Kurowski, B., & Ludwig, B. (1987) *J. Biol. Chem. 262*, 13805-13811.
Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105-132.
Laemmli, U. K. (1970) *Nature 227*, 680-685.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265-275.
Matsudaira, P. (1987) *J. Biol. Chem. 262*, 10035-10038.
Meinhardt, S. W., Kula, T., Yagi, T., Lillich, T., & Ohnishi, T. (1987) *J. Biol. Chem. 262*, 9147-9153.
Nishikimi, M., Hosokawa, Y., Toda, H., Suzuki, H., & Ozawa, T. (1988) *Biochem. Biophys. Res. Commun. 157*, 914-920.
Ohnishi, T. (1979) in *Membrane Proteins in Energy Transduction* (Capaldi, R. A., Ed.) pp 1-87, Dekker, New York.
Ohnishi, T., & Salerno, J. C. (1982) *Iron-Sulfur Proteins 4*, 285-327.
Patel, S. D., Aebersold, R., & Attardi, G. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 4225-4229.
Phillips, M. K., Hederstedt, L., Hasnain, S., Rutberg, L., & Guest, J. R. (1987) *J. Bacteriol. 169*, 864-873.
Pilkington, S. J., & Walker, J. E. (1989) *Biochemistry 28*, 3257-3264.
Pilkington, S. J., Skehel, J. M., Gennis, R. B., & Walker, J. E. (1991) *Biochemistry 30*, 2166-2175.
Ragan, C. I. (1987) *Curr. Top. Bioenerg. 15*, 1-36.
Ragan, C. I., Galante, Y. M., Hatefi, Y., & Ohnishi, T. (1982) *Biochemistry 21*, 590-594.
Raitio, M., Jalli, T., & Saraste, M. (1987) *EMBO J. 6*, 2825-2833.
Raitio, M., Pispa, J. M., Metso, T., & Saraste, M. (1990) *FEBS Lett. 261*, 431-435.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463-5467.
Shine, J., & Dalgarno, L. (1975) *Nature 254*, 34-38.
Skehel, J. M., Pilkington, S. J., Runswick, M. J., Fearnley, I. M., & Walker, J. E. (1991) *FEBS Lett. 282*, 135-138.
Stouthamer, A. H. (1980) *Trends Biochem. Sci. (Pers. Ed.) 5*, 164-166.
Tran-Betcke, A., Warnecke, U., Bocker, C., Zaborosch, C., & Friedrich, B. (1990) *J. Bacteriol. 172*, 2920-2929.
Tuschen, G., Sackmann, U., Nehls, U., Haiker, H., Buse, G., & Weiss, H. (1990) *J. Mol. Biol. 213*, 845-857.
Tzagoloff, A., Capitanio, N., Nobrega, M. P., & Gatti, D. (1990) *EMBO J. 9*, 2759-2764.
Van Spanning, R. J. M., Wansell, C., Harms, N., Oltmann, L. F., & Stouthamer, A. H. (1990) *J. Bacteriol. 172*, 986-996.
von Bahr-Lindström, H., Galante, Y. M., Persson, M., & Jörnvall, H. (1983) *Eur. J. Biochem. 134*, 145-150.
Weiss, H., Friedrich, T., Hofhaus, G., & Preis, D. (1991) *Eur. J. Biochem. 197*, 563-576.
Xu, X., & Yagi, T. (1991) *Biochem. Biophys. Res. Commun. 174*, 667-672.
Xu, X., Matsuno-Yagi, A., & Yagi, T. (1991) *Biochemistry 30*, 6422-6428.
Yagi, T. (1986) *Arch. Biochem. Biophys. 250*, 302-311.
Yagi, T. (1987) *Biochemistry 26*, 2822-2828.
Yagi, T. (1989) *Tanpakushitsu Kakusan Koso 34*, 351-363.
Yagi, T. (1990) *Arch. Biochem. Biophys. 281*, 305-311.
Yagi, T. (1991) *J. Bioenerg. Biomembr. 23*, 211-225.
Yagi, T., & Hatefi, Y. (1988) *J. Biol. Chem. 263*, 16150-16155.
Yagi, T., & Dinh, T. M. (1990) *Biochemistry 29*, 5515-5520.
Yasunobu, K. T., & Tanaka, M. (1980) *Methods Enzymol. 69*, 228-238.

Constraints on Amino Acid Substitutions in the N-Terminal Helix of Cytochrome *c* Explored by Random Mutagenesis†

Douglas S. Auld and Gary J. Pielak*
Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290
Received February 5, 1991; Revised Manuscript Received June 19, 1991

ABSTRACT: The interaction of the N- and C-terminal helices is a hallmark of the cytochrome *c* family. Oligodeoxyribonucleotide-directed random mutagenesis within the gene encoding the C102T protein variant of *Saccharomyces cerevisiae* iso-1-cytochrome *c* was used to generate a library of mutations at the evolutionary invariant residues Gly-6 and Phe-10 in the N-terminal helix. Transformation of this library (contained on a low-copy-number yeast shuttle phagemid) into a yeast strain lacking a functional cytochrome *c*, followed by selection for cytochrome *c* function, reveals that 4–10% of the 400 possible amino acid substitutions are compatible with function. DNA sequence analysis of phagemids isolated from transformants exhibiting the functional phenotype elucidates the requirements for a stable helical interface. Basic residues are not tolerated at position 6 or 10. There is a broad volume constraint for amino acids at position 6. The amino acid substitutions observed to be compatible with function at Phe-10 show that the hydrophobic effect alone is sufficient to promote helical association. There are severe constraints that limit the combinations consistent with function, but the number of functionally consistent combinations observed exemplifies the plasticity of proteins.

