

Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems

Fu Lin and Renxiao Wang*

State Key Laboratory of Bioorganic Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, People's Republic of China

Received August 27, 2009

Abstract: Metal ions are indispensable for maintaining the structural stability and catalytic activity of metalloproteins. Molecular modeling studies of such proteins with force fields, however, are often hampered by the “missing parameter” problem. In this study, we have derived bond-stretching and angle-bending parameters applicable to zinc-containing systems which are compatible with the AMBER force field. A total of 18 model systems were used to mimic the common coordination configurations observed in the complexes formed by zinc-containing metalloproteins. The Hessian matrix of each model system computed at the B3LYP/6-311++G(2d,2p) level was then analyzed by Seminario’s method to derive the desired force constants. These parameters were validated extensively in structural optimizations and molecular dynamics simulations of four selected model systems as well as one protein–ligand complex formed by carbonic anhydrase II. The best performance was achieved by a bonded model in combination with the atomic partial charges derived by the restrained electrostatic potential method. After some minor optimizations, this model was also able to reproduce the vibrational frequencies computed by quantum mechanics. This study provides a comprehensive set of force field parameters applicable to a variety of zinc-containing molecular systems. In principle, our approach can be applied to other molecular systems with missing force field parameters.

1. Introduction

Metalloproteins (or metalloproteinases) are a family of proteolytic enzymes whose catalytic mechanism involves a metal. The metal ion is an indispensable component for maintaining their enzymatic catalysis as well as structural stability.¹ Many metalloproteins are zinc-dependent. The zinc ion in such proteins often functions as a Lewis acid for the stabilization of reactants/intermediates or the occurrence of catalytic reactions. Although it is reported that zinc may adopt other types of coordination geometries,^{2,3} it normally adopts a tetrahedral coordination geometry, in which the zinc ion is linked to the protein via three coordination bonds, while the fourth position is occupied by a labile water molecule or a bound ligand molecule. Matrix metalloproteinase (MMP), carbonic anhydrase (CA), alcohol dehydrogenase (AD), zinc-finger proteins are some well-studied zinc-

containing (Zn-containing) proteins. They play an essential role in the biosynthesis and metabolism of certain bioactive peptides and are relevant to a variety of critical diseases, including arthritis and cancer.^{4,5} For example, MMPs conduct hydrolysis of the amide bonds on certain peptide substrates with the conserved zinc ion in the catalytic site. It has been demonstrated that MMPs regulate degradation of the extracellular matrix and control of angiogenesis, and thus selective inhibitors against MMPs may be used as promising anticancer therapies.⁶

Due to the important biological implications of Zn-containing proteins, molecular modeling is often employed to study the structures and functions of these proteins. Although today’s computers are really powerful, modeling Zn-containing proteins in solvent with high-level quantum mechanics (QM) computations is still not quite possible. Thus, molecular mechanics (MM) is still the dominant approach for such tasks, although some combined QM/MM models^{7–10} have been proposed as well. Unfortunately, most

* Corresponding author. Telephone: 86-21-54925128. E-mail: wangrx@mail.sioc.ac.cn.

today's force fields do not always have appropriate parameters for metal atoms, which has become a practical obstacle for the molecular modeling studies of metalloproteins.

Researchers have developed various methods for tackling this "missing parameter" problem regarding metal atoms, in particular zinc. In general, there are three options: nonbonded, semibonded, and bonded models. The nonbonded model relies basically on electrostatic and van der Waals interactions instead of covalent bonds to maintain the coordination configuration of zinc ion during simulation.¹¹ It is straightforward to incorporate a nonbonded model into an established force field. However, such a model could be sensitive to the choice of atomic charge models. Due to the long-range nature of electrostatic forces, the zinc ion tends to get close to any negatively charged amino acid residues. Consequently, the zinc ion may have problems in retaining a low coordination number or even escapes from the coordination center.¹² Another problem with this type of models is that they cannot take account for charge transfer and polarization effects very well.¹³

Semibonded models were originally proposed by Pang et al.,^{14,15} which are interesting patches to the nonbonded models. A semibonded model places four dummy atoms around the zinc ion, which are covalently connected to the zinc ion in a tetrahedral geometry. The zinc ion is assigned only van der Waals parameters, and its +2e charge is evenly distributed among the four dummy atoms. Interactions between the dummy atoms and amino acid residues are then computed using the conventional electrostatic interaction term. Semibonded models are also relatively convenient to be incorporated into an established force field. Compared to the nonbonded models, they are more suitable for modeling the tetrahedral coordination configuration of a Zn-containing system. However, they basically share the same shortcomings as nonbonded models; they are sensitive to the choice of atomic charge models and cannot be applied to other tasks, such as normal-mode analysis. How to properly set the parameters for the connections between zinc and dummy atoms is another matter of concern.

The bonded models¹⁶ treat the connections between zinc and its ligands as covalent bonds. An obvious advantage of such models is that they can preserve the tetrahedral coordination configuration of zinc even in long-time simulations. If necessary, they can also be applied to other possible coordination configurations of zinc. A disadvantage of bonded models is that it is not convenient to use them to simulate the interconversions between different coordination configurations, since connection tables are kept fixed during simulation. Nevertheless, this is normally not a concern for molecular modeling studies of Zn-containing proteins. So far some researchers have derived force field parameters applicable to Zn-containing systems for bonded models through various approaches.^{17–24} A common practice is to perform frequency analyses on Zn-containing model systems by high-level quantum mechanics computations, and then the diagonal elements in the resulting Hessian matrix are taken as the desired force constants of the bonds or the angles in which zinc is participated. This approach requires the Hessian matrix to be given in internal coordinates, which is

relatively complicated. Besides, a problem observed with this approach is that different settings of internal coordinates may lead to different results.²⁵

Seminario et al. proposed an alternative method^{25–29} by which a force constant may be derived from a Hessian matrix based on the Cartesian coordinates. This method retrieves the 3×3 submatrices relevant to the atom pairs of interests from a given Hessian matrix. Then, the force constant for any internal coordinate (bond stretching, angle bending, and torsional angle) can be derived from the eigenvalues and eigenvectors of these submatrices after some mathematical transformations. Results produced by this method are obviously independent of the choice of internal coordinates. In their original study,²⁵ Seminario et al. has applied this method to some very simple molecules, such as water and nitrogen dioxide. Later, Ryde et al.²⁹ employed this method to derive force field parameters which were used for the refinement of crystal structures of metal-containing enzymes, such as ferrochelatase and iron superoxide dismutase. Bautista et al.²⁶ employed this method to derive force field parameters applicable to polyalanine peptides. Collectively, these studies have demonstrated the value of Seminario's method. Expanding the application of this method to more challenging problems is certainly intriguing.

In our study, we adopted Seminario's method to derive force field parameters applicable to various Zn-containing molecular complexes. These parameters, including equilibrium bond lengths and bond angles as well as force constants of bond stretching and angle bending, are compatible with the popular AMBER force field.³⁰ These parameters were validated on four simple Zn-containing model systems. Our results indicate that application of these parameters to the AMBER force field well reproduced the three-dimensional structures and vibrational frequencies of these model systems. These parameter were further applied to the molecular dynamics simulations on one complex structure formed by carbonic anhydrase II and produced encouraging results. We expect that the parameters derived in our study are applicable to the refinement of crystal and NMR structures of Zn-containing metalloproteins and the molecular modeling studies on the binding of such proteins to their ligand molecules. Compared to other researchers' previous studies, our study provides a more comprehensive set of force field parameters for a variety of Zn-containing complexes so that they can be readily applied without additional adjustment or optimization. More importantly, our study has demonstrated a complete approach to the deduction and validation of force field parameters with Seminario's method. The application of this approach is certainly not limited to Zn-containing molecular systems.

2. Methods

In our study, a total of 18 molecular complexes were used to mimic the typical situations in the binding of Zn-containing proteins with their ligand molecules. Force field parameters relevant to zinc were derived from the outcomes of QM computations on these model systems with Seminario's method. The derived parameters were then evaluated on four model systems to see if they could reproduce the

three-dimensional structures and vibrational frequencies of these model systems. They were also evaluated in molecular dynamics (MD) simulations of one complex structure formed by carbonic anhydrase II, a Zn-containing metalloprotein. QM computations were performed by using the Gaussian 03 software.³¹ All of the other major computations, including energy minimization, MD simulation, and frequency analysis described in the following sections, were performed by using the AMBER software (version 9)³² on a Linux cluster based on Intel Xeon 5345 processors.

2.1. Selection of Zinc-Containing Model Systems and QM Computations. The entire Protein Data Bank (PDB)³³ released by January 1, 2009, consisting of about 55 000 structures, was screened with an in-house computer program to retrieve the Zn-containing metalloproteins of our interests. Only the metalloproteins containing one zinc ion inside the binding pocket and one bound small-molecule ligand, i.e. Zn-containing protein–ligand complexes, were considered during this process. In addition, the zinc ion must be in contact with at least four nonhydrogen atoms within a distance cutoff of 2.8 Å, among which at least one had to be on the ligand molecule. Only crystal structures with overall resolution equal to or better than 2.5 Å were considered in order to impose a control on the quality of these complex structures. The total number of the Zn-containing protein–ligand complexes meeting the above criteria was 1004.

A survey on these complex structures revealed that in most cases the zinc ion was bound with three His residues on the protein side, although Cys, Glu, and Asp residues were observed in some cases. For the sake of convenience, we used three imidazole molecules to mimic the side chains of three His residues on the protein side. On the ligand side, a variety of chemical moieties were found in direct bonding with the zinc ion. Accordingly, a total of 12 small molecules (M-1 to M-12 in Figure 1) were used as models in our study, which covered the majority of such moieties identified in our survey. Six additional model systems (M-13 to M-18 in Figure 1) were considered to represent other mixed coordination centers, in which an acetic acid molecule was used to mimic the side chain of an Asp/Glu residue and an methanethiol molecule was used to mimic the side chain of a Cys residue. In fact, M-1, M-9, and M-10 also can be considered as mixed coordination centers. Therefore, the outcomes of our study can be applied to the modeling of a wider range of Zn-containing proteins.

QM computations were then performed on the 18 model systems summarized in Figure 1. These model systems were optimized by using the B3LYP method,^{34–38} a popular density functional theory (DFT) method, with the 3-21G, 6-31+G(d, p), and then 6-311++G(2d, 2p) basis sets in a stepwise manner. Finally, frequency analysis was performed at the B3LYP/6-311++G (2d, 2p) level to confirm that the optimized structure was a true energy minimum without any imaginary frequency. The frequency analysis also produced the necessary raw data required by the following step.

