Loop Replacement and Random Mutagenesis of \( \Omega \)-Loop D, Residues 70–84, in Iso-1-cytochrome c*

(Published for publication, October 10, 1995, and in revised form, January 5, 1996)

Peggy Mulligan-Pullyblank, Jennifer S. Spitzer, Brad M. Gilden, and Jacquelyn S. Fetrow§

From the Department of Biological Sciences, Center for Molecular Genetics, University at Albany, SUNY
Albany, New York 12222

To study the role of omega loop D, residues 70–84, in the structure and function of yeast iso-1-cytochrome c, this loop was replaced with homologous and heterologous loops. A novel method was developed for rapid insertion of these mutations into the yeast chromosome at the CYC1 locus. The strains containing these loop replacement cytochromes cannot grow on nonfermentable carbon sources, indicating that the proteins are nonfunctional. Whole cell difference spectroscopy shows that no holocytochrome c is present; however, apocytochrome is found by immunoblot analysis. Thus, apocytochrome is present in these mutant strains, but it cannot bind heme and cannot compete with wild type apocytochrome conversion to holocytochrome. This is a unique example of a set of loop replacements that do not produce folded protein, and these results suggest that the loop D amino acid sequence in iso-1-cytochrome c plays a significant role in cytochrome c biosynthesis in vivo. To identify the significant amino acids in loop D, random mutagenesis of six highly conserved loop residues, Tyr-74, Ile-75, Pro-76, Gly-77, Thr-78, and Lys-79, was accomplished. Sequencing of the random mutants shows that strict conservation of none of these residues is required to produce a minimally functional cytochrome c. Preferences are found for small, hydrophilic or aromatic residues at position 74, hydrophobic residues at position 75, glycine and arginine at positions 76 and 77, and \( \beta \)-branched amino acids at position 78. Implications for the role of loop D in the structure and function of iso-1-cytochrome c are discussed.

The protein folding problem is an important, unsolved problem in molecular biology. An understanding of this problem is important to utilization of the data being generated by the human genome project and to further the design and engineering of proteins with specific structures and functions. Loops and turns, the connections between the regular secondary structures, are critical to the design of properly folded proteins. These structures are difficult to characterize because of their range of conformations. The short \( \beta \)-turns and reverse turns were the first to be characterized (1, 2); then the longer \( \Omega \)-loops were described (3); and finally a more comprehensive definition of loops was published (Ref. 4; for review, see Ref. 5). We are interested in defining the role that loops, particularly \( \Omega \)-loops, play in protein folding, function, and stability.

\( \Omega \)-Loops are nonregular protein secondary structures defined as a sequence of six or more residues with a short distance between segment termini and a lack of regular hydrogen bonding between backbone residues (3). However, significant hydrogen bonding and hydrophobic interactions can exist between side chain and backbone atoms within the loop itself.²

\( \Omega \)-Loops are generally found at protein surfaces (3). Because they are positioned at protein surfaces, changes in the loop amino acid sequence, as well as changes in loop length, are generally less disruptive to the rest of the protein than changes made in residues in the protein core. Indeed, previous studies have found that large changes within loop regions rarely affect the ability of the protein to fold, although the function of the protein may be disrupted if the loop plays an important role in protein function (for review, see Ref. 5). Thus, loop replacement or loop swap experiments are good methods with which to analyze the role that a loop plays in protein function; furthermore, the importance of loops for protein folding and stability can be determined by spectroscopic and crystallographic studies of loop replacement proteins (6, 7).

The significance of \( \Omega \)-loop D, residues 70–84, to the structure and function of iso-1-cytochrome c of Saccharomyces cerevisiae is investigated here. Cytochrome c is a small protein found in the electron transport system of most organisms. Iso-1-cytochrome c shuttles electrons from cytochrome c reductase to cytochrome c oxidase within the inter-mitochondrial space (8, 9). It also interacts with cytochrome c peroxidase, cytochrome b₅, and sulfite oxidase. Despite sequence variation, a general tertiary fold is found in c-type cytochromes from prokaryotic and eukaryotic organisms (8, 10). The overall tertiary fold consists of three \( \alpha \)-helices, one at the amino terminus, one at the carboxyl terminus, and one near residues 60–64, and three or four \( \Omega \)-loops, surrounding a heme group that is covalently bound to the protein by two thioether linkages to cysteine residues (8, 10). The positions of the \( \Omega \)-loops are conserved, although the loops are the major sites of insertions and deletions between family members from different species (11). This general tertiary structure is depicted in Fig. 1.

The advantage of studying mutations in S. cerevisiae cytochrome c is that mutants that affect the protein structure and those that affect the function can be distinguished in vivo. Because S. cerevisiae requires a functional cytochrome c to grow on nonfermentable carbon sources, function can be assayed by growth of mutant yeast strains on lactate, glycerol, or ethanol media (12). The presence of holoprotein, cytochrome c to which the heme group is covalently attached, can be detected by whole cell difference spectroscopy following growth of cells in dextrose medium, regardless of the functionality of the pro-

---

* This work was supported National Institutes of Health Grant GM44829 (to J. S. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Present address: Dept. of Medicine, Division of Digestive Diseases, Emory University School of Medicine, Atlanta, GA 30322.

‡ To whom correspondence should be addressed: Dept. of Biological Sciences, University at Albany, SUNY, 1400 Washington Ave., Albany, NY 12222. Tel.: 518-442-4389; Fax: 518-442-4767; E-mail: jacque@isadora.bio.albany.edu.

¹ B.A. Kell and J. S. Fetrow, unpublished results.
tein (12, 13). A lack of holoprotein would indicate that a mutation has altered the fold or stability of apo- or holoprotein so that it is more readily proteolyzed or that it has disrupted a step in the biosynthetic processing of holocytchrome c, such as transcription in the nucleus, translation on free ribosomes, mitochondrial recognition and import, or recognition by cytochrome c heme lyase, the protein that covalently attaches heme to the apoprotein (14, 15).

Previous loop deletion and replacement experiments have been done on Ω-loops in yeast iso-1-cytochrome c. That work established that deletion of a portion of loop A, residues 21–28, or loop D, residues 74–78, resulted in a complete deficiency of holocytochrome c protein; however, holoprotein was still produced when portions of two other loops, loops B and C, were deleted (11). Replacement mutations at the loop A position yielded results similar to the loop replacement experiment results described for other proteins: the mutant proteins are altered in stability or function yet are still able to fold into a native-like holoprotein. In each loop A replacement mutant, a heme-bound native-like cytochrome molecule is found in the cell, although the protein is not always functional (6, 11).

In the current study, loop D was the target of loop replacement experiments so that the role of this loop in protein function and structure could be studied. Here we describe a novel, rapid method for insertion of these mutations into the yeast genome at the wild type iso-1-cytochrome c (CYC1) locus. This protocol was developed so that the mutated cytochromes c are under the same regulatory controls as wild type cytochrome c, and only one copy of the mutant gene exists in the haploid yeast cell. Amino acids at the loop D position in yeast iso-1-cytochrome c were replaced by homologous loop residues from prokaryotic c-type cytochromes and by heterologous loops from other proteins. The ability of the mutant proteins to fold and function was tested in vivo. An in vivo competition study was designed to determine whether the mutated apoproteins could compete with wild type holoprotein production. Because none of the loop replacement mutations produced a folded protein, directed, random mutagenesis was applied to six highly conserved residues within loop D: Tyr-74, Ile-75, Pro-76, Gly-77, Thr-78, and Lys-79. (Five of these six residues are those that serve residues within loop D: Tyr-74, Ile-75, Pro-76, Gly-77, Thr-78, and Lys-79. Five of these six residues are those that serve residues within loop D: Tyr-74, Ile-75, Pro-76, Gly-77, Thr-78, and Lys-79. Five of these six residues are those that serve residues within loop D: Tyr-74, Ile-75, Pro-76, Gly-77, Thr-78, and Lys-79. Five of these six residues are those that serve residues within loop D: Tyr-74, Ile-75, Pro-76, Gly-77, Thr-78, and Lys-79.)