The pairing of α-helices is one of the most fundamental types of protein tertiary structure. The cytochromes *c* are a family of evolutionarily conserved α-helical proteins, and the pairing of the N- and C-terminal helices is found in all cytochromes *c* (Matthews, 1985). Examination of the crystal structure of iso-1-cytochrome *c* from the yeast *Saccharomyces cerevisiae* (Louie & Brayer, 1990) shows that the helical axes are inclined at approximately 90° and their interaction involves the packing

† This work was supported by an NIH FIRST Award (GM42501) and by a grant from the Petroleum Research Foundation (21597-G).
* Address correspondence to this author.

0006-2960/91/0430-8684$02.50/0

of the invariant residues Leu-94 with Gly-6 and Phe-10 with Tyr-97 [throughout this paper, the numbering system for higher eukaryotic cytochromes *c* is used (Cutler et al., 1987)]. Roder et al. (1988) have shown that the association of the N- and C-terminal helices represents an early event along the folding pathway. Helices that pack perpendicularly to one another are found in other proteins including DNA binding proteins (Richardson & Richardson, 1988a; Efimov, 1984). The N- and C-terminal helices of cytochrome *c* were chosen as a model system for the study of this structural motif.

Cytochrome *c* is the penultimate electron-transfer protein of the eukaryotic respiratory chain. *S. cerevisiae* contains two systems for energy production: the Embden–Myerhof pathway in the cytosol and oxidative phosphorylation in the mitochondria. Only the latter requires a functional cytochrome *c*. Yeast grow on fermentable carbon sources (e.g., glucose and sucrose) without a functional cytochrome *c*, but growth on the nonfermentable carbon source glycerol requires a functional cytochrome *c*, and growth on the nonfermentable carbon source lactate requires exclusively a functional iso-1-cytochrome *c* (Sherman et al., 1974). Therefore, genetic selection can be applied to identify altered proteins ranging from nonfunctional to fully functional (Hampsey et al., 1986). In addition, a wealth of structural information is available for this family of proteins (Dickerson et al., 1971; Takano & Dickerson 1981a,b; Louie & Brayer, 1990). The proton nuclear magnetic resonance (NMR)[1] spectrum of the C102T protein variant of iso-1-cytochrome *c* has been nearly completely assigned in both oxidation states (Pielak et al., 1988a,b; Gao et al., 1990). The paramagnetic nature of oxidized cytochrome *c* allows the use of NMR to detect small conformational shifts including differences in hydrogen bonding that occur upon a change in oxidation state (Feng et al., 1990; Gao et al., 1991a; Pielak et al., 1991). The combination of well-defined genetics, a powerful selection strategy, and an array of structural, functional, and evolutionary data makes iso-1-cytochrome *c* an excellent system for the study of protein structure/function relationships.

Random mutagenesis coupled to genetic selection is a powerful tool for the study of protein structure/function relationships (Shortle, 1989; Shortle et al., 1985; Olson-Reidhaar et al., 1990). In the case of eukaryotic cytochrome *c*, the amino acid sequence of over 100 natural protein variants has been determined. This has revealed evolutionarily conserved and invariant residues [see Figure 3 of Cutler et al. (1987)]. However, the progressively selective nature of evolution and the inherent degeneracy of protein folding often make the existence of conserved or invariant residues enigmatic (Dill, 1990). Cytochrome *c* exhibits structural flexibility even at invariant residues. For instance, the evolutionarily invariant residue Phe-82 of iso-1-cytochrome *c* can be replaced with all 19 other naturally occurring amino acids without loss of function (Pielak et al., 1985; Hilgen & Pielak, 1991). The F82S mutation is accommodated by reorganization at both local and distant sites (Pielak et al., 1986; Louie et al., 1988; Gao et al., 1991b). The goal of our research is to understand the interaction of the N- and C-terminal helices at the molecular level. To begin, the ability of the N-terminal helix (Figure 1) to tolerate amino acid substitutions at the helical interface has been explored by random mutagenesis of the codons for Gly-6 and Phe-10.

2	3	4	5	6	7	8	9	10	11	12	13	14
N_{cap}	N1	N2	N3	N4	N5	mid	C5	C4	C3	C2	C1	C_{cap}
S	A	K	K	G	A	T	L	F	K	T	R	C

FIGURE 1: Sequence of the N-terminal helix.

Materials and Methods

Strains and Molecular Biology Techniques. All manipulations were carried out using a yeast shuttle phagemid constructed from the *CEN6*-based vector pRS315 (Sikorski & Hieter, 1989) by cloning the *CYC1* gene into the polylinker region. The phagemid contains the *LEU2* gene for selection in yeast, the Amp^R gene for selection in bacteria, and the f1 origin for the production of single-stranded DNA. The *CEN6* gene ensures mitotic stability and a copy number of 1–2 for the yeast-borne phagemid. This vector is used in bacteria to produce mutants and in yeast to analyze phenotype and express variant proteins.

Escherichia coli strain JM101 [*Δlacpro supE thi F'traD36 proAB lac*^{q} *ZΔM15* (Messing, 1983)] and the helper phage R408, a derivative of K07 (Vieira & Messing, 1987), were used for preparation of DNA for sequence analysis and cloning. Uracil-containing single-stranded template DNA for mutagenesis was made by using the *E. coli* strain RZ1032 [*HfrKL16 PO/45 Zbd-279::Tn10 lysA(61-62) thi1 relA1 supE44 dut1 ung1* (Kunkel et al., 1987)]. Bacterial transformations were carried out by the method of Miller et al. (1989) or by electroporation (Dower et al., 1988) using a home-built electroporator. DNA sequence analysis was performed by the method of Ner et al. (1988) using dideoxy sequencing (Sanger et al., 1977).

