Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T., & Oshima, T. (1984) *J. Biol. Chem. 259*, 2956-2960.
Kim, J.-J. P., & Wu, J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6677-6681.
Kobayashi, K., Doi, S., Negoro, S., Urabe, I., & Okada, H. (1989) *J. Biol. Chem. 264*, 3200-3205.
Jones, T. A. (1985) *Methods Enzymol. 115*, 157-170.
Lienhard, G. E., & Rose, I. A. (1964) *Biochemistry 3*, 185-190.
Londesborough, J. C., & Dalziel, K. (1968) *Biochem. J. 110*, 223-230.
Mas, M. T., & Colman, R. F. (1985) *Biochemistry 24*, 1643-1646.
McKee, J. S., Hlodan, R., & Nimmo, H. G. (1989) *Biochimie 71*, 1059-1064.
Nakamoto, T., & Vennesland, B. (1960) *J. Biol. Chem. 235*, 202-204.
Oppenheimer, N. J. (1987) in *Pyridine Nucleotide Coenzyme: Chemical, Biochemical, and Medical Aspects* (Dolphin, D., Poulson, R., & Avramovic, O., Eds.) Vol. 2A, pp 323-365, John Wiley and Sons, New York.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167-339.
Rossmann, M. G., Moras, D., & Olsen, K. W. (1974) *Nature 250*, 194-199.
Saenger, W., Reddy, B. S., Muhlegger, K., & Weimann, G. (1977) *Nature 267*, 225-229.
Scrutton, N. S., Berry, A., & Perham, R. N. (1990) *Nature 343*, 38-43.
Seltzer, S., Hamilton, G. A., & Westheimer, F. H. (1959) *J. Am. Chem. Soc. 81*, 4018-4024.
Siebert, G., Carsiotis, M., & Plaut, G. W. E. (1957) *J. Biol. Chem. 226*, 977-991.
Steinberger, R., & Westheimer, F. H. (1951) *J. Am. Chem. Soc. 73*, 429-435.
Thorsness, P. E., & Koshland, D. E., Jr. (1987) *J. Biol. Chem. 262*, 10422-10425.
van der Helm, D., Glusker, J. P., Johnson, C. K., Minkin, J. A., Burow, N. E., & Patterson, A. L. (1968) *Acta Crystallogr. Sect. B 24*, 578-592.
Villafranca, J. J., & Colman, R. F. (1974) *Biochemistry 13*, 1152-1160.
Wierenga, R. K., De Maeyer, M. C. H., & Hol, W. G. J. (1985) *Biochemistry 24*, 1346-1357.

Characterization of the 25-Kilodalton Subunit of the Energy-Transducing NADH-Ubiquinone Oxidoreductase of *Paracoccus denitrificans*: Sequence Similarity to the 24-Kilodalton Subunit of the Flavoprotein Fraction of Mammalian Complex I[†,‡]

Xuemin Xu, Akemi Matsuno-Yagi, and Takao Yagi*
Division of Biochemistry, Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037

Received April 23, 1991; Revised Manuscript Received June 14, 1991

ABSTRACT: The NADH dehydrogenase complex isolated from *Paracoccus denitrificans* is composed of approximately 10 unlike polypeptides [Yagi, T. (1986) *Arch. Biochem. Biophys. 250*, 302-311]. Structural genes encoding the subunits of this enzyme complex constitute at least one gene cluster [Xu, X., Matsuno-Yagi, A., & Yagi, T. (1991) *Biochemistry 30*, 6422-6428]. The 25-kDa subunit (*NQO2*), which has been isolated from sodium dodecyl sulfate-polyacrylamide gels, is a polypeptide of this enzyme complex. The partial N-terminal amino acid sequence and amino acid composition of the *NQO2* subunit have been determined. On the basis of the amino acid sequence, the *NQO2* gene was found to be located 1.7 kilobase pairs upstream of the gene for NADH-binding subunit (*NQO1*). The complete nucleotide sequence of the *NQO2* gene was determined. It is composed of 717 base pairs and codes for 239 amino acid residues with a calculated molecular weight of 26 122. The *NQO2* subunit is homologous to the M_r 24 000 subunit of the mammalian mitochondrial NADH-ubiquinone oxidoreductase which bears an electron paramagnetic resonance-visible binuclear iron-sulfur cluster (probably cluster N1b). Comparison of the predicted amino acid sequence of the *Paracoccus NQO2* subunit with those of its mammalian counterparts suggests putative binding sites for the iron-sulfur cluster. In addition, nucleotide sequencing shows the presence of two unidentified reading frames between the *NQO1* and *NQO2* genes. These are designated URF1 and URF2 and are composed of 261 and 642 base pairs, respectively. The possible function of the protein coded for the URF2 is discussed.

Aerobically grown *Paracoccus denitrificans* contains a mitochondrial-type respiratory chain (Stouthamer, 1980). The NADH-ubiquinone (UQ)[1] oxidoreductase of *P. denitrificans* is akin to its mitochondrial counterpart in terms of the presence

[†] This work was supported by U.S. Public Health Service Grants GM33712 to T.Y. and MO1 RR00833 to the General Clinical Research Center. This is Publication No. 6817-MEM from The Scripps Research Institute, La Jolla, CA.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J05337.
* To whom correspondence should be addressed.

[1] Abbreviations: UQ, ubiquinone; Q, quinone; complex I or NDH-1, energy-tranducing NADH-quinone oxidoreductase; complex III, ubiquinol-cytochrome *c* oxidoreductase; bp, base pair(s); FP, IP, and HP, flavoprotein, iron-sulfur protein, and hydrophobic protein fractions of complex I, respectively; SDS, sodium dodecyl sulfate; PVDF, poly(vinylidene difluoride); EPR, electron paramagnetic resonance; FeS cluster, iron-sulfur cluster; URF, unidentified reading frame; kDa, kilodalton(s).

