Eur. J. Biochem. *154*, 569–579 (1986)

Nucleotide sequence and transcription of the *fbc* operon from *Rhodopseudomonas sphaeroides*

Evaluation of the deduced amino acid sequences of the FeS protein, cytochrome *b* and cytochrome c_1

Nadia GABELLINI and Walter SEBALD

Gesellschaft für Biotechnologische Forschung, Braunschweig-Stöckheim

(Received September 2/November 4, 1985) – EJB 85 0954

The *fbc* operon from *Rhodopseudomonas sphaeroides* encodes the three redox carriers of the ubiquinol–cytochrome-*c* reductase (b/c_1 complex): FeS protein, cytochrome *b* and cytochrome c_1 [Gabellini, N. et al. (1985) *EMBO J. 2*, 549–553]. The nucleotide sequence of 3874 bp of cloned *R. sphaeroides* chromosomal DNA, including the three structural genes *fbc*F, *fbc*B and *fbc*C has been determined. The reading frames of the *fbc* genes could be identified readily since the encoded amino acid sequences are highly homologous with the sequences of the corresponding mitochondrial polypeptides.

Initiation and termination points for transcription have been investigated by S_1 nuclease protection analysis. The transcription of the *fbc* operon starts approximately 240 base pairs upstream from the start codon of the *fbc*F gene and terminates 120 base pairs downstream from the stop codon of the *fbc*C gene. Nucleotide sequences resembling recognition signals for the binding and release of the RNA polymerase were identified.

The N-terminal amino acid sequence of the mature cytochrome c_1 was obtained by automated Edman degradation of the isolated subunit, confirming the *fbc*C reading frame and indicating that the bacterial pre-apocytochrome c_1 has a transient leader sequence including 21 residues. The N-terminal sequence of one hydrophilic peptide of the FeS protein has been also obtained confirming the *fbc*F reading frame.

The deduced amino acid sequences are discussed in relation to the known primary structures of the homologous proteins from mitochondria and chloroplasts. The primary structures of the polypeptides are evaluated with respect to (a) their topology in the membrane, (b) their biogenesis, (c) the structure of the catalytic sites and (d) subunit interactions.

In the electron transport chain of the photosynthetic bacterium *Rhodopseudomonas sphaeroides* and of mitochondria, a membrane-bound b/c_1 complex catalyzes the oxidoreduction of the mobile redox components ubiquinol and ferricytochrome *c*. The redox reaction generates an electrochemical potential coupled to ATP synthesis. In chloroplasts the homologous cytochrome b_6/f complex catalyzes the electron transport from plastoquinol to plastocyanin [1]. The functional oxidoreductase which is part of the cyclic photosynthetic and respiratory chain of *R. sphaeroides* comprises three main subunits: cytochrome *b*, cytochrome c_1 and a high potential FeS protein, carrying a cluster of four redox centers [2]. Recently the genes encoding the three main subunits of the bacterial b/c_1 complex have been cloned. The genes are part of one transcriptional unit of *R. sphaeroides* designated the *fbc* operon [3], and occur in the following order of transcription: (5′) *fbc*F, *fbc*B and *fbc*C, encoding the FeS protein, cytochrome *b* and cytochrome c_1. The *fbc* genes are coordinately transcribed in one polycystronic mRNA, and are constitutively expressed under photosynthetic and respiratory conditions of growth [3].

The gene encoding the 10–12-kDa polypeptide which copurifies with the bacterial oxidoreductase [2, 4] was not found in the vicinity of the *fbc* genes. It is in any case unclear whether this protein constitutes part of the b/c_1 complex. The genes encoding the homologous eukaryotic polypeptides are differently organized. The gene for the apocytochrome *b* is part of the mitochondrial genome, whereas the numerous other subunits of the mitochondria b/c_1 complex are encoded by the nuclear DNA and imported post-translationally [5]. Of the polypeptides of the cytochrome b_6/f complex only the FeS protein is encoded by the nuclear DNA, while the others are encoded by the chloroplast genome [6].

Many structural and functional features of the b/c_1 complexes from *R. sphaeroides* and mitochondria are identical. Homologies with the corresponding chloroplast complexes are weaker. One common feature of the cytochrome complexes from photosynthetic membranes is a simpler polypeptide composition, when compared with the mitochondrial counterpart [1].

The amino acid sequences of the three redox carriers FeS protein, cytochrome *b* and cytochrome c_1, deduced from the DNA sequence of the *fbc* operon presented in this work, are the first prokaryotic sequences available. The comparison with the corresponding sequences from eukaryotic systems

Correspondence to W. Sebald, Abteilung Cytogenetik, Gesellschaft für Biotechnologische Forschung, D-3300 Braunschweig-Stöckheim, Federal Republic of Germany

Abbreviations. bp, base pairs; RF, replicative form; SD, Shine and Dalgarno; SDS, sodium dodecyl sulfate.

Enzyme. Ubiquinol–cytochrome-*c* reductase or b/c_1 complex (EC 1.10.2.2).

identified regions of functional significance in terms of the topology and the catalytic structures of the b/c_1 complex subunits.

MATERIALS AND METHODS

Determination of the DNA sequence

The region of *R. sphaeroides* chromosomal DNA including the *fbc* genes was subcloned in three *Pst*I fragments of 604, 994 and 2270 bp respectively in M13 sequencing vectors. The two larger fragments were progressively shortened by two procedures.

The RF of two M13 mp10, bearing the *Pst*I fragment of 994 bp in opposite orientations, were linearized by a single *Hin*dIII cut in the polylinker region of the vector and treated with *Bal*31 exonuclease (0.27 U/μg DNA) at 30°C in a 500-μl assay. During the 15-min incubation period fractions were removed every 3 min, treated with T_4 polymerase and ligated to a *Hin*dIII linker by T_4 ligase. Shortened fragments were isolated from the *Pst*I/*Hin*dIII restriction digest of each fraction by preparative agarose gel electrophoresis, religated in M13 mp10 and used to tranfect *Escherichia coli* [7]. RF DNA was prepared in small scale from 2-ml cultures of 24 clones isolated from each tranfected fraction by standard procedures [8] and analyzed by *Pst*I/*Hin*dIII digestion.

The RF of two M13 mp19 clones bearing the *Pst*I insert of 2270 bp in opposite orientations were opened in the polylinker region at the unique *Bam*HI and *Kpn*I sites and treated with exonuclease III. Unidirectional digestion of the insert [9] from the *Bam*HI site was performed using 150 U exonuclease III/μg DNA at 37°C in a 100-μl assay. During the 280-s incubation period aliquots of 5 μl were removed every 20 s and treated with S_1 nuclease, Klenow DNA polymerase and T_4 ligase as described in [9]. After transfection of *E. coli* with the shortened plasmids, 24 clones were isolated from each sample and the sizes of the inserts were analyzed after *Hin*dIII/*Eco*RI digestion.

DNA sequencing by the dideoxynucleotide chain termination method [10] was carried out using [α-^{35}S]dATP and 6% acrylamide/50% urea gels.

