Exploring the Interface between the N- and C-Terminal Helices of Cytochrome *c* by Random Mutagenesis within the C-Terminal Helix†

Zoey L. Fredericks and Gary J. Pielak*

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290

Received July 13, 1992; Revised Manuscript Received October 21, 1992

ABSTRACT: Buried within cytochrome *c* lies a highly-conserved helix–helix interface formed by the perpendicular packing of the C-terminal helix against the N-terminal helix. This interface involves a peg-in-hole interaction between Gly-6 and Leu-94 and an aromatic–aromatic interaction between Phe-10 and Tyr-97. To gain insight into protein design, we investigated the relationship between the sequence of the interface and the physiological function of yeast iso-1-cytochrome *c*. A library of mutants at positions 94 and 97 of the C-terminal helix was created to examine the effect of novel amino acid combinations. We isolated 45 of the 400 possible amino acid combinations, 32 of which result in a functional cytochrome *c*. Contrary to evolutionary conservation of the peg-in-hole and aromatic–aromatic interactions, we find that side-chain volume and conservation of aromatic residues do not play an essential role in determining function. Additionally, we find negatively-charged residues within the interface that result in a functional cytochrome *c*. Examination of the 45 missense mutants indicates that approximately 120 unique combinations are compatible with function. These results show that the interface is flexible. However, truncation of the C-terminal helix at position 94 abolishes function, suggesting that the interface is essential. The correlation observed between our library of mutants and the mutation matrix compiled by Gonnet et al. [Gonnet, G. H., Cohen, M. A., & Benner, S. A. (1992) *Science 256*, 1443–1445] demonstrates the potential use of the matrix to predict the effect of sequence changes on natural proteins and to optimize the design of novel proteins.

Examination of the protein sequence database indicates that proteins are tolerant of many amino acid substitutions. In three-dimensional structures, substitutions are more often found near the surface than in the interior. These observations underlie the idea that packing of the protein interior is an important determinant of structure and function. Considering this, one vital question in the design of novel proteins concerns the degree to which interior packing can be changed without destroying structures that maintain function. We produced a library of mutants within the buried helix–helix interface of cytochrome *c* to investigate the relationship between sequence and physiological function.

Cytochrome *c*, the penultimate electron transfer protein of the eukaryotic respiratory chain, contains interacting N- and C-terminal helices. The perpendicular pairing of these helices is the most conserved structural motif of both prokaryotic and eukaryotic cytochrome *c* (Mathews, 1985). Figure 1 shows the interface formed by the packing of the N- and C-terminal helices within iso-1-cytochrome *c* from the yeast *Saccharomyces cerevisiae*. This interface involves the conserved residues Leu-94 and Tyr-97 of the C-terminal helix and Gly-6 and Phe-10 of the N-terminal helix (Berghuis & Brayer, 1992). The side chain of Leu-94 packs into the hole created by Gly-6 in a peg-in-hole interaction while Tyr-97 and Phe-10 pack against each other through an aromatic–aromatic interaction. Figure 2 lists the amino acid sequence of these helices in iso-1-cytochrome *c*. Inspection of the 100 known eukaryotic cytochrome *c* sequences reveals the highly-conserved nature of the helix–helix interface (Mathews, 1985).[1] Yet, these helices do not appear to be directly involved in electron transfer or interaction with redox partners. All known eukaryotic cytochromes *c* possess either phenylalanine or tyrosine at positions 10 and 97, and only five of the 400 (20^2) possible amino acid combinations have been observed at positions 6 and 94. Examining the effect of novel amino acids within the interface should elucidate the requirements for helix–helix packing in cytochrome *c*.

Cytochrome *c* is an excellent system for examining the relationship between sequence and function. Using random mutagenesis and a bacterial system, a library of mutants can be produced to study any region of the protein. Introduction of this library into a yeast strain that lacks a functional cytochrome *c* allows functional and nonfunctional mutants to be identified. By coupling genetic selection to mutagenesis, the functional role of evolutionarily invariant and conserved residues can be explored. This approach has been used previously to identify missense mutants in iso-1-cytochrome *c*. Hampsey et al. (1988) have used classical mutagenic techniques to identify functional and nonfunctional missense mutants throughout the gene. Hilgen and Pielak (1991) have used in vitro mutagensis to examine all 19 amino acid substitutions at the evolutionarily-invariant residue Phe-82. Furthermore, Auld and Pielak (1991) have used in vitro random mutagenesis to produce 16 functional missense mutants at Gly-6 and/or Phe-10 of the interface, including the two natural substitutions G6D and F10Y (Amegadzie et al., 1990; Vanfleteren et al., 1990).[1] The advantage of using in vitro random mutagenesis to study cytochrome *c* is that a complete library of mutants can be produced at any specified site. Here we report the production and phenotypic characterization of a library of mutants at Leu-94 and Tyr-97 of the C-terminal helix.

† This work is supported by an NIH FIRST Award (GM42501).

* Address correspondence to this author.

[1] Moore and Pettigrew (1990) compiled the sequences of 96 cytochromes *c*. Four additional sequences have been reported (Amegadzie et al., 1990; Vanfleteren et al., 1990; Ambler et al., 1991; Trimboli et al., 1992).