0006-2960/91/0430-8678$02.50/0

of an energy-coupling site and the presence of electron carriers as noncovalently bound FMN and five EPR-visible FeS clusters (Yagi, 1986, 1989, 1991; Meinhardt et al., 1987). These clusters are N1a, N1b, N2, N3, and N4 according to the nomenclature of Ohnishi (Ohnishi & Salerno, 1982). The potent inhibitors of bovine complex I, rotenone, capsaicin, piericidin A, and DCCD, also inhibit the *Paracoccus* NADH–UQ oxidoreductase activity (Yagi, 1986, 1987, 1990). The NADH dehydrogenase complex purified from this bacterium is composed of approximately 10 unlike polypeptides (Yagi, 1986), suggesting that the *Paracoccus* NADH dehydrogenase complex is structurally simpler than its mitochondrial counterpart (complex I) which has more than 25 unlike polypeptides (Hatefi et al., 1985; Hatefi, 1985; Tuschen et al., 1990; Ragan, 1987). Therefore, the *Paracoccus* NADH dehydrogenase complex appears to provide a useful model system for studying the structure and mechanism of action of the mitochondrial complex I (Yagi, 1989, 1991).

The identification of the NADH-binding subunit of the *Paracoccus* NADH dehydrogenase complex was previously reported by this laboratory (Yagi & Dinh, 1990). In addition, the *NQO1* gene encoding this subunit was cloned and sequenced (Xu et al., 1991). Recently, the cDNA sequence of its bovine counterpart was independently determined by Attardi's and Walker's laboratories (Patel et al., 1991; Pilkington et al., 1991), and Weiss and his co-workers (Weiss et al., 1991) performed the cDNA sequencing of the NADH-binding subunit of *Neurospora crassa* mitochondria. Comparison of the deduced primary structures of the *Paracoccus* NADH-binding subunit and its mitochondrial counterparts exhibited mutual sequence identity as anticipated from chemical and immunochemical studies of the subunits (Yagi & Dinh, 1990; Xu & Yagi, 1991). These NADH-binding subunits have also been shown to have sequence similarities to the α-subunit of the NAD-linked hydrogenase of *Alcaligenes eutrophus* H16 (Tran-Betcke et al., 1990; Xu et al., 1991; Patel et al., 1991; Pilkington et al., 1991). An additional important finding is that the structural genes of the *Paracoccus* NADH dehydrogenase complex constitute a gene cluster (Xu et al., 1991).

The 25-kDa polypeptide is one of the components of the purified *Paracoccus* NADH dehydrogenase complex (Yagi, 1986). This polypeptide is one of the two polypeptides present in the FP-type fraction of this complex (Yagi, 1986; George & Ferguson, 1984). In addition, the antiserum to the bovine complex I cross-reacts with this polypeptide (Yagi, 1986). In view of these facts, the 25-kDa polypeptide is considered to be a bona fide component of the *Paracoccus* enzyme complex rather than a contaminating polypeptide which copurifies with the enzyme complex. Therefore, we attempted to isolate this subunit and to find the locus of the structural gene (*NQO2*) encoding the 25-kDa subunit in the gene cluster of the *Paracoccus* NADH dehydrogenase complex which bears the *NQO1* (NADH-binding subunit) gene (Xu et al., 1991).

This paper presents the complete nucleotide sequence of the *NQO2* gene and the deduced amino acid sequence. The *NQO2* gene is located approximately 1.7 kbp upstream of the *NQO1* gene. It is composed of 717 bp, and codes for 239 amino acid residues with a calculated molecular weight of 26 122. The deduced primary structure of the *NQO2* gene has homology to the 24-kDa subunit of the mammalian complex I (von Bahr-Lindström et al., 1983; Nishikimi et al., 1988; Chomyn & Lai, 1989; Pilkington & Walker, 1989). Comparison of the primary structure of the *Paracoccus* 25-kDa and the mammalian 24-kDa subunits displays the presence of four conserved cysteines and one conserved histidine. On the basis of these conserved residues, putative FeS cluster binding sites for cluster N1b of this subunit are discussed. In addition, two unidentified reading frames (URFs) were found between the *NQO1* and the *NQO2* genes.

Materials and Methods

Isolation of the 25-kDa Subunit of the Paracoccus NADH Dehydrogenase Complex. Paracoccus denitrificans (ATCC 13543) cells were grown aerobically by BIOPURE Fine Chemicals, Inc., Boston, MA, in 550-L cultures with glucose as substrate as described previously (Yagi, 1986). *Paracoccus* membranes and the *Paracoccus* NADH dehydrogenase complex were prepared according to Yagi (1986). The *Paracoccus* NADH dehydrogenase complex was fractionated on SDS–polyacrylamide gels (10%) (Yagi, 1986, 1987). The 25-kDa subunit was recovered from the fractionated complex by electroelution (Yagi & Hatefi, 1988; Yagi & Dinh, 1990).

Amino Acid Analysis of the 25-kDa Subunit. The amino acid analysis of the isolated 25-kDa subunit was conducted by subjecting the purified 25-kDa subunit (10 μg) to lyophilization followed by hydrolysis in an evaporated and sealed tube in 6 N HCl at 100 °C for 25 h. Amino acid analysis was then performed on a Beckman 7300 amino acid analyzer (Yagi & Hatefi, 1988).

