Journal of Bioenergetics and Biomembranes, Vol. 30, No. 3, 1998

Cloning and Sequence Analysis of the Structural Gene for the bc_1-Type Rieske Iron-Sulfur Protein from *Thermus thermophilus* HB8

Domenico L. Gatti,[1,2] George Tarr,[2] James A. Fee,[2,3] and Sharon H. Ackerman[4]

Received September 8, 1997; accepted January 5, 1998

The structural gene encoding the Rieske iron-sulfur protein from *Thermus thermophilus* HB8 has been cloned and sequenced. The gene encodes a protein of 209 amino acids that begins with a hydrophilic N-terminus followed by a stretch of 21 hydrophobic amino acids that could serve as a transmembrane helix. The remainder of the protein has a hydrophobicity pattern typical of a water-soluble protein. A phylogenetic analysis of 26 Rieske proteins that are part of bc_1 or b_6f complexes shows that they fall into three major groups: eubacterial and mitochondrial, cyanobacterial and plastid, and five highly divergent outliers, including that of *Thermus*. Although the overall homology with other Rieske proteins is very low, the C-terminal half of the *Thermus* protein contains the signature sequence CTHLGC-(13X)-CPCH that most likely provides the ligands of the [2Fe-2S] cluster. It is proposed that this region of the protein represents a small domain that folds independently and that the encoding DNA sequence may have been transferred during evolution to several unrelated genes to provide the cluster attachment site to proteins of different origin. The role of individual residues in this domain of the *Thermus* protein is discussed vis-a-vis the three-dimensional structure of the bovine protein (Iwata *et al.*, 1996 *Structure* **4**, 567–579).

KEY WORDS: Rieske iron-sulfur protein; *Thermus thermophilus*; phylogeny of Rieske proteins.

INTRODUCTION

The Rieske iron-sulfur protein is an essential component of the respiratory cytochrome bc_1 (Gennis *et al.*, 1993) and of the photosynthetic cytochrome b_6f complexes (Cramer *et al.*, 1994). Within these integral membrane protein complexes, the [2Fe-2S] cluster of the Rieske protein acts as a one-electron oxidant of reduced quinone (QH_2) bound at the Q_0 site and a one-electron reductant of cytochrome c_1 (or f). In accord with this role, in the crystal structures of the bovine and chicken cytochrome bc_1 complex (Xia *et al.*, 1997; Berry *et al.*, 1997) the Rieske protein occupies a position intermediate between cytochrome b and cytochrome c_1. Site-directed mutagenesis studies (Davidson *et al.*, 1992) and spectroscopic examination of site-specific, isotope-substituted Rieske-type proteins (Cline *et al.*, 1985; Gurbiel *et al.*, 1996) led to the general conclusion that the novel EPR spectrum of the Rieske center (Rieske *et al.*, 1964) originates from unique structural features of the Fe/S cluster, in which one iron is coordinated to two cysteine and the other iron is coordinated to two histidine residues: $(Cys\text{-}S\text{-})_2$[2Fe-2S]$(\text{-}N\text{-}His)_2$. A high-resolution, three-dimensional X-ray structure of the water-soluble fragment of the

[1] Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, Michigan 48201.
[2] Department of Biological Chemistry and Biophysics Research Division, The University of Michigan Medical School, Ann Arbor, Michigan 48109.
[3] Department of Biology, University of California at San Diego, La Jolla, California 92093.
[4] Department of Surgery and Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, Michigan 48201.

0145-479X/98/0600-0223$15.00/0

bovine Rieske iron-sulfur protein, which confirms the predictions of the spectroscopic and mutagenesis studies, was described recently by Michel and coworkers (Iwata *et al.*, 1996; see also Zhang *et al.*, 1996).

The isolation and properties of a Rieske-type protein from plasma membranes of *Thermus thermophilus* HB8 were described by Fee *et al.* (1984). The purified protein, which was detergent solubilized, had a molecular weight of ~20 kDa and appeared to contain 2 [2Fe-2S] clusters.[5] The protein lacked an EPR spectrum in the oxidized form and required one electron to reduce a [2Fe-2S] cluster, after which it exhibited an EPR spectrum that accounted for one unpaired electron per 2Fe. In general, the *T. thermophilus* Rieske protein displayed spectral properties highly similar to purified samples of the bovine Rieske protein (Trumpower *et al.*, 1980; Fee *et al.*, 1984). Notably, although other bacterial sources are known to contain a bc_1 complex with a Rieske protein (Krinner *et al.*, 1982; Gabellini and Sebald, 1986; Yang and Trumpower, 1986), the *T. thermophilus* Rieske protein was not observed in complex with cytochromes *b* or *c*, and there have been no formal reports of the isolation of a bc_1 complex from the *Thermus* species.[6]

In this report we describe the isolation, sequence, and translation of the structural gene encoding the Rieske iron-sulfur protein from *T. thermophilus*. Our analyses suggest that the *Thermus* Rieske iron-sulfur protein is a highly divergent member of the superfamily of the bc_1 type Rieske iron-sulfur proteins.

MATERIALS AND METHODS

T. thermophilus HB8 cells were grown as described by Findling *et al.* (1984), and the Rieske protein was isolated as described by Fee *et al.* (1984). Amino acid composition, cleavage of the protein by cyanogen bromide, and N-terminal protein sequencing were carried out as described by Tarr (1982). Chromosomal DNA was isolated by the methods of Mather and Fee (1990); the DNA oligonucleotide probe used for cloning was designed taking into account the codon preferences of *T. thermophilus* HB8 (cf. Gribskov *et al.*, 1984; Mather *et al.*, 1991). DNA fragments were isolated and cloned as described previously (Mather *et al.*, 1992). DNA was sequenced on both strands using the method of Maxam and Gilbert (1977) or that of Sanger *et al.* (1977). The isolation of plasmid DNA, digestion of DNA with restriction enzymes, electrophoresis of restriction fragments on both analytical and preparative agarose gels, and ligation of restriction fragments were done by standard procedures (Maniatis *et al.*, 1982). The DNA sequence and deduced amino acid sequences were analyzed using the programs of the University of Wisconsin Genetics Computer Group (Devereux *et al.*, 1984). The programs of Feng and Doolittle (1990) were used for phylogenetic analyses.