EXPERIMENTAL PROCEDURES

Strain Designations—The symbol CYC1 represents the wild type gene encoding iso-1-cytochrome c; CYC7 represents the locus encoding iso-2-cytochrome c. When compared with sequences of other c-type cytochromes, iso-1-cytochrome c has a five-residue extension at the amino terminus. In this paper, we follow the tradition of numbering iso-1-cytochrome c as it is aligned with tuna cytochrome c; thus, numbering begins at methionine residue 5 and ends with glutamic acid 103. Loop D is the Ω-loop secondary structure, residues 70–84, in yeast iso-1-cytochrome c. In this study, loop D was replaced by structurally homologous loops from other c-type cytochromes and by heterologous loops from other proteins; the mutant proteins are designated RepD1–D4, where Rep symbolizes a loop replacement mutant, D denotes loop D, and the number represents the individual replacements designated in Table I. Yeast strains generated by standard mutagenesis procedures (see below) are designated by a C followed by the strain number; thus C6 is the yeast strain encoding the wild type cytochrome c, and C93 is the yeast strain encoding wild type cytochrome c containing a cysteine to threonine mutation at residue 102 (C102T). Yeast strains generated by the directed, random mutagenesis protocol are designated by the oligonucleotide number (3R or 4R for Pro-76 and Gly-77, 8R for Thr-78 and Lys-79, or 9R for Tyr-74 and Ile-75) followed by the strain number.

The genotype of the strains used in this study are as follows: Escherichia coli HB101: supE44, hsdS20 (rBmB), rpeA13, ara-14, proA2, lacY1, galK2, rpsL2, xyl-5, met-l (15); E. coli CJ236 CM4, dcm, ung, (17); S. cerevisiae strain B7864: MATa, cyc1-814, ura3-52, his3-11, leu2-3,112, trp1-289, can1-100, cyh2 (obtained from F. Sherman); B2111: MATa, cyc1-239, arg4-77, cyc7 (12) (obtained from F. Sherman); eG2: MATa, trp1, leu2-3, leu2-112, gal1α, cyc::LEU2, his3 (obtained from R. Zitomer); C6 (wild type): MATa, CYC1, LEU2, cyc7-68:cyh; his3-11, ura3-52, leu2-3,112, trp1-289, can1-100, cyh2 (obtained from F. Sherman); eG2: MATa, trp1, leu2-3, leu2-112, gal1α, cyc::LEU2, his3 (obtained from R. Zitomer); C6 (wild type): MATa, CYC1, LEU2, cyc7-68:cyh; his3-11, ura3-52, leu2-3,112, trp1-289, can1-100, cyh2 (obtained from F. Sherman); eG2: MATa, trp1, leu2-3, leu2-112, gal1α, cyc::LEU2, his3 (obtained from R. Zitomer); C6 (wild type): MATa, CYC1, LEU2, cyc7-68:cyh; his3-11, ura3-52, leu2-3,112, trp1-289, can1-100, cyh2. Strain 626 is isogenic with C6 except that it contains the C102T mutation in the CYC1 gene, as described above.

Plasmids Utilized in This Study—The E. coli plasmid, pYC17, containing a deletion of residues Tyr-74 through Thr-78 in loop D and the C102T mutation in iso-1-cytochrome c, was used for site-directed mutagenesis. This plasmid was created from the previously described parent plasmid pAB543 (containing the loop D deletion, residues 74–78), which was created from the wild type plasmid pAB458 (11). It contains an ampicillin resistance marker for selection in E. coli and an f1(IG) origin of replication so that single-stranded DNA can be produced after infection by R408 phage (18).

The incorporation of cytochromes c in the CYC1 locus of the yeast genome, the yeast integration plasmid pYc30 was constructed from plasmid pAB571 (kindly provided by T. Cardillo). pYc30 contains an origin of replication and an ampicillin resistance gene for maintenance and selection in E. coli, a LEU2 marker for selection in yeast, and an f1(IG) origin of replication for single-stranded DNA production after R408 infection. It also contains the 2,530-bp BamHI-EcoRI chromosomal fragment containing the wild type CYC1 gene, although the 3′ EcoRI site was mutated to a BamHI site so that the plasmid could be easily digested with BamHI before integration into the yeast chromosome. The 2.2-kilobase pair Sall-Xhol LEU2 fragment was inserted at the BgII site 515 bp from the start of the BamHI-EcoRI CYC1 fragment. The CYC1 fragment is flanked by Miul sites 262 bp upstream and 198 bp downstream of the beginning and end of the CYC1 locus, respec-

The experiment was performed with the strain C173 cells, a duf. ung strain that utilizes uracil in place of thymi-

During DNA replication (19). Uracil-incorporated single-stranded DNA was isolated by standard protocols (18) 4 hours after infection of CJ236 cells with plasmid pY14. Each of the designed oligonucleotides was annealed to the single-stranded pY14 DNA by heating to 85°C and slowly cooling to 29°C. This annealing mix was combined with an extension/ligation mixing buffer containing 0.1 mM Tris-HCl, pH 8.0, 50 mM MgCl2, 5 mM ATP, 10 mM dithiobiotin, dNTP mix (2.5 mM each dNTP), 50 ng/ml T4 DNA polymerase, and T4 DNA ligase. Chemical suppliers were New England Biolabs, Boehringer Mannheim, and U. S. Biochemical Corp. The reaction was incubated sequentially on ice for 5 min, room temperature for 5 min, and then 37°C for 60 min, at which time a standard E. coli transformation was performed (20) using competent HB101 cells. Cells were plated on YT + ampicillin medium (50 mg/ml ampicillin, 0.5% yeast extract, 0.5% NaCl, and 50 mg/ml ampicillin).

Colonies were screened by purification of the plasmid DNA using the alkaline lysis plasmid preparation (20) and then digested with selected restriction enzymes. Fragments were separated and analyzed by agarose gel electrophoresis.

The abbreviations used are: bp, base pair; SD, synthetic dextrose; PCR, polymerase chain reaction; KU, Klett units; Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.
Mutations of Loop D in Cytochrome c

8635

ose or polyacrylamide gel electrophoresis. Correct insertions resulted in an increase in the size of an AvalI band or a creation of a new restriction site. Positive candidates were confirmed by deoxy sequencing (U. S. Biochemical Corp. Sequenase kit) of NaNH-denatured plasmid DNA prepared by large scale plasmid preparation (Promega).

I Incorporation of Mutant cycl1 Sequences into Yeast CYC1 Locus—To study the effect of the and on cytochrome c gene expression under the same regulatory conditions as wild type CYC1, a fast, simple method to integrate and screen for the mutant gene at the yeast CYC1 locus was developed (Fig. 2). First, mutant cycl1 sequences were cloned into a yeast integration plasmid, pYC31, which contains BamHI-EcoRI fragment encompassing the CYC1 locus and the LEU2 marker inserted into the BglI site of this fragment as described above. The MluI fragment of the plasmid containing the mutant cycl1 sequence was excised from FMC low melting agarose and ligated into pYC31, which had previously been linearized with MluI and dephosphorylated to prevent intramolecular re-ligation.

Yeast cells of strain B7684, in which the CYC1 gene has been replaced with URA3 gene, were made competent by lithium acetate protocol (21) and transformed with pYC31-derived plasmids linearized with BamHI. Transformants were plated on synthetic dextrose (SD) medium (0.67% yeast nitrogen base, without amino acids and ammonium sulfate; 2% dextrose; 2.2% ammonium sulfate) without leucine SD–(leu) to select for integration of the LEU2 marker through homologous recombination. Selected transformants were then screened on SD medium lacking uracil SD–(ura) for removal of the cycl1 fragment by homologous recombination. Several yeast colonies that lost SD–(leu) but not SD–(ura) were selected, subcloned on YPD medium (1% yeast extract, 2% peptone, 2% dextrose), restreaked on SD–(ura), SD–(leu), SD–(trp), and SD–(his) to check strain phenotype. The cycl1 locus of one or two of the colonies that tested correctly was sequenced.

Genetic DNA Sequencing—Yeast genomic DNA from haploid strains was isolated by a glass beads protocol (21) and a first PCR was performed to amplify the CYC1 region of the yeast genomic DNA using the Promega PCR sequencing kit. The first PCR product was purified from nucleotides and primers with Millipore Ultrafree-MC low-binding cellulose filter and protocols supplied by the manufacturer. The purified PCR product was used as the DNA template in a second PCR amplification with Applied Biosystems Taq deoxy terminator cycle sequencing kit. Sequencing reactions were run on an Applied Biosystems DNA Sequencer. For each mutant yeast strain, the entire length of the CYC1 locus was sequenced at least twice.

In Vivo Functional Analysis—Growth curves were done by modification of a previously published procedure (11). The haploid yeast strains produced by the mutagenesis and integration procedures described above were crossed with B2111 and screened for their ability to grow on SD medium. Studying the loop D replacement mutations in diploid strains should eliminate the possibility of interference by background mutations and provide a system in which any effects on growth of the yeast strain can be directly accounted for by the changes made in the cytochrome c gene (12).