S_1 protection analysis

The preparation of single-stranded radioactive DNA probes for hybridization with RNA was carried out as follows. A *Pst*I/*Bst*EII fragment including 319 bp of the DNA region in the vicinity of the 5′ end of the *fbc* transcript was isolated and end-labelled with [γ-^{32}P]ATP. Strand separation was performed by electrophoresis on a denaturating 6% acrylamide gel. Alternative single-stranded probes were obtained by the 'prime cut method' [11] from M13 clones bearing the *Pst*I fragment of 640 bp or a deletion fragment of *Pst*I 2270 bp including the 3′ end of the *fbc* transcript. After hybridization with the sequencing primer, the complementary strand was synthesized by the Klenow DNA polymerase in the presence of [α-^{32}P]dATP. After digestion with an appropriate restriction endonuclease, the probes were separated from the complementary strands by electrophoresis on denaturing gels. Labelled probes were recovered by electroelution of gel strips and ethanol-precipitated together with 75 μg RNA at about 10^5 cpm/assay.

R. sphaeroides total RNA was isolated from exponentially growing cells after lysozyme treatment by the guanidinium isothiocyanate CsCl method [8]. Hybridization performed at 60°C was followed by S_1 nuclease digestion as described in [12].

Determination of the N-terminal sequence of the b/c_1 *subunits*

The b/c_1 complex from *R. sphaeroides* was prepared as described in [2]. Individual polypeptides were isolated by preparative SDS gel electrophoresis and eluted for 4 h at 37°C in 0.5 M ammonium acetate and 0.1% SDS. The extract was dialyzed against 0.1% SDS for 12 h, lyophilized, extracted with ethanol and redissolved in 1 ml of 0.5% SDS.

Aliquots containing 10 nmol of the polypeptides were coupled to 60 mg diisothiocyanate glass beads (23 nm) for 12 h at 20°C and for another 3 h after addition of 100 μl ethanolamine [13]. The glass was washed with water, methanol and diethyl ether, then dried. Samples were deformylated by incubation in 1 M methanolic HCl for 3 h at 20°C. One aliquot of 10 mg glass was used for back-hydrolysis. Automated Edman degradation was carried out with a solid-phase sequencer (model 12, Sequemat, Watertown, USA). The phenylthiohydantoins were identified by high-performance liquid chromatography under conditions described elsewhere [14].

Hydropathy profile of the b/c_1 *subunits*

The gain of free energy during transition of a segment of 20 amino acids from a random coil in water to an α-helix in the membrane was calculated for all sequence positions according to [15].

RESULTS AND DISCUSSION

Nucleotide sequencing strategy

The complete sequence of 3874 bp of *R. sphaeroides* genomic DNA which included the *fbc* operon was determined. Non-random sequence analysis was performed on a set of deletion clones generated by exonuclease processive digestion as described in Materials and Methods. An overlapping set of clones bearing the 994-bp *Pst*I fragment, progressively shortened at both ends by *Bal*31 treatment as shown in Fig. 1, was chosen for DNA sequencing.

The generation of deletion clones of the larger *Pst*I fragment of 2270 bp was more rapidly achieved by exonuclease III unidirectional digestion of the insert, avoiding the step of fragment isolation. About 50% of the clones analyzed preserved the *Eco*RI site located before the primer region, indicating that the vector had been protected from exonuclease III attack. As shown in Fig. 1, a series of clones progressively deleted on average by 220 bp was chosen for DNA sequence analysis. Overlapping regions were easily identified for all of the M13 clones sequenced, allowing non-random continuous collection of sequencing data. The *Pst*I fragment of 604 bp was cloned in both directions in M13 mp19 and the two strands were sequenced as shown in Fig. 1. The DNA sequences of the *Sal*I fragments overlapping the *Pst*I sites were determined as reported previously [3].

The fbc *genes*

The determined DNA sequence shown in Fig. 2 includes three reading frames identified by sequence homology with

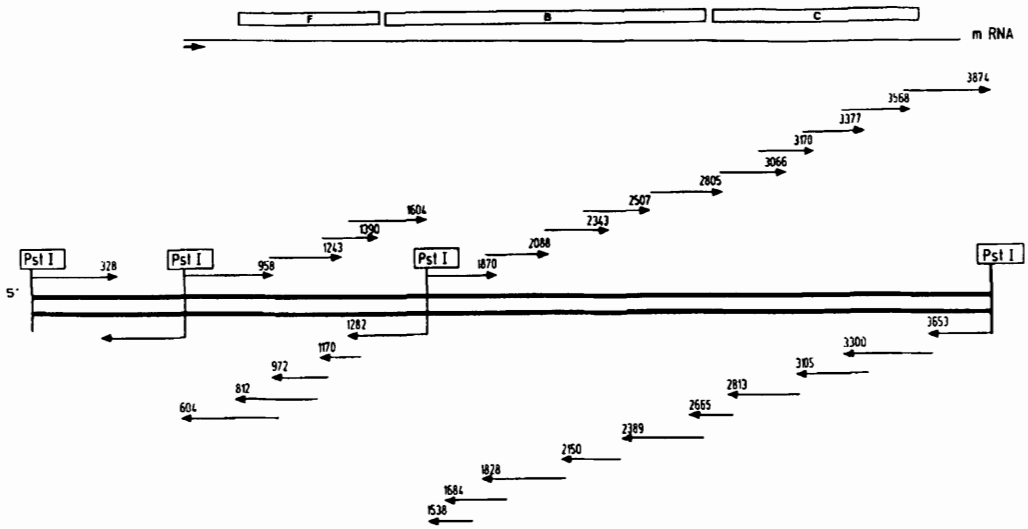

Fig. 1. *Nucleotide sequencing strategy*. The *Pst*I restriction sites of the DNA region including the *fbc* genes are indicated. Arrows indicate the direction and extent of sequencing that was terminated at the given base numbers. The position of genes *fbc*F, *fbc*B and *fbc*C and the polycistronic mRNA are shown in the upper part of the figure, where an arrow indicates the direction of transcription

polypeptides of the b/c_1 complex from mitochondria. The three reading frames correspond to genes *fbc*F, *fbc*B and *fbc*C of the *R. sphaeroides fbc* operon. The genes encode the FeS protein, cytochrome *b* and cytochrome c_1 of the b/c_1 complex from *R. sphaeroides*.

The first gene of the operon is *fbc*F which includes 576 bases encoding 191 amino acids of the FeS protein. The *fbc*F gene is proposed to begin with the GTG codon after base 842 of the determined DNA sequence (Fig. 2). This GTG codon most likely specifies a methionine [16]. It is considered the most probable start of the *fbc*F gene because it is preceded by an A+G-rich sequence strongly complementary to the 3′ end of the 16S ribosomal subunit of *R. sphaeroides* [17]. An SD-like consensus sequence [18] was not found in the region immediately upstream of other ATG or GTG codons in the *fbc*F frame.