The yeast strain GM-3C-2 [*MAT α leu2-3 leu2-112 trp1-1 his4-519 cyc1-1 cyp3-1* (Faye et al., 1981)[2]] was used for both selection and expression. Yeast transformations were carried out by the lithium acetate procedure (Ausubel et al., 1988). Phagemid-bearing colonies were identified by their growth on petri plates containing complete media without leucine (Cm-Leu; Ausubel et al., 1988). Phagemids were isolated from yeast by using a modification of the shperoplast method (Rose, 1987), and electroporation was used to reintroduce phagemids into JM101.

Random Mutagenesis. A combination of a mutagenic oligodeoxyribonucleotide and uracil-containing single-stranded template DNA (Kunkel et al., 1987) was used to produce a library of mutations at the codons for Gly-6 and Phe-10. This primer, 5′-CATCTAGTCTT[G/A/C]NNAAGTGTAGC-[G/A/C/]NNTTTCTTAGCAG-3′ (where N = 25% of each base and G/A/C = 33% of each), is complementary to the coding strand of the *CYC1* gene. Thymine was left out at the 5′ end of each codon in the noncoding strand to eliminate the production of ochre and opal codons. This allows 48 possible codons that encode all 20 amino acids as well as the amber codon at each position.

Mutagenesis was carried out by the two-primer method (Norris et al., 1983). The second primer, 5′-CCTTCAGCTTGACC-3′, anneals 79 nucleotides upstream of the mutagenic oligodeoxyribonucleotide. Oligodeoxyribonucleotides were phosphorylated at their 5′ end by using the method of Zoller and Smith (1983). The reaction mixture for mutagenesis comprised single-stranded uracil-containing template DNA/mutagenic oligodeoxyribonucleotide/upstream

[1] Abbreviations: CM-Leu, complete media lacking leucine; NMR, nuclear magnetic resonance; YPG, rich media containing glycerol; YPL, rich media containing lactate; YPS, rich media containing sucrose.

[2] In the past, *CYC7* and *CYP3* have been used interchangeably to refer to the gene encoding iso-1-cytochrome *c* from *S. cerevisiae* (Verdiere & Petrochilo, 1979; Sherman et al., 1975).

primer in a ratio of 1:1:8. Extension was allowed to proceed for 1 h at 0 °C, followed by 1 h at ambient temperature, and then 2.5 h at 37 °C using T7 DNA polymerase, T4 gene 32 protein, and T4 DNA ligase.

Nomenclature. The term *mutant* is used to describe phagemids that contain mutations and to denote specific alleles. The C102T mutant of the wild-type gene encoding iso-1-cytochrome *c* 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 (S. Betz and G. J. Pielak, unpublished results) but does not have a measurable effect on function (Cutler et al., 1987; Pielak et al, 1988b).

Transformants harboring the *CYC1* gene grow on rich media containing 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: (1) Transformants possess the Cyc^+ phenotype if they grow on YPG and YPL at 30 and 37 °C, (2) the Cyc^{ts} phenotype if they grow on YPG and YPL at 30 °C but fail to grow on YPG and YPL at 37 °C, or (3) the Cyc^- phenotype if they fail to grow on YPG and YPL at 30 °C. Mutants isolated from Cyc^+ and Cyc^{ts} transformants are *functional mutants*, and the amino acids at positions 6 and 10 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 forms of 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., G6V;F10L).

Protein Purification. Protein variants were prepared from Cyc^+ and Cyc^{ts} transformants by a modification of the method of Sherman et al. (1968). A 5-mL overnight culture grown in YPG was inoculated into 0.5 L of YPG and grown to saturation at 30 °C. Protein variants were prepared from Cyc^- transformants by growth in YPS at 30 °C. After batch chromatography, the amount of iso-1-cytochrome *c* was determined by measuring the absorbance at 410 nm using an extinction coefficient of 106 100 M^{-1} cm^{-1} (Margoliash & Frohwirt, 1959). Depending on the mutant, the yield of cytochrome *c* from all transformants exhibiting the Cyc^+ or the Cyc^{ts} phenotype was between 10% and 80% of the yield obtained from yeast harboring the C102T mutant. The protein variants G6A;F10L and G6V were examined for temperature sensitivity in vitro by comparing their visible absorption spectra at 65 and 25 °C to those of the C102T protein obtained under identical conditions.

Results

Properties of the Library. A number of methods exist for the introduction of randomized codon windows into protein coding regions (Dunn et al., 1988; Mandecki, 1990). In the present study, it was desirable to use a method that would generate a large number of mutations at codons 6 and 10 simultaneously. Therefore, the oligodeoxyribonucleotide-directed strategy was chosen. A mixture of G/C/A was incorporated at each wobble base so that the number of codons for each amino acid is still large, but two-thirds of the stop codons are eliminated.

The first round of mutagenesis yielded a library of mutations at codons 6 and 10 with 85% efficiency. To increase the randomness of the library and the percentage of mutations, a second round of mutagenesis was performed using uracil-containing single-stranded template DNA from the first round. This was accomplished by transforming the reaction mixture from the first round into RZ1032 in liquid media and isolating single-stranded DNA. After the second round of mutagenesis, the library contained 93% mutants over the target codons and better approximated a random distribution of mutations.

Table I: Percentages of Nucleotides in the Coding Strand[a]

base	codon 6			codon 10		
C	8	12	17	7	13	37
T	28	30	55	53	43	36
G	45	37	28	25	22	27
A	19	21		15	22	

[a] The wild-type sequence is GGT for position 6 and TTC for position 10.

