JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Comparison of Methods to Obtain Force-Field Parameters for Metal Sites

LiHong Hu[†,‡] and Ulf Ryde*[,‡]

[†]School of Computer Science and Information Technology, North-east Normal University, Changchun, 130024, People's Republic of China

[‡]Department of Theoretical Chemistry, Lund University, Chemical Centre, Post Office Box 124, SE-221 00 Lund, Sweden

Supporting Information

ABSTRACT: We have critically examined and compared various ways to obtain standard harmonic molecular mechanics (MM) force-field parameters for metal sites in proteins, using the 12 most common Zn^{2+} sites as test cases. We show that the parametrization of metal sites is hard to treat with automatic methods. The choice of method is a compromise between speed and accuracy and therefore depends on the intended use of the parameters. If the metal site is not of central interest in the investigation, for example, a structural metal far from the active site, a simple and fast parametrization is normally enough, using either a nonbonded model with restraints or a bonded parametrization based on the method of Seminario. On the other hand, if the metal site is of central interest in the investigation, a more accurate method is needed to give quantitative results, for example, the method by Norrby and Liljefors. The former methods are semiautomatic and can be performed in seconds, once a quantum mechanical (QM) geometry optimization and frequency calculation has been performed, whereas the latter method typically takes several days and requires significant human intervention. All approaches require a careful selection of the atom types used. For a nonbonded model, standard atom types can be used, whereas for a bonded model, it is normally wise to use special atom types for each metal ligand. For accurate results, new atom types for all atoms in the metal site can be used. Atomic charges should also be considered. Typically, QM restrained electrostatic potential charges are accurate and easy to obtain once the QM calculation is performed, and they allow for charge transfer within the complex. For negatively charged complexes, it should be checked that hydrogen atoms of the ligands get proper charges. Finally, water ligands pose severe problems for bonded models in force fields that ignore nonbonded interactions for atoms separated by two bonds. Complexes with a single water ligand can normally be accurately treated with a bonded potential, once it is ensured that the water H atoms have nonzero Lennard-Jones parameters. However, for metal sites with several water molecules, a nonbonded model with restraints (taken from the QM calculations) is more stable.

INTRODUCTION

Molecular mechanics (MM) simulations have become an important complement to experiments for obtaining structural and mechanistic information on biological systems at an atomic level.[1] For example, in the great majority of X-ray and NMR structure determinations, the experimental data are supplemented by MM calculations to give chemically reasonable bond lengths and angles.[2] Moreover, MM calculations and molecular dynamics (MD) simulations have become an important ingredient in biochemical and medicinal chemistry studies.[3,4]

The advantage with MM methods is their speed: With today's computer resources and software, you can study even big protein complexes with full atomic detail and simulate medium-sized proteins for hundreds of nanoseconds. On the other hand, the MM methods need to be parametrized, that is, you need to have MM parameters for all atoms in the system of interest. For biochemical macromolecules, this does not pose any problem, because standard MM parameters are available for all normal amino acids and nucleic acids.[5,6] Moreover, more general force fields are available for other molecules, such as carbohydrates and small druglike molecules.[7−16]

However, metal sites constitute a major problem for force fields.[1,17−22] The reason for this is that the strength of metal−ligand bonds is intermediate between that of covalent bonds and nonbonded interactions, such as hydrogen bonds. Moreover, metals can have many different types of ligands and the number of bonds around a metal is often more than the number of covalent bonds around an atom. Therefore, the metal-coordination sphere is often flexible with several different geometries possible. Finally, for transition metals, quantum mechanical ligand-field, spin-state, trans, and Jahn−Teller effects also become important, which are hard to model in a standard MM force field.[23] Therefore, MM parametrizations for metals have traditionally been restricted to specific metal sites (with a given set of ligands), for which accurate results can be obtained,[24−27] or metal-specific force fields requiring specialized software.[23,28,29]

There are several approaches to incorporate metal ions into MM force fields. The simplest one is to describe the interaction between the ion and its ligand entirely by nonbonded interactions, that is, by electrostatics and Lennard-Jones terms (the nonbonded or ionic method). Such a model has been suggested for Zn, based on formal charges on the ion (+2).[30] This model has also been modified to include polarization and charge transfer,[31] or dummy atoms between the metal and the ligands.[32−34] Sometimes, the nonbonded potential has been supplemented by

Received: December 17, 2010
Published: June 30, 2011

metal−ligand restraints to ensure that the ligands stay bound.[35] The restraints can be of many types, for example, harmonic or flat-bottomed, single- or double-sided.

Alternatively, explicit bonds are defined between the metal and its ligand, a bonded (valence) model. Then, the metal−ligand bonds are treated the same way as covalent bonds, that is, with bond, angle, and dihedral terms (although the latter are often ignored for metal sites).[24,25,36,37] This bonded model is typically combined with a nonbonded potential (electrostatics and Lennard-Jones terms), using either formal charges on the metal or charges derived from quantum mechanical (QM) calculations.[24] The points-on-a-sphere approach is a mixed model in which a metal−ligand stretching function is used, but the ligand−metal−ligand bond angles are replaced by 1,3-nonbonded interactions for the coordinating atoms.[17,38,39] Other more specialized methods to treat metal ions also exist, for example, ligand-field MM,[23] Valbond,[28] the universal force field (UFF),[40,41] YETI,[42,43] or SIBFA.[44−46]

Likewise, there exist several methods to perform a force-field parametrization, which depend on the force field used. For example, for a nonbonded force field with formal charges, only two parameters (the Lennard-Jones parameters for the metal) are available, and they can easily be optimized to reproduce QM data, for example, to give proper metal−ligand distances. However, for a bonded force field, you have many more parameters that need to be optimized. In principle, these parameters can also be obtained from QM calculations, for example, by taking the optimum metal−ligand bond lengths and angles from the values observed in QM-optimized structures of model complexes.[24,25] The corresponding force constants can be obtained from QM scans of the potential surface for the bond or angle, but this is tedious for more than a few parameters. Instead, it is more common to obtain the force constants from a projection of the Hessian matrix (obtained from a QM frequency calculation) into internal coordinates.[16,24,25] Unfortunately, such a procedure is ambiguous because the total number of internal coordinates (bond, angles, and dihedrals) around a metal is more than the degrees of freedom (for example, around a four-coordinate metal, four bonds and six angles can be defined, but there are only $5 \times 3 - 6 = 9$ degrees of freedom). This means that the internal coordinates are not independent and different choices of internal coordinates will give different force constants.[47,48] Moreover, it is not certain that the optimum bond lengths and angles obtained from a QM optimization represent unstrained equilibrium parameters (as is assumed for a MM force field). Instead, they represent an optimum compromise for all interactions (bonded as well as nonbonded) between all atoms in the complex.

In 1996, Seminario[47] suggested an approach to obtain force constants directly from the Hessian matrix, thereby avoiding any use of internal coordinates. This procedure has been employed in automatic parametrization programs by at least three groups, for example, Hess2FF and the metal-center parameter builder.[48−50] It makes the extraction of force constants from the Hessian unambiguous (in fact, the force constants can be obtained in two ways, but they typically give similar results that can be averaged).[48] Test calculations have shown that the Seminario approach gives better force constants than the method involving internal coordinates.[49] Unfortunately, the interdependence of the various internal coordinates still exists. Moreover, this approach involves a double-counting of electrostatic and Lennard-Jones interactions: The QM Hessian matrix contains all interactions, including electrostatic and van der Waals interactions. However, the Hessian is used only to extract the bonded interactions, whereas electrostatics and van der Waals interactions are calculated by separate terms by the MM program. For bonds and angles, this is no problem, because most force fields ignore nonbonded interactions between atoms one or two bonds apart. However, for dihedrals, it is a serious problem, because most force fields complement dihedral terms with nonbonded energies, typically scaled down by a constant factor (for example, in the AMBER force field, used in this paper, 1,4-electrostatics are scaled down by a factor of 1.2 and 1,4-Lennard-Jones interactions by a factor of 2.0). This means that the dihedral parameters will already contain some nonbonded interactions, which then are double-counted in the MM calculations. Thus, the Seminario approach is only approximate.

