Published in final edited form as:

J Comput Chem. 2012 April 30; 33(11): 1142–1151. doi:10.1002/jcc.22944.

Importance of electrostatic polarizability in calculating cysteine acidity constants and copper(I) binding energy of *Bacillus subtilis* CopZ

Timothy H. Click, Sergei Y. Ponomarev, and George A. Kaminski*

Department of Chemistry and Biochemistry, Worcester Polytechnic Institute, Worcester, MA 01609

Abstract

CopZ is a copper chaperone from *Bacillus subtilis*. It is an important part of Cu(I) trafficking. We have calculated pKa values for the CXXC motif of this protein which is responsible for the Cu(I) binding, and the Cu(I) binding constants. Polarizable and fixed-charges formalisms were employed, and solvation parameters for the both models have been refitted. We had to partially redevelop parameters for the protonated and deprotonated cysteine residues. We have discovered that the polarizable force field (PFF) is qualitatively superior and allows a uniformly better level of energetic results. The PFF pKa values for cysteine are within ca. 0.8–2.8 pH units of the experimental data, while the fixed-charges OPLS formalism yields errors of up to tens of units. The PFF magnitude of the copper binding energy is about 10 kcal/mol or 50% higher than the experimental value, while the using the refitted OPLS parameters leads to an overall positive binding energy, thus predicting no thermodynamically stable complex. At the same time, the agreement of the polarizable S...Cu(I) distances with the experimental results is within 0.08 Å range, and the non-polarizable calculations lead to an error of about 0.4 Å. Moreover, the accuracy of the PFF has been achieved without any explicit fitting to either pKa or CopZ...Cu(I) binding energies. We believe that this makes our polarizable technique a choice method in reproducing protein – copper binding and further supports the notion that explicit treatment of electrostatic polarization is crucial in many biologically relevant studies, especially ion binding and transport.

Keywords

polarizable force fields; protein-cation interactions; protein pKa shifts

I. Introduction

Copper is involved in a variety of biologically important processes related to toxicity, electron transfer, photosynthesis and oxygen metabolism in general.¹ It has to be transported without binding to sites of other metal.^{1b} Therefore, selectivity of copper binding is important both in nature and in any computational studies which would lead to biomedically interesting results. This means that binding constants or energies of copper binding to proteins have to be computed accurately. Moreover, the CXXC motif with two cysteine residues is a common system in binding the Cu(I) ion.^{1a,b} The importance of the pH on the copper trafficking has also been demonstrated.^{1c–g} This means that accurate computational studies of copper binding and transport would also benefit greatly from the ability to predict reaction of proteins to the pH level (*i.e.*, protein pKa shifts) within a reasonable error margin

Corresponding author: gkaminski@wpi.edu.

of about 1 – 2 pH units. While cancellation of errors or knowledge-based assignment of ionization states can sometimes help avoid the necessity for accurate calculations of protein acidity constants, the robustness of such computational studies is lesser than that of applying methods rooted in correct physical representation of the underlying physical phenomena.

It has been demonstrated previously that the polarizable force field is superior in reproducing protein pKa shifts, as compared to the fixed-charges force fields, such as the OPLS-AA (the fundamental difference is that polarizable force field includes inducible electrostatic dipoles to respond to changing electrostatic environment and fixed-charges ones are unable to produce such an adjustment). For example, the acidity constants of acidic and basic residues of the turkey ovomucoid third domain (OMTKY3) were determined within 0.6 and 0.7 pH units of the available experimental data, respectively, when PFF was employed. At the same time, the best errors with the fixed-charges OPLS were 3.3 and 2.2 pH units.^{2,3} Absolute pKa values for several small systems were determined within 0.8 units by using PFF, with the fixed-charges counterpart yielding an average error of about 5 pH units.⁴ Moreover, all these results were achieved without any explicit fitting of parameters to the pKa values, therefore we believe that the advantages of using the polarizable force field result from the correct physics of the force field formalism.

Non-polarizable force fields are also known to give a significant error in assessing energies of ions with polarizable systems.⁵ Moreover, while the Cu(I) ion is biologically important, robust theoretical models for it are not very numerous. There are cases when metal ions interacting with proteins are represented simply by point electrostatic charges.⁶ This imposes limitations on the calculations and the accuracy of the results when employing molecular mechanics, molecular dynamics and Monte Carlo simulations. One possible alternative to developing copper parameters for empirical force fields is to carry out quantum mechanical or combined quantum mechanical/molecular mechanical simulations,^{7,8} but they have their limitations as well. In other cases, simulations including these ions have to use a number of additional parameters, such as constraints prohibiting the ion from leaving the prescribed coordinated place.⁹ At the same time, lack of explicitly included electrostatic polarization has been shown to create problems in assessing interaction energies. For example, we have shown that variations of the non-polarizable OPLS force field produce binding energies of Cu(I) with benzene which are two to four times lower in magnitude than the reference number.¹⁰

Overall, comparing performance of polarizable and non-polarizable formalism in calculating CXXC motif acidity constants and binding energies with the Cu(I) ion appears to be an important step in developing the predictive and analytical power of computational methods in the area of ion binding and transport.

