Published in final edited form as:

J Biol Inorg Chem. 2013 August ; 18(6): 599–608. doi:10.1007/s00775-013-1004-6.

Identifying Sequence Determinants of Reduction Potentials of Metalloproteins

Bradley Scott Perrin Jr. and Toshiko Ichiye*

Department of Chemistry, Georgetown University, Box 571227, Washington, DC 20057-1227

Abstract

The reduction potential of an electron transfer protein is one of its most important functional characteristics. While the type of redox site and the protein fold are the major determinants of the reduction potential of a redox active protein, its amino acid sequence may tune the reduction potential as well. Thus, homologous proteins can often be divided into different classes, with each class characterized by a biological function and a reduction potential. Site-specific mutagenesis of the sequence determinants of the differences in the reduction potential between classes should change the reduction potential of a protein in one class to that of the other class. Here, a procedure is presented that combines energetic and bioinformatics analysis of homologous proteins for identifying sequence determinants that are also good candidates for site-specific mutations, using the [4Fe-4S]-ferredoxins and the [4Fe-4S]-HiPIPs as examples. This procedure is designed to guide site-specific mutations or more computationally expensive studies, such as molecular dynamics simulations. To make the procedure more accessible to the general scientific community, it is being implemented into CHARMMing, a web-based portal, with a library of density functional theory results for the redox site that used in the set up of Poisson-Boltzmann continuum electrostatics calculations for the protein energetics.

Keywords

Iron-sulfur proteins; site-specific mutations; HiPIP; ferredoxin; Poisson-Boltzmann continuum electrostatics

Introduction

The standard reduction potential E° of an electron transfer protein is one of its most important functional characteristics since it determines the driving force for electron transfer between a donor and an acceptor. Although the type of redox site is a major factor in determining E° of a protein, differences in the overall protein fold can also affect E° on the order of 1 V while differences in amino acid sequence of homologous proteins may tune E° on the order of 100 mV [1]. Since homologous redox active proteins can often be divided into classes based on differences in both function and E° , identifying the sequence determinants of the differences in E° may also represent sequence differences that alter protein function or activity. In addition, identifying sequence determinants may have potential relevance to understanding genetic disease, genetic engineering, and creating new biomaterials.

Computational studies have played an important role in understanding reduction potentials of metalloproteins. E° is related to the standard free energy of the reduction reaction, ΔG° ,

*To whom correspondence should be addressed. Phone: 202-687-3724, FAX: 202-687-9620, ti9@georgetown.edu.

by $\Delta G^\circ = -nFE^\circ$, where n is the number of electrons and F is Faraday's constant. Furthermore, the energetics can be divided into the inner sphere contribution ΔG_{in} , which is the intrinsic energy required to add an electron to the redox site, and the outer sphere contribution ΔG_{out} , which is the change in the interaction energy of the redox site with the protein and solvent upon reduction of the redox site [2]. Many earlier studies have assumed that ΔG_{in} for a given redox site is independent of its protein so *relative* values of E° between different proteins with the same redox site are calculated [3,4]; our earlier studies also made this assumption [5,6]. However, identifying sequence determinants, which are usually verified by site-specific mutagenesis, is not always straightforward. The most obvious possibilities are charged and polar side chains that affect the electrostatic potential at the redox site by virtue of the electrostatic contributions of the side chain [7]; however, site-specific mutations may lead to unpredictable results [8], or may not even fold properly. In addition, mutations to or from charged side chains are more likely to cause folding problems while mutations involving polar side chains must get the correct orientation of the polar group with respect to the redox site. Other groups have generally focused on the magnitude of the energetic contributions per residue [9], and modeled in various mutants based on changing residue types. However, at times the E° of the modeled mutants have disagreed significantly with experiment [10].

In our previous work, a bioinformatic procedure for identifying candidates for site-specific mutation using energetic analysis of multiple crystal structures plus sequence analysis was developed [2]. In particular, the focus is not on predicting the energetic contribution of the sequence determinant, but whether a site-specific mutation of this determinant is likely to cause a predictable shift in E° . The premise is that when two classes of homologous proteins have different E° , sequence determinants of the difference will be those residues that are different in identity and energetic contribution between the classes but conserved in both within each class. Moreover, the conservation within a class indicates that the contribution of the residue appears robust to many different local protein environments and therefore the residue is a good candidate for site-specific mutation. This is particularly important for polar groups, whose contribution will depend on the orientation with respect to the redox site. Thus, if there are crystal structures available from proteins of both classes, sequence alignments and the energetic contribution of each residue calculated from the crystal structures can be examined for differences between classes and conservation within each class. In addition, since there are many more sequences available than crystal structures, sequence alignments of all available sequences can be examined for consistency with the results based only on the crystal structures, which would further support the identification of the sequence determinant. In addition, by using this energy plus sequence bioinformatic approach, the results are less dependent on the differences in experimental methods used for measuring E° or the accuracy of the crystal structures and calculation methods. For instance, experimental E° are usually collected under different conditions such as ionic concentration, using both voltammetry and potentiometry, which can lead to variations of 20 to 50 mV.

