Development of a Fluctuating Charge Model for Zinc-Containing Metalloproteins

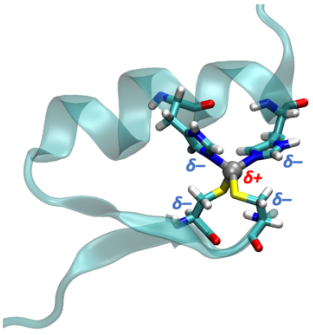
Luke Landry and Pengfei Li

Department of Chemistry, Loyola University Chicago, Chicago, IL 60660, United States

Abstract

The fluctuating charge (FQ) model can generate atomic charges much more efficiently than quantum mechanical methods. The FQ model has been developed for a wide range of applications, but few models were specifically tailored for calculating atomic charges of metalloproteins. Zinc-containing proteins widely exist in biology and play important roles in various processes. In this study, we present a fluctuating charge model for zinc-containing metalloproteins based on the charge equilibration (Qeq) scheme. Our model was parameterized to reproduce CM5 charges, which demonstrated excellent performance in reproducing molecular dipole moments. During our study, we found that adding the Pauling-bond-order-like term (referred to as the "+C term" in a previous study) between the zinc ion and ligating atoms significantly improves the model's performance. Although our model was trained for four-coordinated zinc sites only, our tests indicated it can well describe the atomic charges in five- and six-coordinated zinc sites as well. Finally, we adapted our model to generate partial charges for the metal site in a zinc finger domain. These charges exhibited comparable performance to the widely used restrained electrostatic potential (RESP) charges in molecular dynamics (MD) simulations. The current model can be extended to other metal-containing systems and serve the molecular modeling community.

TOC Graphics



$$E_{sys}(Q_1, Q_2, \ldots, Q_N) = \sum_{k=1}^{N}\left(E_{Ak} + \frac{1}{2}E_{Ik}\right)$$

$$E_{Ak}(Q) = E_{Ak}(0) + \chi_k Q_k + \frac{1}{2}J_k Q_k^2$$

$$Q_k{}^{+C} = Q_k + \sum_{k' \neq k} T_{kk'} B_{kk'}$$

Introduction

Molecular dynamics (MD) simulations are an important approach in computational chemistry and have found wide applications in investigating materials and biochemical systems. In popular force field models used for MD simulations, atomic charge parameters are critical for accurately representing molecular interactions. Quantum mechanical (QM) methods are usually employed to determine these atomic charges. However, QM methods are impractical for modeling large systems, such as proteins or metal-organic frameworks, due to their relatively high computational costs. In contrast, the fluctuating charge (FQ) model is much more computationally efficient and provides an alternative for determining atomic charges in a timely manner, particularly for macromolecular systems.

The fluctuating charge (FQ) model utilizes geometric and atomic properties, mainly atomic coordinates, electronegativity (χ) and chemical hardness (η), to calculate atomic charges. In 1934, Mulliken originally defined absolute electronegativity as χ = (IE + EA) / 2, where IE represents the ionization energy (previously called ionization potential) and EA denotes the electron affinity.[1-2] In 1983, Parr and Pearson introduced the absolute hardness, $\eta = (IP - EA)/2$.[3] In 1991, Rappé and Goddard established the charge equilibration (Qeq) method to calculate atomic charges for MD simulations.[4] The atomic charges predicted by the Qeq method were consistent with experimental dipole moments and with the atomic charges fitted based on electrostatic potentials calculated by ab initio methods, making this approach valuable for simulating biological systems, polymers, and inorganic materials. In 1996, Ramachandran et al. expanded the Qeq method to model periodic systems, introducing the periodic charge equilibration (PQeq) approach. This method considers the impact of the lattice on atomic charges, making it suitable for modeling crystalline systems. By using Ewald summation, the PQeq method derives finite values for electrostatic energy over neutral cells in periodic systems and has been successfully applied to simulate zeolites in solutions.[5]

In 2012, Wilmer et al. proposed the extended charge equilibration (EQeq) method to calculate the atomic charges in large molecular systems.[6] This method used the electronegativity and chemical hardness values based on experimental ionization energies and electron affinities. Although having less *ad hoc* parameters than the original Qeq model, the EQeq model still maintains a reasonable level of accuracy for predicting atomic charges. Moreover, the original Qeq method derived atomic charges iteratively, due to a strategy utilized in calculations of the hydrogen atom charges.[4] The EQeq method requires no iteration, which makes it a more efficient option for screening materials and calculating atomic charges along molecular simulations. However, in 2015, Martin-Noble et al. found that the EQeq method showed limited accuracy in predicting atomic charges of transition metals with high oxidation states, resulting in an unphysical trend where the calculated charges are only weakly influenced by the environment.[7] To address this issue, they developed the EQeq+C scheme,[7] which includes an empirical correction term based on the Pauling bond-order. By adding this correction term, the EQeq+C scheme improved the accuracy of atomic charge predictions for high-oxidation-state metals, without significantly increasing computational time.

In addition to the Qeq model, many other FQ models have been developed. For examples, in the 1980s, Mortier et al. proposed the electronegativity equalization method (EEM), which was evaluated for different molecules and showed promising performance.[8] In 1997, Yang and Wang introduced the atom-bond electronegativity equalization method (ABEEM) and applied it to determine atomic charges in large organic molecules.[9] Chen and Martinez developed

the charge transfer with polarization current equalization (QTPIE) model in 2007.[10] This model correctly described the atomic charges at the bond dissociation limits through a distance-dependent factor that penalizes long-range charge transfer. Overall, these FQ models were developed based on a similar fundamental concept as the Qeq model, but with specific adjustments to suit different applications.

In 2004, Patel et al. introduced the CHARMM-FQ model,[11-12] which was parameterized towards DFT-based calculations to reproduce the behavior of small molecules in the presence of water dipole moments. The nonbonded parameters were also optimized in this model. The CHARMM-FQ model has been employed to simulate different systems, including solute-water systems in the gas phase, properties of bulk liquids, and solvated peptide and protein systems.[11-12] Overall, the CHARMM-FQ model offers valuable insights into how small molecules react with water and it has proven to be a tool which can well simulate complex biomolecular systems.