The percentage of nucleotides at each position in codons 6 and 10 was determined by DNA sequence analysis of 100 phagemids chosen at random from the library contained in *E. coli* (Table I). In addition to the 93 sequences with base pair changes, 1 insertion and 1 deletion were observed. There is an overrepresentation of the wild-type nucleotide at each position. Therefore, amino acids whose codons are close to wild type are overrepresented in the library. This bias is not surprising because oligodeoxyribonucleotides with the least number of mismatches will form the most stable annealed complexes. The lower representation of deoxycytosine is attributed, at least in part, to the lower amount of deoxyguanosine in the oligodeoxyribonucleotide (the noncoding strand) because of the lability of this nucleotide during phosphoramidite synthesis (Dunn et al., 1988). Approximately 80% of the alleles in the library are double missense mutants.

Determination of Phenotype and Relative Growth Rate. S. cerevisiae contains two isoforms of cytochrome *c*. In the strain GM-3C-2, *CYC1* is deleted, and the gene for iso-2-cytochrome *c* (*CYC7*) contains a double point mutation (Faye et al., 1981; Sherman et al., 1974). Therefore, this strain does not grow on nonfermentable carbon sources unless a functional cytochrome *c* is introduced. This potent genetic selection was exploited to identify novel functional mutants in the library without the need to make assumptions about the functionality of any particular mutation.

Of the 678 transformants that were examined for growth on YPG and YPL (all were tested on both YPG and YPL), 217 possess a functional cytochrome *c*, and 461 exhibit the Cyc^- phenotype. The phagemids from 82 of the 217 transformants were isolated and subjected to DNA sequence analysis. Among the 82 sequences, 36 are wild type, 8 are mutants that do not alter the deduced-wild-type amino acid sequence, and 38 are missense mutants. The 14 unique missense mutants among these 38 are listed in Table II. The 14 mutants were further divided into those that give rise to the Cyc^+ phenotype and those that give rise to the Cyc^{ts} phenotype (Table II). To confirm that the Cyc^{ts} phenotype is caused by the mutations found in the phagemids, each was retransformed into GM-3C-2. Fifty colonies from each transformation were again examined for temperature sensitivity. In every case, all transformants exhibit the Cyc^{ts} phenotype. Phagemids were also isolated from 28 Cyc^- transformants and subjected to DNA sequence analysis. All 28 sequences contain exclusively missense mutations at position 6 and/or 10, and each is unique (Table II). The sequence of the entire coding region was determined for the 14 unique functional mutants and for the nonfunctional mutants G6L, G6I;F10G, and G6N;F10K. No additional mutations were

Table II: Characteristics of *CYC1* Alleles with Mutations at Position 6 and/or 10

missense mutants at[a]				
position 6	position 10	positions 6 and 10		
partially TS[b]	normal[f]		TS[c]	
G6S (5^4)	F10Y (2^5)	G6V;F10Y (2^3)		G6S;F10Y (3^3)
TS[c]	F10W (2^4)	G6A;F10L ($2^{2.5}$)		
G6D (1^2)	F10L (2^5)		Nonfunctional[d,e]	
G6V (6^4)	F10I (3^5)	G6V;F10L	G6L;F10I	G6K;F10D
G6M (1^1)	F10V (3^4)	G6V;F10M	G6L;F10C	G6E;F10I
nonfunctional[d,e]	F10M (3^4)	G6V;F10C	G6L;F10L	G6N;F10K
G6L	F10C (3^4)	G6A;F10C	G6L;F10Y	G6D;F10G
G6F	nonfunctional[d,e]	G6A;F10I	G6Y;F10C	G6I;F10G
G6K	F10S	G6S;F10I	G6F;F10S	G6L;F10M
	F10R	G6W;F10T	G6T;F10C	
	F10G	G6W;F10G	G6P;F10S	

[a] The number in parentheses represents the number of times this combination was observed, and its superscript denotes the relative growth rate of the yeast harboring the particular mutant with respect to wild-type and nonfunctional controls on YPG and YPL at 30 °C (5 denotes the rate observed for yeast harboring the C102T mutant, and 0 would represent no growth after 4 days). [b] At 37 °C, this transformant grows more slowly than a transformant harboring the C102T mutant at 37 °C. [c] Growth observed at 30 °C but not at 37 °C. [d] No growth observed after 4 days at 30 °C. [e] All nonfunctional mutants are unique. [f] Growth observed at both 30 °C and 37 °C.

detected. The growth rate of the 14 transformants harboring the unique functional mutants was assessed relative to yeast harboring the *CYC1* gene. These relative growth rates were determined on both YPG and YPL. Although all strains tested grow more slowly on YPL than they do on YPG, the rank order of growth rates was always independent of the nonfermentable carbon source.

Statistical Analysis. The Monte-Carlo-based program written by Olson-Reidhaar et al. (1990) was used to determine the probability that all amino acids compatible with function were observed at a particular codon. This possibility is determined by three experimentally defined parameters: (1) the randomness of the codon (Table I); (2) the number of phagemids isolated from transformants harboring a functional cytochrome *c* and subjected to DNA sequence analysis (i.e., 82; vide ultra); and (3) the number of different amino acids observed at the codon as deduced from the DNA sequence (Table II). When position 6 is used as an example, 6 different amino acids (i.e., Gly, Ala, Ser, Asp, Val, and Met) are observed among the 82 sequences. However, Asn is not observed. The program was used to address the following question: If Asn is compatible with function, what is the probability that all 7 amino acids (the 6 in the list above, *plus* Asn) would have been observed among 82 sequences chosen at random from the library? A probability of 1 suggests that the number of transformants analyzed (in this case, 82) was sufficient to identify the *additional residue* (e.g., Asn) if it was compatible with function. In other words, a high probability suggests that the reason the additional amino acid was not detected is that this amino acid is incompatible with function. Six simulations were carried out for position 6, and five simulations were carried out for position 10 using different amino acids as the additional residue.[3] For both position 6 and position 10, the probability that all functional alleles were observed is greater than 0.7.

