A Multigeneration Analysis of Cytochrome b_{562} Redox Variants: Evolutionary Strategies for Modulating Redox Potential Revealed Using a Library Approach†

Stacy L. Springs,* Susanna E. Bass, Greg Bowman, Ilana Nodelman, C. E. Schutt, and George L. McLendon

Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Received November 19, 2001; Revised Manuscript Received January 23, 2002

ABSTRACT: The redox potential of cytochromes sets the energy yield possible in metabolism and is also a key determinant of the rate at which redox reactions proceed. Here, the heme protein, cytochrome b_{562}, is used to study the in vitro evolution of redox potential within a library of variants containing the same structural archetype, the four-helix bundle. Multisite variations in the active site of cytochrome b_{562} were introduced. A library of variants containing random mutations in place of R98 and R106 was created, and the redox potentials of a statistical sampling of this library were measured. This procedure was carried out for both the low- and high-potential variants of a previously studied F61X/F65X, first-generation library [Springs, S. L., Bass, S. E., and McLendon, G. L. (2000) *Biochemistry 39*, 6075]. The second-generation library reported here has a range of redox potentials which is greater than 40% (160 mV) of the known accessible potential among cytochromes with identical axial ligands (but different folds) and exceeds the range exhibited phylogenetically by the cytochrome *c′* family which internally maintains the same axial ligation and fold. A statistical analysis of the libraries examined reveals that the redox potential of WT cyt b_{562} is found at the high-potential extremum of the distribution, indicating that this protein apparently evolved to differentially stabilize the reduced protein. The 2.7 Å crystal structure of F61I/F65Y/R106L (low-potential variant of the second-generation library) was solved and is compared to the wild-type structure and the 2.2 Å resolution structure of the F61I/F65Y variant (low-potential variant of the first-generation library). The structures indicate that charge−dipole effects are responsible for shifting the redox equilibrium toward the oxidized state in both the F61I/F65Y and F61I/F65Y/R106L variants. Specifically, a new protein dipole is introduced into the heme microenvironment as a result of the F65Y mutation, two new internal water molecules (one in hydrogen-bonding distance of Y65) are found, and in the case of F61I/F65Y/R106L ($\Delta E_m = 158$ mV vs NHE), increased solvent exposure of the heme as a result of the R106L substitution is identified.

One general question in molecular level evolution is: "How does function evolve within a fixed protein architecture to exploit the range of possible chemical reactivity?" For example, consider a heme redox protein, whose reactivity can, in part, be defined by a thermodynamic metric, the redox potential, $E°$. The active site heme has its reactivity altered by incorporation into a protein matrix, which provides a different local environment than does bulk solvent. As Hecht has shown, proteins found in evolutionarily naive libraries can bind heme selectively and with a range of redox potentials.[1] As heme-containing proteins evolve to meet a specific physiological niche, two broad outcomes are possible. A variant may be chosen "robustly" near the center of the chemically permitted range of redox potentials. Alternatively, selection might drive toward an extremum in the range of potentials permitted. Given that recognition of redox partners and tunneling reactivity can be dramatically altered by changing the fold of a cytochrome, here we only address "evolution" within a fixed topology, rather than coevolution of functions such as redox potential and molecular recognition.

We consider here the four-helix bundle motif, which is observed in several cytochromes (cyt),[2] including the cytochrome *c′* family, cytochrome b_{562}, and cytochrome b_{559} (*1*, *2*). Within this set of four-helix bundle proteins, a reduction potential range of ∼400 mV is exhibited: *Rhodospirillum rubrum* cytochrome *c′* has a reduction potential of −8 mV vs NHE, and cyt b_{559} exhibits a potential of 400 mV vs NHE. Cyt b_{562}, the subject of our study, has a potential which lies between these two extremes at 167 mV vs NHE (Figure 1). Nature can access such a range by tuning factors unrelated to global structure: covalent vs noncovalent heme attachment, heme substituents, heme axial ligands, heme solvent exposure, surface charge of the cytochrome, and protein electrostatic effects all play a role in mediating reduction potential (*2*, *3*). The effects of changing axial ligands and

† This work was supported by the NIH (Grants GM18627-02, GM33881, and GM44038).

* To whom correspondence should be addressed. E-mail: ssprings@princeton.edu. Telephone: (609) 258-1371. Fax: (609) 258-6746.

[1] Professor Michael Hecht, personal communication.

[2] Abbreviations: cyt, cytochrome; WT, wild type; NHE, normal hydrogen electrode.

10.1021/bi012066s CCC: $22.00
Published on Web 03/09/2002

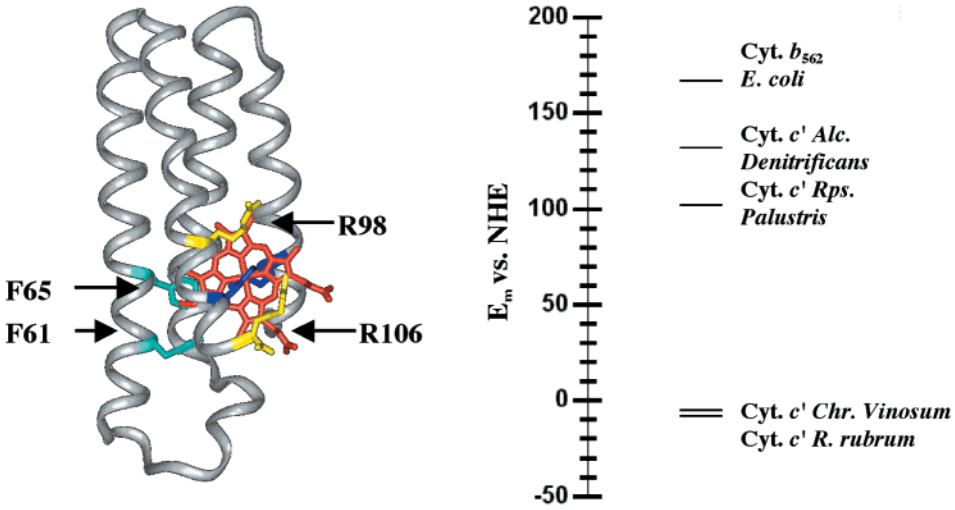

FIGURE 1: The four-helix bundle structural type is represented by cytochrome b_{562}. The second-generation library of cyt b_{562} variants contains randomization at positions 98 and 106 (shown in yellow). Positions 61 and 65 (shown in cyan) were randomized in the first-generation library (*4*). On the right, the range of redox potential in a series of four-helix bundle proteins is displayed. Cytochrome b_{559}, a four-helix bundle cytochrome with bis-Met axial ligation and a midpoint redox potential of 400 mV vs NHE, is not shown (*2*). The range among the cyt *c′* family, which maintain identical axial coordination, is approximately 140 mV.