RESULTS AND DISCUSSION

Nucleotide Sequence of rpt

The protein isolated from *Thermus* cells was found to have an N-terminal amino acid sequence SLRPREEVTP. Treatment with cyanogen bromide followed by liquid chromatography yielded two major peptides, one corresponding to the N-terminus of the isolated protein (see above), while the other had the N-terminal sequence (M)DPKTKVVKSG. Taking advantage of the strong preference of *Thermus* to utilize either a G or C in the third position of the codon (cf. Mather *et al.*, 1991), a DNA probe (5′-ATGGACCCCAAGACCAAGGTGGTGAAGTCCGG-3′) was designed by reverse translation of the latter (internal) protein sequence to search for the Rieske protein structural gene. This "consensus" probe was used to screen *Thermus* chromosomal DNA that had been digested with several different restriction enzymes. A digest with *Bam*HI produced a band of ~4.2 kb that hybridized with the probe. Since it was deemed likely that a fragment of this size would contain the entire gene encoding the Rieske protein, a small genomic library was prepared by isolating *Bam*HI DNA fragments migrating in the agarose gel in the range 3.9–4.5 kb. These were cloned into a *Saccharomyces cerevisiae* / *Escherichia coli* shuttle vector (Hill *et al.*, 1986) and the ligation mixture was used to transform competent *E. coli* RR1 cells. Approximately 1% of the resulting

[5] Based on the amino acid sequence presented here the *Thermus* Rieske protein appears to contain only one cluster per molecule. Therefore, the conclusion reached earlier by Fee *et al.* (1984) is incorrect. A review of the original data has not pointed to a possible source for this error.

[6] Zimmermann (1988) described a cytochrome *b* from *Thermus* plasma membranes having a molecular weight of ~30 kDa, containing 2 heme B and occurring in relatively large amounts. This protein may correspond to the cytochrome *b* of the *Thermus* bc_1 complex.

transformants were found to harbor the Rieske gene by colony hybridization (Grunstein and Hogness, 1975) with the end-labeled oligonucleotide probe. Initial mapping of the 4.2 kb *Bam*HI fragment was carried out by restriction digest and Southern analysis (cf. Maniatis *et al.*, 1982). Genomic maps of the original clone (pRPT/T1) and of the subclone (pRPT/ST1) used for sequencing are shown in the upper portion of Fig. 1.

Several attempts were made to obtain sequence information by the Sanger dideoxy procedure (Sanger *et. al.*, 1977), but only short spans of sequence were obtained. The problem was not relieved by the use of thermostable DNA polymerases and high temperatures for both the annealing and elongation steps or by the use of deoxyinosine or 7-deaza-deoxyguanosine derivatives (cf. Mather *et al.*, 1991). The high G + C content of *Thermus* DNA, noted in previous publications (Mather *et al.*, 1991, 1993), most likely leads to formation of unusually stable tertiary structures in these particular samples. Therefore, the chemical procedure of Maxam and Gilbert (1977) was used to obtain the sequence reported herein. The restriction map of the chromosome region encompassing *rpt* and the sites chosen for 5′ end labeling are shown in the lower portion of Fig. 1.

The DNA and deduced amino acid sequences of the *Thermus* Rieske protein are shown in Fig. 2. The underlined amino acid sequence shown in boldface corresponds to that of the internal peptide used to synthesize the probe for *rpt*. The nucleotide sequence

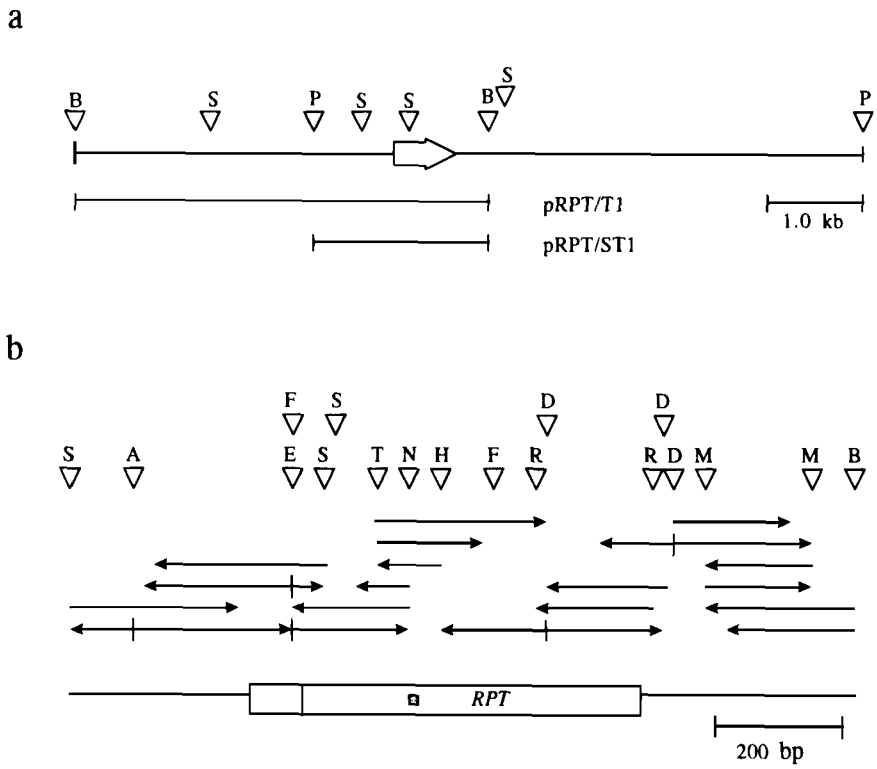

Fig. 1. Gene structure and sequencing strategy of the Rieske iron-sulfur protein gene (*rpt*) in *Thermus thermophilus* HB8. (a) The upper line represents the region of the *Thermus* chromosome containing *rpt* (indicated by a horizontal arrow). The positions of the restriction sites shown above the line were determined by Southern analysis of chromosomal DNA digests using the 5′-labeled probe complementary to the *rpt* sequence. The regions encompassed by the original clone pRPT/T1 and by the subclone pRPT/ST1 used for sequencing are shown below the genomic map. (b) The direction and length of the fragment sequenced by the Maxam–Gilbert procedure are indicated by horizontal arrows. Restriction sites used for 5′ end labeling are marked in the upper part of the panel. The vertical line inside the *rpt* gene (open box) indicates the codon position of the N-terminus of the purified protein. A filled square marks the region of the gene that hybridizes with the synthetic probe used for cloning. B, *Bam*HI; S, *Sma*I; P, *Sph*I; A, *Acc*I; E, *Eco*RII; F, *Sfa*NI; T, *Taq*I; N, *Nco*I; H, *Hae*III; R, *Rsa*I; D, *Dde*I; M, *Mbo*I.

