J|A|C|S

A R T I C L E S

Frozen Density Functional Free Energy Simulations of Redox Proteins: Computational Studies of the Reduction Potential of Plastocyanin and Rusticyanin

Mats H. M. Olsson,* Gongyi Hong, and Arieh Warshel*

Contribution from the Chemistry Department, University of Southern California, 3620 McClintock Avenue, SGM 418, Los Angeles, California 90089-0047

Received September 26, 2002; E-mail: molsson@usc.edu; warshel@usc.edu

Abstract: The evaluation of reduction potentials of proteins by ab initio approaches presents a major challenge for computational chemistry. This is addressed in the present investigation by reporting detailed calculations of the reduction potentials of the blue copper proteins plastocyanin and rusticyanin using the QM/MM all-atom frozen density functional theory, FDFT, method. The relevant ab initio free energies are evaluated by using a classical reference potential. This approach appears to provide a general consistent and effective way for reproducing the configurational ensemble needed for consistent ab initio free energy calculations. The FDFT formulation allows us to treat a large part of the protein quantum mechanically by a consistently coupled QM/QM/MM embedding method while still retaining a proper configurational sampling. To establish the importance of proper configurational sampling and the need for a complete representation of the protein+solvent environment, we also consider several classical approaches. These include the semi-macroscopic PDLD/S-LRA method and classical all-atom simulations with and without a polarizable force field. The difference between the reduction potentials of the two blue copper proteins is reproduced in a reasonable way, and its origin is deduced from the different calculations. It is found that the protein permanent dipole tunes *down* the reduction potential for plastocyanin compared to the active site in regular water solvent, whereas in rusticyanin it is instead tuned *up*. This electrostatic environment, which is the major effect determining the reduction potential, is a property of the entire protein and solvent system and cannot be ascribed to any particular single interaction.

Introduction

Theoretical modeling of protein function in general, and metalloproteins in particular, presents a major challenge for modern computational chemistry. One seemingly reasonable approach is to use an accurate quantum mechanical representation of the active site by a pure high-level ab initio representation of the active site either in a vacuum or with some simplified solvation model. Unfortunately, such a treatment is likely to provide unreliable results since outer-sphere interactions contribute in a major way to the energy changes associated with protein function. Clearly, it is not sufficient to treat the active site with a high-level ab initio method, even though considerable computer time is spent on including electron–electron correlation, when the environment is neglected or, at best, represented in an oversimplified way. Further, it is tempting, and in principle possible, to evaluate the quantum mechanical energy for the active site and the protein in a single protein configuration. However, proteins are not rigid or static systems but highly flexible and assume many different configurations at ambient temperature. Obviously, the potential energy surface for different functional processes of the protein can depend strongly on protein configurations. This makes it essential to average the

energetics of the given process over a substantial amount of protein configurations. Clearly, an accurate and complete quantum mechanical investigation of protein function must balance rigorously the treatment of the active site with a reasonable treatment of the solvation of the active site by its protein and solvent environment.

Metalloproteins can also be studied by more or less elaborate force field methods (see refs 1–3 for representative studies) where both the solute and the solvent are represented by classical point charges and comparatively simple analytical functions. Nuclear quantum mechanical effects can be incorporated in these approaches on different levels of sophistication (see ref 4 for a recent overview of the field). Correctly used, force field methods treat the environment consistently and several interesting questions can be addressed correctly. However, if reliable results are to be obtained for inner-sphere properties, solute parameters such as equilibrium bond distances, charges, force constants, etc., need to be parametrized so that they reproduce observed experimental data. Thus, it is still desirable to improve the description of the system by using a more rigorous ab initio

---

* Corresponding authors. E-mail: (M.H.M.O.) molsson@usc.edu; (A.W.) warshel@usc.edu.

(1) Comba, P. In *Molecular Modeling and Dynamics of Bioinorganic Systems*; Banci, L., Comba, P., Eds.; Kluwer Academic Publishers: P.O.B. 17, 3300 AA Dordrecht, The Netherlands, 1997; pp 21–47.
(2) Comba, P.; Remenyi, R. *J. Comput. Chem.* **2002**, *23*, 697–705.
(3) Åqvist, J.; Warshel, A. *J. Am. Chem. Soc.* **1990**, *112*, 2860.
(4) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467–505.

quantum mechanical approach. Such a density or wave function based representation is more likely to give the correct physical behavior and increases the predictive nature of computer simulation approaches. Therefore, one of the main objectives of the present investigation is to further extend simulation techniques to a more rigorous ab initio quantum mechanical description of the reacting system and to examine how this compares to classical MD simulations. For this purpose, we use a QM/MM approach (see, e.g., refs 5−10) or more precisely a QM/QM/MM approach.[11,12] This approach treats the reaction center and its closest residues and water molecules by an ab initio approach (the central part is treated by a more rigorous approach than the surrounding), whereas the protein residues further away are represented with a classical force field. The quantum mechanical region is treated by the frozen density functional method, FDFT.[13] This method provides one of the most promising option of treating the interface between the QM and MM regions since the frozen part is represented by electron densities rather than point charges and the coupling between the frozen and unfrozen parts is evaluated by a consistent DFT formulation. Since these densities are frozen when we evaluated the active site QM energy and since the interactions between the frozen parts are evaluated classically, a considerable amount of computer time is saved. Thus, the FDFT approach allows us to treat much larger parts of the protein on an ab initio QM level than regular quantum mechanical approaches, yet still allows us to average the properties over a considerable amount of configurations. It should also be noted that the term frozen in FDFT refers to the density and not to the nuclear position; that is, the frozen region is free to move.

Obtaining the free energy correctly, as opposed to just calculating the energy for a single protein configuration, is a very difficult, but crucial task in QM/MM calculations of proteins. Not uncommonly, protein investigations using ab initio QM/MM approaches usually only consider one single configuration where, at best, the system energy has been minimized. Therefore, another main objective of this investigation is to show the importance of configurational averaging for both classical and quantum mechanical approaches. Unfortunately, it turns out to be computationally too expensive even with the FDFT approach to do a meaningful average over the protein configurations needed for proper convergence of the calculated free energies with correct QM/MM gradients. Instead, we have exploited our previous approach of using a classical reference potential for the QM/MM calculations.[14−16] Previously, the FDFT approach has been shown to give very promising results for simpler systems (see, e.g., ref 17) but to the best of our knowledge, this is the first study where it has been used to investigate proteins.
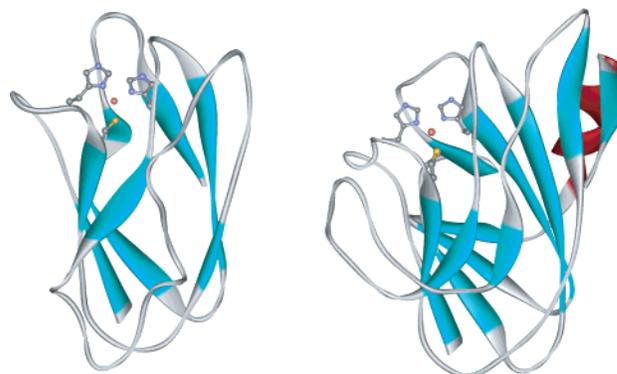
Figure 1. The two blue copper proteins considered in this study: plastocyanin and rusticyanin (PDB IDs 5pcy and 1rcy, respectively, in the Brookhaven data bank). Even though both these proteins have the same type 1 copper active site, their reduction potential differs by more than 300 mV (375 and 680 mV, respectively).

To examine the reliability of the FDFT QM/MM free energy approach, we have studied the reduction potential of the blue copper proteins plastocyanin and rusticyanin, Figure 1. These two proteins have very similar active sites with identical copper ligands, but drastically different reduction potentials: 375 mV for plastocyanin[18,19] and 680 mV for rusticyanin.[20,21] This makes them an ideal case to apply this method to since the error in the quantum mechanical calculation of the active site to a great extent cancels when taking the difference, and thus, we can concentrate on the outer-sphere interactions. Further, these proteins are comparatively small, have a rather small active site, and are well characterized.

The blue copper proteins constitute a group of electron transfer proteins that are characterized by a number of unusual properties, e.g., a bright blue color, a narrow hyperfine splitting in the electronic spin resonance spectra, and an extraordinarily high reduction potential for the Cu^+/Cu^{2+} pair.[22−24] Moreover, crystal structures of the oxidized form of these proteins show a ligand structure distinct from what is normally observed for small inorganic complexes: the copper ion is bound to the protein in an approximate trigonal plane formed by a cysteine thiolate group and two histidine nitrogen atoms. The coordination sphere in most blue copper proteins is completed by one or two axial ligands, typically a methionine thioether group. This trigonal active-site geometry is close to a tetrahedron, which is usually the overall preferred geometry for Cu(I) complexes. Further, the Cu(I)−S_{Cys} distance is unexpectedly short compared to normal Cu(I) complexes and is thought to be of functional significance. Naturally, features that make the active-site geometry similar in the two oxidation states are of functional advantage for an electron transfer protein; since the copper center has similar structure for the two oxidation states, the reorganization energy will be low and the rate of electron transfer high.

(5) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *106*, 421.
(6) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(7) Friesner, R.; Beachy, M. D. *Curr. Opin. Struct. Biol.* **1998**, *8*, 257−262.
(8) Monard, G.; Merz, K. M. *Acc. Chem. Res.* **1998**, *32*, 904−911.
(9) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112* (8), 3483−3492.
(10) Field, M. *J. Comput. Chem.* **2002**, *23*, 48−58.
(11) Vaidehi, N.; Wesolowski, T. A.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 4264.
(12) Vreven, T.; Morokuma, K. *J. Chem. Phys.* **2000**, *113*, 2969.
(13) Wesolowski, T. A.; Warshel, A. *J. Phys. Chem.* **1993**, *97*, 8050.
(14) Bentzien, J.; Muller, R. P.; Florián, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293−2301.
(15) Muller, R. P.; Warshel, A. *J. Chem. Phys.* **1995**, *99*, 17516.
(16) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199.
(17) Hong, G.; Strajbl, M.; Wesolowski, T. A.; Warshel, A. *J. Comput. Chem.* **2000**, *21*, 1554−1561.

(18) Guss, J. M.; Harrowell, P. R.; Murata, M.; Norris, V. A.; Freeman, H. C. *J. Mol. Biol.* **1986**, *192*, 361−387.
(19) Guss, J. M.; Bartunik, H. D.; Freeman, H. C. *Acta Crystallogr.* **1992**, *B48*, 790−807.
(20) Walter, R. L.; Ealick, S. E.; Friedman, A. M.; Blake, R. C.; Proctor, P.; Shoham, M. *J. Mol. Biol.* **1996**, *263*, 730−751.
(21) Harvey, I.; Hao, Q.; Duke, E. M. H.; Ingledew, W. J.; Hasnain, S. S. *Acta Crystallogr.* **1998**, *D54*, 629−635.
(22) Sykes, A. G. *Adv. Inorg. Chem.* **1990**, *36*, 377−408.
(23) Adman, E. T. *Adv. Protein Chem.* **1991**, *42*, 145−197.
(24) Messerschmidt, A. *Struct. Bonding* **1998**, *90*, 37−68.

Interestingly, blue copper proteins with the typical CuCysHis$_2$-Met active site have reduction potentials ranging from 260 (amicyanin) to 680 mV (rusticyanin). Including blue copper proteins with less typical active sites widens the range to cover reduction potentials from 184 mV for stellacyanin,[25] which has a glutamine amide oxygen as the axial ligand, to 1000 mV for the type 1 copper site in domain 2 of ceruloplasmin,[26] where the axial ligand is replaced by a noncoordinating residue.