heme substituents are well described for cytochromes. Thus, it is constructive to consider the range observed in a family of cytochromes (cytochromes *c′*), which have not only the same structural type but consistent axial iron coordination. In such a case, a range of 140 mV is observed phylogenetically. This provides a benchmark for the range that might be expected among a family of structural homologues of cytochrome b_{562}.

The work presented here uses a library approach to examine the impact of electrostatic effects on redox potential within a fixed structural archetype, the four-helix bundle. This approach provides more information than isolated analyses of redox variants. This is because (a) a statistical analysis of the range and variance of redox potential when wild-type residues are varied for all possible amino acid residues in the heme pocket of a cytochrome can be performed and (b) variants with statistically large shifts in the redox potential (ΔE_m) can be identified in an unbiased way. Rather than choosing a variant for study that a priori appears as though it may exhibit a large shift in potential, the library contains essentially complete information about the variations. We have found that the most interesting variations in redox potential may result from apparently conservative replacements. Given this, redox/structure analysis of variants with statistically significant potential shifts can be instructive in a search to understand the relative importance of contributing factors. In addition, such a library provides a data set for testing our quantitative means of predicting reduction potential given a structural input.

In previous work, we examined the effect of simultaneous replacements at amino acid positions F61 and F65 of cytochrome b_{562} (*4*). In this work, we have used the high- and low-potential variants of the first generation of redox variants as parents for a second generation of variants containing mutations at positions R98 and R106 (see Figure 1). The redox potentials of the second-generation library and a redox/structure analysis of the low-potential variant (based on a 2.7 Å X-ray crystal structure) are presented.

EXPERIMENTAL PROCEDURES

Cloning and Mutagenesis. PCR-based mutagenesis was carried out as described in ref *4*. In this case, different oligonucleotide primers (OEP1B, AK55HATAG, Pos98_106_P3, and OEP4B) were used to generate the library of mutations (*4*−*6*). The oligonucleotides used were as follows: OEP1B, 5′-biotin-CGCGCAGAATTCGAGCTCGGTACCCGGGCGAAT-3′; AK55HATAG, 5′-GGCAACAGGAAATGAGGAATTASNNATACTTCTGGTGATAGGCGTTSNNGGTCGTTTTCAG-3′; Pos98_106_P3, 5′-TTCCTCATTTCCTGTTGCCTGCACTCAGG-3′; OEP4B, 5′-biotin-CGCGGCGGATCCCGACGGCAAATTTGTGCAG-3′. All additional cloning procedures were carried out as in ref *4*.

Protein Purification. Purification was carried out as described in ref *4*. The integrity of all mutants was confirmed by electrospray ionization mass spectrometry on a Hewlett-Packard 5989B mass spectrometer and DNA sequencing analysis on an ABI PRISM 377 sequencer.

Spectroelectrochemistry. The spectroelectrochemical titrations were performed under anaerobic conditions in an optically transparent thin-layer cell that has been described previously (*4*). All measurements were made in 20 mM potassium phosphate buffer (pH = 7.2) and 100 mM NaCl with hexaammineruthenium chloride ($E°$ vs NHE = 60 mV), 2,6-dichlorophenolindophenol sodium salt ($E°$ vs NHE = 220 mV), or 2,3,5,6-tetramethylphenylenediamine ($E°$ vs NHE = 260 mV) as the mediator. The protein concentrations were approximately 5.0×10^{-5} M, and the mediator concentrations were between 0.5 and 1.0×10^{-3} M. The experimental procedures and data analysis were carried out as described in ref *4*.

X-ray Crystallography. X-ray quality crystals of the F61I/F65Y/R106L variant were obtained at 4 °C using the hanging drop vapor diffusion method (*7*). X-ray diffraction data were recorded on a Rigaku system using a copper rotating anode electrode and an R-axis II image plate detector. X-ray data used for structure determination were collected at 100 K from a single crystal, grown in 30% PEG 4000, 0.1 M $MgCl_2$,

and 0.1 M Tris-HCl, pH 8.5. The reflections were integrated and merged using DENZO and SCALEPACK in the HKL suite of programs (*8*). The F61I/F65Y/R106L variant crystallizes in the space group $P2_12_12_1$ with unit cell dimensions $a = 42.1$ Å, $b = 49.0$ Å, and $c = 114.4$ Å and contains two molecules in the asymmetric unit.

Using the atomic coordinates of WT cyt b_{562} as a search model for molecular replacement, solutions for both noncrystallographic symmetry mates were obtained with the program AMORE (*9*). To reduce the bias of WT cyt b_{562} in the structure of the variant, side chains of the mutated residues (i.e., positions 61, 65, and 106) in the initial model for building and refinement were replaced with alanine. Initial cycles of refinement were accomplished using simulated annealing, positional refinement, and *B*-factor group minimization within the CNS program suite (*10*), and rebuilding of the model into the electron density maps was performed with the modeling program O, version 6.2 (*11*). After the third annealing, further refinement was achieved using only cycles of positional refinement, *B*-factor group refinement, and building.