The energy plus sequence bioinformatic procedure has been used to identify a Val versus an Asa in the rubredoxins that shifts E° by ~50 mV between classes by differentially shifting the protein backbone near the redox site in the rubredoxins [5], which has been verified experimentally [11]. In addition, this procedure has been used to identify a Cys versus an Ala in the ferredoxins that shifts E° by ~100 mV again by a backbone shift [6], which is consistent with previous experimental results [12] and has been verified experimentally [13]. However, these calculations used simple Coulomb electrostatics, and thus are less useful when the sequence determinants are not immediately adjacent to the cluster so that their effects are subject to dielectric shielding or when the structural homology of the proteins is decreased by differing size loops or additional secondary structure.

Since these early calculations, many advances in computational methodologies have been made. For instance, calculations of E° vs. the standard hydrogen electrode (SHE) by Noodleman, Case, and co-workers using broken symmetry (BS) [14] density functional theory (DFT) [15] to calculate ΔG_{in} and Poisson-Boltzmann (PB) continuum electrostatics to calculate ΔG_{out} [16] are an important step forward in understanding how both the redox site and the protein contribute. Local density approximation (LDA) functionals were used in the DFT calculations while the PB calculations utilized crystal structures with different dielectric regions for the redox site, protein, and solvent. However, more recent studies indicate that LDA functionals give poor energetics especially for transition metals [17].

Recently, our group has predicted E° vs. SHE for [4Fe-4S] proteins that are in excellent agreement with experiment by adding ΔG_{in} from BS-DFT calculations using hybrid functionals of gas phase redox site analogs and ΔG_{out} from PB calculations of the redox site in the protein using partial charges from the DFT calculation [1,18], in an approach similar to Noodleman, Case, and co-workers. The accuracy of this DFT+PB approach is based in part on our extensive testing of the hybrid (rather than LDA) functionals and basis sets [19] used in the BS-DFT calculation of the analog against electrospray ionization - photoelectron spectroscopy (EI-PES) data [20–22] of the same analogs in the gas phase. Our DFT calculations also confirm that ΔG_{in} for the iron-sulfur clusters is relatively independent of environment [19]. In addition, the overall DFT+PB approach has been shown to be accurate for [4Fe4S] proteins with errors of less than 50 mV for crystal structure that are at least 1.5 Å or better [1], which is a significant improvement over earlier work that calculated relative values of E° using the PDL or MD-PDL methods that had errors of ~370 and ~150 mV, respectively [9].

Here, an improved procedure for identifying sequence determinants of E° that are good candidates for site-specific mutation is presented, which combines our DFT+PB approach and new analysis methods with our earlier procedure for identifying sequence determinants. By replacing simple Coulomb electrostatics with PB electrostatics, dielectric screening by local atomic fluctuations in the protein and the effects of different loop sizes or supernumerary secondary structure are accounted for approximately, although no large scale dynamics is accounted for. Again, the aim is not to predict the exact changes in E° but to predict likely site-specific mutations that will shift E° in predictable directions. The procedure is demonstrated for several different scenarios, using small, water-soluble iron-sulfur electron-transfer proteins [23]. For the first two examples, the 2[4Fe-4S] ferredoxin (Fd) are used because extensive mutagenesis data exists. The pseudo-two-fold symmetric 55-residue fold contains two pseudo-symmetric [4Fe-4S] clusters that undergo reductions of the $[\text{Fe}_4\text{S}_4\text{Cys}_4]$ redox sites utilizing the 2-/3- couple. The simple Fd have only the basic fold while another class (referred to here as II-insertion Fd) has an insertion in the cluster II binding motif and a C-terminal helix near cluster II (Figure 1). Because of the pseudo symmetry, the two clusters of the simple Fd are isopotential with $E^\circ = \sim 430$ mV [24], while somewhat surprisingly in the II-insertion Fd, cluster I is ~200 mV lower while cluster II is only slightly lower than the simple Fd value [25]. In the second two examples, the high potential iron-sulfur proteins (HiPIP) are used, which have less mutagenic data. They are characterized by a buried [4Fe-4S] cluster with loops of different lengths surrounding it and undergo a reduction of the $[\text{Fe}_4\text{S}_4\text{Cys}_4]$ redox site that utilizes the 1-/2- couple (Figure 2). While the HiPIPs found in Ectothiorhodospiraceae (referred to as Ecto-type) have E° of ~150 mV, both the smallest HiPIPs found in Rhodospirillaceae (referred to as Rhodo-type) and the largest HiPIPs found in Chromaticeae (referred to as Chroma-type) have larger E° of ~350 mV for the 1-/2- couple [26,27] (remove ref.28). Lastly, these methods are being implemented into CHARMMing, a web-based interface for CHARMM and other methods [28].