Sequence Analysis of the 25-kDa Subunit. The N-terminal amino acid sequence of the 25-kDa subunit was determined according to Matsudaira (1987). Samples of the 25-kDa subunit (10 μg) were lyophilized, dissolved in Laemmli's sample buffer containing 80 mM Tris-HCl (pH 6.8), 6% SDS, 5% 2-mercaptoethanol, 0.005% bromophenol blue, and 20% glycerol, and applied to an SDS–polyacrylamide slab gel (55 × 95 × 0.75 mm) composed of 10% acrylamide as described by Laemmli (1970). The gel was electrophoresed for 1 h at 200 V, and then the subunit was electrophoretically transferred to poly(vinylidene difluoride) (PVDF) membranes (Immobilon, Millipore) as described by Matsudaira (1987). The membranes were stained with 0.1% Coomassie brilliant blue R-250 in 50% methanol for 5 min, destained for 10 min with a solution containing 10% acetic acid and 50% methanol, rinsed with H_2O, and air-dried. The transferred protein on the PVDF membrane was subjected to sequence analysis using an Applied Biosystems 470A gas-phase protein sequencer.

DNA Sequencing Strategy. The sequencing of the DNA of interest in pXT-1 (Xu et al., 1991) was performed by the dideoxynucleotide method of Sanger et al. (1977) using T7 polymerase and 7-deaza-dGTP instead of dGTP. Deletion mutants for nucleotide sequencing were created by using the *Exo*III/mung bean deletion kit from Stratagene. The universal primers, T3 and T7, and internal specific primers (18 bases in length) were used in these experiments.

Nucleotide and Protein Sequence Analysis. As described previously (Xu et al., 1991), the University of Wisconsin Genetics Computer Group's software programs were used to analyze the sequence data (Devereux et al., 1984). The open reading frames were searched by the CODONPREFERENCE program. The comparison of polypeptides was carried out with the BESTFIT program. The FASTA program was used to search the Gen Bank/EMBL Sequence databases for sequences homologous to the 25-kDa subunit and to the open reading frames.

Other Analytical Procedures. Protein was estimated by the method of Lowry et al. (1951) or by the biuret method in the presence of 1 mg of sodium deoxycholate/mL (Gornall et al., 1949). Immunoblotting was carried out as previously described

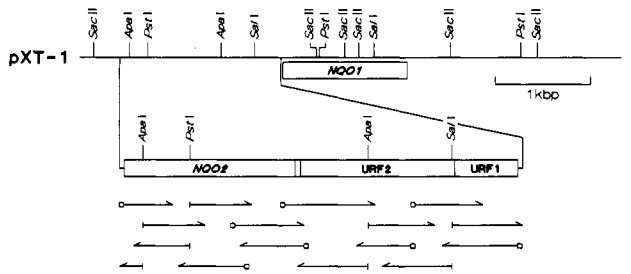

FIGURE 1: Restriction map and sequencing strategy of the *P. denitrificans* DNA fragment containing the 25-kDa subunit-coding region. *NQO1* and *NQO2* indicate the structural genes of the NADH-binding subunit and the 25-kDa subunit of the *Paracoccus* NDH-1, respectively. The direction and extent of the nucleotide sequences are shown by the length of the arrows. The universal and synthetic oligonucleotide primers (18-mer) used for sequencing are displayed by vertical bars and open boxes, respectively. Restriction fragments were ligated into the linker region of pBluescript KS plasmid.

(Yagi & Hatefi, 1988; Yagi & Dinh, 1990; Xu & Yagi, 1991). Details of these procedures are described in the figure legends.

Materials. Acrylamide, SDS, and Coomassie brilliant blue R-250 were from Bio-Rad; 7-deaza-dGTP and T7 polymerase were from Pharmacia LKB; [^{35}S]dATP was from Amersham.

