Protein Thermal Denaturation, Side-Chain Models, and Evolution: Amino Acid Substitutions at a Conserved Helix−Helix Interface†

Gary J. Pielak,*,‡ Douglas S. Auld,‡,§ James R. Beasley,‡ Stephen F. Betz,‡,∥ David S. Cohen,⊥ Donald F. Doyle,‡ Shelley A. Finger,‡ Zoey L. Fredericks,‡,# Sharon Hilgen-Willis,‡,○ Aleister J. Saunders,⊥ and Sonja K. Trojak‡

Department of Chemistry and Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599

Received November 2, 1994; Revised Manuscript Received December 27, 1994⊗

ABSTRACT: Random mutant libraries with substitutions at the interface between the N- and C-terminal helices of *Saccharomyces cerevisiae* iso-1-cytochrome *c* were screened. All residue combinations that have been identified in naturally occurring cytochrome *c* sequences are found in the libraries. Mutants with these combinations are biologically functional. Enthalpies, heat capacities, and midpoint temperatures of denaturation are used to determine the entropy and Gibbs free energy of denaturation (ΔG_D) for the ferri form of the wild-type protein and 13 interface variants. Changes in ΔG_D cannot be allocated solely to enthalpic or entropic effects, but there is no evidence of enthalpy−entropy compensation. The lack of additivity of ΔG_D values for single versus multiple amino acid substitutions indicates that the helices interact thermodynamically. Changes in ΔG_D are not in accord with helix propensities, indicating that interactions between the helices and the rest of the protein outweigh helix propensity. Comparison of ΔG_D values for the interface variants and nearly 90 non-cytochrome *c* variants to side-chain model data leads to several conclusions. First, hydrocarbon side chains react to burial-like transfer from water to cyclohexane, but even weakly polar side chains respond differently. Second, despite octanol being a poor model for protein interiors, octanol-to-water transfer free energies are useful stability predictors for changing large hydrocarbon side chains to smaller ones. Third, unlike cyclohexane and octanol, the Dayhoff mutation matrix predicts stability changes for a variety of substitutions, even at interacting sites. Furthermore, a correlation is observed between stability changes and the growth rates of yeast harboring the variants. In relation to protein evolution, interface variants possessing residue combinations found in naturally occurring cytochrome *c* sequences are the most stable, and the data support the neutral theory of macromolecular evolution.

Our goal is to understand the relationship between the sequence of a protein and its structure, stability, and function. One way of reaching this goal is to seek meaningful relationships between the equilibrium thermodynamics of denaturation, side-chain model compound data, and protein evolution. To this end, we targeted one of the most highly conserved regions of a widely studied protein, the interface between the N- and C-terminal helices of eukaryotic cytochromes *c*.

Eukaryotic cytochromes *c*, the penultimate electron-transfer proteins of the respiratory chain, are well suited for probing relationships between sequence, stability, and function because so much is known about these proteins (Pettigrew & Moore, 1987; Moore & Pettigrew, 1990). Three-dimensional structures from six species and the amino acid sequences from 106 species are available. Iso-1-cytochrome *c* from the yeast *Saccharomyces cerevisiae*, the protein used in the present studies, is a particularly good candidate for examining these relationships. The structures of the reduced and oxidized forms are known (Berghuis & Brayer, 1992) and have been examined by NMR (Gao et al., 1991). The structure of the oxidized form complexed with one of its physiological redox partners, cytochrome *c* peroxidase, is known (Pelletier & Kraut, 1992), and the first-order electron-transfer rates for this and other cytochrome *c* complexes have been determined (Willie et al., 1993). Furthermore, both oxidation states denature by a reversible, two-state process (Betz & Pielak, 1992; Bixler et al., 1992; Hilgen-Willis et al., 1993; Liggins et al., 1994; Cohen & Pielak, 1994, 1995). Finally, random mutagenesis of the cloned gene (Smith et al., 1979) can be used to produce libraries of mutants. Introduction of these libraries into yeast strains which lack cytochrome *c* allows identification of functional and nonfunctional mutants (Wang & Pielak, 1991; Auld & Pielak, 1991; Fredericks & Pielak, 1993).

The helix−helix interface is an ideal site for probing relationships between the equilibrium thermodynamics of thermal denaturation, side-chain model data, and protein evolution for several reasons. The interface comprises two

† Supported by NIH Grant GM42501.

* Author to whom correspondence should be addressed: phone, (919) 966-3671; FAX, (919) 962-2388; E-mail, gjpielak@uncvx1.oit.unc.edu.

‡ Department of Chemistry, University of North Carolina.

§ Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02138.

∥ Present address: Dupont-Merck Pharmaceutical Co., P.O. Box 80328, Wilmington, DE 19880.

⊥ Department of Biochemistry and Biophysics, University of North Carolina.

Present address: Department of Medicine, Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710.

○ Present address: Department of Microbiology, School of Dental Medicine, University of Pennsylvania, Philadelphia, PA 19104.

⊗ Abstract published in *Advance ACS Abstracts,* February 15, 1995.

0006-2960/95/0434-3268$09.00/0

Table 1: Residues at Positions 6, 10, 94, and 97[a]

6	10	94	97	no. of obsd/106 sequences: source
G	F	L	Y	94: including *S. cerevisiae* iso-1[b-g]
D	F	L	Y	1: *Schwanniomyces occidentalis*[h]
G	**Y**	L	Y	1: *Caenorhabdtitis elegans*[i]
G	F	**I**	Y	3: *Euglena gracilis*,[b,j] *Euglena viridis*,[k] *Kluyveromyces lactis*[l]
G	F	**V**	Y	2: *Crithidia fasciculata*,[b,m] *Schizosaccharomyces pombe*[b,n]
G	F	L	**F**	4: *Drosophila melongaster*,[b,o] *Enteromorpha intestinalis*,[b,p] *Arabidopsis thaliana*,[q] *Stellaria longipes*[r]
G	F	**I**	**F**	1: *Neurospora crassa*[b,s]

[a] Sequences are from the sources listed in footnotes *b*–*s*. [b] Moore and Pettigrew (1990). [c] Narita and Titani (1969), Yaoi (1967), and Lederer et al. (1972). [d] GenBank Accession Number M83141. [e] Clark–Walker (1991). [f] Chin et al. (1989). [g] Barber et al. (1993). [h] Amegadzie et al. (1990). [i] Vanfleteren et al. (1990). [j] Pettigrew (1973) and Lin et al. (1973). [k] Ambler et al. (1991). [l] Picos et al. (1993). [m] Hill and Pettigrew (1975). [n] Simon–Becam et al. (1978). [o] Limbach and Wu (1985). [p] Meatyard and Boulter (1974). [q] Kemmerere et al. (1991). [r] Zhang and Chinnappa (1994). [s] Heller and Smith (1966) and Lederer and Simon (1974).

major interactions. In 94 of the 106 known wild-type cytochrome *c* sequences (i.e., the 106 known extant sequences), Gly-6 and Leu-94 form a peg-in-hole interaction and Phe-10 and Tyr-97 form a weakly polar interaction. The remaining 12 sequences (Table 1) represent only six combinations. In five of the six, only one residue is changed and with one exception (aspartate at position 6), the changes are conservative. Comparison of the structures to the known extant sequences suggests that the structure at the interface is invariant (Mathews, 1985). The substitutions occur across the taxonomic hierarchy (i.e., from fungi, to plants, to insects), suggesting that the changes are not adaptive. The four interface residues are inaccessible to solvent in the native state, except for the hydroxyl of Tyr-97. Although the function of cytochrome *c* is to transfer electrons, the interface is neither a physiological electron-transfer path nor a part of the interaction domain between cytochrome *c* and its redox partners (Pelletier & Kraut, 1992). Finally, the interaction between the helices occurs early in folding under certain conditions (Sosnick et al., 1994; Elöve et al., 1994), suggesting that the interaction of the helices directs folding. This observation is supported by the fact that the helices interact in solution if the thioether bonds between heme and the end of the N-terminal helix are intact (Wu et al., 1993). With these facts in mind, we used mutagenesis at the interface to probe relationships between stability, side-chain model data, growth rates, and evolution.