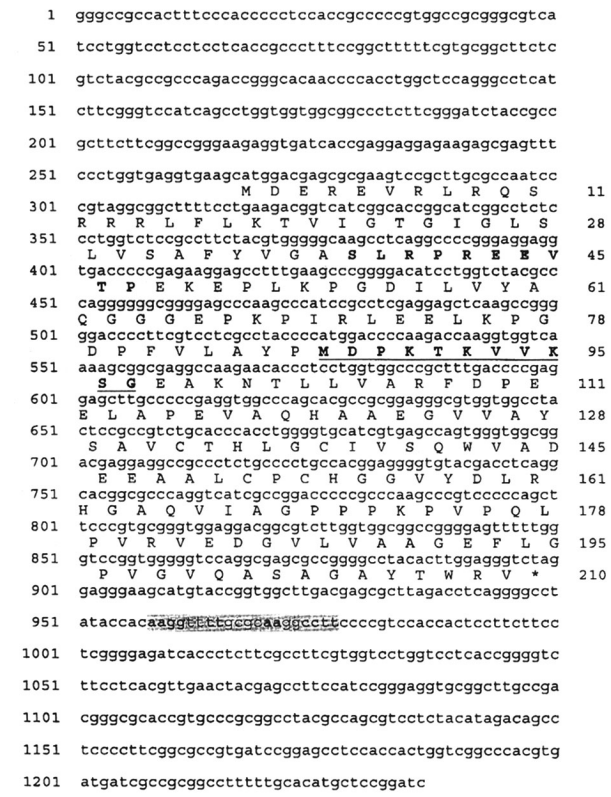

Fig. 2. Sequence of the *T. thermophilus* Rieske protein gene (*rpt*) and its translation. The predicted amino acid sequence is shown below the nucleotide sequence. Nucleotides are numbered on the left and amino acids on the right. Translation begins at 268 and ends at 900. The amino acid sequences shown in boldface are those that were determined by Edman sequencing of cyanogen bromide peptide fragments of the purified protein; the underlined amino acid sequence was used for reverse translation to obtain the nucleotide probe for sequencing. The palindromic nucleotide sequence distal to the 3′ end of the *rpt* gene, which begins at nucleotide 958, is highlighted in a shaded box. The DNA sequence has been deposited in GenBank (Accession Number AF037403).

of the probe matches the gene sequence except at positions 549 (C for G), 552 (A for G), 553 (A for T), and 554 (G for C). Plots of codon utilization frequency (not presented) show clearly that an open reading frame, dominated by codons most frequently used in *Thermus* structural genes (cf. Keightley *et al.*, 1995; Mather *et al.*, 1993), begins at position 268 and continues to 897 followed by a termination codon (tag) at 898. Examination of sequences upstream from position 268 revealed no evidence for sequence patterns suspected to be involved in regulation of transcriptional initiation (cf. Maseda and Hoshino, 1995). Nucleotides 958–977 constitute a palindrome structure that could participate in transcription or translation termination under certain growth conditions. Additional information suggests that this sequence is more likely involved in regulating translation: a second reading frame begins at nucleotide 910,[7] which suggests that *rpt* is part of an operon, as has been observed in other bacterial chromosomes (Gabellini and Sebald, 1986; Kurowski and Ludwig, 1987; Sone *et al.*, 1996).

Sequence Analyses and Evolutionary Relationships

The predicted translation product of *rpt* is 209 amino acids in length. Examination of the deduced amino acid sequence of the *Thermus* Rieske protein reveals an N-terminal, hydrophilic region (residues 1–18) that may contain targeting information important for the localization of the protein in the cell. The putative sorting signal is followed by a 21-residue stretch of relatively hydrophobic amino acids (19–39) that could form a transmembrane helix. The remainder of the protein has a hydrophobicity pattern (Kyte–Doolittle plot, not shown) expected for a water-soluble protein. The sequence of the protein isolated from *Thermus* membranes begins at amino acid 38, which is immediately distal to this hydrophobic region. Therefore, we speculate that the isolated protein, which is obtained in relatively small amounts[8] (Fee *et al.*, 1984), represents a fragment resulting from proteolysis of the mature protein during cell work-up.

BLASTp searches (Altschul *et al.*, 1990) of sequence databases returned a number of "hits" described as Rieske proteins, which show good similarity scores. Interestingly, a search using only the N-terminal portion (excluding the active site region) of the *Thermus* sequence returned no hits, a result that called for a more extensive comparison of the sequences of Rieske proteins from different sources. Twenty-six Rieske protein sequences of bacterial, cyanobacterial, mitochondrial, and chloroplast (plastid) origin were subjected to a phylogenetic analysis using the programs and procedures of Feng and Doolit-

[7] We have translated this partial open reading frame and find it to correspond closely in sequence to a cytochrome *b* as found in other operons containing the structural genes for proteins of the bc_1 and or b_6f complexes.

[8] The yield of Rieske protein in a standard preparation is 0.3 – 0.6 μmol compared with nearly 2 μmol of cytochrome caa_3 present after roughly the same amount of manipulation (Fee *et al.*, 1984; Yoshida *et al.*, 1984).

tle (1990). The resulting tree (cf. Fig. 3 and below) revealed the presence of two strongly diverged clades and a number of even more divergent sequences referred to here as "outliers." Those from eubacteria and mitochondria form one clade while those from cyanobacteria and plastids form the other clade. Within the chloroplast-cyanobacterial clade the sequence from *Anabaena* is the most divergent with only ~45% identity to other members of the clade. The outlier group contains proteins from the green sulfur bacterium *Chlorobium limicola*; two eubacteria, *Bacillus subtilis* and *Thermus thermophilus*; and the archaebacterium, *Sulfolobus acidocaldarius*. This analysis largely confirms the conclusions of an earlier study by Schmidt *et al.* (1996), which was based on a smaller set of Rieske sequences.

