MD simulations of seven zinc enzyme systems with different coordination ligands and distinct chelation modes (four-, five-, and six-fold), including a binuclear zinc active site, yielded zinc coordination numbers and binding distances in good agreement with the corresponding crystal structures as well as ab initio QM/MM MD results. This not only demonstrates the transferability and adequacy of the new SLEF1 force field in describing a variety of zinc proteins but also indicates that this novel SLEF approach is a promising direction to explore for improving force field description of metal ion interactions.

1. Introduction

Zinc proteins constitute approximately 10% of the total human proteome¹ and play a variety of essential biological roles,^{2–5} such as transcription factors, signaling proteins, and transport/storage proteins as well as enzymes. Their function and/or structural organization are critically dependent on the zinc binding site,^{4,6–8} which can be classified as catalytic, structural, inhibitory, and protein interface zinc sites based on the role of the divalent zinc cation. Typical zinc ligands include side chains of Cys, His, Glu, and Asp, water molecules, and other small molecules. A key feature of the

zinc coordination is its flexibility:^{4,9–12} it can adopt multiple binding modes, including tetrahedral-, penta-, or hexacoordination geometry. Especially for the zinc coordination to the carboxylate group, it could be either bidentate or monodentate. This inherent flexibility of zinc coordination poses a daunting challenge for all currently available pairwise atomic force fields to describe zinc interactions,^{13–18} including bonded,^{19–24} nonbonded,²⁵ and semibonded²⁶ models.

In the bonded model,^{19–24} zinc–ligand coordination interactions are modeled as covalent bonds, and the desired zinc coordination geometry is maintained by employing explicit bonding and angle bending terms. This clearly prevents any change of the zinc coordination mode or ligand exchange, therefore not suitable for describing the dynamics of zinc coordination. For the nonbonded model,²⁵ in which

* Corresponding author. E-mail: yingkai.zhang@nyu.edu.

[†] New York University.

[‡] Xiamen University.

interactions between the zinc ion and all other atoms are described by electrostatics and van der Waals (vdW) terms, it has been notoriously known for its failure in describing the tetra- or pentacoordinated zinc cation.^{13–18} Previous simulations of several zinc-containing proteins with nonbonded models have led to very different coordination modes in comparison with corresponding X-ray structures.^{13–15,17} In the semibonded model,^{26,27} virtual fractional charges around a metal atom are employed to mimic valence electrons. It has been shown to describe the tetracoordinated zinc ion well, but its capability to model penta- and hexacoordination has not been demonstrated. Currently, it has been widely thought that pairwise atomic force fields may be inherently unsuitable for describing flexible zinc coordination, and it would be necessary to employ polarizable force fields or quantum mechanical/molecular mechanical (QM/MM) methods to explicitly take account of polarization and charge-transfer effects between Zn²⁺ and its ligands.^{13,17,28–33}

In this work, we are motivated to develop a novel practical strategy to tackle this well-known challenge of modeling the divalent zinc cation in metalloproteins. The working hypothesis is that the main deficiency of existing nonbonded models comes from the 1/r function form for the charge–charge interaction term. It is not appropriate to describe zinc coordination bonding, although it may be reasonable to describe long-range electrostatic interactions between the Zn²⁺ ion and other atoms beyond the first coordination shell. Thus our main idea is to design short–long effective functions (SLEF) to describe charge interactions between the zinc ion and all other atoms. The short-range is designed to describe the coordination bonding between the zinc ion and its ligands, while the other behaves similar to 1/r for long-range electrostatic interactions. Herein by optimizing a total of four parameters through force matching^{34–37} based on Born–Oppenheimer ab initio QM/MM molecular dynamics (MD) simulations,^{10,38–43} we have successfully developed the first SLEF force field (SLEF1) to describe zinc interactions compatible with the amber99SB force field^{44–46} and the TIP3P⁴⁷ water model and demonstrated its good transferability and adequacy in describing a variety of zinc proteins.

2. Methods

A. Nonbonded SLEF Force Field to Model Zinc Interactions. In the current work, we have introduced the following novel short–long effective function (SLEF) to describe charge interactions between a divalent zinc ion *i* and any other atom *j*:

$$E_{\text{es,SLEF}}^{\text{Zn},j}(r_{ij}) = \frac{1}{4\pi\epsilon_0} \left\{ \frac{q_{\text{Zn}}q_j}{\sqrt{r_{ij}^2 + \alpha \times \frac{q_j^2}{(R_i^* + R_j^*)} \times \exp(\beta \times r_{ij}^2)}} + \frac{1}{1 + \exp(-2\left(\frac{2r_{ij}}{3} - 1.0\right))} \times \frac{q_{\text{Zn}}q_j}{r_{ij}} \right\} \quad (1)$$

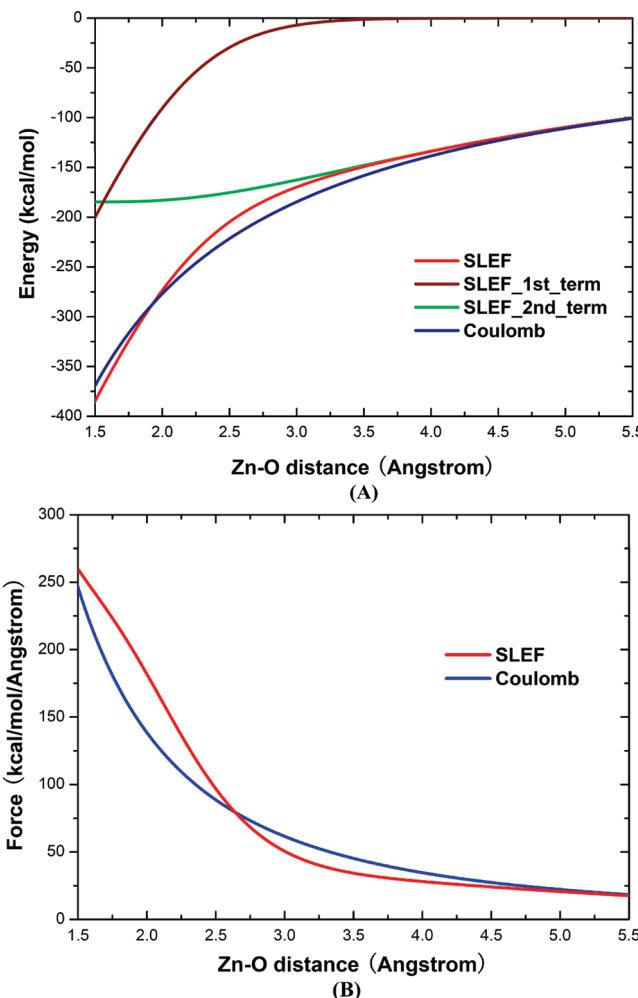


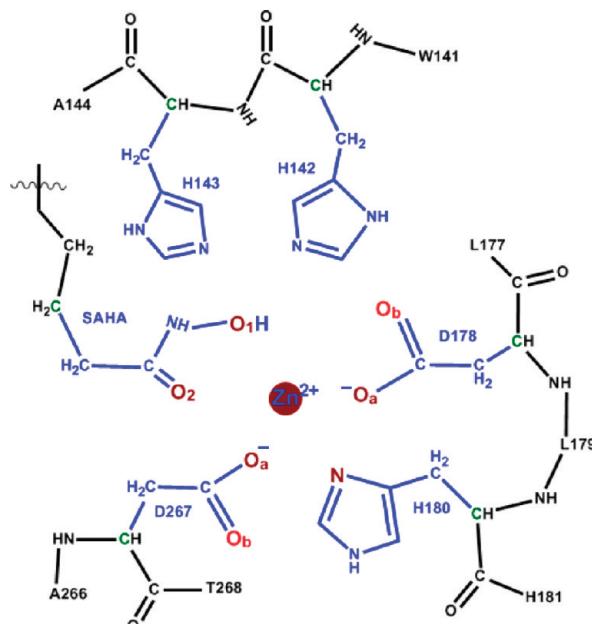
Figure 1. Illustration of the difference between SLEF and the conventional 1/r Coulomb function in describing charge interactions between Zn²⁺ and the oxygen of TIP3P water: (A) energy and (B) force. The parameters in the SLEF1 force field were employed.

