Inorg. Chem. **2007**, *46*, 4492–4503

Comprehensive Molecular Mechanics Model for Oxidized Type I Copper Proteins: Active Site Structures, Strain Energies, and Entatic Bulging

Robert J. Deeth*

Inorganic Computational Chemistry Group, Department of Chemistry, University of Warwick, Coventry CV4 7AL U.K.

Received December 15, 2006

The ligand field molecular mechanics (LFMM) model has been applied to the oxidized Type 1 copper center. In conjunction with the AMBER94 force field implemented in DommiMOE, the ligand field extension of the molecular operating environment (MOE), LFMM parameters for Cu–N(imidazole), Cu–S(thiolate), Cu–S(thioether), and Cu–O(carbonyl) interactions were developed on the basis of experimental and theoretical data for homoleptic model systems. Subsequent LFMM optimizations of the active site model complex $[Cu(imidazole)_2(SMe)(SMe_2]^+$ agree with high level quantum results both structurally and energetically. Stable trigonal and tetragonal structures are located with the latter about 1.5 kcal mol^{-1} lower in energy. Fully optimized unconstrained structures were computed for 24 complete proteins containing T1 centers spanning four-coordinate, plastocyanin-like CuN_2SS' and stellacyanin-like CuN_2SO sites, plus the five-coordinate $CuN_2SS'O$ sites of the azurins. The initial structures were based on PDB coordinates augmented by a 10 Å layer of water molecules. Agreement between theory and experiment is well within the experimental uncertainties. Moreover, the LFMM results for plastocyanin (Pc), cucumber basic protein (CBP) and azurin (Az) are at least as good as previously reported QM/MM structures and are achieved several orders of magnitude faster. The LFMM calculations suggest the protein provides an entatic strain of about 10 kcal mol^{-1}. However, when combined with the intrinsic 'plasticity' of d^9 Cu(II), different starting protein/solvent configurations can have a significant effect on the final optimized structure. This 'entatic bulging' results in relatively large fluctuations in the calculated metal–ligand bond lengths. For example, simply on the basis of 25 different starting configurations of the solvent molecules, the optimized Cu–S(thiolate) bond lengths in Pc vary by 0.04 Å while the Cu–S(thioether) distance spans over 0.3 Å. These variations are the same order of magnitude as the differences often quoted to correlate the spectroscopic properties from a set of proteins. Isolated optimizations starting from PDB coordinates (or indeed, the PDB structures themselves) may only accidentally correlate with spectroscopic measurements. The present calculations support the work of Warshel who contends that adequate configurational averaging is necessary to make proper contact with experimental properties measured in solution. The LFMM is both sufficiently accurate and fast to make this practical.

Introduction

Copper is widespread in biological systems.[1,2] It mediates a variety of processes from oxygen transport to electron transfer (ET), paralleling comparable processes involving iron.[2] The relative accessibility of the $Cu^{I/II}$ redox potential gives copper a particularly important role in ET, especially for the so-called Type 1 (T1) site.

The T1 or 'blue copper' site generally has the metal center strongly coordinated by two histidine (HIS) residues and a thiolate cysteine (CYS) plus a long interaction with a methionine (MET) group (Figure 1). As found in plastocyanin (Pc), for example, the geometry is described as trigonal with the HIS and CYS ligands forming the basal plane and the MET in a distant axial position. The latter can vary. In some cases such as stellacyanin (STC), it is replaced by the oxygen of a glutamine (GLU) residue, while in other cases like azurin (Az), it is augmented by the oxygen of a glycine

* E-mail: r.j.deeth@warwick.ac.uk. Internet: http://www.warwick.ac.uk/go/iccg.

(1) Karlin, K. D.; Tyeklar, Z. *Bioinorganic Chemistry of Copper*; Chapman and Hall: New York, 1993.
(2) Frausto da Silva, J. J. R.; Williams, R. J. P. *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*; Clarendon Press: Oxford, 1991.

10.1021/ic062399j CCC: $37.00
Published on Web 04/27/2007

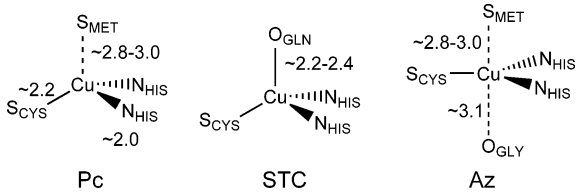

Figure 1. Schematic representation of oxidized Type I copper centers.

(GLY) amide bond to generate a five-coordinate site. The HIS−CYS−MET triad is generally located in a C-terminal loop of the protein, while the other HIS comes from the β-sheet backbone.[3−6]

The T1 site presents a number of unusual spectroscopic features such as the intense charge-transfer band around 600 nm and the small copper hyperfine coupling in the $g_{||}$ region of the EPR spectrum. Extensive experimental and theoretical studies,[7−19] especially from the Solomon group,[4,5,10,13,16−18,20] have generated a more or less complete understanding of these spectroscopic properties, both of which are strongly linked to the strong Cu−S_{CYS} thiolate bond.[16] These and other features of T1 sites are comprehensively discussed in the reviews by Solomon et al.[3,16,17] and Dennison.[6]

A fascinating and controversial aspect of T1 copper centers is the concept of the 'entatic'[21] (or induced rack[22]) state. Prior to the single-crystal X-ray study of poplar plastocyanin,[23] the debate had raged concerning both the number and type of donor groups and their spatial arrangement. The observation of a distorted tetrahedral CuN_2SS' in its oxidized Cu(II) form became the archetypal exemplar of the entatic state—the protein enforces a structure midway between the separate but competing demands of the reduced and oxidized forms, thus minimizing the internal reorganization energy and facilitating efficient ET. This view held sway for two decades until Ryde, in an attempt to quantify the entatic state energy via quantum mechanical (QM) calculations, proposed that the oxidized T1 copper center was not strained.[15] This somewhat controversial conclusion prompted further hot debate with both experimental and computational arguments presented for both sides.[24,25]

What has never been disputed is that T1 sites present some very unusual properties relative to typical Cu(II) complexes. Indeed, small-molecule structural analogues remained an elusive goal for biomimetic chemists for many years,[26] although more recently, model complexes like [Cu(SCPh$_3$)(HB(3,5-iPr$_2$pz)$_3$)] (HB(3,5-iPr$_2$pz)$_3$ = tris(3,5-diisopropylpyrazolyl)hydroborate), while providing a (N_2S + weak axial N) donor set instead of (N_2S + weak axial thioether S) and where the Cu−N bond which corresponds to the thioether donor of Pc has a restricted ability to elongate due to the tripod tris-pyrazolylborate ligand, nevertheless provide a reasonable model of the structural and spectroscopic properties of a true T1 site. It is perhaps because of their unusual properties that proteins containing T1 centers, mutated variants and related systems, continue to attract significant interest.[6,7,17,18,27−31]

Computational studies of these systems are dominated by QM calculations either of models of the active site such as [Cu(imid)$_2$(SMe)(DMS)]$^{+1/0}$ (imid = imidazole, SMe = methyl thiolate, DMS = dimethylsulphide) or QM/MM studies where this QM model is inserted into a molecular mechanics (MM) treatment of the rest of the protein.[4,8,9,11−13,19,28,32−37] In either case, relative to pure MM, the QM calculations are relatively compute intensive and the potentially vital issue of configurational averaging is seldom addressed. Consequently, QM/MM applications are