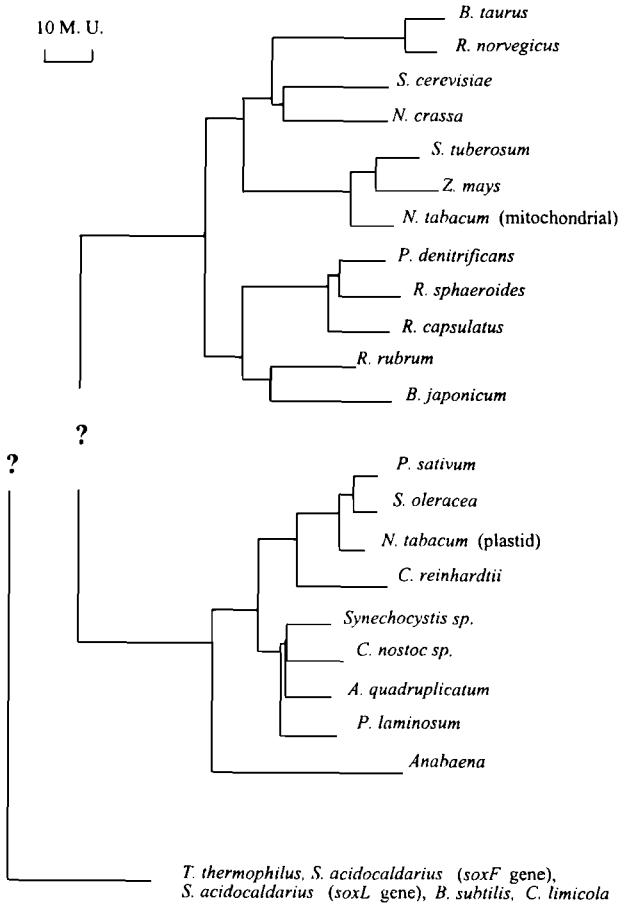

Fig. 3. Putative phylogenetic tree of the Rieske proteins. The tree is based on the alignment of sequences that were truncated for N- and C-terminal sequences to avoid contributions from nonoverlapping ends. Calculations were as described by Feng and Doolittle (1990). The matrix unit (M.U.) is defined as $D = -\ln S$, where S is a measure of sequence similarity. The sequences analyzed are from *Bos taurus, Rattus norvegicus, Saccharomyces cerevisiae, Neurospora crassa, Solanum tuberosum, Zea mays, Nicotiana tabacum* (mitochondrial), *Paracoccus denitrificans, Rhodobacter sphaeroides, Rhodobacter capsulatus, Rhodospirillium rubrum, Bradyrhizobium japonicum, Pisum sativum, Spinacia oleracea, Nicotiana tabacum* (plastid), *Chlamydomonas reinhardtii, Synechocystis sp., Cyanobacterium nostoc sp., Agmenellum quadruplicatum, Phormidium laminosum, Anabaena, Thermus thermophilus, Sulfolobus acidocaldarius* (*soxF* and *soxL* genes) *Bacillus subtilis*, and *Chlorobium limicola*.

Hauska *et al.* (1988) examined Rieske protein sequences available in 1987 and reported significant differences between mitochondrial and plastid sequences ($\leq$ 20% identity). An apparent lack of conservation was noted also in the C-terminal portion of the sequences. These authors suggested that "possibly the Rieske FeS protein has different ancestral domains." We explore that hypothesis here, and suggest a possible evolutionary history of the Rieske proteins. Sequence alignments of the active site and C-terminal region of Rieske proteins from four representative organisms of each of the three phylogenetic groups identified in the present study are shown in Fig. 4. Sequence similarity in the active site region is dominated by the conserved cysteine and histidine residues that serve as ligands (*) to the [2Fe-2S] cluster and that form a novel disulfide linkage (#) (Iwata *et al.*, 1996). These, along with very few other residues (see below), are conserved in the alignment of all 26 sequences (identities are noted by the symbol "|" above the single letter code). Within the mitochondrial/eubacterial and the chloroplast/cyanobacterial groups of sequences there is a high frequency of identities in the carboxyl termini. In contrast the sequences in the outlier group show few (if any true) identities in the C-terminal region, and the partial sequences beyond the rightmost (|) are presented in raw form. The independent alignment of four N-terminal sequences from each of the groups is shown in Fig. 5. The sequences were truncated to remove nonoverlapping N-termini and end immediately proximal to the first conserved Cys residue. Again, there is similarity between Rieske protein sequences from eubacteria and mitochondria (either plants or mammals) and within the group of Rieske proteins from cyanobacteria and chloroplasts, while the sequences within the outlier group are notably dissimilar. Also of note, there are no conserved patterns shared between the clades.

Overall, even when the "active site" residues are excluded, there appears to be a high degree of similarity in both the C-terminal and N-terminal portions of Rieske proteins within the "mitochondrial" clade and within the "plastid" clade, but not between the two

Fig. 4. Aligned sequences of the active site and C-terminal region of four representative Rieske proteins from (a) the mitochondrial-eubacterial group [*Bos taurus, Neurospora crassa, Zea mays, Bradyrhizobium japonicum*], (b) the plastid-cyanobacterial group [*Pisum sativum, Nicotiana tabacum, Synechocystis sp., Cyanobacterium nostoc sp.*], and (c) the outlier group [*Bacillus subtilis, Chlorobium limicola, Thermus thermophilus, Sulfolobus acidocaldarius* (*soxf* gene)]. The (*) symbols indicate the residues that serve as a ligand to the [2Fe-2S] center, the (#) symbol indicates participation in a disulfide bond, and the (|) symbol indicates a residue conserved within each group of four sequences. The sequences beyond the rightmost (|) in the outlier group show very little similarity and are not aligned.

clades or within the outlier group. Since the percent identity between proteins from the two clades (13.43–22.63%) is similar to that expected for random sequences (Doolittle, 1986) it is difficult to determine whether the mitochondrial and the plastid clade share a common ancestor. The high-resolution structure of the soluble fragment of the chloroplast Rieske protein (Zhang *et al.*, 1996) will probably shed further light on this point. A search of the protein data bases using the complete bovine Rieske protein sequence from which the active site sequence (CTHLGCVPIANAG-DFGGYYCPCH) was removed, returned 17 "mitochondrial" hits at high score (smallest sum probability, $P(N = 3) \leq 8 \times 10^{-27}$) followed by eight "plastid" hits at much lower score ($P(N = 1, 2) \geq 1 \times 10^{-3}$) (cf. Karlin and Altschul (1993) for an explanation of

