Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Molecular Graphics and Modelling 24 (2006) 404–411
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Cys_xHis_y–Zn^{2+} interactions: Possibilities and limitations of a simple pairwise force field

Nicolas Calimet, Thomas Simonson *

Laboratoire de Biochimie (UMR 7654 du C.N.R.S.), Department of Biology, Ecole Polytechnique, 91128 Palaiseau, France

Received 30 June 2005; received in revised form 9 October 2005; accepted 9 October 2005
Available online 18 November 2005

Abstract

In zinc proteins, the Zn^{2+} cation frequently binds with a tetrahedral coordination to cysteine and histidine side chains. We examine the possibilities and limitations of a classical, pairwise force field for molecular dynamics of such systems. Hartree Fock and density functional calculations are used to obtain geometries, charge distributions, and association energies of side chain analogues bound to Zn^{2+}. Both ionized and neutral cysteines are considered. Two parameterizations are obtained, then tested and compared through molecular dynamics simulations of two small, homologous proteins in explicit solvent: Protein Kinase C and the Cysteine Rich Domain (CRD) of Raf, which have two Cys_3His–Zn^{2+} groups each. The lack of explicit polarizability and charge transfer in the force field leads to poor accuracy for the association energies, and to parameters—including the zinc charge, that depend on the number of bound cysteines and their protonation state. Nevertheless, the structures sampled with the best parameterization are in good overall agreement with experiment, and have zinc coordination geometries compatible with related structures in the Cambridge Structural Database and the Protein Data Bank. Non-optimized parameters lead to poorer structures. This suggests that while a simple force field is not appropriate for processes involving exchange between water and amino acids in the zinc coordination sphere (e.g. protein unfolding), it can be useful for equilibrium simulations of stable Cys_3His zinc fingers.

Keywords: Force field; Cys_3His–Zn^{2+}; Molecular dynamics

1. Introduction

To help understand the structure and function of zinc proteins, theoretical approaches can be valuable. Many studies using semi-empirical or ab initio quantum calculations have been reported, sometimes in combination with a continuum dielectric or molecular mechanics description of parts of the system [1–10]. Classical mechanical, molecular mechanics approaches are another possibility, which can be used more readily in molecular dynamics simulations [2,11–13]. Since zinc is intermediate between a hard and soft Lewis acid, it can interact with both hard and soft bases. Interactions with hard bases such as water and carboxylates have been successfully modelled with simple, pairwise force fields [12]. Interactions with soft ligands such as thiolates have been studied with the more sophisticated, SIBFA molecular mechanics model, which includes polarizability and charge transfer terms [14–16].

In this paper, we examine the possibility of using a simpler, pairwise, classical mechanical energy function to study the interactions of Zn^{2+} with combinations of cysteine and histidine residues in proteins. Such an energy function could be a useful tool to study the molecular dynamics (MD) of zinc proteins in solution. Long simulations including large quantities of explicit solvent are necessary to generate statistical ensembles for free energy calculations [17], or to study slow dynamical modes that can have functional relevance. Several zinc finger proteins have been studied by MD. Ryde proposed a force field for alcohol dehydrogenase, where zinc is coordinated by a histidine, two cysteinates and a carboxylate [2,11]. This force field was based on high-level quantum chemical calculations. The zinc–ligand interactions were modelled with harmonic, bonded terms, which prevent the ligands from leaving the coordination sphere. In several other studies, the energy function was derived either heuristically, by simply imposing harmonic geometrical restraints around the zinc, or from low level quantum mechanical calculations [18–20].

In a very recent study, Sakharov and Lim derived a force field for zinc–protein complexes including polarizability and

* Corresponding author. Tel.: +33 1 69333881; fax: +33 1 69333013.
E-mail address: thomas.simonson@polytechnique.fr (T. Simonson).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.10.006

charge transfer [13]. This led to good zinc coordination geometries, and probably represents the minimum level of theory for MD studies of zinc protein folding or zinc binding/unbinding. Nevertheless, a simpler approach may be sufficient for problems involving a stable, equilibrium protein structure, where the zinc coordination is maintained.

To determine the feasibility of a simple, pairwise, molecular mechanics energy function, we compute geometries and association energies of Zn^{2+}–Cys_xHis_ygroups with an ab initio supermolecule approach, using both Hartree Fock and density functional methods. We also analyze related experimental geometries found in the Cambridge Structural Database of small molecules and the Protein Data Bank. Based on this data, we parameterize a simple molecular mechanics force field, and examine its behavior in MD simulations of two small proteins in aqueous solution: the Cys_2 activator-binding domain of Protein Kinase C δ (PKC) and the Cysteine Rich Domain of Raf (CRD), which both have two Cys_3His zinc coordination spheres. The lack of explicit polarizability and charge transfer in the force field leads to poor accuracy for the association energies. Nevertheless, the structures sampled with the best parameterization are in good overall agreement with experiment, while non-optimized parameters lead to poorer structures.

As another application, we consider the problem of thiolate protonation in Zn^{2+}–Cys_xHis_y groups in proteins. A common theoretical approach to estimate proton binding constants is to combine molecular mechanics partial charge sets with a continuum electrostatics model of the protein and solvent [21–25]. Although a quantum mechanical model would be necessary to obtain very accurate results [26], a classical mechanical approach should yield a fair qualitative estimate of the protonation state populations. We use the charge sets obtained here, along with finite-difference Poisson–Boltzmann calculations, to determine approximately the dominant protonation state of the cysteine side chains coordinating Zn^{2+} in three proteins: PKC, the HIV nucleocapsid protein p7 (NCp7), and the DNA-binding domain of the nuclear receptor protein NGFI. For the latter protein, we consider both the protein alone and the protein:DNA complex. The calculations predict that binding of the protein to DNA induces proton binding to one of the cysteinates. We also apply the method to a small zinc finger peptide, which corresponds to a 14 aminoacid fragment of NCp7, denoted NCp7-14. For NCp7 and the latter peptide, the protonation states have been measured by mass spectrometry [27,28].