A molecular mass of 21 000 Da was deduced for the Rieske FeS protein of *R. sphaeroides* indicating that the M_r of 25000 previously reported [1] for the subunit in the isolated b/c_1 complex, was slightly overestimated. A similar discrepancy in size was also found for the Rieske FeS protein of *Neurospora crassa* [19]. The *fbc*F frame terminates with the TGA stop codon after base 1415, it is separated from the *fbc*B frame by 12 nucleotides including an SD-like sequence. The *fbc*B frame begins with the ATG codon after base 1430 that opens a long frame of 1314 bases terminating with the TGA stop codon after base 2741. Gene *fbc*B encodes 437 residues of the cytochrome *b* subunit corresponding to a molecular mass of 48100 Da. For this hydrophobic protein an M_r of 40000 was estimated from the relative mobility in SDS gels [20], indicating that the bacterial cytochrome *b*, like the mitochondrial one, also migrates anomalously in SDS gels. The *fbc*C gene starts with the ATG codon after base 2761 and is preceded by an SD-like sequence. The *fbc*C frame includes 843 bp encoding 280 residues of the preapocytochrome c_1 with a molecular mass of 30100 Da. When determined by SDS-gel electrophoresis, the mature cytochrome c_1 exhibits an M_r of 34000 [2], whereas the size of the primary translation product is 1–2 kDa larger [3].

Identification of the fbc *promoter region*

The site of initiation of transcription was investigated by S_1 protection analysis of the DNA region upstream of the *fbc*F gene. A single-stranded probe, carrying 233 bases upstream of the start of the *fbc*F gene and extending to the *Pst*I site at base 609 of the sequence (Fig. 2), was entirely protected from S_1 digestion by *R. sphaeroides* RNA, whereas during a corresponding experiment the upstream *Pst*I fragment of 604 bp was completely digested by S_1 nuclease (not shown). Northern blot analysis revealed that no RNA hybridized to this *Pst*I fragment [3]. These data indicate that the 5′ end of the *fbc* transcript is located in the vicinity of the *Pst*I site at base 609 of the sequence.

The DNA sequence upstream of this *Pst*I site was compared with known promoter sequences. No homology with any *E. coli* promoter was found; however, the DNA sequence in the region of the putative *fbc* promoter can be aligned to the region upstream of the main transcriptional start of the *atp* operon from *Rhodopseudomonas blastica* [21], as shown in Fig. 3. The two sequences from position −1 to −50 showed 30 matches, postulating only minor insertions or deletions. In particular around position −10 the sequence GCCGC is conserved as well as the sequence ACCCGT-TGCG-CGC occurring after position −35. This strong homology suggests that also in *Rhodospirillaceae* the nucleotide sequences at the −10 and −35 regions are consensus sequences for the initiation of transcription [22, 23]. Interestingly, the putative *R. sphaeroides fbc* promoter includes a repeat of this consensus sequence (Fig. 3). The element CCGC, which is part of the −10 region, as well as the sequence ACCCGTTGC of the −35 region are repeated in the DNA region upstream of base 549. The significance of these tandemly repeated sequences is at present uncertain. Such a repeat is not found in the corresponding region of the *atp* promoter. The two Rhodospirillaceae promoters are also found at a different distance from the GTG start codon. The GTG start codon of the *fbc*F gene is located about 240 bases downstream of the *fbc* transcriptional start, whereas the first

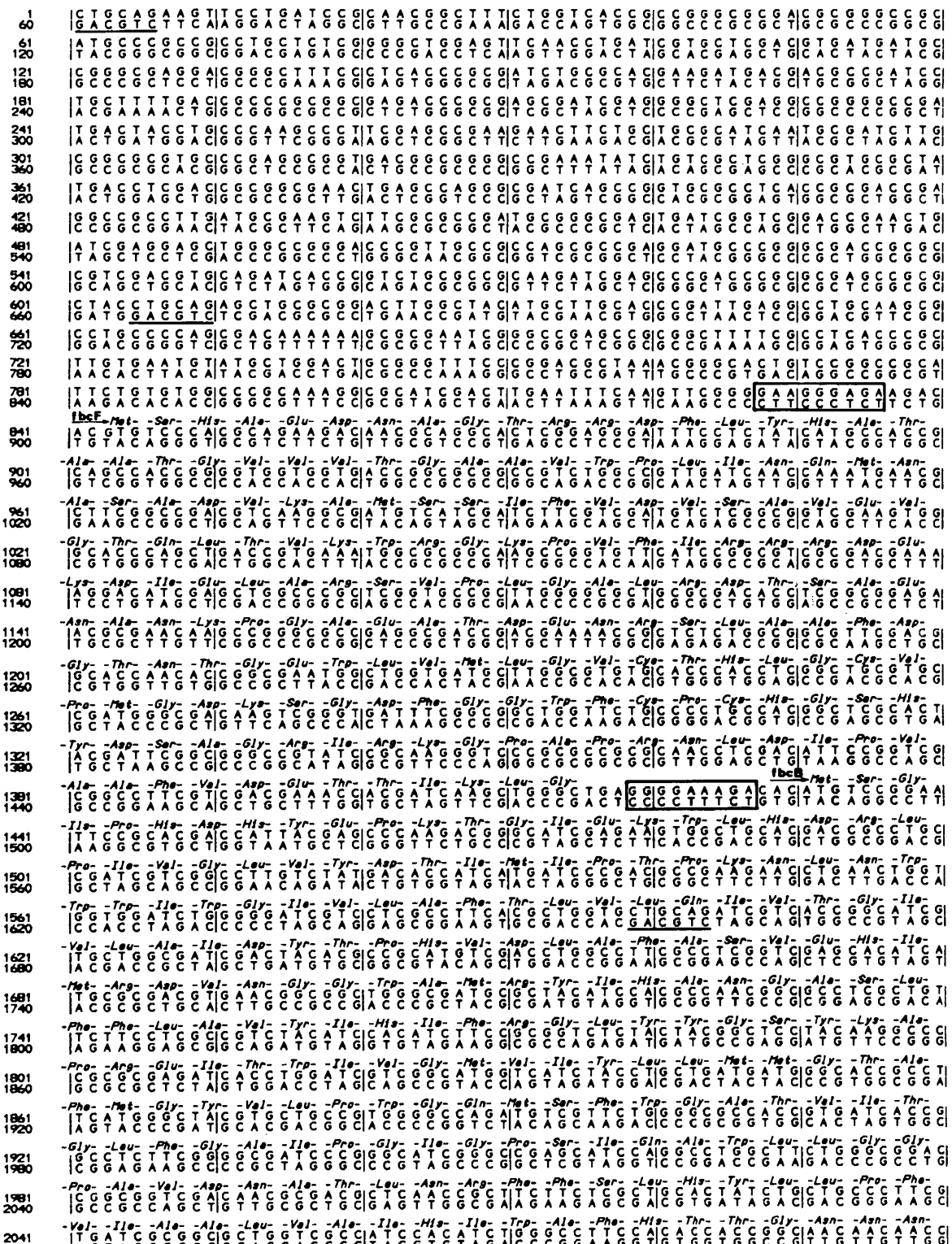

Fig. 2. *Nucleotide sequence of the DNA region including the* fbc *operon. Deduced amino acid sequence.* Start codons of *fbc*F, *fbc*B and *fbc*C genes are indicated. Consensus sequences for the initiation of translation are placed in boxes, the *Pst*I sites are underlined