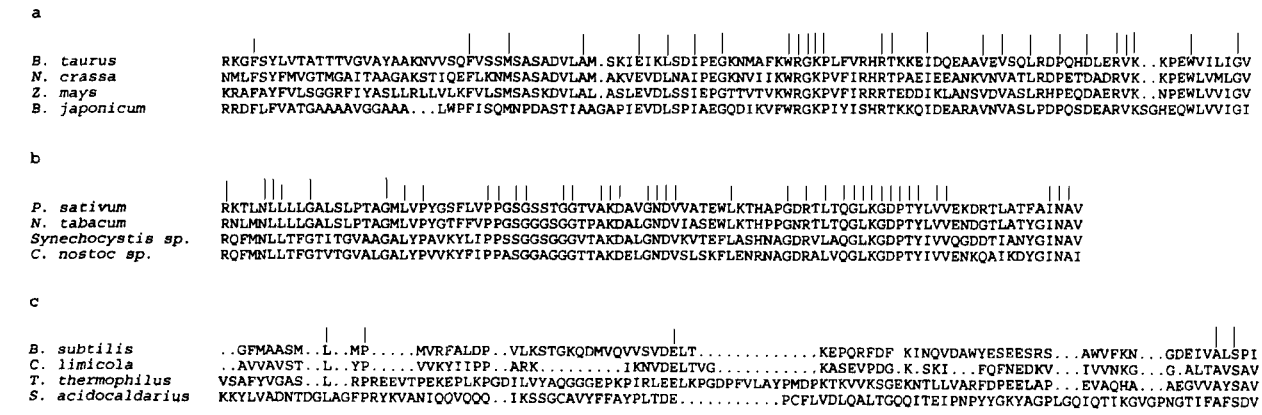

Fig. 5. Aligned sequences of the N-terminal region of four representative Rieske proteins from (a) the mitochondrial-eubacterial group, (b) the plastid-cyanobacterial group, and (c) of the outlier group (species abbreviations as in Fig. 4). The notation (|) indicates a residue conserved within each group of four sequences.

the statistical parameter). Similarly, when the spinach Rieske protein sequence minus the active site sequence (CTHLGCVVPFNAAENKFICPCH) was used as the query, eight plastid hits were returned at high score ($P(N = 2) \leq 1.6 \times 10^{-58}$), followed by eight "mitochondrial" hits with much lower score ($P(N = 2, 1) \geq 4 \times 10^{-2}$). The region of sequence with possible common origin between the two clades begins immediately after the active site sequence and ends at the C-terminus of the bovine protein. When the *Thermus* Rieske protein sequence minus the active site sequence (CTHLGCIVSQWVADEEAALCPCH) was used as the query, there were no hits. One possible explanation for these observations is that a short nucleotide sequence encoding the active site residues was inserted into different genes early in the evolutionary history of the Rieske proteins, thus providing an independent origin for the two clades and for the outliers.

Structural Features of the Thermus Rieske Protein

The crystal structure of a soluble fragment obtained by controlled proteolysis of the bovine Rieske protein has been solved at high resolution (Iwata *et al.*, 1996); this fragment retains the redox and spectral properties of the uncleaved protein (Link *et al.*, 1996). It is of note that the spectral features of isolated Rieske proteins from various sources suggest strong conservation of the three-dimensional structure in the active site region (Trumpower, 1981; Fee *et al.*, 1984; Britt *et al.*, 1991; Gurbiel *et al.*, 1996). Thus, at least some of the structural features of the Fe-S cluster domain of the *Thermus* Rieske protein may be inferred by comparing its amino acid sequence to that of the bovine protein. An alignment of the two amino acid sequences is presented in Fig. 6; the amino terminus of the soluble fragment of the bovine protein (AMSKI sequence) is indicated by an arrow, regions of similarity are shaded, and identities are in bold. In the following discussion, amino acid numbering for the *Thermus* protein will be as per the equivalent residues of the bovine protein shown in Fig. 6. As already pointed out, there is very little sequence similarity between the *Thermus* and the bovine sequence in the region upstream of the cluster binding sequences. The N-terminal region in the soluble fragment of the bovine protein forms one α helix and a long loop (residues 102–130) that connects strands β3 and β4 (Iwata *et al.*, 1996); it is believed that both the helix and the long loop are missing in the Rieske proteins from b_6f complexes, reflecting a different mode of interaction between the iron-sulfur protein and, respectively, cytochrome c_1 or *f*. The lack of sequence similarity between the *Thermus* and bovine protein in this region does not allow one to predict with certainty whether the helix-loop domain is present in the *Thermus* protein. In contrast, the region of the bovine protein between residues 137 and 184 (containing the active site sequences: residues 138–146 and 158–162) is quite similar to that of the *Thermus* protein. Therefore, this domain probably has a similar overall fold in both proteins. The predicted topology of the *Thermus* protein in this region is shown side by side with the known topology of the bovine protein in Fig. 7. A few amino acid differences are of special note. The bovine Pro146 is replaced by a valine in *Thermus*. In yeast the mutation Pro146 → Leu changes the midpoint potential of the iron-sulfur center by −96 mV (Gatti *et al.*, 1989). As pointed out by Iwata *et al.* (1996), the side chain of a leucine at position 146 (in one of its possible rotamers) would be in van der Waals distance of both Tyr165 and Cys158 (both present in *Thermus*) and could produce a conformational change. Since the *Thermus* protein has redox properties similar to the bovine protein (Kuila and Fee, 1986), it would appear that the presence of a valine at position 146 is well tolerated. Notably, none of the outlier sequences listed in Fig. 4 have a proline at the position equivalent to the bovine 146. In the bovine protein, strands β5 and β6 and the intervening loop (residues 147–157) separate the two stretches of residues that hold the cluster. This region of the protein has significant variation in length between the Rieske proteins that are part of bc_1 complexes and the homologous Rieske cluster binding domains of bacterial oxygenases (as revealed by a search of the protein data bases). Significant differences are present also between the plastid and mitochondrial proteins and some of the proteins in the outlier group (Fig. 4) The *Thermus* sequence 147–157 is clearly more hydrophilic than the equivalent bovine β5 and β6 strands (with the intervening loop). This variation may represent differences in the interaction of the bovine and *Thermus* Rieske proteins with other proteins in the respective respiratory complexes. Of particular note is the presence of a glycine at position 163 (bovine numbering) in the *Thermus* protein. In the bovine protein (as well as all the other mitochondrial and plastid proteins) a serine occupies this position, and the crystal structure shows that the serine hydroxyl oxygen donates a hydrogen bond to one of the cluster inorganic sulfur

```
Thermus              MDEREVRLRQSRRRLFLKTVIGTGIGLSLVSAFYVGASLR
Bovine    SHTDIKVPDFSDYRRPEVLDSTKSSKESSEARKGFSYLVTATTTVGVAYA    50

Thermus   PREEVTPEKEPLKPGDILVYAQGGGEPKPIRLEELKPGDPFVLAYPMDPK
Bovine    AKNVVSQFVSSMS.ASADVLAMSKIEIKLSDIPEGKNMAFKWRGKPLFVR    99
                                  ------>

Thermus   TKVVKSGEAKNTLLVARFDPEELAPEVAQHAAEGVVAYSAVCTHLGCIVS
Bovine    HRTKKEIDQEAAVEVSQL..RDPQHDLERVKKPEWVILIGVCTHLGCVPI   147

Thermus   QWVADEEAALCPCHGGVYDLRHGAQVIAGPPPKPVPQLPVRVEDGVLVAA
Bovine    ANAGDFGGYYCPCHGSHYDASGRIRKGPAPLNLEVPSYEFTSDDMVIVG    196
```