Substitutions at Gly-6. Gly-6 is at position N4 of the N-terminal helix (Figure 1) and is at the region of closest approach between the N- and C-terminal helices. The observation that five different amino acid substitutions at this position give rise to a functional cytochrome *c* (Table II) indicates that packing at this site is more flexible than previously thought (Takano & Dickerson, 1981a). The probability that all amino acids compatible with function have been observed at this position is 0.74. In other words, approximately three-quarters of the allowed substitutions have been detected. However, almost all functionally allowed substitutions give rise to the Cyc^{ts} phenotype (Table II). Changes in side-chain volume consistent with function are liberal with increases (relative to the wild-type amino acid Gly)[4] of up 100 Å^3 (e.g., G6M) tolerated.

The G6A and G6C mutants were observed among the 100 phagemids isolated from the library contained in *E. coli*, but were not detected among the 82 phagemids isolated from transformants harboring a functional cytochrome *c*. On the basis of the observation that G6A;F10L is compatible with function (Table II), it is expected that the G6A mutant is also compatible. In addition, Hampsey et al. (1986) found that the G6C mutant gives rise to a low level of growth on YPL. To test the prediction concerning G6A and to investigate the potential conflict concerning G6C, both mutants were separately transformed into yeast. Transformants containing the G6A mutant possess the Cyc$^+$ phenotype and grow at a rate indistinguishable from yeast harboring the C102T mutant. In contrast, transformants harboring the G6C mutant give rise to the Cyc^{ts} phenotype and exhibit a greatly reduced level of growth on both YPG and YPL, consistent with the observation of Hampsey et al. (1986).

Substitutions at Phe-10. Phe-10 is at position C4 of the N-terminal helix, and its side chain interacts with Tyr-97 of the C-terminal helix. Seven different amino acid substitutions at this position give rise to the Cyc$^+$ phenotype (Table II). The probability that all amino acids consistent with function have been observed for position 10 is 0.73, again indicating that nearly three-quarters of the substitutions consistent with function have been detected at this position. Changes in side-chain volume (relative to the wild-type amino acid Phe)[4] consistent with function are extremely broad, ranging from an increase of 38 Å^3 for tryptophan to a decrease of 80 Å^3 for cysteine. Substitutions that result in the abolition of function comprise two groups: those that further decrease the volume and basic residues.

Substitutions at Positions 6 and 10. Of the 14 unique functional alleles, only 3 are double missense mutants, and

[3] The additional amino acids consisted of three groups: those that are known to give rise to the Cyc$^-$ phenotype (Table II), those whose effect on phenotype is unknown and whose codons are present in the library at a low frequency, and those whose effect on phenotype is unknown but whose codons are well represented.

[4] The change in side-chain volume is defined as the volume of the side chain in the wild-type protein minus the volume of the side chain in the protein variant. Volumes are from Zamyatnin (1972).

yeast harboring these double mutants all possess the Cyc^{ts} phenotype. On the other hand, 22 out of 28 double missense mutants give rise to the Cyc$^-$ phenotype (Table II). Substitutions that by themselves are compatible with function often give rise to the nonfunctional phenotype when combined.

Discussion

A library of mutants has been made at codons 6 and 10 of the *CYC1* gene. Genetic selection was applied to identify substitutions compatible with function. A low copy number vector, the natural promoter, and the natural terminator were utilized to more closely approximate physiologically relevant conditions. The fact that only 7% of the library contains the wild-type codons at positions 6 and 10 yet over half of all transformants harboring a functional cytochrome *c* contain the wild-type amino acid sequence shows the power of genetic selection.

Sherman et al. (1974) showed that yeast containing as little as 10% of the wild-type amount of iso-1-cytochrome *c* exhibit normal growth on YPL. Although the G6L and G6L;F10I mutants give rise to the Cyc$^-$ phenotype, the yield of the protein variants obtained from the transformants harboring these mutants would have been sufficient to produce the Cyc$^+$ phenotype had these protein variants been functional. This observation shows that for these nonfunctional mutants, defects in the holoprotein cause the Cyc$^-$ phenotype. However, discussion of the other nonfunctional mutants in terms of effects on the holoprotein cannot be justified because their constituent mutations may act at one or more events in the "life cycle" of cytochrome *c*. These events include mRNA stability and/or legibility, translocation of the apoprotein across the mitochondrial outer membrane (Jordi et al., 1989; Nye & Scrapulla, 1990; Nicholoson et al., 1989), heme insertion, and turnover of the holoprotein (de Jongh & de Kruijff, 1990; Goldberg & Dice, 1974; Goldberg & St. John, 1976). If a decrease in mRNA legibility and/or stability was the cause of the Cyc$^-$ or Cyc^{ts} phenotype, mutants that do not change the amino acid sequence might be expected to be found among phagemids isolated from yeast exhibiting these phenotypes. However, all transformants harboring such mutants exhibit the Cyc$^+$ phenotype, and phagemids isolated from transformants exhibiting the Cyc^{ts} and Cyc$^-$ phenotypes bear exclusively missense mutations. The observations tend to rule out effects on mRNA as the cause of the Cyc^{ts} and Cyc$^-$ phenotypes. Ferrocytochrome *c* from Cyc^{ts} transformants harboring the G6A;F10L and G6V mutants exhibits an altered electronic absorption spectrum at 65 °C compared to the C102T protein variant under identical conditions, showing that the holoproteins themselves are affected. In summary, for the G6L, G6S;F10I, G6V, and G6A;F10L mutants, no matter where else in the life cycle the mutation acts, they affect the holoprotein.