An overnight culture was made by inoculating a single colony of the diploid cytochrome c mutant in YPD medium and growing overnight at 30°C. After growth for 16 h, the cultures were synchronized by placing them at 4°C overnight. Cells were centrifuged and rinsed with liquid lactate medium. Ten ml of liquid lactate medium in side arm flasks (0.67% yeast nitrogen base, without amino acids and ammonium sulfate; 1% lactate; 2% ammonium sulfate) was inoculated with a quantity depending on the experiment, and cell density was measured at 3–4-h intervals over a 25–30-h period. Yeast strains C6 (wild type), C93 (wild type, C102T) were used as positive controls, and C15 (del cycl1) was studied as a negative control representing no cytochrome c.

Whole Cell Protein Extracts and Immunoblotting—Immunoblotting was done to determine whether apoprotein was present in the loop D deletion and replacement mutants. Haploid yeast strains were grown for 15–16 h in YPD medium and then diluted by one-third with YPD medium and grown for approximately 2 h. Total yeast protein extracts were isolated by methods described previously (22). 0.16 ml of 1.85 mM NaOH, 7.4% 2-mercaptoethanol mixture was added to 1 ml of cultures grown to the same cell density. After a 10-min incubation on ice, 0.16 ml of 50% trichloroacetic acid was added, and samples were placed on ice for a second 10-min incubation. After a brief spin, the pellets were washed with 1.5 ml of ice-cold acetone. Pellets were re-suspended in 2 × sample buffer (4% SDS, 24% glycerol, 100 mM Tris, pH 6.8, 4% 2-mercaptoethanol, and 0.02% Sigma blue G dye). Samples were heated for 4 min at 95°C and spun to remove insoluble material. Equal volumes of samples were loaded on a denaturing Tricine-SDS polyacrylamide gel (23).

For Western blots, proteins were transferred to 0.45-μm nitrocellulose-labessheets (Bio-Rad) and washed in blocking buffer (5% low-fat dry milk in PBS-T (80 mM disodium hydrogen phosphate, 20 mM sodium dihydrogen phosphate, 100 mM sodium chloride, and 0.1% Tween 20)). Blots were incubated with anti-cytochrome c rabbit serum, which recognizes both apo- and holocytochrome c (compliments of M. Dumont). Blots were washed with PBS-T and incubated further with anti-rabbit Ig, horseradish peroxidase-linked whole antibody (Amersham Corp.) for 30 min. Further washes with PBS-T were followed by detection with enhanced chemiluminescence reagents (Amersham Corp.). The blots were wrapped immediately in plastic wrap and exposed to autoradiography film (Fuji) for 1–5 min at room temperature.

In Vivo Competition Assay—To analyze the ability of wild type apo- and holocytochrome c to be converted to holocytochrome c in the presence of mutant cytochrome c, haploid yeast strains containing the loop D replacement mutations were crossed with a haploid yeast strain containing wild type cytochrome c (wGH6, CYC1). Diploid yeast strains were used in vivo in spectroscopic and growth assays (as described above) to determine levels of wild type cytochrome c which could be converted to holocytochrome c in the presence of mutant cytochrome c. Diploid strains with mutant cytochrome c and wild type (mutant/CYC1) were compared with diploid strains with cytochrome c deletion mutant and wild type protein (del cycl1/CYC1-)

Directed, Random Mutagenesis of the Yeast Chromosome—Directed, random mutagenesis was performed essentially as described previously (24). Three oligonucleotides, each of which was randomized at two codons, were synthesized on an Applied Biosystems DNA synthesizer. Oligonucleotides 3R and 4R randomized Pro-76 and Gly-77; oligonucleotides 8R and 9R randomized Thr-74 and Ile-75; the sequences are given here. 3R/4R: 5'-GAGACTCTTTGACGATCCAAAAGAATTATTTNGG/C NNG/C NCC ACCAAATGGCCCTTGGGTGTTGGAAG-3'. 8R: 5'-GAGACCTTTGACGATCCAAAAGAATTATTTNGG/T NNG/T ATGGCCTCGGTGTTGGAAG-3'. 9R: 5'-GAGACCTTTGACGATCCAAAAGAATTATTTNGG/T NNG/T CCGTGTACCATAGGTTCTGGGTGTTGGAAG-3'.

30 μg of each oligonucleotide was transformed into yeast strain CS1, which contains a deletion of five residues in loop D and does not produce a folded cytochrome c. Transformed cells were plated on YPD plates and incubated at 25 or 30°C for 4–8 days. Colonies that grew over the background lawn, indicating that this oligonucleotide had integrated the CYC1 locus to produce a functional cytochrome c, were picked, subcloned, checked for phenotype on SD–(leu), SD–(ura), and SD–(trp) plates. The CYC1 locus in these strains was then sequenced by PCR amplification of chromosomal DNA, as described above. The

procedure was repeated, and cells were resuspended in 0.8 ml of 0.1 M sodium phosphate buffer, pH 7.0. To control for the number of cells present in the sample, cell density was measured on a Klett-Summerson colorimeter by making a 1:40 to 1:200 dilution of the stock cells in 0.1 M sodium phosphate buffer, pH 7.0. 250-μl samples, made from the stock cells, gave a cell density of 350 KU based on the cell density was measured by low temperature (19°C) difference spectroscopy on an SLM-Aminco UV-visible DW2000 spectrophotometer. Liquid YPD medium was inoculated with a colony of mutant diploid yeast cells and grown overnight at 30°C. The growth stage of the cultures was synchronized by refrigeration at 4°C for 8 h. One ml of this stationary culture was used to inoculate 25-ml of YPD medium, and these cultures were grown at 25, 30, or 37°C with aeration for 15–16 h. Cultures were spun at 3,500 rpm for 3 min at 4°C, supernatant was discarded, and cells were washed with 10 ml of 0.1 M sodium phosphate buffer, pH 7.0. This wash
Mutations of Loop D in Cytochrome c

Table I

| Amino acid and oligonucleotide sequences of mutated loop D |
|----------------------------------------------------------|
| **Strain Designation** | **Sequence of Mutated Region and Synthetic Oligonucleotide** |
|------------------------|----------------------------------------------------------|
| Wild Type (Parent) | **C6** | `TAC TTO ACT AAC COA AAG AAA TAT GCT GAT GCT GAT AAC AAG ATS GTC TTT GGT GGG` |
| Del D | **C6** | `TAC TTO ACT AAC COA AAG AAA TAT GCT GAT GCT GAT AAC AAG ATS GTC TTT GGT GGG` |
| RepD1 (Parent) | **C6** | `TAC TTO ACT AAC COA AAG AAA TAT GCT GAT GCT GAT AAC AAG ATS GTC TTT GGT GGG` |
| RepD2 (Parent) | **C6** | `TAC TTO ACT AAC COA AAG AAA TAT GCT GAT GCT GAT AAC AAG ATS GTC TTT GGT GGG` |
| RepD3 (Parent) | **C6** | `TAC TTO ACT AAC COA AAG AAA TAT GCT GAT GCT GAT AAC AAG ATS GTC TTT GGT GGG` |
| RepD4 (Parent) | **C6** | `TAC TTO ACT AAC COA AAG AAA TAT GCT GAT GCT GAT AAC AAG ATS GTC TTT GGT GGG` |

Fig. 1. Ribbon structure of iso-1-cytochrome c. The 11-loops and helix secondary structure regions of S. cerevisiae iso-1-cytochrome c(47) are displayed as various shades of gray and labeled. The secondary structures are designated as amino-terminal helix (N-helix), residues 2–11; loop A, residues 18–32; loop B, residues 34–43; loop C, residues 40–54; 60s helix, residues 61–69; loop D, residues 70–84; and carboxy-terminal helix (C-helix), residues 88–101. The side chains of Tyr-74, Ile-75, Pro-76, Gly-77, Thr-78, and Lys-79 are shown as ball-and-stick-figures and are labeled.

entire CYC1 locus of all of these strains was sequenced at least once. To verify sequence accuracy, many were sequenced twice.

RESULTS

Mutagenesis and Integration of Mutations into Yeast CYC1 Locus—Four hybrid iso-1-cytochrome c proteins were created in which loop D (Fig. 1) was replaced by loop sequences from other proteins. Two of these replacements are of homologous loop D (Fig. 1) was replaced by loop sequences from iso-1-cytochrome c in which loop D was replaced by loop sequences from iso-1-cytochrome c. The loop ends from the prokaryotic cytochromes are structurally similar to the ends of loop D in iso-1-cytochrome c. The heterologous proteins were selected because the distance between the loop end points in the tertiary structure was similar to that found in the native loop D. These loop replacements were chosen so as to not perturb the overall protein structure. Loop D from P. aeruginosa cytochrome c(C6: wild type; and C93: wild type, C102T), a complete deletion of cytochrome c(C15, del cyc), and a partial deletion of loop D, residues 74–78, (C51, DelD) were also created and were used for direct comparison with the mutant strains in the in vivo structural and functional studies. A summary of the loop D replacements of these strains is given in Table I.