2. Methods

2.1. Ab initio calculations

A first set of ab initio calculations were performed for a Zn^{2+} cation binding a single ligand: methanethiol or methylthiolate. Calculations were done with the Gaussian98 program [29]. Association energies were calculated at the restricted Hartree Fock level with several basis sets. A second set of calculations involved Zn^{2+} with four ligands: $n = 0$–3 ethanethiols, $3 - n$ ethylthiolates and one imidazole. These complexes were fully optimized with several basis sets (Table 1). Association energies were calculated at the RHF level with basis sets up to 6-311 + G(2d,2p), and with density functional theory at the B3LYP/6-311 + G(2d,2p) level, i.e. using Becke's three-parameter hybrid method [30] along with the Lee, Yang and Parr correlation functional [31]. The 6-311 + G(2d,2p) basis set adds diffuse functions to heavy atoms, and polarization functions to both heavy atoms and hydrogen atoms. Basis set superposition error was neglected, since it has been shown to be small for related Zn^{2+}–ligand complexes at this basis set level [3,16,32].

2.2. Force field parameter fitting

Force field parameters were derived in the case of a Cys_3HisZn^{2+} coordination sphere, with either three cysteines in thiolate form, or two in thiolate and one in thiol form. However, only the all-thiolate parameters were tested by MD simulations of PKC and the CRD (see below). In either case, atomic partial charges were taken directly from the ab initio calculations on small tetrahedral complexes of Zn^{2+} with imidazole and ethanethiol or thiolate. Mulliken charges were used; this choice is reasonably consistent with the development of the Charmm22 force field. [33]. Either the 6-311G** or the

Table 1
Ab initio geometries and energies for tetrahedral zinc–ligand complexes

Ligands[a]	Cys_3^{-}[b]	Cys Cys_2^{-}[c]	Cys_3^{-}[c]	Cys_3^{-}/MM[d]
Zn–ligand distances (Å)				
Zn–N	2.23	2.13	2.29	2.29
Zn–S_1	2.39	**2.81**[e]	2.35	2.25
Zn–S_2	2.43	2.29	2.38	2.29
Zn–S_3	2.39	2.26	2.35	2.25
S–Zn–S angles (°)				
S_1–Zn–S_2	117	97	118	119
S_2–Zn–S_3	116	136	116	119
S_1–Zn–S_3	115	100	116	113
N–Zn–S angles (°)				
N–Zn–S_1	107	97	105	104
N–Zn–S_2	95	107	95	95
N–Zn–S_3	103	112	103	101
				Difference[f]
Dissociation energies (kcal/mol)[g]				
HF/6-311G**	-	600	631	30
HF/6-311 + G(2d,2p)	-	620	652	32
B3LYP/6-311 + G(2d,2p)	-	662	694	32
Molecular mechanics[d]	-	389	430	41

[a] The fourth ligand is imidazole. Cys^{-} represents ethylthiolate; Cys represents ethanethiol.
[b] Optimized at the RHF/6-311G** level.
[c] Optimized at the B3LYP/6-311 + G(2d,2p) level.
[d] Molecular mechanics calculations using the force field parameterized with the 6-311G** charges (''parameter set 2''; see text, Table 2).
[e] Distance to protonated, ethanethiol sulfur in bold face.
[f] Difference between the previous two columns.
[g] At the specified basis set level. Not including zero point energy or thermal corrections; basis set superposition error neglected. Structures optimized at the RHF/6-311G** level.

6-311 + G(2d,2p) charges were used, leading to two distinct force fields. The van der Waals parameters ε and σ for the interactions between Zn^{2+} and the coordinating sulfur and nitrogen were obtained as follows. The structure of the small tetrahedral complex, optimized at the RHF/6-311G** level, was optimized further with the classical mechanical force field, using a range of ε and σ parameters. Values that gave structural deviations of more than a chosen threshold from the ab initio geometry were discarded. From the remaining values, the ones that gave the association energy closest to the ab initio value were chosen. The threshold was based on the rms deviation of the zinc and its four ligating atoms, and was 0.10 Å for the all-thiolate case and 0.25 Å in the case of two thiolates and one thiol. All other van der Waals parameters were taken from the Charmm22 force field [33]. No bonded terms involving the zinc were included in the force field.

2.3. Molecular dynamics protocol

MD simulations were performed for two small, homologous proteins in explicit solvent. The initial PKC structure was the X-ray structure (PDB entry 1PTR [34]). A phorbol ester ligand was removed. The initial CRD structure was model 8 out of the 27 models in the NMR ensemble (PDB entry 1FAR [35]); this model was chosen because it is the closest to the structure obtained by averaging over the 27 models. The protein was placed in a cubic box of water with 60 Å edges, and overlapping waters were deleted. One and two chloride ions were added to PKC and the CRD, respectively, to make each system electrically neutral. They were placed at random positions, 15 Å from the protein. The simulation cells contained $\sim$20,000 atoms in all. Periodic boundary conditions were applied. The Charmm22 force field was used for the protein [33], except for the protein–zinc interactions parameterized in this work. Waters were described by the TIP3P model [36]. Long-range electrostatic interactions were treated with Particle Mesh Ewald summation and tin foil boundary conditions [37]. The system was first minimized for 200 steps with an adopted basis set method, with the protein and zinc fixed. Molecular dynamics were performed for 5 ps with the protein fixed, then for 75 ps with gradually decreasing harmonic restraints on the protein. Temperature and pressure were maintained at 300 K and 1 atm through a Nose–Hoover thermostat and barostat [38]. The lengths of covalent bonds involving hydrogens were held fixed [39]. The simulation was continued without restraints for 1000–1500 ps. Calculations were done with the Charmm program [40].