Metalloproteins and organometallic compounds are vital to many chemical and biochemical processes;[13-14] however, a general force field has yet to be developed for these systems due to the diverse range of metal sites present. To address this issue, Merz and colleagues developed the metal center parameter builder (MCPB) program to streamline the parameterization process for metal sites in metalloproteins.[15] Subsequently, Li and Merz developed the MCPB.py program, which utilizes a Python environment to optimize the workflow and extend its applicability to organometallic compounds.[16] However, this approach still requires QM calculations to determine the atomic charges of the metal site.

While some studies have expanded the FQ model to encompass diverse systems, few have specifically tailored it for classical molecular dynamics simulations of metalloproteins. Zinc containing metalloproteins widely exist in biology. For example, ~10% of the human proteome are zinc containing metalloproteins.[17] They play important roles in various processes such as transcription, signal transduction, enzymatic catalysis, etc.[18-20] In this work, we present a FQ model designed for zinc containing metalloproteins, covering the most common ligands to facilitate the parameterization of atomic charges in the associated metal sites. This model will be integrated into the AMBER software package[21] to better serve the molecular modeling community.

Methods

In a FQ model, the total energy of a molecular system is expressed as a function of the charges on atoms.[4, 8] Our FQ model was based on the work of Wilmer et al. (i.e. the EQeq scheme),[6] without considering the periodic interaction energy aspect. The advantage of this implementation is that one can obtain atomic charges by solving a matrix, instead of with iterative calculations. The method details of our implementation are described below.

In general, the total energy of the system is given as:

$$E_{sys}(Q_1, Q_2, \dots, Q_N) = \sum_{k=1}^{N} (E_{Ak} + \frac{1}{2} E_{Ik}) \qquad (1)$$

Herein E_{Ak} represents the atomic energy of atom k and E_{Ik} represents the electrostatic interaction energy between atom k and any other atoms in the system. The energy of atom k can be represented by a Taylor series truncated at the second order:

$$E_{Ak}(Q) = E_{Ak}(0) + \chi_k Q_k + \eta_k Q_k^2 = E_{Ak}(0) + \chi_k Q_k + \frac{1}{2} J_k Q_k^2 \qquad (2)$$

Herein χ and η are the electronegativity and chemical harness which can be obtained based on the ionization potential (*IP*) and electron affinity (*EA*):

$$\chi = \frac{1}{2}(IP + EA) \quad (3)$$

$$\eta = \frac{1}{2}(IP - EA) = \frac{1}{2}J \quad (4)$$

Where *IP* and *EA* are defined by the following equations:

$$X + IP \rightarrow X^+ + e^- \quad (5)$$

$$X + e^- \rightarrow X^- + EA \quad (6)$$

Moreover, E_{Ik} is given by the summation of two terms, denoted as E_{Ck} and E_{Ok}:

$$E_{Ik} = E_{Ck} + E_{Ok} \quad (7)$$

Where E_{Ck} represents the long-range Coulombic interaction between atom k and all other atoms within the system, while E_{Ok} denotes the short-range damping term (which is also known as the orbital overlap energy) between atom k and all other atoms within the system. Their formulations are shown as follows:

$$E_{Ck} = \sum_{m \neq k}^{N} \frac{K Q_k Q_m}{r_{km}} \quad (8)$$

$$E_{Ok} = \sum_{m \neq k}^{N} K Q_k Q_m E_O(r_{km}) = \sum_{m \neq k}^{N} K Q_k Q_m \left(e^{-\left(\frac{J'_{km} r_{km}}{K}\right)^2} \left(\frac{J'_{km}}{K} - \frac{J'^2_{km} r_{km}}{K^2} - \frac{1}{r_{km}} \right) \right) \quad (9)$$

In the above two equations, Q is in the unit of atomic charge and r is in the unit of Å. K is $1/4\pi\epsilon_r\epsilon_0$, where ϵ_r is the dielectric constant (or relative permittivity), and ϵ_0 is the vacuum permittivity. In the current study, we use ϵ_r of 1.0, providing K as a conversion factor of 14.4 eV·Å/e^2. r_{km} represents the distance between atoms k and m. J'_{km} is equal to the geometric mean of the chemical hardness of atoms k and m: $J'_{km} = \sqrt{J_k J_m}$. Herein the formulation of E_{Ok} well describes the correct asymptotic behavior on the two ends: when r_{km} is large, E_{ok} is negligible so $E_{Ck} + E_{Ok}$ resumes to the Coulombic equation; when r_{km} approaches zero (i.e. atoms k and m overlap with each other), $E_{Ck} + E_{Ok}$ approaches $J_k Q_k^2$ if k and m are the same element, which is cohesive to the hardness term of a single atom.

To keep the total charge Q_T of the system, the following constraint to the atomic charges was applied:

$$\sum_{k=1}^{N} Q_k = Q_T \quad (10)$$

The Lagrange multiplier was employed to minimize the E_{sys} while constraining the total charge:

$$\nabla E_{sys} = \lambda \nabla \left(\sum_{k=1}^{N} Q_k - Q_T \right) \quad (11)$$

Herein ∇ represents the gradient vector along the N dimensional space characterized by the partial charges (N atoms in the system, with each atom has a partial charge). By matching the vector component on each direction, we have:

$$\frac{\partial E_{sys}}{\partial Q_k} = \lambda \quad \forall k \quad (12)$$

Which will give us the following relationship:

$$\frac{\partial E_{sys}}{\partial Q_1} = \frac{\partial E_{sys}}{\partial Q_1} = \cdots = \frac{\partial E_{sys}}{\partial Q_N} \quad (13)$$

With the expansion of $\frac{\partial E_{sys}}{\partial Q_k}$ as follows:

$$\begin{aligned}\frac{\partial E_{sys}}{\partial Q_k} &= \frac{\partial}{\partial Q_k}\left(\sum_{k=1}^{N}\left(E_{Ak} + \frac{1}{2}E_{Ik}\right)\right) \\ &= \frac{\partial}{\partial Q_k}\left(\sum_{k=1}^{N}\left(E_{Ak}(0) + \chi_k Q_k + \frac{1}{2}J_k Q_k^2 + \frac{1}{2}\sum_{m\neq k}^{N}\frac{KQ_kQ_m}{r_{km}} + \frac{1}{2}\sum_{m\neq k}^{N}KQ_kQ_mE_O(r_{km})\right)\right) \\ &= \chi_k + J_kQ_k + \frac{K}{2}\sum_{m\neq k}^{N}\left(\frac{Q_m}{r_{km}} + Q_mE_O(r_{km})\right) \quad (13)\end{aligned}$$