Function of Loop D Replacement Mutations in Vivo—Function of mutant iso-1-cytochrome c proteins was determined by monitoring the ability of diploid strains, which contain a single copy of the mutant cyc1 gene, to grow on lactate or glycerol. To grow on nonfermentable carbon sources, S. cerevisiae strains must contain a functional cytochrome c(25). Mutants that retain a low level of function, to 1% of normal cytochrome c levels, can utilize nonfermentable carbon sources such as glycerol or ethanol (26). Strains lacking both iso-1-cytochrome c and iso-2-cytochrome c do not grow on lactate or glycerol. Our mutant strains do not contain iso-2, so growth on SD medium lacking leucine do not grow on SD(–ura) plates and are colonies that have correctly integrated the mutant gene at the CYC1 locus.

Loop D replacement mutations were created by oligonucleotide-directed mutagenesis of the parent pYC17 (DelD) plasmid, integrated into the yeast chromosome, and confirmed by sequence analysis of a PCR-amplified fragment from the yeast chromosome. Isogenic strains containing wild type cytochrome c(C6: wild type; and C93: wild type, C102T), a complete deletion of cytochrome c(C15, del cyc), and a partial deletion of loop D, residues 74–78, (C51, DelD) were also created and were used for direct comparison with the mutant strains in the in vivo structural and functional studies. A summary of the loop D replacements of these strains is given in Table I.

Growth of mutant strains was determined by monitoring the cell density in liquid lactate for a period of 25–30 h at 30 °C (Fig. 3). These growth curves were compared with those of control strains, C6 (wild type), C93 (wild type, C102T), C15 (del cyc), and C51 (DelD). The small increase in the number of KU at the beginning of the experiment (8 KU) and at the end of monitored growth (~20 KU) observed with the cytochrome c deletion strain may be caused by residual dextrose in the culture medium from the original synchronized culture, by cell
lysis, or by evaporation of medium due to aeration and not as an ability to utilize lactate as a carbon source. Mutant yeast strains that display growth comparable to the cytochrome c deletion yeast strain are interpreted as having a nonfunctional cytochrome c. As reported previously, DelD was unable to grow on lactate (11), and placing this mutation in our background strain produces the same results (Fig. 3). All of the loop D replacement mutations did not grow on lactate, indicating that all mutant strains were deficient in a functional cytochrome c.

To determine if the mutant proteins were unstable and sensitive to temperature, growth curves were also performed at 25°C; no function was observed at this temperature either (Fig. 3).

To assay for a decrease in the holoprotein levels or a decrease in the function of the protein to a level that is 1% of normal cytochrome c (26), cells were plated on glycerol medium. None of the loop D replacement yeast strains were able to grow on glycerol plates, indicating that not even low levels of functional holoproteins are present (data not shown). Possible causes for these results are that no holoprotein is produced, or that holoprotein is produced, but is nonfunctional, or that holoprotein is extremely unstable and is not detectable by these assays.

To determine whether the loop D replacement mutants could enter the mitochondria, bind heme, and thus produce a folded, but nonfunctional, cytochrome c, low temperature difference spectroscopy of intact cells was performed. The loop D deletion (DelD) was previously shown to be deficient in folded holocytochrome c (11), and the same results were obtained in our integration strain (Fig. 4). None of the four loop replacement yeast strains showed the presence of any folded cytochrome c based on absorbance at 548 nm (Fig. 4); thus, these loop D replacement strains did not produce a folded heme-bound holoprotein. Spectroscopic assays were done on cells grown at 25 °C with the same results (Fig. 4A), indicating that the mutant proteins are not simply temperature-sensitive. Five percent of the normal level of cytochrome c in the cell is iso-2-cytochrome c, and this level is detectable by low temperature difference spectroscopy of intact cells. The loop D deletion (DelD) was previously shown to be deficient in folded holocytochrome c (11), and the same results were obtained in our integration strain (Fig. 4). None of the four loop replacement yeast strains showed the presence of any folded cytochrome c based on absorbance at 548 nm (Fig. 4); thus, these loop D replacement strains did not produce a folded heme-bound holoprotein. Spectroscopic assays were done on cells grown at 25 °C with the same results (Fig. 4A), indicating that the mutant proteins are not simply temperature-sensitive. Five percent of the normal level of cytochrome c in the cell is iso-2-cytochrome c, and this level is detectable by low temperature difference spectroscopy of intact cells. The loop D deletion (DelD) was previously shown to be deficient in folded holocytochrome c (11), and the same results were obtained in our integration strain (Fig. 4). None of the four loop replacement yeast strains showed the presence of any folded cytochrome c based on absorbance at 548 nm (Fig. 4); thus, these loop D replacement strains did not produce a folded heme-bound holoprotein. Spectroscopic assays were done on cells grown at 25 °C with the same results (Fig. 4A), indicating that the mutant proteins are not simply temperature-sensitive.
Apoprotein Levels in Loop D Deletion and Replacement Mutants—To determine if the loop D replacement mutations result in a loss of apoprotein synthesis or in unstable apoproteins that are rapidly proteolyzed, the presence of apoprotein in the loop D deletion and replacement mutants was determined by immunoblotting whole cell protein extracts with anti-cytochrome c antibody. Apoprotein corresponding to the correct molecular mass was observed in the loop D deletion and replacement yeast strains (Fig. 5). The half-life of wild type apocytochrome c is short, on the order of minutes, and wild type apoprotein is converted rapidly to holoprotein (27). Low levels of wild type apoprotein are present in a yeast strain that lacks cytochrome c heme lyase demonstrating the short half-life of apo-1-cytochrome c when holoprotein production is blocked (27). The level of apoprotein observed in the RepD1, RepD4, and DelD yeast stains thus appears comparable to the level of apoprotein observed in the heme lyase-deficient yeast strain. Therefore, these mutant apoproteins are present at the level of wild type apocytochrome c but are not rapidly converted to holoprotein. The RepD2 yeast strain contains a higher level of apoprotein than the other mutant strains; however, apparent degradation products appear as smaller molecular weight bands. The RepD3 yeast strain shows a decrease in the level of apoprotein, suggesting that this is a more proteolytically sensitive apoprotein. The presence of apoprotein in each mutant strain suggests that the processing of apocytochrome c to holocytochrome c should occur at least at low levels. The lack of any detectable level of holoprotein suggests that the amino acid changes made in the loop D sequence interfere with the biosynthetic processing of the apoprotein to the holoprotein.

In Vivo Competition Assay—Although the yeast stains containing the loop D replacement mutations were not producing detectable holoprotein, it is possible that the mutant apoproteins could compete with wild type apoprotein conversion to holoprotein. This competition could occur at a step in biosynthesis, such as mitochondrial or heme lyase recognition, mitochondrial import, or heme binding. To test this possibility, a
competition study was designed. A haploid yeast strain containing wild type cytochrome c and the haploid yeast strains containing the loop D replacement mutations were crossed to determine if mutant cytochrome c that was converted to holo-cytochrome c. The growth and amount of cytochrome c in WT/RepD strains were comparable to those of the WT/Delcyc1 control strain, demonstrating that these mutant proteins were not able to compete with wild type apoprotein biosynthesis or were competing at a level that could not be detected by these in vivo assays. In fact, slight increases in the absorbance at 548 nm is observed in several of the replacement mutants (data not shown). These results further suggest that some step in cytochrome c biosynthesis requires the amino acid sequences present in the loop D region.

Random Mutagenesis of Conserved Residues in Loop D—To determine which residues were critical for biosynthesis of iso-1-cytochrome c, directed, random mutagenesis was applied pairwise to six highly conserved residues in loop D. Five of these residues were those that had previously been deleted to produce DeD, the loop D deletion strain. Residues were randomly mutated pairwise using the following oligonucleotides: Tyr-74 and Ile-75, serine, glycine at position 74. With two exceptions, histidine and glutamine, only hydrophobic residues are found 15 times in the data base; 7 times it is encoded by CTG, 4 times by TTG. Arginine codons CTG and GCC and leucine codons CTT and CTG cannot be compared because they were encoded only in one or the other of the oligonucleotides. Thus, overwhelming codon preferences, which would result in apparent amino acid preferences, do not appear to occur with our random mutagenesis protocol. Because of the number of mutants, small, statistically significant codon preferences among the oligonucleotides cannot be ruled out.