(3) Solomon, E. I.; Szilagyi, R. K.; George, S. D.; Basumallick, L. *Chem. Rev.* **2004**, *104*, 419−458.
(4) Szilagyi, R. K.; Solomon, E. I. *Curr. Opin. Chem. Biol.* **2002**, *6*, 250−258.
(5) Solomon, E. I.; Randall, D. W.; Glaser, T. *Coord. Chem. Rev.* **2000**, *200*, 595−632.
(6) Dennison, C. *Coord. Chem. Rev.* **2005**, *249*, 3025−3054.
(7) Yanagisawa, S.; Dennison, C. *J. Am. Chem. Soc.* **2004**, *126*, 15711−15719.
(8) Olsson, M. H. M.; Hong, G. Y.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025−5039.
(9) Ryde, U.; Olsson, M. H. M. *Int. J. Quant. Chem.* **2001**, *81*, 335−347.
(10) Randall, D. W.; George, S. D.; Holland, P. L.; Hedman, B.; Hodgson, K. O.; Tolman, W. B.; Solomon, E. I. *J. Am. Chem. Soc.* **2000**, *122*, 11632−11648.
(11) Olsson, M. H. M.; Ryde, U. *J. Biol. Inorg. Chem.* **1999**, *4*, 654−663.
(12) Pierloot, K.; De Kerpel, J. O. A.; Ryde, U.; Olsson, M. H. M.; Roos, B. O. *J. Am. Chem. Soc.* **1998**, *120*, 13156−13166.
(13) LaCroix, L. B.; Randall, D. W.; Nersissian, A. M.; Hoitink, C. W. G.; Canters, G. W.; Valentine, J. S.; Solomon, E. I. *J. Am. Chem. Soc.* **1998**, *120*, 9621−9631.
(14) De Kerpel, J. O. A.; Pierloot, K.; Ryde, U.; Roos, B. O. *J. Phys. Chem. B* **1998**, *102*, 4638−4647.
(15) Ryde, U.; Olsson, M. H. M.; Pierloot, K.; Roos, B. O. *J. Mol. Biol.* **1996**, *261*, 586−596.
(16) Solomon, E. I.; Gorelsky, S. I.; Dey, A. *J. Comp. Chem.* **2006**, *27*, 1415−1428.
(17) Solomon, E. I. *Inorg. Chem.* **2006**, *45*, 8012−8025.
(18) Hansen, D. F.; Gorelsky, S. I.; Sarangi, R.; Hodgson, K. O.; Hedman, B.; Christensen, H. E. M.; Solomon, E. I.; Led, J. J. *J. Biol. Inorg. Chem.* **2006**, *11*, 277−285.
(19) Sinnecker, S.; Neese, F. *J. Comp. Chem.* **2006**, *27*, 1463−1475.
(20) Randall, D. W.; Solomon, E. I. *J. Inorg. Biochem.* **1999**, *74*, 272−272.
(21) Vallee, B. L.; Williams, R. J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1968**, *59*, 498−505.
(22) Malmstrom, B. G. *Eur. J. Biochem.* **1994**, *223*, 711−718.
(23) Colman, P. M.; Freeman, H. C.; Guss, J. M.; Murata, M.; Norris, V. A.; Ramshaw, J. A. M.; Venkatappa, M. P. *Nature* **1978**, *272*, 319−324.
(24) Gray, H. B.; Malmstrom, B. G.; Williams, R. J. P. *J. Biol. Inorg. Chem.* **2000**, *5*, 551−559.
(25) Ryde, U.; Olsson, M. H. M.; Roos, B. O.; De Kerpel, J. O. A.; Pierloot, K. *J. Biol. Inorg. Chem.* **2000**, *5*, 565−574.
(26) Tolman, W. B. *J. Biol. Inorg. Chem.* **2006**, *11*, 261−271.
(27) Basumallick, L.; Sarangi, R.; George, S. D.; Elmore, B.; Hooper, A. B.; Hedman, B.; Hodgson, K. O.; Solomon, E. I. *J. Am. Chem. Soc.* **2005**, *127*, 3531−3544.
(28) 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.
(29) Dennison, C. *Dalton Trans.* **2005**, *3436*−3442.
(30) Koch, M.; Velarde, M.; Harrison, M. D.; Echt, S.; Fischer, M.; Messerschmidt, A.; Dennison, C. *J. Am. Chem. Soc.* **2005**, *127*, 158−166.
(31) Dennison, C.; Harrison, M. D. *J. Am. Chem. Soc.* **2004**, *126*, 2481−2489.
(32) Sigfridsson, E.; Olsson, M. H. M.; Ryde, U. *J. Phys. Chem. B* **2001**, *105*, 5546−5552.
(33) Ryde, U.; Olsson, M. H. M.; Roos, B. O.; Borin, A. C. *Theor. Chem. Acc.* **2001**, *105*, 452−462.
(34) Solomon, E. I.; LaCroix, L. B.; Randall, D. W. *Pure Appl. Chem.* **1998**, *70*, 799−808.
(35) Olsson, M. H. M.; Ryde, U.; Roos, B. O.; Pierloot, K. *J. Biol. Inorg. Chem.* **1998**, *3*, 109−125.
(36) Olsson, M. H. M.; Ryde, U.; Roos, B. O. *Protein Sci.* **1998**, *7*, 2659−2668.
(37) Comba, P.; Lledos, A.; Maseras, F.; Remenyi, R. *Inorg. Chim. Acta* **2001**, *324*, 21−26.

typically restricted to single geometry optimizations starting from the PDB coordinates despite the demonstration by Olsson, Hong, and Warshel[8] that proper configurational sampling is critical for a reliable estimation of the redox potential in T1 systems. Importantly, the QM/MM optimization is started after surrounding the protein by solvent molecules and then usually performing some kind of MD protocol designed to generate a reasonable starting point representative of the 'true' structure. The effect of this starting point is not always assessed, and even if the starting point is a good one, making proper contact with, say, redox potentials which are measured in solution, requires fuller sampling of configurational space than just one point. What appears to be needed, therefore, is a method with the speed of MM and the accuracy of QM. The latter has generally been considered mandatory for electronically challenging metal centers like d^9 Cu(II), but work in this laboratory has shown that the 'missing ingredient' in conventional MM is the d-electron stabilization energy.[38] Once this is incorporated into MM, in our case via a generalized ligand field theory (LFT) calculation, the physical basis of the model is sufficient to describe accurately even very subtle features of Cu(II) complexes like the Jahn−Teller effect.[39] This paper thus addresses whether the success enjoyed by ligand field molecular mechanics (LFMM[40,41]) for treating simple coordination complexes of Cu(II) can be transferred to the more complicated T1 active site. A preliminary report dealing with five examples of Pc-like sites has already been published.[42] Here, the details are described more fully together with the extension to CuN_2SO and $CuN_2SS'O$ sites.

Computational Details

DFT Calculations. All DFT calculations employed the Amsterdam Density Functional program version 2005.01.[43] Spin-unrestricted geometry optimizations used triple-ζ plus polarization (TZP) STO basis sets on Cu and sulfur and double-ζ plus polarization (DZP) bases on all other atoms. The frozen core approximation[44] was used throughout with up to 2p orbitals frozen for Cu and S and the 1s cores frozen for C, N, and O. The PW91 functional was used with default SCF and geometry convergence criteria.

DommiMOE Calculations. All LFMM calculations employed the DommiMOE program[40] as implemented within MOE 2005.06.[45] The bulk of the FF parameters employed here are based on the AMBER94 (koll94.ff as distributed in MOE). These parameters were augmented by LFMM values for Cu−N, Cu−SR, Cu−SR_2, and Cu−O=CR_2 as given in the Supporting Information. The procedure for deriving parameter values is described later.

Protein Modeling. The typical protocol for preparing each protein for calculation was to load the PDB file into MOE and then to ensure that the reported sequence was matched by actual side-chain coordinates. Missing atoms were replaced using MOE's mutate function and hydrogens were then added. The charge states for ionizable groups were assumed to be −1 for carboxylic acids, +1 for amines, and neutral for histidine residues. Standard AMBER charges were then computed for all atoms barring the active site metal and coordinated residues. The partial atomic charges reported by Comba and Remenyi[46] were used for the active site atoms apart from their reported copper charge of 0.329, which yields an overall charge for a $[Cu(HIS)_2(CYS)(MET)]$ unit of 0.448 instead of the required +1. The copper charge was therefore set to ensure an integer charge for the active site which, for the present T1 systems, gives $\rho(Cu) = 0.777$. The whole protein was further checked to ensure it carried an integral charge overall.

Initially, any solvent molecules reported in the PDB file were retained, but this was subsequently found to generate occasionally extreme differences between X-ray and LFMM active site structures, suggesting that for this particular class of proteins, water molecules apparently located in the experimental X-ray diffraction studies can be unreliable.

The best results were obtained by allowing MOE to solvate the protein automatically. In order to obtain a reasonable starting arrangement of water molecules, the protein was 'soaked' to a depth of 5 Å, which corresponded to 600−800 solvent molecules. The protein atoms were fixed, and the structure of the water layer optimized to an rmsd gradient of 0.05 kcal mol^{-1} $Å^{-1}$. Up to 20 ps of constrained MD was then carried out (fixed H−O−H structure) and the final configuration minimized again to a gradient of 0.05 kcal mol^{-1} $Å^{-1}$. A further 5 Å layer of water molecules was then added giving around 1500−1800 water molecules and minimized to 0.05 kcal mol^{-1} $Å^{-1}$. Finally, the protein atoms were unfixed, and the entire system minimized by LFMM to a gradient of 0.01 kcal mol^{-1} $Å^{-1}$. To gauge the effect of the second layer of water molecules, full LFMM optimizations were also carried out with just the first 5 Å water layer present. The bond lengths to equatorial donors changed on average by 0.025 Å. The difference between the two Cu−N_{HIS} bonds is smaller for 10 Å solvation layer while the average Cu−S_{CYS} distance increases. Conceivably, still more water molecules would have an additional effect, but since the internal nonbonded cutoffs were set to 8−10 Å, any change is estimated to be small. The complete list of observed and computed metal−ligand bond distances and angles is included in the Supporting Information.