The unusual properties of the blue copper proteins have traditionally been explained by protein strain; that is, the structure of the active site is perturbed from its optimal geometry by the protein, thereby altering its chemical and physical properties.[27,28] Moreover, the oxidized form of the active site is forced to adopt a geometry similar to the optimal geometry for the reduced form, thus destabilizing the oxidized state compared to the reduced and raising the reduction potential. However, during the last 5−10 years, theoretical[29−31] as well as experimental[32−34] investigations have provided an alternative picture of the blue copper active site. Quantum mechanical geometry optimizations of active-site model complexes both in a vacuum[35] and in the protein[36] have invariably given structures that overall have a geometry that is very close to what is found experimentally in the protein. The Cu−S$_{Met}$ distance has not been reproduced by these optimizations, but since this interaction is comparatively weak, its equilibrium distance seems to be largely determined by outer-sphere interactions that cannot be captured by this simplified model. Previous investigations have addressed the structural effects, including the Cu−S$_{Met}$ distance, of the blue copper protein on its active site.[2,37] However, as is argued in later sections, it is more important to calculate the energetics correctly rather than obtaining structures that are identical to that obtained by X-ray crystallography. In this investigation, we have applied a small energy constraint to the Cu−S$_{Met}$ bond distance in order to reproduce its experimental distance. This does not influence the result presented in this article. A recent NMR study also indicates that this interaction is not as influential in rusticyanin as might have been expected.[38] In any case, there seems to be an alternative explanation for the unusual properties of the blue copper active site. Though strained models of protein function are still viable in different forms,[39,40] lately, more emphasis has been put on the electrostatic differences in the environment provided by the proteins.[29,41]

In this work, we present a systematic investigation of the difference in reduction potential between plastocyanin and rusticyanin comparing different state of the art simulation techniques. The approaches considered includes all-atom classical linear response approximation (LRA[42]) simulations running long trajectories, PDLD/S-LRA[42−44] calculations, and the use of classical reference potentials in evaluating the FDFT QM/MM free energy.[45] This comparative study allows us to assess the performance of the QM/MM approach and the crucial role of proper configurational averaging and the conditions for proper convergence of the reduction potential. We also use this systematic investigation to examine the protein control of the reduction potential in the blue copper proteins.

Methods

LRA Formulation as a General Way for Evaluating Redox Energies. The relationship between the free energy change and the emf for an electrochemical cell is given by

$$\Delta G = -nF\epsilon \tag{1}$$

Here, F is the Faraday constant and n the number of electrons involved in the redox process. Thus, the reduction potential of proteins can be expressed relative to the corresponding reduction potential in water in terms of its free energy change as

$$\epsilon_{wat}^{\circ} - \epsilon_{wat}^{\circ} = -\frac{\Delta G_{Ox \to Red}^{\circ,p} - \Delta G_{Ox \to Red}^{\circ,w}}{nF} \tag{2}$$

Dropping the $_{Ox \to Red}$ subscript, the free energy difference between the active site in the protein and in water can be divided into intra-redox site, solvation and "solvated electron" contributions and expressed as

$$\Delta\Delta G^{\circ,w \to p} = (\Delta G_{intra}^{\circ,p} + \Delta G_{solv}^{\circ,p} + \Delta G_{e^-}^{\circ,w}) - (\Delta G_{intra}^{\circ,w} + \Delta G_{solv}^{\circ,w} + \Delta G_{e^-}^{\circ,w}) \tag{3}$$

The term $\Delta G_{e^-}^{\circ,w}$ contains the free energy change of the other half-cell reaction, usually taken as the standard hydrogen electrode, but cancels here by formulating the problem as a reduction potential difference.[46]

$$\epsilon_{pro}^{\circ} = \epsilon_{wat}^{\circ} - \frac{\Delta\Delta G_{intra}^{\circ,w \to p}}{nF} - \frac{\Delta\Delta G_{solv}^{\circ,w \to p}}{nF} \tag{4}$$

Similarly, the difference in reduction potential between two proteins can be written as

$$\Delta\epsilon^{\circ} = -\frac{\Delta\Delta G_{intra}^{\circ,p_1 \to p_2}}{nF} - \frac{\Delta\Delta G_{solv}^{\circ,p_1 \to p_2}}{nF} \tag{5}$$

When one considers the difference in reduction potential between two proteins with the same redox site, it is possible to neglect the contribution from $\Delta\Delta G_{intra}^{\circ,p_1 \to p_2}$ if the coupling between the inner and outer reorganization is small. Here, we have assumed that the coupling

(25) Romero, A.; Hoitink, C. W. G.; Nar, H.; Huber, R.; Messerschmidt, A.; Canters, G. W. *J. Mol. Biol.* **1993**, *229*, 1007−1021.
(26) Zaitseva, I.; Zaitsev, V.; Card, G.; Moshkov, K.; Bax, B.; Ralph, A.; Lindley, P. *J. Biol. Inorg. Chem.* **1996**, *1*, 15−23.
(27) Vallee, B. L.; Williams, R. J. P. *Proc. Nat. Acad. Sci.* **1968**, *59*, 498−505.
(28) Malmström, B. G. *Eur. J. Biochem.* **1994**, *223*, 207−216.
(29) Ryde, U.; Olsson, M. H. M.; Pierloot, K.; Roos, B. O. *J. Mol. Biol.* **1996**, *261*, 586−596.
(30) Olsson, M. H. M.; Ryde, U. *J. Biol. Inorg. Chem.* **1999**, *5*.
(31) Ryde, U.; Olsson, M. H. M.; Roos, B. O.; De Kerpel, J. A. O.; Pierloot, K. *J. Biol. Inorg. Chem.* **2000**, *5*, 565−574.
(32) Winkler, J. R.; Wittung-Stafshede, P.; Leckner, J.; Malmström, B. G.; Gray, H. B. *Proc. Nat. Acad. Sci.* **1997**, *94*, 4246−4249.
(33) Leckner, J.; Wittung, P.; Bonander, N.; Karlsson, B. G.; Malmström, B. G. *J. Biol. Inorg. Chem.* **1997**, *2*, 368−371.
(34) Wittung-Stafshede, P.; Hill, M. G.; Gomez, E.; Di Bilio, A. J.; Karlsson, B. G.; Leckner, J.; Winkler, J. R.; Gray, H. B.; Malmström, B. G. *J. Biol. Inorg. Chem.* **1998**, *3*, 367−370.
(35) Ryde, U.; Olsson, M. H. M.; Roos, B. O.; Carlos-Borin, A. *Theor. Chem. Acc.* **2001**, *105*, 452−462.
(36) Ryde, U.; Olsson, M. H. M. *Int. J. Quantum Chem.: Biophys. Quatern.* **2001**, *81*, 335−347.
(37) Comba, P.; Lledós, A.; Maseras, F.; Remenyi, R. *Inorg. Chim. Acta* **2001**, *324*, 21−26.
(38) Donaire, A.; Jiménez, B.; Moratal, J. M.; Hall, J. F.; Hasnain, S. S. *Biochemistry* **2001**, *40*, 837−846.

(39) Guckert, J. A.; Lowery, M. D.; Solomon, E. I. *J. Am. Chem. Soc.* **1995**, *117*, 2817−2844.
(40) Malmström, B. G.; Leckner, J. *Curr. Opin. Chem. Biol.* **1998**, *2*, 286−292.
(41) Libeu, C. A. P.; Kukimoto, M.; Nishiyama, M.; Hornouchi, S.; Adman, E. T. *Biochemistry* **1997**, *36*, 13160−13179.
(42) Lee, F. S.; Chu, Z. T.; Warshel, A. *J. Comput. Chem.* **1993**, *14*, 161−185.
(43) Langen, R.; Brayer, G. D.; Berghuis, A. M.; McLendon, G.; Sherman, F.; Warshel, A. *J. Mol. Biol.* **1992**, *224*, 589−600.
(44) Sham, Y. Y.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 4458−4472.
(45) Wesolowski, T. A.; Muller, R. P.; Warshel, A. *J. Chem. Phys.* **1996**, *100*, 15444.
(46) Churg, A. K.; Warshel, A. *Biochemistry* **1986**, *25*, 1675.
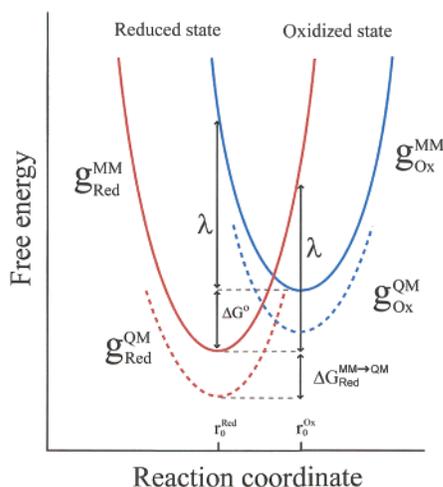
Figure 2. The microscopic free energy functionals, g_{Ox} and g_{Red}, for the classical and quantum mechanical potential energy surface are defined by and obtained from the reorganization energy, λ, and the free energy difference, $\Delta G°$. The LRA free energy difference, eq 7, can easily be derived from the vertical energy differences, $(g_{Red} - g_{Ox})_{r_0^{Red}} = \Delta G° + \lambda$ and $(g_{Red} - g_{Ox})_{r_0^{Ox}} = \Delta G° - \lambda$, assuming that the λ is the same for both oxidation states. The quantum mechanical free energy difference is given similarly by a classical reference potential and the perturbation $\Delta G^{MM \to QM}$.

is indeed small and focus on the second term of eq 5. Thus, we have neglected the sterical effect that the outer sphere might exert on the inner sphere (note that the electrostatic effect was not neglected). This steric coupling would be an interesting issue to address with quantum mechanical optimizations, but this is, however, not a major point of the present study and is clearly out of the scope of this investigation. Our approach seems also especially justified considering that the reaction coordinate mainly constitutes the solvent coordinate. Clearly, it would also be interesting to address the issue about protein strain on the $Cu-S_{Met}$ distance, but this can only partly be addressed since we have included only the electrostatic interaction from the methionine residue on the redox site. The intra-cluster energy term $\Delta\Delta G_{intra}^{o,p_1 \to p_2}$ should, strictly speaking, also include the difference in the polarization energy of the active sites due to their different environment. Though it has not been included in the classical approaches, we found it necessary to include in the quantum mechanical approach since the polarization of the active sites can be quite different in the two proteins. At any rate, our main task is to calculate the change in solvation energy during the redox process according to

$$\Delta\epsilon° = - \frac{\Delta\Delta G_{solv}^{o,p_1 \to p_2}}{nF} \qquad (6)$$

In the present work we evaluate $\Delta G°$ for plastocyanin and rusticyanin in the linear response (LRA) adiabatic charging (AC) framework. The LRA representation for free energy calculations was originally derived by inspection of the microscopic nature of the free energy functional,[42,43,47,48] which corresponds to Marcus parabolas in the macroscopic limit. In the LRA approximation the free energy of a redox process is given by eq 7 (see also the solid lines in Figure 2, which refer to the classical potential energy surface).

$$\Delta G° = \frac{1}{2}(\langle V_{Red} - V_{Ox} \rangle_{Red} + \langle V_{Red} - V_{Ox} \rangle_{Ox}) \qquad (7)$$

Here, $\langle V_{Red} - V_{Ox} \rangle_{Red}$ denotes the average, vertical, energy difference between the reduced and oxidized states evaluated at the optimal