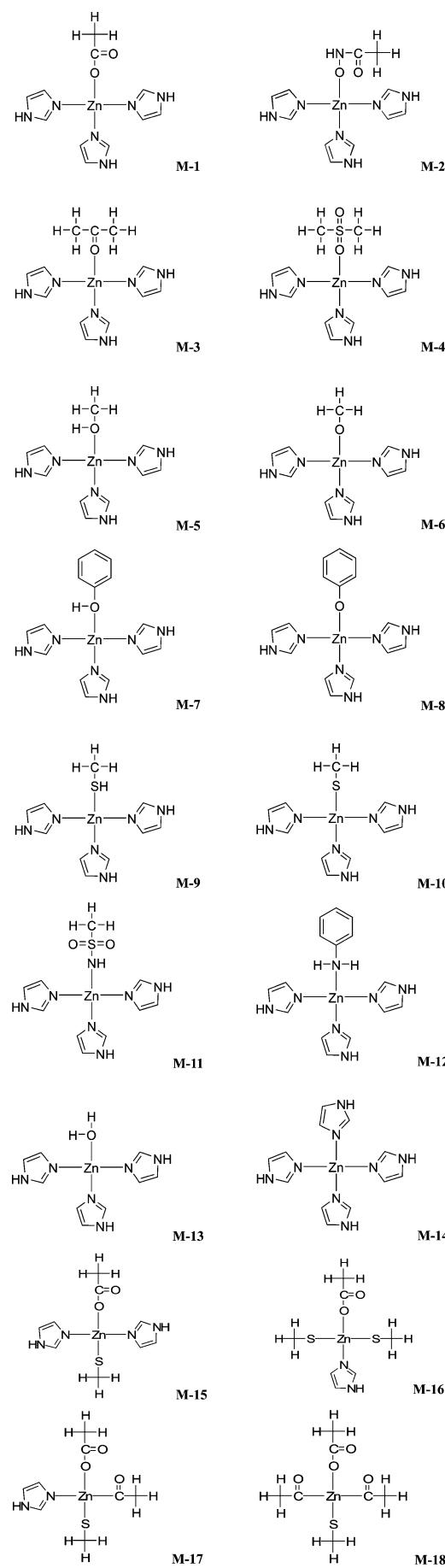


Figure 1. Eighteen Zn-containing model systems considered in this study.

2.2. Derivation of Zinc-Related Force Field Parameters. We aimed at deriving parameters for zinc that are compatible with the AMBER force field. The potential energy function used by AMBER is

$$E = \sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{torsions}} \frac{V_n}{2}(1 + \cos[n\phi - \gamma]) + \sum_{i < j}^{\text{atoms}} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right) + \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{\epsilon R_{ij}} \quad (1)$$

The five terms in the above equation compute the energies of bond stretching, angle bending, torsion angles and nonbonded van der Waals and electrostatic interactions, respectively. Detailed explanations on the parameters in the above equation can be found elsewhere.³⁰ Note that in the current AMBER force field implemented in the AMBER software package (version 9)³² only the van der Waals parameters of nonbonded model for zinc are provided. In this study, we derived bond-stretching and angle-bending parameters of bonded model for zinc. The equilibrium length of each bond and the equilibrium value of each bond angle in which zinc participated were obtained directly from the three-dimensional structures of the 18 model systems listed in Figure 1, which were fully optimized with extensive QM computations. The force constants of bond stretching and angle bending were derived with the method proposed by Seminario et al. A brief description of this method is given below for the sake of readers. More details can be found in the original reference.²⁵

In order to derive the force constants of bond stretching and angle bending required in eq 1, a Hessian matrix in Cartesian coordinates was extracted from the outcomes of frequency analysis of each model system:

$$[\delta F] = -[k] \times [\delta x] \quad (2)$$

Here, $[k]$ denotes for the Hessian matrix of a system composed of N atoms, $[\delta x]$ denotes for the vector of the displacements in Cartesian coordinates, and $[\delta F]$ denotes for the vector of resulting reaction forces. The full form of eq 2 is

$$\begin{bmatrix} \delta F_1 \\ \delta F_2 \\ \delta F_3 \\ \vdots \\ \delta F_{3N} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 E}{\partial x_1^2} & \frac{\partial^2 E}{\partial x_1 \partial x_2} & \frac{\partial^2 E}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 E}{\partial x_1 \partial x_{3N}} \\ \frac{\partial^2 E}{\partial x_2 \partial x_1} & \frac{\partial^2 E}{\partial x_2^2} & \frac{\partial^2 E}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 E}{\partial x_2 \partial x_{3N}} \\ \frac{\partial^2 E}{\partial x_3 \partial x_1} & \frac{\partial^2 E}{\partial x_3 \partial x_2} & \frac{\partial^2 E}{\partial x_3^2} & \cdots & \frac{\partial^2 E}{\partial x_3 \partial x_{3N}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial x_{3N} \partial x_1} & \frac{\partial^2 E}{\partial x_{3N} \partial x_2} & \frac{\partial^2 E}{\partial x_{3N} \partial x_3} & \cdots & \frac{\partial^2 E}{\partial x_{3N}^2} \end{bmatrix} \times \begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \delta x_3 \\ \vdots \\ \delta x_{3N} \end{bmatrix} \quad (3)$$

According to the Seminario's method, the bond-stretching force constant of bond A–B can be derived from the Hessian matrix as a 3×3 matrix, i.e. $[k_{AB}]$:

$$[\delta F_A] = -[k_{AB}] \times [\delta x_B] \quad (4)$$

$$\begin{bmatrix} \delta F_{Ax} \\ \delta F_{Ay} \\ \delta F_{Az} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 E}{\partial x_A \partial x_B} & \frac{\partial^2 E}{\partial x_A \partial y_B} & \frac{\partial^2 E}{\partial x_A \partial z_B} \\ \frac{\partial^2 E}{\partial y_A \partial x_B} & \frac{\partial^2 E}{\partial y_A \partial y_B} & \frac{\partial^2 E}{\partial y_A \partial z_B} \\ \frac{\partial^2 E}{\partial z_A \partial x_B} & \frac{\partial^2 E}{\partial z_A \partial y_B} & \frac{\partial^2 E}{\partial z_A \partial z_B} \end{bmatrix} \times \begin{bmatrix} \delta x_B \\ \delta y_B \\ \delta z_B \end{bmatrix} \quad (5)$$

The differential of force in eq 5, i.e., $[\delta F]$, represents the responding force on atom A due to a displacement in the coordinates of atom B. Diagonalization of the $[k_{AB}]$ matrix gives the eigenvalues λ_i^{AB} and the corresponding eigenvectors v_i^{AB} .

$$k_{AB} = \sum_{i=1}^3 \lambda_i^{AB} |u^{AB} \cdot v_i^{AB}| \quad (6)$$

Here, k_{AB} is the harmonic bond stretching force constant for bond A–B; u^{AB} is the normalized vector pointing from atoms A to B. It should be noted that $k_{AB} = 2K_r$ (K_r is the force constant of bond stretching used in eq 1).

Similarly, the angle-bending force constant k_θ for angle $\angle ABC$ can be derived by considering the responding forces on atoms A and C due to a displacement in the coordinates of atom B:

$$\begin{bmatrix} \delta F_{Ax} \\ \delta F_{Ay} \\ \delta F_{Az} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 E}{\partial x_A \partial x_B} & \frac{\partial^2 E}{\partial x_A \partial y_B} & \frac{\partial^2 E}{\partial x_A \partial z_B} \\ \frac{\partial^2 E}{\partial y_A \partial x_B} & \frac{\partial^2 E}{\partial y_A \partial y_B} & \frac{\partial^2 E}{\partial y_A \partial z_B} \\ \frac{\partial^2 E}{\partial z_A \partial x_B} & \frac{\partial^2 E}{\partial z_A \partial y_B} & \frac{\partial^2 E}{\partial z_A \partial z_B} \end{bmatrix} \times \begin{bmatrix} \delta x_B \\ \delta y_B \\ \delta z_B \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} \delta F_{Cx} \\ \delta F_{Cy} \\ \delta F_{Cz} \end{bmatrix} = - \begin{bmatrix} \frac{\partial^2 E}{\partial x_C \partial x_B} & \frac{\partial^2 E}{\partial x_C \partial y_B} & \frac{\partial^2 E}{\partial x_C \partial z_B} \\ \frac{\partial^2 E}{\partial y_C \partial x_B} & \frac{\partial^2 E}{\partial y_C \partial y_B} & \frac{\partial^2 E}{\partial y_C \partial z_B} \\ \frac{\partial^2 E}{\partial z_C \partial x_B} & \frac{\partial^2 E}{\partial z_C \partial y_B} & \frac{\partial^2 E}{\partial z_C \partial z_B} \end{bmatrix} \times \begin{bmatrix} \delta x_B \\ \delta y_B \\ \delta z_B \end{bmatrix} \quad (8)$$

Then, the angle-bending force constant can be derived as:

$$\frac{1}{k_\theta} = \frac{1}{d_{AB}^2 \sum_{i=1}^3 \lambda_i^{AB} |u^{PA} \cdot v_i^{AB}|} + \frac{1}{d_{CB}^2 \sum_{i=1}^3 \lambda_i^{CB} |u^{PC} \cdot v_i^{CB}|} \quad (9)$$

Here, d_{AB} and d_{CB} are the distances between atoms A–B and C–B, respectively; $u^{PA} = u_N \times u^{AB}$; $u^{PC} = u^{CB} \times u_N$; $u_N = (u^{CB} \times u^{AB}) / (|u^{CB} \times u^{AB}|)$; u^{AB} , u^{CB} , λ_i^{AB} , λ_i^{CB} , v_i^{AB} , and v_i^{CB} have similar meanings as in eq 6. Note that $k_\theta = 2K_\theta$ (K_θ is the angle-bending force constant in eq 1).

All of the 18 model systems summarized in Figure 1 were processed, as described above, using in-house computer programs. Note that, in principle, the harmonic force

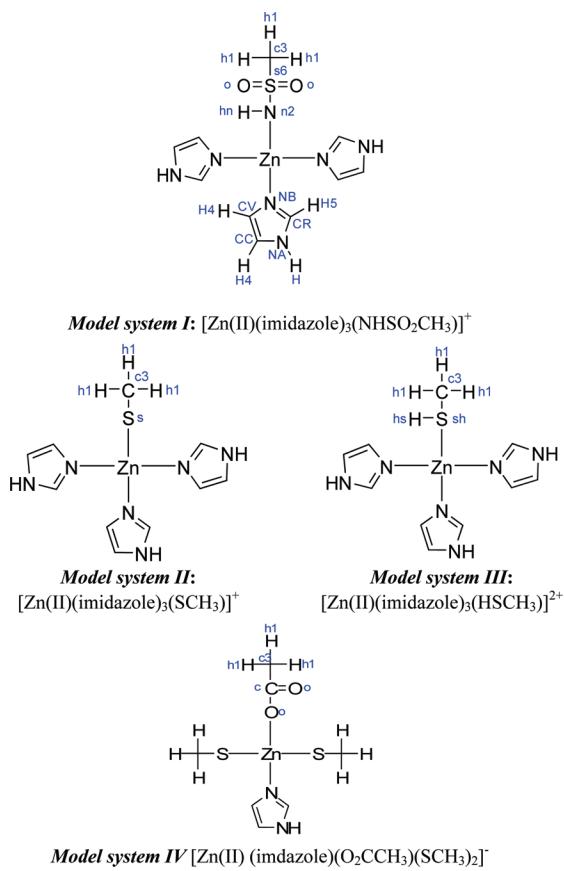


Figure 2. Four model systems used for the validation of force field parameters. The AMBER atom type of each atom is labeled in lower cases.

constants of torsion angles can also be obtained through a similar procedure. However, the harmonic force constants are not compatible with the Fourier form of the torsion energy term in the current AMBER force field. Thus, we did not attempt to derive parameters for torsion angles which include zinc as one of the four component atoms. Neglecting the contributions of such torsion angles is actually not a major problem. After all, the bonded model was adopted in our study for modeling Zn-containing systems. Due to the symmetric tetrahedral geometry of the zinc coordination center, chemical moieties in bonding with the zinc ion are quite rigid and usually devoid of significant torsional freedom.²⁰

2.3. Validation of the Derived Parameters on Model Systems. *Selection of the Model Systems and Force Field Models.* Four model systems, which were referred to as *model systems I–IV* (Figure 2), were selected out of the 18 typical Zn-containing model systems for validating the force field parameters derived in our study. The ligand molecule in *model system I* consists of a sulfonamide moiety, which is normally believed to be deprotonated upon coordination with zinc.³⁷ This moiety is observed in many ligands bound to Zn-containing proteins. For example, the core part of the complex formed between carbonic anhydrase II and 5-dimethylamino-naphtha-1-sulfonamide (PDB entry 1OKL), which was used as an example later in our study for validation purposes, has exactly the same chemicals structure as *model system I*. The ligand molecules in both *model*

systems II and *III* consist of a thiol moiety, another common moiety for forming coordination bonds with zinc. This moiety also mimics the side chain of a Cys residue. The difference between *model systems II* and *III* was that the thiol moiety in *system II* was sketched in the deprotonated form; while the counterpart in *system III* was sketched in the neutral form. *Model system IV* presents a mixed coordination center, in which the three residues in bonding with zinc includes one His and two Cys residues. The ligand molecule in this model system is an acetic acid molecule in the deprotonated form. It can be considered as an Asp or Glu residue as well. In addition, *model system IV* bears an overall negative charge, unlike the other three model systems (Figure 2).