Results and Discussion

Sequence Analysis. The pXT-1 carries the 5.7 kbp *Eco*RI DNA fragment which bears the *NQO1* gene encoding the NADH-binding subunit (Xu et al., 1991). As described previously (Xu et al., 1991), a nucleotide sequence homologous to the 24-kDa polypeptide of the bovine FP was found to be present 1.4 kbp upstream of the *NQO1* gene. This sequence codes for a *Paracoccus* 25-kDa subunit which thus appears to be a counterpart of the bovine 24-kDa subunit (see the introduction). The 25-kDa subunit was purified from *Paracoccus* NADH dehydrogenase complex by electroelution from SDS–polyacrylamide gels. The purified subunit was subjected to amino acid sequencing according to Matsudaira (1987). The sequence of the first 10 amino acids (MLRRLSPIQP) was determined. Using this information, we attempted to locate the nucleotide sequence coding for the N-terminal amino acid sequence of the 25-kDa subunit upstream of the region homologous to the bovine 24-kDa polypeptide. The DNA sequence coding for this primary structure was found 1.7 kbp upstream of the *NQO1* gene (the NADH-binding subunit). These results indicate that the structural gene (*NQO2*) encoding the 25-kDa subunit is present in this region of the pXT-1. Therefore, the nucleotide sequences of the *NQO2* gene

```
                                                    NQO2
-29  CGACATCGTTTTCGGAGAGGTTGACCGCTGATGCTGCGCCGTCTGTCCCCCATCCAGCCGGACTCCTTCGAGTTCACGCCCGCCAACCTC   60
                                                    M  L  R  R  L  S  P  I  Q  P  D  S  F  E  F  T  P  A  N  L

 61  GAATGGGCCCGTGCCCAGATGACCAAATACCCGGAAGGCCGCCAGCAATCGGCGATCATCCCGGTCCTGTGGCGCGCGCAGGAACAAGAG   150
     E  W  A  R  A  Q  M  T  K  Y  P  E  G  R  Q  Q  S  A  I  I  P  V  L  W  R  A  Q  E  Q  E

151  GGCTGGCTGAGCCGCCCCGCCATCGAATATTGTGCCGACCTTCTGGGCATGCCCTATATCCGGGCGCTGGAGGTCGCGACCTTCTATTTC   240
     G  W  L  S  R  P  A  I  E  Y  C  A  D  L  L  G  M  P  Y  I  R  A  L  E  V  A  T  F  Y  F

241  ATGTTCCAGCTGCAGCCGGTCGGCTCGGTCGCGCATATCCAGATCTGCGGCACCACGACCTGCATGATCTGCGGCGCCGAGGATCTGATC   330
     M  F  Q  L  Q  P  V  G  S  V  A  H  L  Q  I  C  G  T  T  T  C  M  I  C  G  A  E  D  L  I

331  CGGGTCTGCAAGGAAAAGATCGCGCCCGAGCCGCATGCGCTGTCGGCGGACGGCCGGTTCAGCTGGGAAGAGGTCGAATGCCTTGGCGCC   420
     R  V  C  K  E  K  I  A  P  E  P  H  A  L  S  A  D  G  R  F  S  W  E  E  V  E  C  L  G  A

421  TGCACCAACGCGCCCATGGCCCAGATCGGCAAGGATTTCTACGAGGACCTGACGGTCGAGAAGCTCGCGGCGCTGATCGACCGCTTTGCG   510
     C  T  N  A  P  M  A  Q  I  G  K  D  F  Y  E  D  L  T  V  E  K  L  A  A  L  I  D  R  F  A

511  GCGGGCGAGGTTCCGGTGCCGGGTCCGCAGAACGGCCGGTTCTCTGCCGAGGCCCTGGGCGGCCCGACGGCGCTGGCCGATCTGAAGGGC   600
     A  G  E  V  P  V  P  G  P  Q  N  G  R  F  S  A  E  A  L  G  G  P  T  A  L  A  D  L  K  G

601  GGCGAGGCGCATAACGCCAGCGTGGCGCGCGCCCTGCGGCTGGGCGACTCGATCAAGCGCATCGACGGCACCGAGGTGCCGATCACGACG   690
     G  E  A  H  N  A  S  V  A  R  A  L  R  L  G  D  S  I  K  R  I  D  G  T  E  V  P  I  T  T

                                                    URF2
691  CCTTGGCTTGCCACGCAGAACGGCGTCTAATCGAAAGGGAGCAGGCGGGTCGCGATGGAACGGGCAGAGTGCAACAGGAACTGCTGGATC   780
     P  W  L  A  T  Q  N  G  V  *                  M  E  R  A  E  C  N  R  N  C  W  I

781  TCGGCGGCAATCGCTGGGGTGGTGGTGCTGTTGTTCACCGCCGGCATCGGCGATCTGCATTGGCTGGCCGGGCTGTTCCTGGGCGTCGTG   870
     S  A  A  I  A  G  V  V  V  L  L  F  T  A  G  I  G  D  L  H  W  L  A  G  L  F  L  G  V  V

871  ACCTTCGTGCTGTTCGGGGGCGCTGATGGTCTGGCTGGTCTGCCATGAACGGCCCGAACTGTTCGAGGAGGGCGCCGGCCTGACCGGCACC   960
     T  F  V  L  F  G  A  L  M  V  W  L  V  C  H  E  R  P  E  L  F  E  E  G  A  G  L  T  G  T

961  GACTGGCAGCGCGCCGCCGTGGACCGCCAGCCCGAGACCCTGCTGGTCGGCGGCTCGCTGGGGCCCGAGCCCTTCACCTCGGAGGCGCAG   1050
     D  W  Q  R  A  A  V  D  R  Q  P  E  T  L  L  V  G  G  S  L  G  P  E  P  F  T  S  E  A  Q

1051 ATGCCCATCGTCGCCGGGGCGATGCCGGCAGAGCCGCTTGTCGAGCACCGGCCCGAACCCAAGCCTGCCGCTGCGGCCAAGGTCGCGCGG   1140
     M  P  I  V  A  G  A  M  P  A  E  P  L  V  E  H  R  P  E  P  K  P  A  A  A  A  K  V  A  R

1141 GGCGCCGACGATCTGAAGCGGATCAAGGGCATCGGCCCCAAGATCTCGGACTGGCTGAACGCGCAGGGCGTGACGCGCTATGACCAGATC   1230
     G  A  D  D  L  K  R  I  K  G  I  G  P  K  I  S  D  W  L  N  A  Q  G  V  T  R  Y  D  Q  I

1231 GCCGCATGGGACCCCGCCACCGTTGACGATTTCGCGCAGCGGCTGGGCCGGATGGGCGGGCGGATCGAGGCCGACGACTGGGTCGGCCAG   1320
     A  A  W  D  P  A  T  V  D  D  F  A  Q  R  L  G  R  M  G  G  R  I  E  A  D  D  W  V  G  Q

                                                    URF1
1321 GCGAAATTGCTGGCGGCGGGGGGCGAGACCGGCCATTCCCGCCGCGTCGACAAGGGCGAGGTCGGCTGATGGCCCGCCCGGCAGGACATG   1410
     A  K  L  L  A  A  G  G  E  T  G  H  S  R  R  V  D  K  G  E  V  G  *  M  A  R  P  A  G  H  A

1411 CGGCGTCGCAGGCGGGTGGGACCGCTGCGGCGGATGCCCGGCAGATGCGTCTGGTGGCGGCGGTGATCGCCGTCACCATGGCGCTGTGGC   1500
       A  S  Q  A  G  G  T  A  A  A  D  A  R  Q  M  R  L  V  A  A  V  I  A  V  T  M  A  L  W  L

1501 TGGGGGTGCAATGGCTGGGGGGCCAGCAGGACTGGCCCGCGAAATATGCCTTTCTCGCCGATCTGGCGGCGATTGGGGCCTTGATCTGGT   1590
       G  V  Q  W  L  G  G  Q  Q  D  W  P  A  K  Y  A  F  L  A  D  L  A  A  I  G  A  L  I  W  S

                                                    NQO1
1591 CGCTTTTGGTGACCTTTCGGATCTGGCGGCGGCGCAAGGCGTCCTCGCAGGGACAAGGATAAGACGATGCTGAACGATCAGGACCGGATC   1680
       L  L  V  T  F  R  I  W  R  R  R  K  A  S  S  Q  G  Q  G  *     M  L  N  D  Q  D  R  I
```

