Inorganic Chemistry
Article

Inorg. Chem. **2010**, *49*, 435–444 **435**
DOI: 10.1021/ic9012735

Protonation of Type-1 Cu Bound Histidines: A Quantum Chemical Study

Peifeng Su and Hui Li*

Department of Chemistry, University of Nebraska-Lincoln, Lincoln, Nebraska 68588

Received July 2, 2009

The protonation of the solvent-exposed histidine ligands of the type-1 Cu sites in five proteins, *Thiobacillus ferrooxidans* rusticyanin, *Pseudomonas aeruginosa* azurin, fern plastocyanin, *Alcaligenes faecalis* pseudoazurin, and *Paracoccus versutus* amicyanin, were studied with quantum chemical methods and conductorlike polarizable continuum model (CPCM). Active site model molecules consisting of ~140 atoms were extracted from X-ray crystal structures and optimized with the homogeneous CPCM/B3LYP/6-31G* method with some atoms fixed. More accurate solvation effects were obtained using a recently developed heterogeneous CPCM method to describe the protein matrix and aqueous solvation of the model molecules. In the heterogeneous CPCM method different effective dielectric constants, 4, 10, and 78.39, were used for different portions of the surfaces encapsulating the active site model molecules. It is found that the two conformations of the protonated histidine, imidazolium unflipped and flipped, show similar energies in the model molecules of these five proteins. The calculated pK_a values are comparable to experimental values. According to the calculations, the main determinants of the pK_a values are local interactions contained in the model molecules and aqueous solvation effects, as well as protein matrix polarization.

1. Introduction

Type-1 Cu centers are critical electron transfer units in biological systems. In a type-1 Cu center, the Cu ion is coordinated by a cysteine thiolate ligand and two histidine $N^{\delta 1}$ ligands. An axial coordination is provided by a methionine thioether ligand in plastocyanin, amicyanin, pseudoazurin, rusticyanin, and azurin. In azurin, a second axial coordination is provided by a glycine backbone carbonyl oxygen ligand.

The solvent-exposed histidine ligands of the reduced type-1 Cu centers in amicyanin, plastocyanin, and pseudoazurin can be protonated. For example, *Paracoccus versutus* (P.v.) amicyanin and *Paracoccus denitrificans* (P.d.) amicyanin show pK_a = 7.2 (297.5 K) and 7.7 (298 K), respectively;[1,2] parsley plastocyanin shows a pK_a = 5.6 at 298 K;[3] poplar plastocyanin shows a pK_a = 4.7 at 298 K;[4] fern plastocyanin shows a pK_a = 4.4 ± 0.1 at 298 K;[5] *Alcaligenes faecalis* (A.f.) pseudoazurin shows a pK_a = 4.8 at 293 K;[6] *Achromobacter cycloclastes* pseudoazurin shows a pK_a = 4.84−4.90 at 298 K;[7,8] and the pK_a of the solvent-exposed histidines in *Pseudomonas aeruginosa* (P.a.) azurin and *Thiobacillus ferrooxidans* (T.f.) rusticyanin have been estimated to be < 2.[9−13] Although these pK_a values have been correlated to the lengths and structures of the ligand-containing loops and the solvent accessibility of the active sites,[8,14−21] the specific interactions that determine these pK_a values have not been identified.

*Corresponding author. E-mail: hli4@unl.edu.

(1) Lommen, A.; Canters, G. W. *J. Biol. Chem.* **1990**, *265*, 2768–2774.
(2) Zhu, Z. Y.; Cunane, L. M.; Chen, Z. W.; Durley, R. C. E.; Mathews, F. S.; Davidson, V. L. *Biochemistry* **1998**, *37*, 17128–17136.
(3) Hunter, D. M.; McFarlane, W.; Sykes, A. G.; Dennison, C. *Inorg. Chem.* **2001**, *40*, 354–360.
(4) Jackman, M. P.; McGinnis, J.; Sykes, A. G.; Collyer, C. A.; Murata, M.; Freeman, H. C. *J. Chem. Soc., Dalton Trans.* **1987**, 2573–2577.
(5) Hulsker, R.; Mery, A.; Thomassen, E. A.; Ranieri, A.; Sola, M.; Verbeet, M. P.; Kohzuma, T.; Ubbink, M. *J. Am. Chem. Soc.* **2007**, *129*, 4423–4429.
(6) Impagliazzo, A. *Trans Protein Interactions: The Case of Pseudoazurin and Nitrite Reductase*. Ph. D. Thesis, Leiden University, The Netherlands, **2005**.
(7) Dennison, C.; Kohzuma, T.; McFarlane, W.; Suzuki, S.; Sykes, A. G. *Inorg. Chem.* **1994**, *33*, 3299–3305.
(8) Dennison, C.; Kohzuma, T.; McFarlane, W.; Suzuki, S.; Sykes, A. G. *J. Chem. Soc.-Chem. Commun.* **1994**, 581–582.
(9) Jeuken, L. J. C.; van Vliet, P.; Verbeet, M. P.; Camba, R.; McEvoy, J. P.; Armstrong, F. K.; Canters, G. W. *J. Am. Chem. Soc.* **2000**, *122*, 12186–12194.
(10) Groeneveld, C. M.; Feiters, M. C.; Hasnain, S. S.; Van Rijn, J.; Reedijk, J.; Canters, G. W. *Biochim. Biophys. Acta* **1986**, *873*, 214–227.
(11) Jeuken, L. J. C.; Ubbink, M.; Bitter, J. H.; van Vliet, P.; Meyer-Klaucke, W.; Canters, G. W. *J. Mol. Biol.* **2000**, *299*, 737–755.
(12) Hunt, A. H.; Toypalmer, A.; Assamunt, N.; Cavanagh, J.; Blake, R. C.; Dyson, H. J. *J. Mol. Biol.* **1994**, *244*, 370–384.
(13) McGinnis, J.; Ingledew, W. J.; Sykes, A. G. *Inorg. Chem.* **1986**, *25*, 3730–3733.
(14) Dennison, C. *Coord. Chem. Rev.* **2005**, *249*, 3025–3054.
(15) Battistuzzi, G.; Borsari, M.; Loschi, L.; Sola, M. *J. Inorg. Biochem.* **1998**, *69*, 97–100.
(16) Battistuzzi, G.; Borsari, M.; Loschi, L.; Sola, M. *J. Biol. Inorg. Chem.* **1997**, *2*, 350–359.
(17) Dennison, C. *Nat. Prod. Rep.* **2008**, *25*, 15–24.
(18) Battistuzzi, G.; Borsari, M.; Canters, G. W.; de Waal, E.; Leonardi, A.; Ranieri, A.; Sola, M. *Biochemistry* **2002**, *41*, 14293–14298.
(19) Yanagisawa, S.; Dennison, C. *J. Am. Chem. Soc.* **2004**, *126*, 15711–15719.
(20) Li, C.; Yanagisawa, S.; Martins, B. M.; Messerschmidt, A.; Banfield, M. J.; Dennison, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 7258–7263.
(21) Li, C.; Banfield, M. J.; Dennison, C. *J. Am. Chem. Soc.* **2007**, *129*, 709–718.

Published on Web 12/14/2009

Another interesting issue is the behaviors of the protonated histidines in these type-1 Cu centers. ^{1}H NMR studies indicate that in low pH solutions, the protonated histidine (the imidazolium group) in spinach plastocyanin[22] and P.v. amicyanin[1,23] can show two conformations: unflipped and flipped. Canters and co-workers further found that for P.v. amicyanin, the ratio of the flipped and unflipped forms in a 297 K solution is ~0.2, corresponding to a ΔG_{flip} = +1.0 kcal/mol.[24] Guss and co-workers found that upon protonation the histidine of the type-1 Cu center in crystalline poplar plastocyanin can flip ~180 degrees.[25] Petratos and co-workers showed that oxidized A.f. pseudoazurin crystal can be reduced and protonated, but the protonated histidine does not flip by ~180 degrees.[26] Instead, it "rotates away" from the Cu^+, as shown by the X-ray structure 1PZB in the PDB.[26] It is not clear whether this is due to an unfavorable energy increase in the flipped form, or simply a high barrier imposed by the rigid environment in the solid state. It is possible that the protonated histidine ligand can show two conformations in a solution. Atomistic insights into this issue have not been obtained from theoretical calculations.