Methods

All crystal structures were obtained from the Protein Data Bank (PDB) [29]. Crystal structures of four Fds were from *Clostridium acidurici* (Ca) at 0.94 Å resolution [2FDN] [30], *Peptostreptococcus asaccharolyticus* (Pa) at 2.00 Å resolution [1DUR] [31], *Pseudomonas aeruginosa* (Pr) at 1.32 Å resolution [2FGO] [25], and *Chromatium vinosum* (Cv) at 2.10 Å resolution [1BLU] [32]. In addition, crystal structures of six HiPIPs were from *Rhodocyclus tenuis* (Rt) at 1.50 Å resolution [1ISU] [33], *Rhodoferrax fermentans* (Rf) at 1.45 Å resolution [1HLQ] [34], *Ectothiorhodospira vacuolata* (Ev2) at 1.80 Å resolution [1HPI] [35], *Ectothiorhodospira halophila* (Eh1) at 2.50 Å resolution [2HIP] [36], *Thermochromatium tepidum* (Tt) at 0.80 Å resolution [1IUA] [37], and *Chromatium vinosum* (Cv) at 1.20 Å resolution [1CKU][38].

Details for the calculations are given elsewhere [18,19] and summarized here. The ΔG_{in} were obtained from BS-DFT calculations at the B3LYP/6-31G(++)_{S**}/B3LYP/6-31G** (energy calculation functional/basis set//geometry optimization functional/basis set) of $Fe_4S_4(SCH_3)_4^{1-2-3-}$ using the program NWChem [39]. The calculations were performed in vacuum for the analog with the dihedral angles C-S-Fe- S_i (where S_i is the cluster sulfur on the opposite plane from the iron) equal to $\sim 60^\circ$ in one plane and $\sim -60^\circ$ in the other plane. ΔG_{in} was -0.232 eV for $Fe_4S_4(SCH_3)_4^{1-2-}$ and 3.452 eV for $Fe_4S_4(SCH_3)_4^{2-3-}$ [18]. The ΔG_{out} for each protein was calculated using Poisson continuum electrostatics for the cluster in the protein surrounded by a continuum solvent utilizing APBS [40], a program for solving the Poisson-Boltzmann equation. A 51.2 Å cubic grid with a grid spacing of 0.2 Å was used for all calculations. The partial charges for the redox site were from the BS-DFT calculations, the atomic radii and the rest of the partial charges were from CHARMM22 parameters [41], and the probe radius for all Connolly [42] surfaces was $r = 1.4$ Å. The dielectric regions were defined as the redox site(s) with $\epsilon_c = 1$, the protein with $\epsilon_p = 4$, and the solvent with $\epsilon_s = 78$, and the ionic concentrations were zero.

The E° in Table 1 for the Fd, which have two clusters, were calculated as follows. For the simple Fd, E° is an average of the values when the opposite cluster is considered as oxidized and reduced, since experiment indicates that the potentials are close enough for either cluster to be reduced first. For the II-insertion Fd, E° for the opposite cluster in the oxidized state was chosen for the cluster with the more positive E° (i.e., cluster II) and E° for the opposite cluster in the reduced state was chosen for the cluster with the more negative E° (i.e., cluster I). However, for the sequence identification in Figures 3 and 4, the opposite cluster was assumed to be in the oxidized state for simplicity. The crystal structure for Pr Fd had two different orientations for Ser 1, one with the side chain oriented toward cluster II and the other with the side chain away; the latter was chosen for the comparison. The burial of the redox site within the protein was described by the parameter R_p and the polarization (or electret) of the entire protein was described by the parameter ϕ_p which were used to approximate E° via the electret-dielectric sphere (EDS) model [43]. The contribution of a residue i was obtained as the difference between ΔG_{out} for the full protein and ΔG_{out} calculated with all of the partial charges for residue i including the backbone set equal to zero. Computational mutations were made by changing the cited residues and minimizing the mutated residues for 50 steps by steepest descent followed by 1000 steps of adopted basis Newton-Raphson minimization while holding the rest of the protein fixed. The Fd sequences were aligned based on previous alignments [6,25] using CaFd for a consensus residue numbering; the HiPIP sequences were also aligned based on previous alignments [44] using CvHiPIP for the consensus residue numbering.

Results and Discussion

Our procedure for identifying sequence determinants that are good candidates for site-specific mutagenesis is demonstrated here. The procedure is based on identifying the sequence determinants that cause the changes in E° between classes of homologous proteins that exhibit different E° . First, E° are calculated using the DFT+PB approach and compared to experiment to check whether PB calculations are likely to capture the origin of the sequence difference. If there are discrepancies for which reasonable explanations cannot be found, PB may not be appropriate for calculating the protein contribution since factors such as dynamics of the protein or change in the redox site may be important. Note that the error in the calculated E° (Table 1) is less than 50 mV for all of the proteins with structures of better than 1.5 Å resolution except for Pr and Cv Fd. The rest of the calculations involve only PB calculations because only the protein contributes. Next, R_p and ϕ_p are examined to see whether the factors responsible for the differences are likely to be in the sequence, indicating potential candidates for site-specific mutagenesis, or in factors that may be harder to control by mutation such as loop length. In particular, if R_p of two classes being examined are different, the differences in E° are likely to be due to difference in loop size or extra secondary structure, which may be difficult to assess by mutation alone. On the other hand, if R_p of two classes being examined are similar but ϕ_p is different, it is likely that the differences in E° are due to the sequence. If a sequence determinant appears likely, sequence alignments of the proteins with crystal structures can then be examined for residues that make a large contribution to E° in one class but not in another. Candidates for sequence determinants are chosen as any residue differing by at least ~50 mV between classes, but are consistent in residue type within a class for the proteins with crystal structure data. Finally, the best candidates for site-specific mutation are chosen from these if they are consistent with sequence data.