PKC and the CRD each have two Zn^{2+}–Cys_3His groups. In all the simulations, the cysteines coordinating the zinc were assumed deprotonated (all-thiolate model). Although there is no direct evidence for this model, the zinc coordination in PKC is approximately symmetric, with all the Zn–S distances between 2.27 and 2.32 Å. This is in contrast to small tetrahedral ab initio structures (below), which become very asymmetrical as soon as one of the thiolates is protonated. The CRD NMR structure does not give precise Zn^{2+}–S distances, but the same, all-thiolate model was adopted on the basis of the structural homology between the CRD and PKC.

2.4. Poisson calculations

The continuum electrostatics approach used here has been described by many authors [21,41,22,23,25]. Briefly, the protein is treated as a low dielectric cavity, with embedded point charges located on all atoms (the molecular mechanics partial charges are used here). A finite-difference method is used to solve the Poisson equation numerically; from the resulting electrostatic potential, the free energy is obtained. Calculations are performed for all the possible protonation states of the cysteine side chains; all other groups are assumed to be in their most common protonation state (e.g. aspartates are negatively charged, lysines are positive). The same calculation is done for an isolated cysteine side chain in solution, providing a reference protonation free energy. Calculations were done with the UHBD program [41]. For NCp7 [42] and the small peptide NCp7-14 [43], the experimental NMR structures were used; calculations were performed for five out of the ensemble of NMR structures. For PKC, the X-ray structure was used [34]. For NGFI, we employed the X-ray structure of the protein–DNA complex [44]. For the protein alone, we simply removed the DNA without modifying the protein structure. For PKC and NCp7, there are two Cys_3His groups, giving a total of 64 possible protonation states. For NCp7-14, there is a single Cys_3His group, giving eight possible protonation states. For NGFI, there are two Cys_4groups and 256 possible protonation states. For each protein, two sets of calculations were done, using protein dielectric constants of 4 and 20, respectively. The former value is usually thought to be more physically realistic [25]; however, the latter is also used by many workers for protein pK_a calculations [45,23].

3. Results

3.1. Ab initio calculations

3.1.1. Small binary complexes

The association energies of methanethiol and thiolate with Zn^{2+} were calculated at the HF/6-311G**//HF/6-311G** level, yielding −100.3 and −372.2 kcal/mol, respectively. The association energy difference is thus 269.9 kcal/mol. The zinc sulfur distances are 2.327 and 2.233 Å, respectively. Thus, protonation produces a 0.1 Å increase in the Zn^{2+}–S distance. Gresh obtained association energies of −112 and −388 kcal/mol with a somewhat larger (“homemade”) basis set [15], giving an association energy difference of 276 kcal/mol. Thus, basis set dependency is still significant at this level; however, it is smaller for the thiol–thiolate difference. El Yazal and Pang calculated the difference in proton dissociation energies for methanethiol with and without a zinc ligand, including thermal and zero point energy contributions [5]. This difference is identical to the association energy difference. They obtained 286 kcal/mol with density functional theory at the B3LYP/6-311 + G(2d,2p) level.

3.1.2. Small tetrahedral complexes

We next consider Zn^{2+} association with one imidazole, 0–1 ethylthiolates and 1–0 ethanethiols. Association energies and geometries are reported in Table 1. In the all-thiolate case, the

three sulfur–zinc distances are equivalent, at 2.36 ± 0.02 Å. With a single thiol, the thiol sulfur–zinc distance increases to 2.81 Å, while the other three ligands form a nearly equilateral triangle at distances of 2.1–2.3 Å from the zinc. A complex with four ethylthiolate ligands was also optimized, yielding Zn^{2+}–S distances of 2.46–2.51 Å; Gresh and co-workers obtained distances of 2.44–2.50 Å with the same basis set [16].

The association energies decrease rapidly as the number of thiolates increases from zero to one and two (by −193 and −118 kcal/mol), then slightly (−30 kcal/mol) for the third thiolate (data not shown). Replacing the imidazole by a fourth thiolate is expected to increase the energy (disfavor association) due to electrostatic repulsion between the thiolates [16]. Including electronic correlation at the B3LYP/6-311 + G(2d,2p) level decreases the energy by another 42 kcal/mol in both the all-thiolate and single-thiol cases, leaving the association energy difference unchanged (Table 1).

3.2. Molecular mechanics calculations

3.2.1. Parameter optimization, association energies

Force field parameters were obtained for Zn^{2+} complexed with imidazole and three thiolates, and for Zn^{2+} complexed with imidazole, two thiolates and one thiol. These ''small tetrahedral complexes'' were optimized at the HF/6-311G** level (see above). Mulliken charges were obtained at the 6-311G** and 6-311 + G(2d,2p) levels for the former, all-thiolate case. For the latter, mixed thiolate/thiol case, Mulliken charges were obtained only at the 6-311G** level. Van der Waals parameters were then derived as described in Section 2, by combinatorially scanning a range of ε and σ values for the nitrogen– and sulfur–zinc van der Waals interactions. For each set of parameters, the structure was optimized (with molecular mechanics), and van der Waals parameters that gave a deviation from the ab initio structure above a chosen threshold were discarded. From the remaining possible parameters, those that gave the best association energy were selected.