Fig. 6. Direct sequence alignment of the bovine and *Thermus* Rieske proteins. Only the sequence corresponding to the mature bovine protein (after cleavage of the mitochondrial targeting sequence) is shown. The N-terminus of the soluble fragment of the bovine protein is indicated by an arrow. The regions of similarity are shaded and identities are in bold. Residue numbering is shown only for the bovine protein.

atoms. Interestingly, two additional outlier sequences shown in Fig. 4 lack a serine at this position. Furthermore, a serine at this position is neither present in ferredoxin-type Rieske proteins, which mostly have a glycine, nor in the bacterial oxygenases that carry a Rieske cluster, which mostly have a tryptophan (data

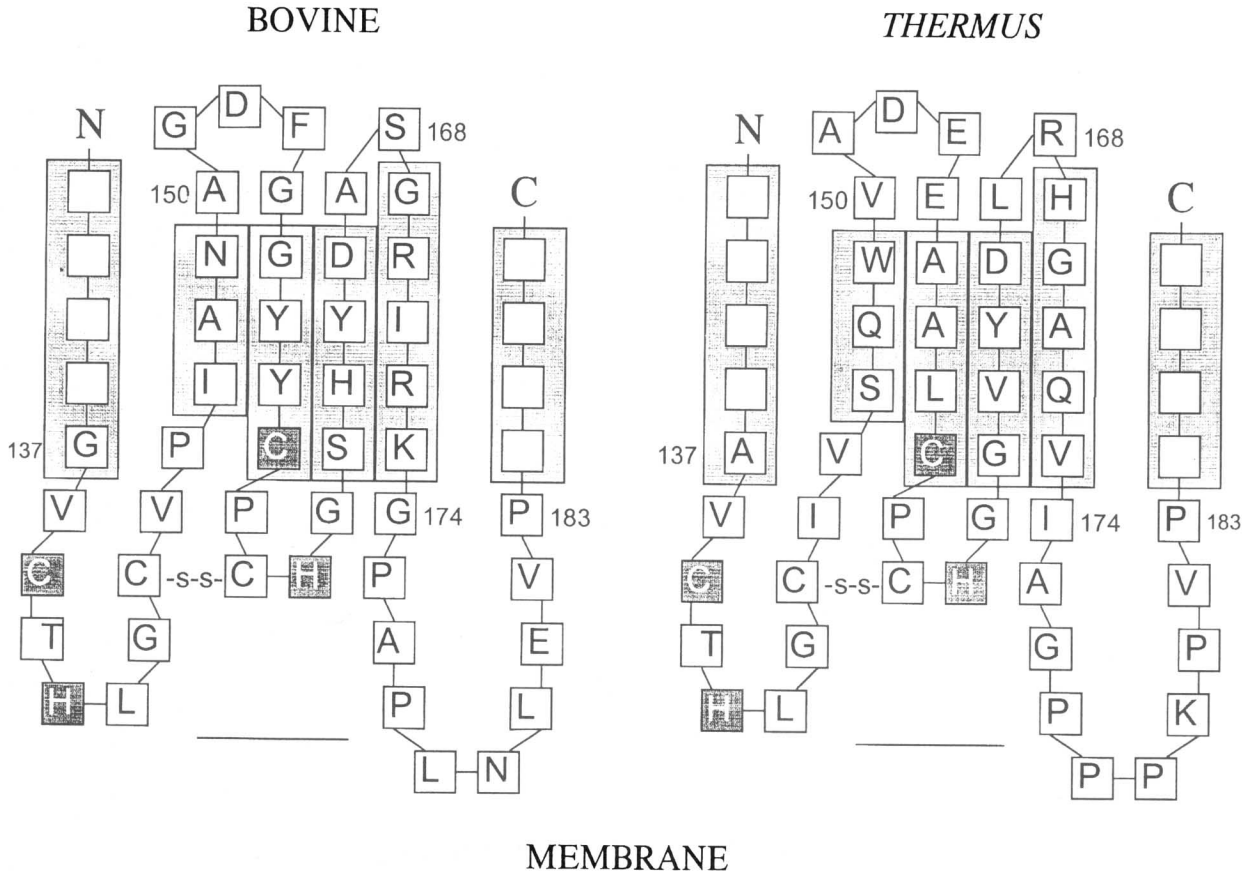

Fig. 7. Topology of the cluster binding domain of the bovine and *Thermus* Rieske proteins. β strands are represented by shaded rectangles. The cluster ligands are shown as dark gray boxes. Residue numbering as in Fig. 6. Adapted from Fig. 3 of Iwata *et al.* (1996).

derived from data base searches). Thus, it would appear that while donation of a hydrogen bond from a serine hydroxyl to the inorganic sulfur might modulate the cluster electronic environment, this action is not required for cluster assembly and/or stability. In the bovine protein the hydroxyl oxygen of Tyr165 donates a hydrogen bond to one of the cysteine sulfurs that ligate the iron. Tyr165 is conserved in *all* the bc_1/b_6f Rieske proteins, but not in the ferredoxin-type Rieske proteins. Another feature that appears universally conserved in the bc_1/b_6f Rieske proteins, but not in the Rieske ferredoxins, is a disulfide bridge (between Cys144 and Cys160 in the bovine sequence) that shields one side of the cluster. Rieske ferredoxins are approximately the same size as the soluble fragment of the bovine Rieske protein and are stable independently from interactions with other proteins. Therefore, it is possible that both the disulfide bridge and the hydrogen bond from Tyr165 to one of the iron ligands are involved primarily in conferring the proper conformation of the polypeptide chain and of the cluster ligands responsible for the much higher redox potential of bc_1 versus ferredoxin type Rieske proteins.