Cluster I of the Ferredoxins: an example of similar folds with sequence determinants

The calculated properties of cluster I in the Fd generally indicate that the difference in the E° of this cluster between the simple and II-insertion Fd lies within the sequence (Table 1). The E° of cluster I of the II-insertion Fd relative to the simple Fd is predicted by DFT+PB to be ~200 mV lower in good agreement with experiment, which indicates it is a good candidate for using PB electrostatics for the protein contribution. In addition, since the R_p are only slightly different between the classes, the slight differences in size do not appear to affect E° of cluster I very much. Moreover, since ϕ_p is different between classes, it appears that the differences in E° are due to the sequence.

Two possible sequence determinants at residues 12 and 51 can be identified based on differences in their contributions to the electrostatic potential via a color-coded sequence alignment (Figure 3). At residue 12, the Gly in the simple Fds contributes 50 to 60 mV due to the backbone while an Asp in the II-insertion Fd contributes -20 mV because the side chain adds a negative contribution. Note that although residue 12 is close to cluster I, the Asp side chain has a relatively small contribution since it is shielded dielectrically. In addition, at residue 51, the Ala in the simple Fds contributes ~55 mV also due to the backbone while the Cys in the II-insertion Fd contributes only ~5 mV, because the Cys side chain causes a larger backbone shift away from the redox site as previously noted [6].

To evaluate these two sequence determinants as candidates for site-specific mutations, comparisons to sequence data are made. In comparison to the sequence data with measured E° (Figure 3), at residue 12, all the simple Fd have a Gly while all but one II-insertion Fds have an Asp and, at residue 51, all the simple Fd have an Ala while all but one II-insertion Fds have a Cys. In addition, the one II-insertion Fd with a Gly12 and an Asn 51 has a measured E° that is ~150 mV higher than Cv and Pr Fd, which is consistent with the

predicted effects for residue 12. In a much larger sequence analysis with 16 simple and 30 II-insertion Fds [45], Gly12 is completely conserved in the simple Fd while Asp12 is semi-conserved (40% Asp, 40% Gly, 20% other) in the II-insertion Fd and Ala51 is completely conserved in the simple Fd while Cys51 is semi-conserved (43% Cys, with 46% of the remaining having a Cys one to two residues further C-terminal) in the II-insertion Fd. Thus, the Gly-Asp conversion at residue 12 and the Ala-Cys conversion at residue 51 appear to be good candidates for site-specific mutations, which agrees with previous experimental results (Table 2). (Note, although a shift in E° has been seen in V13G CvFd [46], it would not be identified by our procedure since residue 13 is an Ala in the simple Fd and a Val in the II-insertion Fd, both with similar contributions.)

Cluster II of the Ferredoxins: an example with different size loops and extra secondary structure without sequence determinants

In this example, a sequence determinant is hard to identify since the II-insertion Fds are only ~40 mV in E° for cluster II lower than the simple Fd experimentally. However, this case is interesting because the insertion in the cluster II binding motif and the C-terminal helix near cluster II might be expected to result in E° being much lower for the II-insertion Fd than the simple Fd. In addition, it is instructive to look at since the E° for cluster II of the II-insertion Fd calculated using DFT+PB are ~100 mV too negative compared to experiment (Table 1).

For this cluster, R_p differ more between the classes, so the lowering of the calculated E° of II-insertion Fd is partially due to the greater burial of the cluster, which can be estimated to lower it by 130 mV using the EDS model. However, other factors that are not accounted for by the simple DFT+PB approach may cause the actual E° to be raised, such as dynamic exposure of the redox site by movement of the cluster II insertion or the C-terminal helix, which can be tested by removing the insertion and the helix and looking for changes in E° . For instance, replacement of the cluster insertion by Ala-Gln resulted in an experimental increase of 44 mV but only a calculated increase of 18 mV, indicating that the calculated cluster insertion contribution is somewhat too negative but not enough to explain the difference between the calculated and experimental E° . On the other hand, removal of the entire C-terminal helix in the P56- mutant (removal of all residues from residue 56 according to consensus numbering to the C-terminus) resulted in a calculated increase of 145 mV although there are no experimental results. This indicates that the C-terminal helix buries D55, E60, and K68 in the calculations so that they are in a low dielectric region. While E60 and K68 form a salt bridge so that their net contributions cancel, movement of the C-terminal helix in solution might expose these residues to solvent so that they are in a high dielectric region in solution and the large negative contribution of D55 may be screened. Although *Bacillus schlegelii* Fd is from a different family of Fd, NMR structures indicate that the C-terminal helix moves to expose D55 and thus decrease its negative contribution. This indicates that a molecular dynamics simulation of CvFd to test the dynamic exposure maybe worthwhile.