where r_{ij} is the distance, q_{Zn} refers the charge of the zinc ion which has a value of 2.0, q_j is the MM charge of the atom *j*, R^* refers to the vdW radii, and α and β are two new positive parameters which need to be determined. As shown in Figure 1, the first term only makes a contribution at the short range, while the second term employing a similar damping function used in DFT dispersion correction approach^{48,49} is relatively flat in the short range but turns into 1/r in the long range (>4.5 Å). Thus the main difference between our introduced SLEF function and the coulomb function form 1/r is at the short-range, where the coordination interaction is expected to be dominant.

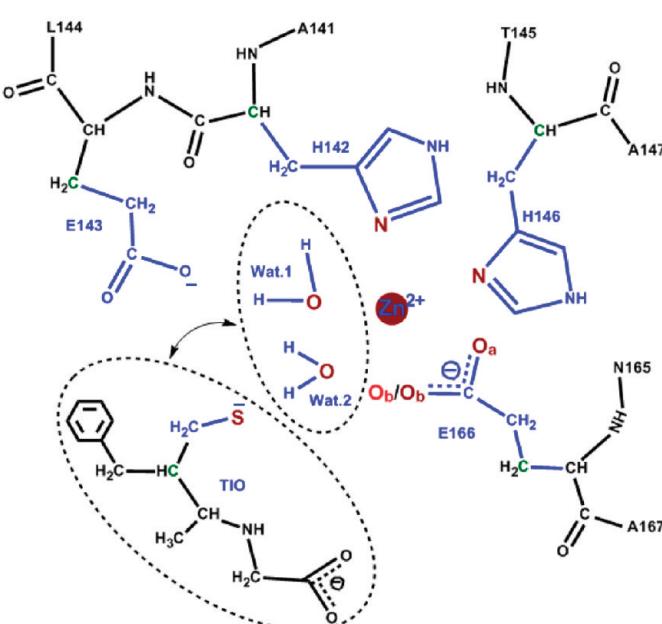
Besides the charge-interaction term, the conventional Lennard-Jones 12–6 function form has been employed to describe the vdW interactions between a zinc ion *i* and any other atom *j*:

$$E_{\text{vdW}}^{\text{Zn},j}(r_{ij}) = \epsilon_{ij} \{(R_{ij}^*/r_{ij})^{12} - 2(R_{ij}^*/r_{ij})^6\}, \quad R_{ij}^* = R_i^* + R_j^*, \quad \epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad (2)$$

We can see that the above SLEF approach to describe zinc interactions has the following desired features: (1) It is



(a) model 1: HDAC8-SAHA



(b) model 2: TLN-apo & model 3: TLN-TIO

Figure 2. Illustration of various coordination shells in three zinc enzyme systems in our training set used for optimizing SLEF1 force field parameters. Those atoms selected for force matching include the zinc cation and directly/potentially coordinated atoms (colored in brown/red).

pairwise, additive, and compatible with widely used atomic pairwise force fields for modeling biomolecules; (2) It only changes interactions between the zinc ion and other atoms and does not affect force field parameters that model other interactions in the system; and (3) It is a nonbonded model that is inherently capable to describe different zinc ligands and coordination modes. Thus to extend the widely employed atomic pairwise force fields to simulate zinc metalloproteins with the SLEF approach, it only needs to determine four additional parameters: α and β in the SLEF function (eq 1) and two zinc vdW parameters: ϵ and R^* (eq 2).

B. Parameterization with Force Matching Based on Ab initio QM/MM MD Simulations. Force matching (FM)^{34–37} has become a powerful and increasingly popular approach to parametrize atomic force fields based on high-level quantum mechanical calculations. Here we have adapted the ab initio QM/MM force matching approach^{36,37} to determine the four parameters (two vdW parameters for Zn, ϵ and R^* and two parameters in SLEF function, α and β) by minimizing the following target function:

$$\chi^2 = \sum_I \sum_J \sum_k \|f_{I,J,k}^{\text{SLEF}} - F_{I,J,k}^{\text{ref}}\|^2$$

where $F_{I,J,k}^{\text{ref}}$ refers the reference force from ab initio QM/MM calculations on the k atom with the J_{th} configuration of the I enzyme system in the training set, and $f_{I,J,k}^{\text{SLEF}}$ is the corresponding force calculated based on the SLEF force field.

In the current study, our training set consists of three zinc enzyme systems: HDAC8-SAHA,⁵⁰ TLN-apo,⁵¹ and TLN-TIO,⁵² as shown in Figure 2. These three systems represent five-, six-, and four-fold zinc coordination, respectively, and the ligands are typical in zinc proteins: His, Glu/Asp, Cys, hydroxamate, and water. All chosen configurations are

snapshots from Born–Oppenheimer ab initio QM/MM MD simulations, as described in detail in our previous work.¹⁰ For each enzyme system, 25 ps B3LYP(SDD,⁵³6-31G*) QM/MM MD simulations had been carried out, and 200 snapshots from the last 20 ps have been chosen for force matching. This level of QM treatment has been extensively tested and employed successfully to describe the zinc coordination shell^{10,43,54–57} and is similar to other recent ab initio QM/MM studies of zinc enzymes.^{58,59} A total of 600 configurations have been employed in parametrization with the amber99SB force field^{44–46} for modeling proteins and the TIP3P⁴⁷ water model. For each configuration, the reference forces on selected atoms, including the zinc cation, all directly and potentially coordinated atoms (illustrated in Figure 2), have been calculated by performing two B3LYP-(SDD,6-31G*) QM/MM calculations. One calculation is on the whole system, and the other is on the same system without the zinc ion. The force difference between two calculations can be considered as the force coming from its interaction with zinc and has been employed as the reference force $F_{I,J,k}^{\text{ref}}$. Correspondingly, the $f_{I,J,k}^{\text{SLEF}}$ is calculated with the SLEF1 force field. The advantage of employing this force difference is that the parametrization of the four parameters would be solely dependent on zinc interactions, which is and should be much desired. In addition, such an ab initio QM/MM force matching approach allows us to employ a large amount of information from a first principle description of zinc interactions, while properly taking account of the heterogeneous enzyme environment and the dynamic fluctuations. All ab initio QM/MM calculations were performed with modified Q-Chem⁶⁰ and Tinker⁶¹ programs, and the QM/MM boundaries were described by the pseudobond approach^{62–65} with the improved parameters.⁶²

Table 1. Resulting Four Parameters of the SLEF1 Force Field to Model Zinc Interactions^a

α	β	R^*	ε
2.23	1.04	1.21	0.23

^a Units: α , Å³/e²; β , 1.04 Å⁻²; R^* , Å; and ε , kcal/mol; α and β are parameters in the SLEF function (eq 1) and R^* and ε are vdW parameters of zinc.