Many methods have been developed for protein pK_a calculation. The continuum electrostatic method, which exists in various forms, is probably the most popular one.[27−39] An empirical method (PROPKA[40]) was also developed. In general, protein pK_a values are determined by protein electrostatic and polarization interactions, as well as the aqueous polarization interactions, as shown by theoretical calculations performed with various methods.[41−45] A recent review by Kamerlin, Haranczyk, and Warshel summarized and compared many, especially QM/MM style methods, for protein pK_a and reduction potential calculations.[46] At the present time, the major difficulty in using accurate QM/MM simulation methods is the daunting computational cost due to extensive sampling. Recently, due to Warshel group's effort, much of the sampling difficulty has been resolved.[46]

Compared to explicit solvation models in QM/MM methods, continuum solvation models contain no specific interaction terms, thus can only empirically predict solvation free energies by using parameterized effective dielectric constants and molecular cavity sizes. However, their simplicity and high efficiency have made them popular choices when only an estimation of the solvation free energy is required. Some authors have explored the possibilities of using continuum models to describe the protein matrix and bulk aqueous solvation on protein active sites.[39,47−49] The current authors' experience in protein pK_a calculations suggests that if the pK_a sites are solvent-exposed, and the model molecules are relatively large (e.g., ~100 atoms), the relative pK_a values of a series of similar ionizable groups can often be well reproduced by a quantum mechanic continuum solvation model.[49] This, of course, is largely due to error cancellations rather than the actual accuracies of the continuum solvation models. In addition, usually only one or a few optimized geometries of the active site model are considered in these QM/Continuum style calculations because extensive dynamic sampling is impractical when the QM parts have > 100 atoms. Therefore, such methods are most useful when dynamic effects are less significant, or when similar dynamic effects can be largely canceled in calculating pK_a and reduction potential shifts.

Another issue in using continuum solvation models for protein active site calculation is the treatment of the microscopic heterogeneous nature of the environment around an active site.[42−44] Unlike QM/MM methods, continuum solvation models cannot be used to accurately describe the heterogeneous protein interactions. However, it can be a good approximation to use different effective dielectric constants for different portions of the system to approximate the heterogeneous polarization of the protein matrix when specific interactions are less significant. A few heterogeneous continuum solvation models have been developed.[50−55] Recently, Iozzi, Cossi, Improta, Rega, and Barone developed a heterogeneous integral equation formalism polarizable continuum model (IEF-PCM)[56,57] to study the pK_a of a

(22) Kojiro, C. L.; Markley, J. L. *FEBS Lett*. **1983**, *162*, 52–56.
(23) Lommen, A.; Canters, G. W.; Beeumen, J. *Eur. J. Biochem*. **1988**, *176*, 213–223.
(24) Canters, G. W.; Lommen, A.; van de Kamp, M.; Hoitink, C. W. G. *BioMetals* **1990**, *3*, 67–72.
(25) Guss, J. M.; Harrowell, P. R.; Murata, M.; Norris, V. A.; Freeman, H. C. *J. Mol. Biol*. **1986**, *192*, 361–387.
(26) Vakoufari, E.; Wilson, K. S.; Petratos, K. *FEBS Lett*. **1994**, *347*, 203–206.
(27) Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219–10225.
(28) Yang, A. S.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. *Proteins-Structure Function and Genetics* **1993**, *15*, 252–265.
(29) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *J. Mol. Biol*. **1994**, *238*, 415–436.
(30) Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. *Biochemistry* **1996**, *35*, 7819–7833.
(31) Nielsen, J. E.; Vriend, G. *Proteins-Structure Function and Genetics* **2001**, *43*, 403–412.
(32) Demchuk, E.; Wade, R. C. *J. Phys. Chem*. **1996**, *100*, 17373–17387.
(33) Georgescu, R. E.; Alexov, E. G.; Gunner, M. R. *Biophys. J*. **2002**, *83*, 1731–1748.
(34) van Vlijmen, H. W. T.; Schaefer, M.; Karplus, M. *Proteins: Struct., Funct. Genet*. **1998**, *33*, 145–158.
(35) You, T. J.; Bashford, D. *Biophys. J*. **1995**, *69*, 1721–1733.
(36) Bashford, D.; Karplus, M.; Canters, G. W. *J. Mol. Biol*. **1988**, *203*, 507–510.
(37) Honig, B.; Nicholls, A. *Science* **1995**, *268*, 1144–1149.
(38) Ullmann, G. M.; Noodleman, L.; Case, D. A. *J. Biol. Inorg. Chem*. **2002**, *7*, 632–639.
(39) Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400–417.
(40) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins* **2005**, *61*, 704–721.
(41) Warshel, A. *Biochemistry* **1981**, *20*, 3167–3177.
(42) Russell, S. T.; Warshel, A. *J. Mol. Biol*. **1985**, *185*, 389–404.
(43) Warshel, A.; Sussman, F.; King, G. *Biochemistry* **1986**, *25*, 8368–8372.
(44) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458–4472.
(45) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *BBA-Proteins Proteom*. **2006**, *1764*, 1647–1676.
(46) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 1253–1272.
(47) Li, J.; Nelson, M. R.; Peng, C. Y.; Bashford, D.; Noodleman, L. *J. Phys. Chem. A* **1998**, *102*, 6311–6324.
(48) Iozzi, M. F.; Cossi, M.; Improta, R.; Rega, N.; Barone, V. *J. Chem. Phys*. **2006**, *124*, 184103–9.
(49) Li, H.; Robertson, A. D.; Jensen, J. H. *Proteins: Struct., Funct., Bioinf.s* **2004**, *55*, 689–704.
(50) Bonaccorsi, R.; Floris, F.; Palla, P.; Tomasi, J. *Thermochim. Acta* **1990**, *162*, 213–222.
(51) Bonaccorsi, R.; Scrocco, E.; Tomasi, J. *Int. J. Quantum Chem*. **1986**, *29*, 717–735.
(52) Bonaccorsi, R.; Hodoscek, M.; Tomasi, J. *J. Mol. Struct.: Theochem*. **1988**, *164*, 105–119.
(53) Bonaccorsi, R.; Ojalvo, E.; Tomasi, J. *Collect. Czech. Chem. Commun*. **1988**, *53*, 2320–2329.
(54) Bonaccorsi, R.; Ojalvo, E.; Palla, P.; Tomasi, J. *Chem. Phys*. **1990**, *143*, 245–252.
(55) Hoshi, H.; Sakurai, M.; Inoue, Y.; Chujo, R. *J. Chem. Phys*. **1987**, *87*, 1107–1115.
(56) Cances, E.; Mennucci, B.; Tomasi, J. *J. Chem. Phys*. **1997**, *107*, 3032–3041.
(57) Mennucci, B.; Cances, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506–10517.

solvent exposed histidine residue in prion protein.[48] In general, despite the approximate nature of the heterogeneous continuum solvation models, they are still useful and efficient tools in protein active site studies, especially for a series of similar solvent-exposed active sites.

In this work, the protonation and pK_a of the solvent-exposed histidine ligands of five type-1 Cu centers in T.f. rusticyanin, P.a. azurin, fern plastocyanin, A.f. pseudoazurin, and P.v. amicyanin were studied with quantum chemical methods. The heterogeneous conductorlike polarizable continuum model (Het-CPCM[58]), recently developed by the authors' group, was incorporated into the B3LYP[59−61] method as a reaction field to describe the protein matrix and aqueous solvation of the active site model molecules, each consisting of ∼140 atoms extracted from X-ray structures. In the Het-CPCM method, different local effective dielectric constants are defined for different surface regions to represent a heterogeneous environment of an active site solvated by protein matrix and by bulk water. The computed structures of the histidine unprotonated and protonated type-1 Cu centers are compared to X-ray and EXAFS data. The computed relative stability of the imidazolium unflipped and flipped states are discussed in terms of hydrogen bonding and solvation energy. The relative pK_a values are also calculated, and their structural determinants are discussed.