In the bovine protein Arg170 forms a salt bridge with Asp166. In yeast, mutation of the Rieske protein Arg170 to Gly suppresses the overall enzyme activity of the bc_1 complex (Graham *et al.*, 1992). Interestingly, in *Thermus* the same replacement (R170 → G) is tolerated, possibly because additional changes are present in the neighboring positions. For example, Asp166 might form an alternative salt bridge with His169. On the opposite side of the disulfide bridge the iron-sulfur cluster is shielded in the bovine protein by a loop (residues 174–183) containing two proline residues (the "Pro"- loop). Mutagenesis studies in yeast and *Rhodobacter* have shown that this loop is critical for the stability of the cluster (Graham *et al.*, 1992; Liebl *et al.*, 1995). Although Pro175 is not conserved in *Thermus*, the proline content of the loop is actually increased (Fig. 7), and it is possible that the additional prolines force an overall conformation of the loop similar to that present in the bovine protein. In view of these similarities it is likely that the loops surrounding the cluster of the *Thermus* Rieske protein, as has been proposed for the bovine protein (Iwata *et al.*, 1996; Link and Iwata, 1996), contact (or are embedded in) the membrane near the site where quinones are oxidized.

In the bovine protein the loop (residue 150–154) between strand β5 and β6 and the turn (residue 167–168) between strands β7 and β8 (Fig. 7) are significantly less hydrophilic than the equivalent elements in the *Thermus* protein. The crystal structure of the soluble fragment of the bovine protein shows that the aforementioned loop and turn interact with the helix-loop domain (residues 102–130, see above). The differences noted in the *Thermus* amino acid sequence of these regions may reflect either a variation in the interaction of the cluster domain with the helix-loop domain or the absence of the latter altogether.

It has been reported that hyperthermostable proteins form a high number of salt bridges and often contain arginine instead of lysine (Hennig *et al.*, 1995; Musafia *et al.*, 1995; Yip *et al.*, 1995). From the sequence alignment of the bovine and *Thermus* proteins shown in Fig. 6 and from the structural considerations just discussed, it is not possible to determine whether the *Thermus* protein forms a larger number of salt bridges than the bovine protein. A comparison of the amino acid content of the two proteins in the aligned regions is shown in Table I. The increased content of alanine, proline, and valine and of alanine *plus* glycine, and the decreased content of basic residues in the *Thermus* protein might be significant. These data argue against the view that the increased thermostability of the *Thermus* protein (Fee *et al.*, 1984) may be attributable to an increase in the number of salt bridges and/or hydrogen bonds. Instead, the overall increase in residues with short side chains (G,A,V,P) suggests that the higher stability of the thermophilic protein may originate from a tighter packing of the polypeptide chain. These observations are generally consistent with the view that thermostability of proteins is the result of an increased number of small stabilizing interactions (Macedo-Ribeiro *et al.*, 1996).

ACKNOWLEDGMENTS

This work was supported by NIH Grant GM48157 to Sharon H. Ackerman and by NIH Grant GM35342 to JAF. Part of the study was conducted by DG as a Research Fellow at the Biophysics Research Division of the University of Michigan and was supported by NIH Grant GM20877 to David P. Ballou and by NIH Grant GM16429 to Martha L. Ludwig. We thank Ms. Karen Findling (University of Michigan) for protein preparation, Dr. Michael W. Mather (Los Alamos) for assistance in probe design, and Dr. Russell Doolittle (University of California at San Diego) for making his phylogenetic analysis programs available to us and for numerous helpful discussions. We thank Drs. David

Table I. Molar Amino Acid Composition (Mole Percent) of the Bovine and of the *Thermus* Rieske Proteins[a]

Residue	Bovine (charge −1, isoelectric point 7.00)	Thermus (charge −6, isoelectric point 5.00)
A = Ala	6.154	13.077
C = Cys	3.077	3.077
D ≡ Asp	6.923	4.615
E = Glu	6.923	8.462
F = Phe	3.077	1.538
G = Gly	8.462	9.231
H = His	3.846	3.077
I = Ile	6.923	2.308
K = Lys	7.692	5.385
L = Leu	6.923	8.462
M = Met	2.308	0.769
N = Asn	2.308	0.769
P = Pro	6.923	11.538
Q = Gln	2.308	3.846
R = Arg	5.385	3.077
S = Ser	5.385	2.308
T = Thr	2.308	2.308
V = Val	8.462	13.077
W = Trp	1.538	0.769
Y = Tyr	3.077	2.308
A + G	14.615	22.308
S + T	7.692	4.615
D + E	13.846	13.077
D + E + N + Q	18.462	17.692
H + K + R	16.923	11.538
D + E + H + K + R	30.769	24.615
I + L + M + V	24.615	24.615
F + W + Y	7.692	4.615

[a] Only the region of the bovine sequence corresponding to the fragment crystallized (last 130 residues) and the corresponding region of the *Thermus* protein (Fig. 6) were considered for this analysis.

Ballou and Martha Ludwig for many stimulating discussions.

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.

Berry, E. A., Huang, L. S., Zhang, Z., Chi, Y. I., Hung, L. W., and Kim, S. H. (1997). Abstract 2468 of the 17th International Congress of Biochemistry and Molecular Biology, *FASEB J.* **11**, A1279.

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

Cline, J. F., Hoffman, B. M., Mims, W. B., LaHaie, E., Ballou, D. P., and Fee, J. A. (1985). *J. Biol. Chem.* **260**, 3251–3254.

Cramer, W. A., Martinez, S. E., Huang, D., Furbacher, P. N., and Smith, J. L. (1994). *Curr. Opin. Struct. Biol.* **4**, 536–544.

Davidson, E., Ohnishi, T., Atta-Asafo-Adjei, E., and Daldal, F. (1992). *Biochemistry* **31**, 3342–3351.