0006-2960/93/0432-929$04.00/0

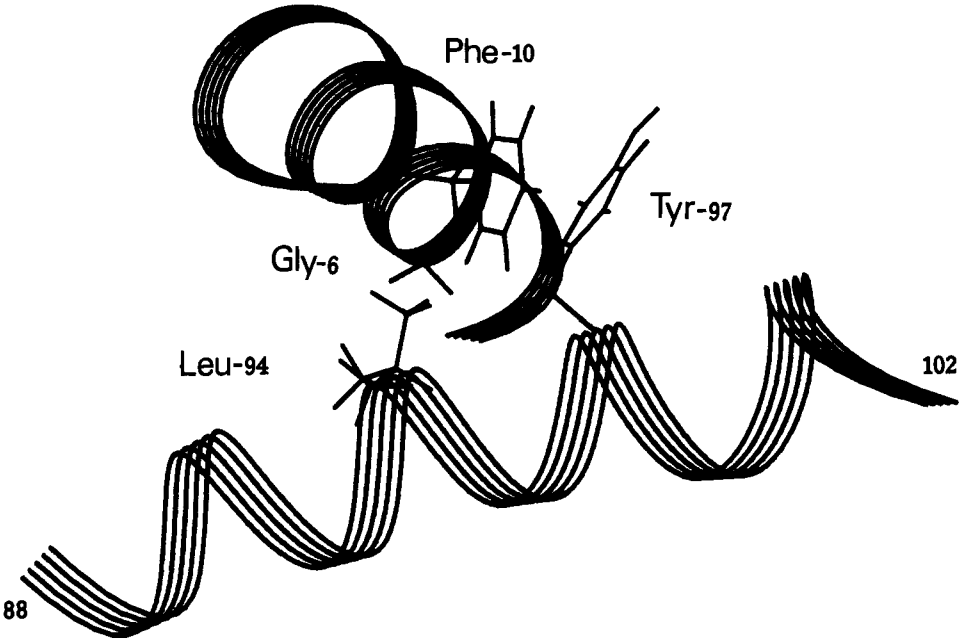

FIGURE 1: Ribbon diagram, constructed from the X-ray crystal structure of iso-1-cytochrome *c* (Louie & Brayer, 1990), showing the interaction between the N- and C-terminal helices. The C-terminal helix is shown at the bottom of the diagram extending from left to right (residues 88–102, respectively) with the N-terminal helix intersecting it (residues 2–14, bottom to top).

N-terminal helix

S	A	K	K	G	A	T	L	F	K	T	R	C
2	3	4	5	**6**	7	8	9	**10**	11	12	13	14

C-terminal helix

E	K	D	R	N	D	L	I	T	Y	L	K	K	A	C
88	89	90	91	92	93	**94**	95	96	**97**	98	99	100	101	102

FIGURE 2: Sequence of the N- and C-terminal helices of wild-type iso-1-cytochrome *c* (Smith et al., 1979).

MATERIALS AND METHODS

Phagemid. All manipulations involving bacteria and yeast were performed using a yeast-shuttle phagemid constructed by cloning the *CYC1* gene into the polylinker region of the PRS315 vector (Sikorski & Hieter, 1989). This phagemid contains the F1 origin for packaging the coding strand of the *CYC1* gene into single-stranded DNA for the purpose of mutagenesis and sequencing, the *LEU2* gene for selection in yeast, the AMP^R gene for selection in bacteria, and the *CEN6* gene for mitotic stability and low copy number (approximately one) in yeast. In bacteria this phagemid is used to produce and identify mutants. In yeast it allows analysis of function and expression of variant proteins.

Strains and Molecular Biology Techniques. DNA preparation for cloning and sequence analysis was accomplished utilizing *Escherichia coli* strains JM101 [*Δlacpro supE thi F' traD36 proAB* $lac^qZ\delta M15$ (Messing, 1983)] and JM109 [*recA1 endA1 gyrA96 thi hsdR17* ($r_K^-m_K^+$) *supE44 Δlacpro relA1 F' traD36 proAB* $lac^qZ\delta M15$ (Yanisch-Perron et al., 1985)] along with the helper phage R408, a derivative of K07 (Vieira & Messing, 1987). Bacterial transformations, DNA sequence analysis, and preparation of uracil-containing single-stranded DNA for mutagenesis were performed as described by Auld and Pielak (1991).

Selection and expression in yeast were carried out using the yeast strain GM-3C-2 [*MATα leu2–3 leu2–112 trp1–1 his4–519 cyc1–1 cyp3–1* (Faye et al., 1981)].[2] This yeast strain has the *CYC1* gene deleted and contains two point mutations in the iso-2-cytochrome *c* gene. Therefore, the only potentially-functional cytochrome *c* present in the cells is that encoded by a phagemid. Yeast transformations were performed by electroporation (Rech et al., 1990), and phagemid-bearing colonies were selected on Petri plates containing complete media minus leucine (CM–Leu; Ausubel et al., 1989). Functional alleles were distinguished from nonfunctional alleles by manipulating the two accessible energy production pathways in yeast. Yeast containing either a functional or a nonfunctional cytochrome *c* grow on fermentable carbon sources (e.g., glucose or sucrose) via glycolysis. However, only yeast containing a functional cytochrome *c* grow on nonfermentable carbon sources (e.g., glycerol or lactate) via oxidative phosphorylation. Additionally a functional iso-1-cytochrome *c* is required for growth on lactate (Sherman et al., 1974).[2] Functional alleles were identified by replica plating the phagemid-bearing colonies onto plates containing nonfermentable carbon sources (Sherman et al., 1965). Phagemids were then isolated from yeast using a modification of the spheroplast method (Rose, 1987) and reintroduced into bacteria by electroporation. Single-stranded DNA (template) was prepared from a single bacterial colony and subjected to DNA sequence analysis to determine the identity of each transformant.

Random Mutagenesis. A library of mutations at the codons for Leu-94 and Tyr-97 was produced in *E. coli* using a mixture of a mutagenic oligodeoxyribonucleotide primer, uracil-containing template DNA (Kunkel et al., 1987), and a second primer. The mutagenic primer 5′-GGCTTTTTTCAA(G/A/C)NNGGTAAT(G/A/C)NNGTCGTTTCTGTCT-3′ (where N = 25% of each base and G/A/C = 33% of each) is complementary to codons 90–101 in the coding strand of the *CYC1* gene. Ochre and opal stop codons are eliminated from the library by omitting thymine from the wobble position.

[2] *S. cerevisiae* contains two isoforms of cytochrome *c*: iso-1-cytochrome *c* (95%) encoded by the *CYC1* gene and iso-2-cytochrome *c* (5%) encoded by the *CYC7* gene (Sherman et al., 1965). In the past, *CYC7* and *CYP3* have been used interchangeably to refer to the gene encoding iso-2-cytochrome *c* (Sherman et al., 1975; Verdiere & Petrochilo, 1979).