Although double missense mutants comprise approximately 80% of the library, only 3 are observed among the 14 unique functional mutants (Table II). Furthermore, some combinations of single missense mutants have additive or synergistic effects on phenotype. For example, transformants containing the G6A, F10L, or F10I mutants possess the Cyc$^+$ phenotype, but the double mutant G6A;F10L give rise to the Cyc^{ts} phenotype, and G6A;F10I results in the Cyc$^-$ phenotype. Such effects have been noted for other proteins (Alber, 1989; Wells, 1990).

How many different combinations of amino acids at positions 6 and 10 are compatible with function? From inspection of the evolutionary record, it appears that only one combination has been maintained. However, Hampsey et al. (1986) have pointed out that evolutionary invariance does not necessarily mean functional invariance. From the data in Table II, the minimum number of different combinations is 14. Other functional combinations probably exist among the transformants that harbor a functional cytochrome *c* but whose phagemids were not subjected to DNA sequence analysis (e.g., G6A and G6C, vide ultra). When the above observations are combined with the volume constraints at position 6 and the observation that certain amino acids and combinations of amino acids result in nonfunctional mutants (Table II), we estimate that a maximum of approximately 40 different combinations give rise to a functional cytochrome *c*.

An independent method for estimating the number of functional combinations is to use a simulation program based on Monte-Carlo analysis (Olson-Reidhaar et al., 1990). These simulations indicate that the probability that all amino acids compatible with function have been identified at positions 6 alone and at position 10 alone is greater than 0.7. Therefore, the probability that all possible functional missense mutants at positions 6 and 10 have been observed is approximately 0.5 (0.7^2). This probability, taken together with the fact that 17 unique combinations of amino acids at positions 6 and 10 are compatible with function (i.e., the wild-type protein, the 14 in Table II, G6A, and G6C), suggests that a maximum of approximately 34 combinations give rise to a functional cytochrome *c*. This number is consistent with that estimated above from examination of the functional and nonfunctional mutants. However, only 9 different combinations give rise to the Cyc$^+$ phenotype (i.e., the wild-type protein, those under the heading "normal" in Table II, and G6A). From this observation, and the statistical argument above, it is concluded that a maximum of approximately 18 different combinations result in the Cyc$^+$ phenotype. In summary, it is estimated that 4–10% of the 400 possible combinations of amino acids at positions 6 and 10 give rise to a functional cytochrome *c* but only approximately half of these result in the Cyc$^+$ phenotype.

Three interrelated explanations are presented to explain why positions 6 and 10 so greatly restrict the polypeptide chain. First, to interact, the stability of the helices must be preserved. Second, the residues are buried in the hydrophobic core of the protein. Third, the orientation of the helices must be preserved because these residues are intimately involved in the earliest detectable folding event (Roder et al., 1988). These explanations are discussed below.

There have been many studies of residues that stabilize isolated α-helices (Pandmanabhan et al., 1990; Degrado & O'Neil, 1990; Lyu et al., 1990). The amino acids substitutions observed in this study are not expected to disrupt the N-terminal α-helix (Richardson & Richardson, 1988b). Furthermore, the invariance to Gly-6 cannot be rationalized from the standpoint of α-helical stabilization because any other amino acid (except proline) at this position will yield a more stable helix. Takano and Dickerson (1981a) have proposed that Gly-6 is invariant because there is no room for a side chain at this position due to the shimming of Leu-94 into this site. Unfavorable interactions with the C-terminal α-helix doubtlessly occur in many of the position 6 protein variants, and these interactions probably explain the Cyc^{ts} phenotype. Stable helical packing is probably produced by satisfying constraints other than those that define helices.

The side chains of Gly-6 and Phe-10 are buried within the hydrophobic core of the protein (Louie & Brayer, 1990). However, one charged amino acid (G6D) and one polar amino acid (G6A) are compatible with function. Although the ionization state of Asp-6 is unknown, stabilization of a negative

charge is possible via interaction with the helix dipole (Baldwin et al., 1989) or with the large number of basic residues found in this region. Increased solvation of the Asp-6 protein variant in its unfolded form could explain the Cyc^{ts} phenotype observed for transformants harboring this mutant. Preservation of hydrophobicity is the only invariant property of functional mutants at position 10. Recently, tyrosine at position 10 has been observed in the cytochrome *c* from the nematode *Caenorhabditis elegans* (Vanfleteren et al., 1990). In addition, two mutants, F10C and F10M, possess sulfur-containing amino acids. Aromatic–aromatic and aromatic–sulfur interactions have been shown to stabilize proteins (Burley & Petsko, 1988; Serrano et al., 1991).

Conformational shifts in interacting helices can allow plasticity at interhelical interfaces. In their study of insulin, Chothia et al. (1983) noted that helical movements of 1–2 Å occur by side-chain rotations and that changes in the helix–helix axis torsion angle occur on a scale of ±15°. Conservative amino acid substitutions such as G6A, F10Y, and F10L are probably accommodated by side-chain rearrangements and small changes in the helix–helix torsion angle. However, for mutants such as F10C and G6M, larger rearrangements must occur. In a comparative study involving cytochrome *c* and cytochrome c_{551} Chothia and Lesk (1985) showed that deformations within the N- and C-terminal helices occur so as to accept amino acid substitutions, but these deformations result in alteration of the heme pocket. Therefore, some of the functionally acceptable substitutions listed in Table II may result in conformational shifts at sites remote from the helical interface. Biophysical studies of the protein variants isolated from the transformants listed in Table II as well as a complementary study of the interfacial residues in the C-terminal helix are underway.