Random mutagenesis was used to search sequence space at positions 6 and 10 (Auld & Pielak, 1991), at positions 94 and 97 (Fredericks & Pielak, 1993), and at all four positions (J. R. Beasley, unpublished observations). Contrary to what is expected by examination of the evolutionary record, at least 78 different amino acid combinations, including all combinations found in known extant sequences, are compatible with function. In an attempt to reconcile this large number of functional mutants with the small number of interface-residue combinations among known extant sequences, we examine how interface residues affect thermal denaturation. We then relate the changes to side-chain model data, growth rates, and theories of protein evolution.

Mutationally induced changes in the free energy of denaturation, ΔG_D,[1] are often compared to changes in the free energy for transferring the corresponding side-chain model compounds from water to a nonpolar solvent. That is, transferring a side chain from the aqueous environment imposed by the denatured state to the inside of a native protein is thought to be equivalent to transferring a side-chain model compound from water to a nonpolar solvent. Pace (1992) reports a linear correlation between accessibility-corrected $\Delta\Delta G_D$ ($\Delta G_{D,var} - \Delta G_{D,wt}$) values and the corresponding changes in octanol transfer free energies, $\Delta\Delta G_{oct\rightarrow H_2O}$ ($\Delta G_{oct\rightarrow H_2O,var} - \Delta G_{oct\rightarrow H_2O,wt}$), after the latter are corrected for volume differences between octanol and the side-chain models (Sharp et al., 1991). Because the size correction is probably not warranted (Holtzer, 1992), we revisit these correlations using uncorrected $\Delta G_{oct\rightarrow H_2O}$ values as well as cyclohexane transfer free energies ($\Delta G_{chx\rightarrow H_2O}$).

Next, we examine relationships between stability and the Dayhoff mutation matrix (Dayhoff et al., 1978). This log-probability matrix was constructed from exhaustive matching of the entire protein sequence database (Gonnet et al., 1992) and is used to assess homology (George et al., 1990). The diagonal elements of the 20-by-20 matrix are related to the probabilities that a residue will not change and the off-diagonal elements are related to the probabilities for the 380 substitutions. We have shown that the matrix predicts the phenotype of interface variants (Fredericks & Pielak, 1993).

Finally, we test the neutral theory of macromolecular evolution, which states that sequence changes in orthologous proteins are caused by fixation of selectively neutral (or nearly neutral) mutations (Kimura, 1983). To test this idea, thermodynamic parameters for variants with known extant sequence combinations were used to construct neutral corridors (Malcolm et al., 1990). Data for all the variants are then examined relative to these corridors. If these parameters are important in selecting the residues in extant sequences, the corridors will demarcate extant and nonextant combinations.

Denaturation experiments were performed under conditions that do not correspond to those found in mitochondria (e.g., ionic strength, pH, specific ions, macromolecular crowding). Nevertheless, the correlations we describe are provocative and suggest experiments to quantify the contribution of in vitro stability to growth rate and phenotype.

[1] Abbreviations: C_{SMe}, *S*-methylcysteine; HS compensation, enthalpy–entropy compensation; P_r, percent probability that a linear correlation coefficient arises from uncorrelated data; r, linear correlation coefficient; T_m, temperature at which half the protein molecules are completely denatured; $T_{m,var}$, T_m of a variant protein; $T_{m,wt}$, T_m of the wild-type protein; ΔC_p, heat capacity of the denatured state minus that of the native state; ΔG_D, Gibbs free energy of the denatured state minus that of the native state; $\Delta G_{D,var}$, ΔG_D for a variant protein; $\Delta G_{D,wt}$, ΔG_D for a wild-type protein; $\Delta G_{chx\rightarrow H_2O}$, free energy change for transfer of a solute from cyclohexane to water; $\Delta G_{oct\rightarrow H_2O}$, free energy change for transfer of a solute from octanol to water; $\Delta\Delta G_{chx\rightarrow H_2O}$, $\Delta G_{chx\rightarrow H_2O,var} - \Delta G_{chx\rightarrow H_2O,wt}$; $\Delta\Delta G_D$, $\Delta G_{D,var} - \Delta G_{D,wt}$; $\Delta\Delta G_{oct\rightarrow H_2O}$, $\Delta G_{oct\rightarrow H_2O,var} - \Delta G_{oct\rightarrow H_2O,wt}$; ΔH_m, van't Hoff (i.e., apparent) enthalpy of denaturation at T_m; ΔH_D, enthalpy of denaturation at temperature T; ΔS_m, $\Delta H_m/T_m$; ΔS_D, entropy of denaturation at temperature T.

MATERIALS AND METHODS

Variants are denoted using the one-letter code with the wild-type residue given first, followed by the position number, and the new residue. For variants with two substitutions, the changes are separated by a semicolon (e.g., L94A;Y97F). The C102T variant is referred to as the wild-type protein, and all variants also contain the C102T mutation. This mutation makes the protein more amenable to biophysical studies but does not change its structure or function (Cutler et al., 1987; Gao et al., 1991; Berghuis & Brayer, 1992). The mutants were made as described by Auld and Pielak (1991) and Fredericks and Pielak (1993). Proteins were purified as described by Willie et al. (1993). Even though the G6D mutant is functional (Auld & Pielak, 1991), it has not been possible to purify enough of this variant to facilitate characterization.

Circular dichroic data were obtained using an Aviv 62DS spectropolarimeter equipped with a thermostated, five-position sample changer. Denaturation was followed at 399 nm by spectrophotometry and at 222 and 416 nm by spectropolarimetry. These and other wavelengths yielded identical values for T_m, the temperature at which half the protein molecules are completely denatured, and ΔH_m, the van't Hoff (i.e., apparent) enthalpy of denaturation at T_m. Denaturation is deemed reversible if two successive thermal denaturations of the same sample yield ΔH_m and T_m values which differ by less than ± 3.9 kcal mol^{-1} and ± 1.1 K, respectively. Thermodynamic parameters and their uncertainties for the wild-type protein and 13 interface variants were obtained between pH 3 and 6 as described by Cohen and Pielak (1994). Values of ΔC_p were determined by linear least-squares fitting of ΔH_m and T_m values to the equation:

$$\Delta H_m = \Delta C_p T_m + b \quad (1)$$

The variants were examined at the following number of pH values: the Y97F variant at 11, the Y97A and L94A;Y97F variants at 10, the L94A variant at 9, the wild-type protein and the F10I, F10Y, and L94I;Y97F variants at 8, the F10W variant at 7, the L94V, F10M, and L94T variants at 6, the L94I variant at 5, and the F10C_{SMe} (*S*-methylcysteine) variant at 2.

The enthalpy of denaturation at temperature T (ΔH_D) was calculated using the equation:

$$\Delta H_D = \Delta H_m + \Delta C_p (T - T_m) \quad (2)$$

The entropy of denaturation at temperature T (ΔS_D) was calculated using the equation

$$\Delta S_D = \Delta S_m + \Delta C_p \ln\left(\frac{T}{T_m}\right) \quad (3)$$

where ΔS_m equals $\Delta H_m/T_m$. Free energies of denaturation at temperature T were calculated using the equation (Elwell & Schellman, 1977)

$$\Delta G_D = \Delta H_m\left(1 - \frac{T}{T_m}\right) - \Delta C_p\left[(T_m - T) + T\ln\left(\frac{T}{T_m}\right)\right] \quad (4)$$

To compare variants to side-chain model data, we used the $\Delta G_{chx \rightarrow H_2O}$ values of Radzicka and Wolfenden (1988) and Saunders et al. (1993) and the $\Delta G_{oct \rightarrow H_2O}$ values of Fauchére and Pliška (1983). The $\Delta G_{chx \rightarrow H_2O}$ and $\Delta G_{oct \rightarrow H_2O}$ values are for pH 7, 300 K. The uncertainties in $\Delta G_{chx \rightarrow H_2O}$ and $\Delta G_{oct \rightarrow H_2O}$ are $\leq \pm 0.1$ kcal mol^{-1}.