FIGURE 2: Nucleotide and deduced amino acid sequences of *NQO2*, URF1, URF2, and *NQO1*. The amino acid sequences determined by protein analysis are underlined. The putative Shine–Dalgarno sequence is doubly lined. The putative 2Fe–2S cluster (cluster N1b) binding sites are marked by dashed lines. *NQO1* shows the structural gene of the NADH-binding subunit of the *Paracoccus* NDH-1.

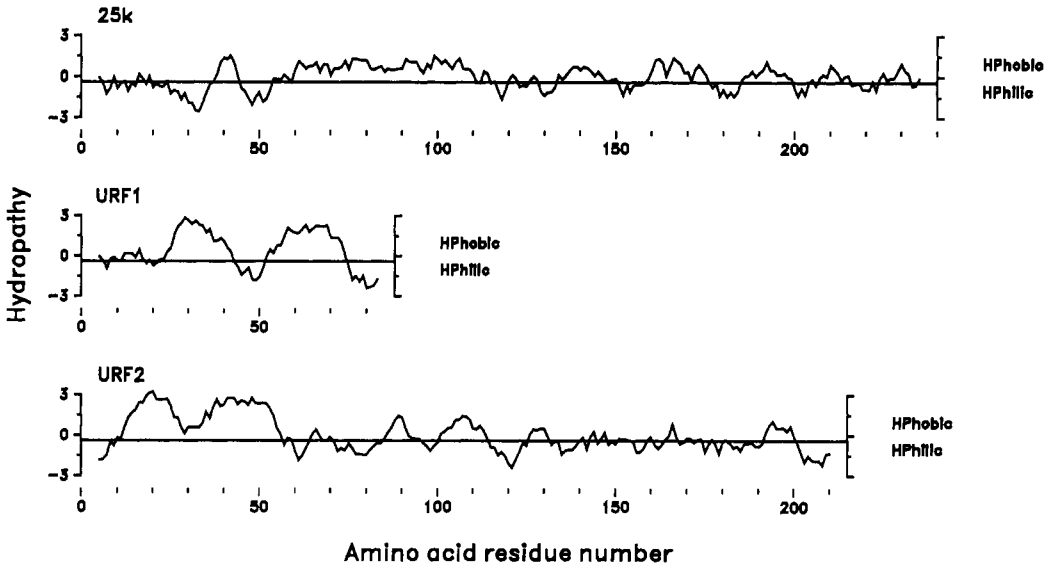

FIGURE 3: Hydropathy plots of the *NQO2*- (A), URF1- (B), and URF2 (C)-encoded polypeptides. On the basis of the data of Kyte and Doolittle (1982), hydropathy indices determined by using the PLOTSTRUCTURE program with a window of nine residues were calculated and plotted on the vertical axis. HPhobic and HPhilic display hydrophobic and hydrophilic areas, respectively. When the window was lengthened to 19 residues, no segment average of the *NQO2* polypeptide exceeded the value of 1.6, which is the critical value for membrane-spanning sequences (Kyte & Doolittle, 1982). On the other hand, the greatest values found in the two hydrophobic regions of the URF1 polypeptide were 2.06 and 2.04, respectively, and the greatest values found in the two hydrophobic regions of the URF2 polypeptide were 2.26 and 2.62, respectively, suggesting that these segments may be membrane-spanning.

and portion between the *NQO2* and the *NQO1* genes have been determined (Figure 1). As shown in Figure 2, the *NQO2* gene is composed of 717 bp and codes for 239 amino acid residues with a calculated molecular weight of 26 122, which is similar to the M_r 25 000 estimated by SDS–polyacrylamide gel electrophoresis (Yagi, 1986).

The CODONPREFERENCE program uses available genes from a particular species to determine the preference of codons used by that species. The program can then be used to look for similar codons in other regions of the gene. Using this program with the codon usage table of the available *Paracoccus* genes (Xu et al., 1991; Kurowski & Ludwig, 1987; Raitio et al., 1987, 1990; Van Spanning et al., 1990; Harms et al., 1987), we identified coding sequences in the DNA sequences in the region between *NQO1* and *NQO2*. The positions of potential initiation and stop codons were also determined as a further aid to delineating potential genes. Two open reading frames were detected between the *NQO1* and *NQO2* genes and were designated URF1 and URF2. No open reading frames were evident on the complementary strand (data not shown). As seen in Figure 2, URF1 and URF2 are composed of 261 and 642 bp and code for 87 and 214 amino acid residues, respectively. The calculated molecular weights of possible proteins encoded by URF1 and URF2 are 9277 and 22 906, respectively.