While the above results indicate uncertainty in the validity of using PB electrostatics to calculate the entire protein contribution, the homologous regions of the simple and II-insertion Fd can be examined for other possible residues that could contribute to the lowering in the II-insertion Fd, with the above caveat. Two possible sequence determinants are weakly identified in the alignment (Figure 3). At residue 20, a Val in the simple Fds contributes ~40 mV while an Asn in the II-insertion contributes ~3 mV and at residue 38, an Ile in the simple Fds contributes ~50 mV while a Thr in the II-insertion Fd contributes ~10 mV; hydrogen bonding of Thr38 to Asn20 stabilizes the orientation that results in the decreased contribution in the II-insertion Fd (Figure 1). In comparison to the other sequences with measured E° , the conservation of the two candidates is less complete than for cluster I and the simple Fd with an Asn20 actually has a slightly higher E° . In the much

larger sequence analysis [45], Val20 is semi-conserved (69% Val) in the simple Fd while Asn20 is semi-conserved (50% Asn) in the II-insertion Fd and Ile38 is semi-conserved (69% Ile) in the simple Fd while Thr38 is semi-conserved (73% Thr) in the II-insertion Fd. Altogether, this indicates that either mutation may have little effect by itself; while a double mutation of Val20Asn and Ile38Thr might be a reasonable candidate.

Rhodospirillaceae vs. Ectothiorhodospiraceae HiPIP: an example of a loop size difference causing differences in E°

The calculated properties of Rhodo- versus Ecto-type HiPIP generally indicate that the difference in their E° do not lie within the sequence (Table 1). The calculated E° are predicted in good agreement with experiment, particularly the ~200 mV greater E° of the Rhodo-type relative to the Ecto-type so DFT+PB appears to be a reasonable method for calculating E° and PB is a reasonable method for calculating the protein contribution. However, a large difference in R_p between the classes indicate the greater exposure of the redox site in the Rhodo-type is responsible for much of the difference, estimated as 40 mV using the EDS model. In addition, the difference in ϕ_p indicates the electret may contribute as well. Examining the color-coded sequence alignment (Figure 4), the N-terminus appears slightly more negative in the Ecto-type and a few other slightly more positive regions are apparent in the Rhodo-type. These results are consistent with earlier calculations indicating the charged side chains maybe generally responsible for the differences [47]. However, no good candidates for site-specific mutagenesis are readily identifiable (Figure 4). Moreover, the features identified as being responsible such as the greater exposure of the redox site in the Rhodo-type appear difficult to test via mutation.

Chromatiaceae vs. Ectothiorhodospiraceae HiPIP: an example with different size loops and a sequence determinant

On the other hand, the calculated properties of Chromo- versus Ecto-type HiPIP generally indicate that the difference in their E° do lie within the sequence (Table 1). The calculated E° are predicted again in good agreement with experiment, particularly the ~200 mV greater E° of the Chromo-type relative to the Ecto-type. However, now the difference in R_p between the classes is much smaller, and in fact, would cause changes in the opposite direction of the difference in E° . Moreover, since the ϕ_p are different and correlate with the observed difference in E° , the cause of the difference appears to lie in the sequences.

Examining the sequence alignment (Figure 4), differences at residue 79 result in a large difference in the contribution to E° between the Chromo- and Ecto-types and thus is a candidate for a sequence determinant of E° . An Ala or Val at residue 79 in the Ecto-type HiPIPs has a small contribution of 10 mV while a Ser in the Chroma-type has a contribution of ~65 mV due mainly to the positive contribution of the side chain. Comparing sequence data of HiPIPs with E° data, residue 79 is also a Ser in the other Chroma-type HiPIP while it is a Val or Ala in the other Ecto-type. In the larger sequence data [44], Ser79 is conserved in the 12 Chroma-type while in the 9 Ecto-type, residue 79 is also semi-conserved for Ala (78%, with the remainder Ser or Val). Thus, the analysis predicts an Ala vs. a Ser at residue 79 as a potential mutable sequence determinant for the HiPIPs (Table 2).

Conclusions

Understanding how sequence differences between homologous proteins lead to differences in reduction potentials is essential for understanding sequence-structure-function relationships in redox-active proteins. Here, an improved procedure for identifying sequence determinants for site-specific mutagenesis is presented using the DFT+PB approach, although other computational methods may be used instead of DFT+PB. First, the E°

calculated using the DFT+PB approach for homologous proteins with crystal structures are compared to experiment E° to test whether PB calculations are appropriate. Next, PB electrostatic calculations of crystal structures are used to localize and identify residues that contribute differently to the calculated E° . Finally, comparisons with sequence data take advantage of the vastly greater amount of sequence versus structural data to examine the robustness of a candidate for mutation.