Twenty-eight colonies obtained from random mutagenesis of Tyr-74 and Ile-75 with oligonucleotide 9R were sequenced, and the data are shown in Tables II–IV. The wild type sequence did not appear in any of these 28 strains, demonstrating that we have not sampled all possible functional sequences. There are five sequences that appear in more than one of the mutant strains sequenced: Arg-74 and Val-75 in 9R-18, 23 and 31, Ser-74 and Leu-75 in strains 9R-7 and 19; Thr-74 and Ala-75 in 9R-3 and 25; Gly-74 and Val-75 in 9R-5 and 24; and Met-74 and Leu-75 in 9R-2 and 20. There is a prevalence of bulky residues, particularly tryptophan, and small residues (serine, threonine, and glycine) at position 74. With two exceptions, histidine and glutamine, only hydrophobic residues are found at residue 75 among the 9R mutant strains.

Of the 195 colonies obtained from random mutagenesis of Pro-76 and Gly-77, the CYC1 locus from the chromosomal DNA sequences shown in Tables II–IV produce a functional protein; however, our current analysis shows that some random mutants discovered using this protocol function at an extremely low level.3

The codon sequences shown in Tables II, III, and IV and summarized in Table V show that individual codons do not appear to be overwhelmingly favored. For example, arginine is encoded 12 times by AGG and nine times by CGG. Leucine is found 15 times in the data base; 7 times it is encoded by CTG, 4 times by TTG. Arginine codons CTG and GCC and leucine codons CTT and CTG cannot be compared because they were encoded only in one or the other of the oligonucleotides. Thus, overwhelming codon preferences, which would result in apparent amino acid preferences, do not appear to occur with our random mutagenesis protocol. Because of the number of mutants, small, statistically significant codon preferences among the oligonucleotides cannot be ruled out.

Random Mutagenesis of Conserved Residues in Loop D—To determine which residues were critical for biosynthesis of iso-1-cytochrome c, directed, random mutagenesis was applied pairwise to six highly conserved residues in loop D. Five of these residues were those that had previously been deleted to produce DeD, the loop D deletion strain. Residues were randomly mutated pairwise using the following oligonucleotides: Tyr-74 and Ile-75, serine, glycine at position 74. With two exceptions, histidine and glutamine, only hydrophobic residues are found 15 times in the data base; 7 times it is encoded by CTG, 4 times by TTG. Arginine codons CTG and GCC and leucine codons CTT and CTG cannot be compared because they were encoded only in one or the other of the oligonucleotides. Thus, overwhelming codon preferences, which would result in apparent amino acid preferences, do not appear to occur with our random mutagenesis protocol. Because of the number of mutants, small, statistically significant codon preferences among the oligonucleotides cannot be ruled out.

Twenty-eight colonies obtained from random mutagenesis of Tyr-74 and Ile-75 with oligonucleotide 9R were sequenced, and the data are shown in Tables II, V, and VI. The wild type sequence did not appear in any of these 28 strains, demonstrating that we have not sampled all possible functional sequences. There are five sequences that appear in more than one of the mutant strains sequenced: Arg-74 and Val-75 in 9R-18, 23 and 31, Ser-74 and Leu-75 in strains 9R-7 and 19; Thr-74 and Ala-75 in 9R-3 and 25; Gly-74 and Val-75 in 9R-5 and 24; and Met-74 and Leu-75 in 9R-2 and 20. There is a prevalence of bulky residues, particularly tryptophan, and small residues (serine, threonine, and glycine) at position 74. With two exceptions, histidine and glutamine, only hydrophobic residues are found at residue 75 among the 9R mutant strains.

Of the 195 colonies obtained from random mutagenesis of Pro-76 and Gly-77, the CYC1 locus from the chromosomal DNA

\(^3\) J. S. Spitzer, B. M. Gilden, and J. S. Fetrow, unpublished results.
of 47 colonies was sequenced. The results are shown in Table II and summarized in Tables V and VI. One of the colonies produced using the 3R and 4R oligonucleotides, 3R-8, returned a wild type sequence. There are several sequence combinations that appear in more than one of the mutant strains sequenced: Gly-76 and Gly-77 in strains 3R-11, 3R-25, and 4R-11; Gly-76 and Arg-77 in strains 3R-10 and 3R-24; Gly-76 and Phe-77 in strains 3R-2 and 4R-43; Arg-76 and Leu-77 in strains 3R-16 and 3R-18; Arg-76 and Gly-77 in strains 3R-5 and 3R-17. This result suggests that an adequate number of mutant strains have been sampled to study the general trends in the Pro-76 and Gly-77 mutants; however, it is probable that other residue combinations that produce a functional cytochrome c would be found if more strains were sequenced.

The data in Tables II and VI show that glycine, arginine, and tryptophan are overwhelmingly preferred at position 76, whereas arginine and glycine are found often at position 77. This is a curious result because glycine is the smallest amino acid, whereas tryptophan and arginine are two of the largest. The distribution of amino acids found at these positions does not seem to change as the isolation temperature is changed from 30 to 25 °C (Table VI); however, the number of mutants sequenced thus far is not sufficient to make a statistical argument.

Forty-one colonies obtained from random mutagenesis of Thr-78 and Lys-79 with oligonucleotide 8R were sequenced (Tables II, V, and VI). One colony, 8R-3, returned a wild type sequence. Several strains returned the same sequence combinations: valine and serine in 8R-8, 26, and 36; alanine and serine in 8R-25, 40, and 47; glycine and alanine in 8R-17 and 19; and serine and arginine in 8R-18 and 23. A preponderance of β-branched residues, isoleucine, threonine and valine, and small residues, glycine, alanine and serine, are found at position 78. Serine and arginine appear to be preferred at position 79. Incubation temperature does not produce a significantly different set of amino acids at these positions (Table VI). Although random mutagenesis does not uncover any residues that are strictly conserved and required for biosynthesis or function of holocytochrome c, a number of preferences were uncovered. Because of the random mutagenesis protocol that we used, all of the residue combinations discovered produce a functional, and thus folded, cytochrome c; however, the mutant
proteins may be minimally functional. An understanding of the exact effect of these mutations on the structure and function of cytochrome c can now be determined by further examination of structure and function in these mutants.

**DISCUSSION**

The importance of the conserved loop D sequences in eukaryotic cytochromes c for function and biosynthesis of this protein has now been examined. Mutant proteins containing loop D replacements were assayed in vivo for function and structure. The chromosomal integration system that we have developed and described here (Fig. 2) allows us to study the effects of loop D replacement mutations inserted at the CYC1 locus in yeast. Moreover, the effects of these mutations can be analyzed at the normal wild-type cytochrome c expression levels to discover the actual effect of the mutation on growth of the yeast.

All loop replacement mutants examined here failed to produce holoprotein, so directed random mutagenesis was performed to identify specifically those residues important for the structure, folding, or biosynthesis of cytochrome c. Of the six highly conserved loop D residues that were randomly mutated, none is strictly required for the folding or structure of cytochrome c. Residue preferences can now be examined in more detail.

A Unique Example of Loop Replacements That Result in Loss of Folded Protein—In this work, loop D (residues 70–84, Fig. 1) was replaced with a series of homologous cytochrome c loops and heterologous loops from other proteins (Table I). One of the significant results is the profound effect that the deletion and replacement of loop D has on the overall structure of iso-1-cytochrome c, which is in direct contrast to the generally predicted and observed results of other loop deletion and replacement experiments. A change in the primary sequence of a surface loop usually does not prevent protein folding. Multiple examples of loop deletions and replacements show that such mutations do not affect the overall structure of the protein, although protein function and stability are often affected. In iso-1-cytochrome c, loop B, residues 34–45, and loop C, residues 40–54, were deleted, and the resulting proteins folded and were functional, although less so than the wild-type protein (11). A six-residue deletion of a mobile loop in Staphylococcal nuclease, located near a putative catalytic site, resulted in a more active, more stable protein (28). NMR experiments demonstrated that this mutation had essentially the same structure as the wild-type protein, except that the larger loop was replaced with a smaller β-turn (28). A deletion of four residues from the mobile loop of triose phosphate isomerase resulted in a folded protein with a significant decrease in specific catalytic activity, suggesting that this mobile loop has a role in the catalytic reaction mechanism (29).