We can get:

$$\begin{aligned}&\chi_1 + J_1Q_1 + Q_2\frac{K}{2}\left(\frac{1}{r_{12}} + E_O(r_{12})\right) + \cdots + Q_N\frac{K}{2}\left(\frac{1}{r_{1N}} + E_O(r_{1N})\right) \\ &= \chi_2 + Q_1\frac{K}{2}\left(\frac{1}{r_{21}} + E_O(r_{21})\right) + J_2Q_2 + \cdots + Q_N\frac{K}{2}\left(\frac{1}{r_{1N}} + E_O(r_{1N})\right) \\ &= \cdots = \chi_N + Q_1\frac{K}{2}\left(\frac{1}{r_{N1}} + E_O(r_{N1})\right) + Q_2\frac{K}{2}\left(\frac{1}{r_{N2}} + E_O(r_{N2})\right) + \cdots + J_NQ_N \quad (14)\end{aligned}$$

Here we define:

$$J_{km} = \frac{K}{2}\left(\frac{1}{r_{km}} + E_O(r_{km})\right) \quad (15)$$

$$J_{kk} = J_k \quad (16)$$

Where the chemical hardness term in eq 2 can be considered as an idempotential (self-Coulombic interaction) term. In addition, we substituted eq 15 by the hardness equation used in the CHARMM-FQ model[11] (eq 17). This hardness equation screens the electrostatic interactions among the 1–2, 1–3, and 1–4 sites. Once the atomic separation is beyond 2.5 Å, the equation resembles the Coulomb potential. This equation provides better fitting results (lower root-mean-square errors) in our test study, and qualitatively agrees with the AMBER force field philosophy (which can be considered as using a scaling factor of 0 for electrostatic interactions among the 1-2 and 1-3 sites, and using a scaling factor of 1.0/1.2 for the electrostatic interactions among the 1-4 sites)[22], so we used this equation in our model.

$$J_{km} = \frac{\frac{1}{2}(J_k + J_m)}{\sqrt{1 + \frac{1}{4}(J_k + J_m)^2 r_{km}^2}} \quad (17)$$

We can rewrite eq 14 as:

$$\begin{aligned}&\chi_1 + J_{11}Q_1 + J_{12}Q_2 + \cdots + J_{1N}Q_N = \chi_2 + J_{21}Q_1 + J_{22}Q_2 + \cdots + J_{2N}Q_N \\ &= \cdots = \chi_N + J_{N1}Q_1 + J_{N2}Q_2 + \cdots + J_{NN}Q_N \quad (18)\end{aligned}$$

If we take subtraction between each of the two neighboring equations, we can get $N-1$ equations:

$$\left(J_{(N-1)1}-J_{N1}\right)Q_1+\left(J_{(N-1)2}-J_{N2}\right)Q_2+\cdots+\left(J_{(N-1)N}-J_{NN}\right)Q_N=\chi_N-\chi_{N-1} \quad (19)$$

The matrix notation of the above equations is:

$$\begin{bmatrix} J_{11}-J_{21} & \cdots & J_{1N}-J_{2N} \\ \vdots & \ddots & \vdots \\ J_{(N-1)1}-J_{N1} & \cdots & J_{(N-1)N}-J_{NN} \end{bmatrix} \cdot \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix} = \begin{bmatrix} \chi_2-\chi_1 \\ \vdots \\ \chi_N-\chi_{N-1} \end{bmatrix} \quad (20)$$

Upon addition of a row to account for the total charge constraint, we obtain the following matrix:

$$\begin{bmatrix} 1 & \cdots & 1 \\ J_{11}-J_{21} & \cdots & J_{1N}-J_{2N} \\ \vdots & \ddots & \vdots \\ J_{(N-1)1}-J_{N1} & \cdots & J_{(N-1)N}-J_{NN} \end{bmatrix} \cdot \begin{bmatrix} Q_1 \\ \vdots \\ Q_N \end{bmatrix} = \begin{bmatrix} Q_T \\ \chi_2-\chi_1 \\ \vdots \\ \chi_N-\chi_{N-1} \end{bmatrix} \quad (21)$$

The partial charges were obtained by solving this matrix. Further study proved that the EQeq method erroneously predicts the charges of highly charged transition metals in amine-templated metal oxides.[7] To solve this issue, a correctional term ("+C" term), which was inspired by the Charge Model 5 (CM5),[23] was added for all atom pairs, yielding the EQeq+C model.[7] Herein, to improve the model performance, we also incorporated the +C term into our FQ model. Due to the charge transfer effect between metal ions and ligands are relatively localized, and to keep the simpleness of our model, we only added the +C term between the metal ions and ligating atoms. Specifically, for the central ion and the ion ligating atoms, the following term was applied:

$$Q_k{}^{+C} = Q_k + \sum_{k' \neq k} T_{kk'}B_{kk'} \quad (22)$$

Where $\sum_{k'\neq k} T_{kk'}B_{kk'}$ is the charge correction term, and Q_k and $Q_k{}^{+C}$ are the Qeq charges of atom k before and after the correction, respectively. $T_{kk'}$ is a parameter for the k and k' atom pair which represents the charge transfer between the two connected atoms (where $T_{kk'} = -T_{k'k}$). $B_{kk'}$ is a Pauling bond-order-like term between the two atoms (see eq 23). If $T_{kk'}$ is negative, the charge value of atom k will decrease due to the correction term, given that the $B_{kk'}$ term is always positive.

$$B_{kk'} = \exp\left[-\alpha\left(r_{kk'}-\left(R_{Z_k}+R_{Z_{k'}}\right)\right)\right] \quad (23)$$

Herein α is a constant set as 2.474 (from the CM5 model[23]), $r_{kk'}$ is the distance between the two atoms, and R_{Z_k} and $R_{Z_{k'}}$ are the covalent radii of atom k and k'. As the distance between two atoms decreases, more partial charge will be transferred due to the "+C" term. The covalent radii for zinc and the ligating atoms were referred from Meng et al.[24] and are provided in **Table S1**.