This protocol represents a compromise between that employed by Barrett et al.[28] who solvate the protein, relax the whole system via MM, and then carry out QM/MM calculations with active site residues optimized but the rest of the protein and the water molecules fixed, i.e., do not use MD, and that used by Sinnecker and Neese[19] who carry out a more complicated procedure involving MD simulations of the solvent and protein, with the active site fixed, prior to subsequent geometry optimization. To a large extent, the protocol depends on the desired property. If one wishes to make contact with data measured in solution, redox potentials for example, then as described by Olsson et al.[8] proper configurational averaging

(38) Burton, V. J.; Deeth, R. J.; Kemp, C. M.; Gilbert, P. J. *J. Am. Chem. Soc.* **1995**, *117*, 8407−8415.
(39) Deeth, R. J.; Hearnshaw, L. J. A. *Dalton Trans.* **2006**, 1092−1100.
(40) Deeth, R. J.; Fey, N.; Williams-Hubbard, B. J. *J. Comp. Chem.* **2005**, *26*, 123−130.
(41) Deeth, R. J. *Coord. Chem. Rev.* **2001**, *212*, 11−34.
(42) Deeth, R. J. *Chem. Commun.* **2006**, 2551−2553.
(43) Baerends, E. J.; Bérces, A.; Bo, C.; Boerrigter, P. M.; Cavallo, L.; Deng, L.; Dickson, R. M.; Ellis, D. E.; Fan, L.; Fischer, T. H.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Groeneveld, J. A.; Gritsenko, O. V.; Harris, F. E.; van den Hoek, P.; Jacobsen, H.; van Kessel, G.; Kootstra, F.; van Lenthe, E.; Osinga, V. P.; Philipsen, P. H. T.; Post, D.; Pye, C. C.; Ravenek, W.; Ros, P.; Schipper, P. R. T.; Schreckenbach, G.; Snijders, J. G.; Sola, M.; Swerhone, D.; te Velde, G.; Vernooijs, P.; Versluis, L.; Visser, O.; van Wezenbeek, E.; Wiesenekker, G.; Wolff, S. K.; Woo, T. K.; Ziegler, T. *ADF 2005.01*; Scientific Computing and Modelling NV, Free University, Amsterdam: Amsterdam, 2005.
(44) Baerends, E. J.; Ellis, D. E.; Ros, P. *Theor. Chim. Acta* **1972**, *27*, 339.
(45) MOE *Molecular Operating Environment*, 2006.06; Chemical Computing Group, Montreal: Montreal, 2006.
(46) Comba, P.; Remenyi, R. *J. Comput. Chem.* **2002**, *23*, 697−705.

is important. In such a case, none of the protocols described above is completely satisfactory since it ultimately leads to a single structure. Full MD on the entire unconstrained system is far more preferable. However, in this paper, we are concerned with reproducing solid-state structures which represent particular 'snap shots' of the energetically accessible conformations. By treating a large number of proteins, the hope is that a wide a range of possible configurations of the T1 center will be sampled which will provide a good test of the flexibility of the LFMM parametrization. That is, the LFMM is being validated against a series of individual structures by carrying out individual optimizations and configurational averaging is not yet required.

Results and Discussion

The central design philosophy behind the LFMM is transferability. That is, although proteins are "remarkably intricate and complex ligands",[3] from a coordination chemist's perspective, they actually form simple metal complexes. Hence, we require the theoretical model to be able to treat Cu−L interactions wherever they occur whether in small molecules or in metalloproteins. This is the significant difference between the present approach and the MM study of Comba and Remenyi[46] who designed FF parameters specifically for the Pc-like CuN_2SS' site. These parameters cannot be used either for small-molecule systems nor any of the other T1 types shown in Figure 1. In contrast, the LFMM captures not only the essence of the individual M−L bond but also how that bond influences and is influenced by the other ligands in the complex. This mutual interplay and lack of bias is the kind of behavior inherent in QM.

This comparison with QM then begs the question of how the LFMM describes covalency. Covalent effects, especially with respect to the Cu−S(CYS) bond are extremely important in T1 centers. However, theoretical descriptions of covalency are necessarily based on an orbital description, i.e., QM, and orbitals are model dependent. For example, density functional theory calculations on planar $[CuCl_4]^{2-}$ give varying degrees of Cu−Cl covalency depending on the amount of Hartree−Fock exchange in the hybrid B3LYP functional.[47] The LFMM approach does not give a direct measure of covalency. Instead, it provides an indirect commentary but only to the extent that covalency and bond lengths are related. By extensively validating the LFMM against experimental structures, the gross covalency is implicitly captured. We might then anticipate that the ligand field part of the LFMM calculations might form the basis of a reasonable description of, say, the EPR *g* values or the d−d transition energies. However, detailed variations of these spectroscopic properties might be too subtle for the LFMM, although a full evaluation will require adequate configurational sampling, which is beyond the scope of the present publication.

The features required in LFMM calculations have been described elsewhere.[40,48] In short, the ligand field stabilization energy (LFSE) is based on d-orbital energies calculated by the angular overlap model (AOM).[49] The ligand-field contribution to each M−L bond is parametrized by e_σ, $e_{\pi x}$, and $e_{\pi y}$ describing, respectively, the M−L σ interaction and the two mutually perpendicular π interactions. A fourth AOM parameter, e_{ds}, monitors the configuration interaction between the valence metal d orbitals and the metal s orbital.[50] The AOM parameters are assumed to vary as some inverse power of the bond length. Electrostatic theory predicts that the crystal field splitting is proportional to $1/r^5$, but in LFMM applications, $1/r^4$ or even $1/r^3$ dependencies are common. Providing the whole force field is correctly balanced, the specific treatment of individual terms is less important.

In addition to the LFSE, the M−L bond is also modeled via a Morse potential requiring definition of the dissociation energy, *D*, the reference bond length, r_0, and the curvature parameter, α. Explicit L−M−L angle bending potentials are removed and replaced by a purely repulsive potential between the ligand donor atoms, modeled as an inverse power of the interatomic distance, *d*. Force constants for torsional twisting around M−L bonds are set to zero. Hence, the only 'conventional' FF parameters, i.e., the only ones which must be added to the distributed MOE force field file, are for M−L−B angle bending. The LFMM and additional AMBER94 parameters are given in the Supporting Information.

LFMM parametrization begins with simple homoleptic complexes containing models of the biologically relevant ligands as the basic reference point. In keeping with other work,[9,37] HIS is modeled as imidazole (imid), CYS as methylthiolate ($[SMe]^-$), and MET as dimethylsuphide (DMS). The carbonyl oxygen of GLU and GLY is modeled using formaldehyde.

Experimental structures for $[Cu(imid)_4]^{2+}$ complexes and related systems are available in the Cambridge Structural Database,[51] and the LFMM has been applied to Cu−imine donors before.[39] The existing parameters were simply mapped onto the AMBER atom types. Model structures for homoleptic thiolate and thioether complexes, $[Cu(SCH_3)_4]^{2-}$ and $[Cu(DMS)_4]^{2+}$, respectively, were computed using DFT. LFMM parameters were derived by assuming 'reasonable' values for the AOM parameters which approximately reproduce the mainly d molecular orbital energy differences from the DFT calculations and then adjusting the Morse function and ligand−ligand repulsion parameters until a good fit between the LFMM and DFT-optimized structures was obtained. As a further test, the structure of the typical active site model $[Cu(imid)_2(SMe)(DMS)]^+$ was also computed and compared to the DFT-optimized structure. This revealed a significant feature of the LFMM. In order to get a good balance between the first-row N and second-row S donors, the power dependence of the ligand−ligand repulsion term for the latter needed to be less severe, i.e., $1/d^4$ as opposed to $1/d^6$ for N donors.

After a degree of trial and error, a suitable set of parameters emerged which gave good structures of both the

(47) Szilagyi, R. K.; Metz, M.; Solomon, E. I. *J. Phys. Chem. A* **2002**, *106*, 2994−3007.
(48) Deeth, R. J.; Foulis, D. L. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4292−4297.
(49) Schaeffer, C. E.; Jorgensen, C. K. *Mol. Phys.* **1965**, *9*, 401.
(50) Riley, M. J. *Inorg. Chim. Acta* **1998**, *268*, 55−62.
(51) Fletcher, D. A.; McMeeking, R. F.; Parkin, D. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 746−749.

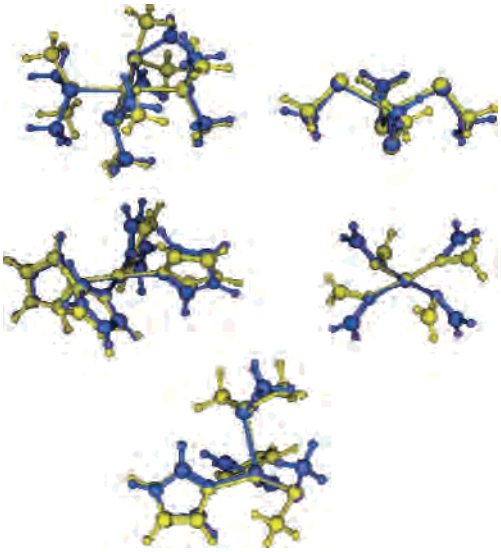

Figure 2. Overlay of DFT-optimized (blue) and LFMM-optimized (yellow) model structures, clockwise from top left: $[Cu(DMS)_4]^{2+}$, $[Cu(SCH_3)_4]^{2-}$, $[Cu(O{=}CH_2)_4]^{2+}$, $[Cu(imid)_2(SCH_3)(DMS)]^{+}$, and $[Cu(imid)_4]^{2+}$. NB: blue structure for $[Cu(imid)_4]^{2+}$ taken from CSD (refcode GADGOH).

homoleptic models and the model active site simultaneously. The overlays for all five systems are shown in Figure 2, and coordinates are provided in the Supporting Information. The imidazole, DMS, and formaldehyde complexes are planar, while the methylthiolate complex has a flattened tetrahedral structure reminiscent of $[CuCl_4]^{2-}$. The significant achievement of the LFMM modeling is that the trigonal structure of the T1 active site model is a local minimum and that the Cu−DMS bond length automatically lengthens by ∼0.6 Å relative to the model homoleptic species while the Cu−S(thiolate) distance shortens by nearly 0.2 Å relative to the homoleptic model. The LFMM model thus faithfully emulates the QM behavior.