The performance of five force field models, including both bonded and nonbonded models (Table 1), were then evaluated on all four model systems. For the three bonded models (FF-1, FF-2, and FF-5), the force field parameters for zinc derived in our study (Table 2) were applied. For the nonbonded models (FF-3 and FF-4), parameters for the residues on the protein side were taken from the AMBER FF03 parameter set,³⁹ while parameters for the ligand molecule ($\text{-NHSO}_2\text{CH}_3$, -SCH_3 , HSCH_3 , and $\text{-O}_2\text{CCH}_3$) were taken from the AMBER GAFF parameter set.^{40,41} The van der Waals parameters for zinc in all five force field models were set as: $\sigma = 1.10 \text{ \AA}$ and $\varepsilon = 0.0125 \text{ kcal/mol}$, which were cited from Merz's study.⁴²

Since the choice of appropriate atomic charges, which are required to compute electrostatic energies, is another common argument in modeling Zn-containing systems, both the bonded and nonbonded models were tested in combination with two atomic partial charge schemes (Figure 3). The first scheme employed the RESP method⁴³ to derive atomic partial charges on the entire molecular system, including zinc, from the outcomes of QM computations at the B3LYP/6-311++G(2d,2p) level. This task was conducted with the RESP fitting protocol implemented in the AMBER program. Note that all QM computations in our study were conducted at the B3LYP/6-311++G(2d,2p) level. Thus, we did not repeat our computations at the HF/6-31G(d) level, which are typically supplied to the RESP method as inputs, to avoid possible inconsistency at other aspects. This charge scheme will be referred to as the “RESP charges” throughout this article. In the second scheme, zinc was assigned a formal charge of +2e. The atomic charges on the bonding residues were taken from the “template charges” for His and Cys residues in the AMBER FF03 parameter set. For the small-molecule ligand, template atomic charges are not available in the AMBER force field. Thus, atomic charges on the ligand molecule ($\text{-NHSO}_2\text{CH}_3$, -SCH_3 , HSCH_3 , and $\text{-O}_2\text{CCH}_3$) were also derived from the outcomes of QM computations at the B3LYP/6-311++G(2d,2p) level by using the RESP method. This scheme is in fact the standard practice employed by most common users of the AMBER program in the molecular modeling studies of metal-containing systems, and it will be referred to as the “formal charges” throughout this article.

Structural Optimizations on Model Systems. The five force field models were applied to all four model systems first to test how well they could reproduce the structures of these

Table 1. Reproduction of the Structures of Four Zn-Containing Model Systems by Five Different Force Field Models

force field model			model system I			model system II			model system III			model system IV		
symbol	bonding model	atomic charges ^b	in vacuum ^c	in vacuum ^d	in water ^d	in vacuum ^c	in vacuum ^d	in water ^d	in vacuum ^c	in vacuum ^d	in water ^d	in vacuum ^c	in vacuum ^d	in water ^d
FF-1	bonded	RESP	0.053	0.053	0.054	0.107	0.107	0.090	0.079	0.076	0.071	0.216	0.205	0.159
FF-2	bonded	formal	0.656	0.604	0.554	0.141	0.141	0.143	0.096	0.096	0.535	0.396	0.383	0.356
FF-3	nonbonded	RESP	1.001	0.293	0.275	0.091	0.085	0.104	16.347	16.424	0.236	1.795	1.778	0.245
FF-4	nonbonded	formal	0.255	0.255	0.240	0.266	0.265	0.267	0.231	0.231	0.237	0.358	0.336	0.287
FF-5 ^e	bonded	RESP	0.054	0.053	0.056	0.107	0.107	0.090	0.083	0.076	0.071	0.218	0.207	0.161

^a Rmsd values were computed by considering the coordinates of zinc and the four atoms in direct bonding with zinc. The structure optimized at the B3LYP/6-311++G(2d, 2p) level was used as the reference. ^b See Figure 4. ^c Structural optimization started from an arbitrary structure. ^d Structural optimization started from a structure preoptimized at the B3LYP/6-311++G(2d, 2p) level. ^e A variant based on FF-1 which was specially optimized to better reproduce vibrational frequencies.

model systems preoptimized at the B3LYP/6-311++G(2d,2p) level. On each model system, the structural optimization was first started from the preoptimized structure by QM. To further test the robustness of the given force field models, the structural optimization was then repeated on an arbitrary structure of the same model system, in which the coordinates of every component atom were scrambled while the connection table was retained (Figure 4). All of the structural optimization computations were performed using the AMBER program. The Newton–Raphson method was applied to energy minimization. The convergence criterion was set to 10^{-8} kcal/(mol·Å). The distance cutoff of nonbonded interactions was set to 999 Å. In each case, the root-mean-squared deviation (rmsd) of the resulting structure was calculated by using the structure preoptimized by QM as the reference. Only zinc and the four atoms in direct bonding with zinc were considered in rmsd calculations.

Molecular Dynamics Simulations on Model Systems. The five force field models were also applied to the molecular dynamics (MD) simulations of all four model systems in vacuum. All MD simulations were also conducted using the AMBER program. In each case, the preoptimized structure by QM was used as the starting structure for the following MD simulation. In order to release the internal strain energies of the entire system gradually, three rounds of restraint MD simulations were carried out first: (1) a 50 ps long simulation with restraints on nonhydrogen atoms (restraint harmonic force constant = 5.0 kcal/mol·Å²); (2) a 50 ps long simulation with restraints on nonhydrogen atoms (restraint harmonic force constant = 0.5 kcal/mol·Å²); and then (3) another 50 ps long simulation without any restraint. After these preparations, the final production run lasted for 10 ns, which was conducted in vacuum under a constant temperature of 300 K. The distance cutoff of nonbonded interaction was set to 999 Å. The periodic boundary condition was not enabled during simulation. The time interval was set to 1 fs during the entire simulation process, and the MD trajectory was also recorded every 1 ps for subsequent analyses.

Since the force field parameters derived in our study may be applied to the modeling of Zn-containing metalloproteins in their physiological environment, the five force field models (Table 1) were also applied to the MD simulations of all four model systems in water. In each case, the preoptimized structure by QM was soaked in a TIP3P water box⁴⁴ with a margin of 14 Å in each dimension. The entire system was

neutralized by adding an appropriate number of counterions, and a three-step minimization was used to release internal strain energies gradually. In each step, 5000 rounds of minimization was performed with the restraint harmonic force constant imposed on all nonhydrogen atoms set to 500.0, 10.0 kcal/mol·Å², and zero, respectively. The entire system was then subjected to the same restraint MD simulation routine as the one performed in vacuum described in the previous paragraph. After all these preparative steps, a production simulation of 10 ns long was performed under constant temperature ($T = 300$ K) and pressure ($P = 1$ atm). Temperature of the entire system was regulated by Langevin thermostat⁴⁵ with the collision frequency $\gamma = 2.0$ ps⁻¹, and pressure of the system was controlled by Berendsen barostat.⁴⁶ The time interval was set to 1 fs. Periodic boundary condition was enabled during simulation. The distance cutoff for nonbonded interactions was 14 Å, and the particle mesh Ewald (PME) method⁴⁷ was used to compute long-range interactions. The MD trajectory was also recorded every 1 ps for subsequent analyses.

2.4. Validation of the Derived Parameters on a Carbonic Anhydrase II Complex. The force field parameters derived in our study were further validated on a complex structure formed by carbonic anhydrase II and 5-dimethylamino-naphthalene-1-sulfonamide (Figure 5). The complex structure was solved by Nair et al⁴⁸ through X-ray diffraction at a resolution of 2.10 Å (PDB entry: 1OKL). In this complex structure, the zinc ion inside the binding pocket is in coordination with three histidine residues (His90, His92, and His115) and with a sulfonamide moiety on the ligand molecule, a chemical configuration identical to *model system I*. Consequently, the force field parameters derived from *model system I* were applied in the following simulations.

Five separate MD simulations of this complex structure were performed to test force field models FF-1 to FF-5, respectively. To set up each simulation, the force field parameters for carbonic anhydrase II were taken from the AMBER FF03 parameter set, and those for the ligand molecule were taken from the AMBER GAFF parameter set. The van der Waals parameters for zinc were also set as: $\sigma = 1.10$ Å and $\epsilon = 0.0125$ kcal/mol. Note that, for all five force field models, the corresponding RESP charges or formal charges indicated in Figure 3 were applied only to the binding center, including the zinc ion, the imidazole moieties on His90/His92/His115, and the entire ligand

Table 2. Bond-Stretching and Angle-Bending Parameters Related to Zinc Derived from Four Zn-Containing Model Systems

bond type ^a	stretching force constant kcal/(mol·Å ²)	equilibrium bond length (Å)
<i>Model System I</i>		
ZN-NB	56.0	2.07
ZN-n2	98.8	1.97
<i>Model System II</i>		
ZN-NB	49.5	2.09
ZN-s	92.8	2.26
<i>Model System III</i>		
ZN-NB	74.8	2.02
ZN-sh	33.2	2.48
<i>Model System IV</i>		
ZN-NB	62.9	2.33
ZN-o	98.8	1.97
ZN-s	62.9	2.34
bond angle ^a	bending force constant kcal/(mol·rad ²)	equilibrium bond angle (°)
<i>Model System I</i>		
NB-ZN-NB	31.1	105.5
NB-ZN-n2	35.7	113.0
ZN-NB-CR	46.6	126.4
ZN-NB-CV	48.7	126.6
ZN-n2-hn	37.6	118.5
ZN-n2-s6	67.0	112.4
<i>Model System II</i>		
NB-ZN-NB	35.4	104.7
NB-ZN-s	27.6	113.8
ZN-NB-CR	49.4	126.2
ZN-NB-CV	49.4	127.0
ZN-s-c3	75.2	104.9
<i>Model System III</i>		
NB-ZN-NB	31.5	111.5
ZN-NB-CR	48.9	126.8
ZN-NB-CV	49.8	126.8
NB-ZN-sh	27.5	107.3
ZN-sh-hs	32.0	100.3
ZN-sh-c3	65.7	110.6
<i>Model System IV</i>		
CR-NB-ZN	43.2	123.21
CV-NB-ZN	43.9	129.78
NB-ZN-o	34.6	91.97
NB-ZN-s	32.7	97.52
o-ZN-s	26.3	113.89
s-ZN-s	21.6	129.12
c-o-ZN	60.7	119.82
c3-s-ZN	70.6	102.40

^a The atom types used in this table are indicated in Figure 2.

molecule. The rest of the parts on the complex were assigned the template charges from the AMBER FF03 parameter set.