Results and Discussion

In the present study, we have employed CM5 charges[23] as the target to fit our FQ model. Meanwhile, the restrained electrostatic potential (RESP)[25] charges are widely used in molecular dynamics simulations, including their use as the standard protocol for the AMBER protein force field[22, 26] and general AMBER force field,[27] where they are computed at the HF/6-31G* level of theory. Nonetheless, the RESP charges are determined based on the electrostatic potentials

surrounding the molecules and may not accurately reflect the intrinsic nature of the atoms. Consequently, this approach could give unphysical negative charges on buried metal ions. In contrast, the CM5 charges are obtained based on the Hirshfeld charges and a Pauling bond order term to further consider the charge transfer, which is due to the Hirshfeld charges often being too small to accurately describe the dipole moments.[23] CM5 charges can effectively reproduce molecular dipole moments, are essentially independent of the basis set used, are less dependent on conformation (compared to ESP charges), and do not suffer from unphysical charges assigned to buried atoms.[25, 28-30] Taking into account these favorable characteristics, we utilize the CM5 charges as the target for model parameterization in the present study.

The zinc binding sites can be classified into different categories: structural, catalytic, and cocatalytic.[19, 31-33] Histidine, glutamic acid, aspartic acid, and cysteine are the most common amino acids that coordinate with zinc in proteins in addition to water typically appearing in catalytic zinc sites.[19, 31-33] For parameter optimization, a total of 13 systems were selected as the training set to represent diverse tetrahedral Zn complexes. These systems encompassed a range of ligands, including water and selected amino acids CYM (deprotonated CYS), HIE (ε-position protonated HIS), HID (δ-position protonated HIS), ASP (which represents both ASP and GLU because their similarity), SER and SED (deprotonated SER). Specifically, among the 13 systems, 11 systems were the same as investigated in Peters et al.'s study,[15] and the other two are Zn-SSSS complexes (**see Table 1**). The initial protein structures were sourced from the Protein Data Bank, except the two Zn-SSSS complexes were manually created. Further details of the selected complexes can be found in **Table 1**.

Table 1. Systems in the training set.

System	PDB entry	Resolution (Å)	Complex	Total charge (e)
1	1A5T	2.20	Zn–CCCC	-2
2	1A73	1.80	Zn–CCCH$_{HIE}$	-1
3	2GIV	1.94	Zn–CCCH$_{HID}$	-1
4	1A1F	2.10	Zn–CCH$_{HID}$H$_{HID}$	0
5	1PB0	1.95	Zn–H$_{HID}$H$_{HID}$H$_{HID}$H$_{HID}$	+2
6	1CK7	2.80	Zn–CH$_{HID}$H$_{HID}$H$_{HID}$	+1
7	1CA2	2.00	Zn–H$_{HID}$H$_{HID}$H$_{HIE}$O	+2
8	1U0A	1.64	Zn–H$_{HIE}$H$_{HIE}$DD	0
9	2USN	2.20	Zn–H$_{HID}$H$_{HID}$H$_{HIE}$D	+1
10	1VLI	2.38	Zn–H$_{HID}$H$_{HID}$OO	+2
11	1L3F	2.30	Zn–H$_{HIE}$OOO	+2
12	*N/A*	*N/A*	Zn–S$_{SED}$S$_{SED}$S$_{SED}$S$_{SED}$	-2
13	*N/A*	*N/A*	Zn–S$_{SED}$S$_{SED}$S$_{SER}$S$_{SER}$	0

To prepare the initial structures for the training set, we extracted the zinc metal sites of Systems 1-11 from the corresponding Protein Data Bank (PDB) files and capped the system by changing the CA atoms to methyl groups. Subsequently, we performed a geometrical optimization for each system in the gas phase at the B3LYP/6-31G* level of theory. The optimized geometries are illustrated in **Figure 1**. These quantum calculations were conducted using the Gaussian 16 program (revision C.01).[34] Afterwards, we optimized the parameters of our FQ model to reproduce the calculated CM5 charges based on the optimized geometries.