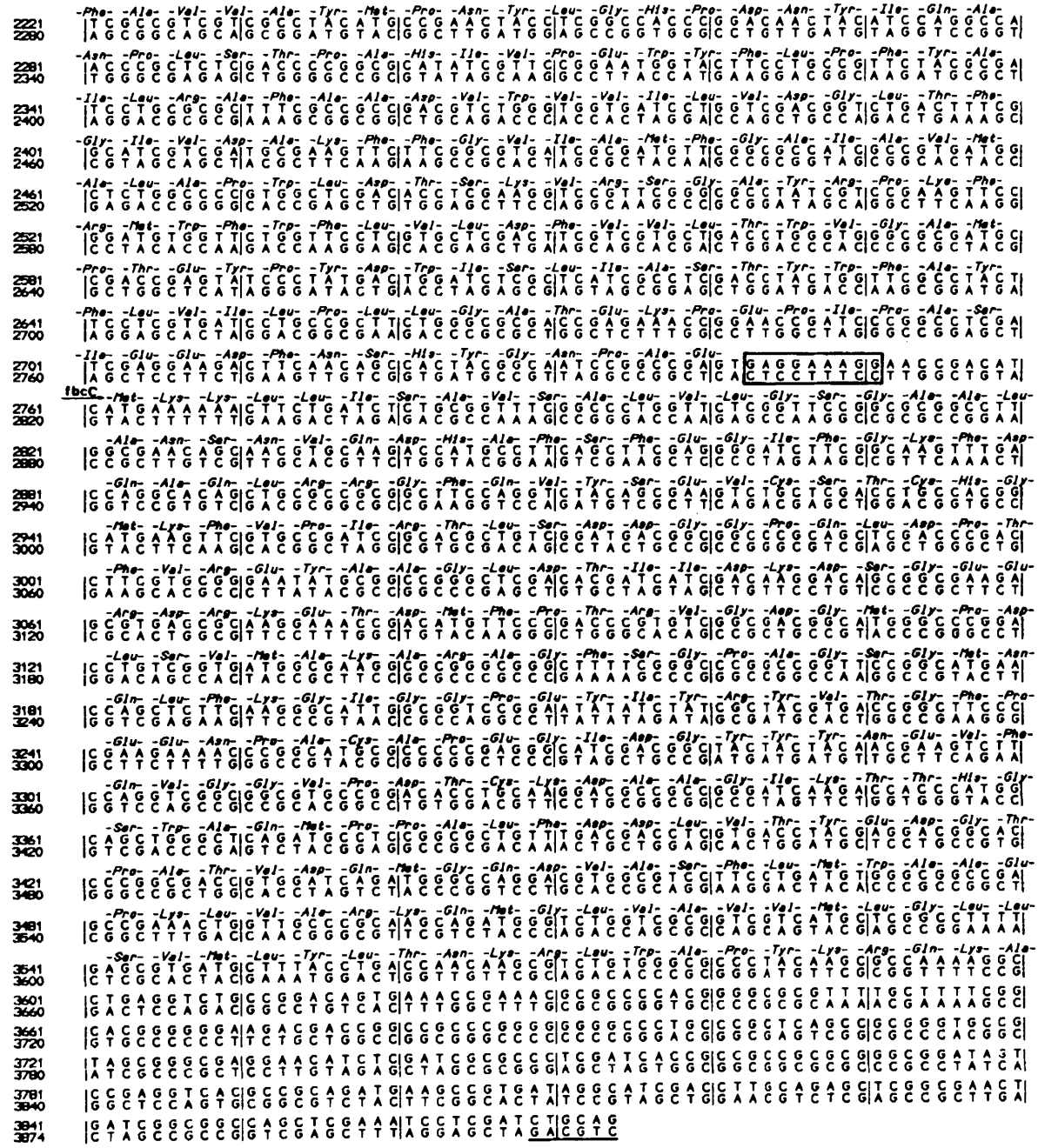

Fig. 2

gene of the *atp* operon encoding the δ subunit of the F_1-ATPase starts with the GTG after base + 87 from the start of transcription.

3′ Mapping of the fbc *transcript*

The 3′ end of the *fbc* transcript was precisely located by S_1 nuclease mapping as follows. From one deletion clone generated by exonuclease III treatment a specific single-stranded probe was isolated and used for hybridization with *R. sphaeroides* RNA followed by S_1 nuclease digestion. The probe covered 328 bp upstream of the *Pst*I site at base 3874 of the sequence (Fig. 2), including the stop codon of the *fbc*C gene. As shown in Fig. 4, a protected DNA fragment hybridizing with the *fbc* transcript was recovered after S_1 digestion (lane C). The size of the protected fragment was determined to be 228 bp by comparison with DNA fragments of known size run in parallel (Fig. 4, lane A). No protected DNA fragment was found in a control experiment employing *E. coli* tRNA (Fig. 4, lane B). Accordingly the 3′ end of the *fbc* transcript is located at base 3774 of the sequence (Fig. 2). Two possible secondary structures resembling termination signals are found in the vicinity of the 3′ end of the transcript (Fig. 5). One stable hairpin which can be formed within the region from base 3619 to 3658, resembles the *rho*-independent *E. coli* terminators, having a region rich in G · C base pairs in dyad

```
a: GGACCCGT TGCCGCCAGCGCCG  AGGATGCCCGG GCGA  CCGCGCCGTCGACGT
b: TCACCCGTCTGC GCCGCAAGATCGAGCCCGACCCGCGCGA GCCGCGCTACCTGCAG
c: CAACCCGT TGC GGCGCTGCAA  AGCCTCTGCTAAG GACGCCGCAAAATCGG
                   |                        |
                  -35                      -10
```

Fig. 3. *Homologies of the nucleotide sequences in the promoter regions of* R. sphacroides fbc *and* R. blastica atp *operon.* Nucleotide sequences of *R. sphaeroides fbc* promoter are aligned (a) upstream of nucleotide T-549, (b) of G-610, with the nucleotide sequence (c) of *R. blastica atp* operon upstream of the main start of transcription [21]. Positions − 10 and − 35 refer to the *atp* operon. Identical sequences are underlined

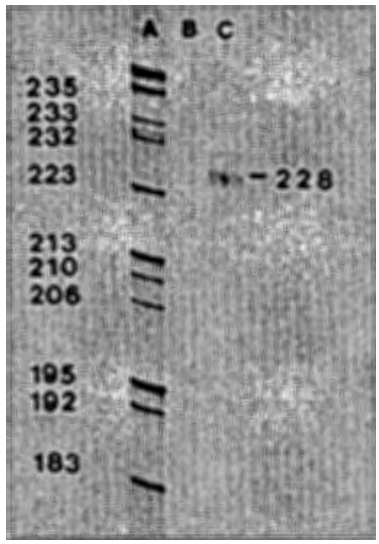

Fig. 4. S_1 *mapping of the 3′ end of the* fbc *transcript.* Lane A shows labelled DNA fragments of known nucleotide sequence used as size markers in base pairs, as indicated on the left side of the figure. In lane B the S_1 digest of the labelled DNA probe incubated with *E. coli* tRNA was loaded as control. In lane C the probe hybridized with *R. sphaeroides* RNA and digested by S_1 nuclease was loaded. The protected DNA fragment of 228 bp is indicated