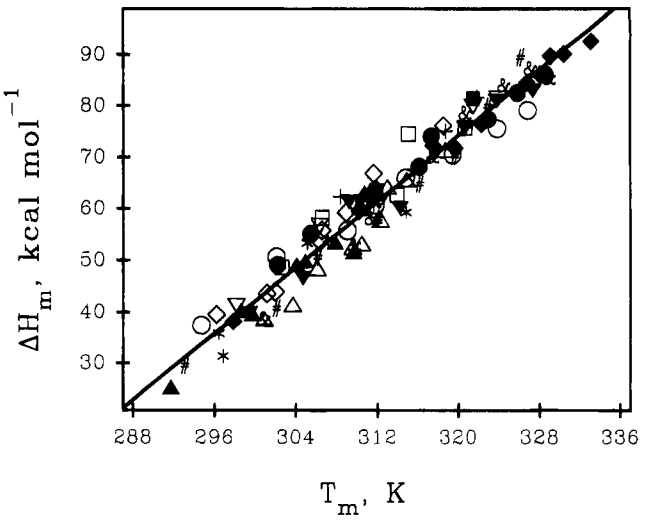

FIGURE 1: ΔH_m versus T_m (♦, Y97F; ●, wild type; ○, F10Y; ▽, F10I; ▼, F10W; ■, F10C_{SMe}; □, F10M; ◊, L94A;Y97F; ▲, Y97A; △, L94A; *, L94T; +, L94V; &, L94I; #, L94I;Y97F). Data for the wild-type protein are from Cohen and Pielak (1994).

To compare data for the interface variants to data for other proteins, we used the $\Delta\Delta G_D$ values (with and without averaging) compiled by Pace (1992). This collection comprises data from 87 different variants representing 11 different types of substitutions in staphylococcal nuclease, gene 5 protein from bacteriophage f1, barnase, and T4 lysozyme. Specifically, the data comprise 9 Ile→Val, 6 Ile→Gly, 9 Ile→Ala, 11 Leu→Gly, 17 Leu→Ala, 9 Val→Gly, 11 Val→Ala, 4 Met→Gly, 4 Met→Ala, 3 Phe→Gly, and 4 Phe→Ala substitutions. Where required, Pace corrected the data to 100% burial. These data were not obtained at a single temperature or pH, but we assume that $\Delta\Delta G_D$ is independent of small differences in conditions.

Values of ΔG_D for the interface variants are uncorrected for solvent exposure. Solvent-accessible surface areas for iso-1-cytochrome *c* were calculated as described by Marmorino et al. (1993) using the coordinates for the oxidized C102T variant (PDB2YCC.ENT; Berghuis & Brayer, 1992), a 1.4-Å-radius probe, and the program GEPOL (Silla et al., 1991). Leu-94 and Phe-10 are solvent inaccessible. Tyr-97 exposes only 12% of tyrosine's maximum accessible surface area (Miller et al., 1987): 20.9 Å^2 from the phenolic oxygen, 5.9 Å^2 from CE1, and 0.4 Å^2 from CD1.

The strength of a linear correlation is defined by r, the linear correlation coefficient. The significance of the correlation is defined by P_r, the percent probability that r arises from uncorrelated data (Taylor, 1982).

RESULTS

Equilibrium Thermodynamics of Protein Thermal Denaturation. Denaturation is reversible for the 14 ferricytochromes *c* at all wavelengths monitored and pH values examined. To determine ΔH_D, we made the common and well-founded assumption that ΔH_D is temperature but not pH dependent (Pfeil & Privalov, 1976). A plot of ΔH_m versus T_m for all the proteins at several different pH values is shown in Figure 1. All the points appear to lie on a single line. The best-fit line for the data in Figure 1 has a slope (ΔC_p) of 1.60 ± 0.03 kcal mol^{-1} K^{-1}, an intercept of −438

Table 2: Growth Rate, T_m, ΔH_m, ΔC_p, and ΔG_D Values for the Ferri Form of the Interface Variants

protein	relative growth rate[b]	T_m, pH 4.6 (±1.1), K	ΔH_m, pH 4.6 (±3.9), kcal mol^{-1}	ΔC_p, kcal mol^{-1} K^{-1}	$\Delta G_{D,325.8K}$ pH 4.6, kcal mol^{-1}
L94I	3	327.1	87.2	1.81 ± 0.11	0.33 ± 0.28
Y97F	3	326.8	84.5	1.53 ± 0.05	0.24 ± 0.28
wild type	3	325.8	82.5	1.37 ± 0.06	0.00 ± 0.28
F10Y	3	323.9	75.5	1.26 ± 0.05	−0.47 ± 0.27
L94I;Y97F	2	322.9	79.9	1.76 ± 0.10	−0.76 ± 0.29
L94V	3	321.8	81.6	1.54 ± 0.20	−1.06 ± 0.31
F10I	3	321.5	79.9	1.57 ± 0.10	−1.14 ± 0.31
F10W	3	320.6	75.6	1.60 ± 0.13	−1.30 ± 0.30
L94T	2	311.9	57.8	1.49 ± 0.24	−3.05 ± 0.34
F10C$_{SMe}$[a]	na[c]	311.9	63.6		−3.35 ● 0.36
F10M	2	310.6	60.1	1.50 ± 0.28	−3.52 ± 0.37
L94A	1	309.6	52.6	1.94 ± 0.13	−3.58 ± 0.37
L94A;Y97F	3	306.4	53.7	1.78 ± 0.11	−4.50 ± 0.42
Y97A	1	304.1	49.2	1.77 ± 0.15	−4.87 ± 0.45

[a] Cysteine at position 10 was modified to form *S*-methylcysteine. Data are from Auld et al. (1993). [b] Data are from Auld and Pielak (1991) and Fredericks and Pielak (1993). [c] Not applicable.

± 3 kcal mol^{-1}, an r of 0.98, and because reduced χ^2 is 0.72 [degrees of freedom, 102; standard deviation of ΔH_m, 3.9 kcal mol^{-1} (Cohen & Pielak, 1994)], the data are well described by a line. Variant-specific values of ΔC_p (Table 2) were determined for all proteins except F10C$_{SMe}$ because only a small amount of this variant was available. For the F10C$_{SMe}$ variant, we used the ΔC_p from fitting all the data with the uncertainty in ΔC_p defined by the average of the uncertainties in Table 2.

Stability is defined as ΔG_D. Because ΔH_D is assumed to be pH independent, the pH dependence of ΔG_D is entirely entropic. Values of ΔG_D and ΔS_D for the interface variants are reported at pH 4.6, the pK_a of the acetate buffer. To minimize the need for extrapolation, ΔG_D, ΔH_D, and ΔS_D values for the interface variants are reported at 325.8 K, the T_m of the wild-type protein at pH 4.6. Under these conditions ΔG_D equals $\Delta\Delta G_D$. Comparisons at other temperatures and pH values do not affect the conclusions.

Values of ΔG_D were determined using eq 4, the T_m and ΔH_m values in Table 2, and, except for the F10C$_{SMe}$ variant (vide supra), the variant-specific ΔC_p values. Using the ΔC_p from fitting all the data changes ΔG_D by ≤0.15 kcal mol^{-1}, which is approximately 2-fold less than the uncertainty in ΔG_D.

Note that ΔG_D values in Table 2 are closely approximated by the equation (Becktel & Schellman, 1987)

$$\Delta\Delta G_D = \frac{\Delta H_{m,wt}}{T_{m,wt}}(T_{m,var} - T_{m,wt}) \quad (5)$$

This simple equation works because the substitutions have small effects on ΔH_D and T_m.