Norrby and Liljefors and co-workers[51,52] have suggested an approach that solves these problems. It involves a complete optimizations of all parameters of the force field in an iterative manner. For every set of parameters, the MM structure is optimized, so there is no risk of double-counting any interactions and it is not assumed that the geometric parameters in the QM structure represent equilibrium values in the MM force field. However, QM data are still used as the reference, e.g. bond lengths, angles, dihedrals from the QM structure, as well as the Hessian elements, and the fitting procedure ensures that the optimized MM structure is as close as possible to the QM structure. This approach was originally developed for the MM3 force field,[53] but it has recently been implemented also for the AMBER force field and software.[54]

In this paper, we compare five different approaches to obtain MM parameters for metal sites in proteins using standard nonpolarizable harmonic force fields, viz., two variants of a nonbonded potential, a restrained nonbonded potential, the Seminario approach,[47] as implemented in the Hess2FF software,[48] the zinc AMBER force field (ZAFF; also based on the Seminario approach),[49] and the ideal procedure of Norrby and Liljefors (NL).[51,54] As test cases, we use 12 simple models for the most common Zn^{2+} sites in proteins. Sites of this type were also used in the previous studies.[49,50] As a reference, we use the corresponding structures optimized by QM methods, and we compare with the structures and Hessian elements of the structures optimized with the various force fields. This comparison also allows us to discuss various problems that are typically encountered during the parametrization of metal sites. The take-home message is that the method of choice depends on the intended use of the force field and that the parametrization of metal sites is seldom an automatic procedure.

METHODS

Structures. As test cases, we used 10 models of the most common Zn^{2+} sites in proteins, taken from the previous investigation by Merz and co-workers:[50] $ZnCys_4$, $ZnCys_3His$, $ZnCys_2His_2$, $ZnCysHis_3$, $ZnHis_4$, $ZnHis_3H_2O$, $ZnHis_2(H_2O)_2$, $ZnHis(H_2O)_3$, $ZnHis_3Asp$, and $ZnHis_2Asp_2$. For two of the models ($ZnCys_3His$ and $ZnHis_3H_2O$), alternative coordinating atoms of the His groups were also tested.[50] In these models, Cys was modeled by CH_3S^-, His by methylimidazole, and Asp by CH_3COO^-. The starting structures for the optimizations were taken from typical crystal structures, selected as in the previous investigation.[50] The structures were protonated by the tleap

routine in AMBER-10[55] and then truncated with hydrogen atoms (these H atoms will be called HT below).

Quantum Mechanical Calculations. The structures were optimized with the hybrid density-functional theory method B3LYP,[56−58] using the 6-31G* basis set[59−61] (again following the previous study).[50] After the geometry optimization, analytic frequencies were calculated at the same level of theory. This structure and the Hessian matrix obtained from the frequency calculation were used as a reference in both the NL and Hess2FF parametrizations and as the target for all the MM minimizations. Charges on all atoms were calculated with the RESP approach, as implemented in AMBER-10,[55] using electrostatic potentials calculated at the B3LYP/6-31G* level and sampled with the Merz−Kollman scheme,[62] albeit with a higher-than-default density of points (10 concentric layers with 17 points/Å^2). All QM calculations were performed with Gaussian-09.[63] Coordinates of the optimized QM structures are given in Table S10 in the Supporting Information.

Force Fields, Parametrizations, and Molecular Mechanical Calculations. All the MM calculations were run with the sander (minimization) and nmode (Hessian calculation) modules of the AMBER-10 software.[55] Two sets of nonbonded force fields were studied (NB1 and NB2). They differed in the charges and Lennard-Jones parameters: One set used a formal charge of +2 on Zn and standard AMBER charges on all the ligand atoms (corrected to a proper net charge for each ligand on the HT atom; NB2). All the other force fields (including the bonded potentials) used the QM RESP charges for all atoms (note that ZAFF normally uses charges calculated for larger models that include full capped amino acids,[50] but for the truncated models we used the RESP charges also for ZAFF). Likewise, one set of nonbonded force field (NB1) used the Stote−Karplus Lennard-Jones parameters for Zn,[30] R = 1.95 Å and ε = 1.046 kJ/mol, whereas all the other force fields employed the Merz parameters, R = 1.1 Å and ε = 0.0523 kJ/mol.[24,50] On all the other atoms, standard AMBER Lennard-Jones parameters were used, except sometimes for water, as will be discussed below.

Atom types and bonded parameters for ZAFF were extracted from the Supporting Information of the original publication.[50] The restrained nonbonded potential used RESP charges, the Merz Lennard-Jones parameters, and a double-sided harmonic potential with the minimum distance taken from the QM structure and the force constant taken from the Hess2FF[48] calculation (distinct distances and force constants were used for all four ligands).

For the Hess2FF and NL force fields, two sets of atom types were tested. The first (AT_{min}) used standard AMBER atom types for all ligands (and the same atom type for the HT atom as the other hydrogen atoms bound to the same carbon atom). A distinct atom type was used for each Zn-ligating atom. This means that only bonded interactions involving Zn were parametrized; all the other bonded parameters were taken from the AMBER-99SB force field.[64,65] The second set (AT_{max}) used individual atom types for all atoms in the complex. Thus, all bonded parameters of the complex were optimized.

The Hess2FF force field was obtained with the Hess2FF program.[48] This program reads the output file of the Gaussian frequency calculation and calculates all the bonded parameters via the Seminario approach.[47] For the AT_{max} set, the program is completely automatic and generates AMBER topology and parameter files. For the AT_{min} set, a file with the desired atom types is provided as input. The program then gives the proper averaging of all parameters. Further instructions for the program can be found in http://www.teokem.lu.se/~ulf/Methods/parm.html.

The NL force field was constructed according to the method developed by Norrby and Liljefors,[51] using the recent implementation for AMBER.[55] This method minimizes a penalty function consisting of the deviation of geometries and Hessian elements between the QM and MM calculations, giving different weights to different kinds of data. The geometries were described as lists of all bonds, angles, and dihedral angles, rather than by absolute coordinates. The weight factors of the various data types were 100 Å^{-1} for bonds, 2 deg^{-1} for angles, 1 deg^{-1} for torsions, and 0.01−0.1 mol·Å^2/kcal for Hessian elements (0.01 for elements involving interactions of an atom with itself, 0.02 for atoms bound to each other, 0.04 for atoms connected by two bonds, 0.1 for atoms connected by three bonds, and 0.01 for all other elements).[51,66]

The iterative NL optimizations were started from the corresponding Hess2FF force field. After convergence, the force field was checked. Typically, some bonds and angles get zero force constants in the first runs of the parametrizations. These were reset to reasonable values and force constants of the other angles around the same central atom were reduced, and then the parametrization was run again. This was repeated until all bonds and angles had nonzero force constants and all other parameters were reasonable. Further instructions for the procedure are found in http://www.teokem.lu.se/~ulf/Methods/ponparm.html. Both the NL and the Hess2FF programs are available from the authors upon request.

Molecular Dynamics Simulations. Two Zn-containing proteins were studied with molecular dynamics (MD) simulations, viz. the δ' subunit of the clamp-loader complex of DNA polymerase III (PDB file 1A5T),[67] which contains a single Cys_4 Zn finger, and the spore coat polysaccharide biosynthesis protein SpsE (PDB file 1VLI),[68] which contains a $ZnHis_2(H_2O)_2$ site. The two structures were protonated with the tleap module in AMBER,[55] with the assumption that all Asp and Glu residues are negatively charged and all Lys and Arg residues are positively charged. The protonation state of the His residues was decided from a detailed study of the hydrogen-bond network and the local surroundings of these sites. The two His residues in 1VLI that coordinate to Zn were protonated on the ND1 atom, whereas the other His residues were assumed to be protonated on the NE2 atom. For 1A5T, His residues 24, 73, 103, and 238 were assumed to be protonated on the NE2 atom and the other residues were assumed to be doubly protonated. A few side-chain atoms that were not resolved in the 1VLI crystal structure were built with the tleap software. Residues 65−73 were also missing in the structure and they were ignored in the calculations.

Both proteins were solvated in a periodic octahedral box with water molecules extending at least 9 Å from the protein on all sides, keeping the crystallographic water molecules. Six simulations were performed. First, the systems were subjected to a 1000-step minimization, keeping all heavy atoms in the proteins restrained toward their positions in the crystal structure with a force constant of 418 kJ·mol^{-1}·Å^{-2}. Then, two 20 ps MD simulations were run with the same restraints. The first simulation was run with a constant volume and the second with a constant pressure. Finally, the box size was equilibrated by a 50-ps MD simulation with a constant pressure and without any restraints. Finally, an equilibration of 200 ps and a production simulation of 5 ns were run with a constant volume. During the latter run, coordinates were collected every 2 ps.