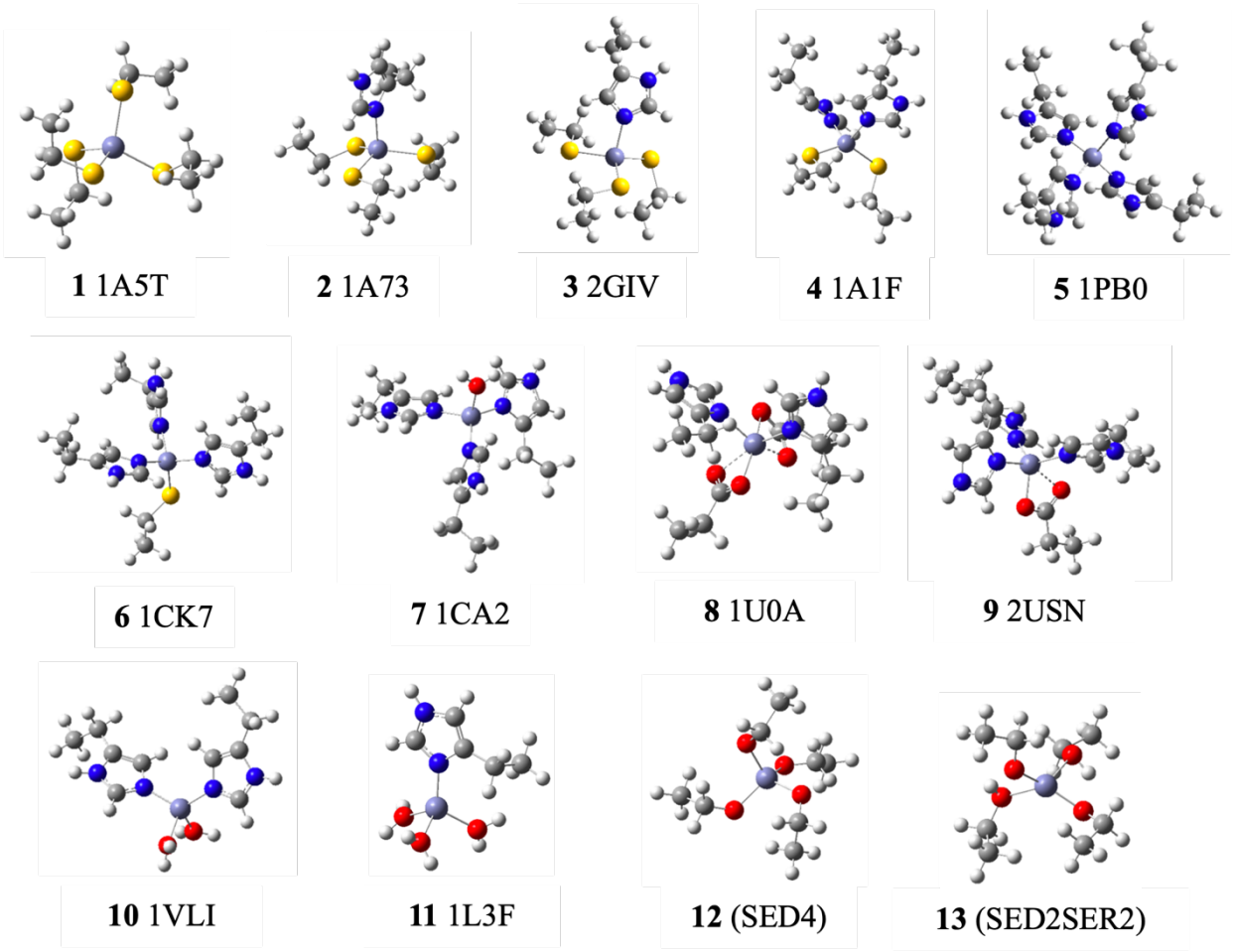


Figure 1. Optimized geometries of the training set systems.

As the CHARMM-FQ model was designed for proteins and our aim is to develop a FQ model for metalloproteins, we followed a similar strategy to the CHARMM-FQ model[11-12] for the parameterization. This was accomplished by utilizing the CT3 atom type as a global reference, with its χ and η parameters were from the CHARMM-FQ model[11] and were fixed during the parameter optimization. We also used the CHARMM-FQ parameters as the initial inputs for the other atom types during our parameter optimization. The relevant details of the CHARMM-FQ parameters are provided in **Table S2**.

In our FQ model, atom types were assigned by referring to the CHARMM-FQ model (**Table S2**), with small adjustments made to improve the model performance (discussed below). To optimize the model parameters, we employed the constrained optimization by linear approximation (COBYLA) method. Initially, we optimized the FQ parameters for each of the 13 systems individually, while fixing the CT3 parameters and allowing identical atom types to have varying parameters across different systems. Our goal was to achieve an agreement within 0.1e (where e is the proton charge) between the Qeq and CM5 charges for each atom. To meet this criterion, new atom types were introduced. For instance, in the histidine ligands, the CG and CD2 atoms (both having the CPH1 atom type in CHARMM force field) may have significantly different CM5 charges. Specifically, in the HIE residue of System 2, the CG and CD2 atoms have CM5 charges of 0.0894e and -0.0313e, respectively. This difference makes it challenging to reproduce

the CM5 charges of these two atoms if they have the same χ and η parameters. Consequently, we assigned different atom types to the CG and CD2 atoms in HIE, denoted as CPHG and CPHD, respectively.

Table S3 provides a weighted average parameter set for χ and η that was obtained using the optimized parameters for each individual residue. The weights were set based on the number of atoms belonging to the same atom type within each system. The atom types for each ligand in our model can be found in **Table S4**, while the 2D structures for atom names and atom types in each ligand are shown in **Table S5**. Using this weighted average parameter set, we optimized $T_{kk'}$ values for each of the five atom types bound to Zn, namely SCD in CYM, NR2 in HIE/HID, OW in water, OC in ASP, and OH1 in SER/SED. We recalculated the Qeq charges for the training set by using the optimized $T_{kk'}$ values and the weighted average χ and η parameters. **Table S7a-m** presents the CM5 and Qeq charge values for the training set, which were based on the established parameters listed in **Tables S3 and S6**. The Qeq charge is within 0.1e of the CM5 charge for each atom in the training set.

We tested the transferability of our FQ model by applying the optimized parameters to the test set (**Table 2**), which were selected based on a previous study.[35] In contrast to the training set, which exclusively comprised four-coordinated systems, the test systems featured zinc complexes with four, five, and six coordination. These test systems were manually constructed and optimized at the B3LYP/6-31G* level of theory as well, and their optimized geometries are presented in **Figure 2**. The CM5 and Qeq charge values for the test set are provided in **Table S8a-f**, and these results also showed that the Qeq charge falls within 0.1e of the FQ charge for each atom in the test set.

Table 2. Systems in the test set.

System	Complex	Total charge (e)
14	Zn–$H_{HID}H_{HID}H_{HIE}$S	+1
15	Zn–$H_{HID}H_{HID}$D	+1
16	Zn–DOO	+1
17	Zn–DOOOO	+1
18	Zn–DOOOOO	+1
19	Zn–OOOOOO	+2

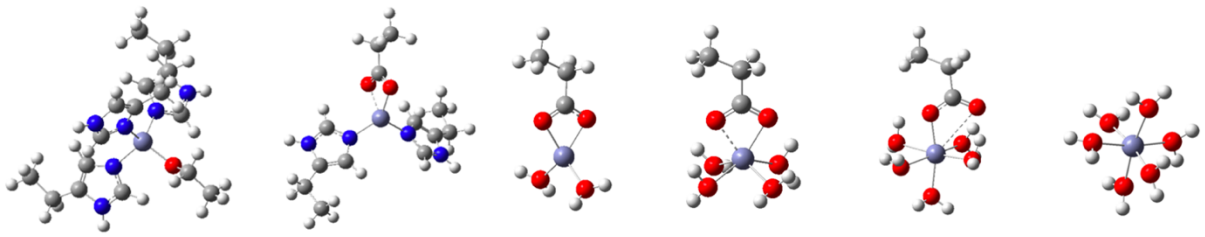


Figure 2. Optimized geometries of test set systems.