CopZ is a copper chaperone from the bacterium *Bacillus subtilis*. Structural and mechanical information is now available, and specific interaction between CopZ and the ATPase CopA has been demonstrated.^{1a,b} Moreover, detailed studies of binding properties of Cu(I) ion and CopZ have recently been reported.^{1a} Therefore, we choose this system as a natural target for our comparison of the validity of the polarizable and fixed-charges force fields as applied to the copper binding and transport studies.

The rest of the paper is organized as follows. Described in Section II are methods employed in this study. Section III contains results of the calculations as well as a discussion of these data. Conclusions are presented in Section IV.

II. Methods

General Scheme of the Calculations

We considered the chain of processes presented on Figure 1, which represent a part of the sequence described in Reference 1a.

First, the two cysteine residues of the CXXC binding cite are deprotonated. We determined the pKa1 and pKa2 acidity constants for this process in the same way as we calculated them in References 2 and 3 (a brief description of these calculations is given in the subsection below). In order to determine which of the two residues (Cys13 or Cys16) is deprotonated first, we calculated the deprotonation energies of the both of them and picked the one with the lower final energy of the system.

After that, we have calculated the binding energy of the doubly deprotonated CopZ with the Cu(I) ion. This step was accomplished by running geometry optimizations of the CopZ...Cu(I) complex, as well as of the copper ion and the CopZ protein alone, and subtracting the resulting energies of the monomers from the optimized energy of the dimer. All the geometry optimizations in this work were carried out with the IMPACT software suite.¹¹

Computing the pKa shifts

We calculated pKa shifts for both Cys13 and Cys16 cysteine residues of the CXXC motif of the CopZ protein. In calculating each of these two values, we considered the following two processes. The first one was the deprotonation energy for CH₃SH, the reference acid for the cysteine residues. All the reactions are occurring in aqueous solution. The reaction is Reference Acid-H → Reference-Acd⁻ + H⁺. The change of free energy for this reaction is

$$\Delta G_1 = G(\text{Reference Acid}^-) + G(\text{H}^+) - G(\text{Reference Acid-H}) \quad (1)$$

Similarly, for the cysteine residue deprotonation reaction: Cys-H → Cys⁻ + H⁺, the free energy change is:

$$\Delta G_2 = G(\text{Cys}^-) + G(\text{H}^+) - G(\text{Cys-H}) \quad (2)$$

The acidity constants pKa(Reference Acid) and pKa(Cys) for these deprotonation processes are:

$$\text{pKa}(\text{Reference Acid}) = \Delta G_1 / (2.303RT) \quad (3a)$$

$$\text{pKa}(\text{Cys}) = \Delta G_2 / (2.303RT) \quad (3b)$$

The pKa shift, *i. e.* the difference between the acidity constants for the residue and the reference acid, is:

$$\Delta \text{pKa} = \text{pKa}(\text{Cys}) - \text{pKa}(\text{Reference Acid}) = [G(\text{Cys}^-) - G(\text{Cys-H}) - G(\text{Reference Acid}^-) + G(\text{Reference Acid-H})] / (2.303RT)$$

or

$$\text{pKa}(\text{Cys}) = \text{pKa}(\text{Reference Acid}) + [G(\text{Cys}^-) - G(\text{Cys-H}) - G(\text{Reference Acid}^-) + G(\text{Reference Acid-H})] / (2.303RT)$$

Equation (5) above is used for calculation of the pKa values for the Cys residues, with the reference compound being CH₃SH and the pKa(Reference) value set to 10.3 pH units.¹²

Values of the energies G(Cys⁻), G(Cys-H), G(Reference Acid⁻), and G(Reference Acid-H) were obtained via geometry optimizations with Poisson-Boltzman (PBF) continuum solvent model. The same solvation formalism was employed in all the other non-gas-phase simulations, including the CopZ···Cu(I) dimerization and parameter fitting. We have previously demonstrated that the PBF model is superior in performance than the Surface Generalized Born (SGB) when accurate values of pKa shifts are desired.² The temperature was set to 298.15 K in all the acidity constant calculations presented in this article.

Force Fields

Although some potential energy parameters have been refitted in the course of this project, the general force field formalism of the fixed-charges OPLS¹³ and polarizable PFF^{4,14} models was retained. The total energy E_{total} is computed as a sum of the electrostatic term $E_{electrostatics}$, the non-electrostatic van-der-Waals part E_{vdW} , bond stretching and angle bending energies E_{bond} and E_{angles} , and the torsional term $E_{torsion}$:

$$E_{total} = E_{electrostatics} + E_{vdW} + E_{bonds} + E_{angles} + E_{torsion} \quad (6)$$

For the polarizable force field (PFF), the total electrostatic energy of the model results from interactions of fixed-magnitudes atomic charges, fixed point dipoles, and inducible dipoles:

$$U(\{q_{ij}\}, \{\mu_i\}) = \sum_i (\chi_i \cdot \mu_i + \frac{1}{2} \mu_i \cdot \alpha_i^{-1} \cdot \mu_i) + \frac{1}{2} \sum_{ij \neq kl} q_{ij} J_{ij,kl} q_{kl} + \sum_{ij,k} q_{ij} \mathbf{S}_{ij,k} \cdot \mu_k + \frac{1}{2} \sum_{i \neq j} \mu_i \cdot \mathbf{T}_{i,j} \cdot \mu_j \quad (7)$$

Here $J_{ij,kl}$ is a scalar coupling between bond-charge increments on sites i,j and k,l q_{ij} and q_{kl} ; $\mathbf{S}_{ij,k}$ is a vector coupling between a bond-charge increment on sites i,j and a dipole on site k μ_k ; a rank-two tensor coupling $\mathbf{T}_{i,j}$ describes interactions between dipoles on sites i and j . α_i is the polarizability of site i . Parameters χ_i describe “dipole affinity” of site i .

