REDOX PROPERTIES OF CYTOCHROME C: NOVEL LINEAR RESPONSE AND HYBRID CONTINUUM–MICROSCOPIC METHODOLOGIES

Thomas Simonson
Institut de Génétique et Biologie Moléculaire et Cellulaire (C.N.R.S.), 1 rue Laurent Fries (C.U. de Strasbourg), 67404 Illkirch France.

Redox properties of yeast cytochrome c are estimated using molecular dynamics combined with a simple linear response approximation, as well as a hybrid continuum–molecular dynamics (COMD) approach. In both approaches, the free energy associated with an electrostatic perturbation (a redox electron) is separated into its relaxation and static (non-relaxation) components. The static component is calculated from the molecular dynamics simulation. The relaxation component is then calculated with a linear response approximation, either from the molecular dynamics, or from a separate continuum calculation. This latter hybrid approach exploits the relative robustness of continuum models for dealing with large perturbations, while avoiding some of their limitations. It is quite general, and could be applied for example to pK_a calculations.

1 Introduction

Electrostatic and dielectric properties of proteins are central to their stability and activity [1]. This is especially obvious in the case of redox and electron transfer proteins, which harvest and orchestrate the flow of much of the energy in the biosphere. A very challenging area is the study of charge screening in these complex, inhomogeneous systems. Various experimental techniques have been used, such as measurements of redox potential shifts [2], pK_a shifts [3], or changes in stability [4], upon mutating charged and polar groups. These techniques provide only indirect information on the protein properties themselves, since they contain a large contribution from solvent relaxation. Dielectric dispersion by dry protein powders [5], and measurements of gas phase basicities of protein ions [6], probe the protein properties directly, but in non-aqueous environments. The presence of a spectroscopic probe can provide a sensitive window into local details of a protein medium. Thus Stark effect measurements have been used to study both the equilibrium electric field [7], and local polarizability [8] at specific protein sites. Time-dependent Stokes shifts of a die complexed with apomyoglobin [9] provide a direct measurement of time-dependent dielectric relaxation. Finally, computer simulations provide a powerful approach, which becomes increasingly attractive as force fields, electrostatic treatments, and computer resources steadily improve [10, 11].

Cytochrome c has served for many years as a model electron transfer pro-

tein, and a large body of experimental and simulation data has been accumulated. We have carried out molecular dynamics simulations of yeast cytochrome c in solution in its oxidized and reduced forms, and we have reported recently detailed analyses of its macroscopic and microscopic dielectric properties [12–14]. We report here further calculations of redox properties, using two novel continuum and molecular dynamics methods. The first is a simple linear response approximation, proposed recently by Simonson *et al.* and Levy *et al.* [15–17]. It is used in conjunction with molecular dynamics simulations to calculate the reorganization free energy associated with heme oxidation and reduction, and compared to more rigorous free energy perturbation calculations. This approximation has already been applied to several problems [12–20]. However applications to proteins have been limited, and it is therefore useful to explore its limits further here.

The second method is a new, hybrid, continuum–molecular dynamics (COMD) method for the calculation of electrostatic properties. In this method, the perturbation free energy associated with, say, heme oxidation is decomposed into two components: a relaxation free energy, and a static free energy [21]. The static free energy is calculated from a molecular dynamics simulation, while the relaxation free energy is obtained from a continuum model. The COMD method is quite general, and can be applied to many electrostatic calculations, including pK_a calculations. As a preliminary test, the method is applied here to the calculation of the redox potential of cytochrome c. Additional applications are underway.