In addition to the parameters obtained by the above procedure, ''standard'' Charmm22 force field parameters were also tried. These parameters were optimized for zinc interacting with hard bases, such as water or carboxylates [12]. The various force field parameters are reported in Table 2. Only the all-thiolate parameters were tested by MD simulations (following section).

The resulting optimized geometries of the ''small tetrahedral complexes'' are reported in Table 1, along with the corresponding association energies. The association energies are in very poor agreement with the ab initio results (Table 1). This results directly from the neglect of polarizability and charge transfer terms in the energy function [14–16,46]. Indeed, charge transfer from the sulfur to the zinc is taken into account in the atomic partial charges, leading to reduced zinc and sulfur charges, but it does not provide any energy gain. On the contrary, the reduced partial charges result in a weakened Coulomb interaction, which cannot account correctly for the strength of the partly-covalent quantum mechanical interaction.

Table 2
Zinc–Cys_3His force field parameters

Group	Atom	Set 1 (Charmm)	Set 2 (6-311G**)	Set 3 (6-311 + G(2d,2p))
Atomic partial charges				
Zinc	ZN	2.00	1.52	1.05
Cys	SG	-0.80	-0.77 [-0.21][a]	-0.75
Cys	HG	-	[0.12]	-
Cys	CB	-0.38	-0.31 [-0.30]	0.20
Cys	HB1/HB2	0.09	0.12 [0.18]	0.02
Cys	CA	0.07	0.07 [0.13]	0.07
Cys	HA	0.09	0.09 [0.22]	0.09
His	NE2	-0.36	-0.38	-0.15
His	HE2	0.32	0.27	0.18
His	CD2	-0.05	0.05	-0.63
His	HD2	0.09	0.13	0.09
His	ND1	-0.70	-0.80	-0.53
His	CG	0.22	0.12	0.62
His	CE1	0.25	0.36	0.38
His	HE1	0.13	0.19	0.16
His	CB	-0.08	-0.21	0.16
His	HB1/HB2	0.09	0.12/0.15	0.09

Atom pair		Set 1 (Charmm)	Set 2 (6-311G**)	Set 3 (6-311 + G(2d,2p))
van der Waals parameters				
ZN–ND1	ε	2.459680	1.565248	1.118034
ZN–ND1	σ	2.30	2.60	2.35
ZN–SG	ε	4.113393	0.685565 [0.335410][a]	2.399479
ZN–SG	σ	2.70	2.80 [3.20]	2.65

Atom names are defined in Fig. 1.
[a] Parameters for the neutral, thiol form of cysteine in brackets.

Despite this deficiency, the geometry of the zinc coordination sphere is accurately reproduced. Furthermore, with the 6-311G** charges, the cohesion of the complex is, in fact, sufficient to prevent intrusion of water into the first coordination shell during MD simulations with explicit solvent (see below).

Table 2 also reports parameters for zinc interacting with histidine, two cysteinates and one cysteine (Fig. 1). These parameters were derived in the same manner, but they have not yet been tested in MD simulations of proteins. The association energy difference between this complex and the all-thiolate complex (41 kcal/mol) is in fair agreement with the ab initio difference (32 kcal/mol).

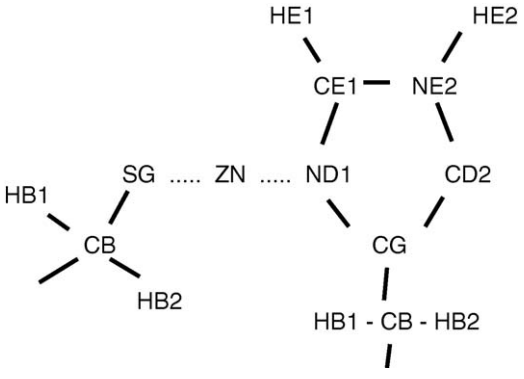

Fig. 1. Atom names used in Table 2.

3.2.2. Molecular dynamics simulations

MD simulations were performed for PKC and the CRD using three different force field parameterizations, corresponding to three different sets of atomic partial charges, along with the corresponding van der Waals parameters. The force field parameters are reported in Table 2. In all cases, an "all-thiolate" model was employed; i.e. the zinc ions are coordinated by three cysteinates (and one His). Some uncertainty exists as to the exact Cys protonation state in the 1PTQ structure of PKC (see above); however in the 1PTR structure used here, the zinc coordination sphere is approximately symmetrical, and has short Zn–S distances. Therefore, an all-thiolate model should be appropriate.