**Table IV**

| Strain no. | Temperature | Amino acid sequence 78 | Amino acid sequence 79 | Codon sequence 78 | Codon sequence 79 |
|------------|-------------|------------------------|------------------------|------------------|------------------|
|            | °C          |                        |                        |                  |                  |
| C93 (wt)   | 30          | Thr                    | Lys                    | ACC              | AAG              |
| 8R-1       | 30          | Gly                    | Asn                    | GGG              | AAT              |
| 8R-3       | 30          | Thr                    | Lys                    | ACT              | AAG              |
| 8R-4       | 30          | Ile                    | Thr                    | ATT              | ACG              |
| 8R-5       | 30          | Ala                    | Arg                    | GCT              | AGG              |
| 8R-7       | 30          | Asn                    | Phe                    | AAT              | TTT              |
| 8R-8       | 30          | Val                    | Ser                    | GTT              | TCG              |
| 8R-9       | 30          | Gly                    | Val                    | GGT              | GTT              |
| 8R-10      | 30          | Gly                    | Leu                    | GGG              | TTG              |
| 8R-11      | 30          | Thr                    | Asp                    | ACT              | GAT              |
| 8R-12      | 30          | Val                    | Met                    | GTG              | ATG              |
| 8R-13      | 30          | Ala                    | Lys                    | GCT              | AAG              |
| 8R-14      | 30          | Ala                    | Ala                    | GCT              | GGG              |
| 8R-16      | 30          | Val                    | Gly                    | GTG              | GGG              |
| 8R-17      | 30          | Gly                    | Ala                    | GGT              | GCG              |
| 8R-18      | 30          | Ser                    | Arg                    | TCG              | CGT              |
| 8R-19      | 30          | Gly                    | Ala                    | GGT              | GCT              |
| 8R-20      | 30          | Thr                    | Ser                    | ACG              | TCG              |
| 8R-21      | 30          | Ser                    | Ser                    | TCT              | AGT              |
| 8R-22      | 30          | Val                    | Arg                    | GTT              | CGG              |
| 8R-23      | 30          | Ser                    | Arg                    | TCG              | AGG              |
| 8R-24      | 30          | Thr                    | Arg                    | ACT              | CGT              |
| 8R-25      | 30          | Ala                    | Ser                    | GCG              | TCG              |
| 8R-26      | 30          | Val                    | Ser                    | GTT              | TCT              |
| 8R-27      | 30          | Val                    | Thr                    | GTT              | ACT              |
| 8R-28      | 30          | Ala                    | Gly                    | GCT              | GGT              |
| 8R-30      | 30          | Met                    | Ser                    | AGT              | AGT              |
| 8R-32      | 30          | Ser                    | Thr                    | TCT              | ACG              |
| 8R-35      | 25          | Gly                    | Arg                    | GGG              | CGG              |
| 8R-36      | 25          | Val                    | Ser                    | GTT              | AGT              |
| 8R-37      | 25          | Gly                    | Gln                    | GGT              | CAG              |
| 8R-38      | 25          | Ser                    | Ala                    | TCG              | GCT              |
| 8R-39      | 25          | Ile                    | Ser                    | ATT              | AGG              |
| 8R-40      | 25          | Ala                    | Ser                    | GCT              | AGT              |
| 8R-42      | 25          | Ile                    | Gly                    | ATT              | GGT              |
| 8R-43      | 25          | Gly                    | Ser                    | GGT              | TCT              |
| 8R-45      | 25          | Ser                    | Leu                    | AGT              | CTT              |
| 8R-46      | 25          | Leu                    | Arg                    | CTT              | CGG              |
| 8R-47      | 25          | Ala                    | Ser                    | GCG              | AGT              |
| 8R-48      | 25          | Asn                    | Gly                    | AAT              | GGG              |
| 8R-49      | 25          | Ile                    | Gln                    | ATT              | CAG              |
| 8R-52      | 25          | Ser                    | Met                    | AGT              | ATG              |
deleting a highly conserved four-residue region within the center of the loop and sealing it with a glycine residue (30). The major effect of this loop deletion was to decrease loop flexibility. In both dihydrofolate reductase and triose phosphate isomerase, structural studies indicate the flexibility of the loop allows conformational changes necessary for performing the function of stabilizing intermediates along the reaction pathway (29, 30).

Loop replacements also generally result in a folded protein, although if the loop is involved in protein function the resulting hybrid protein is not necessarily functional. Replacement of hypervariable loop regions in the immunoglobulins results in folded immunoglobulin proteins with altered antigen specificity (31). Swapping of variable loops of homologous ribonuclease proteins, RNase and angiogenin, results in an exchange of substrate specificity between the two proteins (32, 33). Although deletion of loop A, residues 18–32, in cytochrome c resulted in a complete deficiency of holoprotein, replacement of this loop with various homologous and heterologous loop sequences resulted in folded, cytochrome c-like molecules (6, 11). Loop A can tolerate a large variation in sequence and structure, including one heterologous loop from an unrelated protein, suggesting that any loop-like sequence would fit into this surface loop and that the specific sequence is not important to protein biosynthesis.

Unlike other loops, the replacement of the yeast loop D sequence with sequences that are capable of forming loops in their native protein does not result in the production of a heme-bound protein, even though this is a typical V-loop found at the surface of the protein (Fig. 1). A low level of apoprotein is present in loop D mutant yeast strains (Fig. 5) comparable to the levels of wild type apoprotein found in yeast cells lacking heme lyase (27). Minimally, a low level of holoprotein should be produced in these yeast strains unless apoprotein conversion to holoprotein is blocked or holoprotein is rapidly degraded. To be undetectable in both the in vivo functional and structural assays, the mutant holoproteins would have to be degraded very rapidly to a level below 1–5% (12) of normal cytochrome c.

Thus, this is a unique example of a loop in which replacement of the loop structure with a homologous loop structure does not

| Codon | 5′-Tyr-74-3′ | 5′-Ile-75-3′ | 5′-Thr-78-3′ | 5′-Lys-79-3′ | 5′-Codon-3′ | Pro-76 | Gly-77 |
|-------|-------------|-------------|-------------|-------------|-------------|-------|-------|
| Ala   | GCG         | 1           | 1           | 2           | 2           | GCG   | 2     | 4     |
|       | GCT         | 1           | 5           | 2           | 2           | GCC   | 1     |       |
| Arg   | AGG         | 1           | 3           | 3           | 4           | AGG   | 4     | 4     |
|       | CGG         | 1           | 3           | CGG         | 3           | CGG   | 3     | 2     |
|       | CGT         | 1           | 2           | CGC         | 1           | CGC   | 1     | 1     |
| Asn   | AAT         | 1           | 2           | 1           | 1           | AAC   | 1     |       |
| Asp   | GAT         | 1           | 1           | 1           | 1           | GAC   | 1     | 1     |
| Cys   | TGT         | 1           | TGC         | 1           | TGC         |       |       |       |
| Glu   | CAG         | 1           | 2           | 2           | 1           | CAG   | 5     | 1     |
| Glu   | GAG         | 1           | GAG         | 1           | GAG         | 1     | 2     |       |
| Gly   | GGG         | 1           | 4           | 2           | GGG         | 8     | 7     |       |
|       | GTT         | 3           | 4           | 2           | GGC         | 3     | 5     |       |
| His   | CAT         | 2           | CAT         | 2           | CAT         |       |       |       |
|       | CAC         | 2           | CAC         | 1           | CAC         |       |       |       |
| Ile   | ATT         | 2           | 4           | ATT         | 1           | ATT   | 1     |       |
| Leu   | CTG         | 1           | 3           | CTG         | 1           | CTG   | 1     |       |
|       | CTT         | 2           | 1           | CTT         | 3           | CTT   | 3     |       |
| Lys   | AAA         | 2           | AAG         | 2           | AAG         | 2     |       |       |
| Met   | ATG         | 1           | ATG         | 1           | ATG         | 1     | 2     |       |
| Phe   | TTT         | 2           | TTT         | 1           | TTC         | 1     |       |       |
|       | TTC         | 1           | TTC         | 2           | TTC         |       |       |       |
| Pro   | CCG         | CCG         | CCG         | 1           | CCG         |       |       |       |
| Ser   | AGT         | 2           | 2           | 5           | AGC         | 2     | 4     |       |
|       | TCG         | 1           | 3           | 3           | TCG         | 1     |       |       |
| Thr   | ACT         | 1           | 3           | ACC         | 1           | ACC   | 1     |       |
|       | ACG         | 1           | 2           | ACG         | 1           | ACG   | 1     |       |
| Trp   | TGG         | 5           | 2           | TGG         | 6           | TGG   | 2     |       |
| Tyr   | TAT         | 1           | TAC         | 1           | TAC         | 1     | 1     |       |
| Val   | GTG         | 2           | 4           | 3           | GTG         | 3     | 1     |       |
|       | GTT         | 1           | 5           | 4           | GTT         |       |       |       |

Table V: Codon usage table for the random mutants
A Possible Role for Loop D in Biosynthesis of Holocytochrome c—In support of the hypothesis that some part of loop D plays an important role in biosynthesis, previous competition experiments between wild type rat cytochrome c and cytochrome c protein fragments (34) resulted in the hypothesis that the carboxyl-terminal end of the protein, including loop D, is important for mitochondrial import (35). In this in vitro binding competition experiment, the ability of rat cytochrome c fragments to compete with labeled wild type rat apocytochrome c for binding to rat mitochondria was compared. Only the fragment containing the carboxyl-terminal end of the protein, including the loop D region, was able to compete with wild type protein. Fragments encompassing the amino-terminal region were unable to compete (34). Analogous regions from other eukaryotic systems including rabbit, horse, and yeast were also able to compete with the rat protein in this system (34). However, the homologous fragment from P. denitrificans could not compete, suggesting that some part of the carboxyl-terminal region, which is conserved in eukaryotic cytochromes c, may be important for mitochondrial recognition, heme lyase recognition/reactivity, or mitochondrial import.