(47) Lee, F. S.; Chu, Z. T.; Bolger, M. B.; Warshel, A. *Protein Eng.* **1992**, *5*, 215.
(48) King, G.; Warshel, A. *J. Chem. Phys.* **1990**, *93*, 8682−8692.
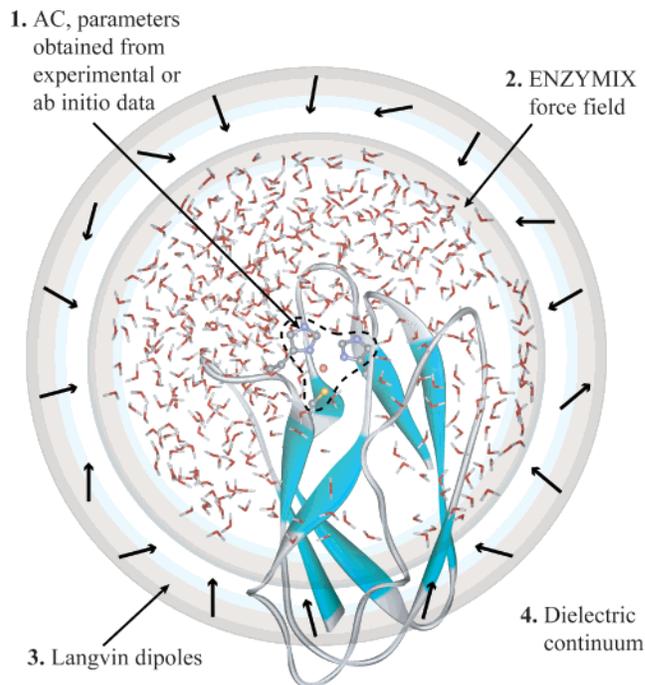
Figure 3. The simulated system is spherical and contains four regions: region 1, the blue copper active site where the charge change is confined during the redox process; region 2, the explicit protein+solvent environment responding to the change in oxidation state of region 1; regions 3 and 4, 2 Å Langevin dipoles and a dielectric continuum, representing the bulk solvation. The explicit solvent in region 2 is subject to a spherical surface constraint.[42]

geometry for the reduced state, i.e., running on the reduced potential energy surface. Thus, we have to run two simulations, one for each electronic state, for each protein in order to get the reduction potential difference. The LRA expression can also be considered as an end-point approximation for the corresponding free energy perturbation (FEP) approach.[49] In principle, the full FEP approach should be more rigorous, but in cases with large solutes, the LRA often outperforms the FEP approach. Furthermore, the LRA approach provides a direct link to key physical parameters such as the reorganization energy, λ, and the free energy change, $\Delta G°$.[50] In the LRA approach, the free energy contributions are additive and are easily applied to different potential energies ranging from the purely classical semi-macroscopic protein dipole Langevin dipole, PDLD/S, method (e.g., refs 44, 51) to all-atom quantum mechanical approaches (e.g., ref 52).

The Classical Simulations. All simulations have been performed according to standard simulation techniques starting from crystal structures, 5pcy for plastocyanin[18] and 1rcy for rusticyanin[20] in the Brookhaven data bank, and allowing the trajectories to evolve in the MD runs according to the given oxidation state. Fundamental for all methods used in this study is the standard surface constrained all-atom solvent (SCAAS) simulation technique with the local reaction field (LRF) long-range treatment as implemented in the MOLARIS program package.[53] The total system (protein and solvent) is spherical and divided into four regions according to Figure 3. The central part of the protein, region 1, contains the blue copper active site. Region 2 comprises the unconstrained protein atoms and explicit water molecules up to a given radius and is represented with the standard ENZYMIX

(49) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004−2010.
(50) Muegge, I.; Qi, P. X.; Wand, A. J.; Chu, Z. T.; Warshel, A. *J. Phys. Chem. B* **1997**, *101*, 825−836.
(51) Stephens, P. J.; Jollie, D. R.; Warshel, A. *Chem. Rev.* **1996**, *96*, 2491.
(52) Strajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106(51)*, 13 333− 13 343.
(53) Chu, Z. T.; Villa, J.; Strajbl, M.; Schutz, C. N.; Shurki, A.; Warshel, A. MOLARIS, version beta9.05; 2002.

Table 1. Mulliken Charge on the Small System

	standard[a]			plastocyanin		rusticyanin	
	q_{Mull}^{Ox}	q_{Mull}^{Red}	δq_{Mull}	Ox	Red	Ox	Red
Cu	0.384	0.229	0.155	0.387	0.244	0.379	0.215
Cys							
C_β	0.213	0.145	0.068	0.221	0.150	0.205	0.141
$H_{\beta 1}$	−0.018	−0.060	0.042	−0.010	−0.065	−0.026	−0.056
$H_{\beta 2}$	−0.040	−0.069	0.029	−0.038	−0.065	−0.042	−0.072
S_γ	−0.066	−0.435	0.369	−0.061	−0.438	−0.071	−0.432
Imidazole 1							
C_γ	0.068	0.016	0.052	0.044	0.006	0.092	0.026
$N_{\delta 1}$	−0.383	−0.335	−0.048	−0.382	−0.309	−0.383	−0.361
$C_{\delta 2}$	0.097	0.050	0.047	0.039	0.000	0.155	0.099
$H_{\delta 2}$	0.033	0.018	0.015	0.049	0.049	0.017	−0.014
$C_{\epsilon 1}$	0.194	0.146	0.048	0.206	0.137	0.181	0.155
$H_{\epsilon 1}$	0.022	−0.016	0.038	0.014	−0.019	0.029	−0.013
$N_{\epsilon 2}$	0.209	0.205	0.004	0.193	0.186	0.225	0.223
$H_{\epsilon 2}$	0.029	0.002	0.027	0.033	0.008	0.024	−0.005
Imidazole 2							
C_γ	0.077	0.034	0.043	0.086	0.032	0.068	0.036
$N_{\delta 1}$	−0.376	−0.332	−0.044	−0.358	−0.346	−0.393	−0.318
$C_{\delta 2}$	0.103	0.063	0.040	0.110	0.076	0.096	0.050
$H_{\delta 2}$	0.043	0.029	0.014	−0.006	−0.020	0.093	0.079
$C_{\epsilon 1}$	0.172	0.129	0.043	0.200	0.173	0.145	0.084
$H_{\epsilon 1}$	0.008	−0.011	0.019	0.020	−0.001	−0.004	−0.020
$N_{\epsilon 2}$	0.218	0.205	0.013	0.237	0.212	0.199	0.198
$H_{\epsilon 2}$	0.013	−0.013	0.026	0.016	−0.010	0.011	−0.015

[a] The Met92/144 is in region 2 and is represented by standard ENZYMIX charges.

force field parameters.[53] The remaining atoms further away than the region 2 radius are kept fixed at their initial position during the simulation. Finally, the system is embedded in a 2 Å shell of Langevin dipoles and a dielectric continuum simulating the bulk solvation.

In general, we have used the system $CuSCH_2(C_3H_3N_2)_2$ to represent the active site, i.e., our region 1, in all calculations. Here, the histidine residues have been represented by their ring systems and the cysteine have been represented with the atoms up to the α carbon. We have used Mulliken[54] charges obtained from quantum mechanical calculations according to the procedure described later in this section to represent region 1 in the classical simulations. It is worth pointing out already here, though, that the polarization of the active site due to the electric field from the environment has been included when calculating these Mulliken charges. These charges are referred to as the standard q_{Mull}^{Ox} and q_{Mull}^{Red} charges and are presented in Table 1.

The LRA simulation procedure that seems to give the most stable results and that we have adopted in most cases is as follows: The simulations on the reduced and oxidized potential energy surface were performed at 300 K with 1 fs time steps using the SCAAS/LRF simulation procedure. Initially, we ran a 50 ps long trajectory with a region 2 radius of 24 Å in order to equilibrate the protein and get 10 starting structures for the simulation (atoms further away than 24 Å are kept fixed at their original X-ray structure). Then, starting from these structures, we ran the actual simulation with a region 2 radius of 18 Å. Before acquiring the $\langle V_{Red} - V_{Ox} \rangle$ averages needed for eq 7, each of these 10 systems was further equilibrated for 500 ps to adjust for the new system condition. Finally, the $(V_{Red} - V_{Ox})$ energy difference was collected at each 10th fs and combined from the 10×400 ps simulations to give an average over 400 000 configurations for the reduced and oxidized states. The simulated 18 Å systems include most parts of the proteins and 500 and 400 water molecules for plastocyanin and rusticyanin, respectively. These water molecules are crucial to get a correct description of the interaction energies and the final result. By using this procedure of spawning 10 simulations, we can in a simple and efficient way take advantage of today's parallel computers. Also, by using a rather large radius of the initial simulation, we ensure that the outer region of the protein (the atoms between 18 and 24 Å from

the simulation center that are free to move in the initial setup but fixed in the final production run) is not fixed at a unrelaxed position and thereby imposes strained conformations of the residues close to the edge of the system. Finally, it should also improve the water penetration.

Since the closest ionizable residues were found to be about 10 Å from the copper ion, we treated the effect of these groups, $\Delta\epsilon_{N\rightarrow I}^\circ$, macroscopically. That is, we followed our standard two-step thermo-dynamic cycle. First we evaluated the reduction potential microscopi-cally, keeping all ionizable residues in their neutral state, $\epsilon_{Q\mu}^\circ$ (here the subscript $Q\mu$ indicates the interaction between the residual charges of region 1 and the protein dipoles of region 2). In the second step we turned on the charges and evaluated their effect using a large screening of the corresponding Coulombic interactions ($\epsilon_{eff} = 60$). Then, the total classical reduction potential is given by

$$\epsilon^\circ = \epsilon_{Q\mu}^\circ + \Delta\epsilon_{N\rightarrow I}^\circ \tag{8}$$

where the first term is calculated according to eq 7 and the second term is calculated from

$$\Delta\epsilon_{N\rightarrow I}^\circ = -\frac{\Delta G_{N\rightarrow I}^{\circ,Red} - \Delta G_{N\rightarrow I}^{\circ,Ox}}{nF} \tag{9}$$

The reliability and justification of this approach has been established elsewhere.[55,56]

The PDLD/S-LRA calculations were performed according to stan-dard procedures,[42] generating configurations for the reduced and oxidized form of the protein with MD simulations using the ENZYMIX force field. The final reduction potential is then obtained with the LRA formulation using the PDLD/S electrostatic energy averaged over these configurations. In this investigation, we have used 20 configurations from a 1000 ps long simulation.

The QM/MM Surface. For the quantum mechanical part of our QM/MM treatment, we have chosen frozen density functional theory (FDFT). This approach allows us to treat large parts of the protein quantum mechanically while retaining proper configurational sampling.

(54) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.

(55) Johnson, E. T.; Parson, W. W. *Biochemistry* **2002**, *41*, 6483−6494.
(56) Schutz, C. N.; Warshel, A. *Proteins: Struct. Funct. Genet.* **2001**, *44*, 400−417.
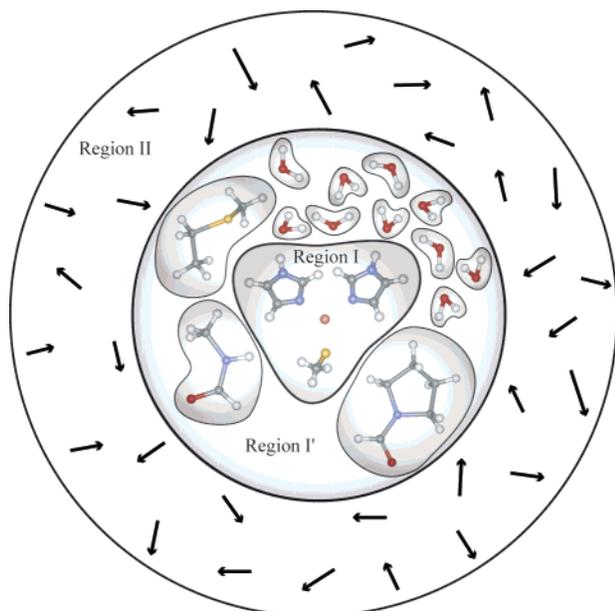
Figure 4. In the QM/MM approaches, region I is treated with regular full DFT, whereas region I′ is represented by frozen densities. Region II, which is depicted schematically by dipoles, is represented by a classical force field. Region I′ is coupled to region I by the FDFT formulation, whereas region II is coupled to region I according to standard QM/MM formulations using point charges. Note that this is a schematic picture to show the division into regions and fragments in the FDFT approach and does not reflect the actual protein + solvent system.