Comparing Protein Data to Side-Chain Model Data. As a first step to determining if the interface responds to amino acid substitutions in the same way as other proteins, we compared the averaged $\Delta\Delta G_D$ values compiled by Pace (1992) to $\Delta\Delta G_{chx\rightarrow H_2O}$, $\Delta\Delta G_{oct\rightarrow H_2O}$, and the off-diagonal elements of the Dayhoff mutation matrix (Gonnet et al., 1992). To compare the mutation matrix to $\Delta\Delta G_D$ values in such a way that the wild-type proteins are at the origin, a score describing each missense mutation is calculated by subtracting the diagonal element for the wild-type residue from the off-diagonal element for the substitution. For proteins with multiple changes, individual scores are added.

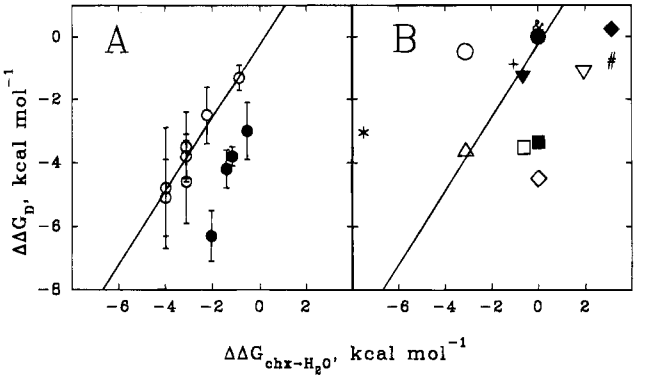

FIGURE 2: (Panel A) Plot of averaged $\Delta\Delta G_D$ values (with their standard deviations) for a collection of non-cytochrome *c* variants (Pace, 1992) versus $\Delta\Delta G_{chx\rightarrow H_2O}$ (Radzicka & Wolfenden, 1988). The open circles, from left to right, represent Ile→Gly/Leu→Gly, Val→Gly/Ile→Ala/Leu→Ala, Val→Ala, and Ile→Val. The closed circles, from left to right, represent Phe→Gly, Met→Gly, Phe→Ala, and Met→Ala. A weighted best-fit line for the hydrocarbon → hydrocarbon side-chain substitutions is shown. (Panel B) Plot of $\Delta\Delta G_D$ for the interface variants versus $\Delta\Delta G_{chx\rightarrow H_2O}$. The symbols are defined in the legend to Figure 1. The line is from panel A.

To compare our analysis to that of Pace (1992), we used the averaged values and their standard deviations for the 11 different types of amino acid substitutions. Plots for the non-cytochrome *c* variants are shown in panels A of Figures 2−4. The parameters from weighted least-squares linear regression (Lyons, 1991) against $\Delta\Delta G_{chx\rightarrow H_2O}$, $\Delta\Delta G_{oct\rightarrow H_2O}$, and the mutation matrix are shown in Table 3. These values are used to produce the lines in Figures 2−4. We also fit the data without averaging, using a uniform standard deviation of ±0.1 kcal mol^{-1} (Shortle et al., 1990). Except for the mutation matrix, this analysis does not significantly change the slopes, intercepts, or P_r values (Table 3). Without averaging, however, r uniformly decreases. Parameters for correlations involving the cytochrome *c* variants are shown in Table 3.

Growth Rates. The growth rates of yeast harboring the mutant genes as their only source of cytochrome *c* on the nonfermentable carbon sources glycerol and lactic acid at 30 °C are shown in Table 2. These rates are not artifacts of recessive mutations introduced during transformations because the rates were determined twice: on the original transformants and after confirming the identity of the

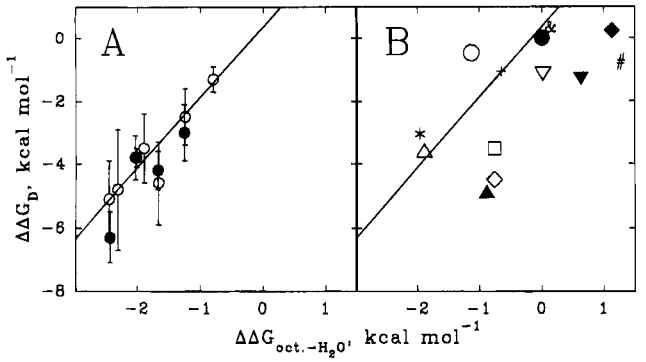

FIGURE 3: (Panel A) Plot of averaged $\Delta\Delta G_D$ values (and their standard deviations) for a collection of non-cytochrome *c* variants (Pace, 1992) versus $\Delta\Delta G_{oct\rightarrow H_2O}$ (Fauchére & Pliška, 1983). The open circles, from left to right, represent Ile→Gly, Leu→Gly, Ile→Gly, Ile→Ala, Leu→Ala, Val→Gly, and Val→Ala. The closed circles, from left to right, represent Phe→Gly, Phe→Ala, Met→Gly, and Met→Ala. A weighted best-fit line for all the data is shown. (Panel B) Plot of $\Delta\Delta G_D$ for the interface variants versus $\Delta\Delta G_{oct\rightarrow H_2O}$. The symbols are defined in the legend to Figure 1. The line is from panel A. The F10C$_{SMe}$ variant is not presented because its $\Delta G_{oct\rightarrow H_2O}$ value is unknown.

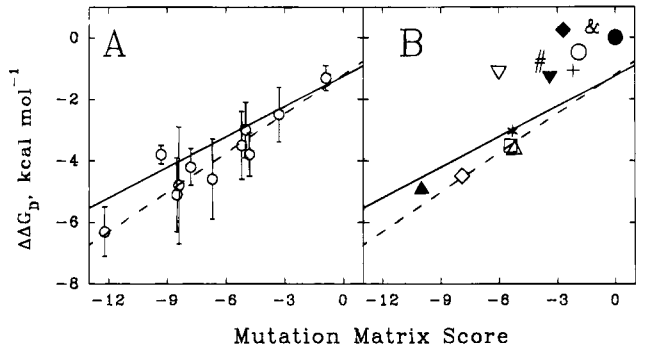

FIGURE 4: (Panel A) Plot of averaged $\Delta\Delta G_D$ values (and their standard deviations) for a collection of non-cytochrome *c* variants (Pace, 1992) versus the mutation matrix score. The points, from left to right, represent Phe→Gly, Phe→Ala, Ile→Gly, Leu→Gly, Met→Gly, Val→Gly, Leu→Ala, Met→Ala, Ile→Ala, Val→Ala, and Ile→Val. The solid line is a weighted fit to the averaged data. The dashed line is an unweighted fit to the averaged data. (Panel B) Plot of $\Delta\Delta G_D$ versus the mutation matrix score for the interface variants. The symbols are defined in the legend to Figure 1. The lines are from panel A. The F10C$_{SMe}$ variant is not presented because *S*-methylcysteine is not a ribosomally encoded amino acid.

mutations followed by retransformation (Auld & Pielak, 1991; Fredericks & Pielak, 1993). Also, the rates are not artifacts of overexpression because the natural promoter is used and the alleles are present on a low-copy-number vector (Sikorski & Hieter, 1989).

DISCUSSION

Denaturation of the Oxidized Proteins Is a Two-State Process. Identical values of ΔH_m and T_m for the interface variants were obtained at two different wavelengths using spectrophotometry and circular dichroism spectropolarimetry. Furthermore, denaturation of the wild-type protein and the F10M, F10C$_{SMe}$, and L94A variants was examined at pH 4.6 using calorimetry (Betz & Pielak, 1992; Auld et al., 1993; Fredericks, 1993). The ratio of ΔH_m to the calorimetric enthalpy is near unity. These observations strongly suggest that denaturation is a two-state process (Sturtevant, 1987).