Our in vivo results concur with these in vitro results. Loop D replacement mutations, where loop D is replaced with homologous loops from prokaryotic cytochromes c, do not produce holoprotein and cannot compete with wild type cytochrome c for mitochondrial import in vivo. In contrast, a carboxyl-terminal helix deletion mutation, which also does not yield holoprotein, was able to compete with wild type apocytochrome c for apoprotein processing.4 Therefore, a mutant apoprotein that contains the wild type loop D sequence (as well as the rest of the protein) can compete with wild type holoprotein production, but apoproteins that lack the wild type loop D sequence are unable to block wild type cytochrome c bioprocessing.

Sequence variability histograms show that the loop D sequence, Asn-Pro-Lys-Tyr-Ile-Pro-Gly-Thr-Lys-Met-Ala-Phe-Gly-Gly, residues 70-84 in yeast, is highly conserved within the eukaryotic cytochrome c sequences (36). This degree of residue conservation is not observed in the smaller sample of prokaryotic sequences. In three of the prokaryotic cytochromes c whose crystal structure has been solved, P. aeruginosa (37), P. denitrificans (38), and Rhodospirillum rubrum (39), the loop D secondary structure is present; however, in these prokaryotic proteins, the amino acid sequence is not conserved, and this loop is the site of insertions and deletions, as are other loops in the c-type cytochromes. The sequence conservation of this loop and the results of the loop replacement experiments support the hypothesis that the loop D sequence or structure is involved at a step in the bioprocessing of apocytochrome c, such as mitochondrial recognition or import, heme lyase recognition or reactivity, heme binding or folding.

Random Mutagenesis of Residues Tyr-74, Ile-75, Pro-76, Gly-77, Thr-78, Lys-79—To define more specifically what residues in loop D are important for biosynthesis, structure, or function of iso-1-cytochrome c, we performed pairwise, directed, random mutagenesis of six residues in loop D, five of which had previously been deleted in the loop D deletion strain (Tyr-74, Ile-75, Pro-76, Gly-77, and Thr-78) and an additional residue (Lys-79). In our analysis of eukaryotic cytochrome c sequences, Pro-76, Gly-77, and Lys-79 are strictly conserved. Thr-78 is highly conserved; the only asparagine is also found at this position. Tyr-74 and Ile-75 are also highly conserved. Tyr-74 can be replaced with the aromatic residues phenylalanine and histidine, which occur three times and one time in the data base, respectively. Ile-75 can be replaced with the hydrophobic residues valine and methionine, each of which occurs three times in the data base.

The 3R and 4R oligonucleotides that produced the Pro-76 and Gly-77 mutants were designed by randomizing the first two bases of each codon with all four nucleotides while the third position was synthesized only with G and C. The 8R and 9R oligonucleotides were designed by randomizing the first two bases of the codons encoding Thr-78 and Lys-79, and Tyr-74 and Ile-75, respectively, with all four nucleotides while the third position was randomized only with G and C. The 8R and 9R oligonucleotides, a more even distribution of codons for each of the 20 amino acids is maintained, and overwhelming preferences for amino acids encoded by 6 codons are eliminated. G and T were utilized at the third position in the 8R and 9R oligonucleotides to maximize yeast codon usage preference.

---

4 P. Mulligan-Pullyblank, J. S. Spitzer, B. M. Gilden, and J. S. Fetrow, unpublished results.

---

**Table VI**

|         | Tyr-74 | Ile-75 | Thr-78 | Lys-79 | Pro-76 | Gly-77 |
|---------|--------|--------|--------|--------|--------|--------|
| 25 °C   |        |        |        |        |        |        |
| Pro     | 0      | 0      | 0      | 0      | 0      | 0      |
| Gly     | 0      | 0      | 0      | 0      | 0      | 0      |
| Ala     | 1      | 0      | 1      | 2      | 5      | 2      |
| Ile     | 0      | 0      | 1      | 3      | 1      | 3      |
| Leu     | 1      | 0      | 3      | 3      | 1      | 0      |
| Val     | 3      | 0      | 4      | 5      | 1      | 6      |
| Phe     | 1      | 1      | 1      | 0      | 0      | 0      |
| Tyr     | 1      | 0      | 0      | 0      | 0      | 2      |
| Trp     | 1      | 4      | 2      | 0      | 0      | 0      |
| His     | 0      | 0      | 1      | 1      | 0      | 0      |
| Asn     | 0      | 1      | 0      | 0      | 0      | 1      |
| Glu     | 0      | 0      | 1      | 2      | 0      | 4      |
| Ser     | 2      | 1      | 0      | 0      | 3      | 4      |
| Thr     | 0      | 1      | 0      | 0      | 4      | 0      |
| Cys     | 0      | 0      | 0      | 0      | 0      | 0      |
| Met     | 1      | 1      | 0      | 0      | 0      | 0      |
| Arg     | 2      | 1      | 0      | 0      | 4      | 6      |
| Lys     | 0      | 0      | 0      | 0      | 2      | 4      |
| Asp     | 0      | 0      | 0      | 0      | 0      | 1      |
| Glu     | 0      | 0      | 0      | 0      | 0      | 1      |

No. of codons:

3 No. of codons is the number of codons that encode each amino acid using the NNG/C or NNG/T schemes described under “Experimental Procedures.”
What can the sequences of these random mutants tell us about the structure and function of loop D in cytochrome c? First, and most obviously, the exact sequence of these six residues is not essential for biosynthesis of a folded cytochrome c. What else can be learned from these sequences? If the allowed sequence space has been sampled sufficiently, these sequences can suggest amino acid characteristics that are necessary for cytochrome c function. Two indications that sequence space has been adequately sampled are if the wild type sequence is found among the mutants and if sequence combinations are found multiple times. As detailed under “Results,” two of the three experiments returned the wild type sequence, and a number of mutant combinations occurred several times in each experiment (Tables II, III, and IV). Thus, sequence space has probably been sufficiently sampled to allow us to suggest amino acid preferences that produce a functional cytochrome c, particularly for Pro-76, Gly-77, Thr-78, and Lys-79 mutants.

The strongest amino acid conservation among the random mutants is found at Ile-75, where only hydrophobic residues are found (Table II). Valine and leucine are strongly favored. This result concurs with the analysis of the sequence data base, where only isoleucine, methionine, and valine are found at this position. 11e-75 is found 98% buried (solvent exposed area (40) divided by the solvent exposed area in an Ala-Xaa-Ala model peptide) in a hydrophobic core that includes Trp-59, Tyr-67, Val-57, and Asn-52 (Fig. 6), implying that hydrophobic packing plays an important role in residue preference at this position.

Thr-78 is 100% buried in iso-1-cytochrome c. Its hydroxyl group participates in a hydrogen bond with the side chain of Asn-63 (Fig. 6). In the random mutants, tryptophan, glycine, arginine, serine, and valine are found more than once (Table VI). Other amino acids, including the native tyrosine, occur once or twice. The side chain of Lys-79 participates in a hydrogen bond with the backbone of residue 47, and this residue is 61% exposed in the iso-1-cytochrome c crystal structure. Positively charged residues occur 10 times at this position; however, serine is also found here 10 times, and alanine and glycine are each found here four times. At both of these positions, preferences for both large and small residues are striking.