To evaluate the applicability of our model to metalloproteins, we conducted molecular dynamics (MD) simulations of a zinc finger domain from the PDB entry 2YRH.[36] The 2YRH entry has multiple structures determined by NMR experiment, the first structure was used in the present

study. 2YRH is comprised of two beta sheets and an alpha helix, as well as a metal site that contains two cysteine and two histidine ligands (**Figure 3**). To make a comparison, we performed MD simulations based on the RESP charges and Qeq charges for the metal site, respectively. The protonation states of the residues were determined by using the H++ webserver[37] under pH=7.0. Afterwards, the protonation states of the metal site residues were corrected if necessary. This is due to H++ not considering waters or metal ions when determining the protonation states of the amino acids. We then used the MCPB.py program[16] to parameterize the metal site, with the force constant calculations and RESP charge calculations performed at the B3LYP/6-31G* level of theory. Specifically, the small model[16] was used to perform the force constant calculation after geometry optimization at the B3LYP/6-31G* level of theory and the Seminario method[38] was used to generated the force constants based on the calculated Cartesian Hessian matrix. The large model[16] was used to perform the RESP fitting, with the hydrogen positions were optimized before performing the Merz-Kollman population analysis. The ChgModB scheme was used to fit the charges of metal site residues during the RESP charge fitting, which restrained the charges of backbone heavy atoms (CA, N, C, O) to the RESP charges in the AMBER force field and showed the best performance among different charge schemes.[15] Note that when we referred to the RESP scheme in the following section, it means the ChgModB scheme specifically.

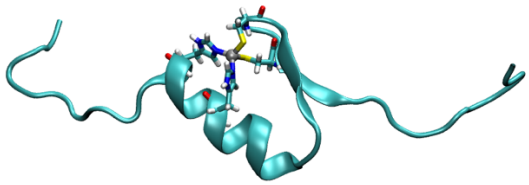


Figure 3. The first structure of the PDB entry 2YRH, with Zn depicted as a silver sphere and the four ligating residues depicted as color sticks.

Because our FQ model was parameterized for CH_3-capped metal sites, adjustments were made for the backbone atoms when applying the Qeq charges to metal site. First, the Qeq charges were calculated for the capped metal sites (using the same small model as mentioned above) based on our FQ model. Then, we assigned partial charges to backbone atoms C, O, N, and H with the RESP charges in the corresponding residues in the AMBER ff19SB force field.[26] These charges are conserved across the neutral residues in the AMBER force field, and they are coupled to the dihedral parameters of the protein backbone.[39] Herein, in order to conserve the total charge of the system, the charges of CA and HA atoms were adjusted by evenly distributing the remaining charge (total charge minus the charge sum of all the atoms) across the eight CA and HA atoms in the metal site. **Table S9** displays the final RESP and Qeq charge parameters used in the MD simulations. For comparison, the Qeq charges obtained for the capped metal site were also listed (i.e. termed as the pre-adjustment set).

The ff19SB force field[26] was used to model the protein, and the additional parameters of the metal site are provided in **Table S10**. A rectangular box of the OPC waters[40] was used to solvate the metalloprotein with a distance from the protein surface to the box edge of at least 10 Å, and Cl^- ions were added to neutralize the system. The Lennard-Jones parameters of Zn^{2+} were from Li et al.,[41] while the parameters of Cl^- ions were obtained from Sengupta et al.[42] After setting up, the system was simulated by the following steps for each of the two charge schemes (RESP and Qeq):

1) 10000 steps of minimization using the steepest descent algorithm followed by 10000 steps of minimization using the conjugate gradient algorithm. The protein heavy atoms were restrained along the minimization procedure.
2) 360 ps NVT MD to heat the system from 0 to 300 K in a gradual manner with a time step of 1 fs
3) 1 ns NPT equilibration at 300 K and 1 atm to correct the density with a time step of 1 fs
4) 100 ns NVT production at 300 K with a time step of 2 fs, with configurations saved every 10 ps, providing 10000 snapshots for analysis
5) The importance of replicas have obtained more attentions nowadays.[43] To test the robustness of these charge parameters, four additional replicas (five replicas in total) were performed with each replica following the above steps

For the above simulations, the particle mesh Ewald (PME) method[44] was used to handle the long-range electrostatics with a cut-off as 10 Å in the real space. The simulations were performed by using the PMEMD.CUDA program[45] in the AMBER 2020 software package.[46] The SHAKE algorithm[47] was used to constrain the bond lengths between heavy atoms and the connected hydrogen atoms, while a three-point SHAKE algorithm[48] was used for the water molecules. The Langevin thermostat was used to control the temperature with a collision frequency of 2 ps^{-1} while the Brendsen barostat was used to control the pressure with a pressure relaxation time of 1.0 ps.

Following the MD simulations, structural analyses were conducted on the trajectories to evaluate the performance of each charge scheme. The zinc finger domain of PDB entry 2YRH contains long coils in its two ends (see **Figure 3**) that exhibit significant fluctuations during MD simulations. To analyze the stability of the zinc finger domain, we calculated the root-mean-square-deviations (RMSDs) over the core region (residues 10-35) only, excluding the fluctuations from the two ends. Specifically, we calculated the RMSDs of the backbone N, CA, and C atoms in residues 10-35 by using the NMR structure as a reference (see **Figure 4**). We also calculated the RMSDs for all the heavy atoms of the metal site residues over the NMR structure, and the results are shown in **Figure 5**.

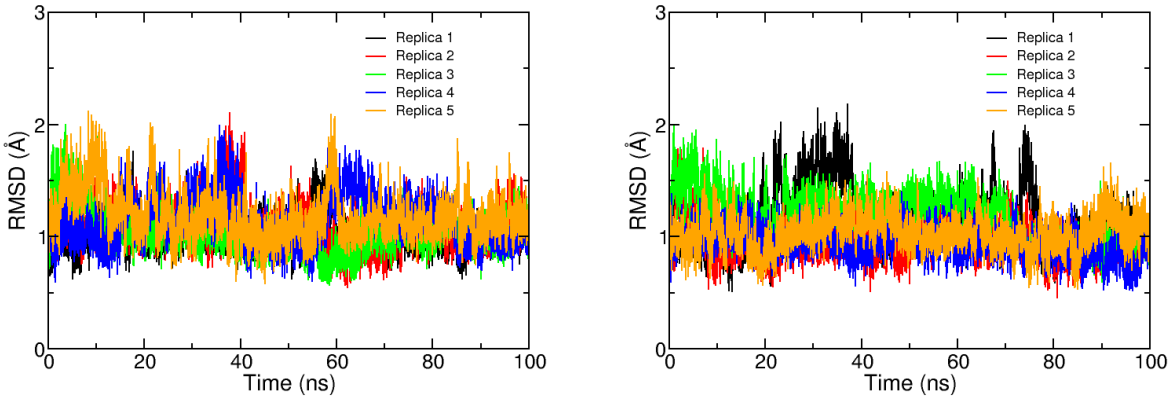


Figure 4. RMSDs of the backbone N, CA, and C atoms in residues 10-35 along the MD simulations based on the (left) Qeq and (right) RESP charge schemes.