Table 1. Performance of the Various Force Fields for the $ZnHis_4$ Complex[a]

					Hess2FF		NL	
	NB1	NB2	restrained	ZAFF	AT_{min}	AT_{max}	AT_{min}	AT_{max}
r^2 Hessian	0.861	0.978	0.967	0.979	0.972	0.963	0.987	0.985
rmsd bonds	0.973	0.045	0.025	0.011	0.016	0.010	0.017	0.006
rmsd angles	7.3	3.1	5.0	2.3	2.2	0.4	2.2	0.2
rmsd dihedrals	14.4	2.3	3.3	2.6	3.7	3.6	3.0	1.0
rmsd coordinates	2.45	0.26	0.33	0.18	0.20	0.20	0.21	0.05
max coordinates	6.05	0.44	0.57	0.40	0.46	0.47	0.52	0.12

[a] The six quality measures are correlation coefficient (r^2) between all QM and MM Hessian elements; root-mean-squared deviations (rmsd) for all bonds, angles, dihedral angles, and coordinates between MM and QM optimized structures; and maximum deviation (max) for coordinates.

The MD calculations were run with the sander module in the Amber software,[55] using the Amber 1999SB force field (FF99).[64,65] For 1A5T, we used the ZAFF charges for the Zn site, whereas for 1VLI, the ZAFF charges are erroneous in the deposited files (e.g., 7×10^{247} e for one of the two H atoms in water).[50] Therefore, we recalculated these charges, following the instructions in the ZAFF paper (the charges are listed in Table S9 in the Supporting Information).[50] The NB2 simulations employed a +2 charge for the Zn ion and Amber 1999SB charges[65] for the other residues. Water was described explicitly with the TIP3P model.[69] Long-range electrostatics were treated with particle-mesh Ewald method[70,71] with a grid size of 80^3, a fourth-order B-spline interpolation, a tolerance of 10^{-5}, and a real-space cutoff of 8 Å. The temperature was kept constant at 300 K and the pressure was kept at 1 atm by use of the Berendsen weak-coupling algorithm[72] with time constants of 1 ps. The nonbonded pair list was updated every 50 fs. Bond lengths involving hydrogen atoms were constrained with the SHAKE algorithm,[73] allowing for a MD time step of 2 fs. However, for some force fields, several of the simulations crashed with SHAKE failure. This was solved by turning off SHAKE and reducing the time step to 0.5 fs.

RESULTS AND DISCUSSION

In this paper, we compare the performance of eight different Zn force fields on 12 model complexes of the most common Zn sites in proteins.[50] We will first discuss the results of a model complex with typical behavior, followed by the average results of all complexes. Then we will describe three complexes that illustrate two types of problems that are encountered when parametrizing metal complexes. The performance of the force fields is judged by how closely they reproduce the structure and the Hessian of the QM-optimized structure used for the parametrization. For the Hessian, we calculated the correlation coefficient (r^2) between all QM and MM Hessian elements. For the structure, we studied the root-mean-squared deviations (rmsd) for all bonds, angles, and dihedral angles (in Ångströms or degrees). Moreover, we will list the root-mean-squared deviations (rmsd) and maximum deviation for the coordinates (after a rmsd fit of the MM and QM structures).

$Zn(His)_4$ Model. The $[Zn(CH_3\text{-imidazole})_4]^{2+}$ model gives results that are typical for most of the complexes. From Table 1, it can be seen that the nonbonded potential with RESP charges and Stote−Karplus Zn Lennard-Jones parameters (NB1) gave the worst results, with a rmsd for the coordinates of 2.45 Å, illustrating that the structure has completely changed. The reason for this is that all Zn−N distances have increased (from 2.01 to 3.35, 3.56, 4.94, and 8.05 Å; cf. Figure 1a), illustrating that the parameters are not optimized for this complex. Much better results were obtained with a formal +2 charge for Zn (and standard AMBER charges for the other atoms), especially when combined with the Merz Zn Lennard-Jones parameters (NB2; cf. Figure 1b; the Stote−Karplus parameters gave appreciably worse results). Of course, even better results could be obtained by tuning the Lennard-Jones parameters for Zn, but because there are only two parameters available, only restricted improvements can be expected, especially for complexes with more than one type of ligands. This illustrates that nonbonded models are very sensitive to the nonbonded parameters and that a single set of parameters will not be optimal for all type of complexes.[17]

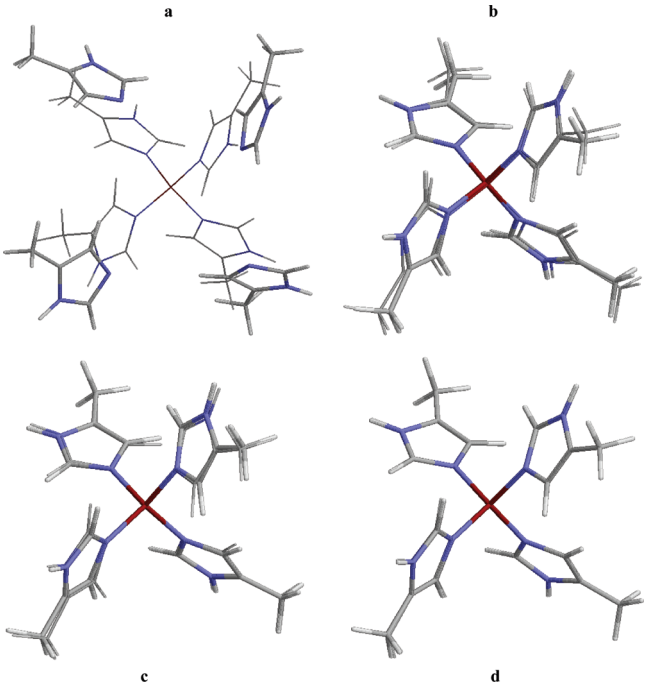

Figure 1. Comparison of QM structure (thin sticks) with MM structures obtained with (a) NB1, (b) NB2, (c) ZAFF, and (d) NL−AT_{max} force fields (thick lines) for the $Zn(His)_4$ model.

The restrained nonbonded potential gave much better result than NB1 but slightly worse than NB2 (because it is based on the RESP charges), although the rmsd for the bonds is better. This shows that the restraints ensure that the ligands do not dissociate.

Table 2. Average Performance of the Various Force Fields Tested for the 12 Model Complexes[a]

					Hess2FF		NL	
	NB1	NB2	restrained	ZAFF[b]	AT_{min}	AT_{max}	AT_{min}	AT_{max}
r^2 Hessian	0.742	0.882	0.833	0.907	0.896	0.928	0.985	0.992
rmsd bonds	0.213	0.048	0.020	0.022	0.032	0.012	0.012	0.003
rmsd angles	3.6	3.2	3.3	2.6	2.9	1.9	1.6	0.4
rmsd dihedrals	11.9	10.1	10.3	8.2	10.3	8.4	5.5	2.3
rmsd coordinates	1.25	0.56	0.50	0.33	0.34	0.26	0.16	0.07
max coordinates	2.31	1.10	1.04	0.79	0.71	0.57	0.39	0.15

[a] Quality measures are the same as in Table 1. Raw data are given in Tables 1 and 3−5 and in Tables S1−S8 in Supporting Information. [b] ZAFF failed for two complexes [$ZnHis_3(H_2O)$ and $ZnHis_2(H_2O)_2$], which are omitted from the average only for ZAFF.

Table 3. Performance of the Various Force Fields for the $ZnCys_4$ Complex[a]

					Hess2FF		NL	
	NB1	NB2	restrained	ZAFF	AT_{min}	AT_{max}	AT_{min}	AT_{max}
r^2 Hessian	0.869	0.836	0.384	0.853	0.901	0.904	0.995	0.993
rmsd bonds	0.330	0.245	0.062	0.076	0.016	0.014	0.006	0.004
rmsd angles	4.4	2.1	3.6	1.2	0.6	0.4	0.4	1.0
rmsd dihedrals	10.0	14.9	37.1	13.1	13.2	13.2	13.2	12.6
rmsd coordinates	0.49	0.46	0.30	0.32	0.32	0.32	0.31	0.29
max coordinates	0.74	0.76	1.29	0.44	0.54	0.54	0.54	0.53

[a] Quality measures are the same as in Table 1.

All the bonded force fields gave better results than the nonbonded ones, with RMSDs for the coordinates of only 0.05−0.21 Å (cf. Figure 1c). As expected, the ZAFF and Hess2FF−AT_{min} approaches gave similar results; they employ essentially the same methodology and differ only in details of the implementation (for example, ZAFF have zeroed all force constants for dihedrals involving Zn). Interestingly, the NL−AT_{min} results are also similar (NL−AT_{min} gives better Hessian and angles, whereas ZAFF gives better bonds, dihedrals, and coordinates). Of course, NL approach can be tuned to give better results for the other quality measures by changing the weights in the NL penalty function.