Changes in ΔH_D, ΔS_D, and ΔG_D. Plots such as those in Figure 1 are generally used for single proteins where ΔH_D and ΔC_p are expected to have a single value (Shortle et al., 1988; Cohen & Pielak, 1994). Examination of Figure 1 shows that data for *different* variants fall on the same line, suggesting that the amino acid substitutions do not drastically alter ΔC_p or ΔH_D over the temperature range examined. A similar conclusion is reported for variants of Arc repressor (Milla et al., 1994). Although variant-specific values of ΔC_p have been determined (Table 2), using these values in combination with the variant-specific values of ΔH_m, T_m, and eq 2 leads to small differences in ΔH_D over the temperature range examined. Given this observation and the differences in ΔG_D, it is tempting to conclude that substitutions affect only ΔS_D. However, the range of ΔG_D values is approximately equal to the uncertainty in ΔH_D (Table 2). This observation does not support the conclusion that changes in ΔG_D are determined solely by changes in ΔS_D. This opens the possibility of enthalpy−entropy (HS) compensation.

HS compensation is an extrathermodynamic relationship in which a small change in ΔG is the result of large and nearly compensatory changes in ΔH and ΔS (Lumry & Rajender, 1970). Obtaining evidence for HS compensation from van't Hoff data is difficult because, for small changes in ΔH and ΔS, correlation of errors leads to a linear relationship in the absence of the extrathermodynamic effect (Krug et al., 1976). Shortle et al. (1988) report a compelling example of HS compensation for a set of staphylococcal nuclease variants which display large changes in ΔH_D and ΔS_D.

A plot of ΔH_D versus $T\Delta S_D$ for the interface variants is shown in Figure 5A, where the diagonal corresponds to perfect compensation (i.e., no change in ΔG_D). The uncertainties in ΔH_D and $T\Delta S_D$ take the form of areas defined by the combinations of uncertainties in ΔH_D, $T\Delta S_D$ (both are $\approx$3.9 kcal mol^{-1}), and ΔG_D ($\approx$0.3 kcal mol^{-1}). To obtain a picture of the uncertainty, separate areas must be drawn for every point. Examination of such a plot, Figure 5B, shows that the first eight variants in Table 2 are clustered in one group and the remainder are clustered in another. The uncertainties in ΔH_D and $T\Delta S_D$ overlap when projected onto the *y*- and *x*-axes, respectively. Thus, there is no compelling evidence for HS compensation. Note, however, that the proteins in the top cluster represent the known extant sequences, revealing a correlation between ΔG_D and the residues found in naturally occurring cytochromes *c* (vide infra).

The Helices Interact Thermodynamically. Examination of crystallographic (Mathews, 1985) and NMR (Auld et al., 1993; Gao et al., 1990) data shows that the N- and C-terminal helices interact in a specific manner. The additivity of ΔG_D values can also be used to assess interactions (Wells, 1990). The L94I;Y97F variant is 1.3 kcal mol^{-1} less stable than predicted from the sum of individual ΔG_D values for the L94I and Y97F variants. The L94A;Y97F variant is 1.2 kcal mol^{-1} less stable than predicted. The discrepancies are consistent with the idea that proteins are stabilized by cooperative interactions (Creighton, 1983) and indicate that the helices interact thermodynamically as well as structurally. Additionally, the stability changes are not correlated with helix propensity. For instance, alanine at position 94 is expected to be stabilizing and isoleucine destabilizing if helix stability is important (O'Neil & DeGrado, 1990), but the

Table 3: Linear Correlations between $\Delta\Delta G_D$ and Side-Chain Model Data

data set	n^a	slope	Y-intercept, kcal mol^{-1}	r^b	P_r^c
non-cytochrome *c* variants					
$\Delta\Delta G_{chx\rightarrow H_2O}$ with averaging					
all substitutions	11	0.7 ± 0.2	−2.3 ± 0.4	0.59	≈5
hydrocarbons	7	1.2 ± 0.3	−0.2 ± 0.6	0.98	<0.1
nonhydrocarbons	4	2.3 ± 0.8	−1 ± 1	0.95	≈5
$\Delta\Delta G_{chx\rightarrow H_2O}$ without averaging					
all substitutions	87	0.72 ± 0.01	−1.80 ± 0.03	0.42	<0.05
hydrocarbons	72	1.2 ± 0.01	−0.025 ± 0.04	0.60	<0.05
nonhydrocarbons	15	2.2 ± 0.05	−1.51 ± 0.07	0.79	<0.05
$\Delta\Delta G_{oct\rightarrow H_2O}$					
with averaging	11	2.2 ± 0.3	0.4 ± 0.6	0.89	<0.05
without averaging	87	2.18 ± 0.02	0.17 ± 0.04	0.59	<0.05
mutation matrix					
with averaging	11	0.33 ± 0.05	−1.2 ± 0.4	0.86	<0.1
without averaging	87	0.427 ± 0.004	−1.19 ± 0.03	0.62	<0.05
cytochrome *c* variants					
$\Delta\Delta G_{chx\rightarrow H_2O}$	14	0.18 ± 0.03	−1.39 ± 0.09	0.23	≈44
$\Delta\Delta G_{oct\rightarrow H_2O}$	13[d]	0.95 ± 0.09	−1.13 ± 0.09	0.63	≈2
mutation matrix	13[e]	0.56 ± 0.04	0.7 ± 0.2	0.89	<0.05

[a] Number of points. [b] Linear correlation coefficient. [c] Percent probability that *r* arises from uncorrelated data. [d] There is no $\Delta\Delta G_{oct\rightarrow H_2O}$ value for *S*-methylcysteine. [e] There is no mutation matrix value for *S*-methylcysteine.

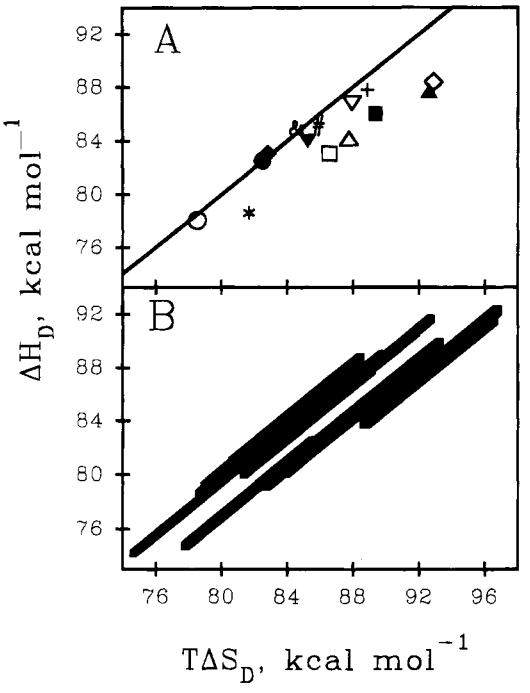

FIGURE 5: (Panel A) Plot of ΔH_D (325.8 K) versus $T\Delta S_D$ (325.8 K, pH 4.6). The solid line has a slope of unity and corresponds to perfect enthalpy−entropy compensation for the wild-type protein. The symbols are defined in the legend to Figure 1. (Panel B) Plot of ΔH_D (325.8 K) versus $T\Delta S_D$ (325.8 K, pH 4.6) with the associated uncertainties (see text).

opposite is observed. Several helix-propensity scales have been proposed [see Munoz and Serrano (1994)], and although values from proteins are more context dependent than those from peptides (Blaber et al., 1993), good correlation has usually been observed. In summary, the N- and C-terminal helices interact both structurally and thermodynamically, and the interaction free energy is more important that the individual helix propensities.

$\Delta\Delta G_D$, $\Delta\Delta G_{chx\rightarrow H_2O}$, and $\Delta\Delta G_{oct\rightarrow H_2O}$ for the Non-Cytochrome c Variants. The linear correlation between $\Delta\Delta G_{chx\rightarrow H_2O}$ and the averaged $\Delta\Delta G_D$ values is weak but significant. Fitting the averaged values for the hydrocarbon → hydrocarbon substitutions separately increases *r* and decreases P_r (Figure 2A). Using the nonaveraged data does not significantly change the slope or intercept for either comparison, but does increase P_r and decrease *r*.