In order to provide the best 'in-protein' description, the model homoleptic structures were assigned the relevant partial atomic charges reported for the complete residues[46] apart from the hydrogen which replaces the side chain carbon, which was given a charge of zero. The copper was assigned a charge of 0.777, which corresponds to the value required for a typical T1 center as described in the Computational Details section. The total charges on the model homoleptic complexes thus do not match their 'true' values. However, this approximation does not have a serious affect on the LFMM structures. For example, the LFMM-optimized structure of $[Cu(imid)_4]^{2+}$ gives a Cu−N distance of 2.02 Å. With no electrostatic interactions at all, the Cu−N distance is only 0.01 Å shorter. The same behavior is found for $[Cu(SMe)_4]^{2-}$, which has a Cu−S bond length of 2.37 Å. Unsurprisingly, the weaker Cu−S bonding in $[Cu(DMS)_4]^{2+}$ leads to greater sensitivity of the computed bond length to the charge scheme. With electrostatics enabled, the optimized distance is 2.37 Å, and with no electrostatics, this decreases to 2.31 Å. Thus, although one would expect the metal charge to vary in line with the large changes in covalency spanned by the different homoleptic systems, the particular charge on the metal does not appear to be significant, at least from a structural point of view. This is

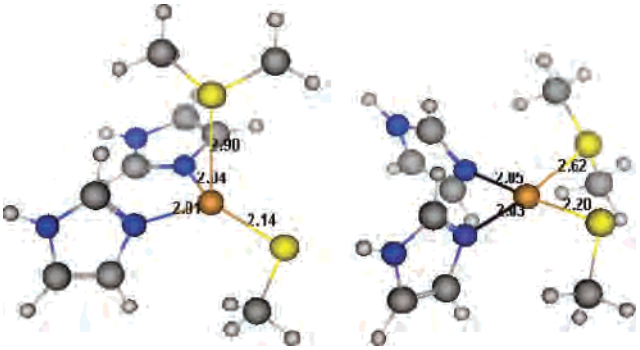

Figure 3. LFMM-optimized active site models: left, trigonal geometry; right, planar geometry.

probably to be expected since in MM, 1−2 and 1−3 electrostatic interactions are explicitly ignored while the 1−4 interactions are usually scaled by some factor less than one.

Previous theoretical calculations on the oxidized active site model complexes employed the B3LYP hybrid functional for geometry optimization followed by CASPT2 single point energies and explored various ligand combinations involving NH_3, imidazole, SH^-, SMe^-, SH_2, and SMe_2. The calculations established both trigonal and 'tetragonal' structures,[35] although the latter could also be viewed as approximately planar.

There is a delicate balance between trigonal and tetragonal geometries with the former favored only when there is a strong bond to a soft, polarizable group such as a thiolate sulfur. The LFMM results for the largest model species are shown in Figure 3. The B3LYP/CASPT2 results with NH_3 instead of imidazole suggests that the trigonal structure is slightly lower in energy than the tetragonal while the LFMM (with imidazole) places the tetragonal system about 1.5 kcal mol^{-1} lower than trigonal. The trigonal structure is strongly influenced by the AOM d−s mixing parameters, especially for the thiolate donor. In fact, until a threshold value is exceeded, the LFMM does not optimize to a trigonal structure and always collapses to the planar geometry. This 'switching' behavior is reminiscent of the LFMM study of Jahn−Teller effects in six-coordinate Cu(II) complexes where the d−s mixing is required to cause a changeover from compressed to elongated structures.[39]

At this stage, then, LFMM parameters have been developed without recourse to any experimental data on actual proteins and closely match the behavior of DFT calculations. Since single-crystal X-ray diffraction experiments are widely regarded as a definitive source of 3-dimensional structural information on T1 copper sites, a wide selection of systems was chosen for further study (Table 1).

The resolution of these data is significantly less than for small-molecule systems. Even with relatively high-resolution data, the uncertainty for strongly bound HIS and CYS ligands is about 0.05 Å, and more like 0.1 Å for lower resolution structures, while the uncertainty for the longer, weaker bonds is about 0.2 Å.

The chosen systems represent multiple examples of the three different donor sets. To the extent that each structure can be considered a 'snap shot' and that each protein provides

Table 1. PDB Codes and Protein Types and X-ray Diffraction Resolutions for the Systems Studied Here[a]

PDB code: protein	diffraction resolution, Å
1AAC: Ami	1.31
1AG6: Pc	1.7
1AZN: Az (av)	2.6
1BQK: Paz	1.35
1BXU: Pc	1.9
1DYZ: AzII	1.75
1E30: Rust (M148Q)	1.5
1F56: PLT (av)	2.05
1ID2: Ami (av)	2.15
1ID2: Ami Chain 1	2.15
1ID2: Ami Chain 2	2.15
1ID2: Ami Chain 3	2.15
1IUZ: Pc	1.6
1JER: STC	1.6
1JZF: Az (no Ru)	1.5
1JZF: Az (Ru)	1.5
1KDJ: Pc	1.7
1PLC: Pc	1.33
1RCY: Rust	1.9
1SF5: Ami (P94A)	1.1
1SFD: Ami (P94F) (av)	0.99
1X9R: Umi (av)	1.9
2CBP: CBP	1.8
2FT6: AzAmi	1.25
4AZU: Az (av)	1.9
4Paz: Paz (P80A)	1.76
6Paz: Paz (P80I)	1.91
8Paz: Paz	1.6

[a] Az = azurin, Paz = pseudoazurin, Pc = plastocyanin, PLT = plantacyanin, RCY = rusticyanin, Ami = amicyanin, STC = stellacyanin, Umi = umicyanin, CBP = cucumber basic protein. Mutants are indicated by the code in parentheses. PDB files with multiple independent molecules in the unit cell where the structural data have been averaged are indicated by (av). For 1KDJ, structures were generated excluding the [Ru(terpy)(bipy)] moiety (no Ru) and including it (Ru) but with the RuN_6 core frozen.

a different local environment for its T1 center, we can expect the structures to provide a sampling of the range of accessible T1 geometries. If the distribution is normal, then the average of a series of systems represents the 'best' structure. Alternatively, it has been argued by Comba and Remenyi that the experimental errors are so large that all the structures (for Pc-type active sites) are essentially the same.[46] Whichever is the case, a gross assessment of the accuracy of the LFMM calculations can be made by comparing the average calculated Cu−L bond lengths with average experimental value. The latter are 2.01, 2.05, 2.16, 2.92, 2.78, and 2.26 Å for $Cu-N_{back}$, $Cu-N_{loop}$, $Cu-S_{CYS}$, $Cu-S_{MET}$, $Cu-O_{GLU}$, and $Cu-O_{GLY}$, respectively, versus computed values of 2.03, 2.03, 2.16, 2.90, 2.69, and 2.26 Å. Here, N_{back} refers to the 'backbone' coordinating histidine while N_{loop} refers to the histidine in the C-terminal loop. Given the uncertainties in the X-ray data, this represents excellent general agreement, especially when we recall that the LFMM parameters yield a Cu−S bond length of 2.36 Å for the model $[Cu(SMe)_4]^{2-}$, 2.38 Å for $[Cu(DMS)_4]^{2+}$, and 1.98 Å for $[Cu(OCH_2)_4]^{2+}$, which are significantly different from the in-protein values and again illustrate the QM-like performance of the LFMM.

Since the LFMM gives the same averaged structure as experiment, significant deviations from this 'norm' can presumably be attributed to local environmental differences. To assess whether the LFMM can also capture these more

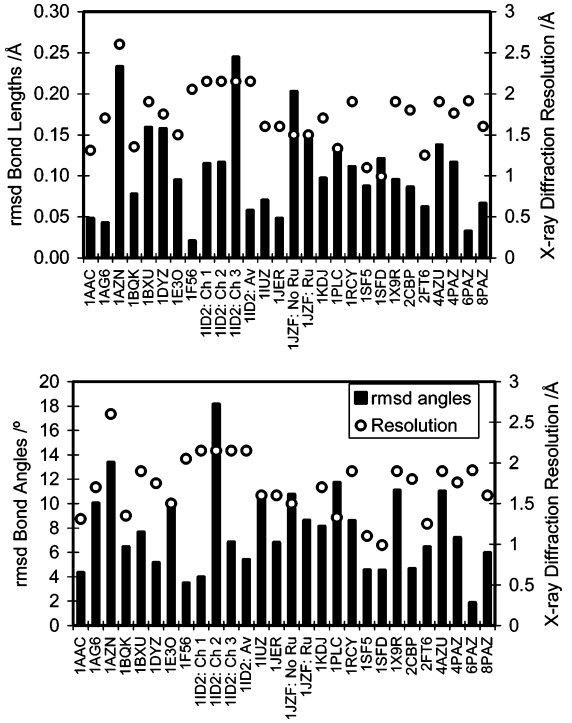

Figure 4. Comparison of rms deviations for complete Cu−L donor sets with X-ray diffraction resolution. Top, bond lengths; bottom, bond angles.

subtle changes, detailed rms deviations for bond lengths and bond angles for each molecule are displayed in Figure 4. The agreement with experiment is generally very good. The average rms deviation across the whole dataset is 0.11 Å for Cu−L bonds and 8° for L−Cu−L′ angles, which is comparable with the experimental uncertainty.