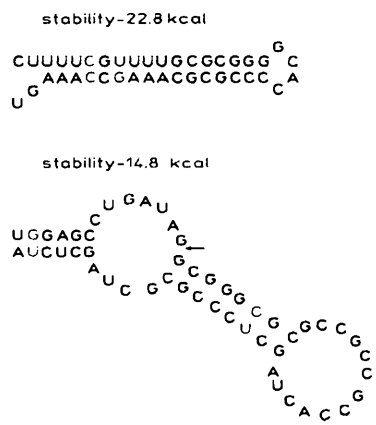

Fig. 5. *Two possible secondary structures of the 3′ end of the* fbc *transcript.* The upper RNA structure corresponds to the sequence from nucleotide T-3619 to C-3658 of the DNA sequence shown in Fig. 2. The RNA structure in the lower part corresponds to the nucleotide sequence from A-3736 to T-3787. The arrow indicates the point of termination of transcription. The stability has been calculated by the computer programme CRUSOE [41] (22.8 kcal = 95.3 kJ; 14.8 kcal = 61.9 kJ)

symmetry, plus a run of uridine residues immediately downstream [24]. This hairpin would be located only 16 bp downstream of the TGA stop codon of the *fbc*F gene, but 118 bases upstream of the actual 3′ end of the transcript, and could be involved in slowing down the RNA polymerase activity. The second structure proposed in Fig. 5 could be formed in the sequence between bases 3736 and 3787 and includes the 3′ end of the transcript. This structure, which includes two loops connected by seven G · C base pairs, could induce the release of the transcript.

EVALUATION OF THE DEDUCED AMINO ACID SEQUENCES

Background

For an evaluation of the deduced amino acid sequences, it is necessary to refer to some well founded structural and functional information about the b/c_1 complex. The three polypeptides encoded by the *fbc* operon are integral membrane proteins carrying the four prosthetic groups of the b/c_1 complex. The hydrophobic cytochrome *b* is embedded in the membrane and carries two hemes, with a midpoint potential E_{m7} of − 90 mV and 50 mV, respectively. The FeS protein and cytochrome c_1 are largely exposed to the water phase on the outer, positive side of the cytoplasmic membrane, and carry respectively a 2 Fe-2S cluster and a *c*-type heme with a similar E_{m7} of about 290 mV [1, 25]. These prosthetic groups form three catalytic sites of the b/c_1 complex at which reactions with the mobile components occur, probably through a Q-cycle mechanism [25, 26]. Ubiquinol is oxidized on the positive side of the membrane through a concerted reaction that involves the FeS cluster and the low-potential heme *b*. It is likely that this catalytic site is formed cooperatively by the FeS and cytochrome *b* subunits. The reduction of the FeS protein results in the transfer of reducing equivalents to cytochrome c_1 which forms the cytochrome *c* reductase site, located on the positive side of the membrane. The third catalytic site of the oxidoreductase faces the negative side of the membrane. It catalyses the reduction of ubiquinone by the high-potential heme *b* in two successive steps, with a semiquinone intermediate. The lipophilic semiquinone has to be stabilized by the binding to a protein [27], which is not yet defined clearly. It could be cytochrome *b* itself, in which case the three polypeptides encoded by the *fbc* operon might form the full catalytic unit of the bacterial b/c_1 complex. It is also possible that the additional 10-kDa subunit is required for the stabilization of the semiquinone as in the mitochondrial b/c_1 complex, in which one of the small subunits is supposed to be a Q-binding protein [28].

Cytochrome b

The distribution of hydrophobic amino acids in the *Rhodopseudomonas* cytochrome *b* sequence (Fig. 6) indicates 9 or 10 possible membrane-spanning regions. The peaks of hydrophobicity match the positions of homologous sequences of mitochondrial cytochrome *b* predicted as transmembrane segments in α-helical conformation [29, 30]. In comparison with the folding pattern predicted for yeast cytochrome *b*, the membrane-spanning segments of *R. sphaeroides* sequence correspond to: I, residues 48−67; II, 95−113; III + IV, 127−184; V, 194−216; VI, 253−275; VII, 330−349; VIII,

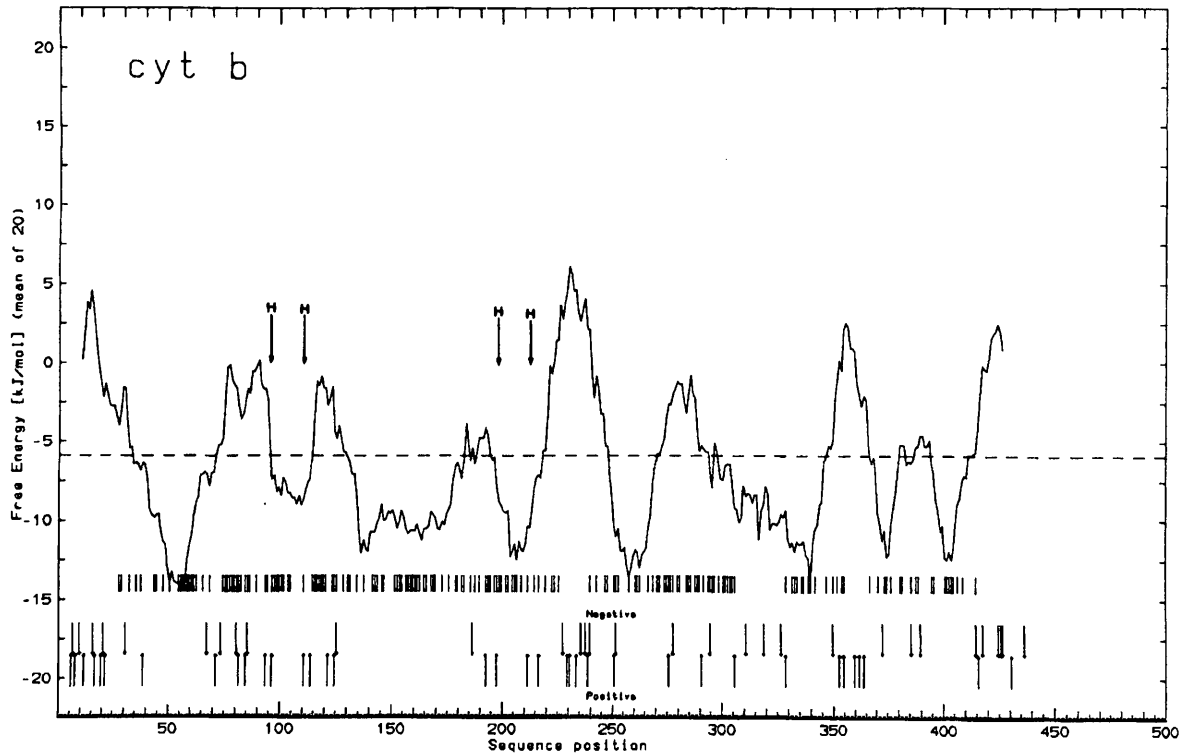

Fig. 6. *Polarity profile of the cytochrome* b *from* R. sphaeroides. The amino acid sequences below the dotted line have a high probability to be located in the membrane. The position of the four histidines (H) that bind the two heme groups are indicated by arrows. The positions occupied by identical amino acid residues in the *R. sphaeroides* and yeast mitochondria cytochrome *b* sequence [31] are marked by vertical bars. At the bottom basic (↑) and acidic (↓) amino acid residues are indicated by arrows