The methods are demonstrated here via the identification of several mutable sequence determinants of the reduction potential. Gly12 vs. Asp12 and Ala51 vs. Cys51 for Cluster I of the simple vs. II-insertion Fd, respectively, have been identified in good agreement with existing experimental data. In addition, Val20, Ileu39 vs. Asn20, Thr 38 for Cluster II of the simple vs. II-insertion Fd, respectively, have been weakly identified and Ala79 vs. Ser79 in the Chromo- vs. Ecto-type HiPIPs, respectively, have been identified; although these mutations have not yet been tested. In addition, the differences between the Rhodo- and Ecto-type appear to be caused by differences in loop sizes, which are much more difficult to test experimentally, and many small changes due to the charged residues, which are also difficult to test experimentally. Finally, the disagreement between calculated and experimental E° for cluster II of the simple vs. II-insertion Fd appears to lie in the dynamics of the supernumerary secondary structure.

The current redox module in CHARMMing allows users to upload coordinates for iron-sulfur proteins and calculate their reduction potentials, all on the server. In addition, although beyond the scope of the work here, molecular dynamics simulations of these proteins can be set up in the CHARMMing portal; for instance, the abovementioned dynamics of the supernumerary secondary structure in the II-insertion Fd can be examined. However, such a study requires considerably more effort and resources than may be desired for simply identifying mutations since to simulate the timescales needed for proper sampling, the input files generated by CHARMMing must be downloaded and run on a machine accessible to the user. Future implementations will permit mutations within the CHARMMing as well as generation of the color-coded sequences shown here and calculation of R_p and ϕ_p .

Acknowledgments

This work was supported by a grant from the National Institutes of Health (GM0453030) and by the Intramural Research Program of the National Institutes of Health, National Heart, Lung, and Blood Institute in the Laboratory of Computational Biology (Z99-TW999999-03). The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government. The continuum electrostatic calculations were performed on computers funded through the William G. McGowan Foundation and Georgetown University. Both authors thank Kelly N. Tran and Mingliang Tan for careful reading of the manuscript.

References

1. Perrin BS Jr, Ichiye T. Proteins: Struct, Funct, Bioinf. 2010; 78:2798.
2. Ichiye, T. Computational Biochemistry and Biophysics. Becker, O.; MacKerell, JA.; Roux, B.; Watanabe, M., editors. Marcel Dekker; New York: 2001. p. 393
3. Langen R, Jensen GM, Jacob U, Stephens PJ, Warshel A. J Biol Chem. 1992; 267:25625. [PubMed: 1464583]
4. Gunner MR, Honig B. Proc Natl Acad Sci U S A. 1991; 88:9151. [PubMed: 1924378]
5. Swartz PD, Beck BW, Ichiye T. Biophys J. 1996; 71:2958. [PubMed: 8968568]
6. Beck BW, Xie Q, Ichiye T. Biophys J. 2001; 81:601. [PubMed: 11463610]
7. Warshel A, Papazyan A, Muegge I. J Biol Inorg Chem. 1997; 2:143.
8. Zeng Q, Smith ET, Kurtz DM, Scott RA. Inorg Chim Acta. 1996; 242:245.