Even better results can be obtained if all atom types are optimized (AT_{max}), with both Hess2FF and NL. This shows that the accuracy of ZAFF and the other AT_{min} force fields is mainly limited by standard AMBER force-field parameters, used for the imidazole ligands. However with AT_{max}, NL gives much better results than Hess2FF in all quality measures. In fact, NL−AT_{max} gives a nearly perfect fit, as can be seen in Figure 1d. This shows that NL is inherently a much better approach for parametrization and that the rather poor results with the AT_{min} atom types mainly reflected problems in the AMBER parameters.

Other Models. Most of the other models gave results similar to that of the $Zn(His)_4$ model. Therefore, the results are not discussed in detail (the performance of each model is shown in Tables S1−S8 in the Supporting Information). Instead we list the average quality measures of each MM force field for all 12 models in Table 2. From these data, it can be seen that NL−AT_{max} gives the best performance for all six quality measures. In fact, the rmsd for the bonds and angles are 4 times better than the second-best method, and the rmsd for the dihedrals and the rmsd and maximum deviation of the coordinates are half that of the second-best method. This shows that this approach is considerably more accurate than the other methods.

The next method is NL−AT_{min}, which is second in all quality measures, even if the rmsd for the bonds is similar to that of Hess2FF−AT_{max}. The latter method is third in performance for all quality measures except the rmsd for dihedrals, for which ZAFF gives slightly better results.

The ZAFF force field gives slightly better results than the Hess2FF−AT_{min} force field in all quality measures except the maximum coordinate error. On the other hand, we will see below that ZAFF failed for two of the complexes, which are omitted from the ZAFF average but not from the averages of the other methods; if these two complexes are omitted also from the Hess2FF−AT_{min} average, the latter method is actually better than ZAFF in all quality measures (giving 0.951, 0.022 Å, 2.66°, 8.2°, 0.32 Å, and 0.44 Å for the six entries in Table 2). This shows that these two complexes are problematic also for the other methods and deteriorate the averages in Table 2.

The restrained and NB2 force fields give similar and somewhat varying results. NB2 gives a better Hessian and better angles and dihedrals, whereas the restrained model is better for the other quality measures. In fact, the restrained model gives better bonds than both ZAFF and Hess2FF−AT_{min}. The NB1 force field gives the worst results in all quality measures. Thus, we can conclude that the performance of the force fields is quite uniform over the 12 tested complexes and follows the order NL−AT_{max}, NL−AT_{min}, Hess2FF−AT_{max}, Hess2FF−AT_{min}, ZAFF, restrained, NB2, and NB1.

$Zn(Cys)_4$ Model. Three of the model complexes gave problems that often are encountered when parametrizing metal sites. Initial calculations on the $[Zn(CH_3S)_4]^{2-}$ complex gave unexpectedly poor results with all methods. In particular, the dihedral angles of the methyl groups were poorly reproduced, giving large

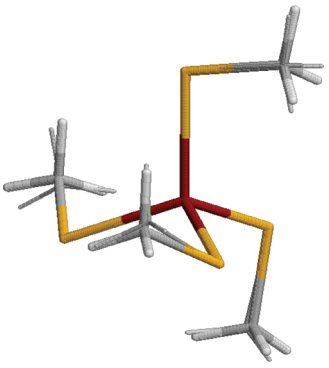

Figure 2. Comparison of QM (thin sticks) and Hess2FF-AT$_{max}$ structures of the ZnCys$_4$ model.

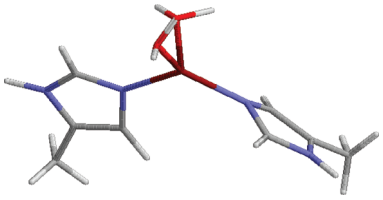

Figure 3. Structure of the Zn(His)$_2$(H$_2$O)$_2$ complex, optimized with the ZAFF force field, the step before the minimization crashes. Note that two of the H and O atoms overlap.

RMSDs of dihedrals and coordinates. The problem was traced to the charges of the hydrogen atoms, which turned out to be negative. In this small model, the general structure (the dihedrals) is completely determined by weak hydrogen bonds between the methyl groups and the sulfur atoms. If the hydrogen atoms have a negative charge, these interactions become repulsive, and the structure will completely change.

The improper charge of the hydrogen atoms is probably caused by the negative net charge (−2) of the complex; negatively charged complexes often give problems in QM calculations.[74] It can be avoided in many different ways, for example, by changing the method or basis set or by performing the calculations in a continuum solvent. In order to keep the calculations as similar as possible to the original ZAFF calculations, we decided to use a slightly different basis set, LanL2DZ.[75,76] It gave a positive charge on the hydrogen atoms and a geometry that was similar to that obtained with the 6-31G* basis set.

Results of the parametrizations obtained from the structures optimized with this basis set are shown in Table 3. The results are quite similar to those obtained for the other models. The nonbonded models gave the worst results, but this time the two nonbonded models gave similar results.

The three AT$_{min}$ force fields gave similar results. Moreover, the two AT$_{max}$ force fields gave slightly better results, but the improvement is not so large, because the model is so small (21 atoms) and the number of AMBER parameters is few (two bonds and one angle). The differences in coordinates are quite high for all bonded potentials because they form a somewhat different pattern of the S−HC hydrogen bonds (Figure 2). However, we have checked with QM calculations that the QM pattern is more stable than that obtained with MM.

ZnHis$_2$(H$_2$O)$_2$ and ZnHis(H$_2$O)$_3$ Models. For the two [Zn(CH$_3$-imidazole)$_2$(H$_2$O)$_2$]$^{2+}$ and [Zn(CH$_3$-imidazole)(H$_2$O)$_3$]$^{2+}$ models,

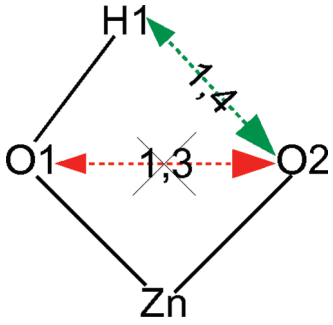

Figure 4. Schematic representation of nonbonded interactions between two Zn-bounded water molecules, illustrating that the O1−O2 1,3-interaction is ignored, whereas the H1−O2 1,4-interaction is included.

the ZAFF minimizations starting from the QM structures crashed because the O and H atoms from different water molecules overlap, as can be seen in Figure 3. This is caused by the Lennard-Jones parameters of the H atoms of water. The default water model in AMBER is TIP3P[69] and it has zeroed Lennard-Jones parameters on the hydrogen atoms (i.e., it interacts with other molecules only through the O atom, to increase computational speed). These parameters were also used in ZAFF. This is problematic for a metal complex, because there are typically other ligating atoms with a substantial negative charge (in our case, the coordinating O atom of the other water ligand; O2 in Figure 4). This atom is two bonds away from the O atom of the first water ligand (O1), implying that the Lennard-Jones and electrostatic interactions between these two atoms are ignored. On the other hand, the O2 atom and the H atom of the first water (H1) are three bonds away, implying that the electrostatics and Lennard-Jones interactions in AMBER are scaled down by factors of 1.2 and 2.0, respectively. However, if the Lennard-Jones parameters of H1 atoms are zero, it means that there is no repulsive interaction between the O2 and H1 atoms and therefore the electrostatic attraction may bring the two atoms together, until they reside on top on each other, giving an infinite electrostatic energy.

This problem can partly be avoided by adding nonzero Lennard-Jones parameters on the H atoms of water. For the Hess2FF and NL force fields, we use the same Lennard-Jones parameters as backbone HN groups (atom type H; this is the only polar H atom type in AMBER with nonzero Lennard-Jones parameters), R = 0.6 Å and ε = 0.657 kJ·mol^{-1}·Å^{-2}. From Table 4, it can be seen that this is enough to avoid problems in the minimizations.

However, it can be seen that the two Hess2FF force fields still give quite large rmsd values for the coordinates (0.9 and 1.6 Å). The reason for this is that the O−O distance becomes only 0.8−0.9 Å, giving strongly distorted structures (Figure 5a). This is caused by a related problem: Even if there are nonzero Lennard-Jones parameters on the H atoms, there is still no repulsion between the O atoms on different water molecules. Therefore, there will be a strong attraction between the H and O atoms on different water molecules, which only will be repelled at short distances, owing to the small radius of the H atom. In fact, there are four H−O interactions (hydrogen bonds) of 1.6 Å in the structure in Figure 5a, but with the wrong orientation (the O atoms are in the middle), owing to the missing O−O repulsion. This shows that the problem cannot be avoided by using a larger Lennard-Jones radius of the water H atom, because it will not

change the orientation (1.6 Å is a reasonable distance for a strong hydrogen bond).