This complex was soaked in a TIP3P water box with a margin of 10 Å and was then neutralized by adding counterions. The subsequent stepwise minimization and restraint MD simulations were performed using the same procedure and settings as the MD simulations on Zn-containing model systems in water. After these preparative steps, a production simulation of 10 ns long was performed under a constant temperature ($T = 300$ K) and a constant pressure ($P = 1$ atm). Temperature of the system was also regulated by Langevin thermostat⁴⁵ with the collision

frequency $\gamma = 2.0$ ps⁻¹, and pressure of the system was controlled by Berendsen barostat.⁴⁶ The time interval for MD simulation was set to 2 fs. Periodic boundary condition was enabled during simulation. The distance cutoff for nonbonded interactions was 12 Å, and the PME method⁴⁷ was used to compute long-range interactions. In addition, the SHAKE algorithm⁴⁹ was applied to constrain all bonds involving hydrogen atoms. The MD trajectory was also recorded every 1 ps for subsequent analyses.

An additional MD simulation was performed on the same complex structure, in which the binding center was modeled by the QM/MM method⁵⁰ implemented in the AMBER program. The semiempirical PM3 method⁵¹ was employed to treat the zinc ion, His90/His92/His115, and the entire ligand molecule. The SCF convergence was set to 10^{-8} kcal/mol. The rest parts of the complex structure were still treated with the AMBER force field using the FF03 parameter set. The complex structure was also soaked in a TIP3P water box with a margin of 10 Å and was then neutralized by adding counterions. The entire system was subjected to the same stepwise minimizations and preparative restraint MD simulations as the other force field models. The final production run lasted for 4 ns under a constant temperature ($T = 300$ K) and a constant pressure ($P = 1$ atm). The time interval for MD simulation was set to 2 fs. The MD trajectory was recorded every 1 ps. All of the other major parameters/settings were the same as those used in the simulations by using other force field models.

2.5. Further Optimization on Derived Parameters for Reproducing Vibrational Frequencies. Producing the correct vibrational frequencies is also an important quality of a good force field model. For each model system illustrated in Figure 2, the vibrational frequencies computed at the B3LYP/6-311++G(2d,2p) level were compared with their counterparts given by AMBER with the FF-1 parameters in normal-mode analysis. These two sets of vibrational frequencies actually fit very well for all four model systems (Figure 6). Some obvious discrepancy was observed only at the high-frequency end (frequency >3000 cm⁻¹). The normal modes given by QM computations on each model system were visually examined in the graphical user interface of the Gaussian 03 program to determine which were responsible for the observed discrepancy in vibrational frequencies. It turned out to be stretching of X–H bonds (X = C, N, O, or S). Accordingly, the bond-stretching force constants of the X–H single bonds in FF-1 were further optimized to better reproduce the vibrational frequencies of *model systems I–IV*, and the outcomes were named as FF-5. All of the other parameters in FF-5 were the same as those in FF-1.

The optimization was carried out through a genetic algorithm (GA) procedure implemented in an in-house computer program. The optimization was carried out on a population of 500 chromosomes. Every chromosome was composed of a certain number of genes, each of which was encoded with the bond-stretching force constant of a particular X–H bond (X = C, N, or S). The initial value of each gene was assigned the original value of the corresponding force constant in FF-1 plus a random perturbation within ±20%. The fitness score of each chromosome was computed

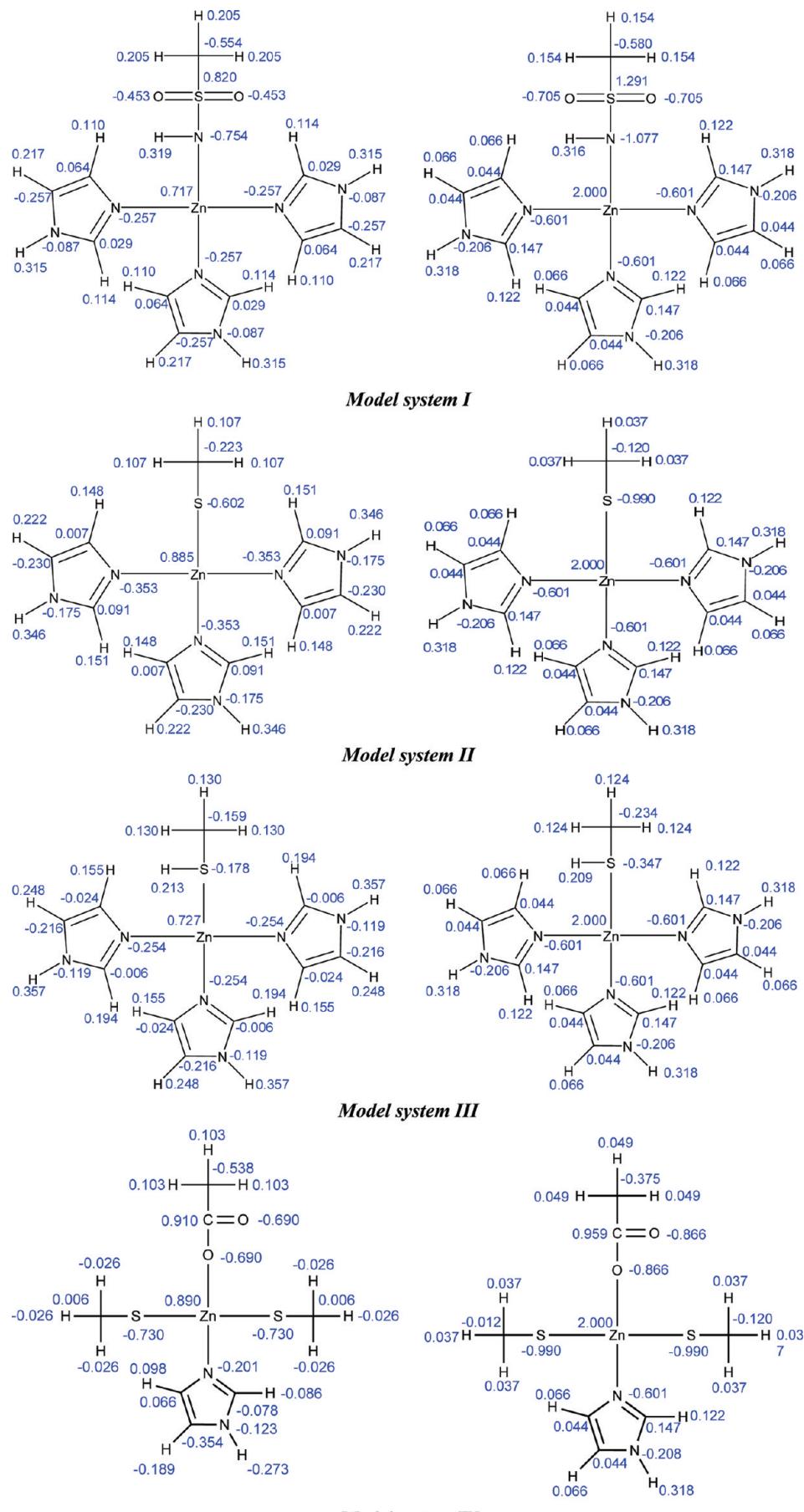


Figure 3. Atomic charges assigned on the four model systems when the RESP charge model was applied (left) or when a formal charge of +2e was applied to zinc (right).

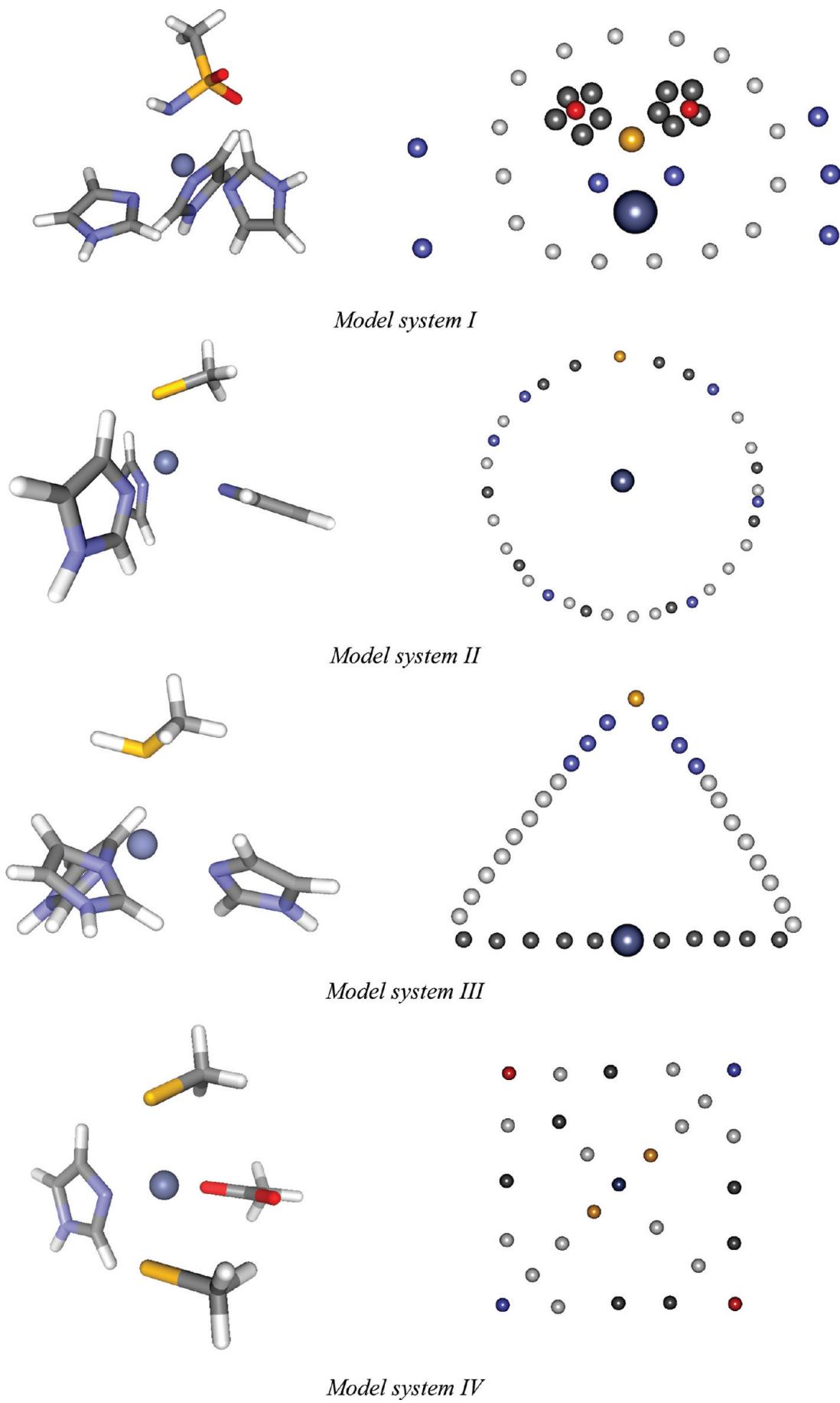


Figure 4. Three-dimensional structures of *model systems* I–IV optimized at the B3LYP/6-311++G(2d, 2p) level (left), and the corresponding arbitrarily sketched structures used for validation (right).