Pro-76 and Gly-77 are strictly conserved throughout the eukaryotic sequences. This conservation is expected because the psi backbone dihedral angle of Gly-77 is 1.5°, a conformation that is not allowed for residues with larger side chains. Thus, as expected, glycine is commonly found at these two positions, 11 times at residue 76 and 12 times at residue 77. However, arginine is also preferred at this position, as it is found eight and seven times at positions 76 and 77, respectively. Once more, the preference for large, positively charged residues, and small residues at the same position is striking and leads to the suggestion that positively charged residues might play a role in the structure or function of cytochrome c in this region. Such contributions will be more fully analyzed by
studying the stability and function of these mutant proteins in vitro.

A Comparison with Mutations in Other Conserved Residues in Loop D and with the Loop Replacement Results—Other mutations have been accomplished at conserved residues in the loop D regions, particularly at Pro-71 (41–43), Lys-72 (trimethyllysine (44)), Lys-73 (45), and Met-80 (46). At each of these positions, one or a small number of mutations has been shown to produce a properly folded cytochrome c. All are also somewhat functional, except those at residue 80, the second heme iron ligand. Our random mutagenesis results within loop D are consistent with these results. No residues are found to be strictly conserved, and, although there are some specific residue preferences, a number of substitutions yield a properly folded cytochrome c.

Why, then, are these residues so strictly conserved during the evolution of eukaryotic cytochromes c? In vivo analysis of the mutants at residues 71 and 80 shows that many of these changes have a significant effect on the function of cytochrome c. Preliminary analysis of several of our random mutants shows drastic effects on protein function in vivo, suggesting that this region of the protein is critical to proper functioning of cytochrome c.

Conclusions—The results presented here provide a unique example of loop replacement mutations that do not produce a folded protein; the results support the hypothesis that the loop D sequence is important for apocytochrome c bioprocessing. Random mutagenesis of six highly conserved residues in loop D demonstrates that none of these residues is strictly required for the folding or function of cytochrome c. This is the first description of random mutagenesis applied to a loop region in a protein and is a first step in identifying those loop-loop and loop-protein interactions that are important in protein folding, structure, and stability. Preferences were observed for the residues in loop D, and further analysis of these mutants, both in vivo and in vitro, will help elucidate the importance of loop D to the structure, folding, and function of cytochrome c.

Acknowledgments—We thank David Miller for useful discussions, which initiated these experiments; Richard Zitomer for helpful advice and yeast strains used in competition experiments; and Diane Schaak and Gerard Fumo for helpful comments and criticisms.

REFERENCES

1. Venkatachalam, C. M. (1958) Biopolymers 6, 1425–1436
2. Sibanda, B. L., Blundell, T. L., and Thornton, J. M. (1989) J. Mol. Biol. 206, 759–778
3. Leszczynski, J. F., and Rose, G. D. (1986) Science 234, 849–855
4. Ring, C. S., Kneller, D. G., Langridge, R., and Cohen, F. E. (1992) J. Mol. Biol. 224, 685–699
5. Fetrow, J. S. (1995) FEBS Lett. 9, 708–717
6. Murphy, M. E. P., Fetrow, J. S., Burton, R. E., and Brayer, G. D. (1993) Protein Sci. 2, 1429–1440
7. Hynes, T. R., Kautz, R. A., Goodman, M. A., Gill, J. F., and Fox, R. O. (1989) Nature 339, 73–76
8. Pettigrew, G. W., and Moore, G. R. (1987) Cytochromes c: Biological Aspects, Springer-Verlag, Berlin
9. Salamé, F. R. (1977) Annu. Rev. Biochem. 46, 299–329
10. Timkovich, R. (1979) in The Porphyins (Dolphin, D., ed.), pp. 241–293, Academic Press, New York
11. Fetrow, J. S., Cardillo, T. S., and Sherman, F. (1989) Proteins Struct. Funct. Genet. 6, 372–381
12. Schweingruber, M. E., Sherman, F., and Stewart, J. W. (1977) J. Biol. Chem. 252, 6577–6580
13. Elliott, W. B., and Margoliash, E. (1971) in Developments in Applied Spectroscopy (Grove, E. L., and Perkins, A. J., eds.), pp. 125–140, Plenum, New York
14. Dumont, M. E., Cardillo, T. S., Hayes, M. K., and Sherman, F. (1991) Mol. Cell. Biol. 11, 5487–5496
15. Nicholson, D. W., Hendersenberg, C., and Neupert, W. (1988) J. Biol. Chem. 263, 19034–19042
16. Bolivar, F., and Backman, K. (1979) Methods Enzymol. 68, 245–250
17. Kunkel, T. A. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 488–492
18. Ruszel, M., Kidd, S., and Kelley, M. R. (1986) Gene (Amst.) 45, 333–338
19. Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) Methods Enzymol. 154, 367–382
20. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., pp. 1,25, 1,82, and 18.60–18.74, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
21. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. M., Seidman, J. G., Smith, J. A., and Struhl, K. (eds) (1994) Current Protocols in Molecular Biology, pp. 13.7.1–13.7.2 and 13.12.2–13.12.3, John Wiley and Sons, New York
22. Yaffe, M. P., and Schatz, G. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 4819–4823
23. Schagger, H., and von Jagow, G. (1987) Anal. Biochem. 166, 368–379
24. Fumo, G., Spitzer, J. S., and Fetrow, J. S. (1995) Gene (Amst.) 164, 33–39
25. Sherman, F., Stewart, J. W., Jackson, M., Gilmore, R. A., and Parker, J. H. (1974) Genetics 77, 255–284
26. Yamamoto, T., Moerschel, R. P., Wakem, L. P., Ferguson, D., and Sherman, F. (1992) Yeast 8, 935–948
27. Dumont, M. D., Mathews, A. J., Nall, B. T., Balm, S. B., Eustice, D. C., and Sherman, F. (1990) J. Biol. Chem. 265, 2733–2739
28. Poole, L. B., Loveys, D. A., Hale, S. P., Gerlt, J. A., Stanczyk, S. M., and Bolton, P. H. (1991) Biochemistry 30, 3621–3627
29. Pompiliano, D. L., Peyman, A., and Knowles, J. R. (1990) Biochemistry 29, 3186–3194
30. Li, J., Falzone, C. J., Wright, P. E., and Benkovic, S. J. (1992) Biochemistry 31, 7826–7833
31. Jones, P. T., Dear, P. H., Foote, J., Neuberger, M. S., and Winter, G. (1986) Nature 321, 522–525
32. Husser, J. W., and Valley, L. L. (1989) Biochemistry 28, 1875–1884
33. Alleman, R. K., Frensel, S. R., and Benner, S. A. (1991) Protein Eng. 4, 831–835
34. Matsuzuka, S., Monique, A., Hannum, C., Margoliash, E., Sabatini, D., and Morimoto, T. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 4388–4372
35. Miller, D. (1988) Binding and Uptake of Apocytochrome c into Mitochondria. Ph.D. thesis, Northwestern University, Chicago, IL
36. Moore, G. R., and Pettigrew, G. W. (1990) Cytochromes c: Evolutionary, Structural, and Physicochemical Aspects, Springer-Verlag, Berlin
37. Matsuzuka, Y., Takano, T., and Dickerson, R. E. (1982) J. Biol. Chem. 257, 409–409
38. Timkovich, R., and Dickerson, R. E. (1976) J. Biol. Chem. 251, 4033–4046
39. Bhattacharya, G. E. (1981) Rearrangement of the Crystal Structure of Oxidized Rhodospirillum rubrum Cytochrome c2. Ph.D. thesis, University of California, San Diego
40. Lee, B., and Richards, F. M. (1971) J. Mol. Biol. 55, 379–400
41. Ernst, J. F., Hampsey, D. M., Stewart, J. W., Rackovsky, S., Goldstein, D., and Sherman, F. (1985) J. Biol. Chem. 260, 13225–13236
42. Ramdas, L., Sherman, F., and Nall, B. T. (1986) Biochemistry 25, 6952–6958
43. Ramdas, L., and Nall, B. T. (1986) Biochemistry 25, 6959–6964
44. Holzhauer, D., Principe, L., Canklin, K. T., Hickey, D. R., Short, J., Tao, R., Mi-Engdon, G., and Sherman, F. (1977) J. Biol. Chem. 262, 7125–7131
45. Bowler, B. E., May, K., Zaragoza, T., York, P., Dong, A., and Caughey, W. S. (1993) Biochemistry 32, 183–190
46. Hampsey, D. M., Das, G., and Sherman, F. (1986) J. Biol. Chem. 261, 3259–3271
47. Louie, G. V., and Brayer, G. D. (1990) J. Biol. Chem. 254, 527–555