RESULTS

Library Design. In the second generation of cytochrome b_{562} variants created here, wild-type cyt b_{562} and F61I/F65Y (the high- and low-potential variants of the first generation, respectively) were used as the parental DNA for new cyt b_{562} progeny which include mutations at positions 98 and 106 (Figure 1). As outlined in our previous report, the necessary criteria for choice of cytochrome include use of a well-characterized protein for which structural analyses of variants might be possible and use of a cytochrome which will allow a straightforward screen for variants with only local structural changes rather than global conformational changes (*4*, *12*). These criteria were used to select cytochrome b_{562} as a physical system in which to study reduction potential mediation. Cyt b_{562} is a 12.3 kDa, 106 amino acid protein consisting of four α-helices connected by three turns to form a four-helix bundle (*13*). It occurs naturally and can be overexpressed in the periplasm of *Escherichia coli* (*14*). The structure of oxidized cytochrome b_{562} has been solved in solution by NMR (*15*) and by X-ray crystallography to 1.4 Å resolution (*16*). The heme of cyt b_{562} is noncovalently bound by axial ligands Met 7 and His 102 (*13*). The heme is oriented such that one face, adjacent to the Met 7 ligand, is buried within the hydrophobic core of the protein, while the face adjacent to the His 102 is partially solvent accessible (*13*). As demonstrated previously, the noncovalently bound heme of cytochrome b_{562} allows a first-order screen for global structural homology between the wild-type protein and variants (*4*, *5*). This is because *E. coli* harboring variants that have maintained their ability to fold and bind heme exhibit a red phenotype, while those which no longer fold properly or bind heme maintain the background color of *E. coli*.

Positions 98 and 106 of cytochrome b_{562} are located in the heme microenvironment (Figure 1) and are therefore expected to play a role in mediating reduction potential. Unlike the nonpolar phenylalanine residues at positions 61 and 65 that were the subject of the first-generation library, WT cyt b_{562} contains charged arginine residues at positions 98 and 106. A priori, it is unclear whether replacement with amino acids that alter the solvent exposure of the heme or those which influence the charge will have the greater impact on redox potential at these positions.

A complete understanding of the effects of mutation at positions 98 and 106 on reduction potential can be gained by randomization of the amino acids at these two positions and characterization of the reduction potentials of the resulting variants. In addition to insight into the specific role of R98 and R106, a multigenerational approach [using the extrema of the first generation of redox variants (F61X/F65X) as parents for the second generation], allows the accessible range and distribution of redox potential available given limited internal variation near the heme of a four-helix bundle cytochrome to be further probed.

To create a library of variants at positions 98 and 106, the DNA of WT cyt b_{562} and of the F61I/F65Y variant was randomized (for all 20 naturally occurring amino acids) at positions corresponding to amino acids 98 and 106, using the overlap−extension PCR-based mutagenesis procedure and color screen previously described.[3] Approximately 30 members exhibiting a red phenotype were randomly selected from the two sublibraries of second-generation variants.

Sequencing Analysis. The diversity of the library at the genetic level was determined by analysis of the DNA sequences of 46 clones prior to library screening.[4] An NNS randomization was used; therefore, the expected nucleotide frequency at N positions is 25% G, A, C, and T, while the expected nucleotide frequency at the S positions is 50% G and 50% C. The actual frequencies were found to differ at N positions with G and T overrepresented (33% and 32%, respectively) and A and C underrepresented (15% and 20%). At S positions, C was present 37% of the time, while G was present 63% of the time.[5] This type of bias is common and must be assessed when considering the distribution of amino acids occurring in the library chosen on the basis of a red phenotype. When the amino acid distribution in a library parented by WT cyt b_{562} (based on an NNS randomizaton and including the bias just described) is analyzed before and after screening for red phenotype (as in ref 4), the data clearly show that the wild-type residue, arginine, is significantly overrepresented relative to its occurrence in the library analyzed prior to screening for red phenotype. The same is true for the library parented by the F61I/F65Y variant. Considerably fewer double (or, in the case of the F61I/F65Y parent, quadruple) variants were observed than would be statistically expected given our sequencing data. This result indicates that R98 and R106, independent of their role in attenuating redox potential, likely stabilize protein folding and/or heme binding in cytochrome b_{562}.

[3] To ensure that a statistical analysis would give meaningful results, the randomization was kept at a maximum of two positions per generation of library. In this way, the maximum possible library size is limited to 400 members, so that one need only sample ca. 15 members to ensure with high confidence levels that the range of redox potential observed corresponds to the range existing in the entire population.

[4] This number excludes several clones which exhibited frame-shift mutations.

[5] Such bias can originate at the solid-phase oligonucleotide synthesis level or as a result of selection at the DNA synthesis level during the polymerase chain reaction. This type of bias is common. (See, for instance, ref *24*.)

Table 1: Reduction Potentials of the Second Generation Variants

variant	E_m (mV vs NHE)	ΔE_m ($E_m^{wt} - E_m^{v}$)
F61Y F65Y R98K R106Q	62 ± 5	105
F61I F65YR98K	66 ± 4	101
F61I F65Y R98G	57 ± 2	110
F61I F65Y R98A	65 ± 5	102
F61I F65Y R98L	7 ± 2	160
F61I F65Y R98F[a]	34 ± 14	133
F61I F65Y R98T[a]	47 ± 8	120
F61I F65Y R106K	79 ± 2	88
F61I F65Y R106W	42 ± 3	125
F61I F65Y R106L	9 ± 1	158
R98W R106K	67 ± 3	100
R98K R106G	103 ± 2	64
R98L R106Q[a]	116 ± 9	51
R98A R106S	60 ± 2	107
R98V R106Q	116 ± 3	51
R98H	120 ± 4	47
R98E	104 ± 3	63
R98A	146 ± 5	21
R98F	125 ± 4	42
R98G	117 ± 1	50
R98L	111 ± 1	56
R98P	157 ± 3	10
R98W	66 ± 4	101
R98K	160 ± 5	7
R106S	107 ± 2	60
R106K	145 ± 2	22
R106M	99 ± 1	68
R106E	77 ± 2	90

[a] These variants were eliminated from the study due to the high error values associated with their reduction potentials. Two variants were eliminated from the study due to anomalous mass spectroscopic data (F61I/F65Y/R98S and R98S/R106C). R106M and R106K were prepared independently and are not bona fide library members.