This method of oligonucleotide-directed random mutagenesis produces 48 codons encoding all 20 natural amino acids at both positions 94 and 97.

The mutagenic efficiency was enhanced using the two-primer method (Norris et al., 1983). The second primer, 5′-ACATGATATCGACAA-3′, anneals 26 nucleotides upstream of the mutagenic primer (Smith et al., 1979). The 5′-end of each primer was phosphorylated (Zoller & Smith, 1983) and the two primers were annealed to the uracil-containing template. The reaction mixture for mutagenesis included uracil-containing template, mutagenic primer, and the second primer in a mole ratio of 1:17:25. The reaction mixture was allowed to extend for 10 min at 0 °C, 5 min at room temperature, followed by 2 h at 37 °C, and finally 1 h at room temperature.

Nomenclature. The term *mutant* is used to describe phagemids that contain mutations and to denote specific alleles. The C102T mutant of the wild-type iso-1-cytochrome *c* gene is denoted *CYC1*. The C102T mutant is used in place of the true wild-type gene because removal of the sole free cysteine sulfhydryl stabilizes the protein in vitro (Betz & Pielak, 1992) but does not have a measurable effect on structure or function (Cutler et al., 1987; Pielak et al., 1988; Gao et al., 1991).

Transformants harboring the *CYC1* gene grow on rich media contaiing glycerol (YPG) and rich media containing lactate (YPL; Sherman et al., 1965) at both 30 and 37 °C. Transformants that grow on YPG and YPL at 30 °C are defined as possessing a functional cytochrome *c*. The terms Cyc^+, Cyc^{ts}, and Cyc^- define three mutually exclusive phenotypes: Transformants possess (1) the Cyc^+ phenotype if they grow on YPG and YPL at 30 and 37 °C, (2) the Cyc^{ts} (temperature sensitive) phenotype if they grow on YPG and YPL at 30 °C but fail to grow at 37 °C after 4 days, or (3) the Cyc^- phenotype if they fail to grow on YPG and YPL at 30 and 37 °C after 4 days. A transformant exhibits a *partially* Cyc^{ts} phenotype if it grows more slowly at 37 °C than yeast harboring phagemid-borne *CYC1*. Mutants isolated from Cyc^+, Cyc^{ts}, and partially Cyc^{ts} transformants are *functional mutants*, and the amino acids at positions 94 and 97 found among these mutants are *compatible with function*. Mutants isolated from Cyc^- transformants are *nonfunctional mutants*. *Missense mutations* alter the deduced wild-type amino acid sequence. Mutationally-altered cytochrome *c* are referred to as *protein variants*. Missense mutants and their concomitant protein variants are denoted by their deduced amino acid substitution(s) using the one-letter code. Double-missense mutations within one allele are separated by a semicolon (e.g., L94A;Y97F).

Protein Purification. A modification of the method of Sherman et al. (1968) was used to prepare protein variants. A 5-mL overnight culture grown in CM–Leu was inoculated into 1 L of rich media containing sucrose (YPS) and grown to saturation at 30 °C. After batch chromatography the absorbance of each variant was measured at 410 nm and 550 nm and converted to protein concentrations using extinction coefficients of 106 100 M^{-1} cm^{-1} and 27 600 M^{-1} cm^{-1}, respectively (Margoliash & Frohwirt, 1959). Several Cyc^+ and Cyc^{ts} transformants were grown in YPS at 30 °C, and the yield of protein was compared to yeast harboring the C102T mutant grown under the same conditions.

RESULTS

Properties of the Library. The library contained in *E. coli* was constructed to allow the generation of all 400 combinations of amino acids at codons 94 and 97 and to maximize the number of double and triple base-pair changes. The first round of mutagenesis yielded a mutagenic efficiency of 88%. To increase the number of mutants and the randomness of the library, a second round of mutagenesis was performed (Auld & Pielak, 1991). The second round yielded a mutagenic efficiency of 91% and a much greater randomization of mutations.

Table I: Nucleotide Percentages at Codons 94 and 97 in the *E. coli* Library[a]

base	codon 94			base	codon 97		
T	49	45	37	T	48	15	27
A	17	18	17	A	27	44	0
C	16	24	25	C	13	28	62
G	18	13	21	G	12	13	11

[a] The wild-type codon is TTA for codon 94 and TAC for codon 97.

After the second round of mutagenesis, 109 phagemids from *E. coli* were chosen at random to determine the types of mutations present within the library. The distribution of nucleotides at each position of codons 94 and 97 was determined by subjecting all 109 phagemids to DNA sequence analysis (Table I). Table I shows that a bias for wild-type nucleotides remains in the library even after two rounds of mutagenesis. This preference exists because oligonucleotides containing the least number of mismatches form the most stable heteroduplexes. The wild-type wobble base of codon 94, adenine, was not overrepresented in the coding strand because thymine was excluded from this position of the mutagenic primer. From DNA sequence analysis, the *E. coli* library was found to contain 9% wild-type alleles, 28% single-missense mutants, 58% double-missense mutants, and 5% amber codons. Seventy-six percent of the missense mutants were unique. These results demonstrate the diversity and abundance of amino acid substitutions within the library.

Determination of Phenotype and Relative Growth Rate. Transformation of yeast with the library was performed to determine which mutants yield a functional cytochrome *c*. After selection for cells harboring the phagemid, 857 phagemid-containing colonies were screened for cytochrome *c* function. Of these, 308 contained a functional cytochrome *c* and 549 displayed the Cyc^- phenotype. Yeast transformants possessing a functional cytochrome *c* were subdivided into those possessing either the Cyc^+ or Cyc^{ts} phenotypes. Of the 269 functional candidates screened, 57 displayed the Cyc^{ts} phenotype. Phagemids isolated from 96 randomly-chosen functional candidates were subjected to DNA sequence analysis: 38% were wild-type, 55% were missense mutants, and 7% were mutants that did not alter the amino acid sequence. To test several hypotheses concerning the interface, yeast were *separately* transformed with three other missense mutants from the *E. coli* library: L94I, Y97N, and Y97D. These three transformants possessed a functional cytochrome *c*. Table II lists the 31 unique missense mutants observed to yield a functional cytochrome *c*. Only 17 of the functional missense mutants were also found among the 71 different missense mutants identified in the *E. coli* library. This demonstrates the power of our genetic selection system.