9. Stephens PJ, Jollie DR, Warshel A. *Chem Rev.* 1996; 96:2491. [PubMed: 11848834]
10. Chen K, Tilley GJ, Sridhar V, Prasad GS, Stout CD, Armstrong FA, Burgess BK. *J Biol Chem.* 1999; 274:36479. [PubMed: 10593945]
11. Eidsness MK, Burden AE, Richie KA, Kurtz DMJ, Scott RA, Smith ET, Ichiye T, Beard B, Min T, Kang C. *Biochemistry.* 1999; 38:14803. [PubMed: 10555962]
12. Iismaa SE, Vazquez AE, Jensen GM, Stephens PJ, Butt JN, Armstrong FA, Burgess BK. *J Biol Chem.* 1991; 266:21563. [PubMed: 1657971]
13. Kümmerle R, Gaillard J, Kyritsis P, Moulis J-M. *J Biol Inorg Chem.* 2001; 6:446. [PubMed: 11372203]
14. Noodleman, L.; Case, DA. *Adv Inorg Chem.* Richard, C., editor. Vol. 38. Academic Press; 1992. p. 423
15. Parr, RG.; Yang, W. *Density-Functional Theory of Atoms and Molecules.* Oxford University Press; Oxford: 1989.
16. Mouesca JM, Chen JL, Noodleman L, Bashford D, Case DA. *J Am Chem Soc.* 1994; 116:11898.
17. Zhao Y, Truhlar DG. *Acc Chem Res.* 2008; 41:157. [PubMed: 18186612]
18. Perrin BS Jr, Niu S, Ichiye T. *J Comput Chem.* 2013; 34:576. [PubMed: 23115132]
19. Niu S, Ichiye T. *Mol Simul.* 2011; 37:572.
20. Zhai H, Yang X, Fu Y, Wang X, Wang L. *J Am Chem Soc.* 2004; 126:8413. [PubMed: 15237997]
21. Fu Y-J, Yang X, Wang X, Wang L-S. *Inorg Chem.* 2004; 43:3647. [PubMed: 15180419]
22. Wang XB, Wang LS. *J Phys Chem.* 2000; 112:6959.
23. Meyer J. *J Biol Inorg Chem.* 2008; 13:157. [PubMed: 17992543]
24. Stombaugh NA, Sundquist JE, Burris RH, Orme-Johnson WH. *Biochemistry.* 1976; 15:2633. [PubMed: 181047]
25. Giastas P, Pinotsis N, Efthymiou G, Wilmanns M, Kyritsis P, Moulis J-M, Mavridis IM. *J Biol Inorg Chem.* 2006; 11:445. [PubMed: 16596388]
26. Luchinat C, Capozzi F, Borsari M, Battistuzzi G, Sola M. *Biochem Biophys Res Commun.* 1994; 203:436. [PubMed: 8074688]
27. Heering HA, Bultink YBM, Hagen WR, Mayer TE. *Biochemistry.* 1995; 34:14675. [PubMed: 7578075]
28. Miller BT, Singh RP, Klauda JB, Hodoscek M, Brooks BR, Woodcock HL III. *J Chem Inf Model.* 2008; 48:1920. [PubMed: 18698840]
29. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. *Nucleic Acids Res.* 2000; 28:235. [PubMed: 10592235]
30. Dauter Z, Wilson KS, Sieker LC, Meyer J, Moulis J-M. *Biochemistry.* 1997; 36:16065. [PubMed: 9405040]
31. Backes G, Mino Y, Loehr TM, Meyer TE, Cusanovich MA, Sweeney WV, Adman ET, Sanders-Loehr J. *J Am Chem Soc.* 1991; 113:2055.
32. Moulis JM, Sieker LC, Wilson KS, Dauter Z. *Protein Sci.* 1996; 5:1765. [PubMed: 8880900]
33. Rayment I, Wesenberg G, Meyer TE, Cusanovich MA, Holden HM. *J Mol Biol.* 1992; 228:672. [PubMed: 1453470]
34. Gonzalez A, Benini S, Ciurli S. *Acta Crystallogr, Sect D: Biol Crystallogr.* 2003; 59:1582. [PubMed: 12925788]
35. Benning MM, Meyer TE, Rayment I, Holden HM. *Biochemistry.* 1994; 33:2476. [PubMed: 8117708]
36. Breiter DR, Meyer TE, Rayment I, Holden HM. *J Biol Chem.* 1991; 266:18660. [PubMed: 1917989]
37. Liu L, Nogi T, Kobayashi M, Nozawa T, Miki K. *Acta Crystallogr, Sect D: Biol Crystallogr.* 2002; D58:1085. [PubMed: 12077426]
38. Parisini E, Capozzi F, Lubini P, Lamzin V, Luchinat C, Sheldrick G. *Acta Crystallographica, Section D.* 1999; 55:1773.
39. Valiev M, Bylaska EJ, Govind N, Kowalski K, Straatsma TP, van Dam HJJ, Wang D, Nieplocha J, Apra E, Windus TL, de Jong WA. *Comput Phys Commun.* 2010; 181:1477.

40. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. *Proc Natl Acad Sci U S A*. 2001; 98:10037. [PubMed: 11517324]
41. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. *J Phys Chem B*. 1998; 102:3586.
42. Connolly ML. *Science*. 1983; 221:709. [PubMed: 6879170]
43. Perrin BS Jr, Ichiye T. *J Biol Inorg Chem*. 2013; 18:103. [PubMed: 23229112]
44. Van Driessche G, Vandenberghe I, Devreese B, Samyn B, Meyer TE, Leigh R, Cusanovich MA, Bartsch RG, Fischer U, Beeumen JJV. *J Mol Evol*. 2003; 57:181. [PubMed: 14562962]
45. Fajardo, MJ. Washington State University; 2004.
46. Kyritsis P, Hatzfeld OM, Link TA, Moulis J-M. *J Biol Chem*. 1998; 273:15404. [PubMed: 9624123]
47. Bertini I, Gori-savellini G, Luchinat C. *J Biol Inorg Chem*. 1997; 2:114.
48. Hochkoeppler A, Ciurli S, Venturoli G, Zannoni D. *FEBS Lett*. 1995; 357:70. [PubMed: 8001683]
49. Moulis JM, Davas V. *Biochemistry*. 1995; 34:16781. [PubMed: 8527453]
50. Battistuzzi G, Mariapina D, Borsari M, Sola M, Macedo A, Moura J, Pedro R. *J Biol Inorg Chem*. 2000; 5:748. [PubMed: 11129002]
51. Clements AP, Kilpatrick L, Lu WP, Ragsdale SW, Ferry JG. *J Bacteriol*. 1994; 176:2689. [PubMed: 8169218]
52. Daas P, Hagen W, Keltjens J, Vogels G. *FEBS Lett*. 1994; 356:342. [PubMed: 7805869]
53. Gao-Sheridan H, Pershad H, Armstrong F, Burgess B. *J Biol Inorg Chem*. 1998; 273:5514.
54. Boll M, Fuchs G, Tilley G, Armstrong FA, Lowe DJ. *Biochemistry*. 2000; 39:4929. [PubMed: 10769152]
55. Saeki K, Tokuda K-i, Fukuyama K, Matsubara H, Nadanami K, Go M, Itoh S. *J Biol Chem*. 1996; 271:31399. [PubMed: 8940149]
56. Przysiecki CT, Meyer TE, Cusanovich MA. *Biochemistry*. 1985; 24:2542. [PubMed: 3925987]
57. Banci L, Bertini I, Capozzi F, Carloni P, Ciurli S, Luchinat C, Piccioli M. *J Am Chem Soc*. 1993; 115:3431.