Redox Potentials. The midpoint redox potentials for wild-type cytochrome b_{562}, the F61I/F65Y variant, and the second-generation variants were determined spectroelectrochemically and are given in Table 1 and graphically displayed in Figure 2. As in the first-generation library, virtually all of the variants (whether parented by the WT or F61I/F65Y structure) exhibited reduction potentials lower than their parent (Figure 2). The second-generation sublibrary parented by F61I/F65Y exhibited a 72 mV range in reduction potential. The sublibrary parented by the WT structure exhibited a range of 107 mV. The overall range exhibited by the second-generation variants is 160 mV. The low-potential variants of the library parented by the F61I/F65Y mutant are F61I/F65Y/R106L and F61I/F65Y/R98L, exhibiting reduction potentials of 9 and 7 mV vs NHE, respectively.

A series of variants containing changes at position 98 (R98E, R98A, R98L, and R98W) are of particular interest. Replacing arginine with a glutamic acid (R98E) replaces a positive charge in the heme microenvironment with a negative charge. The reduction potential of R98E is 104 mV vs NHE, a 62 mV shift from the wild type. The R98A, R98L, and R98W mutations replace a positively charged residue with a nonpolar side chain. These variants exhibit reduction potentials of 146, 111, and 65 mV vs NHE, respectively. This raises several questions. Why does replacing arginine with leucine or tryptophan elicit an equal or greater shift in reduction potential than replacing arginine with glutamic acid? In a series of nonpolar replacements, why does the smaller hydrophobic residue, alanine, elicit less of a response than leucine? What is more important: charge or solvent exposure of the heme (which will be controlled by the volume, hydrophobicity, *and* orientation of an amino acid given local structural constraints)? In the case of position 98, the data suggest that solvent exposure is as important as or more important than charge. This issue will be revisited later in the context of detailed structural information.

One possible effect of replacing a charged amino acid with a large, nonpolar amino acid would be to *decrease* solvent exposure and increase the reduction potential. Such an effect might be expected to counterbalance the effect of removing the positively charged arginine side chain, the magnitude of which is difficult to assess given the positional uncertainty of the arginine side chains in solution. In this scenario, the observation of only marginally shifted redox potentials [certainly more marginal than mutations resulting in total charge reversal, R98E ($\Delta E_m = 62$ mV)] might be expected. However, this is not what is observed. Within experimental error, an R to L mutation is worth as much as an R to E mutation.

In the crystal structure of cyt b_{562}, C_β and C_γ of the R98 and R106 side chains are in van der Waals contact with the solvent-exposed face of the heme, while the remainder of each side chain extends into solvent. Thus, it is possible that the role of the arginine residues at positions 98 and 106 is twofold: the charged guanidinum portion of the side chain destabilizes the oxidized state through charge−charge interactions, and the hydrocarbon portion of the side chain destabilizes the oxidized state by reducing the solvent exposure of the heme. A corollary to this would be that a leucine side chain in position 98 or 106 does not reduce the solvent exposure of the heme as effectively as arginine. Structural data are, of course, necessary to test this hypothesis and shed light on the effect of replacing arginine with a hydrophobic residue such as leucine.

X-ray Crystallographic Analysis of the F61I/F65Y/R106L Variant. In an effort to understand the structural factors giving rise to the significantly decreased reduction potentials of the F61I/F65Y and F61I/F65Y/R106L variants (the low-potential variants of their respective generations), the crystal structures of these two variants were solved. The 2.2 Å crystal structure of the F61I/F65Y variant will be described in detail elsewhere. The salient features of this structure are described here for the purpose of comparison only. Table 2 contains data collection and refinement statistics for the F61I/F65Y/R106L structure.

The global structures of the two molecules contained within the asymmetric unit are essentially identical, with an overall rms deviation of 0.4 Å. No obvious hydrogen-bonding contacts between the interfacial residues exist, and unlike the structure of WT cyt b_{562}, the crystal contacts do not involve the surface with the solvent-exposed heme. In the variant structure, clear interdigitation of some residues, particularly around histidine 63 of molecule B, is observed between the two molecules of the asymmetric unit.

The F61I/F65Y/R106L variant differs from WT cyt b_{562} primarily in the turn regions and in the microenvironment of the heme. Figure 3a shows the structure of the variant molecule B superimposed on the WT structure (PDB access no. 256B). The rms deviation for molecules A and B of the variant superimposed onto WT are 0.60 and 0.67 Å, respectively.