The overall deviations from the experimental structure are reported as a function of time in Fig. 2 for each protein and simulation. The geometries of the zinc coordination spheres are reported in Table 3. For PKC, results with the two force fields derived here are good. Thus, with Mulliken atomic charges at both the HF/6-311G** and HF/6-311 + G(2d,2p) levels, the overall deviations are small: 1.0 Å for PKC with both parameter sets. The geometry of the zinc coordination sphere is also in good agreement with the ab initio results and the Cambridge Structural Database survey, with parameter set 2 giving slightly better results. With both force fields, an approximately tetrahedral zinc coordination is preserved at all times, with the same four aminoacid ligands. Only during one brief, 30 ps segment of the PKC simulation with the 6-311 + G(2d,2p) charges (parameter set 3) does a water enter the zinc coordination sphere. The good agreement with the experimental structures provides indirect evidence in support of the all-thiolate model employed here. The tetrahedral zinc coordination is also consistent with a theoretical study that indicates that in a moderate or low dielectric environment, a tetrahedral zinc coordination is more stable than five- or six-coordinated arrangements [3]. For the CRD, results with parameter set 2 are good, with an rms deviation of 2.0 Å and a stable, tetrahedral, zinc coordination. With parameter set 3, the histidine coordinating one of the two zincs leaves the coordination sphere after about 200 ps of dynamics, being replaced by a water molecule. The overall rms deviation increases rapidly to over 2.5 Å.

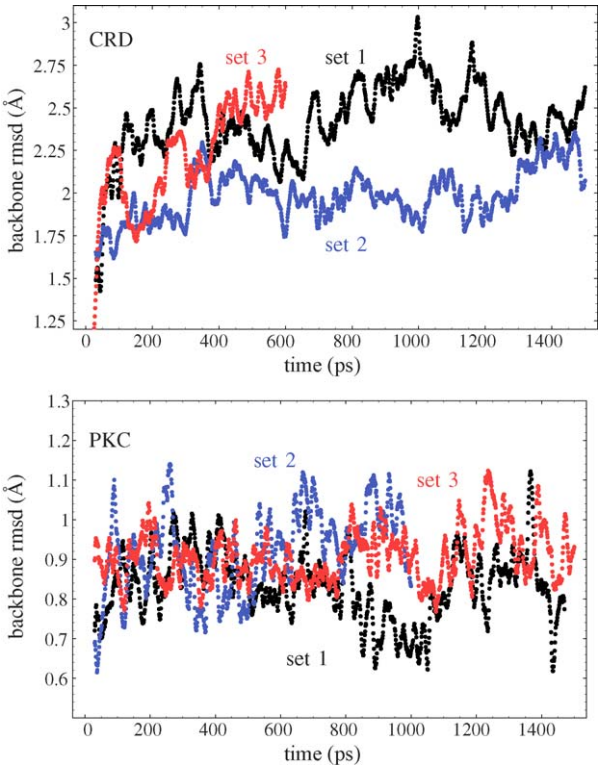

Fig. 2. Rms deviations (Å) from the experimental structure for backbone atoms vs. time (running average over 10 ps segments). Upper panel, CRD; lower panel, PKC. Three simulations were done for each protein. Corresponding force field parameters in Table 2.

Table 3
Zinc–ligand geometries sampled during molecular dynamics simulations

Parameter set	Zn–S distances (Å)	Zn–N distances (Å)	S–Zn–S angles (°)
CRD			
(1) Charmm	**2.49–2.51//2.42 –2.51** (0.06)	2.08//2.07 (0.06)	93–99//93–110 (4)
(2) 6-311G**	2.26–2.28//2.26–2.29 (0.06)	2.25//2.25 (0.06)	110–118//107–117 (4)
(3) 6-311 + G(2d,2p)	2.37–2.39//2.39–2.41 (0.07)	2.21//**5.11** (**0.97**)	113–116//106–117 (7)
PKC			
(1) Charmm	2.36–2.38//**2.41 –2.47** (0.05)	1.97//1.99 (0.04)	108–116//97–121 (5)
(2) 6-311G**	2.26–2.27//2.25–2.26 (0.05)	2.23//2.24 (0.05)	110–113//110–116 (5)
(3) 6-311 + G(2d,2p)	2.38–2.40//2.38–2.43 (0.07)	2.16//2.17 (0.11)	105–114//103–122 (6)
X-ray (1PTR)[a]	2.27–2.32	2.03//2.04	107–114
X-ray (1PTQ)[a]	2.33, 2.45, 2.51//2.35–2.43	2.33//2.38	102–115
Ab initio[b]	2.39–2.43	2.23	115–117
PDB survey[c]	2.335 ± 0.098	2.191 ± 0.190	110.8 ± 6.0

Each protein has two Cys_3His zinc coordination spheres. The Cys ligands were all treated as cysteinates in the simulations. Results for the two zinc coordination spheres are separated by '//'. Standard deviations are in parentheses, averaged over all ligands and both coordination spheres. The range of values listed for the S–Zn–S angles in each coordination sphere correspond to S_1–Zn–S_2, S_2–Zn–$S_3$and S_1–Zn–S_3. Values in bold face indicate cases where the zinc coordinates one or two water molecules, in addition to (or instead of) the usual four protein sidechains.

[a] 1PTR is the structure used to initiate the MD simulations. It contains a phorbol ester ligand, which was removed. 1PTQ is a structure of the apo-protein.

[b] From Table 1.

[c] From Ref. [10].

When non-optimized force field parameters are used (set 1), results are also poor for the CRD. The overall rms deviation of the CRD increases rapidly to 2.5 Å and its two zinc coordination spheres shift rapidly to accomodate five and six ligands, including one and two water molecules, respectively. For PKC, although the overall rms deviation of the backbone is small with the non-optimized parameters, one of the two zincs is coordinated by five ligands, including a water and the usual four protein sidechains. This disagrees with the crystal structure. The ''non-optimized'' parameters were taken directly from the Charmm22 force field; they were designed to model Zn^{2+} interacting with hard bases such as water or carboxylate groups, [12] and not for modelling Zn^{2+}–sulfur interactions. Simulations of this length and size are too expensive to allow many runs and a truly statistical study. Therefore, it remains to be seen whether repeated simulations with these proteins and force fields would consistently give the same good and bad results. Nevertheless, it is encouraging that parameter set 2 gives very good results for both proteins over a total of 3 ns of dynamics.