The four parameters, including two vdW parameters ε and R^* for zinc and two parameters α and β in the SLEF function (eq 1), were determined by the parameter scan combined with local minimization procedure to effectively explore the parameter space. Specifically, ε value has been scanned from 0.05 to 0.50 with 0.01 step size and the other three parameters are optimized at each scan step. The simplex algorithm⁶⁶ implemented in GNU Scientific Library (GSL) and the modified Tinker program⁶¹ were employed in the parametrization procedure. The resulting four parameters for this new SLEF1 force field describing the zinc interactions compatible with the amber99SB force field^{44–46} and the TIP3P⁴⁷ water model are listed in Table 1.

C. Tests. We have implemented the new SLEF1 force field in the modified Tinker program.⁶¹ In order to examine its transferability and performance, we have carried out extensive MD simulations of seven zinc enzyme complexes with different coordination ligands (Asp/Glu, His, Cys, water, and small molecules) and distinct chelation modes (four-, five-, and six-fold), including the binuclear zinc active site. Besides three systems in the training set, as illustrated in Figure 2, the four additional models are: (A) an HDAC8-substrate complex system⁶⁷ which has a five-fold coordinated zinc catalytic site; (B) an HDAC7-SAHA complex⁶⁸ which has two four-fold coordinated zinc binding sites, one catalytic site and one Cys-rich structural site; (C) a carbonic anhydrase (CAII) enzyme system⁶⁹ which has a tetrahedral coordinated

zinc catalytic site; and (D) an L-rhamnose isomerase enzyme⁷⁰ containing a binuclear zinc coordination shell.

For each enzyme system, the initial structure was prepared based on the corresponding crystal structure.^{67–70} Then 4 ns MD simulations with the SLEF1 force field describing zinc interactions were carried out at 300 K with a time step of 1 fs. Amber99SB force field^{44–46} was used for protein residues, TIP3P model⁴⁷ for water molecules, and generalized AMBER force field (GAFF)⁷¹ for the other small molecules. The 18 and 12 Å cutoffs were employed for electrostatic and vdW interactions. For comparison, 4 ns MD simulations with the conventional nonbonded zinc model²⁵ (called as the Coulomb scheme) have also been performed.

3. Results

A. Performance of the SLEF1 Force Field on Three Zinc Enzymes in the Training Set. By optimizing two SLEF parameters as well as zinc vdW parameters through force matching based on ab initio QM/MM MD simulations, we have successfully developed the first SLEF force field to describe zinc interactions compatible with the amber99SB force field and the TIP3P water model. The four parameters are presented in Table 1, with α and β as 2.23 Å³/e² and 1.04 Å⁻², respectively, and the vdW parameters of Zn are $R^*=1.21$ Å; $\varepsilon=0.23$ kcal/mol.

The force errors on the selected ligand atoms from the SLEF1 force field as well as other MM models and QM/MM calculations with different basis sets (denoted as “DBS”) for three zinc enzymes in the training set are summarized in Tables 2–4. Not surprisingly, the Coulomb scheme, in which Stote’s parameters²⁵ for zinc and Amber99SB force field for other atoms were used, gives the largest force errors for each model. For the vdW FM scheme, in which vdW parameters of selected atoms (Zn and four kinds of ligand–atom: His–N, water–O, Glu/Asp–O, S) were optimized by FM,

Table 2. Force Error Calculated for the HDAC8-SAHA System with a Pentacoordinated Zinc Binding Site^a

HDAC8-SAHA(model 1)	Zn (total force)	rms force error (kcal/mol/Å)			
		H180 (N)	D178 (O _a /O _b)	D267 (O _a /O _b)	SAHA (O ₁ /O ₂)
Coulomb + LJ-R _{12,6} (Zn, Stote) ^b (Coulomb scheme)	57.8	11.8	56.7/18.3	30.5/18.2	34.7/21.2
Coulomb + LJ- R _{12,6} (Zn+Ligands, FM) ^c (vdW FM scheme)	48.7	14.5	19.2/15.7	50.1/21.3	19.1/17.1
SLEF(α ; β) + LJ- R _{12,6} (Zn, FM) ^d (SLEF scheme)	23.2	9.5	24.7/10.4	11.1/7.8	14.1/9.1
Different Basie Set in QM/MM ^e (DBS)	3.4	4.0	2.4/2.5	2.6/2.0	5.0/1.8

^a The reference forces are calculated with B3LYP(SDD,6-31G*) QM/MM calculations. ^b Stote’s vdW parameters²⁵ for the zinc ion: $R^*=1.09$ and $\varepsilon=0.25$. ^c Used the Coulomb function to describing charge interactions; vdW parameters of the zinc ion and the 4 types of coordinated atoms (a total of 10) were optimized by force matching. ^d Using the developed SLEF1 force field. (a total of 4 parameters have been optimized: $\alpha=2.23$; $\beta=1.04$; $\sigma=1.21$; and $\varepsilon=0.23$). ^e DBS indicates the force difference derived from using different basis sets (DBS) in QM/MM calculations (level 1: SDD for zinc, other atoms by 6-31G*; and level 2: 6-311G** for all atoms).

Table 3. Force Error Calculated for the TLN-Apo System with a Hexacoordinated Zinc Binding Site^a

TLN-apo (model 2)	Zn (total force)	rms force error (kcal/mol/Å)				
		H142 (N)	H146 (N)	E166 (O _a /O _a)	water1 (O)	water2 (O)
Coulomb + LJ-R _{12,6} (Zn, Stote) ^b (Coulomb scheme)	29.3	10.1	22.3	30.9/39.4	57.1	56.2
Coulomb + LJ- R _{12,6} (Zn+Ligands, FM) ^c (vdW FM scheme)	15.9	5.7	10.6	19.0/5.1	16.7	15.7
SLEF(α ; β) + LJ- R _{12,6} (Zn, FM) ^d (SLEF scheme)	18.4	6.7	12.4	9.3/12.6	18.7	18.3
Different Basie Set in QM/MM ^e (DBS)	2.5	3.9	3.6	1.7/0.9	4.0	3.8

^a The reference forces are calculated with B3LYP(SDD,6-31G*) QM/MM calculations. For other descriptions see Table 2.