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

Doolittle, R. F. (1986). *Of URFS and ORFS: A Primer on How to Analyze Derived Amino Acid Sequences*, University Science Books, Mill Valley, California.

Feng, D. F., and Doolittle, R. F. (1990). *Methods Enzymol.* **183**, 375–387.

Fee, J. A., Findling, K. L., Yoshida, T., Hille, R., Tarr, G. E., Hearshen, D. O., Dunham, W. R., Day, E. P., Kent, T. A., and Munck, E. (1984). *J. Biol. Chem.* **259**, 124–133.

Findling, K. L., Yoshida, T., and Fee, J. A. (1984). *J. Biol. Chem.* **259**, 123.

Gabellini, N., and Sebald, W. (1986). *Eur. J. Biochem.* **154**, 569–579.

Gatti, D. L., Meinhardt, S. W., Ohnishi, T., and Tzagaloff, A. (1989). *J. Mol. Biol.* **205**, 421–435.

Gennis, R. B., Barquera, B., Hacker, B., Van Doren, S., Arnaud, S., Crofts, A., Davidson, E., Gray, K. A., and Daldal, F. (1993). *J. Bioenerg. Biomembr.* **25**, 195–209.

Graham, L. A., Brandt, U., Sargent, J. S., and Trumpower, B. L. (1992). *J. Bioenerg. Biomembr.* **25**, 245–257.

Gribskov, M., Devereux, J., and Burgess, R. T. (1984). *Nucleic Acids Res.* **12**, 539–549.

Grunstein, M., and Hogness, D. S. (1975). *Proc. Natl. Acad. Sci. USA* **72**, 101–106.

Gurbiel, R. J., Doan, P. E., Gassner, G. T., Macke, T. J., Case, D. A., Ohnishi, T., Fee, J. A., Ballou, D. P., and Hoffman, B. M. (1996). *Biochemistry* **35**, 7834–7845.

Hauska, G., Nitschke, W., and Herrmann, R. G. (1988). *J. Bioenerg. Biomembr.* **20**, 211–228.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K., and Jansonius, J. N. (1995). *Structure* **3**, 163–167.

Hill, J. E., Myers, A. M., Koerner, T. J., and Tzagoloff, A. (1986). *Yeast* **2**, 163–167.

Iwata, S., Saynovits, M., Link, T. A., and Michel, H. (1996). *Structure* **4**, 567–579.

Karlin, S., and Altschul, S. F. (1993). *Proc. Natl. Acad. Sci. USA* **90**, 5873–5877.

Keightley, J. A., Zimmermann, B. H., Mather, M. W., Springer, P., Pastuszyn, A., Lawrence, D. M., and Fee, J. A. (1995). *J. Biol. Chem.* **270**, 20345–20358.

Krinner, M., Hauska, G., Hurt, E., and Lockau, W. (1982). *Biochim. Biophys. Acta* **681**, 110–117.

Kuila, D., and Fee, J. A. (1986). *J. Biol. Chem.* **261**, 2768–2771.

Kurowski, B., and Ludwig, B. (1987). *J. Biol. Chem.* **262**, 13805–13811.

Liebl, U., Sled, V., Ohnishi, T., and Daldal, F. (1995). in *Photosynthesis: From Light to Biosphere* (Mathis, P., ed.), Kluwer Academic Publishers, Dordrecht, pp. 749–752.

Link, T. A., and Iwata, S. (1996). *Biochim. Biophys. Acta* **1275**, 54–60.

Link, T. A., Saynovits, M., Assmann, C., Iwata, S., Ohnishi, T., and von Jagow, G. (1996). *Eur. J. Biochem.* **237**, 71–75.

Macedo-Ribeiro, S., Darimont, B., Sterner, R., and Huber, R. (1996). *Structure* **4**, 1291–1301.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Springs Harbor.

Maseda, H., and Hoshino, T. (1995) *FEMS Microbiol.* **128**, 127–134

Mather, M. W., and Fee, J. A. (1990). *Plasmid* **24**, 45–56.

Mather, M. W., Springer, P., and Fee, J. A. (1991). *J. Biol. Chem.* **266**, 5025–5035.

Mather, M. W., Keightley, J. A., and Fee, J. A. (1993). *Methods Enzymol.* **218**, 695–704.

Mather, M. W., Springer, P., Hensel, S., Buse, G., and Fee, J. A. (1993). *J. Biol. Chem.* **268**, 5395–5408.
Maxam, A. M., and Gilbert, W. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
Musafia, B., Buchner, V., and Arad, D. (1995). *J. Mol. Biol.* **254**, 761–770.
Rieske, J. S., Hansen, R. E., and Zaugg, W. S. (1964). *J. Biol. Chem.* **239**, 3017–3022.
Sanger, F., Nicklen, S., and Coulson, A. R. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
Schmidt, C. L., Anemüller, S., and Schäffer, G. (1996). *FEBS Lett.* **888**, 43–46.
Sone, N., Tsuchiya, N., Inoue, M., and Noguchi, S. (1996). *J. Biol. Chem.* **271**, 12457–12462.
Tarr, G. (1982). In *Methods in Protein Sequence Analysis* (Elzinga, M., ed.), Humana Press Inc., Clifton, pp. 223–232.
Trumpower, B. L. (1981). *Biochim. Biophys. Acta* **639**, 129–155.
Trumpower, B. L., Edwards, C. A., and Ohnishi, T. (1980). *J. Biol. Chem.* **255**, 7487–7493.
Xia, D., Yu, C.-A., Kim, H., Xia, J.-Z., Kachurin, A. M., Zhang, L., Yu, L., and Deisenhoffer, J. (1997). *Science* **277**, 60–66.
Yang, X., and Trumpower, B. L. (1986). *J. Biol. Chem.* **261**, 12281–12289.
Yip, K. S. P., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., and Rice, D. W. (1995). *Structure* **3**, 1147–1158.
Yoshida, T., Lorence, R. M., Choc, M. G., Tarr, G. E., Findling, K. L., and Fee, J. A. (1984). *J. Biol. Chem.* **259**, 112–123.
Zhang, H., Carrell, C. J., Huang, D. Sled, V., Ohnishi, T., Smith, J. L., and Cramer, W. A. (1996). *J. Biol. Chem.* **271**, 31360–31366.
Zimmermann, B. H. (1988). Ph.D. Thesis, The University of Michigan.