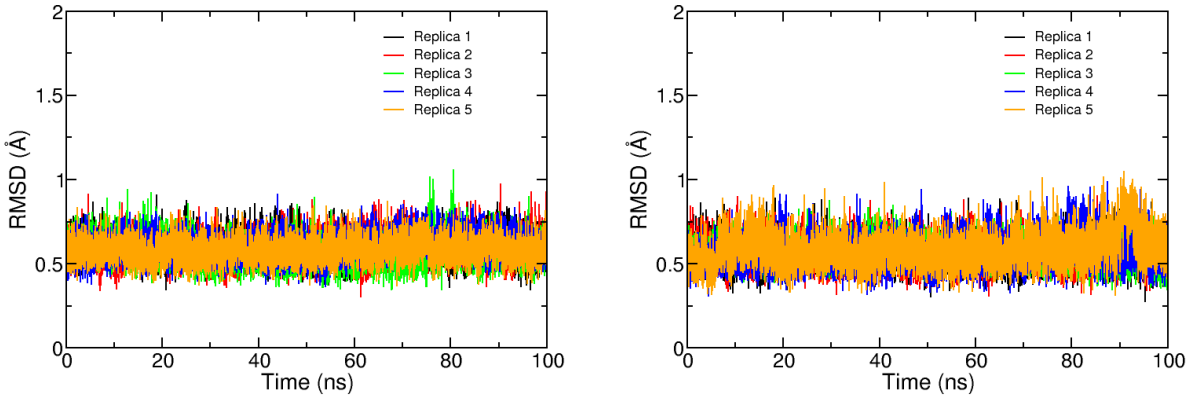


Figure 5. RMSDs of the heavy atoms of the metal site residues (residues CYM13, CYM16, HID29, HID33, and ZN45) along the MD simulations based on the (left) Qeq and (right) RESP charge schemes.

In addition, the root-mean-square-fluctuations (RMSFs) were calculated for each residue based on the backbone N, CA, and C atoms (**Figure S1**). During the RMSF calculations, frames were aligned based on the N, CA, and C atoms in residues 10-35. RMSF can be a measure of residue flexibilities. It is displayed that the terminal ends of the zinc finger domain are highly flexible and residues 10-35 exhibit much larger stability. In general, our RMSD and RMSF analyses indicated the MD simulations based on the Qeq charges showed comparable performance to the MD simulations based on RESP charges, suggesting that our model is transferable to a protein environment.

Conclusions

In comparison to quantum mechanical methods, fluctuating charge (FQ) models offer a computationally efficient alternative of determining atomic charges. Although various FQ models have been developed, few was particularly developed for metal complexes such as organometallic compounds and metalloproteins. In the present study, we developed a FQ model specifically for zinc-containing metalloproteins. Our FQ model has been parameterized to reproduce the CM5 charges. By using the EQeq+C scheme (a variant of the FQ model),[7] we reproduced the CM5 charges within 0.1e for each atom in both the training set systems and test set systems, although the training set only contains only four-coordinated zinc sites while the test set comprises of four-, five-, and six-coordinated zinc sites. These results indicate the excellent transferability of our models across diverse zinc sites. Moreover, we have adapted our FQ model to simulate a zinc finger protein domain in MD simulations, with the determined charges showed comparable performance as the RESP charges. Due to its excellent accuracy, speed, and transferability, this FQ model could be used for simulating zinc containing metalloproteins in MD simulations and molecular docking. Meanwhile, this model can be extended to systems containing other metal ions.

Acknowledgement

This work was supported by the Loyola start-up funds (to P.L.). We thank the XSEDE platform[49] which was funded by the National Science Foundation (grant number: ACI-1548562) for the

computational resources. Specifically, this work utilized the CPU and GPU resources on Expanse at the San Diego Supercomputer Center (allocation numbers: TG-CHE200102 and TG-BIO210105). In addition, we acknowledge ChatGPT (https://openai.com/blog/chatgpt) for optimizing the language of the manuscript.

Supporting Information

The supporting information is available free of charge:

The optimized parameters of the fluctuating charge model; Atom names and atom types for the ligands in the training set and test set; Calculated CM5 charges and Qeq charges for the systems in the training set and test set; RESP and Qeq charges for the metal site of the 2YRH system; Additional force field parameters for the 2YRH system; RMSFs of different residues for the MD simulations employing the RESP charges or Qeq charges (PDF).

Corresponding Author

Pengfei Li, pli4@luc.edu

Funding

This work was funded by the start-up funds from Loyola University Chicago (to P.L.).

ORCIDs

Luke Landry: 0009-0004-9289-4714
Pengfei Li: 0000-0002-2572-5935

Notes

The authors declare no competing financial interest.

References

1. Mulliken, R. S., A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *J. Chem. Phys.* **1934,** *2*, 782-793.
2. Mulliken, R. S., Electronic structures of molecules XI. Electroaffinity, molecular orbitals and dipole moments. *J. Chem. Phys.* **1935,** *3*, 573-585.
3. Parr, R. G.; Pearson, R. G., Absolute hardness: companion parameter to absolute electronegativity. *J. Am. Chem. Soc.* **1983,** *105*, 7512-7516.
4. Rappe, A. K.; Goddard III, W. A., Charge equilibration for molecular dynamics simulations. *J. Phys. Chem.* **1991,** *95*, 3358-3363.
5. Ramachandran, S.; Lenz, T.; Skiff, W.; Rappé, A., Toward an understanding of zeolite Y as a cracking catalyst with the use of periodic charge equilibration. *J. Phys. Chem.* **1996,** *100*, 5898-5907.
6. Wilmer, C. E.; Kim, K. C.; Snurr, R. Q., An extended charge equilibration method. *J. Phys. Chem. Lett.* **2012,** *3*, 2506-2511.
7. Martin-Noble, G. C.; Reilley, D.; Rivas, L. M.; Smith, M. D.; Schrier, J., EQeq+ C: an empirical bond-order-corrected extended charge equilibration method. *J. Chem. Theory Comput.* **2015,** *11*, 3364-3374.
8. Mortier, W. J.; Ghosh, S. K.; Shankar, S., Electronegativity-equalization method for the calculation of atomic charges in molecules. *J. Am. Chem. Soc.* **1986,** *108*, 4315-4320.