As shown in Figure 2, the initiation codons of *NQO2*, URF1, and URF2 were all found to be preceded by sequences resembling the ribosome-binding site (Shine–Dalgarno sequence) (Shine & Dalgarno, 1975). Furthermore, the G-C content of the *NQO2* gene (67.1%), URF1 (69.0%), and URF2 (70.6%) was high as was also found to be the case for other genes of *P. denitrificans* (Xu et al., 1991). Our nucleotide sequence analyses show that the order of the structural genes of the *Paracoccus* NDH-1 is *NQO2*, URF2, URF1, and *NQO1*, followed by the region homologous to the bovine 75-kDa polypeptide (Xu et al., 1991).

As seen in Table II, the *NQO1* and *NQO2* products have sequence similarity, respectively, to residues 186–602 and 1–186 of the α-subunit of the NAD-linked hydrogenase.

Table I: Amino Acid Composition of the 25-kDa (*NQO2*) Subunit of the *Paracoccus* NADH Dehydrogenase Complex

amino acid	by amino acid analysis	from DNA sequence
alanine	29.4	30
arginine	13.1	14
aspartic acid	17.0	15
glutamic acid	33.0	31
glycine	20.9	19
histidine	1.5	3
isoleucine	13.2	15
leucine	21.8	21
lysine	6.5	7
methionine	3.0	6
phenylalanine	8.9	9
proline	15.3	17
serine	10.5	10
threonine	14.8	13
tyrosine	5.3	5
valine	12.7	12

Recently, the bovine 75-kDa polypeptide has reported to be similar to the γ-subunit of the NAD-linked hydrogenase (Xu et al., 1991; Pilkington et al., 1991). In addition, the fact that the arrangements of these structural genes of the *Paracoccus* NDH-1 and the NAD-linked hydrogenase resemble each other may support a hypothesis of an evolutionary relationship between the water-soluble fraction of NDH-1 and the NAD-linked hydrogenase (Xu et al., 1991; Pilkington et al., 1991; Patel et al., 1991).

Structure of the 25-kDa Subunit. The amino acid composition derived from the *NQO2* gene corresponded to the results of the amino acid analysis conducted on the purified 25-kDa subunit (see Table I). The isoelectric point of this subunit was determined experimentally to be approximately 4.8 (Yagi & Dinh, 1990), which agreed with the value (4.8) calculated from the deduced amino acid composition. Although several moderately hydrophobic domains are present in the 25-kDa subunit (residues 35–45, 60–110, 135–145, 160–175, and 185–200), these domains do not appear to be sufficiently hydrophobic to form a membrane-spanning segment (see Figure 3).

Table II: Comparison Matrices for the *NQO1* and *NQO2* Subunits of the Energy-Transducing NADH–Ubiquinone Oxidoreductases and the α-Subunit of the NAD-Linked Hydrogenase of *A. eutrophus* H16

		% identity				
subunit	source	(1)	(2)	(3)	(4)	(5)
NQO2	(1) *P. denitrificans*	100				
	(2) human	41.3	100			
	(3) bovine	40.9	98.6	100		
	(4) rat	42.3	94.9	93.5	100	
	(5) *A. eutrophus*[a]	20.6	26.5	27.1	27.0	100
NQO1	(1) *P. denitrificans*	100				
	(3) bovine	64.3		100		
	(5) *A. eutrophus*[b]	33.3		34.5		100

[a] Residues 1–186 of the α-subunit of NAD-linked hydrogenase. [b] Residues 186–602 of the α-subunit of NAD-linked hydrogenase.

```
Paracoccus    1  MLRRLSPIQ..PDSFEFTPANLEWARAQMTKYPEGRQQSAIIPVLWRAQE  48
                 :::| .| .  ...|:|||.|..:  | :..||||:...|::|||  ||
Bovine       39  FVHRDTPENNPETPFDFTPENYKRIEAIVKNYPEGHKAAAVLPVLDLAQR  88

Paracoccus   49  QEGWLSRPAIEYCADLLGMPYIRALEVATFYFMFQLQPVGSVAHIQICGT  98
                 |:|||. .|:: .|::|.:| :|..||||||  |:. .|||.  |||:|.|
Bovine       89  QNGWLPISAMNKVAEILQVPPMRVYEVATFYTMYNRKPVGKY.HIQVCTT 137

Paracoccus   99  TTCMICGAEDLIRVCKEKIAPEPHALSADGRFSWEEVECLGACTNAPMAQ 148
                 |.||: ..:.:: ....|:: .. . .:|  |.: |||||||||.|||||.|
Bovine      138  TPCMLRNSDSILEAIQKKLGIKVGETTPDKLFTLIEVECLGACVNAPMVQ 187

Paracoccus  149  IGKDFYEDLTVEKLAALIDRFAAGEVPVPGPQNGRFSAEALGGPTALADL 198
                 |..::|||||...:..:|| : ||.:| |||..|||||.|: || |.|.:
Bovine      188  INDNYYEDLTPKDIEEIIDELKAGKIPKPGPRSGRFSCEPAGGLTSLTEP 237

Paracoccus  199  KGGEAHNASVA 209
                 . |.: .. .:
Bovine      238  PKGPGFGVQAG 248
```

FIGURE 4: Comparison of the amino acid sequences of the *NQO2* subunit and the bovine 24-kDa subunit. The upper sequence is the deduced amino acid sequence of the 25-kDa subunit of the *Paracoccus* NDH-1; the lower sequence is the deduced amino acid sequence of the 24-kDa subunit of the FP of bovine complex I. The comparison was carried out by the BESTFIT program. Vertical bars connect residues that are the same; colons and solid dots are put between residues whose comparison values are greater or equal to 0.5 and 0.1, respectively.