Yeast were retransformed with the phagemids containing all 31 missenese mutants to determine growth rates and to divide mutants into phenotypes on the basis of activity in vivo. Four colonies from each transformation were assessed for the ability to grow at 30 and 37 °C on both YPG and YPL. After 4 days, a growth rate and phenotype was assigned to each mutant relative to yeast harboring the C102T mutant (Table II). All four colonies from each transformant yielded

Table II: Missense Mutants at Codons 94 and/or 97 in Yeast *CYC1*

codon 94	codon 97	codons 94 and 97	
Cyc$^+$	Cyc$^+$	Cyc$^+$	Cyc$^-$
L94F (4, 3)[a]	Y97F (2, 3)[b]	L94F;Y97F (3, 3)	L94K;Y97C (3, 0)
L94V (3, 3)[b]	Y97E (1, 3)	L94A;Y97F (1, 3)	L94K;Y97S (1, 0)
L94S (3, 2)	Y97C (6, 2)	L94I;Y97F (2, 2)[b]	L94S;Y97K (1, 0)
L94T (2, 2)	Y97A (1, 1)	L94V;Y97H (1,2)	L94Y;Y97R (1, 0)
L94C (1, 2)	Y97N (0, 2)		L94Y;Y97S (1, 0)
L94M (3, 1)	partially Cyc^{ts}	partially Cyc^{ts}	L94Y;Y97I (1, 0)
L94I (0, 3)[b]	Y97M (1, 2)	L94V;Y97C (1, 3)	L94Y;Y97F (1, 0)
partially Cyc^{ts}	Cyc^{ts}	Cyc^{ts}	L94F;Y97N (1, 0)
L94A (1, 1)	Y97T (3, 3)	L94V;Y97S (1, 2)	L94N;Y97N (1, 0)
Cyc^{ts}	Y97I (1, 2)	L94V;Y97N (1, 1)	L94Q;Y97T (1, 0)
L94G (1, 2)	Y97S (4, 1)	L94M;Y97C (1, 1)	L94A;Y97T (1, 0)
L94N (1,1)	Y97L (1, 1)	L94N;Y97F (1, 1)	
Cyc$^-$	Y97D (0, 1)	L94I;Y97N (2, 1)	
L94D (0, 0)	Cyc$^-$		
	Y97P (0, 0)		

[a] The first number in parentheses represents the number of times the mutant was observed in yeast, and the second number represents the relative growth rate of variants with respect to wild-type (3) and nonfunctional mutants (0). Mutants listed with a 0 as the first number were identified in the *E. coli* library but not identified in the yeast library. Yeast were *separately* transformed with these mutants, and their phenotypes and growth rates were determined. [b] Occurs in nature (see text).

consistent results. The sequence of the entire coding region was determined to ensure that phenotypes and growth rates were not the effect of additional mutations within the *CYC1* gene. Nine different functional mutants were fully sequenced (L94G, L94A, L94M, L94T, L94C, Y97A, Y97C, Y97S, and L94I;Y97F), and no additional mutations were identified.

Phagemids from 15 transformants displaying the Cyc$^-$ phenotype were also isolated and subjected to DNA sequence analysis. Both an amber stop codon at position 94 and a frameshift mutation at position 97 were found among the 15 nonfunctional mutants, showing that the interface is required for protein function. The remaining 13 candidates were double-missense mutants of which 11 were unique (Table II). Only three of the nonfunctional mutants were also observed in the *E. coli* library. Because nonfunctional single-missense mutants were not observed in yeast, the *E. coli* library was examined for possible suspects. Yeast were *separately* transformed with the mutants L94D and Y97P of the *E. coli* library. The transformants were screened for a functional cytochrome *c*, and these mutants were found to cause the Cyc$^-$ phenotype (Table II). The sequence of the entire coding region was determined for three nonfunctional mutants (L94K;Y97C, L94F;Y97N and L94A;Y97T), and no additional mutations were identified.

Amino Acid Substitutions at Leu-94 or Tyr-97. At position 94, we find that 11 different amino acids, including the wild-type residue, are compatible with function. Eight cause the Cyc$^+$ phenotype, two cause the Cyc^{ts} phenotype, and one causes the partially Cyc^{ts} phenotype. Previously, only three different amino acids have been observed at position 94: isoleucine (*Neurospora crassa*, *Euglena gracilis*, and *Euglena viridis*), valine (*Schizosaccharomyces pombe* and *Crithidia fasciculata*), and leucine (in 95 other eukaryotes).[1] We find that yeast harboring the L94I and L94V mutants display the Cyc$^+$ phenotype (Table II), consistent with expected results.

At position 97, we find that 12 different amino acids are compatible with function. Six cause the Cyc$^+$ phenotype, five cause the Cyc^{ts} phenotype, and one causes the partially Cyc^{ts} phenotype. Previously, only two different amino acids have been observed at this position: phenylalanine (*N. crassa*, *Enteromorphia interstinalis*, and *Drosophila melanogaster*) and tyrosine (in 97 other eukaryotes).[1] We find that yeast harboring the Y97F mutant display the Cyc$^+$ phenotype, consistent with expected results.

Amino Acid Substitutions at Leu-94 and Tyr-97. At positions 94 and 97, we find that 11 different combinations of amino acids are compatible with function. Five cause the Cyc$^+$ phenotype, five cause the Cyc^{ts} phenotype, and one causes the partially Cyc^{ts} phenotype. We identified the mutant L94I;Y97F, which occurs naturally in *N. crassa*, among the 31 unique functional mutants. We find that yeast harboring this mutant display the Cyc$^+$ phenotype, consistent with expected results.