Following the Coulomb formalism,

$$J_{ij,kl} = \frac{1}{r_{ik}} - \frac{1}{r_{il}} - \frac{1}{r_{jk}} + \frac{1}{r_{jl}} \quad (8)$$

$$\mathbf{S}_{ij,k} = \frac{\mathbf{r}_{ik}}{r_{ik}^3} - \frac{\mathbf{r}_{jk}}{r_{jk}^3} \quad (9)$$

$$\mathbf{T}_{i,j} = \frac{1}{r_{ij}^3} \left(\mathbf{1} - 3 \frac{\mathbf{r}_{ik} \mathbf{r}_{ik}}{r_{ij}^2} \right). \quad (10)$$

Inducible dipoles are placed on all the heavy atoms and on some polar hydrogens (such as the water hydrogens), as described in Reference 14a. Fixed charges q and permanent dipoles defined by the “dipole affinities” χ are present on all the atoms. “Virtual sites” representing electron lone pairs were placed on the water oxygen atoms with the virtual site – oxygen distances set to 0.47 Å. Dipole-dipole and charge-dipole screening is used to damp the polarization response when the perturbing site is at short distances. Each dipole and each

charge has an individually set screening length. The corresponding length parameters for two atoms are added together to obtain the effective screening length for the pair interaction.

The overall non-electrostatic pair potential form in the PFF is:

$$E_{nb} = \sum_{i < j} A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^6 + C_{ij} \exp(r_{ij}/\alpha_{ij}) \quad (11)$$

The A parameter is set so that the $1/r^{12}$ term is close to zero in the hydrogen bonding region, but is large enough to prevent atoms from being positioned too close and thus penetrating the nonphysical region of the phase space.

Finally the bond stretching and angle bending terms have the standard harmonic functional form, and the torsional energy is described by a sum of Fourier series:

$$E_{bonds} = \sum_{bonds} K_r (r - r_{eq})^2 \quad (12)$$

$$E_{angles} = \sum_{angles} K_{\Theta} (\Theta - \Theta_{eq})^2 \quad (13)$$

$$E_{torsion} = 1/2 \sum_{dihedrals} V_1^i [1 + \cos(\phi_i)] + V_2^i [1 - \cos(2\phi_i)] + V_3^i [1 + \cos(3\phi_i)] \quad (14)$$

Here K_r and K_{Θ} represent the force constants; r , r_{eq} , Θ , and Θ_{eq} are actual and equilibrium values of bond lengths and angles; ϕ are the dihedral angles.

We utilized the values of the parameters from previously developed PFF models for water⁴ and proteins^{14b}, except where explicitly noted.

The non-polarized OPLS-AA force field differs in its non-bonded component, with the $E_{electrostatics}$ and E_{vdW} terms in Equation 6 replaced by the sum of the Coulomb and Lennard-Jones contributions for pairwise intra- and intermolecular interactions:

$$E_{nb} = \sum_{i < j} [q_i q_j e^2 / r_{ij} + 4\epsilon_{ij} (\sigma_{ij}^{12} / r_{ij}^{12} - \sigma_{ij}^6 / r_{ij}^6)] f_{ij} \quad (15)$$

Geometric combining rules for the Lennard-Jones coefficients were employed. The summation runs over all the pairs of atoms $i < j$ on molecules A and B or A and A for the intramolecular interactions. Moreover, in the latter case, the coefficient f_{ij} is equal to 0.0 for any i - j pairs connected by a valence bond (1–2 pairs) or a valence bond angle (1–3 pairs). $f_{ij} = 0.5$ for 1,4-interactions (atoms separated by exactly 3 bonds) and $f_{ij} = 1.0$ for all the other cases. Values of parameters for water and the protein were adopted from the standard OPLS-AA^{13,15} except for the cases explicitly noted below.

Cu(I) Parameters and Parameter Refitting

The PFF parameters and two sets of fixed-charges OPLS-like parameters for the copper(I) ion have been adopted from our previous work¹⁰ without any adjustment. The OPLS-like versions are termed TIP3P and TIP4P. These notations correspond to the two fixed-charges sets of parameters which were originally parameterized to reproduce energies and distances

for the Cu(I) gas-phase dimers with one water molecule in which the latter was simulated with the corresponding OPLS-AA water model.

Simulation of protonated and deprotonated cysteine residues 13 and 16 of the CopZ protein for both the pKa and Cu(I) binding calculations required sulfur and hydrogen parameters for the $-SH$ and sulfur parameters for the deprotonated $-S^-$ groups. We obtained these parameters by refitting CH_3SH and CH_3S^- models so that energies and distances for the gas-phase dimerization complexes with one water molecule shown on Figure 2 would be in agreement with quantum mechanical data.