Equatorial Bonds to HIS and CYS. The average rms deviations between PDB and LFMM bond lengths is 0.07, 0.11, and 0.06 Å for $Cu-N_{back}$, $Cu-N_{loop}$, and $Cu-S_{CYS}$, respectively, while for $Cu-S_{MET}$ the average deviation is 0.18 Å. Given the relatively large uncertainties in protein crystallographic data alluded to above, the agreement is encouraging. However, the LFMM performs even better than suggested by Figure 4. For example, the apparently poor agreement for Azurin (1AZN) is due to a large experimental error in the Cu−N(HIS117), which at 2.36 Å is obviously too long. For chain 3 of Amicyanin (1ID2), the error comes from a large difference between the Cu−S(MET) bonds but since this bond is long and weak, the difference is not energetically significant. A poor fit for Cu−L bond length for Rusticyanin (1RCY) derives from the Cu−N(HIS143) and Cu−S(CYS138) distances which are reported to be 1.89 and 2.26 Å, respectively, versus computed values of 2.07 and 2.15 Å. The experimental distances seems anomalously short and long, respectively, but the crystal only diffracted to 2 Å so the resolution might be expected to be relatively poor. In contrast, the data for the P94F mutant of Amicyanin (1SFD) should be significantly better as measured by the 0.99 Å resolution, but the disagreement between theoretical and experimental bond lengths still appears to be large. However, this is another example of a Cu−S(MET) discrepancy. The Cu−S(MET98) bond length is reported to be 2.80 Å long but calculated to be 3.01 Å.

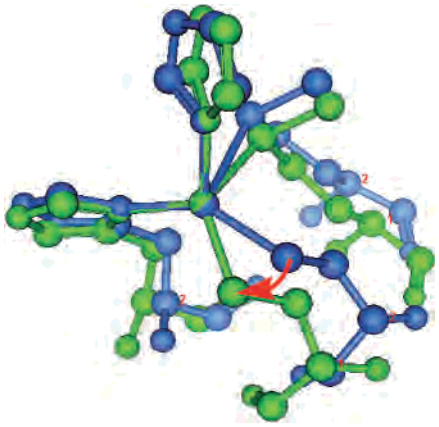

Figure 5. Overlay of active site structure for chain 1 (green) and chain 2 (blue) of Amicyanin (PDB code 1ID2) showing displacement of the cysteine ligand (arrow).

On average, the reproduction of the bond angles at the copper center is also good with the apparent exception of one of the independent molecules in the crystal of Amicyanin (1ID2). In general, the 'trigonal' CuN_2S unit is not rigorously trigonal in the sense that all the in-plane angles at copper are not precisely 120°. Instead, it has unequal $S_{CYS}-Cu-N_{HIS}$ angles with one less than 120° and the other 130° or larger. Since there are two histidine ligands, there are two possible arrangements of the N_2S triad. Although the Ami crystal structure maintains the same angles for all three independent chains, the LFMM optimization keeps chains 1 and 3 the same as experiment but swaps the two $S_{CYS}-Cu-N_{HIS}$ angles in chain 2, leading to an apparently large deviation between theory and experiment. Excising the two active sites to generate model $[Cu(imid)_2(SMe)(DMS)]^+$ centers and computing the LFMM single point energy indicates that the chain 2 active site is only ~1 kcal mol^{-1} higher in energy. The fact that the final structure depends on the starting point has important consequences for correlating spectroscopic data with individual structures rather than with configurational averages. We return to this crucial issue later.

The correlations between experimental and theoretical data for individual Cu−L bonds are shown in Figures 6 and 7. The spread in PDB data is much larger than in the computed data, reflecting the relatively large experimental uncertainties. For small-molecule systems where the experimental error is much less, an analysis of four-coordinate complexes located in the CSD containing at least two Cu−N(imidazole) bonds gives a spread of about 0.1 Å. The obvious outliers in the Cu−N distances from PDB structures (~2.2 and ~2.4 Å) are thus probably crystallographic artifacts and not significant.

At first sight, there appears to be a trend for somewhat shorter $Cu-N_{back}$ distances compared to $Cu-N_{loop}$, as displayed by the average bond lengths of 2.01 and 2.05 Å from the PDB data, although this difference is much reduced in the LFMM calculations (2.027 and 2.034 Å, respectively). However, this observation is of small significance given that individual Cu−N distances carry an experimental uncertainty of the order of 0.1 Å. Moreover, while on average the bond

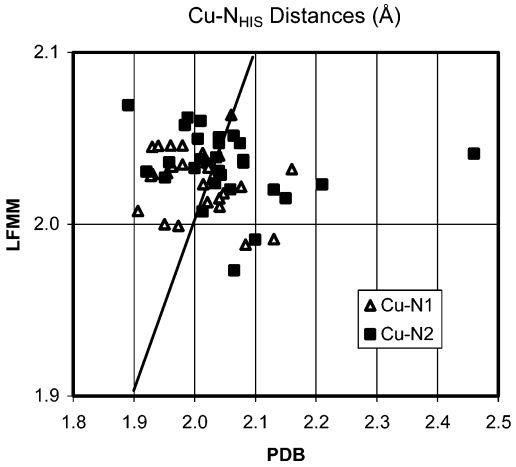

Figure 6. Correlation between experimental (PDB) and calculated (LFMM) Cu−N bond lengths (Å). Cu−N1 corresponds to the backbone histidine and Cu−N2 to the loop histidine.

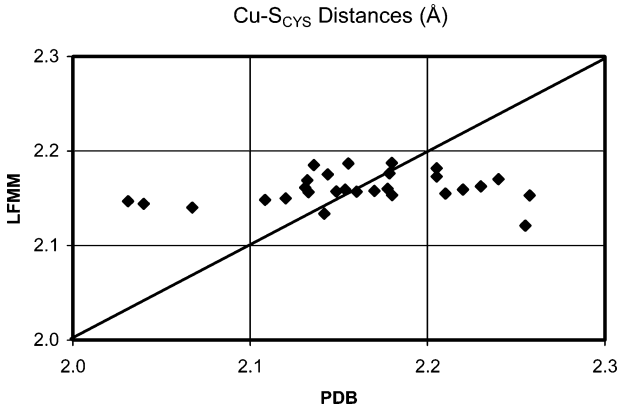

Figure 7. Correlation between experimental (PDB) and calculated (LFMM) $Cu-S_{CYS}$ bond lengths (Å).

lengths to the backbone HIS appear to be shorter than to the loop HIS, there are many examples in both PDB and LFMM structures of active sites with the reverse sense of Cu−N distances. Hence, while both theory and experiment suggest that the two $Cu-N_{HIS}$ distances are seldom the same, there does not appear to be any systematic trend that the shorter contact must always be to the backbone histidine.

The LFMM calculations also give a much smaller spread in the $Cu-S_{CYS}$ distances (0.07 Å) than that reported in the PDB structures (more than 0.2 Å). This seems due again to the uncertainties in the crystallographic data rather than a too stiff LFMM bond stretch potential. In contrast, the Cu−S distance extends to nearly 2.4 Å for the model $[Cu(SCH_3)_4]^{2+}$ so the longer Cu−S distances seen in PDB structures could, in principle, be accommodated if the electronic structure of the site required it. Compressing the bond below the average value of 2.16 Å derived from both LFMM and PDB structures would cost even more energy than extension, suggesting that at least the three points less than 2.1 Å are unreasonable.

Overall, therefore, the experimental data for Cu−N and $Cu-S_{CYS}$ bonds are simply too 'noisy' to reveal any obvious correlations between theory and experiment.

Axial Bonds. The situation for the axial ligands, in particular for axial methionines, is different. As seen in

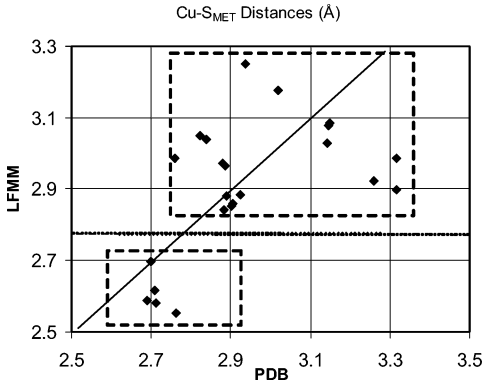

Figure 8. Correlation between experimental (PDB) and calculated (LFMM) Cu−S$_{MET}$ bond lengths (Å).

Figure 8, there is both a loose correlation and a fairly clear separation of axial donors into two distinct groups, i.e., those which have a computed Cu−S$_{MET}$ distance less than 2.7 Å (group G1) and those where it is greater than 2.84 Å (group G2).

The G1 group comprises six proteins: Sea Lettuce Pc from *Ulva Pertusa* (PDB code 1IUZ), Pseudoazurin (Paz) from *Achromobacter Cycloclastes* (PDB code 1BQK), Plantacyanin (PLT) from *Spinacia Oleracea* (PDB code 1F56), Cucumber Basic Protein (PDB code 2CBP), Pseudoazurin P80A mutant (Paz_P80A) from *Alcaligenes Faecalis* (PDB code 4Paz), and Pseudoazurin (Paz) from *Alcaligenes Faecalis* (PDB code 8Paz). G2 comprises Fern *Pc* from *Dryopteris Crassirhizoma* (PDB code 1KDJ), Amicyanin (Ami) from *Paracoccus Denitrificans* (PDB code 1AAC), Ami from *Paracoccus Versutus* (PDB code 1ID2), Spinach *Pc* from *Spinacia Oleracea* (PDB code 1AG6), Poplar *Pc* from *Populus Nigra* (PDB code 1PLC), Rusticyanin (RCY) from *Thiobacillus Ferrooxidans* (PDB code 1RCY), Azurin F114A mutant (Az_F114A) from *Pseudomonas Aeruginosa* (PDB code 1AZN[#]), *Pc* from *Synechococcus Sp.* (PDB code 1BXU), Azurin (Az) from *Pseudomonas Aeruginosa* (PDB code 1JZF[#]), Ami P94A mutant (Ami_P94A) from *Paracoccus Denitrificans* (PDB code 1SF5), Ami P94F mutant (Ami_P94F) from *Paracoccus Denitrificans* (PDB code 1SFD), Az with Ami loop mutant (AzAmi) from *Pseudomonas Aeruginosa* (PDB code 2FT6), Az from *Pseudomonas Aeruginosa* (PDB code 4AZU[#]), and Paz P80I mutant (Paz_P80I) from *Alcaligenes Faecalis* (PDB code 6PAZ). Those marked with [#] are five-coordinate sites.