Protein Isolation. When yeast possessing Cyc$^+$ or Cyc^{ts} mutants (Table II) are grown at 30 °C, the yield of cytochrome *c* is at least 36% of the yield from yeast harboring phagemid-borne *CYC1*. It was not possible to isolate cytochrome *c* from transformants displaying the Cyc$^-$ phenotype. Cytochrome *c* is easily identifiable because of the salmon-pink color of the protein in solution. These results confirm that protein variants can be isolated for structural and thermodynamic studies.

Statistical Analysis. A Monte-Carlo-based program (Reidhaar-Olson & Sauer, 1990) was used to determine the probability that all amino acids compatible with function are observed at a specified codon. Four simulations were performed at each position using only mutants isolated from the functional yeast library. At position 94 there is a 0.56 probability that all functional alleles are observed. At position 97 the probability is 0.62. Of the possible 400 amino acid combinations at these two positions, there is a 0.35 probability (i.e., 0.56×0.62) that all functional combinations have been observed. This indicates that the library contains more novel amino acid substitutions than we have identified.

Analysis of Mutants. We analyzed the side chain volumes, interresidue contact energies, helix-forming tendencies, polarities, and charges of our mutants in an attempt to discover a link between amino acid substitution and phenotype. Changes in volume were determined by subtracting the volume of the wild-type residue (Zamyatnin, 1972) from the volume of each substituted residue. Changes in volume range from +36 Å^3 to −101 Å^3 at position 94 (L94F to L94G) and from −0.2 Å^3 to −112 Å^3 at position 97 (Y97F to Y97A). No absolute volume constraint can be assigned because the volume changes calculated for nonfunctional mutants also lie within the above ranges. Interresidue contact energies (Miyazawa & Jerningan, 1985) were determined between all missense mutants listed in Table II and Gly-6 and Phe-10 of the

N-terminal helix. The helix-forming tendency (O'Neil & DeGrado, 1990) of each missense mutant was also examined. Neither the interresidue contact energies nor helix-forming tendencies appear to influence phenotype. The polarity of mutants was assessed in terms of hydrophobicity and hydrophilicity (Wolfenden et al., 1981). A variety of both hydrophilic and hydrophobic amino acids substitutions were found to be compatible with function at positions 94 and 97 (Table II). However, examination of the double-missense mutants indicates that one hydrophobic residue is required at either position to maintain function. Several charged residues were identified among the missense mutants (Table II). Negatively-charged residues were observed among the functional mutants at position 97, but not at 94. Positively-charged residues were not identified in any of the functional mutants, but occurred with high frequency among the nonfunctional mutants. These findings indicate that negative-charged residues are compatible with function at position 97, but positively-charged residues are not tolerated within the interface. Therefore, in terms of polarity and charge, a correlation exists between amino acid substitution and phenotype.

The relationship between phenotype and amino acid substitutions was also examined using the mutation matrix constructed from exhaustive matching of the entire protein sequence database (Gonnet et al., 1992). This log probability matrix is generally used to align and compare protein sequences (Dayhoff et al., 1978; George et al., 1990). We used the mutation matrix to assign a probability score to each mutant in Table II and each mutant from our examination of Gly-6 and Phe-10 (Auld & Pielak, 1991; J. Beasley, unpublished results). For single-missense mutants the score is the matrix element, and for double-missense mutants the score is the sum of the elements. A positive score indicates that the two residues replace each other more often in related sequences than in random sequences of the same composition. A negative score indicates that replacement occurs less often than in random sequences (Dayhoff et al., 1978). The histogram shown in Figure 3 relates the observed phenotype to the score. In general, as the score of our mutants decreases, the phenotype changes from Cyc^+ to Cyc^-.

DISCUSSION

The helix–helix interface of cytochrome *c* has been linked to essential interior packing. Alteration of the interface should directly affect the packing and, therefore, the function of cytochrome *c*. However, changes within the interface could also affect aggregation and/or degradation of the apoprotein, subsequent heme attachment, and import into the mitochondria. To examine the relationship between the sequence of the interface and the function of the protein, novel amino acid substitutions were produced within the interface using oligonucleotide-directed mutagenesis. In this study, a library of mutations was created in the C-terminal helix at the codons for Leu-94 and/or Tyr-97. Manipulation of this library allowed us to identify 95 of the 400 possible amino acid combinations at positions 94 and/or 97, 45 of which were screened for function and phenotype (Table II).

The results reported here and in a study of the N-terminal helix (Auld & Pielak, 1991) show that the interface is more flexible than the evolutionary record indicates. At positions 94 and/or 97 we isolated 31 unique missense mutants that produce a functional cytochrome *c*. To accommodate so many different amino acids, conformational shifts of the interacting helices must occur through side-chain rotations and changes in the helix–helix torsion angle (Chothia & Lesk, 1985). These 31 missense mutants vary in such aspects as side chain volume, interresidue contact energy, helix-forming tendency, polarity, and charge. Although the interface appears to be extremely flexible, we find constraints on amino acid substitutions that are compatible with function.

A majority of the missense mutants observed at position 97 can be linked to weakly polar interactions with Phe-10. These mutants—Y97F, Y97M, Y97C, Y97N, Y97S, Y97T, Y97D, and Y97E—result in a functional cytochrome *c* (Table II). Our findings are consistent with the suggestion that weakly polar aromatic–aromatic, oxygen–aromatic, nitrogen–aromatic, and sulfur–aromatic interactions stabilize the native state of proteins (Thomas et al., 1982; Reid et al., 1985; Burley & Petsko, 1988). For example, the mutant Y97F conserves the aromatic–aromatic interaction with Phe-10 and appears to stabilize the L94A mutant (i.e., L94A;Y97F) with respect to function (Table II). This stabilizing effect may be explained by the results of Burley and Petsko (1985), who found that Phe pairs occur more frequently than Tyr pairs possibly because Phe is better suited to the protein interior. We also find that substitution of Tyr-97 with Arg or Lys results in loss of function. Similar results were reported for substitutions at Phe-10 (Auld & Pielak, 1991). Burley and Petsko (1986) suggest that interactions between Arg or Lys and aromatic residues stabilize proteins, indicating that Y97R and Y97K may disturb other interactions within the interface in addition to the aromatic–aromatic interaction (Serrano et al., 1991).