The quantum mechanical target fitting data were obtained using a previously developed extrapolation procedure which utilizes LPM2/cc-pVTZ(-f) and LMP2/cc-pVQZ data and which was shown to demonstrate excellent results in assessing geometry and energy of hydrogen-bonded complexes.¹⁶ The calculations were carried out using Jaguar 7.6 software.¹⁷

The next step in refitting potential energy parameters was in fitting the parameters for the PBF continuum solvation model employed in the pKa and CopZ...Cu(I) dimerization calculations. We adjusted the reaction field radii and cavity radii for the copper(I) ion models and for the CH_3SH and CH_3S^- models for both the PFF and OPLS-like potential energy functions. This fitting was done in order to achieve a good level of agreement with the reference hydration energy. The copper(I) free energy of hydration was determined in the project described in Reference 10, and the reference data for the protonated and deprotonated methanethiol are from Reference 18.

Geometry Optimizations

Geometry optimizations for both the OPLS-like fixed-charges and polarizable PFF models were carried out with the IMPACT software suite.¹¹ The CopZ and CopZ...Cu(I) simulations involved only a part of the protein immediately neighboring the two Cys residues used in the pKa calculations and in the copper(I) ion binding, as shown on Figures 3 and 4 (presented on Figure 3 is only the sequence of the residues included in these calculations, while Figure 4 demonstrates the special structure). We used apo- and holo-CopZ geometries from the PDB structures 1P8G and 1K0V, respectively.

Moreover, since only a rather small part of the protein was explicitly present, its geometry was fixed with the exception of the two cysteine residues (protonated or deprotonated). In case of the CopZ...Cu(I) complex, the Cu(I) ion was allowed to move freely.

The method of explicit inclusion of only a part of a protein close to its active site was used to speed up the calculations and to avoid unnecessary noise in the total energy value. We had employed this technique before in other protein acidity constant calculations^{2,3} following the original specific technique for choosing such a fragment presented in Reference 19.

III. Results and Discussion

Non-Bonded Parameter Refitting for CH_3SH and CH_3S^-

OPLS and PFF parameters for the sulfur atom in the $-CH_2SH$ and $-CH_2S^-$ groups were refitted with the goal to achieve a close agreement with the quantum mechanical $S\cdots O$ distances and dimerization energies for the complexes with one water molecule shown on Figure 2. In case of the fixed-charges fitting, the TIP3P version of water^{13a} was used. For the PFF, we employed a previously developed polarizable water model.⁴ The final parameters are shown in Table 1. In case of the OPLS-like fixed charges parameters, only

the Lennard-Jones σ and ϵ were refitted and only for the sulfur atoms. For the PFF formalism, a new atomtype for the sulfur atom in the CH_3S^- system was created. The van-der-Waals parameters for the CH_3S^- sulfur were produced, while these parameters for the CH_3SH molecule (and the Cys residue) were adopted without any change from the standard values in References 14a and 14b, except for the value of the van-der-Waals parameter alpha. In addition, the following changes were made to the PFF parameter values: (i) the C-S bond-charge increment for the neutral CH_3SH was set to 0.35 (the positive sign means electron transfer from C to S); (ii) the new sulfur atomtype for the charged CH_3S^- molecule has the same electrostatic parameters as the standard PFF S in CH_3SH , except that the C-S bond-charge increment is set to -0.10 (electron transfer from S to C) and a formal charge of -1 electron added to S.

The resulting values of the S...O distances and dimerization energies between the protonated and deprotonated methanethiol and one water molecule (the systems shown on Figure 2) are introduced in Table 2. This procedure of fitting gas-phase dimerization distances and energies is typical for our force field development.^{4,14} Our aim was to achieve the accuracy of ca. 0.05 Å in the distance and 0.5 – 1.0 kcal/mol in the energy. Imposing tighter criteria would not be justified given the intrinsic error present in the quantum mechanical calculations, as had been observed previously for sulfur-containing hydrogen bonding.^{14a} As can be seen from Table 2, we have successfully refitted the sulfur parameters, for both the fixed-charges and the polarizable formalisms, and could thus proceed to the next step.

Refitting Continuum Solvation Parameters for the Cu(I) ion and Cys and Cyx Residues

Since the actual CopZ and CopZ...Cu(I) calculations had to be done in aqueous solution, it was important to make sure that the parameters for the copper(I) ion and the CH_3SH and CH_3S^- groups employed in the PBF Poisson-Boltzmann solvation model were adequate in reproducing the free energy of hydration for these species. The resulting adjusted parameters (reaction field radii and cavity radii) are shown in Table 3. For the protonated and deprotonated methanethiol, the only cavity radius that was changed was that for the sulfur in the OPLS CH_3SH , while all the reaction field values were changed. At the same time, we had to both refit the reaction field radii for the Cu(I) ion and eliminate the cavity radius, both for the OPLS-like and the PFF formalism. As a result, the solvation free energies which are shown in Table 4 are in a very close agreement with the reference data. The methanethiol solvation energies are accurate to 0.05 kcal/mol, and the other values are no more than 0.07 kcal/mol off, with the exception of the OPLS(TIP3P) copper(I) ion which has the greatest error of only 0.13 kcal/mol. Overall, we can conclude that our fitting was successful.