On the other hand, the NL force fields gave excellent structures, especially with AT_{max}, as can be seen in Figure 5b. Apparently, a thorough parametrization of the bonded parameters can overcome the problems caused by water Lennard-Jones parameters, for example, by compensating the missing nonbonded O−O interaction by a large O−Zn−O force constant.

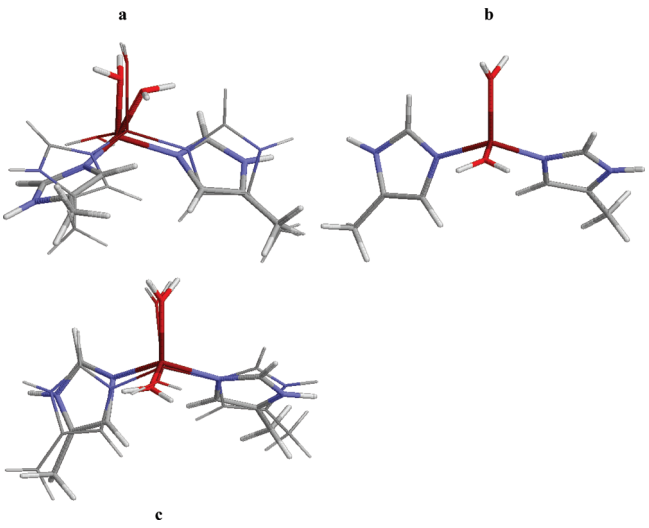

Figure 5. Comparison of QM (thin lines) with (a) Hess2FF−AT_{min}, (b) NL−AT_{max}, or (c) restrained nonbonded structures of the $Zn(His)_2(H_2O)_2$ models. Panel a shows that the Hess2FF−AT_{min} structure has a close O−O interaction and, hence, a strongly distorted structure. In panel b, the fit is so perfect (the rmsd of the coordinates is 0.003 Å) that the thin lines cannot be discerned.

It is also notable that the restrained nonbonded force field gave a better structure than Hess2FF, as can be seen in Table 4 and Figure 5c. The reason for this is that, in a nonbonded force field, there are no bonds to the Zn ligands, so the O−O interaction is fully active, ensuring that there is no great distortion of the structure. Moreover, the restraints on the Zn−ligand distances ensure that these distances are well reproduced, better than in the Hess2FF structures. On the other hand, there are no angles involving Zn, so the structure is far from perfect (the rmsd for the coordinates is 0.5 Å). This shows that the restrained nonbonded model is an alternative to a bonded model with more than one water molecule.

Similar results apply also to the $ZnHis(H_2O)_3$ complex (Table 5): The ZAFF minimization crashed with overlapping H and O atoms. Hess2FF-AT_{min} (with nonzero H parameters for water) managed the minimization but gave a quite poor structure, this time caused by decreased Zn−O bond lengths (~1.6 Å, to improve the H−O hydrogen bonds). On the other hand, Hess2FF-AT_{max}, NL−AT_{min}, and especially NL−AT_{max} gave accurate structures. The restrained nonbonded model is better than both nonbonded models.

A natural question is why the problems with the water molecule and the Lennard-Jones parameters are seen only with two or more water molecules and not for the $ZnHis_3(H_2O)$ complex (Table S1, Supporting Information). The reason is most likely that, for the latter complex, van der Waals interactions between the water O atom and the nonligating atoms in the imidazole ring prohibit the H and O atoms of water from coming too close to the N atom. However, our experience with other metal complexes involving a water ligand is that you often encounter problems with unstable MD trajectories if you use zeroed Lennard-Jones parameters for the water H atoms.

Table 4. Performance of the Various Force Fields for the $ZnHis_2(H_2O)_2$ Complex[a]

					Hess2FF		NL	
	NB1	NB2	restrained	ZAFF	AT_{min}	AT_{max}	AT_{min}	AT_{max}
r^2 Hessian	0.924	0.914	0.920	crashed	0.413	0.669	0.971	0.997
rmsd bonds	0.375	0.048	0.014		0.060	0.011	0.011	0.002
rmsd angles	4.6	5.6	2.7		10.8	10.6	1.6	0.1
rmsd dihedrals	8.2	7.1	2.2		49.1	40.6	1.6	0.1
rmsd coordinates	0.95	0.52	0.50		1.58	0.94	0.10	0.00
max coordinates	1.51	0.87	0.81		3.48	2.44	0.21	0.01

[a] Quality measures are the same as in Table 1.

Table 5. Performance of the Various Force Fields for the $ZnHis(H_2O)_3$ Complex[a]

					Hess2FF		NL	
	NB1	NB2	restrained	ZAFF	AT_{min}	AT_{max}	AT_{min}	AT_{max}
r^2 Hessian	0.781	0.759	0.750	crashed	0.821	0.967	0.972	0.997
rmsd bonds	0.405	0.080	0.024		0.163	0.015	0.015	0.002
rmsd angles	4.4	5.5	4.5		2.6	2.0	2.5	0.1
rmsd dihedrals	21.2	23.2	24.4		14.8	2.9	3.6	0.1
rmsd coordinates	0.89	0.54	0.31		0.37	0.09	0.11	0.00
max coordinates	1.32	1.33	0.93		0.74	0.19	0.23	0.01

[a] Quality measures are the same as in Table 1.

Molecular Dynamics Simulations. The tests in the previous sections were performed only for small model systems. Such tests have the advantage of having a well-defined reference, viz., the QM calculation. Therefore, small differences between the various force fields can be easily discerned. However, it is also of interest to see how the various Zn force fields behave in simulations of full proteins. Therefore, we have performed two sets of MD simulations with all eight force fields on two Zn-containing proteins: the δ' subunit of the clamp−loader complex of DNA polymerase III (1A5T),[67] which contains a single Cys_4 Zn finger, and the spore coat polysaccharide biosynthesis protein SpsE (1VLI),[68] which contains a $ZnHis_2(H_2O)_2$ site. These two proteins were selected among those used for the model systems because they contain a single metal site and no other unusual ligands. One of them contains a Zn site, for which we do not expect any problems in the simulations, whereas the other one is a hard test case. The proteins were simulated for 5 ns (after equilibration) and the structure of the Zn site was examined for 2500 snapshots sampled every 2 ps. The results are collected in Tables 6 and 7.

For the protein with the $ZnCys_4$ site, all force fields except NB1 gave stable Zn sites throughout the whole simulation. However, for the NB1 force field, the Zn−S bonds were always long (~3.2 Å); after 100 ps the ligands started to dissociate, and after 800 ps Zn had dissociated from all four Cys ligands and drifted around in the solvent. Such a dissociation of the metal ion is always a risk with a nonbonded model, especially if the Zn−ligand interactions are too weak, as in the NB1 force field.

On the other hand, the NB2 force field gave too strong Zn−S bonds, with average and maximum distances of 1.98 Å and 2.17 Å, respectively; these are appreciably shorter than the QM distances of 2.53 Å (Table 6). This shows that a nonbonded force field needs to be thoroughly calibrated for the complex of interest to give accurate results. A nonbonded force field with RESP charges and Merz Lennard-Jones parameters gave average Zn−S bond lengths of 2.43 Å (NB3 in Table 6), closer to the QM reference.

The other six force fields gave rather similar results with average Zn−S bonds of 2.40−2.52 Å (Table 6). It is notable that the NL force fields gave longer Zn−S bonds that are closer to the QM bond lengths than the other four force fields. The reason for this is that the other force fields simply use the QM bond length as the equilibrium value, which owing to interactions with the other atoms in the complex leads to too short bonds. On the other hand, the NL method optimizes the structure with MM in every step of the parametrization, so that if the other interactions in the complex tend to shorten the Zn−S bonds, this is compensated by longer equilibrium Zn−S bond lengths

Table 6. Variation of Zn−S Bond Lengths in Simulation of δ' Subunit of Clamp−Loader Complex of DNA Polymerase III[a]

force field	avg	std dev	min	max
NB1[b]				
NB2	1.98	0.04	1.86	2.17
NB3[c]	2.32	0.13	1.99	3.40
restrained	2.43	0.09	2.11	2.74
ZAFF	2.40	0.09	2.08	2.74
Hess2FF−AT_{min}	2.47	0.09	2.13	2.79
Hess2FF−AT_{max}	2.48	0.09	2.11	2.85
NL−AT_{min}	2.50	0.09	2.15	2.85
NL−AT_{max}	2.52	0.08	2.20	2.85

[a] Average, minimum, and maximum distances are given for the four Zn−S bonds, as well as their standard deviation (std dev). The QM distances were all 2.53 Å. [b] Zn dissociates. [c] A nonbonded force field with RESP charges, but Merz Zn Lennard-Jones parameters.