2. Computational Methods

The X-ray structures 2CAK,[62] 1E5Y,[63]1KDI,[64] 1PZA,[26] and 1ID2[65] available in the Protein Data Bank (PDB)[66] were selected for T.f. rusticyanin, P.a. azurin, fern plastocyanin, A. f. pseudoazurin, and P.v. amicyanin, respectively. All these structures are in base form, i.e., the solvent-exposed histidine ligands unprotonated. These crystalline structures were obtained at pH = 3.8, 5.5, 4.5, 5.7, and 4.5, respectively. It must be noted here that the protonation states of the solvent-exposed histidine ligands in the solid state may not necessarily be the same as in solutions because protonated histidine ligands are less stable in the solid state due to the loss of some solvent stabilization. As indicated by the authors of the X-ray structures, 2CAK, 1E5Y, 1KDI, and 1PZA are in reduced forms (Cu^+), 1ID2 is in oxidized form (Cu^{2+}). The authors of the current paper did not find a suitable reduced (and with the solvent-exposed His ligand unprotonated) amicyanin X-ray structures in the PDB. The structures of the reduced A.f. pseudoazurin (1PZA and 1PZB) at low pH were obtained from the oxidized structure by lowering the pH first and then reducing it using Na-ascorbate; the reduced forms were not crystallized directly from solution.[26] Hydrogen atoms were added to the X-ray structures with the WHAT IF web interface.[67,68] Model molecules consisting of ∼140 atoms were extracted from the X-ray structures (Figure 1). These model molecules contain the first layer ligands around the Cu ion and the solvent-exposed histidine, with necessary extensions to make chemically meaningful structures. Nearby backbone amide groups (most of them are in hydrogen bonding network) were always kept and capped with methyl groups. The initial geometries of the two acid forms, histidine flipped and unflipped, were prepared by manually adding protons to the histidine rings and adjusting their positions.

All the quantum chemical calculations were performed with the GAMESS package.[69,70] The main method used in this study is B3LYP[59−61] with 96 radial and 302 Lebedev angular grid points, and the standard 6-31G* basis set. In the quantum chemical calculations, the oxidation state of the Cu is +1 and the electron spin multiplicity is one, i.e., closed-shell wave functions were used.

Two CPCM methods were used. One is the commonly used homogeneous CPCM (Homo-CPCM[71,72]) in which one dielectric constant is used to represent a homogeneous and isotropic solvent, such as bulk water. The other is a recently developed heterogeneous CPCM (Het-CPCM[58]) in which different local effective dielectric constants can be defined for different surface regions to represent a heterogeneous environment, such as an active site solvated by protein matrix and bulk water. Practically, this is realized by defining different effective dielectric constants for different spheres used to form the solute cavity. In the current implementation, the CPCM boundary elements or tesserae on the same sphere have the same dielectric constant.

In both the Homo-CPCM and Het-CPCM methods, the simplified united atom model for Hartree−Fock (SUAHF) radii of 0.0, 2.124, 2.016, 1.908, 2.52, and 2.76 Å, respectively, were used for H, C, N, O, S, and Cu atoms except for the atoms on the imidazolium and imidazole rings of the concerned histidines, for which the united atom model for Hartree−Fock (UAHF[73]) radii 2.340, 2.286, 2.286, 2.016, 2.232, 2.016 Å were used, respectively, for C^{β}, C^{γ}, $N^{\delta 1}$, $C^{\delta 2}$, $C^{\varepsilon 1}$, and $N^{\varepsilon 2}$. The absolute solvation energy, and the difference between the acid and base forms of the same molecule, is sensitive to the selection of radii in CPCM calculation. However, the relative pK_a values calculated in the current work for a series of similar solvent-exposed histidine ligands are not very sensitive to the choice of radii. The FIXPVA tessellation scheme[74] recently developed by the authors was used with 240 initial tesserae per sphere. Only the electrostatic solvation free energy was considered; cavitation, dispersion, and repulsion free energies were not included.

In the Homo-CPCM method, a dielectric constant of 78.39 was used. In the Het-CPCM calculations, a dielectric constant of 78.39 was used for the spheres associated with the solvent-exposed histidine ligands (Figure 1), but 4 or 10 was used for the spheres associated with the remaining atoms in the model molecules. Using dielectric constants like 4 or 10 is reasonable because these values have been commonly used for protein interiors in continuum electrostatic methods. It is

(58) Si, D.; Li, H. *J. Chem. Phys.* **2009**, *131*, 044123–8.
(59) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345–351.
(60) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
(61) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(62) Barrett, M. L.; Harvey, I.; Sundararajan, M.; Surendran, R.; Hall, J. F.; Ellis, M. J.; Hough, M. A.; Strange, R. W.; Hillier, I. H.; Hasnain, S. S. *Biochemistry* **2006**, *45*, 2927–2939.
(63) Nar, H.; Messerschmidt, A., To be Published.
(64) Kohzuma, T.; Inoue, T.; Yoshizaki, F.; Sasakawa, Y.; Onodera, K.; Nagatomo, S.; Kitagawa, T.; Uzawa, S.; Isobe, Y.; Sugimura, Y.; Gotowda, M.; Kai, Y. *J. Biol. Chem.* **1999**, *274*, 11817–11823.
(65) Romero, A.; Nar, H.; Huber, R.; Messerschmidt, A.; Kalverda, A. P.; Canters, G. W.; Durley, R.; Mathews, F. S. *J. Mol. Biol.* **1994**, *236*, 1196–211.
(66) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(67) Vriend, G. *J. Mol. Graphics* **1990**, *8*, 52–&.
(68) Rodriguez, R.; Chinea, G.; Lopez, N.; Pons, T.; Vriend, G. *Bioinformatics* **1998**, *14*, 523–528.
(69) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
(70) Gordon, M. S.; Schmidt, M. W., Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005.
(71) Barone, V.; Cossi, M. *J. Phys. Chem. A* **1998**, *102*, 1995–2001.
(72) Li, H.; Jensen, J. H. *J. Comput. Chem.* **2004**, *25*, 1449–1462.
(73) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210–3221.
(74) Su, P. F.; Li, H. *J. Chem. Phys.* **2009**, *130*, 074109.