At this point, all the parameters needed for the calculations involving the CopZ protein, its cysteine residues and its complex with the Cu^+ particle are present, and we can proceed with description of the calculations of the Cys13 and Cys16 pKa shifts.

Calculating Cysteine pKa shifts

We used Equation 4 to calculate the shifts in the acidity constants. First of all, we calculated pKa shift of a simple methyl-capped cysteine dipeptide. Protonated and deprotonated forms of this dipeptide are shown on Figures 5 and 6, respectively. Both systems were completely flexible with no constraints in the geometry optimizations.

It should be noted that these structures are not true isolated dipeptides but rather methyl-capped forms of the dipeptides utilized to mimic cysteine residues in polypeptides. We have been using such capped structures in our procedure of developing force fields for proteins, for example, in work described in References 14b and 15.

We calculated energies of the both forms, as well as of the reference acid CH₃SH, and they are all presented in Table 5. Then Equation 4 was employed to find out the resulting pKa values of the methyl-capped cysteine dipeptide, based on the experimental value of the CH₃SH pKa of 10.3¹² and differences of the hydration energies for the dipeptide and this reference acid. The results were as follows. The PFF value of the pKa was found to be 7.35 pH units, or 0.8 pH units smaller than the experimental value of 8.14.²⁰ At the same time, the OPLS-like fixed charges number was 19.05 units, or about 11 units greater than the reference. This result is not entirely unexpected given our previously calculated OPLS pKa errors for the protein OMTKY3 residues which could be wrong by up to 9.3 pH units.²

Having calculated the pKa shift for the simple cysteine dipeptide, we then proceeded to the cysteine residues 13 and 16 of the CopZ protein. Once again, methanethiol was used as the reference compound, its pKa value assumed to be 10.3 pH units.¹² Both pKa1 and pKa2 were computed. To find the correct order of deprotonating these two cysteine residues, we calculated deprotonating energies for the following two processes:

The reaction with the lower final energy was assumed to be the one corresponding to the experimental pKa1 ≤ 4.^{1a} We used the fragment of the CopZ protein shown on Figure 4 and the PBF continuum solvation model for the hydration component of the Hamiltonian. Values of all the calculated energies for the first cysteine deprotonation are given in Table 6. Also shown in the same table are the energies employed in simulating deprotonation of the reference CH₃SH molecule.

The first fact to notice is that the energy of singly deprotonated CopZ is significantly lower for the Cyx13/Cys16 than Cys13/Cyx16 system, as computed with both the OPLS (about 60 kcal/mol difference) and PFF (40 kcal/mol difference) techniques. That is, the Cys13 residue is clearly deprotonated first. This can be physically explained by noticing the following about the structure on Figure 7.

While the Cys13 residue is exposed to the solvent and its side-chain has no hydrogen bonds to the rest of the protein, the Cys16 residue is more buried and its H(S) hydrogen is coordinated by two oxygen atoms of the charged Asp66 residue. This lowers the energy of the protonated form of Cys16, and thus converting Cys16 to the negatively charged deprotonated Cyx16 is less favorable energetically than removing the H(S) hydrogen from the Cys13 part of the molecule. Therefore, we assume that the first deprotonation occurs in accordance with Equation 16a.

While the order of deprotonation is the same with the OPLS and PFF, the resulting energy differences are very different. While the OPLS predicts that the deprotonation of the Cys13 residue lowers the energy of the system by −53.03 kcal/mol, the PFF result is −82.72 kcal/mol. The difference is qualitative, since the change in the energy produced by the OPLS is smaller than that of the simple Cys dipeptide (−61.19 kcal/mol), while the magnitude of the PFF result is greater than the PFF cysteine dipeptide difference of −80.47 kcal/mol. Thus, the PFF reproduces this property better than the OPLS, since the CopZ Cys pKa is known from experimental studies to be lower than that of a simple cysteine.^{1a} And the OPLS shows the same tendency of overestimating the acidity constant values as was demonstrated before.^{2,3} The overall pKa1 value is 25.03 pH units as computed with the OPLS and 5.70

units from our polarizable calculations, which is within a much more reasonable error margin of ca. 2 pH units from the experimental result of ≤ 4 .^{1a}

Values of energies for deprotonation of the second cysteine residue (Cys16) in the Cys13/Cys16 pair are given in Table 7. Two numbers are given for the Cyx13/Cyx16 energies (and thus for the energy differences) for both OPLS-AA and PFF. These numbers correspond to the Cyx/Cyx geometries found in the apo- and holo- forms of the CopZ protein, respectively. The reason is that the PDB apo-geometry corresponds to the doubly-protonated CopZ system, while the CopZ in the complex with the Cu(I) ion is doubly deprotonated. Therefore, it is not unreasonable to suggest that the holo-conformation corresponds to the lower energy case and thus should be used for the Cyx/Cyx energy in the pKa calculations. This hypothesis is confirmed by the fact that both the OPLS and the PFF results are much better if the holo-configuration is used (the second number in the table data for Cyx13/Cyx16).