Charged residues buried within the interface affect the function of cytochrome *c*. The negatively-charged residues Asp and Glu are compatible with function at position 97, but not at position 94. This may be explained by the existence of a weakly-polar oxygen–aromatic interaction between these residues at position 97 and Phe-10 that does not exist between position 94 and Phe-10. Charged residues buried within a hydrophobic region which do not disturb the structure of a protein have been reported previously (Dao-pin et al., 1991; Stites et al., 1991). No positively-charged residues were observed among the functional mutants, yet positively-charged residues occur with high frequency among the nonfunctional mutants (Table II). These results indicate that positively-charged residues within the interface abolish function. Similar results were reported for substitutions at Gly-6 and Phe-10 (Auld & Pielak, 1991). Like all mutations in Table II that abolish function, these substitutions may adversely affect transcription, translation, transport, posttranslational modification, and/or the free energy of denaturation.

Side chain volume does not strongly influence protein function even though the peg-in-hole and aromatic–aromatic interactions of the interface are conserved in nature. As was noted for substitutions at Phe-10 of the N-terminal helix (Auld & Pielak, 1991), there is a broad volume constraint for functional mutants at positions 94 and/or 97. Tyrosine-97 can be replaced with amino acids as small as Ala while still maintaining function. The cavity-creating mutants L94G and L94A also produce functional proteins, contrary to what is expected for a peg-in-hole interaction. However, a solution of the L94G variant changed from salmon-pink to yellow after less than 48 h at 4 °C while solutions of the wild-type protein and other variants remained pink. This color change suggests that L94G is less stable than other mutants. Similar cavity-creating mutants destabilize phage T4 lysozyme by 2.7 to 5.0 kcal mol^{-1} at room temperature (Eriksson et al., 1992). In contrast to the apparent destabilizing effect of L94G, the growth rate increased and the phenotype changed from