Though using a purely quantum mechanical method is very attractive, the use of a proper sampling is just as important when studying protein function since the protein assumes many different configurations that are close in energy and usually have very different values for the energies associated with the specific function.

The FDFT method divides the system into three regions (see Figure 4): region I contains the active site, in our case the copper ion and its closest ligands (Im₂CuSCH₃), and is treated with a normal high-level DFT calculation; region I′ typically contains the second coordination sphere, residues, and water molecules that are close or comparatively strongly interacting, and needs to be represented by densities; and finally, in region II, which contains the majority of the protein residues and water molecules, the explicit environment is treated as classical point charges. The division between the QM and MM regions in various QM/MM approaches is a well-known problem that has been considered in numerous studies (for a more recent review see ref 57). The FDFT approach minimizes this problem. Further, in the present investigation, the connection between regions I and I′ involves only the C_β–C_γ bonds for the two coordinating histidine residues and the C_α–C_β bond for the cysteine residue. These are rather far from the region that undergoes large chemical change, which implies that the active site retains its physical properties and the error due to the charge transfer between the innermost regions is minimized. Further, previous experience with CDFT has shown that similar results are obtained also with different degrees of charge transfer.[17] Therefore, our treatment should be reasonable also according to the rather strict rules in the recent review by Bersuker.[57]

The density of the total system is expressed according to

$$\rho(\mathbf{r}) = \rho_I(\mathbf{r}) + \rho_{I',1}(\mathbf{r}) + \rho_{I',2}(\mathbf{r}) + ... + q_{II,1}(\mathbf{r}) + q_{II,2}(\mathbf{r}) + ... \quad (10)$$

and the overall energy of region I and its surroundings is evaluated with the FDFT formulation (here the subscripts 1, 2, etc. designate the fragment numbering).[13] This is done using the nonadditive kinetic
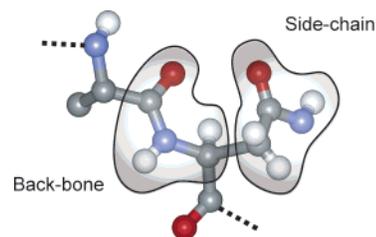
Figure 5. The protein parts included in region I′ have in general been divided into two regions, one backbone and one side-chain fragment. The backbone fragment has not been divided according to its residue numbering but, as shown in the figure, into CONHCH₂ fragments.

energy functional for the interaction between region I and region I′. The remaining interaction between region I and region II is done according to standard QM/MM treatment. Further technical details of the FDFT method is described in length elsewhere.[52]

In general, we have divided the protein into fragments such that each residue is represented by two fragments, one for the backbone atoms and one for the side-chain atoms, as is outlined in Figure 5. Gly, Ala, and Pro, however, were kept as entire residues in all cases. Fragments closer than 6 Å from the region I center were truncated with one hydrogen for each "broken bond", to make a closed shell system, and represented as a region I′ density. It might seem a bit aggravating to cut these bonds and make two presumably strongly interacting fragments. However, it should not pose any considerable problems since the interaction between fragments in region I′ is always treated classically. Fragments further than 6 Å from the region I center were kept as point charges since their interaction with the active site seems to be adequately described with this simplified model. The backbone fragments have not been divided in accordance with their residue numbering, but as CH_3NHCHO fragments, which can be seen in Figure 5. This division seems natural since it does not cut through the peptide partial double bond and does not create any additional polar NH bonds. To us, this division seems chemically sound and gives the best performance versus accuracy, and initial test calculations of interaction energies were more similar to full DFT calculations than other procedures. Then, the calculation proceeds as follows:

(1) The density of region I is calculated with the remaining atoms (regions I′ + II) given as point charges.

(2) The density of each region I′ fragment is calculated with region I as density and with the rest (the remaining region I′ fragments + region II) given as point charges.

(3) The density of region I is recalculated with all region I′ fragments as densities and region II as point charges.

In principle, it is necessary to repeat steps 2 and 3 until all densities are self-consistent, the so-called freeze and thaw method. However, initial test calculations showed that the noniterative procedure is sufficient if the point charges in the initial step are a good approximation to the density of the frozen fragments.

The FDFT approach has been shown in previous studies to give reliable results, e.g., ref 58. Here, we present another system that shows that the FDFT is capable of giving results in good agreement with regular full DFT. In Figure 6, we compare the potential energy surfaces for two hydrogen-bonded systems using FDFT with regular DFT. In all cases the hydrogen bond acceptor is a water molecule and constitutes region I. The hydrogen bond donor is CH₃OH for the red curve and CH₃SH for the green curve and is treated as the frozen density, i.e., region I′. These weakly interacting systems are representative of the type of interactions we intend to treat with the frozen density functional theory.

We have used a modified version of the DEMON program package[59,60] to calculate the quantum mechanical energies where the frozen density

(57) Bersuker, I. B. In *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 2001; pp 69–135.

(58) Wesolowski, T. A.; Goursot, A.; Weber, J. *J. Chem. Phys.* **2001**, *115*, 4791–4797.
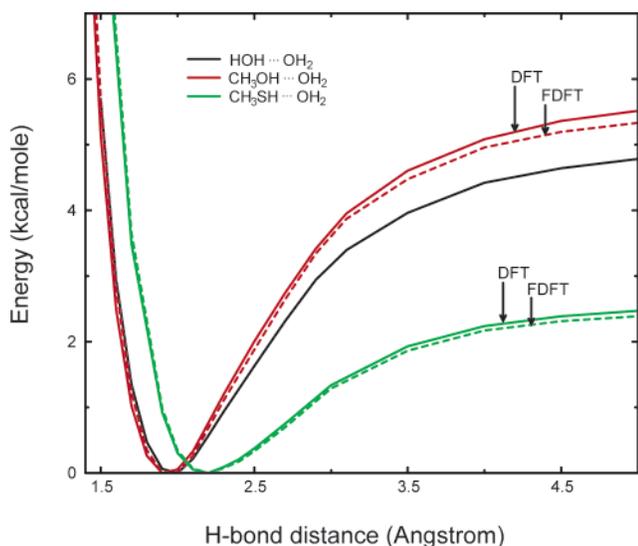
Figure 6. FDFT potential energy surface for two hydrogen-bonded systems, where the water molecule constitutes region I and the hydrogen-bond donor is in region I′, the frozen density.

functional method has been implemented such that we can represent the amino acids surrounding the active site as individual frozen densities. The Becke88 exchange potential and Perdew86 correlation potential was adopted in all calculations. The nonlocal kinetic energy functional developed by Wesolowski was used for the nonadditive kinetic energy term. In region I we have used the (63321/5211*/41+) basis set for copper, (73111/6111/1*) for sulfur, (5211/411/1) for oxygen, nitrogen, and carbon, and (41/1) for hydrogen, whereas for the frozen density fragments we have used the smaller (6321/521) basis set for sulfur, (521/41) for oxygen, nitrogen, and carbon, and (41) for hydrogen.

Calculating the QM/MM Free Energy. One of the key elements of our approach evaluating the QM/MM free energy is the use of a classical reference potential. In this way, our main task is to evaluate the free energy of moving from the reference potential to the QM/MM surface, $\Delta G^{MM \rightarrow QM}$. The perturbation to the quantum mechanical surface can be done by a single step FEP[14] using

$$e^{-\Delta G^{MM \rightarrow QM}/kT} = \langle e^{-\Delta E^{MM \rightarrow QM}/kT} \rangle_{MM} \quad (11)$$

Here, $\langle...\rangle_{MM}$ designates that the average has been obtained when running trajectories on the molecular mechanical rather than quantum mechanical potential energy surface. In principle, we can also use the LRA expression.[42]

$$\Delta G^{MM \rightarrow QM} = \frac{1}{2}(\langle E^{QM} - E^{MM} \rangle_{MM} + \langle E^{QM} - E^{MM} \rangle_{QM}) \quad (12)$$

This approach requires, however, an expensive average over the QM (more specifically the QM/MM) surface. A simplified but useful alternative is provided by

$$\Delta G^{MM \rightarrow QM} = \langle E^{QM} - E^{MM} \rangle_{MM} \quad (13)$$

This approximation is justified if the averages on the QM and MM surfaces are similar. In this study, we performed the averages given by eqs 11 and 13 over 51−220 configurations. Then, the quantum mechanical reduction potential is given by

$$\epsilon^{\circ}_{QM} = \epsilon^{\circ,ref}_{Q\mu} + \Delta\epsilon^{\circ}_{N \rightarrow I} + \Delta\epsilon^{\circ,MM \rightarrow QM} \quad (14)$$

(59) St-Amant, A.; Salahub, D. R. *Chem. Phys. Lett.* **1990**, *169*, 387.
(60) Sim, F.; Salahub, D. R.; Chin, S.; Dupuis, M. *J. Chem. Phys.* **1990**, *95*, 4317.

where $\Delta\epsilon^{\circ,MM \rightarrow QM}$ is given by

$$\Delta\epsilon^{\circ,MM \rightarrow QM} = -\frac{\Delta G^{MM \rightarrow QM}_{Red} - \Delta G^{MM \rightarrow QM}_{Ox}}{nF} \quad (15)$$

At present, we have chosen to include only the electrostatic interaction energy in this perturbation. Since the active sites are the same in the two proteins, the effect from its potential energy surface, the quantum mechanical term corresponding to $\Delta\Delta G^{\circ,p_1 \rightarrow p_2}_{intra}$ in eq 5, will most likely cancel; besides, the uncertainty introduced from this term would probably be greater than the effect itself.

To make the reference potential as similar as possible to the quantum mechanical potential energy surface, we have in some calculations used Mertz−Kollman charges, q_{MK}, of the frozen fragments averaged over at least 10 configurations calculated with the B3LYP functional, a 6-31G* basis set, and an extended grid rather than using the standard force field point charges. To account for the polarization effect due to the protein on these fragments, we added the difference of the Mulliken charges, q_{Mull}, calculated in a vacuum and in its protein environment. Thus, the charges are evaluated as

$$q = q^{vac}_{MK} + q^{pro}_{Mull} - q^{vac}_{Mull} \quad (16)$$

Results and Discussion

Ionizable Groups. To include the effect of the charges from the ionizable residues on the reduction potential macroscopically rather than explicitly in the simulation might intuitively seem less accurate. However, to reproduce the charge−charge compensation from the solvation energy by the protein+solvent system correctly is extremely challenging, e.g., see ref 55. This solvent compensation effect involves water penetration and other large protein reorganizations that probably would require extremely long simulations to treat correctly. In addition, if the charges are far from the center of the simulated system, the effective size screening this interaction is considerably smaller than the actual simulation size. Thus, we evaluated the effects of the ionizable groups using an effective dielectric constant of 60 for charge−charge interactions. Using this approach while ionizing all groups that would be ionized at pH = 7 in water and averaging over 40 reduced and 40 oxidized protein configurations lowers the reduction potential by 113 and 33 mV for plastocyanin and rusticyanin, respectively. This seems to be intuitively correct since plastocyanin is more negatively charged and, thus, stabilizes the oxidized state over the reduced state and lowers the reduction potential. The overall effect increases the reduction potential difference between these two proteins by 80 mV, $\Delta\epsilon^{\circ}_{N \rightarrow I}$. A similar magnitude was found by Case et al.,[61] who calculated the reduction potential difference with their Poisson−Boltzmann (PB) approach at 100 mM salt concentration. Note, however, that PB studies with small dielectric constants for the protein tend to overestimate the effect of the ionizable residues.[55,56]