Figure 1.
(a) Pa Fd (left) and Pr Fd (right) are shown as ribbons with the redox site and sequence determinants as balls and sticks. Cluster I is the upper cluster in both.

Figure 2. Rhodo-type (top), Ecto-type (middle), and Chroma-type (bottom) HiPIPs are shown as ribbons with the redox site and sequence determinants as ball and sticks.

Figure 3.

Sequence alignment of Fds colored by contribution of each residue to the E° of Cluster I (Top) and Cluster II (Bottom). Colors indicate contributions of entire residue to E° . Dark blue are >50 mV, light blue is 50 to 25 mV, white -25 to 25 mV, pink -50 to -25 mV, and red is less than -50 mV. Residues are numbered relative to Ca Fd. Residues ligated to the redox site are in bold with a line below the alignment connecting the residues. References for reduction potentials are given in Table I except for Ct1 [49], Cp [50], Met [51], Mb [52], AvvI [53], Ta [54], and Rcl [55].

Figure 4.

Sequence alignment of HiPIPs colored by contribution of each residue to the E° as in Figure 3. Residues are numbered relative to Cv HiPIP. Residues ligated to the redox site are in bold with a line below the alignment connecting the residues. References for reduction potentials are given in Table I except for Rge [26], Rt3 [56], Rgl [26], Ev2[26], Eh1[26], Tr [56], Cg [27], and Tp [57].

Table 1

Predicted E° , R_p and ϕ_p , and experimental E° for homologous ferredoxins (Fd) and homologous high-potential iron-sulfur proteins (HiPIP).

Protein	Cluster	Exp E° (V) ^a	Cal E° (V)	R_p (Å)	ϕ_p (eV)
Ca Fd	I	-0.43	-0.47	8.7	0.67
Pa Fd	I	-0.43	-0.32 ^b	8.7	0.74
Pr Fd	I	-0.66	-0.70	9.0	0.51
Cv Fd	I	-0.66	-0.70 ^b	9.1	0.46
Ca Fd	II	-0.43	-0.44	8.8	0.68
Pa Fd	II	-0.43	-0.40 ^b	8.9	0.62
Pr Fd	II	-0.48	-0.59	10.3	0.50
Cv Fd	II	-0.46	-0.56 ^b	10.1	0.55
Rt HiPIP	I	0.36	0.42 ^b	9.8	0.48
Rf HiPIP	I	0.32	0.35	9.9	0.46
Ev2 HiPIP	I	0.16	0.22 ^b	10.3	0.30
Eh1 HiPIP	I	0.13	0.10 ^b	10.5	0.13
Tt HiPIP	I	0.33	0.33	10.6	0.45
Cv HiPIP	I	0.35	0.38	10.7	0.52

^aCaFd and PaFd from Ref. [24], PrFd from Ref. [25], CvFd from Ref. [46], Rt HiPIP from Ref. [27], Rf HiPIP from Ref. [48], Ev2 and Eh1 from Ref [26], TtHiPIP from Ref. [37], Cv HiPIP from Ref. [26]

^bResolution of crystal structure is 1.5 Å or poorer.

Table 2

Predicted sequence determinants and ΔE° , the deviation in calculated and experimental reduction potential relative to protein type 1.

Residue Number	Residue in Protein Type1	Residue in Protein Type 2	ΔE° (mV)	
			Cal	Exp
12	Gly in simple Fd, cluster I	Asp in II insertion Fd, cluster I	-73 ± 7	-77^a
51	Ala in simple Fd, cluster I	Cys in II insertion Fd, cluster I	-54 ± 8	$-72^b, -60^c$
20+38	Val+Ile in simple Fd, cluster II	Ans+Thr in II insertion Fd, cluster I	-77 ± 11	NA
79	Ala/Val in Ecto HiPIP	Ser in Chroma HiPIP	$+41 \pm 13$	NA

^a Asp12Gly in CvFd[46]

^b Cys51Ala in CvFd[13]

^c Ala51Cys in *Pyrococcus furiosus* (personal communication from MWW Adams)