The slope of the hydrocarbon fit is near unity, indicating that hydrocarbon side chains respond to burial like transfer from water to cyclohexane (Rose & Wolfenden, 1993). Substitutions involving changes from methionine and phenylalanine to alanine and glycine, however, describe a line with greater slope, suggesting that side chains with even weakly polar atoms are not in cyclohexane-like environments (Radzicka & Wolfenden, 1988). This suggestion is consistent with the idea that such side chains are often involved in weakly polar interactions (Burley & Petsko, 1988). The dichotomy between hydrocarbon and weakly polar residues is absent from the plot (Figure 3A) of $\Delta\Delta G_{oct\rightarrow H_2O}$ versus $\Delta\Delta G_D$ (independent of averaging; Table 3), suggesting that octanol [or the 2.3 M H_2O in saturated octanol (Wolfenden & Radzicka (1994)] participates in weakly polar interactions (Radzicka & Wolfenden, 1988). Although the fit is strong and significant, the slope is not unity, indicating that side-chain transfer from water to octanol does not adequately describe burial.

In summary, $\Delta\Delta G_{chx\rightarrow H_2O}$ values for changing larger hydrocarbon side chains to smaller ones are good predictors of protein stability, and cyclohexane is a good model for the average environment of hydrocarbon side chains in folded proteins. Cyclohexane, however, is not a good model for the protein environment of even weakly polar residues. On the other hand, $\Delta\Delta G_{oct\rightarrow H_2O}$ values are a good and more general predictor of stability for a broader range of substitutions, but the attenuation of changes in ΔG_D shows that octanol is a poor model for protein interiors.

$\Delta\Delta G_D$, $\Delta\Delta G_{chx\rightarrow H_2O}$, and $\Delta\Delta G_{oct\rightarrow H_2O}$ for the Interface Variants. There is no significant linear correlation between $\Delta\Delta G_{chx\rightarrow H_2O}$ and $\Delta\Delta G_D$ (Figure 2B; Table 3). The relationship between $\Delta\Delta G_{oct\rightarrow H_2O}$ and $\Delta\Delta G_D$ (Figure 3B; Table 3) is stronger and more significant, but the slope is 2.3 times the slope for the non-cytochrome *c* data.

There are at least three reasons for the weaker correlations between $\Delta\Delta G_D$ and $\Delta\Delta G_{chx\rightarrow H_2O}$ for the cytochrome *c* variants compared to the non-cytochrome *c* variants. First, the non-cytochrome *c* data comprise averaged $\Delta\Delta G_D$ values for 11 different types of substitutions at many different sites in five different proteins. This is contrasted by the interface variants which are from a single protein and, although they comprise 13 different types of substitutions, occur at only three different positions. Thus, the cytochrome *c* data are not as well averaged. In an attempt to address this problem, we fit the non-cytochrome *c* data without averaging the $\Delta\Delta G_D$ values. As expected, *r* decreases (Table 3). Even without averaging, the correlation between $\Delta\Delta G_{chx\rightarrow H_2O}$ and $\Delta\Delta G_D$ for non-cytochrome *c* variants is stronger than the correlation between $\Delta\Delta G_{chx\rightarrow H_2O}$ and $\Delta\Delta G_D$ for the cytochrome *c* variants (Table 3). Therefore, averaging is not the only reason for the weak fit. Second, there are no examples of changes in hydrogen bonding in the non-cytochrome *c* data set, but both addition and deletion of potential hydrogen-bonding groups occur among the cytochrome *c* variants. Note that variants which gain a potential hydrogen bond donor (L94T, F10Y) are more stable than expected from examination of $\Delta\Delta G_{chx\rightarrow H_2O}$ and $\Delta\Delta G_{oct\rightarrow H_2O}$ values for hydrocarbon → hydrocarbon swaps, and variants that lose a potential hydrogen-bonding group (Y97F, L94A;Y97F, L94I;Y97F, Y97A) are less stable than expected. This observation suggests that proteins exploit hydrogen-bonding potential that is not reflected in $\Delta G_{chx\rightarrow H_2O}$ or $\Delta G_{oct\rightarrow H_2O}$. This is consistent with the observation that proteins tend to maximize their hydrogen-bonding potential (Chothia, 1975; Richards, 1977; Stickle et al., 1992; McDonald & Thornton, 1994). Third, positions 10, 94, and 97 interact both structurally and thermodynamically (vide supra).

Despite these observations, the points for the hydrocarbon → hydrocarbon substitutions in Figures 2B and 3B (L94I, L94V, and L94A) are close to the lines defined by the non-cytochrome *c* data in Figures 2A and 3A. In summary, the weak correlation between the stability of the interface variants and $\Delta G_{chx\rightarrow H_2O}$ and $\Delta G_{oct\rightarrow H_2O}$ suggests that, in addition to polarity, the hydrogen-bonding potential and side-chain−side-chain interactions are important in stabilizing the interface between the N- and C-terminal helices. These additional properties are not well modeled by transfer free energies. Rose et al. (1985) reach a similar conclusion from examination of their empirical hydrophobicity scale.

$\Delta\Delta G_D$ and the Mutation Matrix. There is strong and significant correlation between averaged $\Delta\Delta G_D$ values for the non-cytochrome *c* variants and the mutation matrix scores which describe the changes (Figure 4A; Table 3). Without averaging, the slope increases and *r* decreases. The slope increases because the weighted fit is strongly influenced by the small standard deviation for the Phe→Ala substitutions (Figure 4A). This small standard deviation may be an artifact because there are only four variants in this set. For this reason both weighted and unweighted best-fit lines are shown in Figure 4.

For the interface variants, comparison of the sequences in Table 1 to the data in Figure 4B reveals that variants with substitutions found in known extant sequences possess both the highest stability and the highest mutation matrix scores. Unlike correlations between $\Delta\Delta G_{oct\rightarrow H_2O}$ or $\Delta\Delta G_{chx\rightarrow H_2O}$ and $\Delta\Delta G_D$, the correlation between the mutation matrix score and $\Delta\Delta G_D$ for the cytochrome *c* variants is stronger than that for the non-cytochrome *c* variants. Furthermore, the slope for the cytochrome *c* variants is similar to that for the non-cytochrome *c* variants. Thus, we have found a strong and significant correlation between in vitro protein stability and the mutation matrix.

To our knowledge this is the first example of correlations between any mutation matrix and protein stability. The matrix is a better predictor than $\Delta\Delta G_{chx\rightarrow H_2O}$ and $\Delta\Delta G_{oct\rightarrow H_2O}$ because both polarity and packing are important for the stability of natural proteins, and the matrix is made from a compilation of sequence data for natural proteins. Thus, the mutation matrix is a more generally applicable predictor of stability than $\Delta\Delta G_{chx\rightarrow H_2O}$ or $\Delta\Delta G_{oct\rightarrow H_2O}$. The fact that the points in Figure 4 for amino acid combinations found in nature lie above the best-fit lines suggests a relationship between known extant sequences and $\Delta\Delta G_D$. This relationship is discussed in the section about neutral corridors.

$\Delta\Delta G_D$, Growth Rates, and Extant Sequences. The most stable protein (L94I) does not possess the most common extant combination (Gly-6, Phe-10, Leu-94, and Tyr-97; Table 1). This observation is in accord with the idea that most wild-type proteins are not at their maximum stability (Zuckerkandl, 1976). Comparison of rates to ΔG_D values (Table 2) shows that the more stable variants possess a higher growth rate. We also looked for a correlation with phenotype (Fredericks & Pielak, 1993). All the mutants, however, give rise to the wild-type phenotype, except for L94A, which gives rise to a temperature-sensitive phenotype.

Stability, however, cannot fully explain growth rate. For instance, L94A;Y97F is one of the least stable variants but exhibits a wild-type growth rate. Furthermore, it is known that only 10% of the wild-type protein level is required to support function in vivo (Sherman et al., 1974) and ≈90% of the L94A variant is folded at 30 °C in vitro, but this mutant gives rise to a slow growth rate and a temperature-sensitive phenotype. This raises interesting questions regarding other factors including folding rate, reduction potential, lethality of the denatured state, and differential stability of the oxidized and reduced forms (Komar-Panicucci et al., 1994). These factors need to be examined to determine the cause of this behavior. Nevertheless, our analysis shows that known extant sequences give rise to the most stable ferricytochromes *c* and the fastest growth rates. These observations indicate that in vitro stability, growth rate, and biological function are related.