2 Methods

2.1 Linear response approximation

When a set of perturbing charges q_i (such as a redox electron) is introduced into the vicinity of a protein, the perturbation free energy has the well-known expression [22]

$$\Delta G = -kT \ln\langle \exp(-U/kT)\rangle_0 \,, \tag{1}$$

where U is the perturbation energy, and the brackets indicate an ensemble average over the unperturbed ensemble. If the perturbation consists of a single redox electron, for example, ΔG is the redox potential. Expanding the above expression in powers of U/kT gives

$$\begin{aligned} \Delta G &= \langle U\rangle_0 - \frac{1}{2kT}(\langle U^2\rangle_0 - \langle U\rangle_0^2) + O([U/kT]^3) \\ &= G_1 + G_2 + G_{nl} \end{aligned} \tag{2}$$

The first term G_1 represents the static component of the perturbation free energy, while the remaining terms G_2 and G_{nl} represent the relaxation free energy [13,21]. If the perturbation consists of a redox electron, the relaxation free energy is identical to the reorganization free energy of electron transfer theory [23]. The linear response approximation truncates the free energy expansion after the second order term G_2, giving a quadratic free energy with respect to the perturbing charges. The remaining G_{nl} is the 'non-linear' component. This truncation is equivalent [17] to assuming a gaussian probability distribution of the perturbing potential (over the unperturbed ensemble). The relaxation free energy can also be written

$$\Delta G - G_1 = G_{rlx} = -kT \ln\langle \exp(-\delta U/kT)\rangle_0, \tag{3}$$

where δU is the deviation of U from its mean. The ensemble averaging is normally done using a microscopic simulation technique—molecular dynamics in the present work.

2.2 COMD: a hybrid continuum–molecular dynamics model

The decomposition of the perturbation free energy into static and relaxation components (above) can be made in continuum models as well as molecular dynamics. It is easy to show (*e.g.* [13, 15]) that the relaxation free energy is identical to the self-energy of the perturbing charges. Thus for a redox electron, for example, the relaxation free energy for reduction is the free energy to insert the perturbing electron, with all the protein permanent charges removed. Proton self-energies are routinely calculated as a part of pK_a calculations.

We propose here a novel, hybrid, method, termed COMD, in which the relaxation component of the perturbation free energy is calculated from a continuum model, while the static component ($\langle U\rangle_0$) is calculated from a molecular dynamics simulation. This additional complexity aims to exploit optimal features of each method. Thus, the molecular dynamics model should provide a reasonably accurate estimate of the static component of the free energy, obtained by simply averaging the perturbation over the simulation. While this procedure is known to have significant statistical uncertainty, it is nevertheless free from the more serious approximations of continuum models; the neglect of specific ordering of discrete water moleules for example. On the other hand, the continuum model provides a simple and robust approach to relaxation free energies, which has been tested specifically for this purpose in some detail [13]. Microscopic free energy calculations, either using the linear approximation or the full exponential form (1), are limited with respect to the size of the perturbation that can be tackled (see below, as well as *e.g.* [18,12]),

precisely because of lack of convergence of the relaxation free energy. The continuum model, while approximate, is known to be more robust in this respect. It is also completely independent, in this context, of any choice of permanent charge distribution, since the self-energy is calculated without any protein permanent charges; and it is fairly robust with respect to the choice of atomic radii. It does require an assumption for the protein dielectric. Here, the COMD method exploits an important feature of proteins: the optimal dielectric constant to describe relaxation properties is normally different from the low value (1–2) used to describe equilibrium properties [24–26]. This is physically fairly obvious, but has been neglected in most applications of continuum models. By explicitly separating out the relaxation part of the free energy, the COMD method relies clearly on the Fröhlich-Kirkwood dielectric constant, which is unambiguously defined as a linear response coefficient. The Fröhlich-Kirkwood dielectric constants of eight proteins have been calculated recently from molecular dynamics simulations [27, 12, 26]. It was found that while the overall values are surprisingly large (ranging from 11 to 35), values in the protein interior are much lower, ranging from 2 to 4, consistent with earlier empirical estimates [24] as well as powder experiments [5]. For cytochrome c, in particular, the polarizability in the heme vicinity is low [12, 13].

At first glance, the self-energy calculation with the continuum model would appear to require an assumption for the radius of the perturbing charges. This could introduce strong artefacts into the calculation. However in practice, one transfers the perturbing charges from a reference compound, such as a heme in solution, to the compound of interest—the protein-bound heme. In that case there is a systematic cancellation of the radius-dependent part of the free energy.