9. Yang, Z.-Z.; Wang, C.-S., Atom− bond electronegativity equalization method. 1. Calculation of the charge distribution in large molecules. *J. Phys. Chem. A* **1997,** *101*, 6315-6321.
10. Chen, J.; Martínez, T. J., QTPIE: Charge transfer with polarization current equalization. A fluctuating charge model with correct asymptotics. *Chem. Phys. Lett.* **2007,** *438*, 315-320.
11. Patel, S.; Brooks III, C. L., CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. *J. Comput. Chem.* **2004,** *25*, 1-16.
12. Patel, S.; Mackerell Jr, A. D.; Brooks III, C. L., CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **2004,** *25*, 1504-1514.
13. Finkelstein, J., Metalloproteins. *Nature* **2009,** *460*, 813-813.
14. Noffke, A. L.; Habtemariam, A.; Pizarro, A. M.; Sadler, P. J., Designing organometallic compounds for catalysis and therapy. *Chem. Commun.* **2012,** *48*, 5219-5246.
15. Peters, M. B.; Yang, Y.; Wang, B.; Fusti-Molnar, L.; Weaver, M. N.; Merz Jr, K. M., Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *J. Chem. Theory Comput.* **2010,** *6*, 2935-2947.
16. Li, P.; Merz Jr, K. M., MCPB. py: a python based metal center parameter builder. *J. Chem. Inf. Model.* **2016,** *56*, 599-604.
17. Andreini, C.; Banci, L.; Bertini, I.; Rosato, A., Counting the zinc-proteins encoded in the human genome. *Journal of proteome research* **2006,** *5*, 196-201.
18. Anzellotti, A.; Farrell, N., Zinc metalloproteins as medicinal targets. *Chem. Soc. Rev.* **2008,** *37*, 1629-1651.
19. Maret, W.; Li, Y., Coordination Dynamics of Zinc in Proteins. *Chem. Rev.* **2009,** *109*, 4682-4707.
20. Parkin, G., Synthetic analogues relevant to the structure and function of zinc enzymes. *Chem. Rev.* **2004,** *104*, 699-768.
21. Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005,** *26*, 1668-1688.
22. Cornell, W.; Cieplak, P.; Baylyl, C.; Gould, I.; Merez, J.; KM, F.; DM, S.; DC, F., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc* **1995,** *117*, 5179.
23. Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G., Charge model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* **2012,** *8*, 527-541.
24. Meng, E. C.; Lewis, R. A., Determination of molecular topology and atomic hybridization states from heavy atom coordinates. *J. Comput. Chem.* **1991,** *12*, 891-898.
25. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993,** *97*, 10269-10280.
26. Tian, C.; Kasavajhala, K.; Belfon, K. A.; Raguette, L.; Huang, H.; Migues, A. N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q., ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **2019,** *16*, 528-552.
27. Wang, W.; Wolf, R., J.; Caldwell, JW; Kollman, PA; Case. *J. Comput. Chem.* **2004,** *25*, 92.

28. Laio, A.; VandeVondele, J.; Rothlisberger, U., D-RESP: Dynamically generated electrostatic potential derived charges from quantum mechanics/molecular mechanics simulations. *J. Phys. Chem. B* **2002,** *106*, 7300-7307.
29. Francl, M. M.; Carey, C.; Chirlian, L. E.; Gange, D. M., Charges fit to electrostatic potentials. II. Can atomic charges be unambiguously fit to electrostatic potentials? *J. Comput. Chem.* **1996,** *17*, 367-383.
30. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A., Application of RESP charges to calculate conformational energies, hydrogen bond energies, and free energies of solvation. *J. Am. Chem. Soc.* **2002,** *115*, 9620-9631.
31. Lee, Y.-m.; Lim, C., Physical basis of structural and catalytic Zn-binding sites in proteins. *J. Mol. Biol.* **2008,** *379*, 545-553.
32. 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: Structure, Function, and Bioinformatics* **2007,** *69*, 466-475.
33. Auld, D. S., Zinc coordination sphere in biochemical zinc sites. *Zinc biochemistry, physiology, and homeostasis: recent insights and current trends* **2001**, 85-127.
34. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
35. Ahlstrand, E.; Hermansson, K.; Friedman, R., Interaction energies in complexes of Zn and amino acids: a comparison of Ab Initio and force field based calculations. *J. Phys. Chem. A* **2017,** *121*, 2643-2654.
36. Qin, X. R.; Hayashi, F.; Yokoyama, S., Solution structure of the C2H2-type zinc finger domain (699-729) from zinc finger protein 473. **2007**, DOI: 10.2210/pdb2YRH/pdb.
37. Gordon, J.; Myers, J.; Folta, T.; Shoja, V., Health LS, Onufriev A. *H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res* **2005,** *33*, 368-371.
38. Seminario, J. M., Calculation of intramolecular force fields from second-derivative tensors. *Int. J. Quantum Chem* **1996,** *60*, 1271-1277.
39. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995,** *117*, 5179-5197.
40. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *J Phys Chem Lett* **2014,** *5*, 3863-3871.

41. Li, Z.; Song, L. F.; Li, P.; Merz Jr, K. M., Systematic parametrization of divalent metal ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *J. Chem. Theory Comput.* **2020,** *16*, 4429-4442.
42. Sengupta, A.; Li, Z.; Song, L. F.; Li, P.; Merz Jr, K. M., Parameterization of monovalent ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. *J. Chem. Inf. Model.* **2021,** *61*, 869-880.
43. Knapp, B.; Ospina, L.; Deane, C. M., Avoiding false positive conclusions in molecular simulation: the importance of replicas. *J. Chem. Theory Comput.* **2018,** *14*, 6127-6138.
44. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
45. Explicit, G., Routine microsecond molecular dynamics simulations with Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput* **2013,** *9*, 3878-3888.
46. D.A. Case, K. B., I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz,R Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev, F. Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York and P.A. Kollman, AMBER 2020. **2020**.
47. JP, C. R., Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977,** *23*, 15.
48. Miyamoto, S.; Kollman, P., SETTLE: an analytical version of the shake an RATTLE algorithms for molecular simulation. *J. Comput. Chem* **1992,** *13*, 952-962.
49. Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D., XSEDE: accelerating scientific discovery. *Computing in science & engineering* **2014,** *16*, 62-74.