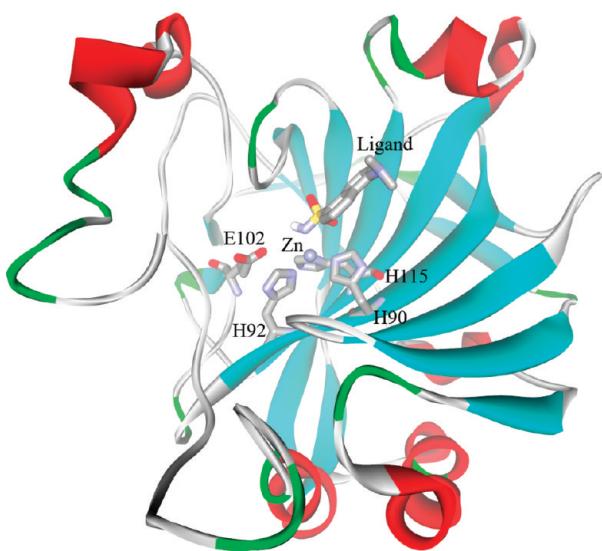


Figure 5. Crystal structure of carbonic anhydrase II in complex with 5-dimethylamino-naphtha-lene-1-sulfonamide (PDB entry: 1OKL), which was used for the validation of force field parameters in this study.

as the rmsd between the vibration frequencies given by QM computations and those given by the normal-mode analysis in AMBER applying the force constants encoded in the given chromosome. Thus, the smaller was its fitness score, the better was the given chromosome. The entire population undertook optimization in the steady-state mode⁵² for a total of 50 000 GA rounds. At each round, two types of genetic operations, including single-point mutation and crossover, may occur at a probability of 30% and 70%, respectively. The single-point mutation occurred randomly at a particular gene on a given parent chromosome in which that gene was altered randomly within $\pm 20\%$ of its original value. The single-point crossover occurred between two parent chromosomes in which the two chromosomes exchanged their genes starting from a random point. For both the mutation and the crossover operations, the parent chromosomes were selected using the roulette-wheel method,⁵² so that better chromosomes had a greater chance to produce offsprings. At each GA round, the newly generated chromosomes (one or two) were compared with the worst chromosome in the entire population. If the new one had better fitness scores, then they would replace the worst chromosome in the population; otherwise the population would remain as the same to enter next GA round. The average fitness score of the entire population was monitored along the entire GA process. For all four model systems, the average fitness score of the entire population actually reached convergence well before 50 000 GA steps. After the GA process was completed, the force constants encoded in the best chromosome were retrieved as the final parameters in FF-5.

To make comparison with other force field models, FF-5 was also applied to the structural optimizations and MD simulations on *model systems I–IV* as well as the MD simulations on the carbonic anhydrase II complex, using the same settings described earlier in this manuscript.

3. Results and Discussion

3.1. Validation of Force Field Models on Four Simple Model Systems. As described in the Methods Section, all five force field models (FF-1 to FF-5) were applied to the structural optimizations of *model systems I–IV* (Figure 2) in both vacuum and water. The rmsd values between the structures optimized by FF-1 to FF-5 and the corresponding structures optimized at the B3LYP/6-311++G(2d, 2p) level are summarized in Table 1. As one example, the three-dimensional structures of *model system I* optimized by these force field models in vacuum are illustrated in Figure 7.

One can see that FF-1 (bonded model + RESP charges) performed very well in reproducing the correct three-dimensional structures of all four model systems, no matter if the structural optimization was started from a preoptimized structure or a ridiculous structure. The rmsd values of the structures optimized by FF-1 are generally below 0.25 Å on all four model systems. In contrast, the performance of FF-2 (bonded model + formal charges) was generally inferior to that of FF-1. For example, FF-2 produced a relatively large rmsd value in handling the structure of *model system I* in both vacuum and water. This may be attributed to the strong secondary electrostatic interactions between the zinc ion (atomic charge = +2.000e) and the two oxygen atoms (atomic charge = -0.705e) on the sulfonamide moiety on the ligand molecule. Indeed, one can see in the structure produced by FF-2 (Figure 7) that the zinc ion tends to get closer to these two oxygen atoms and results in an obvious distortion of the tetrahedral geometry of the coordination center. Another disadvantage of FF-2 is that, unlike FF-1, its performance is somewhat unpredictable; its performance was not so good on *model system I*; whereas it was acceptable on *model system II*. As for *model system III*, in which the ligand molecule is in the neutral form, it produced reasonable structures in vacuum but not in water.

As for the two nonbonded models, FF-4 (nonbonded model + formal charges) demonstrated a relatively robust performance across all four model systems. The rmsd values of the structures produced by FF-4 are generally between 0.2–0.4 Å no matter if the optimization was performed in vacuum or water. This level of accuracy, however, is still inferior to the one produced by FF-1. The performance of FF-3 (nonbonded model + RESP charges) was not consistent across four model systems; it produced reasonable structures of *model system II*, whereas it had some obvious problems in reproducing the structures of I, III, and IV in vacuum. In particular, for *model system III*, in which the ligand molecule is in the neutral form, the electrostatic interaction between the zinc ion (charge = +0.727e) and the sulfur atom (charge = -0.178e) on the thiol group is not strong enough for maintaining these two atoms in a bonding range. The structure of this model system was basically disrupted upon optimization with FF-3, resulting in a large rmsd value of over 16 Å. These results suggest that nonbonded models, which rely only on electrostatic and van der Waals interactions for modeling the coordination configuration of zinc, are less capable than bonded models. If a nonbonded model

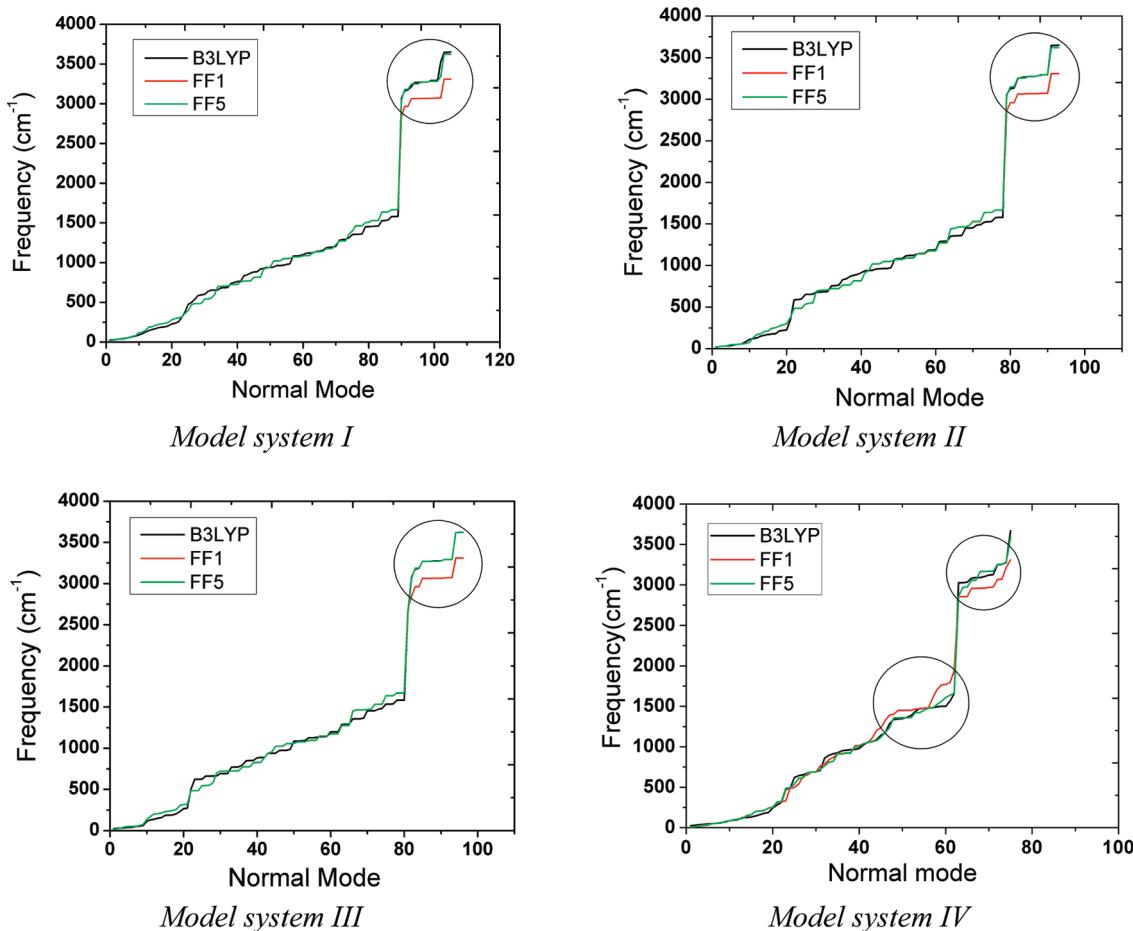


Figure 6. Vibrational frequencies of four model systems computed at the B3LYP/6-311++G(2d, 2p) level (black), force field model FF-1 (red), and the specially optimized force field model FF-5 (green). In each figure, the normal modes are sorted by their vibrational frequencies computed by B3LYP as the x-axis.

is chosen for this purpose anyway, application of RESP charges will not improve its performance.

The five force field models were also tested in extensive MD simulations on the four Zn-containing model systems. When the MD simulations were performed in vacuum, all force field models except FF-3 were able to basically maintain the stable structures of all four model systems (data not shown). In the case of FF-3, the structures under simulation went complete wrong rapidly (<20 ps). The rmsd curves monitored along the entire MD trajectory produced by FF-1 to FF-5 in explicit water are illustrated in Figure 8. One can see clearly that both nonbonded models (FF-3 and FF-4) were not able to maintain the stable structures of all four model systems. Note that the oxygen atom in a TIP3P water molecule carries a substantial amount of partial charge (-0.834e). The water molecule thus acts as a strong competitor for bonding with the zinc ion, which could be challenging for nonbonded models for maintaining the desired coordination configuration of zinc. Bonded models have obvious advantages in this aspect. As revealed in Figure 8, the desired tetrahedral coordination configuration of zinc was highly stable during the entire simulation by FF-1. As for FF-2, the four model systems underwent some noticeable structural fluctuations from time to time. This may also be attributed to the exaggerated secondary electrostatic interac-

tions between the zinc ion and the water molecules due to the formal charge assigned on the zinc ion ($+2\text{e}$).

The results observed in the structural optimizations and the MD simulations of the four model systems are basically consistent. They both indicate that the performance of bonded models (FF-1 and FF-2) is superior to that of nonbonded models (FF-3 and FF-4) in reproducing the desired coordination configuration of zinc. Note that the major difference between FF-1 and FF-2 is whether to employ the RESP or the formal charge model in the computation of electrostatic interactions. The formal charge model assumes that a form charge of $+2\text{e}$ is localized on the zinc ion, and the rest of the parts are not affected by it. In contrast, the charge carried by the zinc ion is dispersed on the entire system according to the RESP model, which better mimics the charge transfer effect between zinc and the surrounding chemical moieties. As indicated by our results, FF-1 is generally more accurate and more robust than FF2. We thus conclude that FF-1 (bonded model + RESP charges) is the better choice for modeling Zn-containing molecular systems.