Table 4. Force Error Calculated for the TLN-TIO System with a Tetraordinated Zinc Binding Site^a

TLN-TIO (model 3)	Zn (total force)	rms force error (kcal/mol/Å)			
		ligand atom (the force derived from Zn)			
		H142 (N)	H146 (N)	E166 (O _a /O _b)	TIO (S)
Coulomb + LJ-R _{12,6} (Zn, Stote) ^b (Coulomb scheme)	47.3	18.6	13.6	35.5/15.2	43.6
Coulomb + LJ- R _{12,6} (Zn+Ligands, FM) ^c (vdW FM scheme)	29.6	13.2	10.6	37.5/14.7	26.8
SLEF(α ; β) + LJ- R _{12,6} (Zn, FM) ^d (SLEF scheme)	26.1	10.3	7.9	13.3/11.0	9.5
Different Basie Set in QM/MM ^e (DBS)	11.5	5.6	5.5	8.1/3.2	10.6

^a The reference forces are calculated with B3LYP(SDD,6-31G*) QM/MM calculations. For other descriptions see Table 2.

the force errors are reduced for all models. The SLEF scheme gives the smallest force errors overall among the three MM schemes. It should be noted that there are a total of 10 parameters optimized in the vdW FM scheme, while only 4 parameters optimized in the SLEF1 force field. As shown in Tables 2 and 4, the force errors of several reference atoms in the vdW FM scheme are significantly larger than those in the SLEF1 force field, such as those of Zn and D267(O_a/O_b) in the model 1 which has a five-fold zinc coordination and E166 (O_a/O_b) and TIO (S) in the model 3 which has a four-fold zinc coordination. Therefore, the SLEF function plays an important contribution to decrease the force errors for four- and five-fold zinc coordination shells. As a result, we found that MD simulations with the vdW FM scheme could not reproduce the similar zinc coordination as observed in crystal structures and in ab initio QM/MM MD simulations for models 1 and 3, while the SLEF1 force field yields good results in MD simulations of all three systems, as shown in Figure 3. These results lend further support for our working hypothesis that the difficulty of the conventional nonbonded zinc model in describing zinc-coordination may come from the $1/r$ function form for the charge interaction term. Meanwhile, we can see that the error with the SLEF1 force field is still significantly larger than the DBS error, which indicates that there is significant room to further improving the description of zinc interactions.

The test results of the amber99SB-SLEF1 force field in describing the zinc coordination shell for three enzymes in the training set are presented in Figure 3. We can see that for all three systems, simulations with the amber99SB-SLEF1 force field yield zinc coordination geometries consistent with both experimentally determined X-ray structures and ab initio QM/MM simulation results. On the other hand, for the conventional nonbonded model with the $1/r$ form for zinc charge interactions, it yields very different coordination geometries. Meanwhile, with the corresponding crystal structure as the reference, we can see that the root-mean-square deviation (rmsd) of heavy atoms in the first zinc coordination shell is significantly smaller for simulations with the amber99SB-SLEF1 force field.

Model 1 (HDAC8-SAHA System). As shown in Figure 3, a five-fold coordination geometry is observed in the crystal structure⁵⁰ and has been reproduced well by ab initio QM/MM MD simulations.¹⁰ A similar coordination structure as well as important hydrogen-bond interactions between SAHA and His142/His143 are also well maintained in our simulations with the new SLEF1 force field. But for the conventional Coulomb scheme, it yields a six-fold coordination geometry for zinc and significantly changes the active site

geometry. Due to the wrong bidentate chelation of the two Asp residues, O₂ of SAHA is no longer bonded with zinc in simulations with the conventional nonbonded zinc model, and there is no hydrogen bond between SAHA and His142.

Model 2 (TLN-apo System). The six-fold zinc coordination structure⁵¹ was reproduced well in our previous QM/MM simulations and with the SLEF scheme. But for simulations with the conventional Coulomb scheme, the coordination interaction between His142 and Zn was replaced by another water molecule instead, and its Zn coordination shell is much different from the crystal structure, which is also demonstrated by the rmsd curve. It should be noted that the flexible behavior of Glu166 observed in our ab initio QM/MM MD simulations was not be observed in our simulations with the amber99-SLEF1 force field, which indicates its limitation.

Model 3 (TLN-TIO System). Both ab initio QM/MM MD simulations¹⁰ and the SLEF scheme yielded a similar tetrahedral coordination geometry as in the crystal structure.⁵² But the Glu166 was bidentate with zinc in the Coulomb scheme, resulting in a five-fold zinc coordination geometry. Meanwhile, the Zn–S coordination bond, which is obviously too short in the Coulomb scheme (2.02 Å), is significantly improved in simulations with the SLEF1 force field (2.24 Å).

B. Tests of the Transferability of SLEF1 on Other Zinc Enzyme Models. To further test the transferability and the performance of the resulting amber99SB-SLEF1 force field, we have further carried out MD simulations on four additional zinc enzyme systems including: Model A, a HDAC8–substrate complex system which has a five-fold coordinated zinc catalytic site; Model B, a HDAC7–SAHA complex which has two four-fold coordinated zinc binding sites—one catalytic and one Cys-rich structural site; Model C, a carbonic anhydrase (CAII) enzyme system which has a four-fold coordinated zinc catalytic site; and Model D, a L-rhamnose isomerase enzyme containing a binuclear zinc coordination site. The results are presented in Figures 4–7. We can see that simulations with the SLEF scheme yield zinc coordination geometries consistent with crystal structures and are significantly better than those results from MD simulations using the conventional Coulomb scheme. The various coordination numbers observed in crystal structures are well maintained in simulations with the SLEF scheme, while the Coulomb scheme tends to yield higher coordination numbers for most zinc coordination shells.

Model A (HDAC8–Substrate). As shown in Figure 4, Model A is a substrate-bound HDAC8 system. Our previous QM/MM simulations⁴³ yielded a five-fold zinc coordination shell, consistent with the observations from the crystal

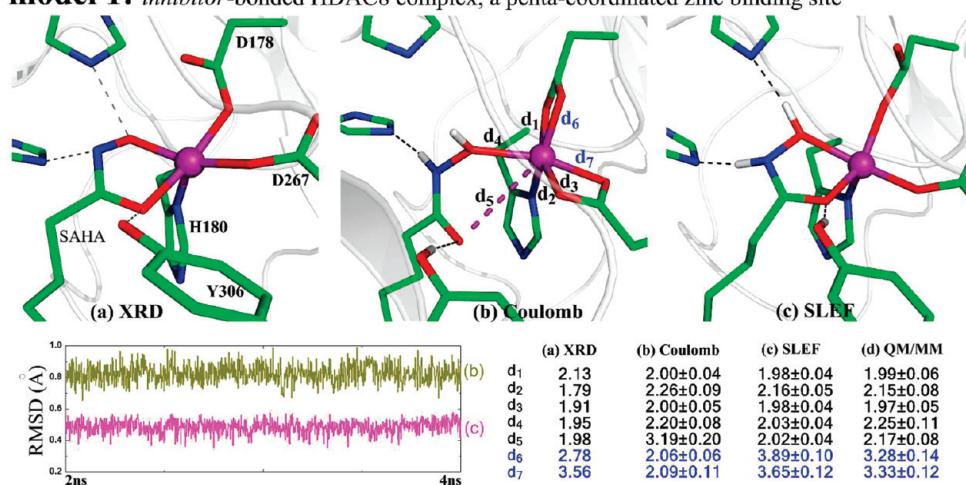
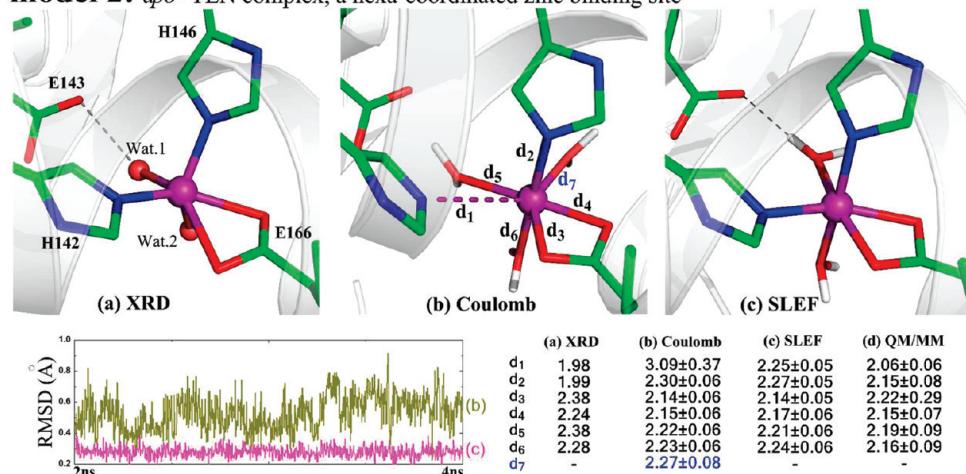
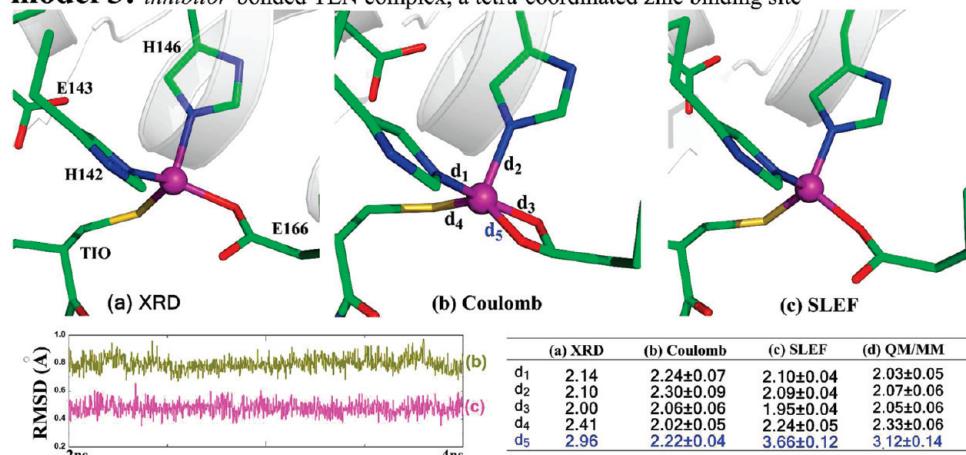
model 1: inhibitor-bonded HDAC8 complex, a penta-coordinated zinc binding site**model 2:** apo-TLN complex, a hexa-coordinated zinc binding site**model 3:** inhibitor-bonded TLN complex, a tetra-coordinated zinc binding site