Classical Semi-macroscopic Approaches. In Table 2, we present the results of the classical simulations where the reduction potentials have been shifted so that we obtain the experimental value for the total reduction potential ($\epsilon^{\circ}_{Q\mu} + \Delta\epsilon^{\circ}_{N \rightarrow I}$) of plastocyanin. The most stable and accurate way of calculating the reduction potential appears to be the PDLD/S-LRA approach. The PDLD/S-LRA procedure requires that the reduction potential is averaged over a number of configurations

(61) Botuyan, M. V.; Toy-Palmer, A.; Chung, J.; Blake II, R. C.; Beroza, P.; Case, D. A.; Dyson, H. J. *J. Mol. Biol.* **1996**, *263*, 752−767.
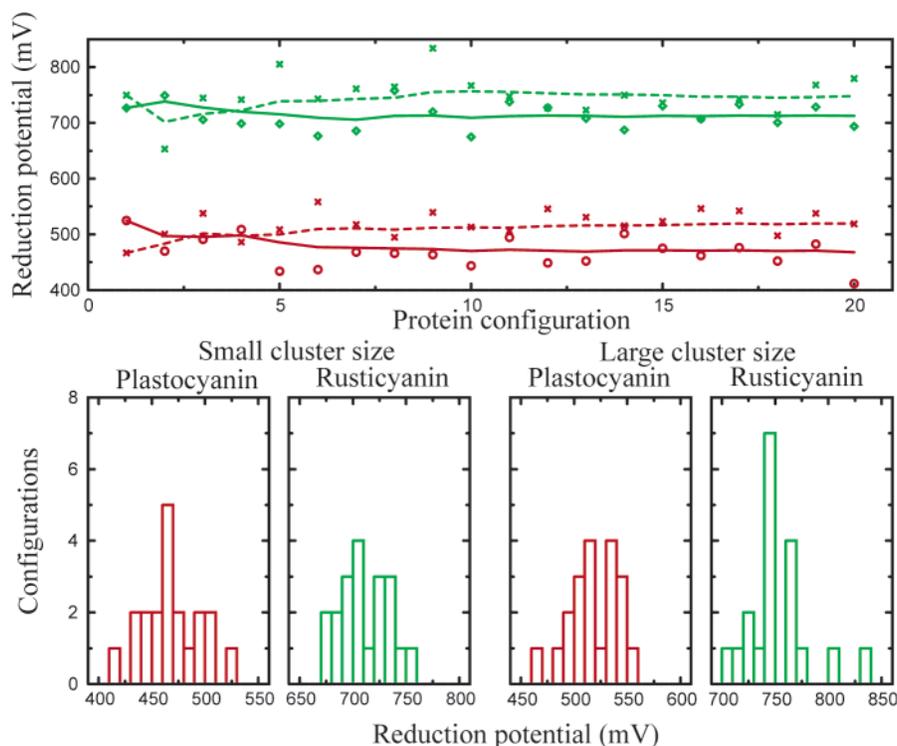
Figure 7. Dipolar contribution, $\epsilon_{Q\mu}^{\circ}$, to the reduction potentials of plastocyanin (red) and rusticyanin (green) calculated for the CuCysHis$_2$ (solid lines) and CuCysHis$_2$Met (dashed lines) models of the active site with the semi-macroscopic PDLD/S-LRA method. The curve shows the cumulative average, the average over all points from the beginning up to a certain simulation time. The points and the curve have been shifted so that the average reduction potential of the active-site cluster in water is 500 mV. The histograms of the lower half show the distribution of the points for the four cases clearer.

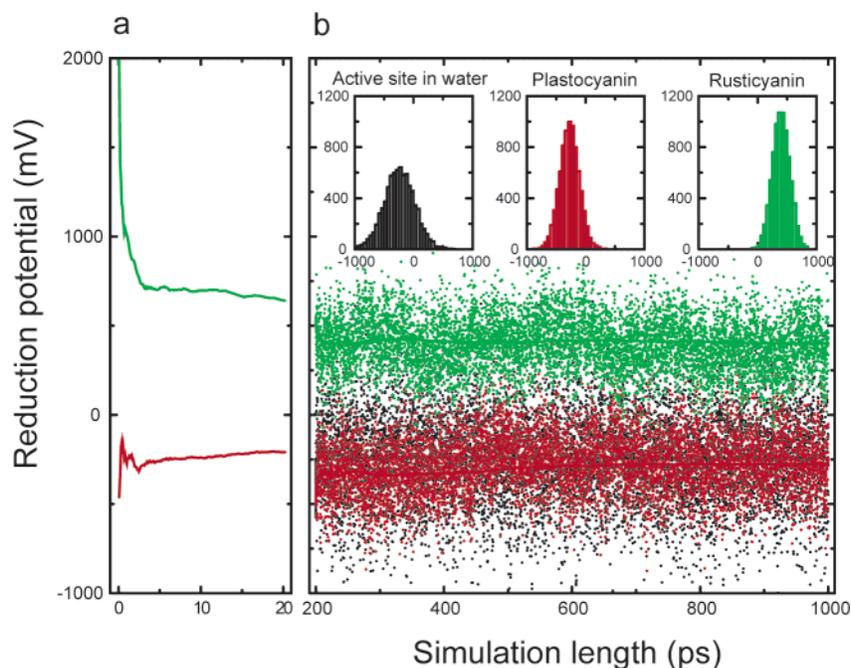**Table 2.** Reduction Potential (mV) of the Active Site in Water, Plastocyanin, and Rusticyanin

	active site	plastocyanin[a]		rusticyanin		
	ϵ_{Tot}°	$\epsilon_{Q\mu}^{\circ}$ [b]	ϵ_{Tot}° [c]	$\epsilon_{Q\mu}^{\circ}$ [b]	ϵ_{Tot}° [c]	$\Delta\epsilon^{\circ}$ [d]
PDLD/S-LRA						
CuCysHis$_2$[e]	520	488	375	732	699	324
CuCysHis$_2$Met[e]	469	488	375	717	684	309
LRA all-atom, regular force-field						
q_{Mull}[f]	393	488	375	1108	1075	700
δq_{Mull}[g]		488	375	845	812	437
LRA all-atom, polarizable force-field						
q_{Mull}[f]	569	488	375	677	644	269
δq_{Mull}[g]		488	375	721	688	313
expt			375		680	305

[a] The reduction potentials have been shifted so that plastocyanin gets its experimental value. [b] The reduction potential from simulations where all residues are kept in their neutral form. [c] The total ($\epsilon_{Q\mu}^{\circ} + \Delta\epsilon_{N\to I}^{\circ}$) reduction potential. [d] The total ($\epsilon_{Q\mu}^{\circ} + \Delta\epsilon_{N\to I}^{\circ}$) reduction potential difference $\epsilon_{Rc}^{\circ} - \epsilon_{Pc}^{\circ}$. [e] For the PDLD/S-LRA calculations, we have used two different sizes of region 1, CuCysHis$_2$ and CuCysHis$_2$Met. Since both models give very similar results, we have used only the smaller model in the remaining investigation. [f] Using the standard Mulliken charges, q_{Mull}^{Ox} and q_{Mull}^{Red} in Table 1. [g] Using the Mulliken charge difference, δq_{Mull} in Table 1 for the oxidized state and zero residual charges for the reduced state.

generated with the reduced and oxidized charge distributions. This average should include a significant number of protein configurations since the electrostatic potential changes significantly with protein fluctuations. The LRA treatment of the PDLD/S energy captures explicitly a significant part of the protein and solvent reorganization processes including some of the "slow processes" such as water penetration. Obviously, the longer the simulation is, the better the chance of capturing the energetics of such processes. However, since we cannot simulate all the slow rearrangement processes in a reasonable simulation

time, we have to use a protein dielectric constant, ϵ_p, which represents all the effects that cannot be simulated explicitly (see discussion in ref 56). Here we use $\epsilon_p = 4$. For these calculations, we have used two models of the active site, CuCysHis$_2$ and CuCysHis$_2$Met, solid and dashed lines respectively in Figure 7. These systems give reduction potential differences between the two proteins of 324 and 309 mV, respectively, which are both very close to what is observed experimentally, 305 mV. Though the small model gives a slightly worse agreement with the experimental reduction potential shift, it is still within the error limit of the method. This is very encouraging since it means that we can use this representation without losing much accuracy in the other methods. This is especially appreciable for the study with the otherwise very expensive quantum mechanical approach since it reduces the computational time considerably. The similarity of the results obtained with the two models also indicates that the charge transfer between the copper atom and the S$_{Met}$ atom is small and that the Cu$-$S$_{Met}$ interaction contributes very little to the reduction potential difference between plastocyanin and rusticyanin.

The 20 configurations evaluated over 1000 ps considered in this study fluctuate between 351 and 441 mV for plastocyanin and between 642 and 734 mV for rusticyanin (the red and green curves, respectively, in Figure 7). Thus, just considering one of these configurations rather than averaging could actually give a reduction potential difference anywhere between 200 and 380 mV. For most purposes, this is very similar to solving the Poison$-$Boltzmann equation for a set of charges representing the protein environment, which has been applied to this and many similar problems by Case et al.[61] They get a reduction potential difference of 228 and 389 mV for the NMR and X-ray

Figure 8. Dipolar contribution, $\epsilon^{\circ}_{Q\mu}$, to the reduction potentials of plastocyanin and rusticyanin calculated with the all-atom method, where the solvent is treated explicitly. The points are for the active-site cluster in water (black dots), plastocyanin (red dots), and rusticyanin (green dots). The picture to the left shows the cumulative average of the first 20 ps of the simulation, where the protein is unrelaxed and unequilibrated. As discussed in the text, the histogram insets show that the calculated reduction potential spans a 1000 mV range for each protein, giving a reduction difference between 0 and 2000 mV for a one-configuration calculation.

structure, respectively. This clearly shows that the results obtained with a single configuration have a considerably larger error range than those produced by rigorous protein simulations with proper treatment of the environment and averaging over long trajectories.

It is also interesting to find that the worst result obtained with this method is by using the original X-ray crystal structure without relaxing the environment. With our method, just considering this structure gives a reduction potential difference of 400 mV, which is slightly worse than 380 mV. Similarly, the results obtained in ref 61 are close to the extreme values in our PDLD/S-LRA calculations. This is due to not only experimental uncertainties or differences between crystal structures and the solution structures but also a problem with theoretical modeling of proteins (this will be even more pronounced in later sections). Today there is no force field that exactly reproduces the interactions in the natural protein. Therefore, any force field combined with an experimentally determined structure will give an artificial electrostatic potential at the active site that would correspond to an energetically excited structure. In fact, as much as energy calculations are concerned, it is essential to be in the minimum of the given potential energy surface even if they do not strictly reproduce the X-ray structure.[43]

The results presented here clearly show that it is possible to get a wide range of results with only one protein configuration and that it is crucial to average over a number of structures. However, the result is not disastrous since the Langevin dipoles and the dielectric of the Poisson–Boltzmann approach replace much of the explicit interactions of the environment, especially the solvent, with an average interaction. On the other hand, these methods require a protein dielectric "constant", ϵ_p, which is not uniquely defined. In the PDLD/S-LRA approach, the protein reorganization is considered explicitly, and thus, a relatively

small ϵ_p can be used, whereas the Poisson–Boltzmann method normally requires a larger ϵ_p. In principle, the semi-macroscopic methods are less rigorous than the microscopic models. Therefore, the next step is to consider all interactions explicitly.