2.3 Simulations

Two molecular dynamics simulations were performed, of yeast ferro- and ferricytochrome c (abbreviated 1YCC and 2YCC respectively), solvated by 1400 TIP3P waters [28], at 293 K, starting from the crystal structures [29], lasting one nanosecond each. Electrostatic interactions were truncated beyond 12 Å (1 Å = 0.1 nm). The Charmm/Param19 empirical force-field was used [30]. A soft spherical boundary potential of radius 24 Å was used to confine the system. Simulations were done with the program X-PLOR [31].

Continuum calculations were performed with the program Delphi [32] using the protein crystal structure.

In the redox calculations, the redox electron is assumed to be distributed equally on the four nitrogen atoms that bind the heme iron.

3 Results

3.1 Heme oxido-reduction: linear response free energy calculations

Free energies to oxidize reduced cytochrome c (1YCC), and to reduce oxidized cytochrome c (2YCC), are reported in Table 1. The free energies are decomposed into the first-order term G_1, the second-order term G_2, and higher-order terms G_{nl} (eq. 2). By averaging over these 'forward' and 'backward' free energy estimates, we obtain another estimate of the 1YCC oxidation free energy. This estimation is equivalent to performing the oxidation in two steps, *i.e.* through a half-oxidation of 1YCC and a half-reduction of 2YCC. This two-step protocol is closer to a full-fledged free energy calculation, and should be more accurate.

We first observe the performance of the linear response approximation. The quality of the linear response approximation is measured by the magnitude of the non-linear terms, which are significant for the full reduction of 1YCC, but negligible for the half-oxidation and half-reduction. Thus the linear response approximation to the relaxation free energy is accurate, in this case, for a perturbation as large as a half-electronic charge; it is inaccurate for perturbation by a full electronic charge

Second, we can analyze further the accuracy of the relaxation free energy by assuming that the probability distribution of the equilibrium electrostatic potential is gaussian on the heme. This was shown to be the case at all the C_α positions in cytochrome c, over the thermally accessible range of conformations [13]. It has also been observed in the active site of histidyl tRNA synthetase (T. Simonson, unpublished). The analysis goes as follows. Calculation of G_{rlx} (eq. 3) involves integrating the product of $\exp(-\delta U/kT)$ by the probability distribution $p(\delta U)$, which we are assuming to be gaussian. This product is another gaussian, $w(\delta U)$, displaced with respect to $p(\delta U)$ by $-\sigma^2/kT$, where σ is the standard deviation of U. In order to calculate the integral accurately, the peaks of p and w must overlap sufficiently. In other words, our simulation must sample the regions that contribute most to the integral. This will be the case if the shift σ^2/kT is not much larger than the width σ of p. Thus the condition to reliably integrate is

$$\sigma \sim kT. \tag{4}$$

If on the contrary the fluctuations of δU are much larger than kT, we cannot be sure of the accuracy of a single step calculation of the relaxation free energy; the unperturbed ensemble does not sufficiently sample the tails of the probability distribution $p(\delta U)$. This does not necessarily mean that linear response is inaccurate. Its accuracy appears in several cases to hold up beyond the region

where $\sigma \sim kT$ [18]. Thus we find empirically that we can calculate G_{rlx} in a single step with linear response, even in cases when the simulation data would not by itself give us confidence to do so. Indeed, this empirical robustness is precisely what makes linear response useful.

For the present data, σ = 1 kcal/mol for the 1YCC simulation (considering the perturbation to 2YCC, *i.e.* δq = 1). Thus the integration of the exponential in (1) should be moderately accurate, and we can compare the total relaxation free energy to the linear response approximation with some confidence. Therefore the deviation from linear response seen in this case is probably significant, even though the statistical error in G_1 is of the same magnitude.