Acknowledgments

We thank J. Beasley, S. Betz, R. Bruik, H. Chui, D. Doyle, Z. Fredericks, L. L. Garcia, R. Green, S. Hilgen, and X. Wang for advice and encouragement, John Reidhaar-Olson for providing us with the simulation program, M. Smith for providing PING4, and the staff of the instrument and electronic shops in the department for building much of the equipment used in this study.

Registry No. Gly, 56-40-6; Phe, 63-91-2; cytochrome *c*, 9007-43-6.

References

Alber, T. (1989) *Annu. Rev. Biochem. 58*, 765–798.

Ausubel, F., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., & Struhl, K. (1988) in *Current Protocols in Molecular Biology*, Vol. 2, John Wiley & Sons, New York.

Baldwin, R. L., Stewart, J. M., York, E. J., Shoemaker, K. R., York, E. J., & Fairman, R. (1989) *Proteins: Struct., Funct., Genet. 5*, 1–7.

Burley, S. K., & Petsko, G. (1988) *Adv. Protein Chem. 39*, 125–189.

Chothia, C., & Lesk, A. M. (1985) *J. Mol. Biol. 182*, 151–158.

Chothia, C., Lesk, A. M., Dodson, G. G., & Hodgkin, D. C. (1983) *Nature (London) 302*, 500–505.

Cutler, R. L., Pielak, G. J., Mauk, A. G., & Smith, M. (1987) *Protein Eng. 1*, 95–99.

Degrado, W. F., & O'Neil K. T. (1990) *Science 250*, 646–651.

de Jongh, H. H. J., & de Kruijff, B. (1990) *Biochim. Biophys. Acta 1029*, 105–112.

Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Copper, A., & Margoliash, E. (1971) *J. Biol. Chem. 246*, 1511–1535.

Dill, K. A. (1990) *Biochemistry 29*, 7133–7155.

Dower, W. J., Miller, J. G., & Ragsdale, C. W. (1988) *Nucleic Acids Res. 16*, 6127–6145.

Dunn, I. S., Cowan, R., & Jennings, P. A. (1988) *Protein Eng. 2*, 283–291.

Efimov, A. V. (1984) *FEBS Lett. 166*, 33–38.

Faye, G., Leung, D. W., Tatchell, K., Hall, B. D., & Smith, M. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2258–2262.

Feng, Y., Roder, H., & Englander, S. W. (1990) *Biochemistry 29*, 3494–3504.

Gao, Y., Boyd, J., Williams, R. J. P., & Pielak, G. J. (1990) *Biochemistry 29*, 6944–7003.

Gao, Y., Boyd, J., Pielak, G. J., & Williams, R. J. P. (1991a) *Biochemistry 30*, 1928–1934.

Gao, Y., Boyd, J., Pielak, G. J., & Williams, R. J. P. (1991b) *Biochemistry 30*, 7033–7040.

Goldberg, A. L., & Dice, J. F. (1974) *Annu. Rev. Biochem. 43*, 835–869.

Goldberg, A. L., & St. John, A. C. (1976) *Annu. Rev. Biochem. 45*, 747–803.

Hampsey, D. M., Das, G., & Sherman, F. (1986) *J. Biol. Chem. 261*, 3259–3271.

Hilgen, S. E., & Pielak, G. J. (1991) *Protein Eng. 4*, 575–578.

Jordi, W., Li-Xin, Z., Pilon, M., Demel, R. A., & de Kruijff, B. (1989) *J. Biol. Chem. 264*, 2292–2301.

Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367–382.

Louie, G. V., & Brayer, G. D. (1990) *J. Mol. Biol. 214*, 527–555.

Louie, G. V., Pielak, G. J., Smith, M., & Brayer, G. D. (1988) *Biochemistry 27*, 7870–7876.

Lyu, C. P., Liff, M. I., Marky, L. A., & Neville, K. R. (1990) *Science 250*, 669–672.

Mandecki, W. (1990) *Protein Eng. 3*, 221–226.

Margoliash, E., & Frohwirt, N. (1959) *Biochem. J. 71*, 570–572.

Matthews, F. S. (1985) *Prog. Biophys. Mol. Biol. 45*, 1–56.

Messing, J. (1983) *Methods Enzymol. 101*, 20–78.

Miller, R. H., Niemela, S. L., & Chung, C. T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2172–2175.

Ner, S. S., Goodin, D. B., Pielak, G. J., & Smith, M. A. (1988) *BioTechniques 6*, 408–412.

Nicholoson, D. W., Stuart, R. A., & Neupert, W. (1989) *J. Biol. Chem. 264*, 10156–10168.

Norris, K., Norris, F., Christiansen, L., & Fill, N. (1983) *Nucleic Acids Res. 11*, 5103–5112.

Nye, S. H., & Scarpulla, R. C. (1990) *Mol. Cell. Biol. 10*, 5763–5771.

Olson-Reidhaar, J. F., & Sauer, R. T. (1990) *Proteins: Struct., Funct., Genet. 7*, 306–316.

Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., & Baldwin, R. L. (1990) *Nature 344*, 268–270.

Pielak, G. J., Mauk, G. A., & Smith, M. (1985) *Nature 313*, 152–154.

Pielak, G. J., Kimio, O., Mauk, G. A., Smith, M., & Cyril, M. K. (1986) *J. Am. Chem. Soc. 108*, 2724–2727.

Pielak, G. J., Atkinson, R. A., Boyd, J., & Williams, R. J. P. (1988a) *Eur. J. Biochem. 177*, 179–185.