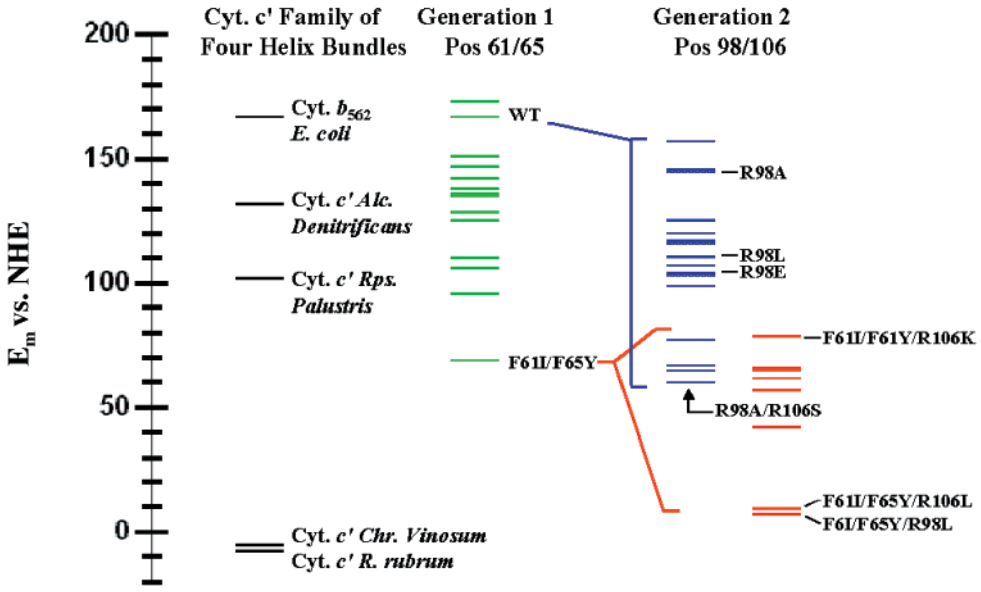

FIGURE 2: Range of redox potential in the first- and second-generation libraries. The overall range exhibited is 166 mV. With the exception of one, all variants studied exhibit lower redox potentials than wild-type cytochrome b_{562}. See text for discussion of the redox potentials of specific variants.

Table 2: Data Collection and Refinement Statistics for the F61I/F65Y/R106L Structure[a]

(a) data statistics	
space group	$P2_12_12_1$
unit cell (Å)	$a = 42.1$, $b = 49.0$, and $c = 114.4$
resolution (Å)	50−2.7
molecules per asymmetric unit	2
R_{merge} (%)	8.5 (28.7)
no. of reflections	49249
completeness (%)	98.2 (98.4)
$I/\sigma(I)$	8.2
(b) refinement statistics	
$R_{working}$ (%)	23.2
R_{free} (%)	33.3
average *B*-factor (Å^2), all atoms	39.6

[a] The numbers in parentheses are values for the highest resolution bin (2.8−2.7).

Positions 61, 65, and 106. In the variant structure, there are clear, but subtle differences between the two molecules of the asymmetric unit with respect to the side chains at positions 61 and 106. Figure 3b depicts a CPK rendering of the heme of both molecules of the unit cell, as compared to the WT structure. At position 106, the leucine residues are oriented with torsion angles which are opposed to one another in molecules A and B. In molecule B, the leucine residue is oriented such that the terminal methyl groups are pointed up, away from the heme. This creates a solvent-exposed channel, leading from the iron atom at the center of the heme to the solvent. In molecule A, the terminal methyl groups are pointed "down" toward the face of the heme ligated by histidine. This orientation creates a pocket of open space between the histidine and the leucine large enough to accommodate a water molecule. There is also variation between the isoleucine residues at position 61. Again, the origin of this variation is the orientation of the terminal methyl groups.

At position 65, the replacement of the WT phenylalanine with a tyrosine side chain in the variant induces a slight shift in the placement of this residue. This is a similar feature of the F61I/F65Y structure. With the additional hydroxyl group, the tyrosine is no longer accommodated in the same space as the phenylalanine which it replaced. The residue is therefore shifted partially into the space which was occupied by the phenylalanine at position 61 in the WT structure. In the variant, however, the phenylalanine at position 61 has been replaced with an isoleucine. The combination of the isoleucine and the tyrosine leaves a small space open in which a solvent channel could form, as it does in the F61I/F65Y variant (see Figure 3c).

The shift of the tyrosine residue at position 65 in the F61I/F65Y/R106L variant is consistent with the crystal structure of the F61I/F65Y variant (Figure 3c). This structure was solved to 2.2 Å resolution, and clear electron density corresponding to water molecules (shown in red in Figure 3c) is observed in this case. In the F61I/F65Y structure, the hydroxyl group of the tyrosine is hydrogen bonded to a water molecule (2.85 Å), which is, in turn, hydrogen bonded to a second water molecule (2.85 Å). This creates a hydrogen-bonding network and introduces two molecules of water into a space that, in the WT cyt b_{562}, is completely hydrophobic. The introduction of these water molecules into the heme microenvironment provides a rationale for the low-reduction potential of this variant. Given that the F61I/F65Y and F61I/F65Y/R106L variants exhibit a nearly identical structural arrangement near I61 and Y65, it is probable that this variant is also participating in a hydrogen-bonding network with buried water molecule(s).

In the WT crystal structure, the terminal guanidinium groups of arginine 98 and 106 project into the solvent, but C_β and C_γ are in van der Waals contact with the heme, protecting the heme from solvent. In the F61I/F65Y/R106L variant, there are two critical differences in the new amino acid side chain at position 106: (1) leucine is completely nonpolar and (2) leucine does not contain an extended hydrocarbon chain, as does arginine. Analysis of the F61I/F65Y/R106L structure shows that the heme is more solvent exposed in the variant than in the wild-type structure. As mentioned, the orientation of the leucine residues in the two molecules of the asymmetric unit is slightly different, but in