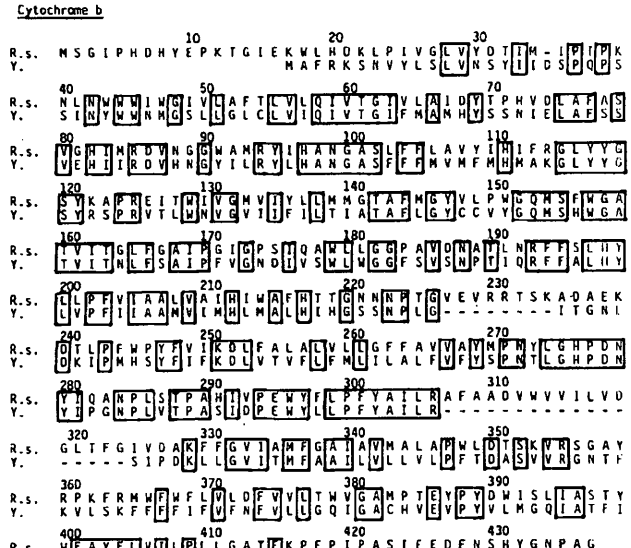

Fig. 7. *Alignment of the amino acid sequence of* R. sphacroides *and yeast mitochondria cytochrome* b. R. s., *Rhodopseudomonas sphaeroides*; Y, yeast. Identical residues are placed in boxes. The homology is even more extensive because of isofunctional substitutions

365–383; IX, 389–409. This arrangement places two aspartic acids at position 373 and 390 into membrane-spanning domains, however. The alignment of *R. sphaeroides* and yeast cytochrome *b* sequences [31] shown in Fig. 7, demonstrates the existence of many homologous regions. Of particular interest is the conservation also in the bacterial sequence of the four histidines postulated as ligands of the two heme groups on the basis of sequence homologies [29, 30]. The histidines occur in the *R. sphaeroides* sequence at residues 97, 111, 198 and 212. They are located in two hydrophobic domains corresponding to membrane-spanning segments II (97–111) and V (198–212) and are separated by 13 amino acids in both segments.

The structural relationship of the four histidine ligands suggests that the two hemes are coordinately bound between membrane-spanning segments II and V [29, 30], by using the histidine pairs His-97(II)/His-212(V) and His-111(II)/His-198(V) in the *Rhodopseudomonas* sequence. Remarkably four positive charges (Arg-94, Arg-114, Arg-193, His-217) in the vicinity of the heme binding sites are also conserved in the *R. sphaeroides* sequence. The identity of the two heme binding sites in terms of their redox potentials and the orientation of the polypeptide in the membrane cannot yet be predicted. A number of conserved residues are found also in the polar sequences, linking membrane-spanning domains as shown in Fig. 7. The hydrophilic regions could form pockets on the membrane surfaces around the catalytic regions, or might interact with other subunits of the complex.

The alignment of *R. sphaeroides* and yeast cytochrome *b* sequences implies that the bacterial cytochrome is larger then the mitochondrial one, including additional sequences at the two ends of the protein plus two extra stretches in the central domain. Remarkably the stretch of thirteen amino acids from Gly-226 includes six extra charged residues. Another addi-

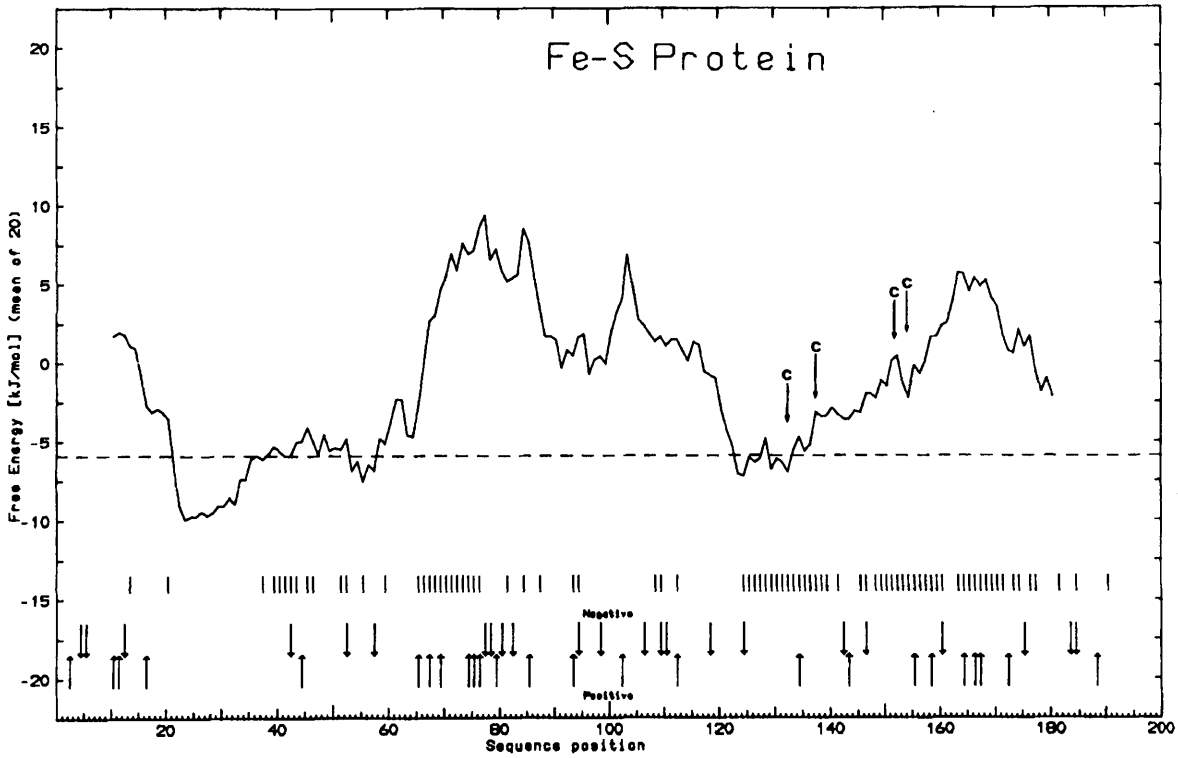

Fig. 8. *Polarity profile of the Rieske FeS protein from* R. sphaeroides. The positions of the four cysteines (C) involved in the binding of the FeS center are marked by arrows. Amino acids occupying identical position in *R. sphaeroides* and *N. crassa* FeS protein sequences [19] are marked by vertical bars. Basic (↑) and acidic (↓) residues are indicated at the bottom of the figure

tional hydrophobic sequence including two negative charges is observed between Arg-306 and Ile-325.