3.3. Application: proton binding to proteins

As another application of the force field, we compute the probabilities of all possible protonation states of the cysteines coordinating the zinc, using a dielectric continuum model. The atomic charges and radii are taken from the above force fields (parameter sets 1 and 2). Results are reported in Table 4. We list the mean populations of the dominant protonation states, which are related to the protonation free energies by a standard Boltzmann relation [22]. In NCp7 and NCp7-14, when a low protein dielectric constant of four is used, the most probable state is the one where all cysteines are ionized (all-thiolate state). Other states are negligibly populated ($\leq$1.5%). However, with a larger protein dielectric of 20, the populations shift, and the most probable state is the one with a single thiol. Notice that the dielectric constant of NCp7-14 computed from MD simulations is about 11 [20,47]. Summing over all the possible locations of the single thiol (there are three possible positions in each Cys_3His group), the aggregate weight of the single-thiol states is 54% in NCp7 and 57% in NCp7-14. States with two thiols have weights of 6 and 37%, respectively (with the protein dielectric of 20). From mass spectrometry, the dominant states are expected to have one thiol per zinc.

In PKC, with a low protein dielectric of four, the dominant state is the single-thiol state (one thiol and five thiolates for the two zinc fingers), with an aggregate population of 61%, compared to 12% for the all-thiolate state. Note that in the MD simulations above, we considered only the all-thiolate state. The two-thiol state has a population of 27%; other states are negligible. In the two-thiol state, there is predominantly one thiol in each Cys_3His group (i.e. states with two thiols in the same group have negligible populations; not shown), so that each group is electrically neutral. With a larger protein dielectric of 20, the populations shift towards states with three thiols (40%) and four thiols (40%), with $\sim$10% occupancy each for two- and five-thiol states. Experimentally, we expect that the all-thiolate state is predominant, based on the symmetrical zinc coordination geometry seen in the crystal structure. A lower protein dielectric (1–2) would presumably be needed to reproduce this result.

For NGFI, we used the original, Charmm22 parameters. This is because the parameters developed here (sets 2 and 3) are optimized for Cys_3His groups, whereas NGFI has two Cys_4 groups. Since the zinc charge is non-integer with sets 2 and 3, the Cys charges are not directly transferable to this case. With the Charmm22 charges, and with a dielectric of four for the protein (and the DNA), the populations are: all-thiolate, 3%; one thiol, 45%; two thiols, 51%. The mean number of bound protons is 1.47. For the protein alone, the populations shift towards more highly charged states: all-thiolate, 61%; one thiol, 39%; two thiols, 0%. The mean number of bound protons is 0.39. Removing the DNA from the protein raises the dielectric constant locally, and this stabilizes the thiolate states. Overall, when the protein binds to the DNA, the zinc finger groups are predicted to bind an additional 1.08 protons on

Table 4
Probabilities of Cys protonation states in proteins from continuum model

Protein	Protein dielectric	Number of protons bound						
		0	1	2	3	4	5	6
PKC	4	0.12	**0.61**	0.27	0.00	0.00	0.00	0.00
PKC	20	0.00	0.01	0.10	**0.40**	**0.40**	0.10	0.00
NCp7	4	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00
NCp7	20	0.40	**0.54**	0.06	0.00	0.00	0.00	0.00
NCp7-14	4	**0.99**	0.01	0.00	0.00			
NCp7-14	20	0.06	**0.57**	0.37	0.00			
NGFI	4	**0.61**	0.39	0.00	0.00	0.00	0.00	0.00
NGFI:DNA	4	0.03	0.45	**0.51**	0.00	0.00	0.00	0.00
NGFI	20	**0.43**	0.30	0.27	0.00	0.00	0.00	0.00
NGFI:DNA	20	0.02	0.23	**0.51**	0.21	0.02	0.00	0.00

Values corresponding to the most probable state of each system are in bold face. Calculations were done with the ab initio 6-311G** charges derived above, except for the NGFI calculations, which employed Charmm22 charges; this is because the 6-311G** charges were derived for Cys_3His groups, whereas NGFI contains two Cys_2groups.

average (1.47–0.39), predominantly on the zinc finger closest to the DNA. With a protein/DNA dielectric of 20, the same trends are observed. The number of additional protons taken up when NGFI binds to DNA is 1.88–0.84 = 1.04, almost exactly the same as with $\varepsilon = 4$. This robustness suggests that the proton binding is also robust with respect to the exact charge set used. The prediction of a coupling between proton-binding and DNA binding could be experimentally tested through the pH-dependency of the protein:DNA binding constant.

4. Concluding discussion

Zinc proteins play an important and widespread role in structural biology. To study their structure, dynamics and thermodynamics, it is necessary to develop reliable computational tools. In particular, there is a need for efficient force fields, suitable for long molecular dynamics simulations. Here, we have focused on one major class of Zn^{2+} binding sites, in which the cation binds with a pseudo-tetrahedral coordination to cysteine and histidine side chains. For these systems, we have examined some of the possibilities and limitations of a simple, pairwise, classical mechanical force field. In addition, we have considered the related question of cysteine protonation states in different protein environments.