When the amino acid sequence of the *Paracoccus* 25-kDa subunit was used to search the GenBank/EMBL database, this subunit was found to have significant similarity (58% similarity and 41% identity) to the 24-kDa subunits of the mammalian mitochondrial NADH–Q oxidoreductase (von Bahr-Lindström et al., 1983; Pilkington & Walker, 1989; Nishikimi et al., 1988; Chomyn & Lai, 1989) as shown in Table II and Figure 4. Bovine complex I can be resolved into a water-soluble fraction and a water-insoluble fraction (HP) (Hatefi et al., 1985; Hatefi, 1985). The water-soluble fraction can be separated into a flavoprotein fraction (FP) and an iron–sulfur protein fraction (IP) by ammonium sulfate fractionation (Han et al., 1988, 1989; Yagi & Hatefi, 1988). FP is composed of three subunits of M_r 51K, 24K, and 9K (Galante & Hatefi, 1979; Hatefi et al., 1985; Hatefi, 1985) which can be further resolved into a 51-kDa subunit fraction and a 24-kDa + 9-kDa subunit fraction. Both fractions are water-soluble (Ragan et al., 1982; Hatefi et al., 1985; Hatefi, 1985). Furthermore, topological studies using a monospecific antibody suggest that the bovine 24-kDa subunit is exposed to the matrix side but not to the cytosolic side of the inner mitochondrial membrane (Han et al., 1988). These results together with the apparent lack of a hydrophobic membrane-spanning sequence support the hypothesis that neither the *Paracoccus* 25-kDa subunit nor its mammalian counterpart is an integral membrane protein.

Bovine FP contains two EPR-visible FeS clusters, the binuclear center N1b and the tetranuclear center N3. Both the 51-kDa subunit fraction and the 24-kDa + 9-kDa subunit fraction contain non-heme iron and acid-labile sulfide (Ragan et al., 1982). On the basis of the non-heme iron content, the tetranuclear cluster is in the 51-kDa subunit, and the binuclear cluster is in the 24-kDa + 9-kDa subunit fraction (Ragan et al., 1982). Other data described previously (Xu et al., 1991; Hatefi et al., 1985) also suggest that the cluster N3 of bovine complex I may be assigned to the 51-kDa subunit. Cluster N1b may then be located in the 24-kDa + 9-kDa subunit fraction. Since the 9-kDa subunit does not contain any cysteine residue (Masui et al., personal communication; Skehel et al., 1991), the cluster N1b would most likely be found in the 24-kDa subunit. The cluster N1b ($E_{m,7}$ = −260 mV) also appears to be present in the *Paracoccus* NADH–UQ oxidoreductase segment (Meinhardt et al., 1987). An FP-type fraction of the *Paracoccus* NADH dehydrogenase complex is composed of the NADH-binding subunit (50 kDa) and the 25-kDa subunit (George & Ferguson, 1984). A subunit corresponding to the 9-kDa subunit was not detected in this fraction. Therefore, it is conceivable that the 25-kDa subunit bears the cluster N1b. If that is the case, it should be possible to locate the residues involved in binding the cluster.

The 2Fe–2S clusters of bacterial and chloroplast ferredoxin are known to bind to four cysteine residues which construct the sequence motif $CxxxxCxxC(x)_{20}C$ (Yasunobu & Tanaka, 1980). However, such a sequence motif is not present in the deduced primary structure of the *NQO2* subunit or its mammalian counterparts. In the case of Rieske FeS protein of complex III which also bears a binuclear FeS cluster, all the subunits for which amino acid sequences are available have two conserved histidine residues and four conserved cysteine residues in the following arrangement near the C-terminus: $CxHxxC(x)_{13-14}CxCH$. Recently, Britt et al. (1991) reported that the results of electron spin-echo envelope modulation studies of the Rieske FeS centers of the cytochrome b_6f complex of spinach and complex III of bovine heart mitochondria, *Rhodospirillum rubrum* and *Rhodobacter sphaeroides* strain R-26, suggest a coordination of nitrogen ligands from two histidines to the Reiske FeS centers. Therefore, we might consider not only cysteine but also histidine residues as can-

didates for N1b binding sites.