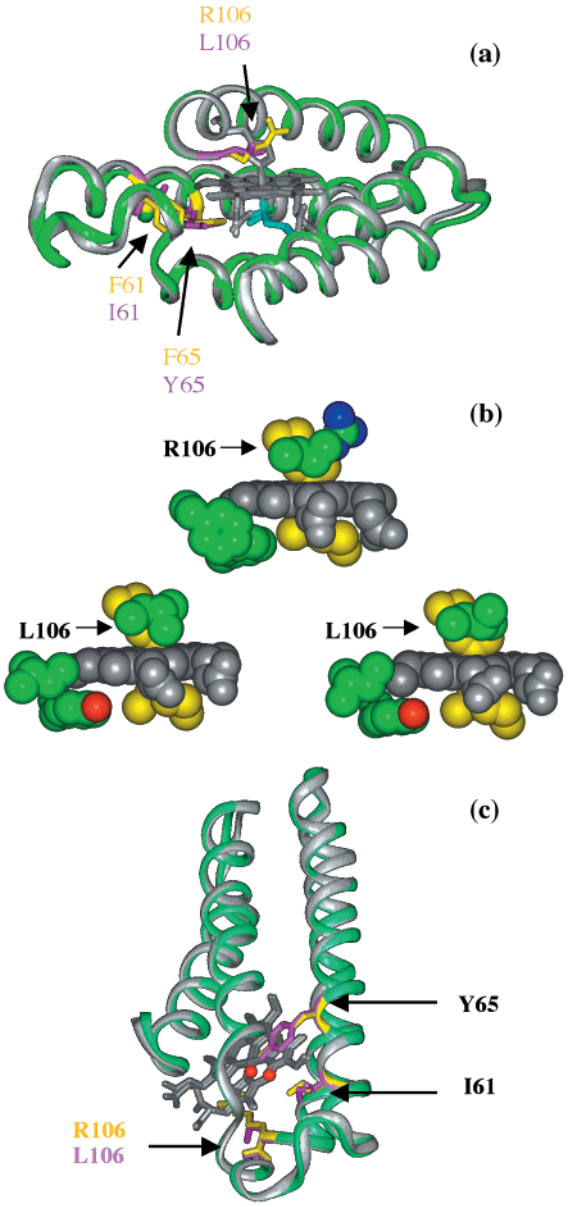

FIGURE 3: (a) Superimposition of the 1.4 Å resolution structure of wild-type cytochrome b_{562} and the 2.7 Å resolution structure of the F61I/F65Y/R106L variant. WT b_{562} is shown with a green ribbon, and residues 61, 65, and 106 are highlighted in yellow. The F61I/F65Y/R106L variant is shown with a gray ribbon and residues 61, 65, and 106 highlighted in purple. This superimposition shows the global homology between these two structures, which have redox potentials which differ by 160 mV. (b) CPK rendering of the heme and three sites of mutation for WT cytochrome b_{562} (PDB access no. 256B) (top) and the variant F61I/F65Y/R106L molecule A (lower left) and molecule B (lower right). At position 106, the replacement of an arginine side chain with leucine leads to increased solvent exposure of the heme. (c) Superimposition of the 2.2 Å resolution structure of the F61I/F65Y variant with the 2.7 Å resolution structure of the F61I/F65Y/R106L variant. The F61I/F65Y variant is depicted with a green ribbon and yellow side chains, and the F61I/F65Y/R106L variant is depicted with a gray ribbon and purple side chains. Two internal water molecules (shown in red) can be easily identified in the 2.2 Å structure of F61I/F65Y. Given the similarity between these structures, it is possible that such a network of water contributes to the low potential of the F61I/F65Y/R106L variant also.

both, the structure suggests the possibility of a solvent channel existing in the variant structure that is not present in the WT structure. Due to the lower resolution of the variant structure, density corresponding to specific water molecules in the pocket of space created by the location of leucine 106 was not observed. However, a water molecule occupies a box with rough dimensions of 1.19 Å × 0.91 Å × 0.45 Å; the space left open by the "pocket" in molecule A is large enough to encompass this volume.

The solvent accessibilities of the heme for both molecules of WT cyt b_{562} and the F61I/F65Y/R106L variant were calculated using the Surface subroutine in the CCP4 suite of programs (*17*). The C_β, C_γ, O_1, and O_2 atoms of the heme propionates were not included in the calculation, due to the uncertainty of their location in solution. The solvent-exposed surface area for the two molecules of wild-type cyt b_{562} and the F61I/F65Y/R106L variant are 93.8, 102.2, 116.2, and 114.0 Å^2, respectively. Thus, the heme group of the F61I/F65Y/R106L variant (in both molecules of the unit cell) is more solvent exposed than the WT heme. It is interesting to speculate that because leucine is a γ-branched amino acid, this mutation leaves the heme even more exposed than an analogous mutation to alanine (the shift in redox potential for R98A was only 21 mV; that for R98L was 56 mV).

DISCUSSION

We have generated a library of cytochrome b_{562} variants and examined the redox potentials of a statistical sampling of this library. For a statistical analysis, the sample data need to be corrected for bias in the population. An intrinsic bias is introduced by the differing codon frequencies, but additional bias is introduced by the mutagenesis procedures. Such bias, since it can be independently assessed, can be corrected by probability weighting the sample to estimate the mean, the variance, and the range (*18*). The statistics must be calculated independently for each of the two sublibraries that comprise the second generation. For variants which resulted from the wild-type parent, the unweighted mean is 111 mV, and the probability weighted mean is 113 mV. For the sublibrary originating from the F61I/F65Y parent, the unweighted mean is 52 mV, and the probability weighted mean is 65 mV. For the library parented by WT cyt b_{562}, the WT redox potential is clearly at the extremum of the distribution. The most conservative possible estimate of the range is given by Chebyshev's theorem (*19*), which makes no assumption about the form of the distribution. By this estimate, the sample data reported here represent at least 75% of the total possible range for each sublibrary.

The redox potential of heme proteins with the same axial ligands can be found to vary by approximately 400 mV (*3*). The second-generation library of cyt b_{562} mutants which maintain identical axial coordination and global structure exhibits a range of 160 mV. This means that, through mutation of only four residues, greater than 40% of the known potential possible has been accounted for in this library. When the comparison is restricted to the structurally homologous family of cytochromes c' (which contain identical axial ligands among one another), a range of 140 mV is observed phylogenetically. This range is easily recaptured by only two generations of mutation of WT cyt b_{562}.