The resulting PFF energy difference between the singly and doubly deprotonated CopZ energies is -78.51 kcal/mol, which is only slightly lower than the methanethiol deprotonation energy change of -76.45 kcal/mol. At the same time, the OPLS result is only -10.74 kcal/mol, which leads to a disastrous overall fixed-charges OPLS pKa2 of 56.03 pH units. The polarizable PFF yields a total pKa2 value of 8.79 units. It is greater than the reference experimental ~ 6 pH units, but is in a much better agreement.

All the calculated values of the acidity constants are summarized in Table 8.

Cu(I) binding to CopZ

The final set of the calculations presented in this work was in finding the binding energy of the Cu(I) ion with the doubly deprotonated CopZ protein. The complex is shown on Figures 8 (for OPLS, TIP3P-compatible set) and Figure 9 (PFF).

Only the part of the protein explicitly employed in the simulations is shown on these Figures. The energy results are given in Table 9. Overall, the binding energy (calculated as the difference between the total complex energy and the energies of the separate components, the CopZ fragment and the Cu(I) ion) obtained with the PFF technique was found to be -33.05 kcal/mol. This number is somewhat greater than the experimental result of -22.48 kcal/mol. However, the agreement here is qualitatively better than that with the fixed-charges OPLPS force field. Both versions of the OPLS give reasonably close results with binding energies of ca. $+10$ kcal/mol. The positive sign of the binding energy means that the complex corresponds to a local, not global minimum, and it is thus thermodynamically unstable. Thus, even the qualitative existence of the complex is predicted with the polarizable PFF, but not fixed-charges OPLSS force field.

Moreover, the average Cu(I)···S⁻ distance produced with the OPLS (and shown on Figure 8) is ca. 2.56 Å. This is much longer than the experimentally observed value of 2.14 – 2.15 Å.⁶ At the same time, the polarizable methodology produced an average distance of 2.18 Å, with deviations from this number of only 0.06 Å. This is so even though the CH₃SH and CH₃S⁻ complexes with Cu(I) were optimized similarly with the polarizable and fixed-charges methodologies. The polarizable technique is clearly superior and demonstrates a greater degree of transferability of the parameters. We believe that the reason for that is in the basis of the polarizable force field being more physically sound. Thus, even the very structure of the complex is reproduced much better with the polarizable force field.

There is one point which still has to be discussed. The values of the PFF acidity constants for the cysteine residues of the CopZ protein, although qualitatively much better than those

produced with the fixed-charges calculations, are somewhat high, and the PFF binding energy for the CopZ...Cu(I) complex is too attractive. We believe that this can be explained as follows. The conformation used for the doubly protonated Cys13/Cys16 simulations is taken from the PDB structure for this specific ionization state, and thus the Cys13/Cys16 system has its energy close to the global minimum. Once we deprotonate one of the residues, the conformation would change. Therefore, using the same backbone geometry for the Cyx13/Cys16 configuration imposes some penalty as the structure is not optimal for this singly deprotonated state. And the holo-conformation employed for the Cyx13/Cyx16 version, while obviously better than the apo-structure, is still not optimal. This is why the energies of both singly and double deprotonated CopZ are likely to be somewhat overestimated, which increases the pKa values. And, since the Cyx13/Cyx16 energy is overestimated, the Cyx13/Cyx16...Cu(I) binding seems to be more negative in energy than it should be. Therefore, the errors (which are still not excessive) produced with the polarizable methodology, are explained not by insufficient accuracy of the polarizable energy calculations, but rather by imperfections in sampling the conformational space. At the same time, including the whole protein into the geometry optimizations would lead to a level of noise which would be too high and correspond to fluctuations in conformational and solvation energies unrelated to the cysteine deprotonation and Cu(I) binding processes. To fully resolve this issue, one would probably need to perform these simulations with Monte Carlo or molecular dynamics technique and the statistical perturbation theory.

IV. Conclusions

We have calculated cysteine pKa shifts and Cu(I) binding energies for CopZ, a copper chaperone from *Bacillus subtilis* and an important part of Cu(I) trafficking. The cysteine acidity constants were addressed in this study since it is the CXXC motif of this protein which is responsible for the Cu(I) binding. Both polarizable (PFF) and fixed-charges partially reparameterized OPLS-AA force fields were employed. We have found the PFF values of the cysteine acidity constants to be within ca. 0.8–2.8 pH units of the experimental data, while the fixed-charges OPLS formalism yielded errors of up to tens of units. The magnitude of the copper(I) binding energy as obtained with the PFF was overestimated by about 50%, while the fixed-charges formalism lead to an overall positive binding energy. The distances between cysteine sulfur atoms and the bound Cu(I) ion calculated with the polarizable methodology were found to fall within 0.08 Å range from the experimental values of 2.14 – 2.15 Å, and the fixed-charges errors were ca. 0.4 Å.

The above results demonstrate that (i) accurate simulations of energetics of processes related to copper binding and transport with empirical force fields are within our reach; (ii) such simulations require explicit treatment of electrostatic polarization; (iii) Monte Carlo or molecular dynamics with the statistical perturbation theory is likely to be needed for the next qualitative improvement of the results, even though very useful binding information can already be obtained with the current methodology. We hope that our work will permit further advances in the area studies of copper binding and transport, as well as in other areas where biological process involving proteins and ions are investigated.