Figure 3. Test results on the three zinc enzyme systems in the training set. XRD refers to results in crystal structures;^{50–52} Coulomb refers to results calculated from 4 ns MD simulations with the amber99SB force field and the nonbonded Coulomb model for zinc; and SLEF denotes results from 4 ns MD simulations with the amber99SB force field and our parametrized SLEF1 model for zinc interactions. QM/MM indicates the results from 25 ps B3LYP(SDD, 6-31G*) QM/MM MD simulations.¹⁰

structure.⁶⁷ With the SLEF scheme, both the coordination number and the important hydrogen-bonds around the p-53 peptide substrate are kept very well during the MD simulation. On the other hand, for simulations with the Coulomb scheme, although the rmsd value is also small, the penta-coordinated structure is not maintained due to the bidentate

chelation of Asp178, and there is no hydrogen bond between Y306 and the p-53 peptide.

Model B (HDAC7–SAHA). In comparison with the Model 1 (HDAC8–SAHA) complex in the training set, the coordination residues and inhibitors are the same, but different SAHA–zinc chelation modes have been observed in crystal

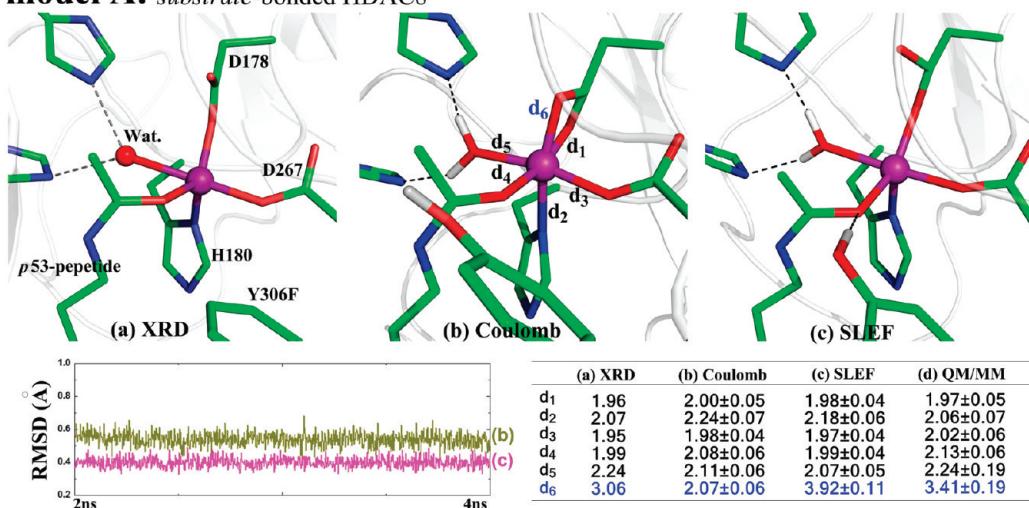
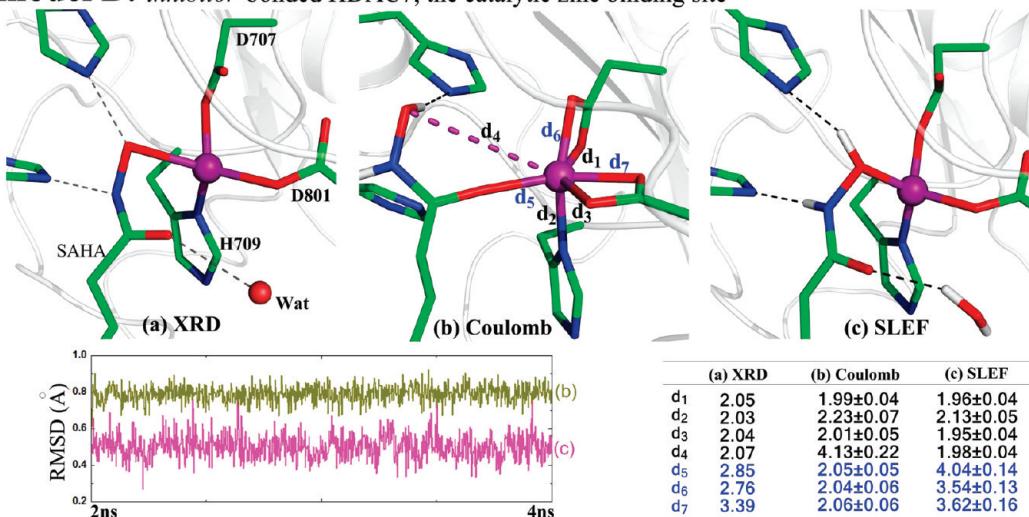
model A: substrate-bonded HDAC8

Figure 4. Test results on the Model A system. XRD refers to the crystal structure.⁶⁷ QM/MM indicates the results from 25 ps B3LYP(SDD, 6-31G*) QM/MM MD simulations.⁴³ For other descriptions see Figure 3.

model B: inhibitor-bonded HDAC7, the catalytic zinc binding site

the structural zinc binding site in HDAC7 (distances observed in crystal structure are shown in bracket)