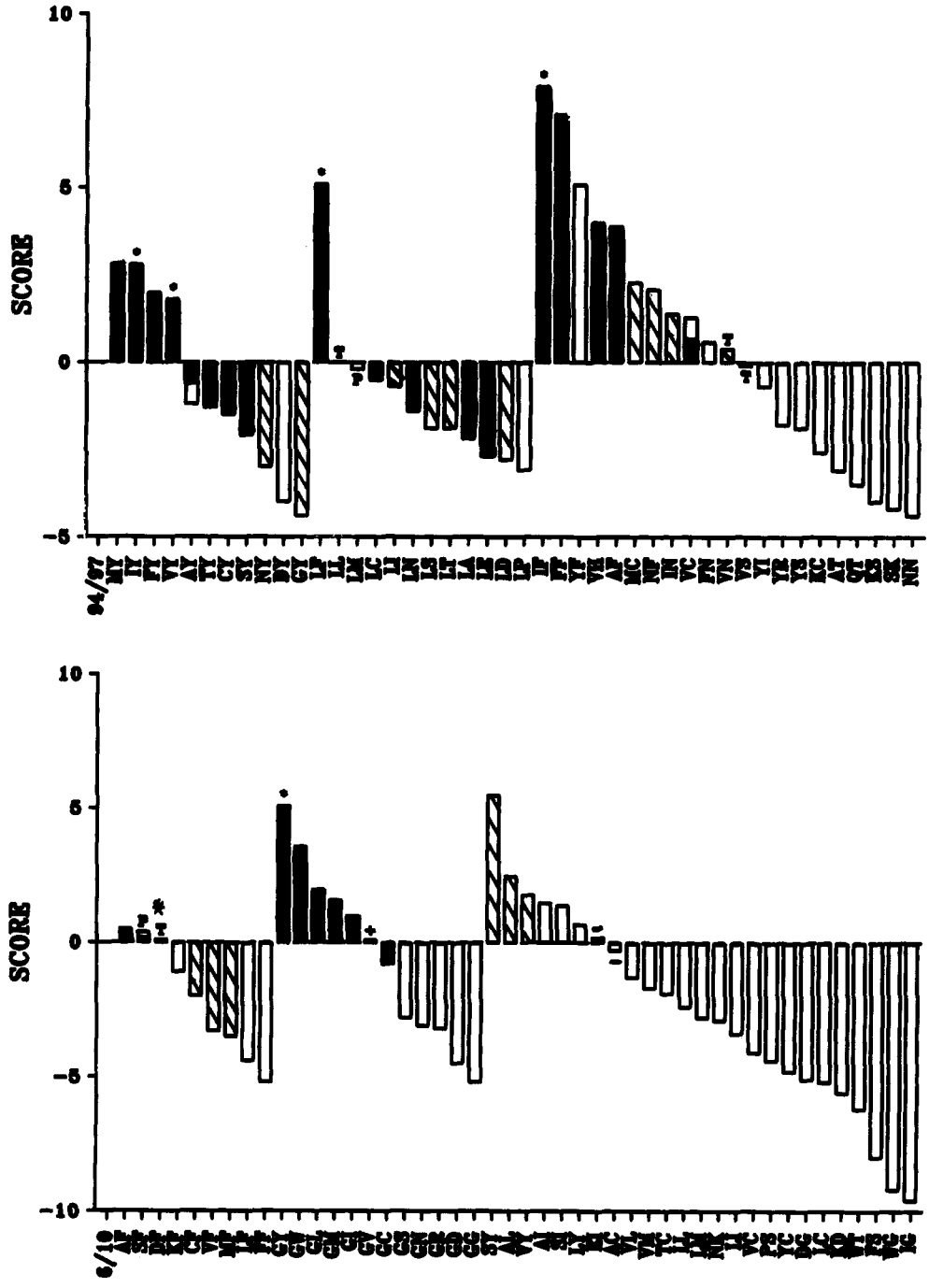

FIGURE 3: Histogram relating the score calculated from the mutation matrix (see text) to the observed phenotype for mutants at positions Leu-94 and/or Tyr-97 (top panel) and Gly-6 and/or Phe-10 (bottom panel). Filled, half-filled, striped, and open bars indicate the Cyc$^+$, partially Cyc^{ts}, Cyc^{ts}, and Cyc$^-$ phenotypes, respectively. Where the matrix element is near zero, the phenotype is indicated (+, Cyc$^+$; **T**, Cyc^{ts}; **P**, partially Cyc^{ts}; –, Cyc$^-$). Mutants that correspond to naturally-occurring cytochromes *c* are indicated by an asterisk.

partially Cyc^{ts} to Cyc$^+$ when a second mutation, Y97F, was combined with L94A (Table II). These changes are interesting because there is not a significant difference in side chain volume or polarity when Tyr-97 is replaced with Phe. This type of functional effect has also been noted in second site revertants (Shortle & Lin, 1985), including those found in different areas of iso-1-cytochrome *c* (Schweingruber et al., 1979; Das et al., 1989).

Our results indicate that the missense mutants listed in Table II are not the only mutants capable of producing a functional cytochrome *c*. This is apparent when one considers that only a fraction of phagemids present in the library were subjected to genetic selection and DNA sequence analysis. In fact, statistical analysis suggests that there is only a 35% chance that we have identified all the amino acid combinations that yield a functional protein. The observation that all functional double-missense mutants consisted of two functional single-missense mutants led us to predict other mutants that result in a functional cytochrome *c*. For example, we predicted and subsequently found the mutants L94I and Y97N to be functional (Table II). However, we noted that not all functional single-missense mutants yield a functional mutant when combined (e.g., L94F;Y97N, L94N;Y97N, and L94A;Y97T). This suggests that certain substitutions have additive or synergistic effects, perhaps because of stereochemical restraints at the interface. The fact that single-missense mutants comprise only 28% of the missense mutants identified in the *E. coli* library but nearly 75% of functional

missense mutants provides evidence for additive or synergistic effects. Using the results listed in Table II, we estimate that a maximum of 120, or 30%, of the 400 possible amino acid combinations will yield a functional cytochrome *c*. Despite the evolutionary conservation of the interface, many more sequences are compatible with function than are observed in nature.

The maximum number of amino acid combinations compatible with function was also estimated using the histogram (Figure 3) derived from the mutation matrix (Gonnet et al., 1992). A probability score was determined for all 400 amino acid combinations at positions 6 and/or 10, and at positions 94 and/or 97. All combinations possessing a score greater than the cutoff between functional and nonfunctional mutants (Cyc^-) were considered compatible with function, except at position 97 where only the observed functional mutants were included. On the basis of these criteria, we estimate that 34% of the possible substitution at positions 94 and/or 97 and 14% of the possible substitutions at positions 6 and/or 10 are compatible with function. These estimates are remarkably similar to those made solely by inspection of the functional mutants [30% at positions 94 and/or 97; 10% at positions 6 and/or 10 (Auld & Pielak, 1991)]. The agreement between our experimental results and the mutation matrix (Gonnet et al., 1992) shows that the elements of the matrix are intimately related to the biological function of cytochrome *c*, even though the matrix was derived for sequences more distantly related than eukaryotic cytochromes *c*. Therefore, our results demonstrate that the mutation matrix can be used to predict the effect of sequence changes in natural proteins. The large percentage of substitutions that are compatible with function indicates that prediction of a native fold from an amino acid sequence is a complex task. However, we suggest that the mutation matrix can be useful in optimizing the sequence of novel proteins and that prediction endeavors will be achieved.

Examination of our results at Leu-94 and Tyr-97 and at Gly-6 and Phe-10 (Auld & Pielak, 1991) reveals new ideas concerning the construction of the interface. Clearly the C-terminal helix is more tolerant of substitutions than is the N-terminal helix. We suggest that a maximum of 30% of the 400 possible substitutions at Leu-94 and Tyr-97 result in a functional cytochrome *c*, and Auld and Pielak (1991) suggest a maximum of 10% for substitutions at Gly-6 and Phe-10. Additionally, we find that Tyr-97 accepts four more amino acid substitutions than Phe-10, and Leu-94 accepts five more substitutions than does Gly-6. There are probably several reasons for the increased plasticity of the C-terminal helix. First, stricter constraints are imposed on substituting larger residues for Gly-6. Second, intramolecular interactions between Phe-10 and other residues are essential for protein structure and stability, such as its interaction with Val-20 (Fetrow & Farid, 1992). Third, the fact that the N-terminal helix is terminated by the cysteines that form the two thioether linkages to the heme may also explain stricter constraints imposed on positions 6 and 10. Finally, there is a greater allowed range of motion for the C-terminal helix because the C-terminal helix is more stable than the N-terminal helix. In fact, in their comparison of several cytochromes *c*, Chothia and Lesk (1985) note that relative to the heme the C-terminal helices shift to a significantly greater degree than do the N-terminal helices. Evidence for stability differences between the two helices has been obtained by determining amide proton exchange rates. The rate for residues within the C-terminal helix is approximately 3 orders of magnitude slower than that of the N-terminal helix (D. Auld, unpublished results).

Our results suggest that although packing interactions are important, stringent criteria for interior packing are not essential to produce a functional protein. This is consistent with the idea (Behe et al., 1991) that the native fold determines packing, but packing does not determine the native fold. Contrary to the observations of Lim and Sauer (1989), however, we find a much weaker interdependence between hydrophobicity and function. Our results also show that the mutation matrix (Gonnet et al., 1992) can be used to predict the effect of sequence changes in natural proteins and to optimize the sequence of novel proteins. To examine quantitatively the relationship between protein sequence, packing, and stability, we have begun thermodynamic and structural studies of the protein variants isolated from the mutants listed in Table II.