The color screen employed in these studies returns members of the redox library on the basis of structural integrity and heme-binding affinity. If either of these is compromised severely, a variant is selected against. Given

this and the fact that the redox potential of a heme/protein complex can be linked in a thermodynamic cycle to the oxidized/reduced binding equilibrium, it is important to consider whether the color screen preferentially selects variants which bind heme better than wild type and if this skews distribution of redox potentials returned. The color screen is performed on whole cells overexpressing cytochrome b_{562} and variants in the periplasm. The environment of the periplasm is significantly more reducing than the redox potential of WT cyt b_{562}; thus, the proteins are present in the reduced oxidation state when subjected to the color screen. If bias existed among the variants returned, it would be for variants which bind heme in the reduced state with a higher affinity than wild type, thus returning potentials artificially skewed to higher values of the redox potential than the WT protein. Our experimental data do not indicate that such bias exists. While it is theoretically possible that the heme affinity threshold of the reduced variants in the periplasm could put a lower limit on the potentials returned in the selection, this does not appear to be the case as the lowest potential variant of the second generation of redox mutants has a redox potential of 7 mV vs NHE, which is approaching the value for a solvent-exposed heme with identical axial ligands.

No correlation between redox potential and side-chain hydrophobicity (or volume) was observed among as many as 8−10 variants at positions 61, 65, and 98, respectively (data not shown). Each protein variant must be analyzed in structural detail in order to fully understand the consequences of a specific mutation. For example, an analysis of the R98E, R98A, R98L, and R98W variants shows that the change from a positively charged arginine residue to a neutral tryptophan residue produces a greater shift in reduction potential than does the change from arginine to a negatively charged glutamic acid. This certainly could not be predicted on the basis of simple concepts such as side-chain hydrophobicity or volume.

In the absence of structural data, it is difficult to rationalize reduction potential shifts such as those mentioned above. An analysis of the crystal structure of the low-potential variant from the second-generation library, the F61I/F65Y/R106L variant, reveals the types of structural rearrangements that can lead to statistically large changes in redox potential. This variant exhibited a midpoint reduction potential of 9 mV vs NHE, a 158 mV shift from WT, and a 58 mV shift from the F61I/F65Y variant. One might predict that such a shift in reduction potential has occurred because the positive charge on the arginine, which destabilizes the oxidized state of the heme, has been replaced with an uncharged residue. However, such a prediction would be incomplete. Reduction potential mediation is governed not only by the presence/absence of charged residues in the heme microenvironment but by many factors, including the heme solvent exposure (*3*, *20*). The 2.7 Å resolution crystal structure of this variant, solved here, indicates that the placement of a leucine at position 106 leaves the heme face more exposed to solvent than the arginine it replaced. The structure suggests that heme solvent exposure may be increased in two locations: at position 106 and also near positions 61 and 65. The crystal structure of the F61I/F65Y variant (the low-potential variant from the first generation) reveals that the tyrosine at position 65 is slightly shifted from the position of the phenylalanine in the WT structure, allowing room for two water molecules near the heme. While no water molecules were observed directly in the F61I/F65Y/R106L structure, the similarity between the two structures just mentioned provides an indicator that a solvent channel may exist near position 61/65 of F61I/F65Y/R106L also. Calculation of the solvent-accessible area surrounding the heme reveals that the F61I/F65Y/R106L heme is more solvent exposed. In both mutants for which structural data exist, there is a significant increase in the interaction of the heme with water. Water molecules in close proximity to the heme stabilize the charged oxidized state of the heme relative to the neutral reduced state and result in a lowering of the reduction potential.

While individual variants, such as the low-potential variants discussed above, shed light on the physical aspects of reduction potential mediation, the library of variants, as a unit, could be useful for studying quantitative means of analyzing reduction potential. This is because the library identifies variants that exhibit large shifts in potential. The composition of variants defining the redox extrema of each generation is not a priori chemically intuitive. Variants identified with the library should make interesting case studies for fine-tuning our ability to calculate redox potential. Such calculations are in progress.

While it is possible, we believe it unlikely that subsequent mutagenesis at additional sites in the heme microenvironment would reveal a dramatically greater range of accessible reduction potential. Cytochrome b_{562} can only tolerate limited mutations at the heme binding site before heme binding and protein stability become severely compromised. There is already some indication of this in the second-generation library, where only one variant containing substitutions at positions 61, 65, 98, and 106 survived the color screen and analysis. By contrast, the majority of randomly selected variants in the first-generation library contain mutations at both position 61 and position 65.

CONCLUSIONS

Exhaustive multigenerational evolution has been employed to explore the range of redox potential available with fixed ligation and within a single structural archetype. Our model archetype is the four-helix bundle of the type exhibited by cytochrome b_{562} and the cytochrome c' family. Through this process, we discovered that the entire phylogenetic range (140 mV) for homologous four-helix bundle structures which maintain self-consistent axial ligation (i.e., cytochomes c') can be recapitulated through a very restricted set of mutations (positions 61, 65, 98, 106) in cytochrome b_{562}. These changes cannot be predicted a priori from the type of substitution: "conservative" nonpolar → nonpolar substitutions can have larger effects than do polar → nonpolar or even substitutions resulting in charge reversal. Consequently, structural data are currently required to predict redox potential.

We suggest that the range observed through two generations of "in vitro" evolution represents the intrinsic chemical structural limits possible within the cytochrome b_{562} family (variation to axial ligands excluded). The wild-type protein represents one extreme in potential: evolution has driven the structure to be as reducing as this four-helix bundle allows. The triple mutant F61I/F65Y/R98L approaches an alternate extremum. This variant shows stable heme incor-

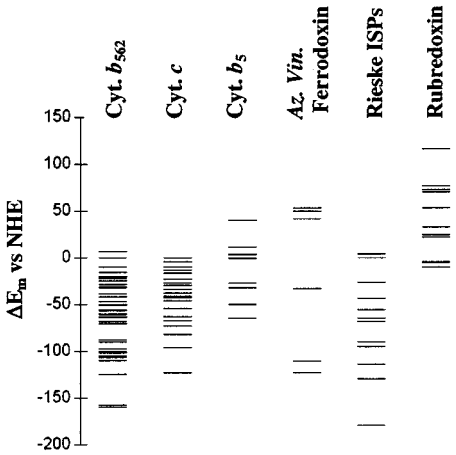

FIGURE 4: Shift in redox potential from the wild-type value for variants of several different cytochromes and iron−sulfur cluster proteins. In the case of cytochrome b_{562}, the potentials shown are from a statistical sampling of variants from three combinatorial libraries. In the case of cyt b_{562}, it is strictly true that the wild-type potential lies at near the extremum of the distribution. In the case of cytochrome *c* and the other proteins shown, the potentials have resulted from many independent studies, and therefore, statistical analysis of these distributions is not rigorous. Despite this, the data provide a strong indication that some of these proteins may be evolving toward a redox potential extremum.