There have been several attempts to correlate axial and equatorial Cu−S bond lengths with spectroscopic data.[3] For example, the relative intensity of the absorption band at about 430 nm increases relative to the ∼600 nm band in the series Pc, CBP, and nitrite reductase (NiR), and this has been attributed to a 'coupled distortion'[52] whereby a progressive shortening of the Cu−S$_{MET}$ bond is accompanied by a compensating lengthening of the Cu−S$_{CYS}$ bond and a change in the angle between the CuN$_2$ and CuSS′ planes. This relationship between the spectroscopy and structure is compelling but the LFMM results *based on optimizations*

(52) Basumallick, L.; Szilagyi, R. K.; Zhao, Y. W.; Shapleigh, J. P.; Scholes, C. P.; Solomon, E. I. *J. Am. Chem. Soc.* **2003**, *125*, 14784−14792.

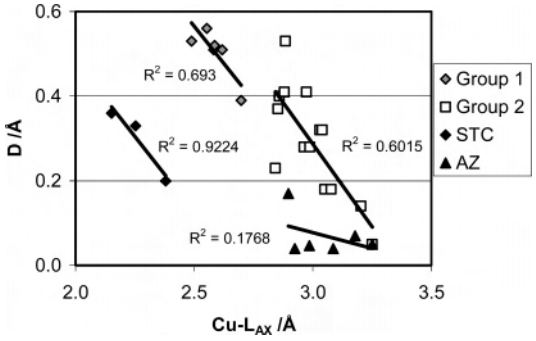

Figure 9. Out-of-plane displacements of Cu atom from N$_2$S$_{CYS}$ plane as a function of the bond length to the axial methionine S (or glutamine O for stellacyanin-type systems) distance. Group 1 have short Cu−S$_{MET}$ bonds, Group 2 long Cu−S$_{MET}$ bonds, AZ corresponds to five-coordinate Azurin-type systems, and STC corresponds to four-coordinate active sites with axial glutamines.

starting from PDB coordinates do not appear to confirm it. However, this does not necessarily imply that the coupled distortion model is invalid. Rather, it highlights the care required to select appropriate solid-state structural data for comparison with experimental data measured in solution.

For the coupled distortion correlation to work, plastocyanin needs to have the shortest, CBP an intermediate, and NiR the longest Cu−S(MET) distance. In poplar Pc (1PLC), this is 2.07 Å, but in spinach Pc (1AG6), it is 2.15 or even 2.26 Å for fern Pc (1KDJ). Of course, the 1PLC crystals diffract to 1.33 Å, as opposed to only 1.7 Å for the other two proteins, and so represents the best data. The Cu−S(MET) distance in CBP needs to be longer than in Pc and is reported to be 2.16 Å, but here the resolution is 1.8 Å. on the basis of structures of comparable resolution, this *could* have come out rather differently.

Therefore, we should not expect subtle correlations to be apparent with the present protocol, although there does appear to be a relationship between the displacement of the copper out of the N$_2$S plane, *D*, and the axial ligation (Figure 9). The G1 systems with short Cu−S$_{MET}$ contacts show larger displacements, while the longer Cu−S$_{MET}$ distances of the G2 systems tend to have smaller displacements. The five-coordinate centers all have nearly zero displacements. The apparent outlier with $D = 0.17$ Å corresponds to the Az structure (1JZF) where the attached ruthenium complex had been removed. When this is left in, the *D* value drops to 0.05 Å. Interestingly, the *D* values for axial glutamine ligands vary from 0.2 to 0.36 Å and correlate with the computed Cu−O distances, although there are only three data points. The experimental values are all about 0.3 Å. It is generally quoted that axial glutamine is a better donor than axial methionine, but the present calculations would appear to contradict this. It should be noted that apart from a small difference in the oxygen charge, the parameters describing Cu−O$_{GLU}$ and Cu−O$_{GLY}$ are the same, yet the former is very much shorter than the latter. The modeling suggests that the differences between GLU and GLY oxygen donors is not a difference in intrinsic binding abilities.

Variable Coordination Numbers. An important success of the LFMM modeling is that a single set of Cu−L parameters applies to a range of active sites, that is, Pc-like

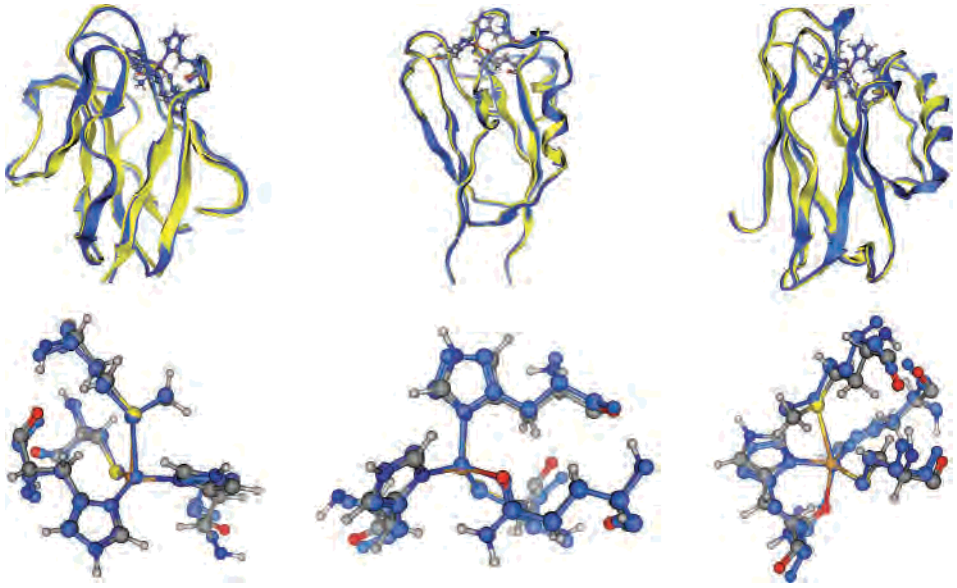

Figure 10. Rmsd overlays of experimental (blue) and computed (yellow or CPK) backbone carbons (top) and active sites (bottom) for Amicyanin (left), Stellacyanin (middle) and Azurin (right) (PDB codes 1AAC, 1JER, 1DYZ).

Table 2. Comparison of PDB and LFMM Cu−L Bond Lengths for Proteins Shown in Figure 10

1AAC	PDB	LFMM	1JER	PDB	LFMM	1DYZ	PDB	LFMM
HIS53	1.954	2.029	HIS46	1.960	2.045	GLY45	2.720	2.666
CYS92	2.108	2.148	CYS89	2.178	2.176	HIS46	2.040	2.014
HIS95	2.033	2.023	HIS94	2.043	2.028	CYS112	2.135	2.184
MET98	2.904	2.859	GLN99	2.209	2.253	HIS117	1.988	2.062
						MET121	3.260	2.924

Table 3. Comparison of PDB and LFMM Bond Angles at the Metal Center for the Proteins Shown in Figure 10

1AAC	PDB	LFMM	1JER	PDB	LFMM	1DYZ	PDB	LFMM
HIS53−CYS92	136	141	HIS46−CYS89	133	139	GLY45−HIS46	78	82
HIS53−HIS95	104	97	HIS46−HIS94	101	98	GLY45−CYS112	104	98
HIS53−MET98	84	86	HIS46−GLN99	94	87	GLY45−HIS117	86	83
CYS92−HIS95	112	108	CYS89−HIS94	117	113	GLY45−MET121	148	156
CYS92−MET98	110	112	CYS89−GLN99	101	112	HIS46−CYS112	132	140
HIS95−MET98	100	104	HIS94−GLN99	101	95	HIS46−HIS117	106	103
						HIS46−MET121	73	75
						CYS112−HIS117	121	116
						CYS112−MET121	105	105
						HIS117−MET121	88	92

CuN_2SS', STC-like CuN_2SO, and Az-like $CuN_2SS'O$ centers. Overlays of the protein backbones and active sites are shown for an example of each of the representative T1 sites in Figure 10 with a detailed comparison of Cu−L bond lengths and L−Cu−L bond angles in Tables 2 and 3, respectively. The performance of the LFMM model is good and contrasts with the previous purely MM treatment of Comba and Remenyi, which can only work for CuN_2SS' sites typified by that found in the plastocyanins. In fact, the LFMM behaves just like QM in that each Cu−L bonding interaction influences, and is influenced by, all the other interactions. Hence, the model works equally well for these T1 active sites and simple coordination complexes even though the structures vary a great deal. The electronic energy inherent in the LFSE term provides this connection and confers quantum-like behavior but at a MM price.