Classical All-Atom Simulations. In all-atom models, all interactions are treated explicitly instead of averaged as in Langevin dipole or related approaches. With the all-atom approach (in the nonpolarizable version), we obtain a considerably higher reduction potential difference between rusticyanin and plastocyanin than with the previous PDLD/S-LRA method, 700 compared to 300 mV. This method is used to obtain the result in Figure 8. In Figure 8a, the cumulative average of the reduction potentials is shown for the first 20 ps simulation; it can be seen that the first few picoseconds are considerably worse than the following. This is most conspicuous for the very first point of the simulation, corresponding to the X-ray structure (hydrogen atoms have been placed by the automated procedure in MOLARIS), where the difference in reduction potential between the two proteins is close to 2500 mV. These points are far from representative for the "correct" simulation (Figure 8b) and do not capture the correct physics of the natural system. This artificial situation arises since the same protein structure has been used for both oxidation states of the protein and, as was discussed in previous section, due to the unrelaxed protein structure (parts of this difference, however, come from poorly equilibrated water molecules, which may be unjustified to include here). Needless to say, this will not represent the natural protein. Already after a few picoseconds, however, this artificial situation has been relaxed, giving sensible interactions and starting to give reasonable values of the reduction potential. In many cases, though, the required equilibration time is much longer than in this particular run. Also, the X-ray structure is usually available for only one of the oxidation states and, clearly,

any correct treatment should use structures equilibrated for each state.

From Figure 8b, it can also be seen that the fluctuations with the all-atom method are much larger than with PDLD/S-LRA. Thus, just taking any one point from the simulation in Figure 8b would give reduction potentials in a 1000 mV range for each protein, giving a difference between the two proteins in a 2000 mV range. Therefore, calculating the reduction potential with an explicit all-atom method for only one configuration will give meaningless results.

Even though we have run long trajectories and sampled the reduction potential over a substantial amount of configurations, it seems that this approach gives results considerably different from what is found experimentally. Interestingly, using charge differences, i.e., zero residual charges for the reduced form and the charge difference δq_{Mull} for the oxidized form, rather than the absolute charges gives much better results. This might reflect the difficulty of obtaining stable results for the neutral polar reduced form, or other yet unclear relations between the quality of the sampling and the simplicity of the charge distribution. As can be seen in Table 2, using the charge differences rather than the absolute values brings the reduction potential difference down below 500 mV, which is clearly in better agreement with experiments than 700 mV.

In these simulations, we have used the same Mulliken charges to represent the active site of both proteins, i.e., averaged over several plastocyanin and rusticyanin configurations. If we instead calculate the charges specifically for each proteins, i.e., obtaining the plastocyanin charges by averaging the corresponding QM/MM charges over protein configurations generated only from the plastocyanin simulations and similarly for the rusticyanin, the reduction potential difference decreases from 700 mV to about 650 mV. These charges can be found in Table 1. However, if this is done, the term $\Delta\Delta_{intra}^{o,p_1\rightarrow p_2}$ in eq 5 will not cancel and the polarization energy needs to be estimated. In these simulations this term should be small, less than 50 mV. Using the energy of the array of point charges used to represent the active site does not give a trustworthy estimate. Another option is to use the quantum mechanical polarization energy, i.e., the increase in the solute energy relative to its gas phase energy.[16] However, this treatment overestimates the contribution in the classical system since the redox site seems to be more polarized in the QM/MM calculation than in the classical calculation (see the QM/MM section). At any rate, including this term here should not change the result significantly, whereas it seems to increase the uncertainties. Thus, there is no reason to consider it at this stage.

Interestingly, using a polarizable force field[5] gives a considerably lower reduction potential difference between the two proteins. The procedure of averaging over 10 trajectories that was described for the nonpolarizable force field gives a difference of 269 mV with the polarizable force field. This is a significant improvement over the 700 mV obtained for the simulation with the regular force field. Similarly, using the Mulliken charge difference rather than the absolute charges brings us closer to the experimental reduction potential difference, 310 versus 305 mV, which is in even better agreement with the experimental difference than can be expected from this method. The drastic improvement is probably due to the fact
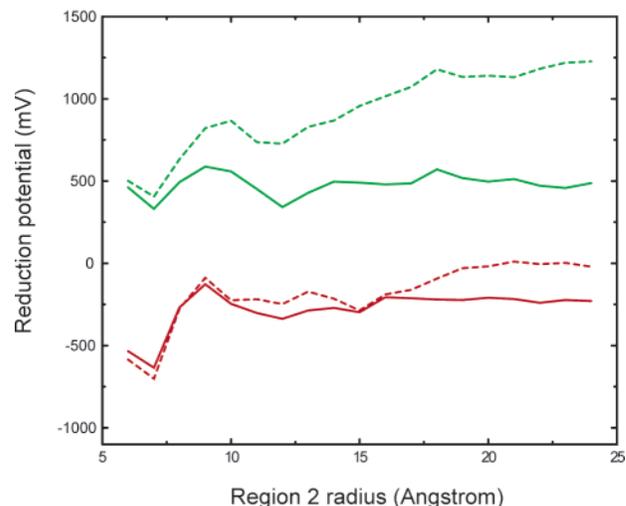
Figure 9. Reduction potential for different system sizes with the regular all-atom approach for plastocyanin (red curve) and rusticyanin (green curve) in regular water (solid lines) and nonpolar water (dashed lines).

that the induced dipoles help to stabilize charges in regions where the permanent dipoles are not so effective.

In general, including explicit water molecules when simulating proteins is of great importance. This is extra important when calculating reduction potentials since the increase in charge of the active site and the response from the environment almost exclusively constitutes the reaction coordinate. As can be seen in Figure 9, where the reduction potentials have been shifted so that plastocyanin gets its experimental value for the largest system size, the reduction potential simulated without the explicit water molecules strongly overestimates the reduction potential of both proteins. Needless to say, increasing the part of the simulated protein will not compensate for excluding the water environment. From the solid lines, it can also be seen that the calculations with explicit water molecules seem to converge with a sphere size of about 16 Å.

FDFT QM/MM Approach. The above classical approaches provide the background and a point of reference for the main task of performing reliable QM/MM redox calculations. Interestingly, and in contrast to what one might expect, moving to a quantum mechanical description does not automatically improve the energetics of protein modeling. Just taking the crystal structure and calculating the vertical quantum mechanical energy difference of the two states in the presence of protein point charges gives an energy difference of 97 and 156 kcal/mol for the plastocyanin and rusticyanin, respectively. This corresponds to a reduction potential difference of 2500 mV between the two proteins. If more care is taken at the QM/MM interface, i.e., using the FDFT formulation, the difference goes down to 2100 mV. The error in the absolute reduction potential, however, is probably even larger since it usually cancels to a certain extent by taking the difference between two similar systems. The result can be further improved by optimizing the solute either in a vacuum or, even better, in the protein. However, as long as the solvent, or the reorganization of the solvent, is not considered, this description will give equally bad results as the one-configuration unequilibrated classical approach. This procedure cannot be expected to represent the natural system and will result in arbitrary reduction potentials.

We have tried several different approaches to calculate the QM/MM reduction potentials using different ways of modeling

Table 3. Averaged Electrostatic Contributions Used in the MM → QM Perturbation (kcal/mol)

	plastocyanin		rusticyanin	
	Red	Ox	Red	Ox
ENZYMIX region I'[a]				
$\langle V_{Red} - V_{Ox}\rangle_{Red/Ox}$	−12.50	24.09	−30.31	10.16
$\langle E_{Q\mu}^{MM}\rangle$[b]	−47.99	−81.45	−52.24	−65.71
$\langle E_{Q\mu}^{QM}\rangle$	−58.16	−88.61	−42.59	−71.13
$\langle E_{pol}^{QM} - E_{intra}^{MM}\rangle$	37.20	26.72	24.89	20.45
$\Delta G^{MM\rightarrow QM}$[c]	27.04	19.56	34.54	15.03
$\Delta G_{Boltzmann}^{MM\rightarrow QM}$	16.16	9.20	22.43	1.96
adapted region I'[d]				
$\langle V_{Red} - V_{Ox}\rangle_{Red/Ox}$	−11.33	23.95	−21.29	14.13
$\langle E_{Q\mu}^{MM}\rangle$[b]	−42.32	−83.24	−45.17	−65.54
$\langle E_{Q\mu}^{QM}\rangle$	−56.86	−98.14	−45.94	−83.45
$\langle E_{pol}^{QM} - E_{intra}^{MM}\rangle$	34.65	23.68	26.86	28.44
$\Delta G^{MM\rightarrow QM}$[c]	20.11	8.79	26.09	10.59
$\Delta G_{Boltzmann}^{MM\rightarrow QM}$	11.41	−1.95	18.73	−0.81

[a] The standard ENZYMIX force field has been used to represent the reference potential of region I'. [b] This average is taken only over the configurations used for the $G^{MM\rightarrow QM}$ perturbation. [c] The term $\Delta G^{MM\rightarrow QM}$ can easily be calculated from this table as $\langle E_{Q\mu}^{QM}\rangle - \langle E_{Q\mu}^{MM}\rangle + \langle E_{pol}^{QM} - E_{intra}^{MM}\rangle$, whereas $\Delta G_{Boltzmann}^{MM\rightarrow QM}$ is given by eq 9. [d] The region I' has been adapted to the QM/MM potential by using Merz−Kollman charges calculated according to eq 16.

Table 4. QM/MM Reduction Potential (kcal/mol if not given explicitly)

	ϵ_{QM}°		$\Delta\epsilon_{QM}^{\circ}$	
	Pc	Rc	kcal/mol	mV
ENZYMIX region I'				
$\epsilon_{Q\mu}^{\circ,\,ref}$	−5.79	10.07	15.87	687.95[b]
$\epsilon_{Q\mu}^{\circ,\,ref} + \Delta\epsilon_{N\rightarrow I}^{\circ}$	−8.40	9.30	17.70	767.11
$\epsilon_{Q\mu}^{\circ,\,ref} + \Delta\epsilon_{N\rightarrow I}^{\circ} + \Delta\epsilon^{\circ,MM\rightarrow QM}$	−15.88	−10.21	5.67	245.61
$\epsilon_{Q\mu}^{\circ,\,ref} + \Delta\epsilon_{N\rightarrow I}^{\circ} + \Delta\epsilon_{Boltzmann}^{\circ,MM\rightarrow QM}$	−15.36	−11.17	4.18	181.38
adapted region I'				
$\epsilon_{Q\mu}^{\circ,\,ref}$	−6.31	3.58	9.89	428.72[b]
$\epsilon_{Q\mu}^{\circ,\,ref} + \Delta\epsilon_{N\rightarrow I}^{\circ}$	−8.91	2.81	11.72	507.88
$\epsilon_{Q\mu}^{\circ,\,ref} + \Delta\epsilon_{N\rightarrow I}^{\circ} + \Delta\epsilon^{\circ,MM\rightarrow QM}$	−20.23	−12.69	7.54	326.83
$\epsilon_{Q\mu}^{\circ,\,ref} + \Delta\epsilon_{N\rightarrow I}^{\circ} + \Delta\epsilon_{Boltzmann}^{\circ,MM\rightarrow QM}$	−22.27	−16.74	5.53	239.89

[a] The reference reduction potential, ϵ_{ref}°, is calculated from $\langle V_{Red} - V_{Ox}\rangle_{Red/Ox}$ in Table 3 using eqs 1 and 7. [b] The large difference in ϵ_{ref}° between the two force fields comes mainly from E_{intra}^{MM} and is, as discussed in the text, to a great extent artificial.

the protein, different ways of including the perturbation, and different reference potentials. Depending on these approaches, we obtain reduction potential differences between 180 and 340 mV. Two representative calculations are presented in Table 3 and Table 4. In Table 3, the $G^{MM\rightarrow QM}$ perturbation and its decomposition are presented for each protein and oxidation state. The upper half refers to calculations using the standard region I' ENZYMIX force field and the standard region I Mulliken charges presented in Table 1. The lower half, on the other hand, refers to the calculation where the reference potential energy surface has been adapted to that of the QM/MM approach by using charges for region I' obtained with eq 16 and region I Mulliken charges calculated separately for each protein (see Table 1). The latter approach should be more accurate since the potential energy surfaces are more similar. This can also be seen from Table 3, where the free energy difference between

the classical potential energy surface and the QM/MM potential energy surface, $\Delta G^{MM\rightarrow QM}$ and $\Delta G_{Boltzmann}^{MM\rightarrow QM}$, is smaller in the lower half of the table. Using the average, this adaptation brings the reference potential 8 kcal/mol closer to the QM/MM potential energy surface. Similarly, the resulting reduction potentials for the two calculations can be found in Table 4.