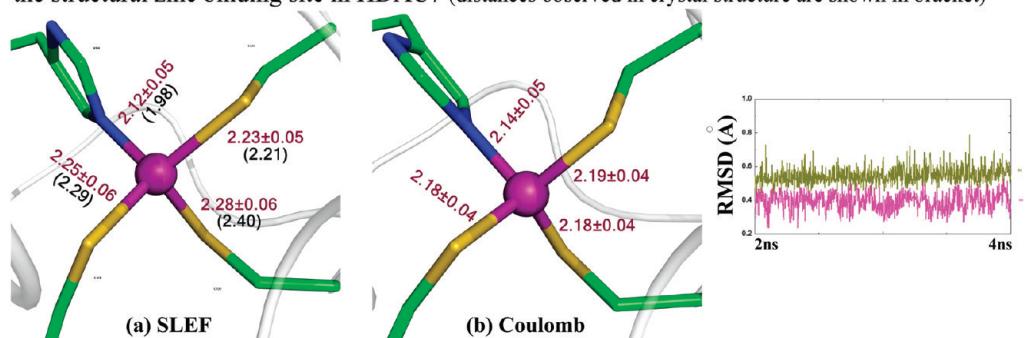


Figure 5. Test results on the Model B system. XRD refers to the crystal structure.⁶⁸ For other descriptions see Figure 3.

structures: monodentate and a four-fold zinc coordination shell in HDAC7,⁶⁸ while bidentate and a five-fold coordination in HDAC8.⁵⁰ Such a distinct coordination mode has also been confirmed by ab initio QM/MM MD simulations and thus serves as a stringent test for the force field description. From Figure 5, we can see that MD simulations with the SLEF scheme yield a zinc coordination shell consistent with the X-ray structure and ab initio QM/MM MD

simulations, maintain the important hydrogen-bond network, and show a smaller rmsd value. On the other hand, the conventional Coulomb scheme leads to a six-fold coordinated structure.

Besides the catalytic zinc site, there is another Cys-rich structural zinc coordination motif in HDAC7. The SLEF scheme can describe this coordination shell very well, as shown in Figure 5. Although the Coulomb scheme can also

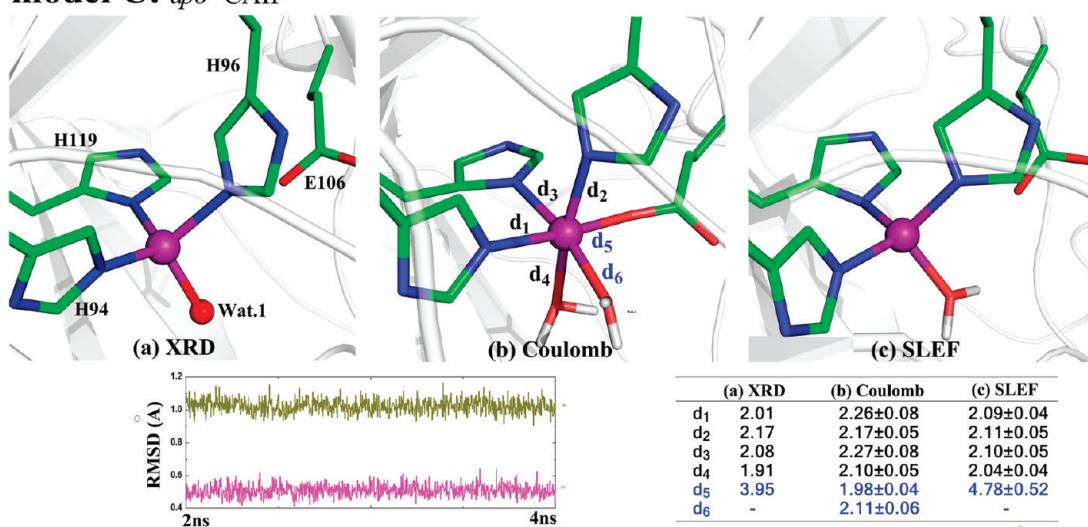
model C: apo-CAII

Figure 6. Test results on the Model C system. XRD refers to the crystal structure.⁶⁹ For other descriptions see Figure 3.

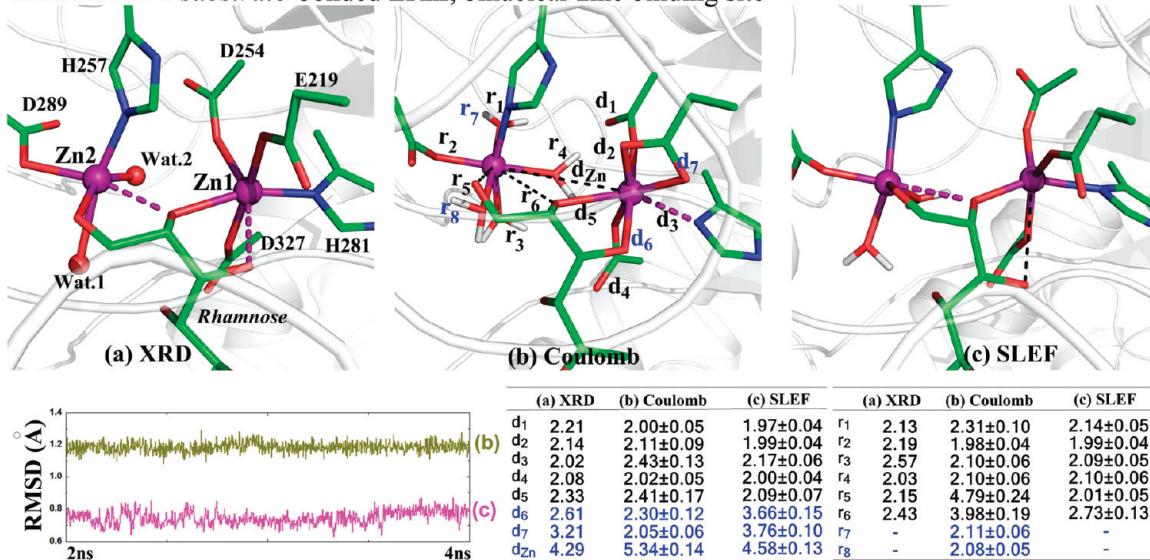
model D: substrate-bonded LRhI, binuclear zinc binding site

Figure 7. Test results on the Model D system. XRD refers to the crystal structure.⁷⁰ For other descriptions see Figure 3.

obtain the correct coordination number, the Zn–S coordination distance is very short.

Model C (CAII-Apo). The apo structure of carbonic anhydrase (CAII) has a tetrahedral zinc active site,⁶⁹ as shown in Figure 6. Herein our SLEF scheme also reproduces the four-fold coordination shell very well, but the Coulomb scheme leads to the hexacoordination. It seems that Coulomb scheme overestimates the electrostatic interaction between E106 and the divalent zinc cation, which is ~4 Å apart in the crystal structure.⁶⁹

Model D (L-RhI). L-Rhamnose isomerase, which can efficiently catalyze the isomerization between various aldoses and ketoses, has a binuclear zinc coordination shell.⁷⁰ Although no such binuclear zinc active site has been employed in the parameter optimization, the resulted SLEF1 force field can describe this challenging case⁷² relatively well, including the zinc–zinc distance. As seen in Figure 7, it improved significantly against the conventional coulomb

scheme in terms of both the zinc coordination spheres and the rmsd from the crystal structure. Meanwhile, these test results indicate that the SLEF1 force field still needs to be further improved, and a binuclear zinc active site should also be included in the training set in the future development.