3.2. MD Simulation of a Carbonic Anhydrase II Complex in Explicit Solvent. Our force field parameters are derived from some simple model systems. The really meaningful application of these parameters will be the modeling of the complexes formed by Zn-containing met-

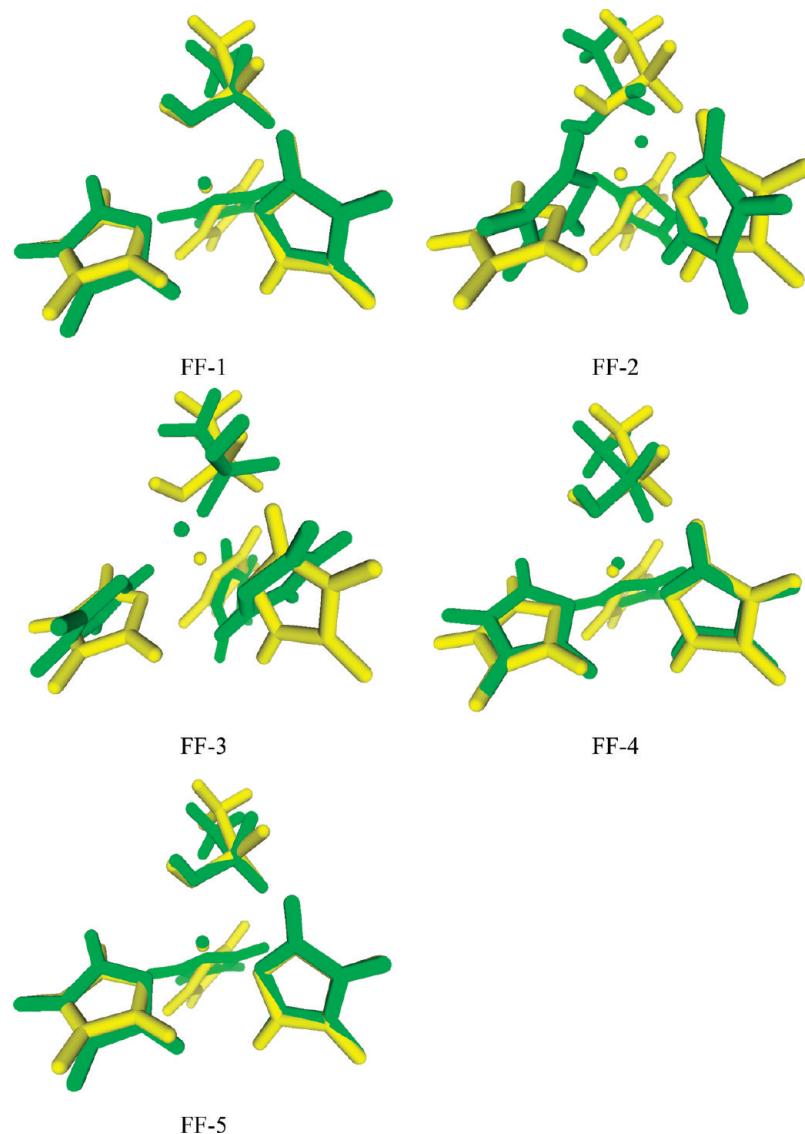


Figure 7. Three-dimensional structures of *model system I* optimized by FF-1 to FF-5 in vacuum based on a ridiculous initial structure. These structures are superimposed with the one optimized at the B3LYP/6-311++G(2d, 2p) level (the structure in yellow).

alloproteins. Thus, we have chosen a protein–ligand complex formed by carbonic anhydrase II, a typical Zn-containing enzyme with pharmaceutical implications, to further validate all five force field models. In this complex structure, the zinc ion inside the binding pocket is in coordination with three histidine residues on the protein (His90, His92, and His115) and a sulfonamide group on the ligand molecule, a chemical configuration identical to *model system I*. In addition, an interesting feature of this complex is that Glu102 is near the coordination center (Figure 5), which is not in direct bonding with the zinc ion ($Zn-O$ distance = 3.9 Å). This feature makes the complex more challenging for simulation since the negatively charged side chain of Glu102 could disrupt the desired coordination configuration of the zinc ion inside the binding pocket.

The five force field models were applied to the MD simulations of this complex structure in explicit water. The force field parameters derived from *model system I* were applied to the simulations by the bonded models, i.e., FF-1, FF-2, and FF-5. Rmsd values of the coordination center (the

zinc ion plus four atoms in direct coordination with it) were monitored along each MD trajectory (Figure 9). One can see that these three bonded models were able to maintain the tetrahedral geometry of the coordination center very well during the entire simulation. In particular, the average rmsd values produced by FF-1 and FF-5 are as small as ~0.4 Å. The average rmsd values produced by FF-2 (~0.7 Å) are slightly larger than those of FF-1 and FF-5. In the last snapshot on the MD trajectory of FF-2, the carboxyl group on the side chain of Glu102 tends to get closer to the zinc ion as compared to the structures produced by FF-1 and FF-5. In fact, one of the oxygen atoms on that carboxyl group is already in a bonding range with the zinc ion ($Zn-O$ distance = 1.80 Å) in this structure. This observation further proves that the significant formal charge assigned on the zinc ion is not completely reasonable.

As for the two nonbonded models (FF-3 and FF-4), relatively large rmsd values are observed on their MD trajectories (Figure 9). One can see clearly in the structures produced by FF-3 and FF-4 that the zinc ion has moved out

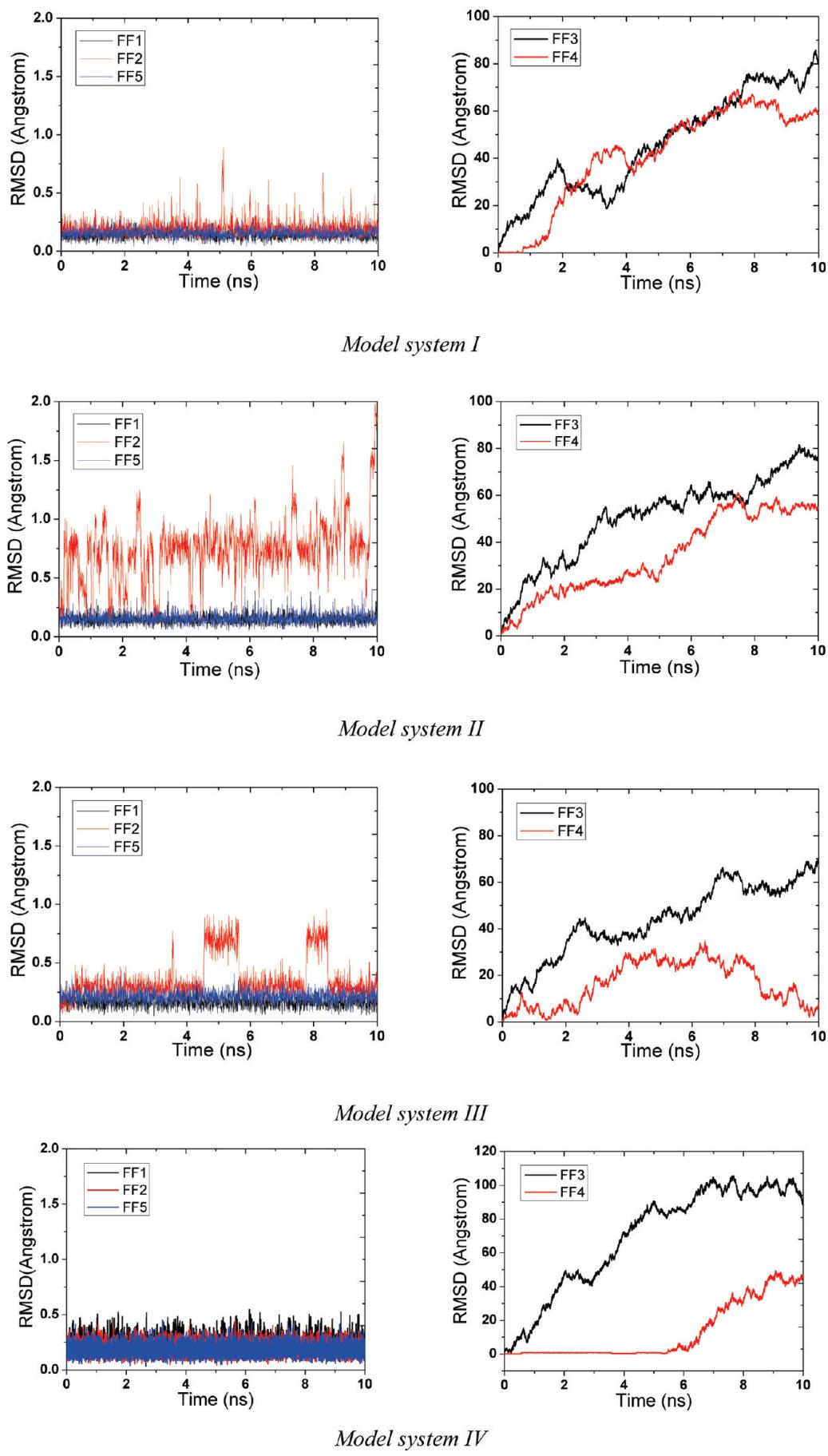


Figure 8. Rmsd values monitored along the MD trajectories produced by five force field models in explicit water. In each case, the rmsd values are computed by considering only the coordination center, while the structure optimized at the B3LYP/6-311++G(2d, 2p) level is used as the reference.

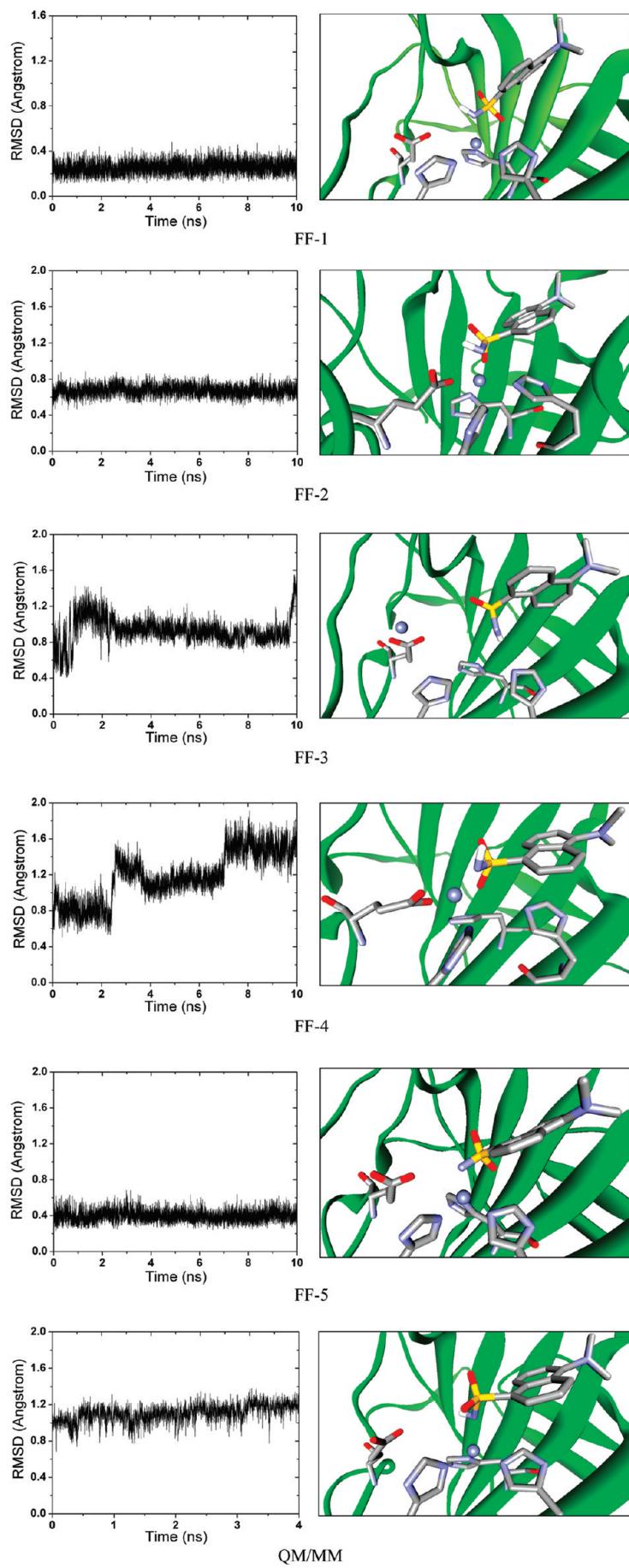


Figure 9. Rmsd values of the coordination center on a carbonic anhydrase II complex structure monitored along each MD trajectory (left), and the last snapshot retrieved from each MD trajectory (right) produced by six different methods.

the coordination center and is now close to the carboxyl group on the side chain of Glu102. Consequently, the tetrahedral geometry of the coordination center is disrupted considerably. It seems that the complex structure under simulation has not reached equilibrium at 10 ns, and it would undergo even more significant changes if the simulation was extended. The potential problem of nonbonded models in modeling Zn-containing protein–ligand complexes is clearly demonstrated in this test.