ACKNOWLEDGMENT

We thank P. Anderson, D. Auld, J. Beasley, S. Betz, S. Hilgen, and A. Saunders for helpful discussions and encouragement.

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.
Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1989) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York.
Behe, M. J., Lattman, E. E., & Rose, G. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4195–4199.
Berghuis, A. M., & Brayer, G. D. (1992) *J. Mol. Biol. 223*, 959–976.
Betz, S. F., & Pielak, G. J. (1992) *Biochemistry 31*, 12337–12344.
Burley, S. K., & Petsko, G. A. (1985) *Science 229*, 23–28.
Burley, S. K., & Petsko, G. A. (1986) *FEBS Lett. 203*, 139–143.
Burley, S. K., & Petsko, G. A. (1988) *Adv. Protein Chem. 39*, 125–189.
Chothia, C., & Lesk, A. M. (1985) *J. Mol. Biol. 182*, 151–158.
Cutler, R. L., Pielak, G. J., Mauk, A. G., & Smith, M. (1987) *Protein Eng. 1*, 95–99.
Dao-pin, S., Anderson, D. E., Baase, W. A., Dahlquist, F. W., & Matthews, B. W. (1991) *Biochemistry 30*, 11521–11529.
Das, G., Hickey, D. R., McLendon, D., McLendon, G., & Sherman, F. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 496–499.
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.
Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P., & Matthews, B. W. (1992) *Science 255*, 178–183.
Faye, G., Leung, D. W., Tatchell, K., Hall, B. D., & Smith, M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2258–2262.
Fetrow, J., & Farid, L. (1992) *Biophys. J. 61*, A214.
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.
Hampsey, D. M., Das, G., & Sherman, F. (1988) *FEBS Lett. 231*, 275–283.
Hilgen, S. E., & Pielak, G. J. (1991) *Protein Eng. 4*, 575–578.
Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367–382.
Lim, W. A., & Sauer, R. T. (1989) *Nature 339*, 31–36.

Louie, G. V., & Brayer, G. D. (1990) *J. Mol. Biol. 214*, 527–555.
Margoliash, E., & Frohwirt, N. (1959) *Biochem. J. 71*, 570–572.
Mathews, F. S. (1985) *Prog. Biophys. Mol. Biol. 45*, 1–56.
Messing, J. (1983) *Methods Enzymol. 101*, 20–78.
Moore, G. R., & Pettigrew, G. W. (1990) *Cytochromes c: Evolutionary, Structural and Physicochemical Aspects*, pp 116–127, Springer-Verlag, New York.
Miyazawa, S., & Jerningan, R. L. (1985) *Macromolecules 18*, 534–552.
Norris, K., Norris, F., Christiansen, L., & Fiil, N. (1983) *Nucleic Acids Res. 11*, 5103–5112.
O'Neil, K. T., & DeGrado, W. F. (1990) *Science 250*, 646–651.
Pielak, G. J., Boyd, J., Moore, G. R., & Williams, R. J. P. (1988) *Eur. J. Biochem. 177*, 167–177.
Rech, E. L., Dobson, M. J., Davey, M. R., & Mulligan, B. J. (1990) *Nucleic Acids Res. 18*, 1313.
Reid, K. S. C., Lindley, P. F., & Thorton, J. M. (1985) *FEBS Lett. 190*, 209–213.
Reidhaar-Olson, J. F., & Sauer, R. T. (1990) *Proteins: Struct., Funct., Genet. 7*, 306–316.
Rose, M. D. (1987) *Methods Enzymol. 152*, 481–504.
Schweingruber, M. E., Stewart, J. W., & Sherman, F. (1979) *J. Biol. Chem. 254*, 4132–4143.
Serrano, L., Bycroft, M., & Fersht, A. R. (1991) *J. Mol. Biol. 218*, 465–475.
Sherman, F., Taber, H., & Campbell, W. (1965) *J. Mol. Biol. 13*, 21–39.
Sherman, F., Stewart, J. W., Parker, J. H., Inhaber, E., Shipman, N. A., Putterman, G. J., Gardisky, R. L., & Margoliash, E. (1968) *J. Biol. Chem. 243*, 5446–5456.
Sherman, F., Stewart, J. W., Jackson, M., Gilmore, R. A., & Parker, J. H. (1974) *Genetics 77*, 255–284.
Sherman, F., Jackson, M., Liebman, S. W., Schweingruber, A. M., & Stewart, J. W. (1975) *Genetics 81*, 51–73.
Shortle, D., & Lin, B. (1985) *Genetics 110*, 539–555.
Sikorski, R. S., & Hieter, P. (1989) *Genetics 122*, 19–27.
Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L., & Hall, B. D. (1979) *Cell 16*, 753–761.
Stites, W. E., Gittis, A. G., Lattman, E. E., & Shortle, D. (1991) *J. Mol. Biol. 221*, 7–14.
Thomas, K. A., Smith, G. M., Thomas, T. B., & Feldmann, R. J. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4843–4847.
Trimboli, C. A., Barber, M. J., Clark, M., Young, C., & Neame, P. J. (1992) *Biophys. J. 61*, A201.
Vanfleteren, J. R., Evers, E. A. I. M., Van de Werken, G., & Van Beeumen, J. J. (1990) *Biochem. J. 271*, 613–620.
Verdiere, J., & Petrochilo, E. (1979) *Mol. Gen. Genet. 175*, 209–216.
Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3–11.
Wolfenden, R., Andersson, L., Cullis, P. M., & Southgate, C. C. B. (1981) *Biochemistry 20*, 849–855.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103–119.
Zamyatnin, A. A. (1972) *Prog. Biophys. Mol. Biol. 24*, 107–123.
Zoller, M. J., & Smith, M. (1983) *Methods Enzymol. 100*, 468–500.