Our third observation is that the average perturbation energies, or G_1, appear to be quite noisy, as has been seen in other studies. Thus our 'forward' and 'backward' free energy estimates (lines 1 and 2 in Table 1) agree closely for their relaxation components, but disagree completely for the static components. The 'two-step' estimate (line 5) is effectively the average of the previous two. The real uncertainty in G_1 is therefore at least $\pm$5 kcal/mol, despite the small apparent statistical error.

Our estimate of the oxidation free energy of 1YCC is thus 9$\pm$5 kcal/mol. Our estimate of the relaxation, or reorganization free energy is much more precise: G_{rlx} = -0.80$\pm$0.02 kcal/mol. This low reorganization free energy reflects the low polarizability of the protein matrix around the heme, and the absence of nearby solvent. Indeed, recall that this relaxation free energy includes the contribution of explicit solvent. Bulk solvent is not present however in our simulations, since we simulated the protein in 24 Å water droplets, surrounded by vacuum. A simple Born estimate of the bulk solvent contribution to the relaxation free energy is -7 kcal/mol, giving a total of G_{rlx} = -7.8 kcal/mol, including now the effect of bulk solvent.

In classical electron transfer theory, this reorganization free energy is directly related to the activation free energy G^* for transfering an electron from ferro-cytochrome c to ferri-cytochrome c: $-2G_{rlx} = 4G^*$. Thus $G^* = -G_{rlx}/2$ = 3.9 kcal/mol. In contrast, the free energy to introduce a negative charge into a heme-sized cavity in bulk water would be on the order of -80 kcal/mol, giving an activation free energy of 40 kcal/mol. Thus the low polarizability of the protein matrix around the heme has a dramatic effect on the electron transfer kinetics; we expect that qualitatively similar results would hold for electron transfer between cytochrome c and its physiological partners cytochrome c oxidase and reductase. Similar results were obtained by Warshel and coworkers with a Protein Dipole Langevin Dipole model [33].

Table 1: Heme oxido-reduction: linear response results

	δq^e	G_1	G_2	G_{nl}	ΔG
1YCC[a] $\rightarrow$ 2YCC[b]	-1	4.7 (4)	-0.8 (2)	-0.15	3.8 (4)
2YCC $\rightarrow$ 1YCC	1	-13.6 (4)	-0.8 (2)	0.14	-14.3 (4)
1YCC $\rightarrow$ 1/2YCC[c]	-0.5	2.4 (2)	-0.2 (1)	-0.02	2.1 (2)
2YCC $\rightarrow$ 1/2YCC	+0.5	-6.8 (2)	-0.2 (1)	-0.01	-7.0 (2)
1YCC $\rightarrow$ 1/2YCC $\rightarrow$ 2YCC	-1				9.1[d] (3)

Results are in kcal/mol. Uncertainties on the last digits in parentheses. [a]Ferri-cytochrome c. [b]Ferro-cytochrome c. [c]Half-oxidized, half-reduced intermediate state. [d]Obtained by adding together the two previous half steps. [e]Magnitude of perturbing charge.

3.2 COMD hybrid methodology

We now adopt the hybrid COMD viewpoint to attack the same problem: calculation of the free energy to transfer an electron from ferro-cytochrome c to ferri-cytochrome c. We therefore retain the molecular dynamics estimate of the static component of the free energy, but use a continuum model to calculate the relaxation component. We just saw how well the linear response approximation and the molecular dynamics approach appears to work for this calculation, so that a continuum model would not actually present any practical advantage here. We perform the calculation merely as an illustration of the hybrid methodology; other applications will be presented in the future.

We consider separately the radius-dependent contribution to the self-energy, and the remaining contribution, as discussed above (Methods). The radius-dependent contribution is in fact the Born free energy to embed the perturbing charge in a homogeneous, infinite medium having the dielectric constant of the protein (see [13] for details). In ordinary applications, as pointed out above, the radius-dependent contribution would systematically cancel when a reference compound is considered. In the present application, there is no reference compound, and we will actually explicitly calculate the radius-dependent part, keeping in mind its artificial character. For the protein dielectric constant ϵ_p we assume various values between one and two, based on theoretical calculations of the dielectric constant in the heme region [12].