4. Discussion

The above tests clearly demonstrated that the conventional coulomb scheme has two main deficiencies in describing zinc coordinations: (1) Its strong preference of water coordination and bidentate chelation of Asp/Glu residues leads to higher coordination numbers for most zinc coordination shells; and (2) its coordination to the neutral His residue can be substituted by a water molecule or a carboxylate ligand, such as in Model 2. Both deficiencies have been overcome by our developed SLEF1 force field, which has yielded zinc coordination structures in very good agreement with the corresponding crystal structures as well as ab initio QM/

Table 5. Mean and Maximum Deviation between the Coordination Distances in Crystal Structures and those from SLEF1 MD Simulations^a

	mean deviation (Å)	maximum deviation(Å)
negative ligands	-0.09	+0.07/-0.24
neutral ligands	0.04	+0.41/-0.48

^a The coordination distances from the SLEF1 MD simulations are the averaged values from the MD trajectories. The coordination distances in seven zinc enzyme complexes (training and test models) are all considered.

MM MD results. As summarized in Table 5, the mean deviation between the coordination distances in crystal structures and the corresponding average MD value from our SLEF1 simulations is 0.04 and -0.09 Å, respectively, and the largest deviation is +0.41/-0.48 Å, which is from the binuclear zinc binding site of L-RhI. For all seven zinc enzyme complexes, the coordination modes observed in crystal structures are well reproduced in simulations with the SLEF1 force field. In particular, for HDAC8-SAHA (Model 1) and HDAC7-SAHA (Model B) systems, their coordination ligands are the same, but different coordination modes have been observed in crystal structures: a four-fold zinc coordination shell in HDAC7,⁶⁸ while a five-fold coordination in HDAC8.⁵⁰ Such two distinct coordination modes with the same coordination ligands have been well reproduced in our simulations with the SLEF1 force field. Meanwhile, this would pose a fundamental challenge for bonded models^{19–24} to describe zinc interactions, which usually assumes that the same set of coordination ligands would adopt the same coordination mode.

In comparison to the conventional Coulomb function $1/r$, the key difference of our introduced SLEF function is in the short range (<4.5 Å), where the coordination interaction is expected to be dominant. From Figure 1, we can see that the resulted energy and force from the SLEF approach do not parallel those from the Coulomb function in the short-range regime. Meanwhile, the difference of the SLEF function to $1/r$ is varied with the magnitude of charges since the charge also appears in the denominator of the short-range function in eq 1. Thus, the SLEF function cannot be considered as simply scaling the charge in the short-range regime and then returning to the $1/r$ form in the long range.

Since the SLEF approach is a nonbonded model to model zinc interactions and only changes interactions between the zinc ion and other atoms, it would be quite straightforward to implement it into typical MD simulation packages: the replacement of the Coulomb function with the SLEF function to describe charge interactions between the zinc ion and all other atoms, and the employment of vdW parameters developed here for the zinc ion.

5. Conclusion

In this work, we have introduced a novel practical strategy to meet the challenge of describing zinc interactions: the design of new short-long effective functions (SLEF) to treat charge interactions between the zinc ion and all other atoms. By optimizing a total of four parameters based on ab initio QM/MM MD simulations and force matching, we have

developed the first transferable nonbonded pairwise SLEF force field to describe zinc interactions for modeling zinc metalloproteins compatible with the amber99SB force field and the TIP3P water model. We have carried out MD simulations with the amber99SB-SLEF1 force field on seven different enzymes complexes (a total of nine zinc coordination shells), which include four common kinds of ligands (His, Asp/Glu, Cys, and Water) and various coordination numbers (4, 5, or 6). Most simulations yielded zinc coordination numbers and binding distances in very good agreement with the corresponding crystal structures as well as ab initio QM/MM MD results. These very encouraging results indicate that this novel SLEF approach is a promising and attractive direction to explore for further improving force field description of metalloproteins.

Acknowledgment. This work was supported by NIH (R01-GM079223), NSF (CHE-CAREER-0448156), and the China Scholarship Council. We thank NCSA and NYU-ITS for providing computational resources and Dr. Shenglong Wang for computing support.

References

- (1) Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. *J. Proteome Res.* **2006**, *5*, 196–201.
- (2) Parkin, G. *Chem. Rev.* **2004**, *104*, 699–767.
- (3) Anzellotti, A. I.; Farrell, N. P. *Chem. Soc. Rev.* **2008**, *37*, 1629–1651.
- (4) Maret, W.; Li, Y. *Chem. Rev.* **2009**, *109*, 4682–4707.
- (5) Sensi, S. L.; Paoletti, P.; Bush, A. I.; Sekler, I. *Nat. Rev. Neurosci.* **2009**, *10*, 780–791.
- (6) Auld, D. S. *Biometals* **2001**, *14*, 271–313.
- (7) Lee, Y. M.; Lim, C. J. *Mol. Biol.* **2008**, *379*, 545–553.
- (8) Tamames, B.; Sousa, S. F.; Tamames, J.; Fernandes, P. A.; Ramos, M. J. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 466–475.
- (9) McCall, K. A.; Huang, C. C.; Fierke, C. A. *J. Nutr.* **2000**, *130*, 1437S–1446S.
- (10) Wu, R.; Hu, P.; Wang, S.; Cao, Z.; Zhang, Y. *J. Chem. Theory Comput.* **2010**, *6*, 337–343.
- (11) Kuppuraj, G.; Dudev, M.; Lim, C. J. *Phys. Chem. B* **2009**, *113*, 2952–2960.
- (12) Babor, M.; Greenblatt, H. M.; Edelman, M.; Sobolev, V. *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 221–230.
- (13) Banci, L. *Curr. Opin. Chem. Biol.* **2003**, *7*, 143–149.
- (14) Donini, O. A. T.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 4180–4188.
- (15) Koca, J.; Zhan, C. G.; Rittenhouse, R. C.; Ornstein, R. L. *J. Comput. Chem.* **2003**, *24*, 368–378.
- (16) Dal Peraro, M.; Spiegel, K.; Lamoureux, G.; De Vivo, M.; DeGrado, W. F.; Klein, M. L. *J. Struct. Bio.* **2007**, *157*, 444–453.
- (17) Zimmer, M. *Coord. Chem. Rev.* **2009**, *293*, 817–826.
- (18) Li, X.; Hayik, S. A.; Merz, K. M. *J. Inorg. Biochem.* **2010**, *104*, 512–522.
- (19) Vedani, A.; Huhta, D. W. *J. Am. Chem. Soc.* **1990**, *112*, 4759–4767.