Table 7. Variation of Zn−Ligand Bond Lengths in Simulation of Spore Coat Polysaccharide Biosynthesis Protein SpsE[a]

	His1			His2			Wat1			Wat2			O−O		
	avg	min	max	avg	min	max	avg	min	max	avg	min	max	avg	min	max
QM	1.95			1.95			2.05			2.05			3.08		
NB1[b]															
NB2[c]	2.11	1.87	2.52	2.20	1.89	2.84	1.90	1.73	2.18	1.92	1.73	2.34			
Simulations with Merz Zn Lennard-Jones Parameters															
restrained	1.98	1.84	2.13	1.97	1.84	2.13	2.03	1.84	2.28	2.03	1.85	2.25	3.85	2.54	4.36
ZAFF[d]															
Hess2FF−AT_{min}	1.92	1.75	2.18	1.94	1.73	2.13	1.73	1.47	1.92	1.72	1.49	1.96	0.83	0.69	1.05
Hess2FF−AT_{max}	1.95	1.74	2.16	1.98	1.76	2.16	2.02	1.9	2.16	2.03	1.89	2.18	0.82	0.66	1.09
NL−AT_{min}	1.92	1.75	2.12	1.94	1.76	2.12	2.02	1.86	2.14	2.02	1.89	2.14	0.97	0.77	1.43
NL−AT_{max}	1.99	1.79	2.22	1.96	1.78	2.17	2.02	1.92	2.13	2.02	1.88	2.14	1.74	1.41	2.12
Simulations with Stote−Karplus Zn Lennard-Jones Parameters															
restrained	2.49	2.40	2.58	2.49	2.42	2.59	2.57	2.43	2.71	2.56	2.45	2.70	4.67	2.93	5.25
ZAFF[d]															
Hess2FF−AT_{min}	1.87	1.84	2.54	1.90	1.84	2.09	1.64	1.84	1.91	1.79	1.85	1.95	0.85	0.71	1.18
Hess2FF−AT_{max}	1.90	1.70	2.08	1.90	1.71	2.10	2.01	1.87	2.13	2.02	1.88	2.14	0.85	0.68	1.10
NL−AT_{min}	1.89	1.76	0.66	1.91	1.90	2.11	2.01	1.89	2.13	2.02	1.77	2.15	1.08	0.79	1.85
NL−AT_{max}	1.94	2.16	0.21	1.91	2.16	2.11	2.02	2.18	2.14	2.00	2.36	2.13	1.74	1.32	2.35

[a] Average, minimum, and maximum distances are given for the four Zn−ligand bonds, as well as the O−O distance between the two water ligands. [b] Zn dissociates. [c] The Wat2 ligand for this simulation dissociated and was replaced by three new water ligands. The results in the Wat2 column are the average, minimum, and maximum values for these three ligands. [d] Simulation crashed.

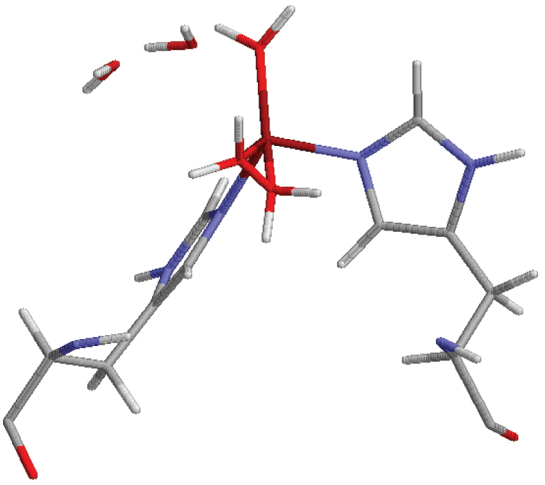

Figure 6. Snapshot of MD simulation of spore coat polysaccharide biosynthesis protein SpsE with NL−AT_{min} force field. The Zn ligands, as well as two second-sphere water molecules, are shown. Note the short O−O interaction between the two original water ligands (0.89 Å).

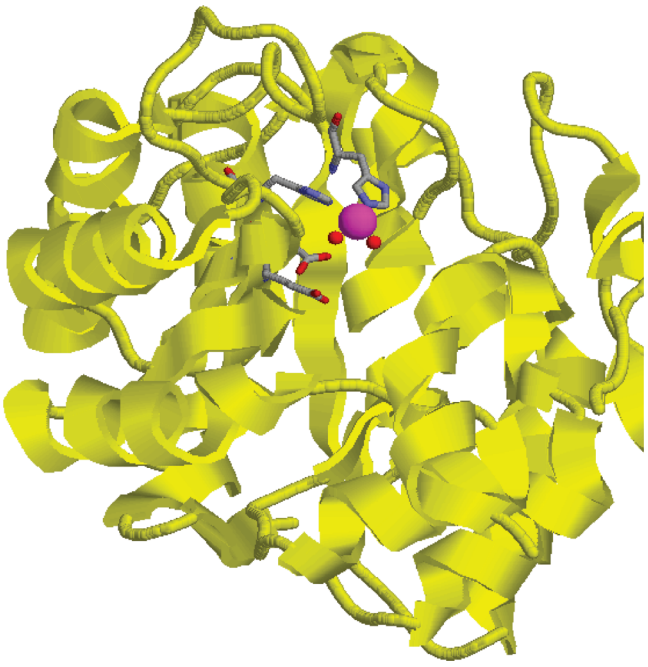

Figure 7. Crystal structure of spore coat polysaccharide biosynthesis protein SpsE with Zn ion (magenta), two water ligands (red balls), two His ligands (sticks), and Glu-22 and Glu-234 (sticks) emphasized.

(e.g., ∼2.56 Å with NL−AT_{min}). Consequently, the average Zn−S bond length is within 0.01 Å of the QM value for the NL−AT_{max} simulation and 0.03 Å too short with NL−AT_{min}, whereas it is 0.05−0.06 Å too short with Hess2FF, 0.10 Å too short for the restrained model, and 0.13 Å too short with ZAFF. This illustrates one of the advantages of the NL approach.

The Zn site in this protein is on the surface, exposed to the solvent. Therefore, it is not unexpected that water molecules occasionally come rather close to the Zn ion. This happen in a few snapshots for all force fields except NB2, but the water molecule never stays close to the Zn ion for more than one or two snapshots.

Finally, we can compare the Zn−S distances observed in the simulations with those reported in the crystal structure, 2.27−2.34 Å, with an average of 2.31 Å. Thus, the experimental distances are 0.22 Å shorter than the theoretical ones. This discrepancy between theory and experiment for a $ZnCys_4$ site has been discussed before;[74] it is caused by missing solvation effects in the vacuum QM calculations for this highly negatively charged complex. Of course, such problems need to be considered in an accurate parametrization of the metal site. In our case, it could be solved by performing the QM optimization and frequency calculations in a continuum solvent with a high dielectric constant. Alternatively, for the restrained nonbonded potential and the bonded potentials based on the Seminario approach, experimental data can easily be included in the force field, as the restraint target or the equilibrium bond lengths or angles. However, for the NL approach this is more problematic, because a target function needs to be defined and it is likely that if the QM bond lengths are changed, the Hessian elements and the angles will also change. Therefore, for NL it is better to use a QM method that gives accurate results.

The results of the simulation of the $ZnHis_2(H_2O)_2$ site in the spore coat polysaccharide biosynthesis protein spsE are described in Table 7. As expected, we encountered severe problems with the stability of the simulations with ZAFF, owing to overlapping H and O atoms from the water ligand (like the structure in Figure 3), and the resulting structures were distorted. However, the other bonded force fields also showed stability problems, but these could be solved by running some (or in a few cases all) of the simulations without any bond-length constraints and a short time step (0.5 fs).

As for the other protein, the Zn ion dissociates from the original ligands with the NB1 force field, in this case already during the equilibration. It diffuses around in water solution and does not bind to the same ligand for more than 400 ps.