Recently, Lim et al developed a nonbonded force field model with a potential energy function, including terms for charge transfer and polarization effects.¹³ They conducted molecular dynamics simulations of Zn²⁺ bound to Cys and His residues in proteins using both conventional force field and their modified model. In their study, simulations with the conventional force field yielded a nontetrahedral Cys₂His₂ Zn-binding configuration and significantly overestimated the experimental Zn–S bond length. In contrast, simulations with their new potential energy function better reproduced the experimentally observed tetrahedral Cys₂His₂ and Cys₄ Zn-binding configurations. Lim's study is another good example for demonstrating the limitations of conventional force field models in the simulation of Zn-containing molecular systems. Some appropriate considerations on charge transfer and polarization effect are thus desired if nonbonded force field models are to be used to study such systems.

Nevertheless, in Lim's study the charge transfer was restricted between the zinc ion and the atoms in direct bonding with it, and the amount of transferred charge needed to be recomputed at each time step during simulation. More importantly, adding a polarization term into an existing force field model certainly requires careful reparameterizations and validations. It can be fairly complicated especially when diverse chemical structures have to be considered. In contrast, our approach only needs to define a number of new bond types and to supply the necessary parameters, which is technically more practical. In this study, we have demonstrated the application of this approach to a variety of ligand molecules bound with zinc. It was also unclear in Lim's study if their new potential energy function could reproduce vibrational frequencies, a more challenging goal for force field models. As described later in this article, the force field parameters derived in our study are capable for this purpose.

Besides force field models, the QM/MM method implemented in the AMBER program was also applied in our study to modeling of the same carbonic anhydrase II complex. We found that the MD simulation by employing QM/MM was about five times slower than the MD simulation by employing the AMBER force field. Thus, the MD simulation with QM/MM was performed only for 4 ns due to the significant computation cost. The tetrahedral geometry of the zinc coordination center is basically maintained during this simulation (Figure 9). It, however, undergoes a notable distortion from the original crystal structure (rmsd ~1.3 Å). The same trend is also indicated by the rmsd values computed considering the entire ligand molecule (see the Supporting Information), where the structure produced by the QM/MM simulation exhibits a greater change compared to those produced by FF-1 and FF-5.

It is interesting to find that the force field parameters derived in our study (FF-1/FF-5) outperformed the QM/MM method in modeling this carbonic anhydrase complex. It prompts that, although the concept of QM/MM is appealing, this method is not automatically more accurate than a conventional MM approach. The less encouraging performance of QM/MM observed in our study may be attributed to some technical reasons. For example, the QM computations in QM/MM actually employed the semiempirical PM3 method, which may not produce satisfactory results in this particular case. There are also some other parameters which may affect the final outcomes of the QM/MM method. An optimal set of these parameters may lead to better outcomes. Nevertheless, exploring the QM/MM method implemented in the AMBER program is certainly beyond the scope of this study.

3.3. Reproduction of the Vibrational Frequencies of Zinc-Containing Model Systems. Besides three-dimensional structures, producing the correct vibrational frequencies is also a desired quality of a force field. The vibrational frequencies of four model systems computed at the B3LYP/6-311++G(2d, 2p) level and the force field models FF-1 and FF-5 are illustrated in Figure 6. One can see that, although most vibrational frequencies produced by FF-1 match well with their counterparts produced by QM computations, some notable discrepancies at the high-frequency end (>3000 cm⁻¹) still exist. As described in the Methods Section, we found that stretching motions of the X–H bonds (X = C, N, or S) were largely responsible for this. Consequently, the bond-stretching force constants of such bonds in FF-1 were further optimized through a genetic algorithm approach to better reproduce vibrational frequencies. The resulting parameters, named as FF-5, are tabulated in the Supporting Information together with the original ones in FF-1 for all four model systems.

FF-5 was also applied to optimize the structures of all four model systems. In each case, the structure preoptimized at the B3LYP/6-311++G(2d,2p) level was used as the starting structure, and the final optimized structure was subjected to normal model analysis by using the AMBER program. The vibrational frequencies computed thereby are compared with those produced by QM computations in Figure 6. One can see that FF-5 indeed reproduces the vibrational frequencies at the high-frequency end very well on all four model systems. This observation indicates that our method for optimizing FF-1 is effective; the desired goal can be achieved simply by adjusting the stretching force constants of the X–H bonds within a reasonable range ($\pm 20\%$) of their original values.

Thus, FF5 is more accurate than FF1 in terms of producing correct vibrational frequencies. The only difference between FF1 and FF5 lies in some bond-stretching force constants of X–H bonds. Thus, it is not surprising that, as demonstrated in all of our tests, FF1 and FF5 are equally capable of reproducing the correct three-dimensional structures of simple Zn-containing model systems as well as a Zn-containing protein–ligand complex. In fact, accurate modeling of high-frequency vibrations is rarely a matter of concern for modeling biological macromolecules. For example, the

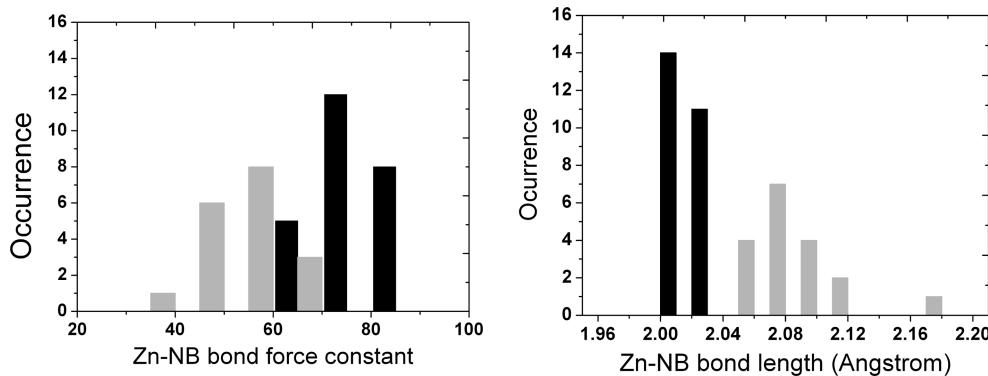


Figure 10. Distribution of the bonding stretching parameters for the Zn–NB bond derived from 18 typical Zn-containing model systems. Columns in black and gray represent the parameters for model systems in which the ligands are neutral and deprotonated, respectively.

popular SHAKE approximation in the AMBER program fixes the lengths of all X–H bonds during MD simulation in order to reduce computational costs. Thus, one can safely apply FF-1 instead to Zn-containing metalloproteins, saving an extra amount of efforts on parameter optimization.

3.4. On the Derived Force Field Parameters. A notable feature of our study is that we constructed a variety of Zn-containing model systems and then applied Seminario's method to derive force field parameters for each of them. As described in the Methods Section, an extensive survey was performed on the entire PDB to identify common chemical moieties in bonding with zinc, including acid, ketone, sulfone, alcohol, thiol, amide, and amine groups. The model systems summarized in Figure 1 are designed to mimic the binding of these chemical groups with Zn-containing proteins. Note that, for alcohol and thiol groups, both of their neutral and deprotonated forms are considered (M5/M6, M7/M8, and M9/M10) in order to explore the possible difference in their bonding with zinc. In addition, six more model systems (M-13 to M-18) are used to mimic some mixed zinc coordination centers observed in Zn-containing proteins. The complete list of the bond-stretching and angle-bending parameters derived in our study for all 18 Zn-containing model systems are summarized in the Supporting Information. They are readily applicable to the molecular modeling studies of Zn-containing proteins, either in bound or unbound states, with the AMBER program.

We have observed in our results that the force field parameters for the same type of bond or angle derived from different model systems are not identical. Instead, they scatter in a certain range. Since most model systems considered in our study contain three imidazole moieties in binding with zinc, we use the Zn–NB bond (NB is a nitrogen atom on the imidazole ring, see Figure 2) here as an example to illustrate this issue. Distributions of the parameters for the Zn–NB bond derived on all model systems are given in Figure 10. One can see that the equilibrium bond lengths of the Zn–NB bond are between 2.00–2.18 Å, while the corresponding stretching force constants are between 30–90 kcal/(mol·Å²). In fact, different sets of these parameters are also reported in literature. For example, the corresponding data reported by Lu et al.¹⁹ were 2.27 Å and 26.0 kcal/(mol·Å²), the data reported by Tuccinardi et al.²⁰ were 2.08 Å and 99.0 kcal/(mol·Å²), and the data derived from the

Raman spectrum of [Zn(NH₃)₄]I₂ crystals were 2.10 Å and 40.0 kcal/(mol·Å²).^{19,53}

In particular, one can see the parameters of the ZN–NB bond shown in Figure 10 can be divided into two distinct groups. The ZN–NB bond tends to be shorter (2.00–2.04 Å) and its stretching force constant is larger (60–90 kcal/mol·Å²) when the ligand molecule binding with zinc is in its neutral form. In contrast, the ZN–NB bond tends to be longer (2.04–2.18 Å) and its stretching force constant is smaller (30–70 kcal/mol·Å²) when the ligand molecule is in its deprotonated form. This can be easily understood: when the ligand is deprotonated, i.e., negatively charged, the bonding between the zinc ion and the ligand is stronger and is associated with a shorter bond length. Consequently, the bonds between zinc and imidazole rings are weakened and become slightly longer. This actually can be proven by comparing the relevant bond-stretching parameters derived from *model systems* II and III (Table 2). The Zn–S bond between zinc and the –SH group is longer than the counterpart between zinc and –S[–] by 0.22 Å, and it is much weaker than the latter. Results derived from other model systems in a neutral/deprotonated pair also reveal the same trend (see the Supporting Information).

Our results, together with the previous results reported by other researchers, suggest that zinc-related force field parameters are not always transferable across different systems. As revealed in some previous studies,^{19,20} parameters originally derived from other model systems had to be optimized on the specific system under study through trial-and-error efforts; otherwise it could be risky. For this reason, we have considered a variety of Zn-containing model systems and produced a more comprehensive set of parameters. When one wants to model a specific Zn-containing system, he/she may choose and apply the appropriate force field parameters derived from a corresponding model system, which will in turn reduce the efforts on parameter optimization and produce more accurate results. This is a noteworthy advantage of our study as compared to previous studies.

Another noteworthy advantage of our study, which is perhaps more important, is that we have demonstrated a complete procedure for the deduction and validation of force field parameters. We have chosen Seminario's method for deriving force field parameters from the outcomes of QM computation, which has obvious technical advantages. First,

it is based on a Hessian matrix in the Cartesian coordinates. Thus, it avoids the troubles in setting internal coordinates as seen in some previous studies.²⁵ Second, the desired force constants of multiple bonds and bond angles can be derived simultaneously in one job rather than through an iterative trial-and-error approach. Very little human inference is needed during the whole process. In fact, once the Hessian matrix of a given model system is available, the rest of the steps can be largely automated by computer programs. Due to this advantage, we were able to process a variety of model systems and obtain the desired parameters. Application of Seminario's method is not limited to the 18 selected Zn-containing systems considered in this study. The same approach certainly can be employed to handle other Zn-containing molecular systems whenever necessary. It, in principle, can be applied to other types of metal ions or uncommon chemical moieties lacking appropriate force field parameters as well.