- (20) Hoops, S. C.; Anderson, K. W.; Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 8262–8270.
- (21) Bredenberg, J.; Nilsson, L. *Int. J. Quantum Chem.* **2001**, *83*, 230–244.
- (22) Li, W. F.; Zhang, J.; Wang, J.; Wang, W. *J. Am. Chem. Soc.* **2008**, *130*, 892–900.
- (23) Lin, F.; Wang, R. X. *J. Chem. Theory Comput.* **2010**, *6*, 1852–1870.
- (24) Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 2935–2947.
- (25) Stote, R. H.; Karplus, M. *Proteins: Struct., Funct., Bioinf.* **1995**, *23*, 12–31.
- (26) Pang, Y. P. *Proteins: Struct., Funct., Bioinf.* **2001**, *45*, 183–189.
- (27) Pang, Y. P.; Xu, K.; El Yazal, J.; Prendergast, F. G. *Protein Sci.* **2000**, *9*, 1857–1865.
- (28) Sakharov, D. V.; Lim, C. *J. Am. Chem. Soc.* **2005**, *127*, 4921–4929.
- (29) Sakharov, D. V.; Lim, C. *J. Comput. Chem.* **2009**, *30*, 191–202.
- (30) de Courcy, B.; Piquemal, J. P.; Gresh, N. *J. Chem. Theory Comput.* **2008**, *4*, 1659–1668.
- (31) Gresh, N.; Piquemal, J. P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113–1130.
- (32) Wu, J. C.; Piquemal, J. P.; Chaudret, R.; Reinhardt, P.; Ren, P. Y. *J. Chem. Theory Comput.* **2010**, *6*, 2059–2070.
- (33) Elstner, M.; Cui, Q.; Munih, P.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Comput. Chem.* **2003**, *24*, 565–581.
- (34) Ercolessi, F. *Europhys. Lett.* **1994**, *26*, 583–588.
- (35) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896–10913.
- (36) Akin-Ojo, O.; Song, Y.; Wang, F. *J. Chem. Phys.* **2008**, *129*, 064108.
- (37) Maurer, P.; Laio, A.; Hugosson, H. W.; Colombo, M. C.; Rothlisberger, U. *J. Chem. Theory Comput.* **2007**, *3*, 628–639.
- (38) Hu, P.; Wang, S.; Zhang, Y. *J. Am. Chem. Soc.* **2008**, *130*, 3806–3813.
- (39) Hu, P.; Wang, S.; Zhang, Y. *J. Am. Chem. Soc.* **2008**, *130*, 16721–16728.
- (40) Ke, Z.; Zhou, Y.; Hu, P.; Wang, S.; Xie, D.; Zhang, Y. *J. Phys. Chem. B* **2009**, *113*, 12750–12758.
- (41) Wang, S.; Hu, P.; Zhang, Y. *J. Phys. Chem. B* **2007**, *111*, 3758–3764.
- (42) Zhou, Y.; Wang, S.; Zhang, Y. *J. Phys. Chem. B* **2010**, *114*, 8817–8825.
- (43) Wu, R.; Wang, S.; Zhou, N.; Cao, Z.; Zhang, Y. *J. Am. Chem. Soc.* **2010**, *132*, 9471.
- (44) 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. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (45) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (46) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
- (47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (48) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- (49) Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515–524.
- (50) Somoza, J. R.; Skene, R. J.; Katz, B. A.; Mol, C.; Ho, J. D.; Jennings, A. J.; Luong, C.; Arvai, A.; Buggy, J. J.; Chi, E.; Tang, J.; Sang, B. C.; Verner, E.; Wynands, R.; Leahy, E. M.; Dougan, D. R.; Snell, G.; Navre, M.; Knuth, M. W.; Swanson, R. V.; McRee, D. E.; Tari, L. W. *Structure* **2004**, *12*, 1325–1334.
- (51) Holland, D. R.; Hausrath, A. C.; Juers, D.; Matthews, B. W. *Protein Sci.* **1995**, *4*, 1955–1965.
- (52) Roderick, S. L.; Fourniezeluski, M. C.; Roques, B. P.; Matthews, B. W. *Biochemistry* **1989**, *28*, 1493–1497.
- (53) Dolg, M.; Wedig, U.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1987**, *86*, 866–872.
- (54) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Biophys. J.* **2005**, *88*, 483–494.
- (55) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *J. Am. Chem. Soc.* **2007**, *129*, 1378–1385.
- (56) Xiao, C.; Zhang, Y. *J. Phys. Chem. B* **2007**, *111*, 6229–6235.
- (57) Corminboeuf, C.; Hu, P.; Tuckerman, M. E.; Zhang, Y. *J. Am. Chem. Soc.* **2006**, *128*, 4530–4531.
- (58) Blumberger, J.; Lamoureux, G.; Klein, M. L. *J. Chem. Theory Comput.* **2007**, *3*, 1837–1850.
- (59) Xu, D. G.; Guo, H. *J. Am. Chem. Soc.* **2009**, *131*, 9780–9788.
- (60) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J., O. C., Brown, S. T., Gilbert, A. T., Slipchenko, L. V., Levchenko, S. V., O'Neill, D. P., DiStasio, R. A., Lochan, R. C., Wang, T., Beran, G. J., Besley, N. A., Herbert, J. M., Lin, C. Y., Van Voorhis, T., Chien, S. H., Sodt, A., Steele, R. P., Rassolov, V. A., Maslen, P. E., Korambath, P. P., Adamson, R. D., Austin, B., Baker, J., Byrd, E. F., Dachsel, H., Doerksen, R. J., Dreuw, A., Dunietz, B. D., Dutoi, A. D., Furlani, T. R., Gwaltney, S. R., Heyden, A., Hirata, S., Hsu, C. P., Kedziora, G., Khaliulin, R. Z., Klunzinger, P., Lee, A. M., Lee, M. S., Liang, W., Lotan, I., Nair, N., Peters, B., Proynov, E. I., Pieniazek, P. A., Rhee, Y. M., Ritchie, J., Rosta, E., Sherrill, C. D., Simmonett, A. C., Subotnik, J. E., Woodcock, H. L., Zhang, W., Bell, A. T., Chakraborty, A. K., Chipman, D. M., Keil, F. J., Warshel, A., Hehre, W. J., Schaefer, H. F., Kong, J., Krylov, A. I., Gill, P. M., Head-Gordon, M. *Q-Chem*, version 3.0; Q-Chem, Inc.: Pittsburgh, PA, 2006.
- (61) Ponder, J. W. *TINKER, Software Tools for Molecular Design*, version 4.2; 2004.
- (62) Zhang, Y. *J. Chem. Phys.* **2005**, *122*, 024114.
- (63) Zhang, Y. *Theor. Chem. Acc.* **2006**, *116*, 43–50.
- (64) Zhang, Y.; Lee, T. S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46–54.
- (65) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483–3492.
- (66) Nelder, J. A.; Mead, R. *Comput. J.* **1965**, *7*, 308–313.

- (67) Vannini, A.; Volpari, C.; Gallinari, P.; Jones, P.; Mattu, M.; Carfi, A.; De Francesco, R.; Steinkuhler, C.; Di Marco, S. *Embo Rep.* **2007**, *8*, 879–884.
- (68) Schuetz, A.; Min, J.; Allali-Hassani, A.; Schapira, M.; Shuen, M.; Loppnau, P.; Mazitschek, R.; Kwiatkowski, N. P.; Lewis, T. A.; Maglathin, R. L.; McLean, T. H.; Bochkarev, A.; Plotnikov, A. N.; Vedadi, M.; Arrowsmith, C. H. *J. Biol. Chem.* **2008**, *283*, 11355–11363.
- (69) Budayova-Spano, M.; Fisher, S. Z.; Dauvergne, M. T.; Agbandje-McKenna, M.; Silverman, D. N.; Myles, D. A. A.; McKenna, R. *Acta Crystallographica Section F-Struc-*
tural Biology and Crystallization Communications **2006**, *62*, 6–9.
- (70) Yoshida, H.; Yamada, M.; Ohyama, Y.; Takada, G.; Izumori, K.; Kamitori, S. *J. Mol. Biol.* **2007**, *365*, 1505–1516.
- (71) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (72) Wu, R.; Xie, H.; Mo, Y.; Cao, Z. *J. Phys. Chem. A* **2009**, *113*, 11595–11603.

CT100525R