Neutral Corridors. In their study of lysozyme, Malcolm et al. (1990) defined a *selectively neutral* corridor with an in vitro system using T_m values for extant sequences. The breadth of the corridor was defined as the region between two standard deviations (i.e., a 95% confidence interval) above the highest T_m, and two standard deviations below the lowest T_m of variants possessing residues from known extant sequences. We adopt the same rationale, but in addition to T_m, we use ΔH_D, ΔS_D, ΔC_p, and ΔG_D.

The enthalpy corridor at 325.8 K runs from 95 kcal mol^{-1} for the L94V variant to 68 kcal mol^{-1} for the F10Y variant (see Figure 5). The ΔH_D values for all the interface variants are within this corridor. The same result is found for ΔS_D and ΔC_p. These observations do not necessarily conflict with neutral theory but suggest that either nature measures ΔH_D, ΔS_D, and ΔC_p more accurately or there is no direct selective pressure on these parameters.

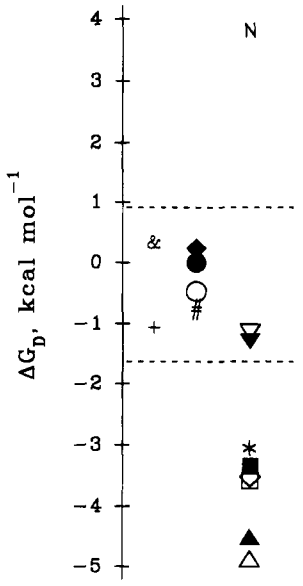

FIGURE 6: Notional free energy corridor. Symbols are defined in the legend to Figure 1 (N, N52I). Values of ΔG_D are from Table 2. The left column comprises the variants which define the neutral corridor (see text). The center column comprises variants with sequence combinations found in other extant cytochromes *c*. The right column comprises variants with sequence combinations that have not been found in known extant cytochromes *c*.

As shown in Figure 6, the free energy corridor for positions 10, 94, and 97 lies between 0.89 kcal mol^{-1} (L94I) and −1.68 kcal mol^{-1} (L94V). Most of the variants with substitutions not found among known extant sequences fall outside (below) the corridor, although two, F10W and F10I, fall within the corridor. The observation that F10W and F10I fall within the corridor does not necessarily conflict with neutral theory but suggests that nature measures ΔG_D more accurately (both are near the lower boundary of the corridor) or that Trp-10 and/or Ile-10 will be found when additional extant sequences become available. Taken together, these observations suggest that extant sequences represent amino acids that are neutral with respect to stability. The same conclusion is reached if T_m is used to define the corridor, but it is important to bear in mind that the simple relationship between T_m and ΔG_D (eq 5) is valid only when changes in ΔH_D and T_m are small [for an exception see Betz and Pielak (1992)]. For the interface variants we have shown this to be true (Figure 1).

Another observation involves the nonhelix−helix interface substitution, Ile-52, which is not found among extant sequences. The N52I variant, although functional and more stable than the wild-type protein (Hickey et al., 1991; D. Doyle, unpublished observation), does not fall within the free energy corridor. In summary, to the degree that the free energy corridor matches the neutral corridor, the data suggest that nature acts upon ΔG_D, keeping it within a range narrower than that defined by being substantially folded under physiological conditions.

ΔG_D and Base Changes. The known extant sequences (Table 1) can be inter converted by single base-pair changes. The F10Y and F10I mutants are one base pair removed from Phe-10 and their stabilities are within the free energy corridor, yet Ile-10 is not observed in known extant sequences. On the other hand, variants with nonextant combinations both possess stabilities outside the corridor and are at least two base pairs from the combinations in the corridor. These observations show that although there need not be a correlation between the genetic code and extant sequences (Gonnet et al., 1992), such correlations are common, especially at low divergence.

CONCLUSIONS

Variants with combinations of interface residues found in wild-type eukaryotic cytochromes *c* are more stable than those with residues that have not been observed in wild-type eukaryotic cytochromes *c*. This may explain why at least 78 different combinations of residues at the helix−helix interface give rise to a functional cytochrome *c*, but only six of these combinations are observed in wild-type eukaryotic cytochromes *c*.

Stability changes cannot solely be allocated to either ΔH_D or ΔS_D, but there is no evidence for enthalpy−entropy compensation.

The helices interact thermodynamically, and tertiary interactions determine their stability.

Amino acid side-chain model $\Delta G_{chx\rightarrow H_2O}$ values are good predictors of ΔG_D when a small hydrocarbon side chain replaces a larger one. Furthermore, cyclohexane is a good model for the environment of buried hydrocarbon side chains in proteins.

Amino acid side-chain model $\Delta G_{oct\rightarrow H_2O}$ values are good predictors of ΔG_D when a small hydrocarbon side chain replaces a larger nonpolar one. However, octanol is not a good model for protein interiors.

Neither $\Delta G_{chx\rightarrow H_2O}$ nor $\Delta G_{oct\rightarrow H_2O}$ values are good predictors or models when substitutions involve changes in potential hydrogen-bonding or side-chain−side-chain interactions.

There is a linear correlation between ΔG_D and the Dayhoff mutation matrix. The matrix is less affected by changes in potential hydrogen bonding and side-chain−side-chain interactions.

The neutral theory of macromolecular evolution accurately separates the known extant sequences from a pool of functional cytochrome *c* variants when ΔG_D is used to define the neutral corridor, but ΔH_D, ΔS_D, and ΔC_p fail to partition the groups.

In general, variants whose genes differ from wild type by a single base are more stable than those with multiple changes.

ACKNOWLEDGMENT

We thank R. Wolfenden for helpful discussions of transfer free energies.

REFERENCES

Ambler, R. P., Kamen, M. D., Bartsch, R. G., & Meyer, T. E. (1991) *Biochem. J. 276*, 47−52.
Amegadzie, B. Y., Zitomer, R. S., & Hollenberg, C. P. (1990) *Yeast 6*, 429−440.
Auld, D. S., & Pielak, G. J. (1991) *Biochemistry 30*, 8684−8690.
Auld, D. S., Young, G. B., Saunders, A. J., Doyle, D. F., Betz, S. F., & Pielak, G. J. (1993) *Protein Sci. 2*, 2187−2197.
Barber, M. J., Trimboli, A. J., Clark, M., Young, C., & Neame, P. J. (1993) *Arch. Biochem. Biophys. 301*, 294−298.
Becktel, W. J., & Schellman, J. A. (1987) *Biopolymers 26*, 1859−1877.
Berghuis, A. M., & Brayer, G. D. (1992) *J. Mol. Biol. 223*, 959−976.
Betz, S. F., & Pielak, G. J. (1992) *Biochemistry 31*, 12337−12344.