The predicted primary structure of the *Paracoccus* 25-kDa subunit contains seven cysteines and three histidines. Before the primary structure of this subunit was determined, the only available primary structures of the 24-kDa subunits of NADH–UQ oxidoreductases were from human, bovine, and rat mitochondria. As seen in Table II, these three subunits show more than 93% identity to each other, and all five cysteines and three histidines are conserved except for one additional cysteine in the rat 24-kDa subunit. A comparison of the *Paracoccus* 25-kDa subunit with the mammalian 24-kDa subunits (Figure 4) reveals the presence of four conserved cysteines (C_{96}, C_{101}, C_{137}, and C_{141}) and one conserved histidine (H_{92}). Four of these five residues have the potential to be associated with the cluster N1b. The fact that the sequence $C_{137}LGAC_{141}$ is similar to the sequence $C_{160}LEAC_{164}$ of the putative FeS cluster binding site of the FeS subunit of the *Bacillus subtilis* succinate dehydrogenase (Phillips et al., 1987) suggests that C_{137} and C_{141} may be associated with cluster N1b. Further clues to the identity of the residues responsible for binding cluster N1b came from both EPR work and comparison of redox potentials. The g_{av} of the Rieske FeS cluster is approximately 1.90, whereas g_{av} values for most ferredoxins is in the range 1.94–1.98. Furthermore, the midpoint potential of the Rieske FeS cluster (260–280 mV) is 400–600 mV more positive than those of ferredoxins. As described above, the ligands of the Rieske FeS cluster are two cysteines and two histidines, as opposed to four cysteine ligands in ferredoxins. Since histidine is less strongly electron-donating than sulfur, this difference in ligand could account for the high midpoint potential of the Rieske FeS cluster and may also account for the atypical *g* value. That one or more of the ligands of this FeS cluster must be less electron-donating than sulfur was predicted on theoretical grounds by Blumberg and Peisach (1974). The fact that the E_m value for bovine N1b is −245 mV (Ohnishi, 1979) and for *Paracoccus* N1b is −260 mV (Meinhardt et al., 1987) and that the g_{av} values of the N1b centers of bovine and *P. denitrificans* are approximately 1.94 (Ohnishi, 1979; Meinhardt et al., 1987) could suggest that the N1b center, like the ferredoxins, binds to the four conserved cysteines. Nevertheless, this assignment needs to be verified directly.

Structures of URF1 and URF2. The polypeptides encoded by the remaining open reading frames (URF1 and URF2) were not found in the isolated *Paracoccus* NADH dehydrogenase complex. Furthermore, no significant homology was found between URF1 and URF2 and any of the sequences in the GenBank (release 66.0, 12/90)/EMBL (release 26.0, 2/91) database. Nor was any significant relationship found between URF1 and URF2 and the known amino acid sequences of any of the other proteins of NADH–UQ oxidoreductases.

URF1 appears to code for an integral membrane protein as shown in Figure 3. The hydropathy plot of the URF1 product suggests the presence of two highly hydrophobic regions in this protein. These regions appear long enough to span a membrane. The calculated isoelectric point of the URF1 product is 12.2, suggesting that if this protein is synthesized, it is extremely basic.

The putative polypeptide encoded by URF2 contains the hydrophobic domain located in residues 1–56 (see Figure 3) which may be long enough to span the membrane twice. The calculated isoelectric point of the URF2 product is 5.02. It is possible that the URF2 product may be involved in the transfer of FeS clusters to the NADH dehydrogenase complex because the sequence C_6xxxC_{10} of the URF2 product is capable of chelating Fe ion. Another possibility stems from the relationship between the URF2 and ORF3 genes located between the *Paracoccus COXII* and *COXIII* genes (Raitio et al., 1987). Not only are the hydropathy plots of these proteins similar, but meaningful sequence similarity (47% similarity and 23.6% identity) was detected in the URF2 and ORF3 gene products:

URF2	14–22	AAIAGVVVL
		| | | | | | |
ORF3	14–22	AMLAGVVVL
URF2	29–38	GDLHWLAGLF
		| | | | |
ORF3	24–33	GALSWAAVPF
URF2	192–197	QAKLLA
		| | | |
ORF3	182–187	QAALDA

Recently, Tzagoloff and co-workers (Tzagoloff et al., 1990) reported that the *COX11* product of *Saccharomyces cerevisiae* has sequence identity with the *Paracoccus* ORF3 product and the *COX11* product is essential for assembly of the cytochrome oxidase complex although the *COX11* product is not a subunit of this enzyme complex. As described previously (Xu et al., 1991), a gene located in an operon carrying a particular enzyme complex should code for essential polypeptides of that enzyme complex. Together, these data suggest that the URF2 product may be required for the assembly of the *Paracoccus* NADH dehydrogenase complex. The expression of URF1 and URF2 and studies on URF deletion mutants may provide useful information about the structural and/or functional role of these two predicted polypeptides.

ACKNOWLEDGMENTS

We thank Tri M. Dinh and Nanette C. Wong for isolation of the 25-kDa subunit of the *Paracoccus* NADH dehydrogenase complex and Dr. Youssef Hatefi for critical reading of the manuscript.

REFERENCES

Blumberg, W. E., & Peisach, J. (1974) *Arch. Biochem. Biophys. 162*, 502–512.

Britt, R. D., Sauer, K., Klein, M. P., Knaff, D. B., Kriauciunas, A., Yu, C.-A., Yu, L., & Malkin, R. (1991) *Biochemistry 30*, 1892–1901.

Chomyn, A., & Lai, S. S.-A. T. (1989) *Curr. Genet. 16*, 117–125.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Galante, Y. M., & Hatefi, Y. (1979) *Arch. Biochem. Biophys. 192*, 559–568.

George, C. L., & Ferguson, S. J. (1984) *Eur. J. Biochem. 143*, 567–573.

Gornall, A. G., Bardawill, C. J., & David, M. M. (1949) *J. Biol. Chem. 177*, 751–766.

Han, A.-L., Yagi, T., & Hatefi, Y. (1988) *Arch. Biochem. Biophys. 267*, 490–496.

Han, A.-L., Yagi, T., & Hatefi, Y. (1989) *Arch. Biochem. Biophys. 275*, 166–173.

Harms, N., De Vries, G. E., Maurer, K., Hoogendijk, J., & Stouthamer, A. H. (1987) *J. Bacteriol. 169*, 3969–3975.

Hatefi, Y. (1985) *Annu. Rev. Biochem. 54*, 1015–1069.

Hatefi, Y., Ragan, C. I., & Galante, Y. M. (1985) in *The Enzymes of Biological Membranes* (Martonosi, A. N., Ed.) pp 1–70, Plenum Publishing, New York.

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