With $\epsilon_p = 1$ we obtain G_{rlx} = -16.9 kcal/mol. The radius-dependent part of the self-energy is obviously zero (ϵ_p=1), so only the surrounding solvent contributes to the relaxation free energy. This result can be compared with the molecular-dynamics-plus-Born-correction result of -7.8 kcal/mol, which contains however a contribution from the protein medium itself.

With $\epsilon_p = 2$, the self-energy is -8.3 kcal/mol, not including the radius-dependent term. If we incude the radius-dependent term, the total self-energy

is much too large. To reproduce the molecular dynamics result, we require ϵ_p smaller than 2. Assuming $\epsilon_p = 1.2$, we obtain agreement with a radius of 4 Å. This low protein dielectric is not unreasonable, when comparing to molecular dynamics results, which do not contain explicit electronic polarizability. A cavity radius of 4 Å is close to values used in continuum treatments of charged perturbations (photoexcitation) on an indole ring [34].

4 CONCLUSION

We have presented an application of a simple linear response method, proposed earlier [15–17], to calculate electrostatic perturbation free energies in proteins. The linear response approximation appears to be accurate for a perturbing charge inserted on the heme group of cytochrome c. The low polarizability seen here for the heme environment is of obvious functional relevance.

We have also proposed a novel hybrid continuum–molecular dynamics approach for electrostatic free energies. While the application considered was somewhat artificial, the method is quite general, and should be useful for pK_a calculations in particular. It has two advantages in principle: first, it should be able to exploit the robustness of continuum models for handling rather large perturbations, while still allowing the inclusion of important discrete water molecules through the static free energy term, where they are likely to contribute most. Second, it incorporates automatically the distinction between the static and relaxation free energies: it has been pointed out in the past and recently [24, 13, 25] that these two components should probably not be calculated with a single dielectric model, but should each have a specific treatment, as in the present approach.

Acknowledgments

Simulations were done on the C90 computer of the IDRIS supercomputer center of the Centre National de la Recherche Scientifique.

References

1. M. Perutz. Electrostatic effects in proteins. *Science*, 201:1187–1191, 1978.
2. R. Varadarajan, T.E. Zewert, H.B. Gray, and S.G. Boxer. Effects of buried ionizable amino acids on the reduction potential of recombinant hemoglobin. *Science*, 243:69–72, 1989.

3. A.J. Russell and A.R. Fersht. Rational modification of enzyme catalysis by engineering surface charge. *Nature*, 328:496–500, 1987.
4. H. Nicholson, W.J. Becktel, and B.W. Matthews. Enhanced protein thermostability from designed mutations that interact with α helix dipoles. *Nature*, 336:651–656, 1988.
5. S. Bone and R. Pethig. Dielectric studies of protein hydration and hydration-induced flexibility. *J. Mol. Biol.*, 181:323–326, 1985.
6. P.D. Schnier, D.S. Gross, and E.R. Williams. Electrostatic forces and dielectric polarizability of multiply protonated gas-phase cytochrome c ions probed by ion/molecule chemistry. *J. Am. Chem. Soc.*, 117:6747–6757, 1995.
7. D.J. Lockhardt and P.S. Kim. Internal Stark effect measurement of the electric field at the amino terminus of an α helix. *Science*, 257:947–951, 1992.
8. M.A. Steffen, K. Lao, and S.G. Boxer. Dielectric asymmetry in the photosynthetic reaction center. *Science*, 264:810–816, 1994.
9. D.W. Pierce and S.G. Boxer. Dielectric relaxation in a protein matrix. *J. Phys. Chem.*, 96:5560–5566, 1992.
10. C.L. Brooks III, M. Karplus, and M. Pettitt. Proteins: a theoretical perspective of dynamics, structure and thermodynamics. *Adv. Chem. Phys.*, 71:1–259, 1987.
11. A. Warshel. *Computer modelling of chemical reactions in enzymes and solutions*. John Wiley, New York, 1991.
12. Thomas Simonson and David Perahia. Internal and interfacial dielectric properties of cytochrome c from molecular dynamics simulations in aqueous solution. *Proc. Natl. Acad. Sci. USA*, 92:1082–1086, 1995.
13. Thomas Simonson and David Perahia. Microscopic dielectric properties of cytochrome c from molecular dynamics simulations in aqueous solution. *J. Am. Chem. Soc.*, 117:7987–8000, 1995.
14. Thomas Simonson and David Perahia. Dielectric properties of proteins from simulations: tools and techniques. *Comp. Phys. Comm.*, 91:291–303, 1995.
15. T. Simonson, D. Perahia, and G. Bricogne. Intramolecular dielectric screening in proteins. *J. Mol. Biol.*, 218:859–886, 1991.
16. T. Simonson, D. Perahia, and A. T. Brünger. Microscopic theory of the dielectric properties of proteins. *Biophys. J.*, 59:670–90, 1991.
17. R. Levy, M. Belhadj, and D. Kitchen. Gaussian fluctuation formula for electrostatic free energy changes. *J. Chem. Phys.*, 95:3627–3633, 1991.
18. P. Smith and W.F. van Gunsteren. Predictions of free energy differences from a single simulation of the initial state. *J. Chem. Phys.*, 100:577–584,