We first analyzed the interactions of zinc with small aminoacid analogues through ab initio calculations. Both ionized and neutral cysteine analogues were considered. The Cys_3His complexes are approximately symmetrical when all the cysteines are ionized; when one is protonated, they become markedly asymmetrical, with short zinc–thiolate and a long zinc–thiol distances.

We next examined the possibilities of a classical, pairwise force field, suitable for long MD simulations. Several parameterizations were obtained. No bonded terms involving the zinc were included; only van der Waals and Coulomb terms. Atomic partial charges were taken from a Mulliken population analysis of the small tetrahedral complexes using two different basis sets; the Zn–S and Zn–N van der Waals parameters were then calibrated to reproduce (primarily) the ab initio geometry. These sets, as well as the ''standard'' Charmm parameters, were tested and compared through 0.6–1.5 ns MD simulations of Protein Kinase C and the Cysteine Rich Domain of Raf. The zinc was assumed to be coordinated by three cysteinates and one histidine, consistent with the PDB analysis discussed above. Force field parameters were also derived for the case of one thiol and two thiolates, but they have not yet been tested in MD simulations.

The lack of explicit polarizability and charge transfer in the force field leads to poor accuracy for the association energies, and to parameters—including the zinc charge, that depends on the number of bound cysteines and their protonation state (thiol or thiolate). Nevertheless, the structures sampled with one of the two parameterizations derived here are in good overall agreement with experiment. In contrast, non-optimized parameters lead to poorer structures for the CRD, where the tetrahedral zinc coordination is lost, and the overall protein fold is not well-maintained. This shows that even though the simple force field gives a crude description of the zinc–cysteine association energetics, it should be a useful tool for equilibrium simulations of stable zinc fingers. To obtain a highly accurate force field, it would be necessary to take into account the polarizability of the zinc and sulfur [13], and possibly to include a zinc–sulfur charge transfer term in the energy function.

Another application of the parameters is the calculation of protonation states and other electrostatic properties with a continuum electrostatic model. Although a quantum mechanical model would be necessary for quantitative accuracy, the results above suggest that reasonable qualitative results can be obtained. The predictions for NCp7 and NCp7-14 agree with mass spectrometry data; for NGFI, a coupling between DNA binding and proton binding is predicted.

Acknowledgments

We thank Martin Karplus for the Charmm program. Some of the simulations were carried out at the supercomputer center C.I.N.E.S. of the French Ministry of Education (supercomputer allocation to T.S.). N.C. was the recipient of a Ph.D. fellowship from the French Ministry of Research.

References

[1] U. Ryde, The coordination chemistry of the catalytic zinc ion in alcohol dehydrogenase studied by ab initio quantum chemical calculations, Int. J. Quantum Chem. 52 (1994) 1229–1243.
[2] U. Ryde, On the role of Glu68 in alcohol dehydrogenase, Prot. Sci. 4 (1995) 1124–1132.
[3] T. Dudev, C. Lim, Tetrahedral vs. octahedral zinc complexes with ligands of biological interest: a DFT/CDM study, J. Am. Chem. Soc. 122 (2000) 11146–11153.
[4] T. Dudev, C. Lim, Modeling Zn^{2+}–cysteinate complexes in proteins, J. Phys. Chem. B 105 (2001) 10709–10714.
[5] J. El Yazal, Y. Pang, Ab initio calculations of proton dissociation energies of zinc ligands: hypothesis of imidazolate as zinc ligand in proteins, J. Phys. Chem. B 103 (1999) 8773–8779.
[6] A. Maynard, M. Huang, W. Rice, D. Covell, Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density functional theory, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 11578–11583.
[7] J. Aqvist, A. Warshel, Computer simulation of the initial proton transfer step in human carbonic anhydrase I, J. Mol. Biol. 224 (1992) 4–14.
[8] J. Chen, L. Noodleman, D. Case, D. Bashford, Incorporating solvation effects into density functional electronic structure calculations, J. Phys. Chem. 98 (1994) 11059–11068.
[9] J. Jackman, K. Merz, C. Fierke, Disruption of the active site solvent network in carbonic anhydrase II decreases the efficiency of proton transfer, Biochemistry 35 (1996) 16421–16428.
[10] T. Simonson, N. Calimet, Cys_xHis_y–zn^{2+} interactions: thiol vs. thiolate coordination, Proteins 49 (2002) 37–48.
[11] U. Ryde, Molecular dynamics simulations of alcohol dehydrogenase with a four- or five-coordinate catalytic zinc ion, Proteins 21 (1995) 40–56.
[12] R. Stote, M. Karplus, Zinc binding in proteins and in solution: a simple but accurate non-bonded representation, Proteins 23 (1995) 12–31.
[13] D.V. Sakharov, C. Lim, Zn protein simulations including charge transfer and local polarization effects, J. Am. Chem. Soc. 127 (2005) 4921–4929.
[14] N. Gresh, W. Stevens, M. Krauss, Mono- and polyligated complexes of Zn^{2+}: an ab initio analysis of the metal–ligand interaction energy, J. Comp. Chem. 16 (1995) 843–855.