Acknowledgments

This project was supported by Grant Number R01GM074624 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health. The authors express gratitude to Schrödinger, LLC for the Impact program and to Professor José Argüello for fruitful discussions.

References

1. See for example: (a) Zhou L, Singleton C, Le Brun NE. *Biochem. J.* 2008; 413:459–465. [PubMed: 18419582] (b) Badarau A, Dennison C. *J. Am. Chem. Soc.* 2011; 133:2983–2988. [PubMed: 21323310] (c) Waldon KJ, Robinson NJ. *Nat. Rev. Microbiol.* 2009; 7:25. [PubMed: 19079350] (d) Robinson NJ, Winge DR. *Annu. Rev. Biochem.* 2010; 79:537–562. [PubMed: 20205585] (e) Banci L, Bertini I, Ciofi-Baffoni S, Kozyreva T, Zovo K, Palumaa P. *Nature.* 2010; 465 645-U145. (f) Banci L, Bertini I, Cantini F, Ciofi-Baffoni S. *Cell. Mol. Life Sci.* 2010; 67:2563–2589. [PubMed: 20333435] (g) Bertini I, Cavallaro G, McGreevy KS. *Coord. Chem. Rev.* 2010; 254:506–524.
2. MacDermaid CM, Kaminski GA. *J. Phys. Chem. B.* 2007; 111:9036–9044. [PubMed: 17602581]
3. Click TH, Kaminski GA. *J. Chem. Theory Comput.* 2009; 5:2935–2943. [PubMed: 20209038]
4. Kaminski GA. *J. Phys. Chem. B.* 2005; 109:5884–5890. [PubMed: 16851640]
5. Caldwell JW, Kollman PA. *J. Am. Chem. Soc.* 1995; 117:4177–4178.
6. Banci L. *Curr. Opinion Chem. Biology.* 2003; 7:143–149.
7. Fuchs J-F, Nedev H, Poger D, Ferrand M, Brenner V, Dognon J-P, Crouzy S. *J. Comput Chem. B.* 2006; 27:837–856.
8. Venkateswarlu D. *BMS Struct. Biology.* 2010; 10
9. Rodriguez-Granillo A, Crespo A, Estrin DA, Wittung-Stafshede P. *J. Phys. Chem. B.* 2010; 114:3698–3706. [PubMed: 20166696]
10. Ponomarev SY, Click TH, Kaminski GA. *J. Phys. Chem. B.* 2011; 115:10079–10085. [PubMed: 21761909]
11. Impact version 3. New York NY: Schrödinger, LLC; 2004.
12. Klicic JJ, Friesner RA, Liu S-Y, Guida WC. *J. Phys. Chem. A.* 2002; 106:1327–1335.
13. (a) Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J. Chem. Phys.* 1983; 79:926–935. (b) Jorgensen WL, Severance DL. *J. Am. Chem. Soc.* 1990; 112:4768–4774. (c) Jorgensen WL, Maxwell DS, Tirado-Rives J. *J. Am. Chem. Soc.* 1996; 118:11225–11236.
14. (a) Kaminski GA, Stern HA, Berne BJ, Friesner RA. *J. Phys. Chem. A.* 2004; 108:621–627. (b) Maple JR, Cao YX, Damm WG, Halgren TA, Kaminski GA, Zhang LY, Friesner RA. *J. Chem. Theory. Comput.* 2005; 1:694–715.
15. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. *J Phys Chem B.* 2001; 105:6474–6487.
16. Kaminski GA, Maple JR, Murphy RB, Braden D, Friesner RA. *J. Chem. Theory Comput.* 2005; 1:248–254.
17. Jaguar version 7.6. New York, NY: Schrodinger LLC; 2009.
18. Kelly CP, Cramer CJ, Truhlar DG. *J Phys Chem B.* 2006; 110:16066–16081. [PubMed: 16898764]
19. Jensen JH, Li H, Robertson AD, Molina PA. *J. Phys. Chem. A.* 2005; 109:6634–6643. [PubMed: 16834015]
20. Lide, DR. *CRC handbook of chemistry and physics : a ready-reference book of chemical and physical data.* 87th ed. Boca Raton, Fla: CRC Press; 2006.

Figure 1. Processes considered in this work. Cyx stands for a deprotonated Cys residue with a charge of $-1e^-$.

Figure 2.
Complexes of CH_3SH (a) and CH_3S^- (b) particles with one water molecule used in fitting the sulfur and hydrogen parameters.

Figure 3.

The part of the CopZ protein molecule used in the simulations. HID stands for histidine-delta (hydrogen on the delta-nitrogen), ACE is $-C(=O)-CH_3$ and NMA denotes $-N(-H)-CH_3$.

Figure 4.

The part of the CopZ protein molecule used in the simulations. The system conformation is from the PDB data for the apo-structure.

Figure 5.
Protonated form of cysteine dipeptide used in calculating Cys pKa shift.

Figure 6.
Deprotonated form of cysteine dipeptide used in calculating Cys pKa shift.