However, with the NB2 force field, the Zn ion remains in the original site but one of the water ligands dissociates during the equilibration. It is replaced by three additional water ligands, giving a six-coordinate Zn ion throughout the production simulation. The Zn−N bond lengths are 0.16−0.25 Å longer than in the QM calculations, whereas the Zn−O bonds are 0.13−0.15 Å too short. This shows that it is hard with a nonbonded model to get all metal−ligand bond lengths correct if there are several types of ligands, because only two Lennard-Jones parameters are available (unless you change the Lennard-Jones parameters also of the ligands, which will influence their interactions with the surrounding protein).

All five bonded potentials retained the bonds between Zn ion and the original ligands. However, in all simulations, an additional water molecule also bound to the Zn ion, although with the NL−AT_{min} force field it was frequently exchanged during the simulation. Sometimes, the coordination number increased even to six. All average Zn−ligand bond lengths are within 0.04 Å of the QM bond lengths, except the two Zn−O distances for Hess2FF−AT_{min} (0.33 Å too short). This may give the impression that the simulations are successful, but this is not the case, as the statistics of the O−O distance between the two water ligands show (also included in Table 7). In the QM structure it is 3.08 Å, but in the simulations this distance is never longer than 2.18 Å, and it is 0.8 Å on average in the Hess2FF simulations. This shows that the structures are strongly affected by the problem of the missing 1,3 O−O repulsion, discussed above, and the structures

are actually totally unrealistic, as is shown by the typical snapshot in Figure 6.

In fact, the only simulation that gives realistic results is the restrained nonbonded potential, for which the Zn−ligand distances are within 0.03 Å of the QM distances. However, also in this simulation, an additional water ligand tends to bind to the Zn site, exchanging several times during the 5 ns simulation (one water ligand stays for 2 ns; the others stay for less than 0.6 ns). On the other hand, such an increase in the coordination number of the Zn ion is not totally unrealistic considering the crystal structure. As can be seen in Figure 7, the four observed ligands are all nearly on the same hemisphere of the Zn ion, with a seemingly empty coordination site on the opposite side, with over 6 Å to the closest protein atom. Thus, it is possible that the simulation with the restrained nonbonded potential actually gives a good picture of the metal coordination.

The coordination number can be determined by the force field. The Zn ion is almost hidden by the ligands. Therefore, the Zn Lennard-Jones parameters for a bonded potential can be used to avoid undesired ligand. With the Zn charge used for this site (1.33 e), the Merz Zn Lennard-Jones parameters gives a minimum for the interaction with a water molecule of ∼1.95 Å. On the other hand, the Stote−Karplus parameters give an ∼0.9 Å longer interaction. Therefore, we rerun the bonded and restrained nonbonded simulations also with the latter parameters. For the bonded potentials, the results did not change significantly, except that the extra water ligand disappeared (i.e., the structures were still unrealistic), although there were still occasional water molecules approaching the Zn site, with distances down to 2.62 Å. However, for the restrained nonbonded potential, reasonable four-coordinate structures were obtained. Unfortunately, the Zn Lennard-Jones parameters also affect the distances of the four original ligands so that they became ∼0.5 Å too long. The reason for this is, of course, that the restraints are harmonic (a r^2 term) whereas the repulsive Lennard-Jones term is a much steeper r^{-12} term. Of course, we can tune these interactions by changing the restraint or by using specific Lennard-Jones terms for the Zn−solvent interactions, but we did not pursue such calculations any further.

This shows that the MD simulations will give the results dictated by the parameters used. If we want to know whether the Zn site if four- or five-coordinate in this protein, we could optimize the structure of $ZnHis_2(H_2O)_3$ with QM methods and then tune the Zn Lennard-Jones parameters so that we get the correct Zn−O bond lengths with the Zn charge used in the simulations (note that the minimum distance depends on the Zn charge, and therefore a new calibration has to be done for each new metal site when RESP charges are used). If the simulations then show an increased coordination number, the calculations indicate that that the site should be five-coordinate.

On the basis of the available results, the Merz parameters give Zn−O bond lengths closer to the QM results [for the $ZnHis_2(H_2O)_2$ model] than the Stote−Karplus parameters. Therefore, it is likely that the site actually should be five-coordinate, although additional calculations are needed to settle this issue. The reason why only two of the three water molecules are observed in the crystal structure (at 2.38 Å resolution) is probably that the two observed water molecules are stabilized by two Glu residues (Glu-22 and Glu-224, cf. Figure 7), whereas the third water molecule does not form any strong interactions with the surrounding protein.

CONCLUSIONS

Recently, two softwares have been presented to obtain MM force-field parameters for metal sites.[49,50] These may give the false impression that any metal site may be rapidly and accurately parametrized in an automatic manner. The intention of this paper is to give a more nuanced picture of the matter and to show that the parametrization of metal sites involves several pitfalls.

First, there are several levels of approximation for the parametrization of metal sites. The lowest is a nonbonded model with standard charges for the ligands and formal charges for the metal. It can be used without any parametrization for any metal site but is unlikely to give accurate results for a general metal site, as has been seen in Tables 1−7. In particular, there is a great risk for an unwanted exchange of ligands.[49,77] Therefore, such a model is not recommended.

At the next level, the QM calculations can be used to obtain a full bonded model for the metal site. Such calculations can be made automatic with the approach of Seminario,[47] as has been suggested several times.[48−50] Once the QM calculations are done, the parametrization takes only seconds. However, it must be remembered that this approach is only approximate: It assumes that the bond lengths, angles, and dihedrals observed in the QM structure are optimal, although they actually are a compromise of strain and nonbonded interactions caused by the other ligands, and it involves a double counting of nonbonded interactions, in particular for dihedrals.

These problems are avoided by the ideal iterative method by Norrby and Liljefors.[51] In this approach, a penalty function is set up describing the goal of the force field and the importance of the various terms, for example, the reproduction of QM bond lengths, angles, dihedrals, and Hessian elements. Then the parameters are optimized by full minimization at each step, numerical derivatives, and nonlinear optimization algorithms. This is a much more involved method, taking hours to several days for complicated systems. Moreover, the optimization typically have to be run several times before all parameters are acceptable, and significant human intervention and judging are needed. On the other hand, essentially a perfect fit to the QM data can be obtained (cf. Figures 1d and 5b), which is also reflected in the Zn−ligand distances during the simulations. As a results, reliable energies can be extracted from such simulations of metal complexes.[52,54]

Furthermore, the accuracy of a parametrization is affected by the choice of atom types. For a nonbonded model, it is natural to use standard atom types for all atoms. However, for a bonded model, the choice of atom types is more crucial. A reasonable choice for the Seminario approach is standard atom types for all ligand atoms, except those binding to the metal, for which new atom types are employed, different for all ligands (but using standard parameters within the ligand). For parametrizations with the Norrby−Liljefors approach, a more thorough consideration of atom types is necessary. This method is so accurate that the accuracy of the final model is typically limited by the force field of the ligands. Therefore, it is preferable to use new atom types also for all (not symmetry-equivalent) atoms of the ligands.

This study has also illustrated several possible pitfalls during the parametrization of metal sites. First, the $ZnCys_4$ complex showed that for negatively charged complexes, there is a risk that improper charges are obtained. In particular, it should be checked that ligand hydrogen atoms get a proper (positive) charge. For this model, QM calculations in vacuum also give poor results compared to experiments.

Second, we have showed that water molecules pose a serious problem for bonded models in force fields (like AMBER) that ignore nonbonded interactions for atoms separated by two bonds. The reason for this is that the repulsion between the ligating atoms is ignored. If these bind hydrogen atoms (as for water), these H atoms may form hydrogen bonds to the other ligating atoms with a wrong orientation, as is shown in Figures 5a and 6, leading to distorted structures.

With a single water molecule, it is possible to obtain reasonable structures, once it is ensured that the water H atoms have nonzero Lennard-Jones parameters. However, with several water molecules even the NL method gives strange structures in MD simulations. The only method that works in such a case is the restrained nonbonded model. It combines the best aspects of the nonbonded and bonded models and avoids the problem with the 1,3-interactions by not explicitly define any bonds. Clearly, it is only an approximate method (cf. Figure 5c), but it can be improved by adding more restraints, for example, for angles around the metal, at the end giving a model that is similar to the Seminario approach but avoiding the problem of ignored 1,3-interactions. The restrained nonbonded model is similar to the points-on-a-sphere (POS) model, frequently used for coordination compounds.[17,38,39] Both methods include metal−ligand bonds, exclude ligand−metal−ligand angles, and include ligand−ligand nonbonded interactions. However, the POS model normally retains all the other bonded terms involving the metal, whereas the restrained model excludes them. Moreover, the POS model typically ignores many nonbonded interactions involving the metal, whereas the restrained model includes them all.