Maybe a bit surprisingly, it turns out that the result does not depend terribly much on which of these reference potentials is used. It seems that the same ENZYMIX parameters as for the classical simulation gives a too low reduction potential difference, 250 and 180 mV depending on how the perturbation is included, whereas using the "adapted reference function" gives a result that is closer to the experimental value. However, it is not really known if this is a general fact or just reflects the uncertainties of the method. Further, it should in principle be physically more correct to include the $G^{MM\rightarrow QM}$ perturbation as Boltzmann weights, $\Delta G_{Boltzmann}^{MM\rightarrow QM}$, but if it is not sampled properly, low extreme values of this perturbation can give an unbalanced perturbation. On the other hand, using the average gives too large weight to perturbations high in energy. In the two cases presented here, it seems that averaging the Boltzmann weights rather than the energy difference results in a 60−90 mV larger reduction potential difference. At any rate, though the numerical values of the quantum mechanical approach might be less stable and have larger error ranges than the classical, it seems clear that moving to the quantum mechanical surface decreases the reduction potential difference considerably. This could in part reflect that the frozen densities are polarizable and should be more similar to the simulation with the polarizable force field. It should also be noted that the difference in the reference energies between the two force fields presented in Table 4, 690 compared to 430 mV, is artificial. This is because different sets of point charges have been used for the two proteins in the calculations corresponding to the lower half of the table. This will include a nonphysical value of $\Delta\Delta G_{intra}^{\circ,p_1\rightarrow p_2}$ in $\epsilon_{Q\mu}^{\circ,ref}$, E_{intra}^{MM} in the notation of Table 3, that does not cancel in eq 5. This term, however, is subtracted when including the polarization term ($E_{pol}^{QM} - E_{intra}^{MM}$) in the perturbation.

In Figure 10, the time-dependent perturbation ($E^{QM} - E^{MM}$) is depicted for each oxidation state and protein with the adapted reference potentials. In general, this energy difference spans a range of about 30 kcal/mol for each protein and oxidation state (see the histograms in Figure 10). This reflects how difficult it is to couple the QM and MM regions when the environment is included correctly and that it is necessary to average over a large number of protein configurations in order to get a meaningful perturbation. Clearly, running directly on the QM/MM potential would be preferable since the distribution in the histograms would be narrower and, thus, the number of configurations needed for the $G^{MM\rightarrow QM}$ perturbation considerably less. At the moment, though, this would not allow us to sample the protein configurations properly and would introduce errors on the order of tens of kcal/mol.

The polarization energy of the solute in the quantum mechanical calculations cannot always be neglected since the electric field from the different protein environments induces different dipoles in the active sites. If it is neglected, the perturbation is overestimated and the reduction potential difference becomes negative. Again, this reflects how challenging
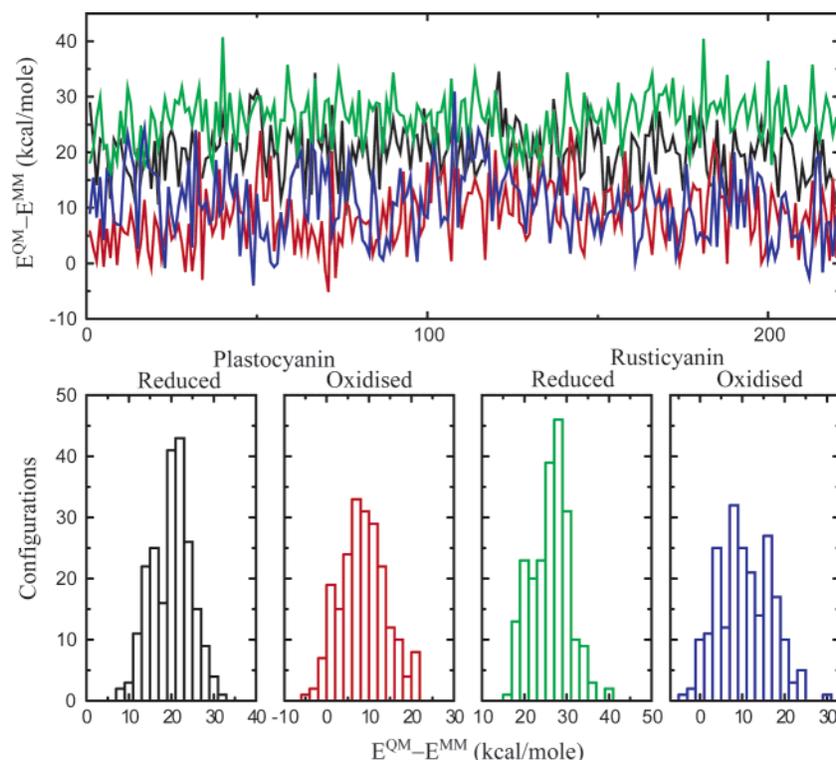
Figure 10. Perturbation ($E^{QM} - E^{MM}$) calculated for the 220 configurations for each oxidation state and protein: reduced (black) and oxidized (red) plastocyanin; reduced (green) and oxidized (blue) rusticyanin.

it is to calculate reduction potential differences with a QM/MM approach correctly.

Origin of the Reduction Potential Difference. When discussing the protein function, it is helpful to choose a reference system. As such, it falls natural to choose the active site cluster, Im$_2$CuSCH$_3$, in an aqueous environment. In essence, we cut away the effect from the protein, but retain the water environment. Interestingly, the calculations of this reference system invariably end up between the two proteins in our calculations, 400−600 mV. This means that the intrinsic reduction potential of the active site, here defined as the region 1 cluster in water, must be rather high from the beginning. This result might seem a bit counterintuitive considering that the blue copper proteins were early noted for having a considerably higher reduction potential than the Cu$^+$/Cu^{2+} couple in an aqueous medium. Clearly, it also contradicts the assumption that the primary function of the blue copper proteins is to encapsulate the redox system in a hydrophobic patch.[40] However, the first coordination sphere in the protein is quite different from that of the aqueous system, and the rather high reduction potential of the active site in water is supported by small inorganic copper complexes that are experimentally found to have comparatively high reduction potentials.[62,63] These are, however, rather poor models for the blue copper active site, and the comparison can at the most be qualitative.

Another closely related issue is the solvent exposure of the active site as a way of tuning the reduction potential. In rusticyanin the active site is buried deeper in the protein than in plastocyanin and should therefore be better protected from the solvent. Again, it has been speculated that the solvent, being

a high dielectric environment, stabilizes the oxidized form more than the reduced and lowers the reduction potential. Thus, one can easily assume that the higher reduction potential of rusticyanin would be ascribed to the active site being buried deeper in the protein and less exposed to ambient water. In Table 5, we have divided the reduction potential for the two proteins into Born, Langevin, and permanent and induced water and protein dipole contributions. The actual values vary depending on what method is used and how the system is modeled, but the overall picture is the same: the protein permanent dipoles raise the difference between the proteins by 335, 1431, and 613 mV for the PDLD/S-LRA and the all-atom simulations with the regular and polarizable force field, respectively. The water contribution, however, lowers the difference. The reduction potential difference is lowered by 95, 784, and 83 mV with the different methods. (In the PDLD/S-LRA, the solvent water molecules are represented as Langevin dipoles rather than explicit water contributions.) Following the reasoning above, the latter might seem a bit counterintuitive since plastocyanin is the more solvent exposed of the two proteins. However, since the protein dipoles of rusticyanin are oriented such that the reduction potential is raised, it seems natural that the water molecules, which orient themselves according to the total electrostatic potential, would counteract the effect of the protein dipoles.

As was indicated in the previous section, the overall effect of the protein is quite different in the two cases. In plastocyanin, the reduction potential is tuned down, whereas it is tuned up in rusticyanin. Again, the amount is different depending on how the proteins are modeled and what method is used, whereas the direction is the same with all approaches. To understand this difference, it is very instructive to consider the contributions

(62) Krylova, K.; Kularilleke, C. P.; Heeg, M. J.; Salhi, C. A.; Ochrymowycz, L. A.; Rorabacher, D. B. *Inorg. Chem.* **1999**, *38*, 4322−4328.
(63) Karlin, K. D.; Yandell, J. K. *Inorg. Chem.* **1984**, *23*, 1184−1188.

Table 5. Protein and Water Permanent and Induced Dipolar Contributions to the Reduction Potential (mV) from the All-Atom Calculation

	permanent dipoles		induced dipoles				
	protein	water[a]	protein	water	Langevin[a]	Born	total[b]
PDLD/S-LRA							
rusticyanin	169	0	0	0	−145	−81	−57
plastocyanin	−166	0	0	0	−50	−85	−301
difference	335	0	0	0	−95	4	244
LRA with regular force field							
rusticyanin	776	−866	0	0	−35	−351	−476
plastocyanin	−655	−82	0	0	−2	−358	−1097
difference	1431	−784	0	0	−33	7	621
LRA with polarizable force field							
rusticyanin	699	−836	−529	−173	−36	−341	−1216
plastocyanin	86	−763	−197	−163	−12	−358	−1407
difference	613	−73	−332	−10	−24	17	191

a In the PDLD/S-LRA method, the water contribution is described by Langevin dipoles and is considered as such as well. *b* The PDLD/S-LRA and the microscopic LRA corresponds to different reference potential. That is, the PDLD/S-LRA calculations give the difference between the energy of the cluster in the protein and in water. The microscopic LRA calculations, on the other hand, give the full solvation relative to vacuum.
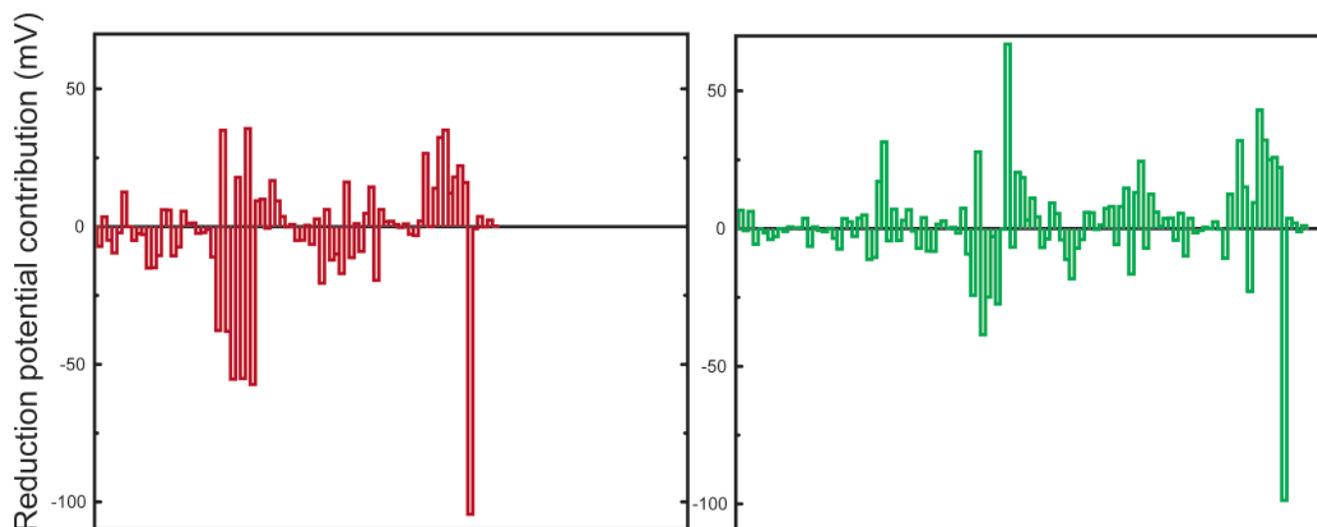
Figure 11. Group contributions to the reduction potential in plastocyanin (red) and rusticyanin (green) calculated with the PDLD/S-LRA method.