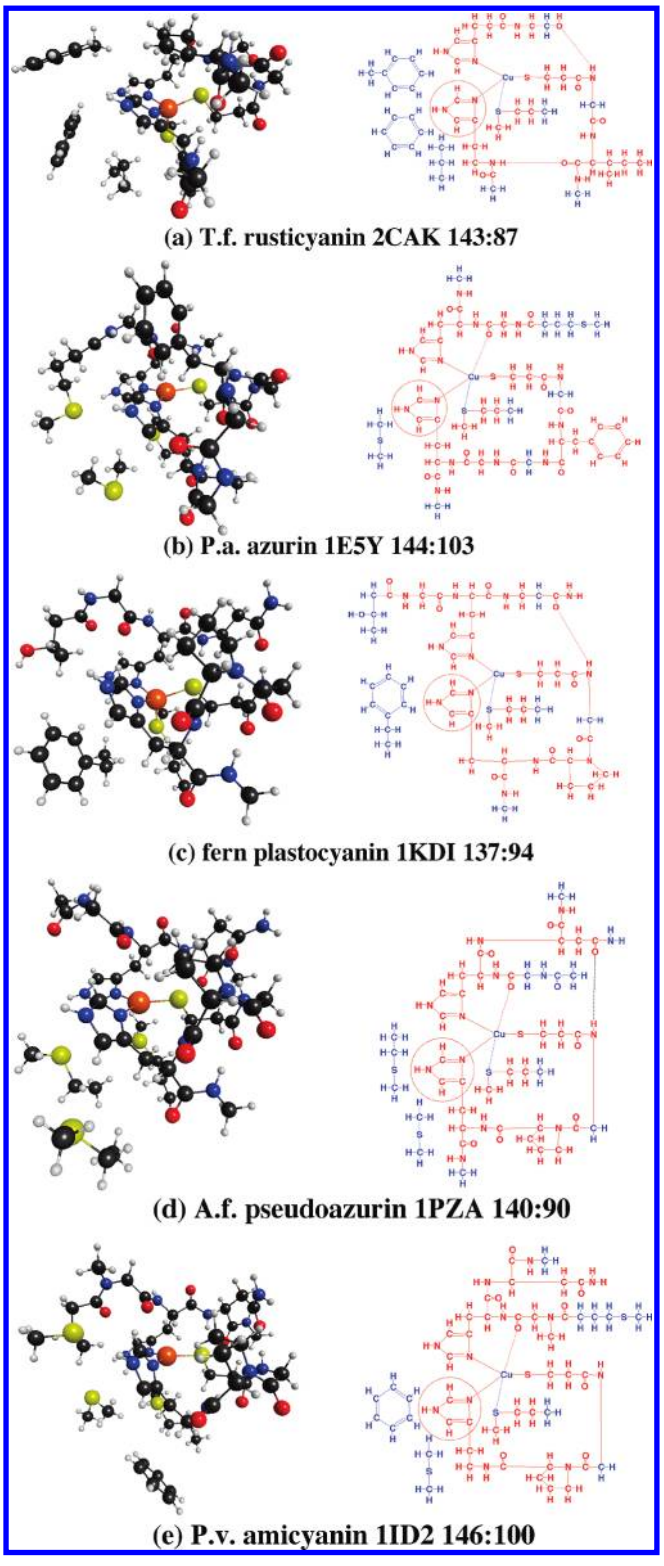

Figure 1. Base form model molecules of the type-1 Cu centers in T.f. rusticyanin (2CAK), P.a. azurin (1E5Y), fern plastocyanin (1KDI), A.f. pseudoazurin (1PZA), and P.v. amicyanin (1ID2). The total number of atoms in the model molecules and the number of optimized atoms are indicated after the PDB file names. In the schematic structures blue atoms are fixed while the red atoms are optimized with the Homo-CPCM/B3LYP/6-31G* method. The solvent-exposed His is cycled.

known experimentally that the dielectric constants of "dry protein powders" are ~1.2, and those of "water-adsorbed protein powders" are typically 2−5.[75−77] It is emphasized here that the accurate effective dielectric constants are unknown, and could be significantly different at different portions around the type-1 centers. In addition, it should be noted that effective dielectric constants for protein interiors are physically meaningful parameters for free energy estimations rather than rigorously defined physical quantities.

Geometry optimizations were performed with the Homo-CPCM/B3LYP/6-31G* method, with some atoms fixed (Figure 1) to mimic the forces imposed by the proteins. The general philosophy is to fix the C^{α} atoms and the hydrogen atoms on them, and optimize other atoms. This is reasonable and unbiased because the protein backbone skeleton can be effectively maintained and the side chains can be fully optimized. In addition, the groups around the solvent-exposed histidine ligands (Figure 1), which are either side chain fragments or backbone peptide groups, are also fixed. This is necessary to maintain the microscopic structures similar to those in the X-ray structures. For the base form models, the Cu^{+} ions and the S atoms of the Met ligands were fixed in their X-ray coordinates. The reason is that if they are not fixed, the optimized Cu^{+}−S(Met) distances are either too short or too long. More discussions will be presented in Section 3.1. For the acid form models, the Cu^{+} ions and the S(Met) atoms were not fixed because it is expected that their coordinates will change significantly upon protonation. Based on the Homo-CPCM/B3LYP/6-31G* optimized structures, single point energies were calculated with the Het-CPCM/B3LYP/6-31G* and gas phase B3LYP/6-31G* methods.

Fern plastocyanin (Plc) is used as the reference for relative pK_a calculation. For a protein (Pro) the free energy change for the following proton exchange reaction

$$\mathrm{ProH} + \mathrm{Plc} \xrightarrow{\Delta G} \mathrm{Pro} + \mathrm{PlcH} \qquad (1)$$

is approximated with the electronic energies (including nuclear repulsion and solvation free energies from CPCM calculation) computed for the model molecules:

$$\Delta G \approx \Delta E^{ele} = \Delta E^{ele}_{\mathrm{Pro}} + \Delta E^{ele}_{\mathrm{PlcH}} - \Delta E^{ele}_{\mathrm{ProH}} - \Delta E^{ele}_{\mathrm{Plc}} \qquad (2)$$

The relative method is based on the assumption that the differences in the zero-point-energies, thermal energies, and entropies of the model molecules make minor contributions to the relative pK_a. Although not general, this is indeed a good approximation when a series of similar protein pK_a sites are considered. The free energy contributions due to the protein matrix and aqueous solvent are included in the CPCM solvation free energy.

The pK_a of the protein at $T = 298.15$ K is computed by

$$\mathrm{p}K_a = 4.40 + \frac{\Delta G}{RT\ln 10} \qquad (3)$$

where R is the gas constant, and 4.40 is the experimental pK_a of fern plastocyanin.

3. Results and Discussion

3.1. Base Form Structure. The Homo-CPCM/B3LYP/6-31G* optimized Cu^{+}-ligand distances in the model molecules are listed in Table 1. The corresponding distances in the PDB files are also listed for comparison. It is noted that the Cu^{+} ions and the S(Met) atoms are fixed in the X-ray coordinates.

(75) Rosen, D. *Trans. Faraday Soc.* **1963**, *59*, 2178.
(76) Bone, S.; Pethig, R. *J. Mol. Biol.* **1982**, *157*, 571–575.
(77) Bone, S.; Pethig, R. *J. Mol. Biol.* **1985**, *181*, 323–326.

Table 1. CPCM/B3LYP/6-31G* Optimized Cu^+-Ligand Distances (Å)

	$Cu-S_{Cys}$	$Cu-N1^a$	$Cu-N2^b$	$Cu-S_{Met}$	$Cu-O^c$
T.f. rusticyanin (2CAK), Cu^+, base	2.234	2.059	2.078	2.917	
base model	2.234	1.992	1.993	2.917	
acid unflipped model	2.211	1.921	3.729	2.225	
acid flipped model	2.183	1.931	5.360	2.297	
P.a. azurin (1E5Y), Cu^+, base	2.284	2.055	2.192	3.326	2.970
base model	2.280	1.974	1.982	3.326	2.892
acid unflipped model	2.148	1.896	3.059	3.775	2.457
acid flipped model	2.155	1.868	5.206	3.532	2.968
Fern plastocyanin (1KDI), Cu^+, base	2.207	1.949	2.103	2.914	
base model	2.231	1.977	1.981	2.914	
acid unflipped model	2.175	1.926	3.219	2.280	
acid flipped model	2.178	1.925	5.337	2.297	
Poplar plastocyanin (6PCY), Cu^+, acid flipped	2.127	2.115	5.166	2.509	
A.f. pseudoazurin (1PZA), Cu^+, base	2.169	2.162	2.287	2.909	
base model	2.223	1.946	2.033	2.909	
acid unflipped model	2.169	1.917	3.153	2.322	
acid flipped model	2.179	1.921	5.321	2.306	
A.f. pseudoazurin (1PZB), Cu^+, acid unflipped	2.157	2.190	3.089	2.423	
P.v. amicyanin (1ID2), Cu^{2+}, base	2.108	1.954	2.034	2.905	
base model	2.221	1.975	1.980	2.905	
acid unflipped model	2.217	1.919	3.935	2.199	
acid flipped model	2.180	1.912	5.327	2.354	
P.d. amicyanin (1BXA), Cu^+, acid flipped	2.093	1.910	5.451	2.901	

[a] N1 represents the His ligand with lower sequence number. [b] N2 represents the His ligand with higher sequence number, which is solvent-exposed. [c] Backbone carbonyl O ligand, only for azurin.