Bixler, J., Bakker, G., & McLendon, G. (1992) *J. Am. Chem. Soc. 114*, 6938−6939.
Blaber, M., Zhang, X.-J., & Matthews, B. W. (1993) *Science 260*, 1637−1640.
Burley, S. K., & Petsko, G. A. (1988) *Adv. Protein Chem. 39*, 125−189.
Chin, C. C. Q., Niehaus, W. G., & Wold, F. (1989) *J. Protein Chem. 8*, 165−171.
Chothia, C. (1975) *Nature 254*, 304−308.
Clark-Walker, G. D. (1991) *Curr. Genet. 20*, 195−198.
Cohen, D. S., & Pielak, G. J. (1994) *Protein Sci. 3*, 1253−1260.
Cohen, D. S., & Pielak, G. J. (1995) *J. Am. Chem. Soc.* (in press).
Creighton, T. E. (1983) *Biopolymers 22*, 49−58.
Cutler, R. L., Pielak, G. J., Mauk, A. G., & Smith, M. (1987) *Protein Eng. 1*, 95−99.
Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure* (Dayhoff, M. O., Ed.) Vol. 5, Suppl. 3, pp 345−352, National Biomedical Research Foundation, Washington, DC.
Elöve, G. A., Bhuyan, A. K., & Roder, H. (1994) *Biochemistry 33*, 6925−6935.
Elwell, M. L., & Schellman, J. A. (1977) *Biochim. Biophys. Acta 494*, 367−383.
Fauchére, J.-L., & Pliška, V. (1983) *Eur. J. Med. Chem. 18*, 369−375.
Fredericks, Z. L. (1993) Doctoral Dissertation, University of North Carolina, Chapel Hill.
Fredericks, Z. L., & Pielak, G. J. (1993) *Biochemistry 32*, 929−936.
Gao, Y., Boyd, J., Williams, R. J. P., & Pielak, G. J. (1990) *Biochemistry 29*, 6994−7003.
Gao, Y., Boyd, J., Pielak, G. J., & Williams, R. J. P. (1991) *Biochemistry 30*, 1928−1934.
George, D. G., Barker, W. C., & Hunt, L. T. (1990) *Methods Enzymol. 183*, 333−351.
Gonnet, G. H., Cohen, M. A., & Benner, S. A. (1992) *Science 256*, 1443−1445.
Heller, J., & Smith, E. L. (1966) *J. Biol. Chem. 241*, 3165−3180.
Hickey, D. R., Berghuis, A. M., Lafond, G., Jaeger, J. A., Cardillo, T. S., McLendon, D., Das, G., Sherman, F., Brayer, G. D., & McLendon, G. (1991) *J. Biol. Chem. 266*, 11686−11694.
Hilgen-Willis, S., Bowden, E. F., & Pielak, G. J. (1993) *J. Inorg. Biochem. 51*, 649−653.
Hill, G. C., & Pettigrew, G. W. (1975) *Eur. J. Biochem. 57*, 265−271.
Holtzer, A. (1992) *Biopolymers 32*, 711−715.
Kemmerer, E. C., Lei, M., & Wu, R. (1991) *J. Mol. Evol. 32*, 227−237.
Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge, England.
Komar-Panicucci, S., Weis, D., Bakker, G., Qiao, T., Sherman, F., & McLendon, G. (1994) *Biochemistry 33*, 10556−10560.
Krug, R. R., Hunter, W. G., & Greiger, H. A. (1976) *J. Phys. Chem. 80*, 2335−2351.
Lederer, F., & Simon, A. M. (1974) *Biochem. Biophys. Res. Commun. 56*, 317−323.
Lederer, F., Simon, A. M., & Verdiere, J. (1972) *Biochem. Biophys. Res. Commun. 47*, 55−58.
Liggins, J. R., Sherman, F., Mathews, A. J., & Nall, B. T. (1994) *Biochemistry 33*, 9209−9219.
Limbach, K. J., & Wu, R. (1985) *Nucleic Acids Res. 13*, 631−644.
Lin, D. K., Niece, R. L., & Fitch, W. M. (1973) *Nature 241*, 533−535.
Lumry, R., & Rajender, S. (1970) *Biopolymers 9*, 1125−1227.
Lyons, L. (1991) in *A Practical Guide to Data Analysis for Physical Science Students*, Cambridge University Press, Cambridge, U.K.
Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F., & Wilson, A. C. (1990) *Nature 345*, 86−89.
Marmorino, J. L., Auld, D. S., Betz, S. F., Doyle, D. F., Young, G. B., & Pielak, G. J. (1993) *Protein Sci. 2*, 1966−1974.
Mathews, F. S. (1985) *Prog. Biophys. Mol. Biol. 45*, 1−56.
McDonald, I. K., & Thornton, J. M. (1994) *J. Mol. Biol. 234*, 777−793.
Meatyard, B. T., & Boulter, D. (1974) *Phytochemistry 13*, 2777−2782.
Milla, M. E., Brown, B. M., & Sauer, R. T. (1994) *Struct. Biol. 1*, 518−523.
Miller, S., Janin, J., Lesk, A. M., & Chothia, C. (1987) *J. Mol. Biol. 196*, 641−656.
Moore, G. R., & Pettigrew, G. W. (1990) in *Cytochromes c Evolutionary, Structural and Physiochemical Aspects*, Springer-Verlag, Berlin.
Munoz, V., & Serrano, L. (1994) *Struct. Biol. 1*, 399−409.
Narita, K., & Titani, K. (1969) *J. Biochem. 63*, 226−241.
O'Neil, K. T., & DeGrado, W. F. (1990) *Science 250*, 646−651.
Pace, C. N. (1992) *J. Mol. Biol. 226*, 29−35.
Pelletier, H., & Kraut, J. (1992) *Science 258*, 1748−1755.
Pettigrew, G. W. (1973) *Nature 241*, 531−533.
Pettigrew, G. W., & Moore, G. R. (1987) *Cytochrome c: Biological Aspects*, Springer-Verlag, Berlin.
Pfeil, W., & Privalov, P. L. (1976) *Biophys. Chem. 4*, 41−50.
Picos, M. A. F., Torres, A. M. R., Ramil, E., Cerdan, M. E., Breunig, K. D., Hollenberg, C. P., & Zitomer, R. S. (1993) *Yeast 9*, 201−204.
Radzicka, A., & Wolfenden, R. (1988) *Biochemistry 27*, 1664−1670.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
Rose, G. D., & Wolfenden, R. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 381−415.
Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985) *Science 229*, 834−838.
Saunders, A. J., Young, G. B., & Pielak, G. J. (1993) *Protein Sci. 2*, 1183−1184.
Sharp, K. A., Nicholls, A., Fine, R. F., & Honig, B. (1991) *Science 252*, 106−109.
Sherman, F., Stewart, J. W., Jackson, M., Gilmore, R. A., & Parker, J. H. (1974) *Genetics 77*, 255−284.
Shortle, D., Meeker, A. K., & Freire, E. (1988) *Biochemistry 27*, 4761−4768.
Shortle, D., Stites, W. E., & Meeker, A. K. (1990) *Biochemistry 29*, 8033−8041.
Sikorski, R. S., & Hieter, P. (1989) *Genetics 122*, 19−27.
Silla, E., Tuñón, I., & Pascual-Ahuir, J. L. (1991) *J. Comput. Chem. 12*, 1077−1088.
Simon-Becam, A. M., Claisse, M., & Lederer, F. (1978) *Eur. J. Biochem. 86*, 407−416.
Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L., & Hall, B. D. (1979) *Cell 16*, 753−761.
Sosnick, T. R., Mayne, L., Hiller, R., & Englander, S. W. (1994) *Struct. Biol. 1*, 149−156.
Stickle, D. F., Presta, L. G., Dill, K. A., & Rose, G. D. (1992) *J. Mol. Biol. 226*, 1143−1159.
Sturtevant, J. M. (1987) *Annu. Rev. Phys. Chem. 38*, 468−488.
Taylor, J. R. (1982) *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, University Science Books, Mill Valley, CA.
Vanfleteren, J. R., Evers, E. A. I. M., Van de Werken, G., & Van Beeumen, J. J. (1990) *Biochem. J. 271*, 613−620.
Wang, X., & Pielak, G. J. (1991) *J. Mol. Biol. 221*, 97−105.
Wells, J. A. (1990) *Biochemistry 29*, 8509−8517.
Willie, A., McLean, M., Liu, R.-Q., Hilgen-Willis, S., Saunders, A. J., Pielak, G. J., Sligar, S. G., Durham, B., & Millett, F. (1993) *Biochemistry 32*, 7519−7525.
Wolfenden, R., & Radzicka, A. (1994) *Science 265*, 936−937.
Wu, L. C., Laub, P. B., Elöve, G. A., Carey, J. A., & Roder, H. (1993) *Biochemistry 32*, 10271−10276.
Yaoi, Y. (1967) *J. Biochem. 61*, 54−58.
Zhang, X. H., & Chinnappa, C. C. (1994) *Mol. Biol. Evol. 11*, 365−375.
Zuckerkandl, E. (1976) *J. Mol. Evol. 7*, 269−312.

BI942564J