FeS protein

The FeS protein of the b/c_1 complex [32] belongs to a new class of protein carrying a high-potential 2-Fe-2S cluster. The hydropathy pattern of the FeS protein from *R. sphaeroides* (Fig. 8) shows that the polypeptide has a large hydrophilic domain, indicating that the major portion of the FeS sequence is exposed to the water phase. The polypeptide includes only one hydrophobic stretch from Phe-14 to Ala-42 that could form a membrane-spanning region. A similar distribution of hydrophobicity was also found in *N. crassa* FeS protein sequence [19] although the composition of the N-termini of the two polypeptides is not identical. The alignment of the two sequences shown in Fig. 9 indicates that the hydrophobic region is shifted by about 10 residues towards the N-terminus in *N. crassa*. This situation probably reflects the different biogenetic pathways of the two polypeptides. The N-terminus of the bacterial FeS protein resembles a leader sequence for the export through the cytoplasmatic membrane [33]. It contains six charged residues after the first methionine with a net charge of +2, followed by an hydrophobic sequence. The existence of a precursor form of *R. sphaeroides* FeS protein with an apparent molecular mass about 1 kDa greater than that of the mature subunit had been suggested by *in vitro* expression of the *fbc* genes [3]. Possibly only the small hydrophilic portion of the N-terminal signal sequence is cleaved off during secretion, whereas the following hydrophobic sequence is retained in the mature FeS protein. In an attempt to determine the N-terminal sequence of the mature FeS protein,

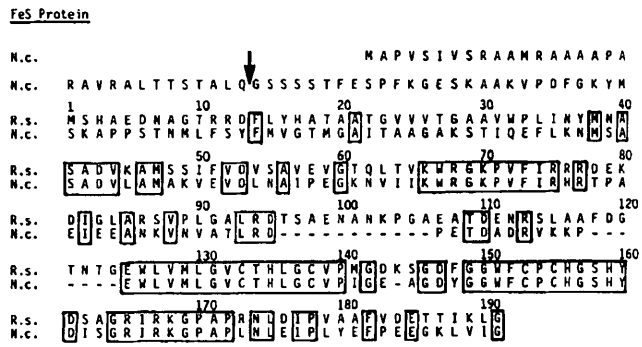

Fig. 9. *Alignment of the amino acid sequence of the Rieske FeS protein from* R. sphaeroides *and* N. crassa. R. s., *Rhodopseudomonas sphaeroides*; N. c., *Neurospora crassa*. Identities have been placed in boxes. The arrow indicates the N-terminus of the mature subunit of *N. crassa* [19]

the 25000-Da subunit of the *R. sphaeroides* b/c_1 complex was isolated and submitted to Edman degradation after a deformylation cycle. The sequence of the nineteen residues identified starts from Ser-49, which belongs already to the polar domain. The sequence was obtained only from 5% of the total coupled material indicating that only a small fraction of the FeS protein had a free N-terminus. Therefore it is likely that the sequenced peptide originated by a proteolytic cleavage during isolation of the protein. Similarly, the polar domain of the Rieske FeS protein from *N. crassa* can be cleaved off by chymotrypsin [19].

The C-terminal hydrophilic portion of this FeS protein is found to be highly conserved between *R. sphaeroides* and *N.*

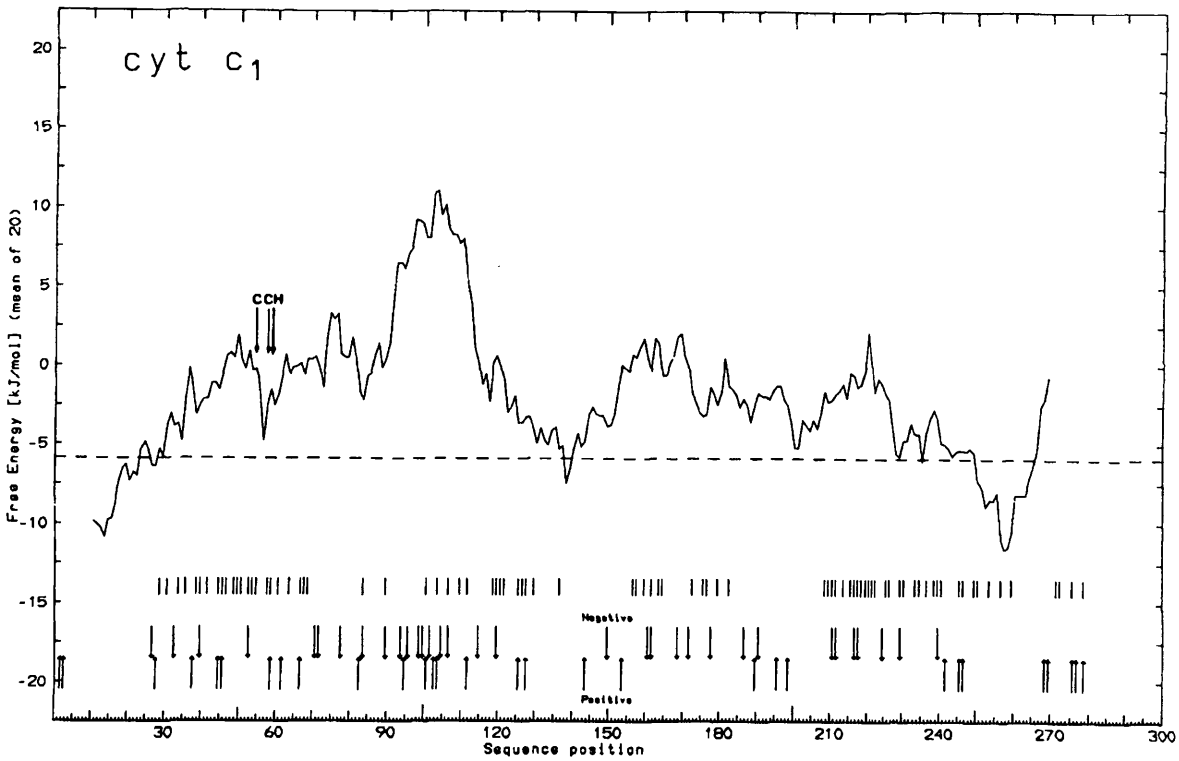

Fig. 10. *Polarity profile of the preapocytochrome* c_1 *from* R. sphaeroides. The positions of the two cysteines (C) and of the histidine (H) involved in the binding of the heme are indicated by arrows. Positions occupied by identical residues in the preapocytochrome c_1 sequence of *R. sphaeroides* and yeast mitochondria [34] are indicated by vertical bars. Basic (↑) and acidic residues (↓) are indicated by arrows

crassa (Fig. 9). The sequence includes four conserved cysteines occurring at positions 133, 138, 153 and 155 of *R. sphaeroides* which are the likely ligands of the 2Fe-2S cluster, although the involvement of other conserved residues cannot at present be excluded. Three conserved histidines are also found in this region at positions 135, 156 and 159. These conserved residues are located in a moderately hydrophobic sequence, suggesting that the FeS cluster could be included in a hydrophobic pocked. In particular the residues Trp-126, Trp-151, Tyr-160 and Phe-182 could form a water-repellent environment around the FeS cluster. This highly conserved region carrying the FeS cluster is expected to be in the vicinity of the low-potential heme cytochrome *b* to form the catalytic site for the oxidation of ubiquinol. Many conserved charged residues are also found in this region carrying in total seven negative and eight positive charges that could be potentially involved in establishing interactions with hydrophilic segments of the cytochrome *b*.