Comparison with QM/MM. The hybrid QM/MM approach has emerged as a useful and versatile method for modeling proteins.[53,54] However, any computation which involves a quantum part will be significantly slower than MM. Now, since the LFMM reproduces the experimental crystal structures and since this criterion is often used to judge the quality of QM/MM studies, it follows that the LFMM should be a fast, but equally good, alternative to QM/M.

Several QM/MM studies involving T1 centers have appeared.[9,19,28,37] As with all QM/MM schemes, there is the issue of how to couple the quantum part to the classical part. The LFMM circumvents this issue since the whole protein is treated on the same footing. However, the QM/MM embedding scheme is of secondary importance. The major issue with QM/MM is that proper configurational averaging is seldom, if ever, undertaken which, according to Olsson et al.,[53] is due to the excessive computation time. We return to this issue below. Meanwhile, in order to make a sensible comparison, the present analysis is restricted to comparing the LFMM results with QM/MM optimizations for Pc, CBP, and Az reported by Ryde and Olsson.[9] (Nitrite reductase is omitted since it contains a second copper center.)

(53) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.
(54) Ryde, U. *Curr. Opin. Chem. Biol.* **2003**, *7*, 136−142.

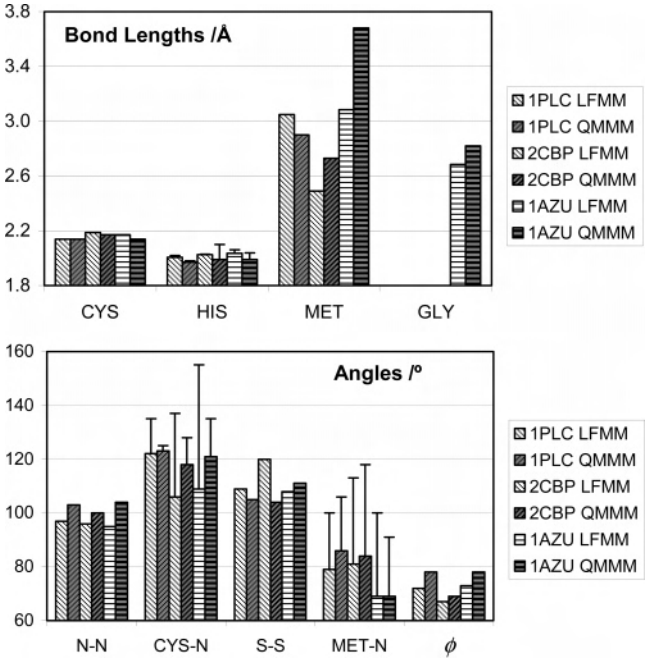

Figure 11. Comparison of LFMM and QM/MM bond lengths and angles around the copper centers in oxidized Pc, CBP, and Az. The T-bars represent the larger value when there are two values represented by a single bar. The QM/MM data have been extracted from Table 2 of ref 9 and refer to oxidized copper, with the protein not fixed and the active site connected to the backbone. ϕ is the angle between the CuNN and CuSS′ planes.

Table 4. Optimized Cu−L Bond Lengths for the Three Independent Chains of Amicyanin (1ID2)

chain	Cu−N_{back}	Cu−S_{CYS}	Cu−N_{loop}	Cu−S_{MET}	Cu−S_{MET}: X-ray
1	2.036	2.144	2.023	2.973	2.88
2	2.032	2.158	2.015	3.203	2.75
3	2.045	2.157	2.031	2.880	2.86

As expected, the LFMM gives essentially the same structures as QM/MM (Figure 11). A conservative estimate based on the QM/MM study of Comba et al.[37] suggests the LFMM is at least 1000 times faster than the code they used.

Strain Energy and 'Entatic Bulging'. There is a lively debate concerning whether the T1 copper center is strained and, if so, by how much. Computational estimates suggest a strain energy, ΔE_{str}, for an oxidized T1 center of the order of 8−10 kcal mol^{-1}.[9] That is, the model active site [Cu(imid)$_2$(SMe)(DMS)]$^+$ at its 'in-protein' geometry is 8−10 kcal mol^{-1} higher than the fully relaxed in vacuo structure. It is of interest to compare this value with the LFMM estimates.

Most of the proteins studied here have a single molecule in the unit cell. Amicyanin (PDB code 1ID2) is unusual in having three independent molecules and has been discussed above in terms of the bond angles around the copper centers.[55] Each independent molecule is chemically the same, and the optimized Cu−L bond lengths for each chain vary by only 0.02 Å for the strongly coordinated equatorial groups. However, and not unexpectedly, much larger changes are evident for the long axial contact: about 0.3 Å (Table 6).

Table 5. Strain Energies (kcal mol^{-1}) Computed for the Three Independent Chains of Amicyanin

1ID2	in-protein	in vacuo	ΔE_{str}
chain 1	−182.031	−192.25872	10.2 (0.44 eV)
chain 2	−181.134	−192.25894	11.1 (0.48 eV)
chain 3	−184.117	−192.25824	8.1 (0.35 eV)

Table 6. Comparison of Active Site Geometries from the Averaged Solvation Simulation, the Initial LFMM Optimization, and the PDB X-ray Diffraction Structure for 1PLC

metric	simulation average	LFMM single	X-ray (1PLC)
r(H37)	2.020	2.008	1.906
r(C84)	2.168	2.140	2.067
r(H87)	2.017	2.020	2.059
r(M92)	2.826	3.050	2.822
a(H37−C84)	123.0	119.4	131.7
a(H37−H87)	101.1	97.0	97.2
a(H37−M92)	78.9	74.1	88.5
a(C84−N87)	123.2	141.0	121.0
a(C84−M92)	107.3	101.8	109.9
a(H87−M92)	115.0	101.1	100.6
plane angle	68.5	71.5	81.6
D	0.42	0.18	0.36

The final structure is therefore relatively sensitive to the starting point, and there are small but potentially significant fluctuations in the local structure around the copper center, as shown in Table 4. Copper(II) is notorious for its so-called 'plasticity'.[56] An obvious manifestation is the Jahn−Teller effect in six-coordinate species where the difference between short equatorial contacts and long axial bonds can be up to 0.6 Å.[57−59] Depending on the system, all three potential axes of elongation can be active since there is often only a small energy barrier separating successive elongated minima. Hence, Cu(II) complexes can exhibit apparently massive structural changes, but these cost very little energy. Other coordination geometries are equally plastic. The upshot for T1 centers is that structural variability for Cu(II) systems is natural and that the 10 kcal mol^{-1} or so of strain in a T1 center can manifest in a variety of structures depending on the details of the local coordination. That is, within certain limits, the coordination geometry 'bulges' in response to minor external perturbations and, for amicyanin, this results in small but significant fluctuations in the Cu−L bond distances as a function of the protein configuration. Clearly, in order to compute a reliable structure for comparison with experimental spectroscopic data, configurational averaging will be necessary.

The relatively subtle differences between the three otherwise identical chains also lead to small variations in the local strain energy (Table 5). An estimate can be obtained by excising the active site from the optimized structure and computing the energy difference between the in-protein and in vacuo structures. A representative overlay is shown in Figure 12. The structure does not change much. The axial

(55) Romero, A.; Nar, H.; Huber, R.; Messerschmidt, A.; Kalverda, A. P.; Canters, G. W.; Durley, R.; Mathews, F. S. *J. Mol. Biol.* **1994**, *236*, 1196−1211.
(56) Gazo, J.; Bersuker, I. B.; Garaj, J.; Kabesowa, M.; Kohout, B. J.; Langfelderova, V.; Melnik, M.; Serator, M.; Valach, F. *Coord. Chem. Rev.* **1976**, *19*, 253.
(57) Burton, V. J.; Deeth, R. J. *J. Chem. Soc., Chem. Commun.* **1995**, 573−574.
(58) Hitchman, M. A. *Comments Inorg. Chem.* **1994**, *15*, 197.
(59) Deeth, R. J.; Hitchman, M. A. *Inorg. Chem.* **1986**, *25*, 1225−1233.

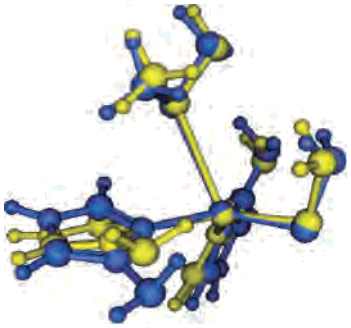

Figure 12. Representative overlay between in-protein (blue) and in vacuo (yellow) active site.

bond length alters by 0.06 Å, with the others changing by less than 0.02 Å, and yet the computed energy difference is around 10 kcal mol^{-1} of which about a third is down to changes in the LFSE. Note that the strain energy fluctuates by ±15% (∼ ±1.5 kcal mol^{-1}) for the different Ami chains again suggesting that an accurate 'in solvent' estimate will require more exhaustive sampling.

The magnitude of the LFMM strain energy is in line with QM estimates and is consistent with the recent analysis of protein folding free energies.[60] Note that these strain energies should not be confused with reorganization energies which await the development of a suitable force field for Cu(I).