[15] N. Gresh, Energetics of Zn^{2+} binding to a series of biologically relevant ligands: a molecular mechanics investigation grounded on ab initio self-consistent field supermolecular computations, J. Comp. Chem. 16 (1995) 856–882.
[16] G. Tiraboschi, N. Gresh, C. Giessner-Prettre, L. Pedersen, D. Deerfield, Parallel ab initio and molecular mechanics investigation of polycoordinated Zn(II) complexes with model hard and soft ligands: variation of binding energy and its components with number and charges of ligands, J. Comp. Chem. 21 (2000) 1011–1039.
[17] T. Simonson, Free energy calculations, in: O. Becker, A. Mackerell, B. Roux, M. Watanabe, Jr. (Eds.), Computational Biochemistry & Biophysics, Marcel Dekker, N.Y., 2001 (chapter 9).
[18] M. Eriksson, T. Härd, L. Nilsson, Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution, Biophys. J. 68 (1995) 402–426.
[19] T. Bishop, K. Schulten, Molecular dynamics study of glucocorticoid receptor DNA-binding, Proteins 24 (1996) 115–133.
[20] G. Löffler, H. Schreiber, O. Steinhauser, Calculation of the dielectric properties of a protein and its solvent: theory and a case study, J. Mol. Biol. 270 (1997) 520–534.
[21] D. Bashford, M. Karplus, The p K_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model, Biochemistry 29 (1990) 10219–10225.
[22] M. Schaefer, H.V. Vlijmen, M. Karplus, Electrostatic contributions to molecular free energies in solution, Adv. Protein Chem. 51 (1998) 1–57.
[23] J. Warwicker, Simplified methods for p K_a and acid pH-dependent stability estimation in proteins: removing dielectric and counterion boundaries, Protein Sci. 8 (1999) 418–425.
[24] T. Simonson, Macromolecular electrostatics: continuum models and their growing pains, Curr. Opin. Struct. Biol. 11 (2001) 243–252.
[25] T. Simonson, Electrostatics and dynamics of proteins, Rep. Prog. Phys. 66 (2003) 737–787.
[26] W. Richardson, C. Peng, D. Bashford, L. Noodleman, D. Case, Incorporating solvation effects into density functional theory: calculation of absolute acidities, Int. J. Quantum Chem. 61 (1997) 207–217.
[27] D. Fabris, J. Zaia, Y. Hathout, C. Fenselau, Retention of thiol protons in two classes of protein zinc ion coordination centers, J. Am. Chem. Soc. 118 (1996) 12242–12243.
[28] D. Fabris, Y. Hathout, C. Fenselau, Investigation of zinc chelation in zinc-finger arrays by electrospray mass spectrometry, Inorg. Chem. 38 (1999) 1322–1325.
[29] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople Jr., Gaussian98, Revision A, Gaussian, Inc., Pittsburgh, PA, 1998.
[30] A. Becke, Density functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648.
[31] C. Lee, W. Yang, R. Parr, Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density, Phys. Rev. B 37 (1988) 785–789.
[32] T. Dudev, J. Cowan, C. Lim, Competitive binding in magnesium coordination chemistry: water versus ligands of biological interest, J. Am. Chem. Soc. 121 (1999) 7665–7673.
[33] A. Mackerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, An all-atom empirical potential for molecular modelling and dynamics study of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[34] G. Zhang, M. Kazanietz, P. Blumberg, J. Hurley, Crystal structure of the Cys_2 activator binding domain of protein kinase C delta in complex with phorbol ester, Cell 81 (1995) 917.
[35] H. Mott, J. Carpenter, S. Zhong, S. Ghosh, R. Bell, S. Campbell, The solution structure of the Raf-1 cysteine rich doma: a novel Ras and phospholipid binding site, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 8312.
[36] W. Jorgensen, J. Chandrasekar, J. Madura, R. Impey, M. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[37] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald: an $N\log(N)$ method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.
[38] O. Becker, M. Watanabe, Dynamics methods, in: O. Becker, A. Mackerell, B. Roux, M. Watanabe, Jr. (Eds.), Computational Biochemistry & Biophysics, Marcel Dekker, N.Y., 2000 (chapter 3).
[39] J. Ryckaert, G. Ciccotti, H. Berendsen, Numerical integration of the cartesian equations of motion for a system with constraints: molecular dynamics of *n*-alkanes, J. Comp. Phys. 23 (1977) 327–341.
[40] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations, J. Comp. Chem. 4 (1983) 187–217.
[41] J. Madura, J. Briggs, R. Wade, M. Davis, B. Luty, A. Ilin, J. Antosiewicz, M. Gilson, B. Baheri, L. Scott, J. McCammon, Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program, Comp. Phys. Comm. 91 (1995) 57–95.
[42] N. Morellet, H. Demene, V. Teilleux, T. Huynh-Dinh, H. de Rocquigny, M. Fournie-Zaluski, B. Roques, Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide dACGCC, J. Mol. Biol. 283 (1998) 419.
[43] T. South, M. Summers, Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger domain of the HIV nucleocapsid protein: NMR structure of the complex with the psi-site analog, dACGCC, Protein Sci. 2 (1993) 3.
[44] G. Meinke, P. Sigler, DNA-binding mechanism of the monomeric orphan nuclear receptor NGFI-B, Nat. Struct. Biol. 6 (1999) 471.
[45] J. Antosiewicz, J. McCammon, M. Gilson, Prediction of pH dependent properties of proteins, J. Mol. Biol. 238 (1994) 415–436.
[46] L. Hemmingsen, P. Amara, E. Ansoborio, M. Field, Importance of charge transfer and polarization effects for the modelling of uranyl–cation complexes, J. Phys. Chem. A 104 (2000) 4095–4101.
[47] T. Simonson, The dielectric constant of cytochrome c from simulations in a water droplet including all electrostatic interactions, J. Am. Chem. Soc. 120 (1998) 4875–4876.