Figure 7. Positions of Cys13 and Cys16 residues in the apo-form of CopZ. The latter is coordinated by oxygen atoms of the Asp66 residue.

Figure 8.
OPLS complex of CopZ with Cu(I).

Figure 9.
PFF complex of CopZ with Cu(I).

Table 1

Final van-der-Waals sulfur parameters produced for the sulfur atoms in CH_3SH and CH_3S^- models. Parameters σ and α are in Å, and ϵ and C are in kcal/mol.

Atom	OPLS		PFF		
	σ	ϵ	σ	ϵ	C
S in CH_3SH	3.64	0.800	1.160208	420.25	450000
S in CH_3S^-	4.50	0.066	1.160208	420.25	300000
					0.310

Table 2

Binding energies in kcal/mol and S...O distances in Å for the gas-phase complexes of CH₃SH and CH₃S⁻ with one water molecule shown on Figure 2.

System/Property	OPLS	PFF	QM
CH ₃ SH, Energy	-3.07	-3.89	-3.61
CH ₃ SH, R(S...O)	3.39	3.40	3.40
CH ₃ S ⁻ , Energy	-13.86	-13.93	-12.93
CH ₃ S ⁻ , R(S...O)	3.21	3.20	3.16

Table 3

Final refitted PBF hydration parameters used in the CH_3SH , CH_3S^- , CopZ and CopZ...Cu(I) calculations. Radii are in Å.

Atom/Method	Reaction Field Radius	Cavity Radius
Cu(I)		
OPLS	1.169	0.000
PFF	1.000	0.000
S in CH_3SH		
OPLS	1.710	1.710
PFF	3.115	1.800
S in CH_3S^-		
OPLS	2.075	1.800
PFF	2.255	1.800

Table 4

Hydration energies calculated with the refitted PBF parameters, kcal/mol.

System/Method	ΔG_{hyd}	ΔG_{hyd} , reference
Cu(I)		
OPLS(TIP3P)	-152.71	-152.58 ^a
OPLS(TIP4P)	-152.51	-152.58 ^a
PFF	-152.52	-152.58 ^a
CH ₃ SH		
OPLS	-1.23	-1.24 ^b
PFF	-1.19	-1.24 ^b
CH ₃ S ⁻		
OPLS	-73.83	-73.80 ^b
PFF	-73.76	-73.80 ^b

^aReference 10.^bReference 18.

Table 5

Energies used in calculating pKa shift for the Cys dipeptide with respect to CH₃SH acidity constant, in kcal/mol.

System/Process	Energy	
	OPLS	PFF
CH ₃ SH	0.46	2.21
CH ₃ S ⁻	-72.67	-74.24
CH ₃ SH → CH ₃ S ⁻	-73.13	-76.45
Cys dipeptide (protonated)	-5.71	-8.79
Cyx dipeptide (deprotonated)	-66.90	-89.26
Cys dipeptide → Cyx dipeptide	-61.19	-80.47

Table 6Energies used in calculating pK_a of the CopZ Cys13/Cys16 residues, in kcal/mol.

System/Process	Energy	
	OPLS	PFF
CH ₃ SH	0.46	2.21
CH ₃ S ⁻	-72.67	-74.24
CH ₃ SH → CH ₃ S ⁻	-73.13	-76.45
CopZ, Cys13/Cys16	654.59	-42.90
CopZ, Cys13/Cys16	601.56	-125.62
CopZ, Cys13/Cys16	661.75	-84.70
CopZ, Cys13/Cys16 → Cys13/Cys16	-53.03	-82.72

Table 7

Energies used in calculating pKa2 of the CopZ Cys13/Cys16 residues, in kcal/mol.

System/Process	Energy	
	OPLS	PFF
CopZ, Cyx13/Cys16	601.56	-125.62
CopZ, Cyx13/Cyx16	607.84 ^a / 590.82 ^b	-167.12 ^a / -204.13 ^b
CopZ, Cyx13/Cys16 → Cyx13/Cyx16	6.28 ^a / -10.74 ^b	-41.50 ^a / -78.51 ^b

^aUsing the apo-configuration.^bUsing the holo-configuration.

Table 8

Calculated cysteine pKa's values, in pH units.

System/Process	pKa		
	OPLS	PFF	Experiment
CH ₃ SH			10.3 ^a
Cys dipeptide	19.05	7.35	8.14 ^b
CopZ Cys/Cys → Cys/Cyx	25.03	5.70	≤ 4 ^c
CopZ Cys/Cyx → Cyx/Cyx	56.03	8.79	~ 6 ^c

^aReference 12.^bReference 20.^cReference 1a.

Table 9Energies used in calculating CopZ binding to Cu⁺, in kcal/mol

System/Process	Energy			
	OPLS TIP3P	OPLS TIP4P	PFF	Experiment
Cu ⁺	−152.53	−152.53	−152.52	−152.58 ^a
CopZ Cyx13/Cyx16	590.82	590.82	−204.13	
CopZ Cyx13/Cyx16...Cu ⁺	448.95	448.27	−389.70	
CopZ Cyx13/Cyx16 + Cu ⁺ → CopZ Cyx13/Cyx16...Cu ⁺	10.66	9.98	−33.05	−22.48 ^b

^aReference 10.^bReference 1a.