Configurational Averaging. An important influence on the optimized structure turns out to be the arrangement of solvent molecules. To better quantify the effect, a 40 ps MD simulation on solvated Pc (PDB code 1PLC) was carried out with the solvent allowed to move while the protein is kept frozen, taking a snapshot every picosecond. The first 15 ps were discarded and the remaining 25 configurations energy minimized using the LFMM with all atoms free. The local structure at the metal fluctuates with the Cu−N_{HIS} bonds spanning 0.06 Å, Cu−S_{CYS} 0.04 Å, and Cu−S_{MET} 0.35 Å.

Table 6 gives a comparison between various metrics of the active site averaged over the 25 individual configurations, the single LFMM optimization starting from the PDB coordinates, and the PDB metrics themselves. While all three sources give broadly similar results, there are some relatively large differences between the simulation average and the other two. For example, this particular PDB structure seems to have an anomalously short Cu−S_{CYS} bond length of 2.07 Å despite its relatively high resolution. The LFMM Cu−S distance of around 2.15 Å is much more consistent with the other PDB structures and the most recent DFT calculations of Hansen et al.s' large T1 model[18] and Sinnecker and Neese's QM/MM calculations, both of which give a Cu−S distance of 2.2 Å. The LFMM single optimization gives too small a displacement of copper from the equatorial plane compared to the PDB structure, but the agreement is enhanced in the simulation average.

Another interesting feature of the simulation average is that the bond length variations display a kind of 'coupled distortion'.[52] The latter asserts that (a) shortening the Cu−S_{MET} axial bond should result in the copper being lifted more out of the trigonal plane, (b) a shorter Cu−S_{CYS} bond should lead to a larger angle between the CuNN and CuSS′ planes,

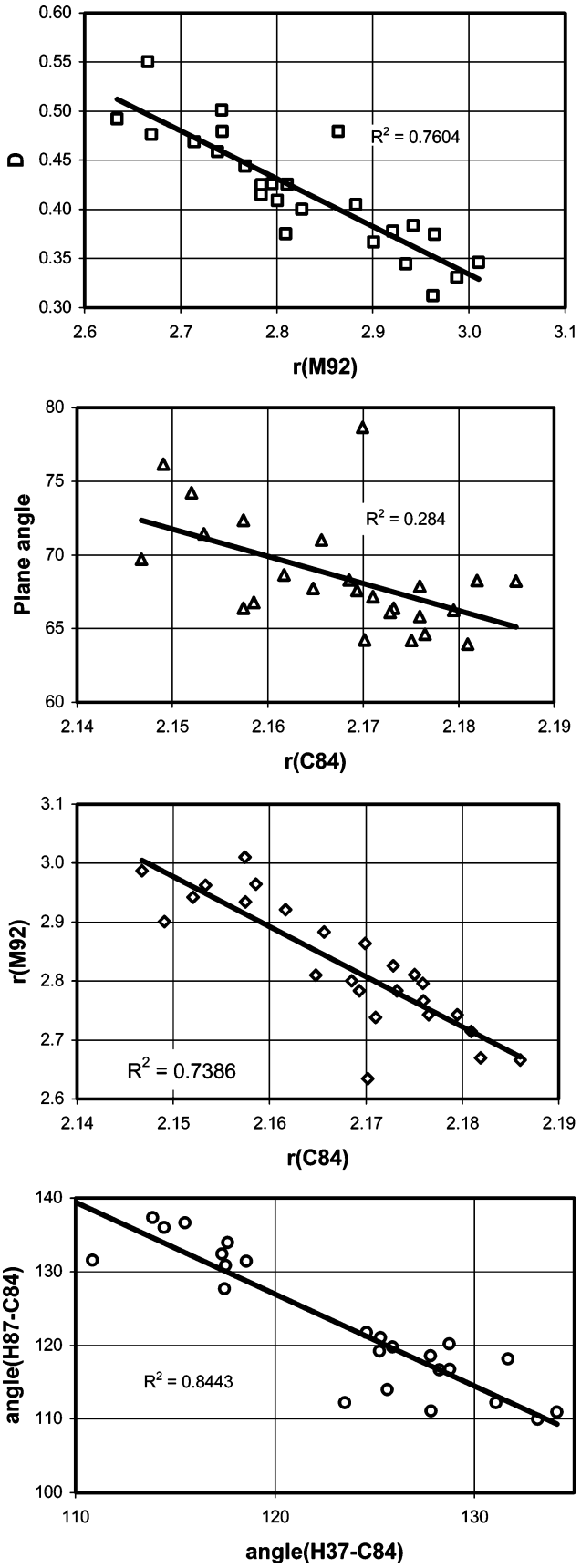

Figure 13. Correlations between calculated geometrical parameters in plastocyanin from different solvation configurations.

(60) Frank, P.; Benfatto, M.; Szilagyi, R. K.; D'Angelo, P.; Della Longa, S.; Hodgson, K. O. *Inorg. Chem.* **2005**, *44*, 1922−1933.

and (c) a shorter $Cu-S_{CYS}$ bond should lead to a longer $Cu-S_{MET}$ contact. The data in Figure 13 support the first and last, but there does not appear to be a relationship between the $Cu-S_{CYS}$ distance and the CuNN/CuSS′ interplane angle, at least not for Pc. On balance, however, the calculations support the idea that a coupled distortion could well be at work across a series of related T1 centers and that we will need to carry out configurational averaging in order to see it properly. Moreover, the calculations show two configurations for the in-plane ligands. One of the $S_{CYS}-Cu-N$ angles is less than 120°, and the other is greater than 120°, with the CYS ligand able to switch between the two possibilities, a feature which was already noted for Amicyanin (Figure 5). This feature is hidden by the simple average of the simulation data, which gives the same value for both angles (∼123°) even though each individual contributing structure has an asymmetric in-plane coordination geometry. This could have important consequences for interpreting crystallographic data since the experimental structure will be a Boltzman-weighted average of the two possibilities, assuming the barrier between them is relatively small. The spectroscopy will probe the instantaneous asymmetric structure and will not correlate with the averaged crystallographic structure. This effect is very well known in copper(II) coordination chemistry. For example, the simple amine complex $[Cu(tach)_2]X_2$ (tach = 1,3,5-triaminocyclohexane) gives a 'normal' elongated structure at room temperature when X = ClO_4^- but six nearly equal Cu−N bond lengths for X = NO_3^-.[61] However, the d−d spectra of latter displays two broad bands indicative of a tetragonal geometry rather than the single d−d absorption which would be expected if the nitrate salt were truly octahedral.

Conclusions

LFMM has been applied to a wide range of copper proteins containing a single oxidized Type I center. The model is extensively validated by comparing a series of individual experimental structures with a series of individual LFMM optimizations. Fully optimized unconstrained structures were computed for 24 complete proteins containing T1 centers spanning four-coordinate, plastocyanin-like CuN_2SS' and stellacyanin-like CuN_2SO sites plus the five-coordinate $CuN_2SS'O$ sites of the azurins. The agreement is very encouraging and well within experimental error, particularly given that the instances of worst agreement are normally due to gross errors in the PDB structures.

The geometry optimizations have been carried out with a 10 Å sheath of water molecules, and the detailed structure around the metal center is found to depend on the starting solvent configuration. Since Cu(II) is intrinsically plastic, the ∼10 kcal mol^{-1} of strain energy imposed by the protein backbone can be manifested in relatively large structural variations for relatively minor differences in protein/solvent configurations. This 'entatic bulging', akin to Solomon's coupled distortion coordinate, can result in bond length and angle variations in a single protein which are of the same order of magnitude as the differences between a series of proteins based on individual PDB structures. Twenty-five separate LFMM optimizations of plastocyanin starting from different solvent configurations establishes a fairly wide range of structures is accessible to the system such that reliance on a single structure—whether it be a PDB file or a single LFMM or QM/MM optimization—may not be a good representation of the true structure in solution. As suggested by Warshel, proper configurational averaging is important for making contact with experimental observables measured in solution such as redox potentials, and it may be important for other properties as well. We are, therefore, in the process of implementing ligand field molecular dynamics (LFMD).

The LFMM gives essentially the same structures as QM/MM but several orders of magnitude faster and without the issue of how to join the quantum region to the classical region. In the LFMM, the metal center is placed on the same footing as all other atoms, giving a uniform theoretical treatment across the entire system. For the oxidized T1 center, the LFMM captures the important electronic effects arising from the metals' d^9 configuration and thus appears to be a viable alternative to QM/MM. To the author's knowledge, this is the first empirically based model capable of treating all three commonly observed variants of T1 copper sites using a single set of transferable parameters.

The successful validation of the LFMM parameters presented here gives us confidence that subsequent LFMD simulations will be soundly based.

Acknowledgment. The author acknowledges the EPSRC Chemical Database Service, Daresbury Laboratories, UK for access to the Cambridge Structural Database and Dr Christopher Dennison for many enlightening discussions.

Supporting Information Available: Cartesian coordinates for the molecules shown in Figure 2 in *XYZ* format—one file for all five LFMM structures and one for all five exp/DFT structures. AMBER94 MOE force field parameters specific to oxidized Type 1 centers. LFMM parameters—these require the DommiMOE executable available from the author on request. Full listing comparing Cu−L distances and L−Cu−L angles derived from PDB structures and LFMM calculations. Complete author list for ref 39. This material is available free of charge via the Internet at http://pubs.acs.org.

IC062399J

(61) Ammeter, J. H.; Burgi, H. B.; Gamp, E.; Meyer-Sandrin, V.; Jensen, W. P. *Inorg. Chem.* **1979**, *18*, 733−750.