The optimized $Cu-S^-(Cys)$ distances are 2.22−2.28 Å in the base form model molecules, as compared to 2.11−2.28 Å in the X-ray structures and 2.19−2.22 Å from EXAFS studies;[62,78] the $Cu^+-N(His)$ bond lengths are 1.95−1.99 Å in the base form model molecules, as compared to 1.95−2.29 Å in the X-ray structures and 1.96−2.12 Å from EXAFS studies.[62,78] All of these X-ray structures are base forms (the solvent-exposed histidine unprotonated) and were determined for the reduced form (Cu^+) except for 1ID2, which is an oxidized form (Cu^{2+}), and has the shortest Cu−S(Cys) and Cu−N(His) distances of 2.11 and 1.95 Å. It is well-known that reduced and oxidized type-1 Cu centers show similar geometries, especially the Cu−S(Cys) and Cu−N(His) distances. These three bonds are relatively strong and rigid, and show small variations in different proteins and model molecules. Because the Cu^+ ions and the S atoms of the Met ligands are fixed in their X-ray coordinates, the Cu−S(Met) distances are exactly the same as in the X-ray structures. Due to the same reason, the overall structures of the optimized base form type-1 Cu centers are very similar to the X-ray structures.

If the Cu^+ ions and the S atoms of the Met ligands are not fixed, the optimized $Cu^+-S(Met)$ bond lengths show significant deviations from those in the X-ray structures. For T.f. rusticyanin (2CAK) and fern plastocyanin (1KDI), the optimized $Cu^+-S(Met)$ distances are ∼0.5 Å shorter than those in the X-ray structures. This has been noticed and discussed in the literature, for example, by Ryde et al.,[79] Solomon et al.,[78] and Ando et al.[80] For P.a. azurin (1E5Y), A.f. pseudoazurin (1PZA), and P.v. amicyanin (1ID2), if not fixed, the optimized $Cu^+-S(Met)$ distances are much longer than the experimental values. B3LYP calculations performed for reduced P.a. azurin model molecules by Corni et al.[81] and Ryde et al.[82] revealed a similar trend: the $Cu^+-S(Met)$ distance is significantly elongated and the Met ligand tends to leave the Cu^+ ion. It is well-known that if no protein interactions are considered, the optimum $Cu^+-S(CH_3)_2$ distance is ∼2.3 Å; it is the protein interactions that dictate the actual distances in different proteins or at different temperatures, as shown by Solomon et al.[83] Since in the current work, only some protein interactions are included by using model molecules with ∼140 atoms, it is better to fix the Cu and S atoms of the Met ligands in their X-ray coordinates.

3.2. Acid Form Structure. After protonation, the solvent-exposed histidine is no longer a Cu^+ ligand. The Cu^+ ion coordinates to only three ligands: S^-(Cys), N(His), and S(Met). The $Cu^+-S^-(Cys)$ and $Cu^+-N(His)$ distances in the optimized acid form model molecules are slightly shorter than the corresponding values in the base forms. For example, in the unflipped acid form of the fern plastocyanin (1KDI) model molecule, the $Cu^+-S^-(Cys)$ and $Cu^+-N(His)$ distances are 2.175 and 1.926 Å, shorter than those in the base forms by 0.056 and 0.051 Å, respectively (Table 1). The shortening of these bond lengths suggests that their strengths are increased. This trend is

(78) Solomon, E. I.; Szilagyi, R. K.; George, S. D.; Basumallick, L. *Chem. Rev.* **2004**, *104*, 419–458.
(79) Ryde, U.; Olsson, M. H. M.; Pierloot, K.; Roos, B. O. *J. Mol. Biol.* **1996**, *261*, 586–596.
(80) Ando, K. *J. Phys. Chem. B* **2004**, *108*, 3940–3946.
(81) Corni, S.; De Rienzo, F.; Di Felice, R.; Molinari, E. *Int. J. Quantum Chem.* **2005**, *102*, 328–342.
(82) De Kerpel, J. O. A.; Pierloot, K.; Ryde, U. *J. Phys. Chem. B* **1999**, *103*, 8375–8382.
(83) Ghosh, S.; Xie, X. J.; Dey, A.; Sun, Y.; Scholes, C. P.; Solomon, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4969–4974.

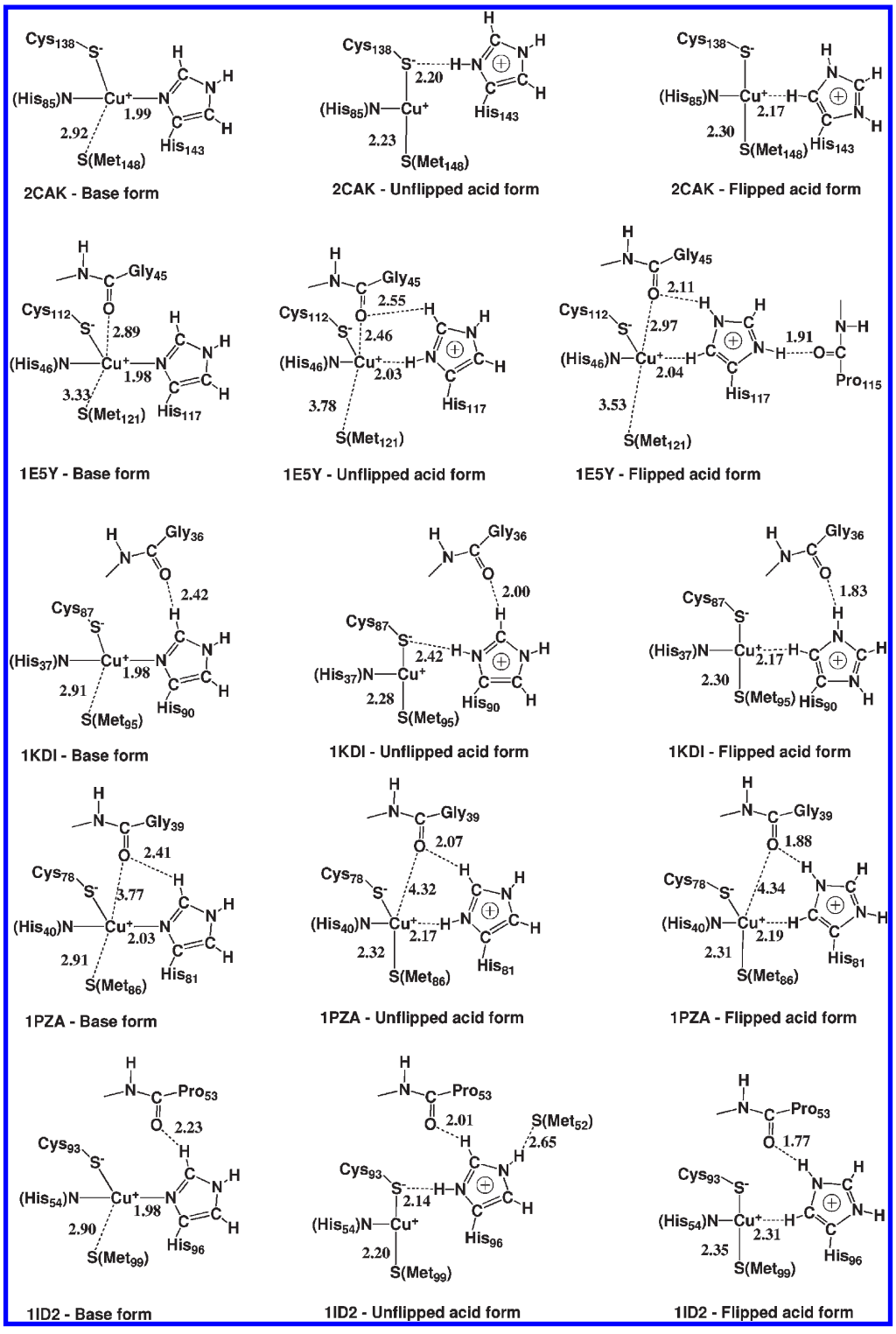

Figure 2. Bond lengths (Å) in the homogeneous CPCM/B3LYP/6-31G* optimized model molecules for T.f. rusticyanin (2CAK), P.a. azurin (1E5Y), fern plastocyanin (1KDI), A.f. pseudoazurin (1PZA), and P.v. amicyanin (1ID2).

not clear in the X-ray structures, probably due to experimental errors.