1994.
19. G.S. Del Buono, F.E. Figueirido, and R. Levy. Intrinsic pKa's of ionizable residues in proteins: an explicit solvent calculation for lysozyme. *Proteins*, 20:85–97, 1994.
20. J. Aqvist, C. Medina, and J.E. Samuelsson. A new method for predicting binding affinity in computer-aided drug design. *Prot. Eng.*, 7:385–391, 1994.
21. L. Landau and E. Lifschitz. *Statistical Mechanics*. Pergamon Press, New York, 1980.
22. R. H. Fowler and E. A. Guggenheim. *Statistical Thermodynamics*. Cambridge University Press, 1939.
23. R. Marcus. *Ann. Rev. Phys. Chem.*, 15:155, 1964.
24. M. Gilson, A. Rashin, R. Fine, and B. Honig. On the calculation of electrostatic interactions in proteins. *Journal of Molecular Biology*, 184:503–551, 1985.
25. L. Krishtalik, A.M. Kuznetsov, and E.L. Mertz. Electrostatics of proteins: a description in terms of two dielectric constants simultaneously. *Biophys. J.*, 70:A225, 1996.
26. Thomas Simonson. Accurate calculation of the dielectric constant of water from simulations of a microscopic droplet in vacuum. *Chem. Phys. Lett.*, 250:450, 1996.
27. P. Smith, R. Brunne, A. Mark, and W.F. van Gunsteren. The dielectric constant of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. *J. Phys. Chem.*, 97:2009–2014, 1993.
28. W. Jorgensen, J. Chandrasekar, J. Madura, R. Impey, and M. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
29. A. M. Berghuis and G. D. Brayer. Oxidation state-dependent conformation changes in cytochrome c. *J. Molec. Biol.*, 223:959–976, 1992.
30. B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. Charmm: a program for macromolecular energy, minimization, and molecular dynamics calculations. *J. Comp. Chem.*, 4:187–217, 1983.
31. A. T. Brünger. X-PLOR *version 3.1, A System for X-ray crystallography and NMR*. Yale University Press, New Haven, 1992.
32. K. Sharp. *DelPhi, Version 3.0*. Columbia University, 1988.
33. A.K. Churg and A. Warshel. Control of the redox potential of cytochrome c and microscopic dielectric properties in proteins. *Biochemistry*, 25:1675–1681, 1986.

34. P.L. Muino and P.R. Callis. Hybrid simulations of solvation effects on electronic spectra: indoles in water. *J. Chem. Phys.*, 100:4093–4109, 1994.