poration into a structure in which the main chain fold is essentially identical to wild type. Nonetheless, the limiting potential approaches that of a fully exposed heme with equivalent substituents and axial ligands ($E_m = \sim -50$ mV). Such studies suggest how readily (redox) function can be tuned by a very few substitutions within a fixed archetype. Initially, incorporation of heme into a protein may have a minimal effect on redox potential (relative to a protein-free heme with equivalent ligation). A challenge for biology, then, is to attenuate heme redox potential by judicious combinations of charge and solvation effects. Our studies show that, within the fixed b_{562} structural archetype, the system can rapidly converge toward maximum stabilization, which the wild-type protein seems to represent.

All of the variants in the b_{562} library, save one, exhibited a reduction potential lower than that of their parent. This characteristic is striking and points toward a fundamental characteristic of cytochrome evolution. Furthermore, this result is not unique to the b_{562} protein. An analysis of redox variants of several cytochromes and iron−sulfur proteins[6] indicates that, in some cases (i.e., cyt *c*, cyt b_{562}, Rieske ISP's, and rubredoxin), the wild-type protein has likely evolved toward a redox potential extremum (see Figure 4). Having said this, evolution toward an extremum need not be the case for all redox proteins. A counter example can be found in variants of the *Rhodobacter sphaeroides* reaction center, where the wild-type potential lies near the midpoint of the variants examined thus far (*25*, *26*). In the case of cytochrome b_{562} it is likely that further evolution toward higher redox potentials (by alteration of electrostatics in the heme region) will be limited. For the other proteins examined in Figure 4 it is not obvious if this is the case as the data were not collected in a statistically rigorous manner. Nonetheless, the pattern appears to be clear and poses an interesting challenge for evolutionary theory.

[6] The ΔE_m values shown in Figure 4 were calculated from a variety of literature sources. (See for instance, refs *21*−*23*.)

ACKNOWLEDGMENT

Some of the issues presented in this work build on pioneering studies by Barbara Burgess. We therefore wish to dedicate this paper to her memory. She will be missed.

REFERENCES

1. Brayer, G. D., and Murphy, M. E. P. (1996) in *Cytochrome c, A Multi-Disciplinary Approach* (Scott, R. A., and Mauk, A. G., Eds.) pp 101−166.
2. Moore, G. R., and Pettigrew, G. W. (1990) *Cytochromes c: Evolutionary, Structural and Physiochemical Aspects*, Springer-Verlag, Berlin.
3. Mauk, A. G., and Moore, G. R. (1997) *J. Biol. Inorg. Chem. 2*, 119.
4. Springs, S. L., Bass, S. E., and McLendon, G. (2000) *Biochem*istry *39*, 6075.
5. Kadowaki, H., Kadowaki, T., Wondistord, F. L., and Taylor, S. I. (1989) *Gene 76*, 161.
6. Reidharr-Olsen, J. F., et al. (1991) *Methods Enzymol. 208*, 564.
7. McRee, D. E. (1993) *Practical Protein Crystallography*, Academic Press, New York.
8. Otwinowski, Z., and Minor, W. (1997) *Methods in Enzymology, Vol. 276, Macromolecular Crystallography, Part A* (Carter, C. W., Jr., and Sweet, R. M., Eds.) pp 307−326, Academic Press, New York.
9. Navaza, J. (1994) *Acta Crystallogr. A50*, 157−163.
10. Brunger, A. T., et al. (1998) *Acta Crystallogr. D54*, 905.
11. Jones, T. A., Zou, J.-Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110.
12. Brunet, A., Huang, E. S., Huffine, M. E., Loeb, J. E., Weltman, R. J., and Hecht, M. H. (1993) *Nature 364*, 355.
13. Mathews, F. S., Bethge, P. H., and Czerwinski, E. W. (1979) *J. Biol. Chem. 254*, 1699.
14. Nikkila, H., Gennis, R. B., and Sligar, S. G. (1991) *Eur. J. Biochem. 202*, 309.
15. Arnesano, F., Banci, L., Bertini, I., Faraone-Mennella, J., and Rosato, A. (1999) *Biochemistry 28*, 8657.
16. Hamada, K., Bethge, P. H., and Mathews, F. S. (1995) *J. Mol. Biol. 247*, 947.
17. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol. 55*, 379.
18. Kish, L. (1965) *Survey Sampling*, John Wiley and Sons, New York.
19. Triola, M. (1998) *Elementary Statistics*, Addison-Wesley, Reading, MA.
20. Tezcan, F. A., Winkler, J. R., and Gray, H. B. (1998) *J. Am. Chem. Soc. 120*, 13383.
21. Geuergova-Kuras, M., Kuras, R., Ugulava, N., Hadad, I., and Crofts, A. R. (2000) *Biochemistry 39*, 7436.
22. Chen, K., et al. (1999) *J. Biol. Chem. 274*, 36479.
23. Xiao, Z., et al. (2000) *J. Biol. Inorg. Chem. 5*, 75.
24. Reidharr-Olsen, J. F., and Sauer, R. T. (1988) *Science 241*, 53.
25. Williams, J. C., Haffa, A. L. M., McCulley, J. L., Woodbury, N. W., and Allen, J. P. (2001) *Biochemistry 40*, 15403.
26. Ivancich, A., Artz, K., Williams, J. C., Allen, J. P., and Mattioli, T. A. (1998) *Biochemistry 37*, 11812.

BI012066S