The Cu^+−S(Met) distances in the acid forms are correlated to the Cu^+−O(carbonyl) distances (Table 1 and Figure 2). For T.f. rusticyanin (2CAK), fern plastocyanin (1KDI), A.f. pseudoazurin (1PZA), and P.v. amicyanin (1ID2), which have no axial O(carbonyl) ligands, the Cu^+−S(Met) distances are 0.6−0.7 Å shorter than those in the base forms. Again, these suggest that the Cu^+−S(Met) interaction becomes slightly stronger due to the loss of a N(His) ligand. The optimized Cu^+−S(Met) distance in the unflipped acid form model of A.f. pseudoazurin is 2.322 Å, in good agreement with 2.423 Å in the corresponding X-ray structure 1PZB[26] of unflipped acid form A.f. pseudoazurin (Table 1). The optimized Cu^+−S(Met) distance in the flipped acid form model of

fern plastocyanin is 2.297 Å, comparable to 2.509 Å in the X-ray structure 6PCY[25] for flipped acid form poplar plastocyanin, a very similar protein (Table 1). The optimized $Cu^+-S(Met)$ distance in the flipped acid form model of P.v. amicyanin is 2.354 Å, much shorter than the 2.901 Å in the X-ray structure 1BXA[2] and the 2.88 Å obtained by Davidson et al.[2] for the flipped acid form P.d. amicyanin, a protein very similar to P.v. amicyanin. As discussed in Section 3.1, the $Cu^+-S(Met)$ distances in the base form proteins are dictated by protein interactions. This could also be true for the acid form proteins. The 1ID2[65] and 1BXA[2] structures suggest that upon protonation, the $Cu^+-S(Met)$ distances are almost unchanged (both ~2.90 Å). Therefore, to reproduce the experimental results, very large model molecules that include the second and third layer groups around the Cu^+ ion are necessary. Clearly, the short $Cu^+-S(Met)$ distance in the calculated structure will affect the relative energetics and lead to a higher relative pK_a value for P.v. amicyanin. The authors have compared the two sets of pK_a values calculated with the base form model molecules in that the Cu^+ and S(Met) are either fixed or optimized (see discussions in Section 3.1), and found that the pK_a values can be affected by at most ±1.4, a significant but not extraordinarily large value. Therefore, it is likely that the calculated pK_a of P.v. amicyanin will be at most overestimated by 1.4 due to this reason.

For P.a. azurin (1E5Y), which has a strong axial O(Gly45) ligand, the optimized $Cu^+-S(Met)$ distance shows some changes: base form 3.326 Å (as in the X-ray structure), unflipped acid form 3.775 Å, flipped acid form 3.532 Å. The optimized axial $Cu^+-O(Gly45)$ distance in the base form is 2.892 Å, slightly shorter than the 2.970 Å in the X-ray structure. The optimized Cu−-O(Gly45) distances are 2.457 and 2.968 Å, respectively, in the unflipped and flipped acid forms of the model molecules. This longer distance in the flipped form is due to the formation of a hydrogen bond between O(Gly45) and the $H-N^{\varepsilon 2}$(His), which becomes available after the flipping (Figure 2). Apparently, the competition from the hydrogen bond weakens the $Cu^+-O(Gly)$ bond.

Upon protonation but without flipping, the $Cu^+-N^{\delta 1}$(His) distances in the five model molecules vary from 3.1 to 3.9 Å. For example, it is 3.153 Å in the A.f. pseudoazurin model, in good agreement with the 3.089 Å in the X-ray structure 1PZB[26] for protonated (not flipped) A.f. pseudoazurin. In the flipped forms, all $Cu^+-N^{\delta 1}$(His) distances are around 5.3 Å (Table 1). In the X-ray structure 6PCY[25] and 1BXA[2] for histidine-protonated and flipped poplar plastocyanin and P.d. amicyanin, the $Cu^+-N^{\delta 1}$(His) distances are 5.166 and 5.451 Å, respectively. In the refinement of a crystal structure (2RAC) of P.d. amicyanin, Davidson et al. also obtained a similarly long $Cu^+-N^{\delta 1}$(His) distance, which can be estimated as ~6.0 Å according to the 3.92 Å $Cu^+-C^{\delta 2}$(His) distance reported in their paper.[2] Therefore, the Homo-CPCM/B3LYP/6-31G* optimized $Cu^+-N^{\delta 1}$(His) distance agrees well with experimental results.

In the optimized acid form model molecules of fern plastocyanin, A.f. pseudoazurin, and P.v. amicyanin, the $N^{\varepsilon 2}$ of the flipped histidines form hydrogen bonds with the backbone carbonyl oxygen atoms (Figure 2). When the His is flipped, the $O-N^{\varepsilon 2}$(His) distance in the fern plastocyanin model is 2.815 Å, in good agreement with the 3.101 Å in the X-ray structure 6PCY[25] for poplar plastocyanin. When the His is unflipped, the $O-C^{\varepsilon 1}$(His) distance in the A.f. pseudoazurin model (unflipped) is 2.960 Å, in good agreement with the 3.062 Å in the X-ray structure 1PZB.[26] These good agreements suggest that the Homo-CPCM/B3LYP/6-31G* method is reliable for the model molecules. However, the $O-N^{\varepsilon 2}$(His) distance in the P.v. amicyanin model is 2.685 Å, significantly different from the 4.284 Å in the X-ray structure 1BXA[2] for P.d. amicyanin. The difference is mainly caused by the relative position of the protonated His96: it is moved toward the carbonyl oxygen in the model molecule, but is moved toward the solvent in the crystal structure 1BXA. In the 1BXA file, there is a water molecule (O 878) between the backbone carbonyl O and $N^{\varepsilon 2}$(His), presumably forming hydrogen bonds with them. According to the calculation, it is likely that a direct hydrogen bond between the backbone carbonyl O and $N^{\varepsilon 2}$(His) will form because the water is unlikely a structural one.

3.3. Histidine Flipping. The relative energies of the imidazolium-flipped acid forms as compared to the unflipped forms are listed in Table 2. The solvation free energies calculated with CPCM methods are listed in Table 3. Clearly, solvation effects can affect the relative energies. For example, the flipped form of the T.f. rusticyanin model molecule (2CAK) gains ~5 kcal/mol more solvation free energy than the unflipped form (Table 3). In general, the Het-CPCM/B3LYP relative energies are similar to the Homo-CPCM/B3LYP ones. This is not surprising because a dielectric constant of 78.39 was used for the solvent-exposed histidine ligands for these model molecules. In the following discussions, only the Homo-CPCM/B3LYP results are used for simplicity.

For T.f. rusticyanin (2CAK model), the gas phase B3LYP energy of the flipped acid form is higher than the unflipped form by 2.76 kcal/mol, due to the breaking of the hydrogen bond between $H-N^{\delta 1}$(His143) and S^-(Cys138), as shown in Figure 2. However, according to the Homo-CPCM/B3LYP/6-31G* calculation, the His143 flipped form is likely to gain ~6 kcal/mol more solvation free energy (Table 3), implying that when the His143 is protonated, it will flip (if the protein is not denatured).

For P.a. azurin (1E5Y model), the gas phase B3LYP energy of the flipped acid form is lower than that of the unflipped form by 3.13 kcal/mol, due to the formation of two hydrogen bonds: one is between O(Gly45) and $H-N^{\varepsilon 2}$(His117), one is between O(Pro115) and the $H-N^{\delta 1}$(His117), as shown in Figure 2. However, according to the Homo-CPCM/B3LYP/6-31G* calculation, the flipped acid form is likely to lose ~4 kcal/mol solvation free energy (Table 3), thus becoming 1.15 kcal/mol less stable than the unflipped form. If this value is accurate, ~13% flipped form should be observed at ~300 K and very low pH (if the protein is not denatured).