of the protein residues to the reduction potential. A convenient approximation to these contributions is obtained by the PDLD/S-LRA group contributions,[50] which are evaluated according to

$$\Delta G^{(i)} = \frac{\langle V_{Q\mu}^{(i)} \rangle}{\epsilon_{eff}} \qquad (17)$$

where $V_{Q\mu}^{i}$ is the Coulombic interaction between the dipole moment of the ith residue and the residual charges of region 1. ϵ_{eff} is the effective dielectric for the charge−dipole interaction and is taken as $\epsilon_{eff} = 4$. The average $\langle ... \rangle$ is taken over the 20 protein configurations used in this investigation. The result for residues closer than 20 Å from the system center is presented in Figure 11, and the 10 biggest positive and negative individual contributions are collected in Table 6. From these data, it is indicated that the Cu−S_{Met} interaction actually lowers the reduction potential of both proteins, but does not contribute to the difference. This is clearly in accordance with the higher reduction potential found for ceruloplasmin, where the methionine ligand is replaced by a noncoordinating residue. However, it is unknown to what extent this effect comes from this particular ligand and how much comes from the remaining electrostatic protein environment.

Table 6. The Ten Largest Positive and Negative Contributions to the Reduction Potential (mV) for the PDLD/S Calculation

	plastocyanin				rusticyanin			
1	HIE37	36	MET92	−105	HIE81	67	MET144	−99
2	HIE87	35	ASN38	−57	HIE139	43	ASN76	−38
3	ASN32	35	GLY34	−55	ALA140	32	PHE79	−27
4	PRO86	32	PRO36	−55	GLN135	32	LYS77	−25
5	TYR83	27	ALA33	−38	PRO48	32	ASN74	−24
6	ALA90	22	ASN31	−38	THR75	28	PRO137	−23
7	GLY89	18	MET57	−21	THR142	26	VAL94	−18
8	PHE35	18	GLU68	−20	ALA141	25	GLY106	−17
9	ASP42	17	ASP61	−17	SER108	24	ALA93	−11
10	LEU62	16	SER11	−15	GLY143	22	PHE45	−11
sum				−166				170

Interestingly, experimental denaturation of azurin, a blue copper protein with an additional axial ligand and reduction potential close to that of plastocyanin, raises the reduction potential from about 310 to about 450 mV rather than lowers it.[34] This is consistent with plastocyanin lowering the reduction potential relative to the active site in an aqueous environment, 375 compared to 500 mV, which was found in this study. To make any quantitative comparison, however, it is necessary that the active site is intact (if not, the $\Delta\Delta G^\circ_{intra}$ term will not cancel in eqs 4 and 5). Further, it is also questionable if the denatured state could be considered similar to the active site in water,

even though the solvent exposure is probably increased in the denatured state. An alternative explanation could be that the mild denaturation disrupts the preorganized dipoles, which then does not tune down the reduction potential any longer. At any rate, our calculations of plastocyanin show qualitatively the same behavior.

Discussion

The present work addresses both the general computational challenge of ab initio QM/MM calculations of proteins and the specific issue of the control of the reduction potential in blue copper proteins. These two aspects of our work will be discussed below.

Requirements of Reliable Calculations of Reduction Potentials in Proteins. Undoubtedly, experimentally determined crystal and NMR structures provide much better overall structural information than theoretical simulations of proteins. It is, however, not necessary that these structures provide the best option for theoretical investigations of protein function. This manifests itself in several problems. First, the electrostatic free energy, which we believe is the quantity of interest for protein function, reflects the average over many configurations (both charged and uncharged forms of the redox site), while the experimental structure is a single average structure. Second, we do not yet have a "perfect force field" that follows the natural system strictly and that reproduces the experimental average structure exactly. This makes experimentally determined structures often high-energy configurations on our potential energy surface, which, even more seriously, can lead to unphysically large differences between the energy of the reduced and oxidized forms of the protein. Running a molecular dynamics simulation, or relaxing the system with the given force field, will on the other hand give structures that differ slightly from the experimental. However, these structures reflect the interactions within the protein and, more importantly, the electric field at the active site correctly. To get sensible results, it is essential to reproduce the interactions in the protein+solvent environment and its field at the active site. This can only be done by using a protein structure that corresponds to the force field that is used for the given structure−function correlation.

Analogous to the protein force field, the choice of charges used to represent the active site is less important than its interaction with the protein. Thus, provided that the system is thoroughly equilibrated, different charges will correspond to different protein dipole orientations but similar interaction energies. If the system is not thoroughly equilibrated, on the other hand, the result usually depends strongly on the region 1 charges. From this, it also follows that it is far more important to reproduce the difference in charge density due to the redox process rather than the absolute values. Naturally, even if these issues have not been considered, it is possible to get results that are in good agreement with experimental results. This is, however, fortuitous and will probably not give stable results with respect to the parameters used.

At present, it seems clear that the PDLD is easier to use and that a more modest number of protein configurations is needed. In the all-atom types of approaches, however, it is much more difficult to obtain converged results since the environment is not averaged for each protein configuration. Thus, we have to average not only over the "slow dynamics" but also over the

"fast dynamics". There are advantages to this method as well though. Using explicit interactions rather than an arbitrary dielectric constant makes the approach more rigorous and greatly helps to pinpoint practical and conceptual problems. Thus, one of the objectives of this investigation has been to try to establish how the accuracy and stability of the all-atom approach compares to methods using averaged interactions.

For the computational point of view, however, the main point of the present work is the development of an effective and accurate ab initio QM/MM approach for evaluation of reduction potentials. To appreciate our advance, it is important to realize that none of the currently reported ab initio QM/MM investigations of reduction potentials and related pK_a calculations sample the protein configurations properly. This includes the studies by Noodleman, Case, and co-workers, who represent the protein by a macroscopic approach. Though this is a reasonable treatment of the question, as is evident from the above PDLD/ S-LRA study, it is not a true QM/MM approach since the protein is not treated on a microscopic level. A more closely related study is the attempt by Jensen and co-workers to evaluate the pK_a of LYS 55 in Turkey ovomucoid third domain.[64] This study involve an ab initio QM/MM treatment of the protein and a PCM continuum treatment of the solvent. However, the calculations do not include an averaging over the protein configurations. Considering the excellent results obtained in ref 64, it could perhaps be assumed that the average over the protein configurations is not needed. However, the LYS residue is practically immersed in water or the PCM solvation model, which tends to give reasonable pK_a values. As is clear from the results presented in this article, the electrostatic energies of charged groups in the protein interactions are extremely sensitive to the protein configurations. Thus, it is essential to average over a substantial number of protein configurations. Finally, it is important to recognize that the importance of proper averaging and the difficulties in obtaining converging results persist regardless if the potential energy surface is represented by classical or quantum mechanical methods. Furthermore, additional uncertainties can easily be introduced if proper care is not taken with the coupling between the QM and MM regions. The importance of a proper averaging might not be familiar to some who are struggling with the enormous time requirement of ab initio QM/MM calculations. However, it is easy to verify our assertion by considering a small model system whose fluctuations can be evaluated both by a full QM treatment and by a MM treatment. In this case, it will be easy to verify the sensitivity of the conformational fluctuations.

Reduction Potential Control of Blue Copper Proteins. Frequently, the active site of rusticyanin is described as "highly hydrophobic" (e.g., ref 38), implying that rusticyanin has a considerably lower effective dielectric constant than plastocyanin. This would stabilize the oxidized form of plastocyanin more than rusticyanin and increase the reduction potential difference. A highly hydrophobic protein would correspond to negligible protein electrostatic contributions. As can be seen from Figure 11 and Tables 2, 5, and 6, the protein residues considerably raise the reduction potential in rusticyanin for all methods. This is consistent with having a protein with preorganized dipoles, which actively *raise* the reduction potential. In plastocyanin,

(64) Minikis, R. M.; Kairys, V.; Jensen, J. H. *J. Phys. Chem. A* **2001**, *105*, 3829−3837.

on the other hand, the preorganized dipoles actively *lower* the reduction potential. More precisely, the dipoles of the amino acids making up the protein (CO, NH, OH groups, etc.) are positioned by the protein fold such that their reorganization during the reaction is small and at the same time give an electrostatic potential at the active site that is advantageous for the protein function. From this, it also follows that the water contribution will decrease the reduction potential difference, which was also found and discussed in the previous section. Further, it should be realized that the copper ion would lose a majority of its solvation energy going from the water phase into a "highly hydrophobic" protein active site. Since the copper ion spontaneously goes into the apo form of the protein in its oxidized state, it can be assumed that there is no substantial loss in free energy by incorporating the ion in the protein active site. Finally, it should be kept in mind that also hydrophobic residues have large backbone dipoles. This is also confirmed in Table 6, where it can be seen that some of the larger contributions to the reduction potential (both positive and negative) come from amino acids that are typically classified as hydrophobic.

Concluding Remarks

Throughout the computational chemistry community, several research groups are working to improve the accuracy of regular quantum mechanical approaches without considering the solvation. Sometimes these accurate approaches are used to study the active site of proteins, assuming that their accuracy is similar in the condensed phase that constitutes a protein. Unfortunately, this not only ignores the role of the protein in most cases but will also give a rather poor description of the active site properties.

We have put considerable effort into trying to obtain an accurate ab initio quantum mechanical description of the active-site region, where the quantum mechanical and classical molecular mechanical systems are coupled correctly with free energies. This is not a trivial task since the physics of the system needs to be treated correctly while retaining an accurate description of the active site; that is, the free energy needs to be sampled properly along the reaction coordinate rather than just minimizing the energy. This requirement is very demanding since ab initio quantum mechanical calculations are considerably more time-consuming than regular molecular dynamics simulations. More specifically for the results in this investigation, the fluctuations of both the protein+solvent and MM → QM perturbation are on the order of 20−30 kcal/mol. Since we are addressing an energy difference considerably less than 10 kcal/mol, it is essential to average both these properties over a large number of configurations where QM and MM regions are coupled rigorously. The requirement of proper sampling is ususally neglected in today's ab initio QM/MM approaches, where the system energy is at the best minimized. It is also instructive to recognize that the field of simulating complex biochemical processes is not yet at the stage where we can obtain results with chemical accuracy from QM/MM ab initio simulations. In fact, at present, more accurate results are obtained with the classical approaches than with QM/MM, which can be seen from the results presented here. However, the challenges addressed in the present work concern a large QM region and evaluation of extensive configurational averages. Both of these issues appear to be addressed effectively by the FDFT method and the use of a classical reference potential.

All calculations presented here indicate that the electrostatic interactions are of great importance for protein function. More specifically, this means that the reduction potential optimal for the proteins has been attained by orienting the protein dipoles, e.g., CO, NH, OH groups, such that the electrostatic potential at the active site is appropriate for its function, i.e., both creating a stable protein, which is attractive for the copper ion, and giving the appropriate reduction potential. This is a general assumption that should be true for all electron transfer proteins.

JA0212157