4. Conclusions

In this study, we have derived bond-stretching and angle-bending parameters applicable to zinc-containing systems through a systematic approach. A total of 18 Zn-containing model systems were considered, and Seminario's method was applied to analyze the Hessian matrix of each model system computed at the B3LYP/6-311++G(2d,2p) level to derive the desired force constants. Then, the derived parameters were validated extensively in structural optimization and molecular dynamics simulations of four model systems as well as one protein–ligand complex formed by carbonic anhydrase II. With the application of these parameters, the bonded model in combination with the RESP charges (FF-1) was founded to be most robust in reproducing the three-dimensional structures of these Zn-containing systems; whereas the performance of nonbonded models was generally inferior. The performance of FF-1 was even better than the quantum mechanics/molecular mechanics (QM/MM) method implemented in the AMBER program in the MD simulations of a carbonic anhydrase II complex in explicit water. After some necessary optimizations on the force constants of X–H bonds, i.e., FF-5, it was also able to reproduce the vibrational frequencies of each model system provided by QM computations. Thus, the force field parameters derived in our study seem to be very reliable. Our approach, which is based on Seminario's method, has certain technical advantages. It can derive the parameters for all relevant chemical bonds in one batch rather than through an iterative procedure. Application of this approach is certainly not limited to the Zn-containing systems considered in this study. It is, in principle, applicable to molecular systems containing other types of metal ions or uncommon chemical moieties in which appropriate force field parameters are currently not available.

Acknowledgment. The authors are grateful to the financial supports from the Chinese National Natural Science Foundation (grant no. 20772149 and 90813006), the Chinese Ministry of Science and Technology (grant no. 2006AA02Z337 and 2009ZX09501-002), and the Science and Technology Commission of Shanghai Municipality (grant no. 074319113).

The authors also thank Dr. Eddy Bautista at Texas A&M University in College Station, Texas and Prof. Ulf Ryde at Lund University in Sweden for their helpful instructions on Seminario's method.

Supporting Information Available: A complete list of the force field parameters derived in this study. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

- (1) Sarkar, B. Metal protein interactions. *Prog. Food Nutr. Sci.* **1987**, *11*, 363–400.
- (2) Dudev, T.; Lin, Y. L.; Dudev, M.; Lim, G. First-Second Shell Interactions in Metal Binding Sites in Proteins: A PDB Survey and DFT/CDM Calculations. *J. Am. Chem. Soc.* **2003**, *125*, 3168–3180.
- (3) Auld, D. S. Zinc coordination sphere in biochemical zinc sites. *BioMetals* **2001**, *14*, 271–313.
- (4) Lipscomb, W. N.; Strater, N. Recent Advances in Zinc Enzymology. *Chem. Rev.* **1996**, *96*, 2375–2433.
- (5) Brinckerhoff, C. E.; Matrisian, L. M. Matrix metalloproteinases: a tail of a frog that became a prince. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 207–214.
- (6) Tallant, C.; Marrero, A.; Gomis-Rüth, F. X. Matrix metalloproteinases: Fold and function of their catalytic domains. *Biochim. Biophys. Acta* **2010**, *1803*, 20–28.
- (7) Ryde, U. Combined quantum and molecular mechanics calculations on metalloproteins. *Curr. Opin. Chem. Biol.* **2003**, *7*, 136–142.
- (8) Ryde, U. The coordination of the catalytic zinc ion in alcohol dehydrogenase studied by combined quantum-chemical and molecular mechanics calculations. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 153–164.
- (9) Brancato, G.; Rega, N.; Barone, V. Microsolvation of the Zn(II) ion in aqueous solution: A hybrid QM/MM MD approach using non-periodic boundary conditions. *Chem. Phys. Lett.* **2008**, *451*, 53–57.
- (10) Hartsough, D. S.; Merz, K. M. Dynamic force field models: Molecular dynamics simulations of human carbonic anhydrase II using a quantum mechanical/molecular mechanical coupled potential. *J. Phys. Chem.* **1995**, *99*, 11266–11275.
- (11) Stote, R. H.; Karplus, M. Zinc Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 12–31.
- (12) Donini, O. A.; Kollman, P. A. Calculation and prediction of binding free energies for the matrix metalloproteinases. *J. Med. Chem.* **2000**, *43*, 4180–4188.
- (13) Sakharov, D. V.; Lim, C. Zn Protein Simulations Including Charge Transfer and Local Polarization Effects. *J. Am. Chem. Soc.* **2005**, *127*, 4921–4929.
- (14) Pang, Y. P. Novel zinc protein molecular dynamics simulations: Steps toward antiangiogenesis for cancer treatment. *J. Mol. Model.* **1999**, *5*, 196–202.
- (15) Pang, Y. P.; Xu, K.; El Yazal, J.; Prendergast, F. G. Successful molecular dynamics simulation of the zinc-bound farnesyl-transferase using the cationic dummy atom approach. *Protein Sci.* **2000**, *9*, 1857–1865.
- (16) Vendani, A.; Huhta, D. W. A New Force Field for Modeling Metalloproteins. *J. Am. Chem. Soc.* **1990**, *112*, 4759–4767.

- (17) Merz, K. M. Insights into the function of the zinc hydroxide-Thr199-Glu106 hydrogen bonding network in carbonic anhydrases. *J. Mol. Biol.* **1990**, *214*, 799–802.
- (18) Suarez, D.; Merz, K. M. Molecular Dynamics Simulations of the Mononuclear Zinc- β -lactamase from *Bacillus Cereus*. *J. Am. Chem. Soc.* **2001**, *123*, 3759–3770.
- (19) Lu, Q.; Tan, Y. H.; Luo, R. Molecular Dynamics Simulations of p53 DNA-Binding Domain. *J. Phys. Chem. B* **2007**, *111*, 11538–11545.
- (20) Tuccinardi, T.; Martinelli, A.; Nuti, E.; Carelli, P.; Balzano, F.; Uccello-Barretta, G.; Murphy, G.; Rossello, A. Amber force field implementation, molecular modelling study, synthesis and MMP-1/MMP-2 inhibition profile of (R)- and (S)-N-hydroxy-2-(N-isopropoxybiphenyl-4-ylsulfonamido)-3-methylbutanamides. *Bioorg. Med. Chem.* **2006**, *14*, 4260–4276.
- (21) Li, W.; Zhang, J.; Wang, J.; Wang, W. Metal-Coupled Folding of Cys₂His₂ Zinc-Finger. *J. Am. Chem. Soc.* **2008**, *130*, 892–900.
- (22) Falconi, M.; Altobelli, G.; Iovino, M. C.; Politi, V.; Desideri, A. Molecular dynamics simulation of Matrix Metalloproteinase 2: fluctuations and time evolution of recognition pockets. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 837–848.
- (23) Lu, D. S.; Voth, G. A. Molecular dynamics simulations of human carbonic anhydrase II: Insight into experimental results and the role of solvation. *Proteins* **1998**, *33*, 119–134.
- (24) Ryde, U. Molecular dynamics simulations of alcohol dehydrogenase with a four- or five-coordinate catalytic zinc ion. *Proteins: Struct., Funct. Genet.* **1995**, *21*, 40–56.
- (25) Seminario, J. M. Calculation of Intramolecular Force Fields from Second-Derivative Tensors. *Int. J. Quantum Chem.* **1996**, *60*, 59–65.
- (26) Bautista, E. J.; Seminario, J. M. Harmonic force field for glycine oligopeptides. *Int. J. Quantum Chem.* **2008**, *108*, 180–188.
- (27) Brandt, P.; Norrby, T.; Åkermark, B.; Norrby, P. O. Molecular mechanics (MM3*) parameters for ruthenium(II)–polypyridyl complexes. *Inorg. Chem.* **1998**, *37*, 4120–4127.
- (28) Norrby, P.; Rasmussen, T.; Hallen, J.; Strassmen, T.; Houk, K. N. Rationalizing the Stereoselectivity of Osmium Tetroxide Asymmetric Dihydroxylations with Transition State Modeling Using Quantum Mechanics-Guided Molecular Mechanics. *J. Am. Chem. Soc.* **1999**, *121*, 10186–10192.
- (29) Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. An automatic method to generate force-field parameters for hetero-compounds. *Acta Crystallogr.* **2003**, *59*, 274–289.
- (30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (31) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. C.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; V Rotiz, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskora, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; AlLaham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Gordon, M. H.; Replogle, E. S.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- (32) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T.; Wang, J.; Ross, W. S.; Simmerling, C.; Darden, T.; Merz, K. M.; Stanton, R. V.; Cheng, A.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. *AMBER 9*; University of California: San Francisco, CA, 2006.
- (33) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nuc. Acid. Res.* **2008**, *28*, 235–242.
- (34) Wachters, A. J. H. Gaussian basis set for molecular wavefunctions containing third-row atoms. *J. Chem. Phys.* **1970**, *52*, 1033–1036.
- (35) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Farnesyltransferase—new insights into the zinc-coordination sphere paradigm: evidence for a carboxylate-shift mechanism. *Biophys. J.* **2005**, *88*, 483–494.
- (36) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Theoretical studies on farnesyl transferase: evidence for thioether product coordination to the active-site zinc sphere. *J. Comput. Chem.* **2007**, *28*, 1160–1168.
- (37) Tamames, B.; Sousa, S. F.; Tamames, J.; Fernandes, P. A.; Ramos, M. J. Analysis of zinc-ligand bond lengths in metalloproteins: trends and patterns. *Proteins* **2007**, *69*, 466–475.
- (38) Amin, E. A.; Truhlar, D. G. Zn Coordination Chemistry: Development of Benchmark Suites for Geometries, Dipole Moments, and Bond Dissociation Energies and Their Use To Test and Validate Density Functionals and Molecular Orbital Theory. *J. Chem. Theory Comput.* **2008**, *4*, 75–85.
- (39) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (40) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (41) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.
- (42) Hoops, S. C.; Anderson, K. W.; Merz, K. M. Force Field Design for Metalloproteins. *J. Am. Chem. Soc.* **1991**, *113*, 8262–8270.
- (43) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **1995**, *16*, 1357–1377.
- (44) Jorgensen, W. L.; Chandrasekhar, J.; Madurs, J.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (45) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin dynamics of peptides: The frictional dependence of isomerization rates of N-acetylanyl-N'-methyl amide. *Biopolymers* **1992**, *32*, 523–535.

- (46) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Nola, A. D.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (47) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. The Smoothed Particle Mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (48) Nair, S. K.; Elbaum, D.; Christianson, D. W. Unexpected binding mode of the sulfonamide fluorophore 5-dimethylamino-1-naphthalene sulfonamide to human carbonic anhydrase II. Implications for the development of a zinc biosensor. *J. Biol. Chem.* **1996**, *271*, 1003–1007.
- (49) Ryckaert, J. P.; Ciccotti, G.; Berendsen, J. C. Numerical integration of the Cartesian equation of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Chem. Phys.* **1977**, *67*, 327–341.
- (50) Walker, R. C.; Crowley, M. F.; Case, D. A. The Implementation of a Fast and Accurate QM/MM Potential Method in AMBER. *J. Comput. Chem.* **2008**, *29*, 1019–1031.
- (51) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods Applications. *J. Comput. Chem.* **1989**, *10*, 221–264.
- (52) Holland, J. *Adaptation in Natural and Artificial Systems*; University of Michigan Press: Ann Arbor, MI, 1975.
- (53) Nakamoto, K.; Takemoto, J.; Chaw, T. L. Metal Isotope Effect on Raman Spectrum of $[Zn(NH_3)_4]I_2$ Crystals. *Appl. Spectrosc.* **1971**, *25*, 352–355.

CT900454Q