For fern plastocyanin (1KDI model), the gas phase B3LYP energy of the flipped acid form is lower than the unflipped form by 0.45 kcal/mol, due to the formation of a hydrogen bond between O(Gly36) and $H-N^{\varepsilon 2}$(His90), as shown in Figure 2. According to the Homo-CPCM/B3LYP/6-31G* calculation, the solvation free energies of

Table 2. Relative B3LYP Energies (kcal/mol) of Imidazolium-Flipped Acid Forms As Compared to Unflipped Forms

			dielectric constants used in CPCM			
protein	PDB	gas phase	78.39 and 4[a]	78.39 and 10[b]	78.39[c]	exp.
T.f. rusticyanin	2CAK	+2.76	−2.02	−2.62	−2.97	
P.a. azurin	1E5Y	−3.13	+1.88	+1.41	+1.15	
fern plastocyanin	1KDI	−0.45	−1.15	−0.88	−0.73	
A.f. pseudoazurin	1PZA	−1.52	+0.25	+0.54	+0.71	
P.v. amicyanin	1ID2	+1.84	+0.57	+0.83	+0.98	+1.0

[a] ε = 78.39 used for the concerned His groups, ε = 4 for the remaining atoms. [b] ε = 78.39 used for the concerned His groups, ε = 10 for the remaining atoms. [c] ε = 78.39 used for all the atoms.

Table 3. CPCM Computed Solvation Free Energies (kcal/mol) for the Model Molecules (B = Base Form; A-u = Acid Unflipped; A-f = Acid Flipped)

		dielectric constants used in CPCM								
		ε = 78.39 and 4[a]			ε = 78.39 and 10[b]			ε = 78.39[c]		
		B	A-u	A-f	B	A-u	A-f	B	A-u	A-f
T.f. rusticyanin	2CAK	−37.90	−61.55	−66.33	−45.60	−71.47	−76.85	−50.27	−77.43	−83.17
P.a. azurin	1E5Y	−42.38	−68.17	−63.16	−51.13	−80.09	−75.55	−56.43	−87.25	−82.97
fern plastocyanin	1KDI	−43.59	−66.12	−66.82	−52.60	−77.17	−77.60	−58.06	−83.81	−84.09
A.f. pseudoazurin	1PZA	−46.79	−66.65	−64.88	−56.86	−78.08	−76.02	−62.99	−84.96	−82.72
P.v. amicyanin	1ID2	−40.49	−57.44	−58.71	−49.44	−68.36	−69.44	−54.88	−74.93	−75.79

[a] ε = 78.39 used for the concerned His groups, ε = 4 for the remaining atoms. [b] ε = 78.39 used for the concerned His groups, ε = 10 for the remaining atoms. [c] ε = 78.39 used for all the atoms.

these two acid forms are very similar (−83.81 versus −84.09 kcal/mol), with the flipped acid form being likely to gain ~0.2 kcal/mol more solvation free energy (Table 3). Therefore, upon protonation, flipped His90 is favored.

For A.f. pseudoazurin (1PZA model), the gas phase B3LYP energy of the flipped acid form is lower than the unflipped form by 1.52 kcal/mol, due to the formation of a hydrogen bond between O(Gly39) and H−$N^{\varepsilon 2}$(His81), as shown in Figure 2. According to the Homo-CPCM/B3LYP/6-31G* calculation, the flipped acid form is likely to lose ~2 kcal/mol solvation free energy (Table 3), thus becoming 0.71 kcal/mol less stable than the unflipped form. This still makes the flipped form observable. Although solid-state interactions and solution-phase interactions are different, the general trends should be similar. Indeed, in the crystalline state of the protonated and reduced (Cu^+) form of A.f. pseudoazurin, the protonated histidine does not flip by ~180 degrees. Instead, it "rotates away" from the Cu^+, as shown by the X-ray structure 1PZB.[26] As already discussed in Section 3.2, the optimized Cu^+−S(Met) distance in the acid form A.f. pseudoazurin model is 2.322 Å, in good agreement with the 2.423 Å seen in the X-ray structure 1PZB (Table 1).

For P.v. amicyanin (1ID2 model), the gas phase B3LYP energy of the flipped acid form is higher than the unflipped form by 1.84 kcal/mol, due to the breaking of two hydrogen bonds: one between H−$N^{\delta 1}$(His96) and S^-(Cys93), one between H−$N^{\varepsilon 2}$(His96) and S(Met52), as shown in Figure 2. A new hydrogen bond between H−$N^{\varepsilon 2}$(His96) and O(Pro53) is formed after flipping, and is likely to compensate some of the energy penalty in breaking two hydrogen bonds. In addition, the flipped acid form is likely to gain ~1 kcal/mol more solvation free energy (Table 3), thus only 0.98 kcal/mol less stable than the unflipped form. This value agrees very well with the experiments performed by Canters and co-workers, who found that for P.v. amicyanin, the ratio of the flipped and unflipped forms in a 297 K solution is ~0.2, which corresponds to a ΔG_{flip} = 1.0 kcal/mol.[24]

3.4. pK_a Determinants. Because there are two possible acid forms for each protein, the relative pK_a should be calculated using the statistically averaged free energy of the acid form for each protein.[49] If the two acid forms have the same free energy, the computed pK_a should be corrected by +0.3 (i.e., log2) due to the lowering in the free energy of the acid form. Therefore, using the lower-energy acid form will at most introduce an error of −0.3 pH units, which is insignificant considering the errors pertaining to the computational methods used in the current study. Using only the lower-energy acid form allows for simpler discussions, and is used here.

For comparison, the pK_a values calculated using either all unflipped acid forms or all flipped acid forms are also listed in Table 4. Such a comparison is meaningful because there are always uncertainties in the calculated relative stabilities of the two acid forms. It is certainly possible that the method used in the current study leads to a few kcal/mol errors. Usually, the errors in the relative energies are smaller if the structures of the molecules are more similar (i.e., all unflipped or all flipped), so the pK_a values calculated using all unflipped or all flipped acid forms are "more reliable" if the errors are of concern. Using all unflipped acid forms, the gas phase relative pK_a values are 2.0, −1.1, 4.4, 9.9, and 11.8, respectively, for T.f. rusticyanin, P.a. azurin, fern plastocyanin (reference), A.f. pseudoazurin, and P.v. amicyanin. Using all flipped acid forms, the values are −0.4, 0.9, 4.4, 10.6, and 10.1, respectively. Using the lower-energy acid form for each model molecule, the values are 1.6, 0.9, 4.4, 10.6, and 11.5, respectively. All these values are similar to each other, implying that there are no significant errors that are larger than 2 pK_a units.

An important piece of information provided by the gas phase pK_a values in Table 4 is that the experimental

Table 4. B3LYP Computed Relative pK_a of the Solvent-Exposed Histidine Ligands of Type-1 Cu Centers

			dielectric constants used in CPCM				
protein	PDB	gas phase	78.39 and 4[b]	78.39 and 10[c]	78.39[d]	mix[e]	exp[a]
			use all unflipped acid form				
T.f. rusticyanin	2CAK	2.0	2.8	2.9	3.0	1.3	< 2
P.a. azurin	1E5Y	−1.1	1.3	2.2	2.7	2.2	< 2
fern plastocyanin	1KDI	4.4	4.4	4.4	4.4	4.4	4.4
A.f. pseudoazurin	1PZA	9.9	7.9	7.4	7.1	7.4	4.8
P.v. amicyanin	1ID2	11.8	7.7	7.6	7.6	7.6	7.2
			use all flipped acid form				
T.f. rusticyanin	2CAK	−0.4	3.4	4.2	4.6	2.1	< 2
P.a. azurin	1E5Y	0.9	−0.9	0.5	1.3	0.5	< 2
fern plastocyanin	1KDI	4.4	4.4	4.4	4.4	4.4	4.4
A.f. pseudoazurin	1PZA	10.6	6.9	6.4	6.0	6.4	4.8
P.v. amicyanin	1ID2	10.1	6.4	6.4	6.4	6.4	7.2
			use lower-energy acid form				
T.f. rusticyanin	2CAK	1.6	3.4	4.2	4.6	2.1	< 2
P.a. azurin	1E5Y	0.9	0.5	1.5	2.1	1.5	< 2
fern plastocyanin	1KDI	4.4	4.4	4.4	4.4	4.4	4.4
A.f. pseudoazurin	1PZA	10.6	7.1	6.7	6.5	6.6	4.8
P.v. amicyanin	1ID2	11.5	6.8	7.0	7.1	7.0	7.2