Pielak, G. J., Boyd, J., Moore, G. R., & Williams, R. J. P. (1988b) *Eur. J. Biochem. 177*, 167–177.

Pielak, G. J., Auld, D. S., Betz, S. F., Hilgen, S. E., & Garcia, L. L. (1991) in *The Cytochrome c Sourcebook* (Mauk, A. G., & Scott, R. A., Eds.) University Science Books, Mill Valley, CA (in press).

Richardson, J. S., & Richardson, D. C. (1988a) *Proteins: Struct., Funct., Genet. 4*, 229–239.
Richardson, J. S., & Richardson, D. C. (1988b) in *Prediction of protein structure and the principles of protein conformation* (Fasman, Ed.) G. D., Plenum, New York.
Roder, H., Elöve, G. A., & Englander, W. (1988) *Nature 335*, 700–704.
Rose, M. D. (1987) *Methods Enzymol. 152*, 481–507.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
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., Gildord, R. A., & Parker, J. H. (1974) *Genetics 77*, 255–284.
Sherman, F., Jackson, M., Liebman, S. M., Schweingruber, A. M., & Stewart, J. W. A. (1975) *Genetics 81*, 51–73.
Shortle, D. (1989) *J. Biol. Chem. 264*, 5315–5318.
Shortle, D., & Lin, B. (1985) *Genetics 110*, 539–555.
Sikorski, R. S., & Hieter, P. (1989) *Genetics 122*, 19–27.
Takano, T., & Dickerson, R. E. (1981a) *J. Mol. Biol. 153*, 79–94.
Takano, T., & Dickerson, R. E. (1981b) *J. Mol. Biol. 153*, 94–115.
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) *Genetics 175*, 209–216.
Vieira, J., & Messing, J. (1987) *Methods Enzymol. 153*, 3–11.
Wells, J. A. (1990) *Biochemistry 29*, 8509–8517.
Zamyatnin, A. A. (1972) *Prog. Biophys. Mol. Biol. 24*, 107–123.
Zoller, M. J., & Smith, M. (1983) *Methods Enzymol. 100*, 468–500.

Quantitating and Engineering the Ion Specificity of an EF-Hand-like Ca^{2+} Binding Site†

Joseph J. Falke,* Eric E. Snyder, Kay C. Thatcher, and C. Stefan Voertler
Department of Chemistry and Biochemistry, University of Colorado, Boulder, Colorado 80309-0215
Received March 28, 1991; Revised Manuscript Received June 13, 1991

ABSTRACT: The *Escherichia coli* D-galactose and D-glucose receptor, an aqueous periplasmic receptor that triggers sugar sensing and transport, possesses a single Ca^{2+} binding site similar in structure and specificity to the EF-hand class of sites found in eukaryotic Ca^{2+} signaling proteins including calmodulin and its homologues. A universal feature of these sites is the use of a pentagonal bipyramidal array of seven oxygens to coordinate bound Ca^{2+}. Here we investigate the mechanisms used by this coordinating array to control ion specificity. To vary the cavity size and charge of the array, we have replaced axial glutamine 142 in the prokaryotic site with asparagine, glutamate, and aspartate. The ion selectivities of the resulting engineered sites have been quantitated by measuring dissociation constants for a series of spherical metal ions, differing in increments of radius and charge, from groups Ia, IIa, and IIIa and the lanthanides. Dramatic specificity changes are observed: sites containing an engineered smaller side chain (Asn or Asp) bind the largest cations up to 50-fold more tightly than the native site; and sites containing an engineered negative side chain (Glu or Asp) exhibit preferences for trivalent over divalent cations up to 1900-fold higher than the native site. The results indicate that the cavity size and negative charge of the coordination array play key roles in selective Ca^{2+} binding and that the array can be engineered to preferentially bind other cations.

Protein Ca^{2+} binding sites selectively bind Ca^{2+} even in the presence of up to 10^5-fold higher concentrations of Na^+, K^+, and Mg^{2+}. This specificity is particularly important for regulatory sites in Ca^{2+} signaling proteins, which must remain unoccupied until a Ca^{2+} signal appears (Persechini et al., 1989), and for Ca^{2+} channel sites, which provide selective Ca^{2+} fluxes (Tsien et al., 1987; Alsobrook, 1988; Catteral et al., 1990). We are using protein engineering to probe the molecular basis of Ca^{2+} selectivity. Ca^{2+} sites are well suited for the protein engineering approach for two reasons. First, they exhibit substantial heterogeneity in primary structure and are thus likely to tolerate engineered sequence changes, and, second, the spherical ion is the simplest type of substrate. Such a substrate requires optimization of relatively few geometrical variables, and its electrostatic field is likely to impose optimal ordering of the coordinating side chains. In a favorable system such as this, it may be possible to apply rational design methods to generate predictable changes in substrate selectivity.

The D-galactose and D-glucose receptor of *Escherichia coli* possesses a single Ca^{2+} site that shares key structural elements with the important EF-hand class of Ca^{2+} sites, the latter commonly found in eukaryotic Ca^{2+} signaling proteins [see previous structural comparisons by Vyas et al. (1987) and Snyder et al. (1990)]. The structure of the receptor, determined to 1.9-Å resolution by Quiocho and co-workers (Vyas et al., 1987), reveals a coordinating array of seven oxygens arranged in a pentagonal bipyramid geometry around the bound Ca^{2+}, as typically observed in EF-hand sites (Strynadka & James, 1989). In the prokaryotic site the coordinating array, illustrated in Figure 1A, consists of four coordinating oxygens provided by three side-chain carboxylates (mono-

† Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, and to the National Institutes of Health for support of this work.

0006-2960/91/0430-8690$02.50/0