In conclusion, we have seen that the parametrization of metal sites is not an automatic approach and, as usual in computational methods, there is a trade-off between accuracy and speed. We would suggest the following general strategy: If the metal site is not of central interest (e.g., a structural metal far from a catalytic or ligand-binding site), a bonded parametrization based on the Seminario method is recommended. If the metal site involves more than one water ligand, a restrained nonbonded model must be used. However, if the metal site is of central interest, either in structural or energy terms, a more thorough parametrization is needed, based on the Norrby−Liljefors approach. Finally, we want to emphasize the need of testing the parametrization before use, that is, how well it reproduces the QM calculations used for the parametrization in terms of structures and Hessian elements.

■ ASSOCIATED CONTENT

Supporting Information. Ten tables showing performance of various force fields, charges used in MD simulations, and coordinates of QM-optimized structures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Ulf.Ryde@teokem.lu.se. Tel: +46-46 2224502. Fax: +46-46 2228648.

■ ACKNOWLEDGMENT

This investigation has been supported by grants from the Swedish Research Council (Project 2010-5025) and the Wenner-Gren Foundation. It has also been supported by computer resources of Lunarc at Lund University.

■ REFERENCES

(1) Boeyens, J. C. A.; Comba, P. *Coord. Chem. Rev.* **2001**, *212*, 3–10.
(2) Kleywegt, G. J.; Jones, T. A. *Methods Enzymol.* **1997**, *227*, 208–230.
(3) Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644–2676.
(4) Gubbins, K. E.; Moore, J. D. *Ind. Eng. Chem. Res.* **2010**, *49*, 3026–3046.
(5) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27–85.
(6) MacKerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604.
(7) Lifson, S.; Warshel, A. *J. Chem. Phys.* **1968**, *49*, 5116–5129.
(8) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127–8134.
(9) Fabricius, J.; Engelsen, S. B.; Rasmussen, K. *J. Carbohydr. Chem.* **1997**, *16*, 751–772.
(10) Pérez, S.; Imberty, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K.; Jimenez-Barbero, J.; Poveda, A.; Espinosa, J.-F.; van Eyck, B. P.; Johnson, G.; French, A. D.; Kouwijzer, M. L. C. E.; Grootenuis, P. D. J.; Bernardi, A.; Raimondi, L.; Senderowitz, H.; Durier, V.; Vergoten, G.; Rasmussen, K. *Carbohydr. Res.* **1998**, *314*, 141–151.
(11) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.
(12) Maple, J. R.; Hwang, M.-J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. *J. Comput. Chem.* **1994**, *15*, 162–182.
(13) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.
(14) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(15) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655.
(16) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671–690.
(17) Hay, B. P. *Coord. Chem. Rev.* **1993**, *126*, 177–236.
(18) Comba, P.; Zimmer, M. *J. Chem. Educ.* **1996**, *73*, 108–110.
(19) Norrby, P.-O.; Brandt, P. *Coord. Chem. Rev.* **2001**, *212*, 79–109.
(20) Hambley, T. W.; Jones, A. R. *Coord. Chem. Rev.* **2001**, *212*, 35–59.
(21) Marques, H. M.; Brown, K. L. *Coord. Chem. Rev.* **2002**, *225*, 123–158.
(22) Zimmer, M. *Coord. Chem. Rev.* **2009**, *253*, 817–826.
(23) Deeth, R. J.; Anastasi, A.; Diedrich, C.; Randell, K. *Coord. Chem. Rev.* **2009**, *253*, 795–816.
(24) Hoops, S. C.; Anderson, K. W.; Merz, K. M. *J. Am. Chem. Soc.* **1991**, *113*, 8262.
(25) Ryde, U. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 40–56.
(26) Gresh, N.; Piquemal, J. P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113.
(27) Nilsson Lill, S. O.; Forbes, A.; Donoghue, P.; Verdolino, V.; Wiest, O.; Rydberg, P.; Norrby, P.-O. *Curr. Org. Chem.* **2010**, *14*, 1629–1645.
(28) Root, D. M.; Landis, C. R.; Cleveland, T. *J. Am. Chem. Soc.* **1993**, *115*, 4201–4209.
(29) Sabolovic, J.; Gomzi, V. *J. Chem. Theory Comput.* **2009**, *5*, 1940–1945.
(30) Stote, R. H.; Karplus, M. *Proteins: Struct., Funct., Genet.* **1995**, *23*, 12–31.
(31) Sakharov, D. V.; Lim, C. *J. Am. Chem. Soc.* **2005**, *127*, 4921.
(32) Åqvist, J.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 7.
(33) Pang, Y. P. *J. Mol. Model.* **1999**, *5*, 196–202.
(34) Pang, Y. P.; Xu, K.; Yazal, J. E.; Prendergas, F. G. *Protein Sci.* **2000**, *9*, 1857.
(35) Singh, N.; Warshel, A. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1705–1723.
(36) Hancock, R. D. *Prog. Inorg. Chem.* **1989**, *37*, 187.
(37) Hancock, R. D. *Acc. Chem. Res.* **1990**, *23*, 253.

(38) Kepert, D. L. *Inorg. Chem.* **1972**, *11*, 1561–1567.
(39) Hambley, T. W.; Hawkins, C. J.; Palmer, J. A.; Snow, M. R. *Aust. J. Chem.* **1981**, *34*, 45–53.
(40) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
(41) Sirovatka, J. M.; Rappe, A. K.; Finke, R. G. *Inorg. Chim. Acta* **2000**, *300*, 545.
(42) Vedani, A.; Huhta, D. W. *J. Am. Chem. Soc.* **1990**, *112*, 4759.
(43) Vedani, A.; Huhta, D. W.; Jacober, S. P. *J. Am. Chem. Soc.* **1989**, *111*, 4075.
(44) Gresh, N. *Curr. Pharm. Des.* **2006**, *12*, 2121.
(45) Gresh, N.; Piquemal, J.-P.; Krauss, M. *J. Comput. Chem.* **2005**, *26*, 1113–1130.
(46) Piquemal, J.-P.; Chevreau, H.; Gresh, N. *J. Chem. Theory Comput.* **2007**, *3*, 824–837.
(47) Seminario, J. M. *Int. J. Quantum Chem.* **1996**, *60*, 1271.
(48) Nilsson, K.; Lecerof, D.; Sigfridsson, E.; Ryde, U. *Acta Crystallogr. D* **2003**, *59*, 274–289.
(49) Lin, F.; Wang, R. *J. Chem. Theory Comput.* **2010**, *6*, 1852–1870.
(50) Peters, M. B.; Yang, Y.; Wang, B.; Fürsti-Molnár, L.; Weaver, M. N.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 2935–2947.
(51) Norrby, P.-O.; Liljefors, T. *J. Comput. Chem.* **1998**, *19*, 1146–1166.
(52) Norrby, P.-O.; Brandt, P. *Coord. Chem. Rev.* **2001**, *212*, 79.
(53) Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.
(54) Rydberg, P.; Olsen, L.; Norrby, P.-O.; Ryde, U. *J. Chem. Theory Comput.* **2007**, *3*, 1765–1773.
(55) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. AMBER 10, University of California, San Francisco, 2008.
(56) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.
(57) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
(58) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(59) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.
(60) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. *J. Am. Chem. Soc.* **1982**, *104*, 2797.
(61) Dobbs, K. D.; Hehre, W. J. *J. Comput. Chem.* **1987**, *8*, 861.
(62) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431–439.
(63) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT, 2009.
(64) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
(65) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
(66) Norrby, P.-O. *J. Mol. Struct. (THEOCHEM)* **2000**, *506*, 9–16.
(67) Guenther, B.; Onrust, R.; Sali, A.; O'Donnell, M.; Kuriyan, J. *Cell* **1997**, *91*, 335–345.
(68) Joint Center for Structural Genomics; to be published; PDB structure 1VLI.
(69) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(70) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(71) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8592.
(72) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
(73) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(74) Ryde, U. *Eur. Biophys. J.* **1996**, *24*, 213–221.
(75) Dunning, T. H.; Hay, P. J. *Modern Theoretical Chemistry*; Schaefer, H. F., Ed.; Plenum Press: New York, 1976; Vol. 3., p 1.
(76) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 270.
(77) Donini, O. A.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 4180–4188.