[a] The actual pK_a values are not known for T.f. rusticyanin and P.a. azurin. [b] $\varepsilon = 78.39$ used for the concerned His groups, $\varepsilon = 4$ for the remaining atoms. [c] $\varepsilon = 78.39$ used for the concerned His groups, $\varepsilon = 10$ for the remaining atoms. [d] $\varepsilon = 78.39$ used for all the atoms. [e] $\varepsilon = 78.39$ used for the concerned His groups; for the remaining atoms in T.f. rusticyanin, $\varepsilon = 4$ is used; for the remaining atoms in the other four proteins, $\varepsilon = 10$ is used.

relative pK_a values and the gas phase B3LYP computed values show the same trend, except that the gas phase values span a range of ~10 pK_a units, much wider than the experimental range of ~5 pK_a units for the five proteins. This has almost already explained why the solvent-exposed histidine ligands in these type-1 Cu proteins show such relative pK_a values in solution. Clearly, the differences in the gas-phase pK_a values are caused by the local interactions contained in the model molecules. It is desirable to establish correlations between the model structures and the relative gas-phase pK_a values. For example, it is often to correlate the pK_a shifts to the number of hydrogen bonds on the pK_a sites.[40] However, the gas-phase pK_a values (Table 4) of the five model molecules cannot be simply correlated to the number of the hydrogen bonds formed by the considered histidine in the model molecules studied in the current work. For example, in the P.a. azurin model molecule the concerned histidine forms more hydrogen bonds than that in the T.f. rusticyanin model molecule (Figure 2), but they show almost the same gas-phase pK_a value. Therefore, without a quantitative interaction analysis, no apparent correlations between the pK_a values and specific interactions can be established. However, as discussed in the following, the total interactions contained in the model molecules can be analyzed.

The Cu−His coordination and hydrogen bonding of the histidine in the model molecules (Figure 2) can significantly change the histidine deprotonation energy. For comparison, the gas-phase deprotonation energy (ΔE^{gas}, only the electronic energy and nuclear repulsion are considered) for 4-methy-imidazolium was calculated at the B3LYP/6-31G* level of theory using the Homo-CPCM/B3LYP/6-31G* optimized geometry, with a result of 240.93 kcal/mol. The ΔE^{gas} for the P.v. amicyanin (using unflipped acid form 1ID2 model in Figure 2) calculated at the same level of theory is 274.38 kcal/mol, higher by 33.45 kcal/mol. Therefore, the interactions in the 146-atom P.v. amicyanin (1ID2) model molecule preferentially stabilize the unflipped acid form by 33.45 kcal/mol. These interactions include the hydrogen bonds to the histidine and the Cu-ligand interactions. The charge−charge interaction is unlikely a major factor because the positive charge of the Cu^+ is neutralized by the negative charge of the S^-(Cys). Compared to the P.v. amicyanin model, other model molecules show smaller ΔE^{gas} (see gas-phase pK_a in Table 4), indicating that they provide less stabilization for the acid forms. For example, the ΔE^{gas} for the T.f. rusticyanin 2CAK model is 261.33 kcal/mol, 13.05 kcal/mol smaller than the value for the P.v. amicyanin model, but still 20.4 kcal/mol larger than that for 4-methyl-imidazoloium. Recently, Dennison et al. found that $\pi-\pi$ interactions in fern plastocyanin can affect the pK_a value of its solvent-exposed histidine.[84] Unfortunately, because the B3LYP method used in the current study does not contain long-range dispersion interactions, such an effect cannot be analyzed.

Solvation can significantly affect the relative pK_a values. Since all the structures were optimized using the Homo-CPCM/B3LYP/6-31G* method with $\varepsilon = 78.39$, the first set of relative pK_a values was obtained with this method. As shown in Table 4, if all unflipped acid forms are used, the computed pK_a values are 3.0, 2.7, 4.4, 7.1, and 7.6, respectively, for T.f. rusticyanin, P.a. azurin, fern plastocyanin (reference), A.f. pseudoazurin, and P.v. amicyanin. If all flipped acid forms are used, the values become 4.6, 1.3, 4.4, 6.0, and 6.4. If lower-energy

(84) Yanagisawa, S.; Crowley, P. B.; Firbank, S. J.; Lawler, A. T.; Hunter, D. M.; McFarlane, W.; Li, C.; Kohzuma, T.; Banfield, M. J.; Dennison, C. *J. Am. Chem. Soc.* **2008**, *130*, 15420–15428.

acid forms are used, the values are 4.6, 2.1, 4.4, 6.4, and 7.1, respectively. Compared to the gas-phase results, these values are much closer to the experimental values, which are <2, <2, 4.4, 4.8, and 7.2, respectively.

Using the Het-CPCM method, the protein matrix solvation of these model molecules can be better described because different effective dielectric constants can be used. Although the solvation free energies change significantly for each individual model molecule (Table 3), using $\varepsilon = 4$ and 10 for protein interior for all the cases leads to relative pK_a values that are similar to those from using $\varepsilon = 78.39$ (Table 4), mainly due to the fact that $\varepsilon = 78.39$ has been used for the solvent-exposed histidine ligands in all the model molecules. Using the lower-energy acid forms and $\varepsilon = 4$ for the protein interior the relative pK_a values are 3.4, 0.5, 4.4, 7.1, and 6.8, respectively, for T.f. rusticyanin, P.a. azurin, fern plastocyanin (reference), A.f. pseudoazurin, and P.v. amicyanin. Using the lower-energy acid forms and $\varepsilon = 10$ the relative pK_a values are 4.2, 1.5, 4.4, 6.7, and 7.0, respectively. In all of the three cases, large solvation effects of $\sim$4 pK_a units are seen for A.f. pseudoazurin and T.v. amicyanin.

The relative pK_a value computed for T.f. rusticyanin is too high. A heteronuclear (^{1}H, ^{15}N) nuclear magnetic resonance (NMR) study suggests that the protein matrix around the rusticyanin (2CAK) type-1 Cu center is highly hydrophobic and rigid, which corresponds to a low effective dielectric constant.[85] Using the lower-energy acid forms, $\varepsilon = 4$ for the protein interior of T.f. rusticyanin, $\varepsilon = 10$ for the protein interior of other proteins, and $\varepsilon = 78.39$ for the solvent-exposed histidine ligands, the relative pK_a values are 2.1, 1.5, 4.4, 6.6, and 7.0 for the five proteins, compared very well to the experimental values <2, <2, 4.4, 4.8, and 7.2 (fern plastocyanin value 4.4 is the reference).

(85) Jimenez, B.; Piccioli, M.; Moratal, J. M.; Donaire, A. *Biochemistry* **2003**, *42*, 10396–10405.

4. Conclusion

To conclude, the following points are highlighted:

(1) CPCM/B3LYP/6-31G* calculations with model molecules consisting of $\sim$140 atoms revealed that upon protonation, the solvent-exposed histidine ligands of the reduced type-1 Cu centers in T.f. rusticyanin, P.a. azurin, fern plastocyanin, A.f. pseudoazurin, and P.v. amicyanin will decoordinate from the Cu^+ ion.

(2) For the five studied proteins, the two conformations of the protonated histidines, unflipped and flipped, show similar energies, consistent with the experimental observation on spinach plastocyanin and P.v. amicyanin in solution (Table 2).

(3) Using the Het-CPCM/B3LYP/6-31G* method and fern plastocyanin as the reference (pK_a = 4.4), the calculated pK_a values are 2.1, 1.5, 6.6, and 7.0 (Table 4), comparable to experimental values <2, <2, 4.8, and 7.2, respectively, for the solvent-exposed histidine ligands of the type-1 Cu centers in T.f. rusticyanin, P.a. azurin, A.f. pseudoazurin, and P.v. amicyanin.

(4) The main determinants of the solvent-exposed histidine pK_a values are local interactions contained in the model molecules, which can create pK_a differences up to $\sim$10 units. Aqueous solvation of the histidine ligands and protein matrix polarization can affect the relative pK_a values by as large as 4−5 units (Table 4). These interactions can well explain the observed 5−6 pK_a range for the five proteins.

Acknowledgment. This work was supported by startup funds from the University of Nebraska-Lincoln.