Another conserved region is found in the central hydrophilic domain of the protein that includes six positive charges: Lys-66, Arg-68, Lys-70, Arg-75, Arg-76 and Arg-77. This cluster of positive charges could be the region of interaction with the negative cytochrome c_1 (see below) at the reducing site of the FeS protein.

Cytochrome c_1

The hydropathy pattern of preapocytochrome c_1 from *R. sphaeroides* shown in Fig. 10 indicates two small hydrophobic domains, located near the N-terminus and the C-terminus of the protein, plus a large central hydrophilic domain. The N-terminal sequence of the cytochrome c_1 was determined by Edman degradation. The sequence of 18 residues obtained matches the sequence of the *fbc*C reading frame starting with the Asn-22. This result confirms the identity of the *fbc*C frame and furthermore indicates that the polypeptide is synthesized as a precursor form including a transient leader sequence of 21 amino acids.

The leader sequence has only two positively charged amino acids (Lys-2 and Lys-3) and it is rich in Ala and Val residues in the hydrophobic part. This is reminiscent of the leader peptide of the FeS protein (see above) and of the signal peptide from other secreted proteins [33]. Interestingly enough the precursor of the evolutionarily related mitochondrial cytochrome c_1, and also of the FeS protein, are imported post-translationally. Accordingly, they contain a different type of addressing signal (see Figs 9 and 11).

The stretch of hydrophobic amino acids 249–269 in the vicinity of the C-terminus most probably forms an α-helical segment which anchors the polypeptide to the membrane. A similar arrangement has been postulated on the basis of the amino acid sequences of the mitochondrial cytochrome c_1 [34] and of the chloroplast cytochrome *f* [35, 36].

The alignment of *R. sphaeroides* cytochrome c_1 with the corresponding polypeptide from yeast mitochondria, shown in Fig. 11, reveals a number of conserved regions of functional significance. The typical sequence that binds the heme covalently [37] is located near the N-terminus of the polypeptide corresponding to Cys-55, Cys-58 and His-59. A number of identical residues are found in this region, indicating that the overall structure of the heme binding peptide has been highly conserved. If the sixth ligand of the cytochrome c_1 heme is a methionine, as suggested in [38], it could only be Met-205 which is conserved not only in yeast but also in the horse-

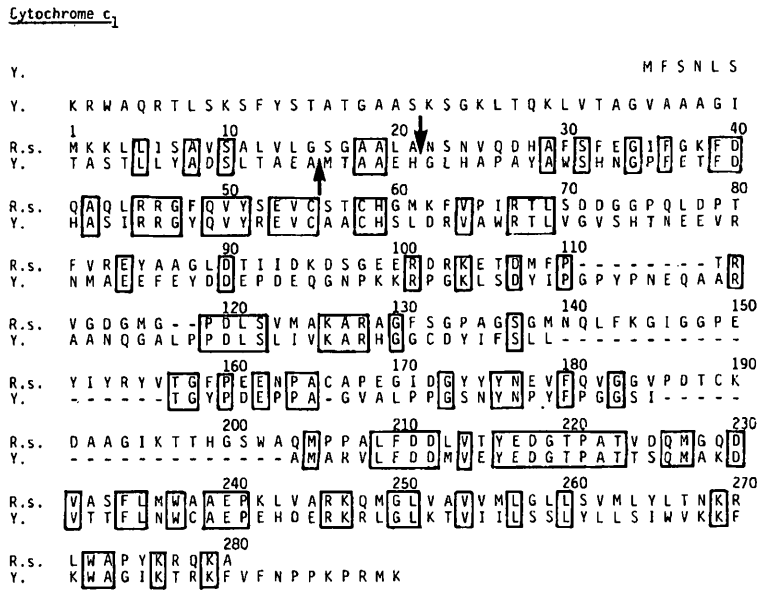

Fig. 11. *Alignment of the preapocytochrome* c_1 *sequence from* R. sphaeroides *and yeast mitochondria.* R. s., *Rhodopseudomonas sphaeroides*; Y, yeast. Boxes have been placed around identical residues. The arrows indicate the N-terminus of the mature subunits

heart sequence. This methionine occurs just before a highly conserved region near the C-terminus of the protein and is preceded by an insertion of 18 residues in *R. sphaeroides*.

Previously, the interaction of cytochrome c_1 from horse heart mitochondria with the soluble cytochrome *c* was studied during labelling experiments with a polar carbodiimide [39]. Cytochrome *c* protected a highly acidic sequence of horse heart cytochrome c_1, corresponding to amino acids 63–81. In this region three carboxyl groups, corresponding to Glu-84, Asp-90 and Glu-94 of the *R. sphaeroides* sequence, have been conserved in all three sequences available.

During cross-linking experiments cytochrome *c* was found to bind to a peptide corresponding to amino acids 165–174 of horse heart cytochrome c_1 [40]. In this region two negative charges corresponding to Glu-217 and Asp-218 of the *R. sphaeroides* sequence have been conserved. The two acidic regions of cytochrome c_1 identified in these studies could constitute the cytochrome c_2 binding domain.

Significant sequence homology with cytochrome *f* from spinach is found only in the heme binding region, although the general topology of the polypeptide is conserved.

Finally it may be noted that all three bacterial b/c_1 subunits are larger then their mitochondrial counterparts. The extra sequence elements found in the bacterial polypeptides might perform certain functions that in the mitochondrial b/c_1 complex are performed by the numerous additional subunits.

We are grateful to J. Hoppe, J. E. G. McCarthy and H. U. Schairer for help and stimulating discussions and to P. Cotoras for the excellent technical assistance. Nadia Gabellini thanks the Commission of the European Communities for the award of a training contract in Biomolecular Engineering of one year duration.

REFERENCES

1. Hauska, G., Hurt, E., Gabellini, N. & Lockau, W. (1983) *Biochim. Biophys. Acta 726*, 97–133.
2. Gabellini, N., Bowyer, J. R., Hurt, E., Melandri, A. & Hauska, G. (1982) *Eur. J. Biochem. 126*, 105–111.
3. Gabellini, N., Harnisch, U., McCarthy, J. E. G., Hauska, G. & Sebald, W. (1985) *EMBO J. 2*, 549–553.
4. Yu, L. & Yu, C. (1982) *Biochem. Biophys. Res. Commun. 108*, 1285–1292.
5. Von Jagow, G. & Sebald, W. (1980) *Annu. Rev. Biochem. 49*, 281–314.
6. Alt, J., Westhoff, P., Sears, B. B., Nelson, N., Hurt, E., Hauska, G. & Herrmann, R. G. (1983) *EMBO J. 2*, 979–986.
7. Hanahan, D. (1983) *J. Mol. Biol. 166*, 557–580.
8. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, New York.
9. Henikoff, S. (1984) *Gene 28*, 351–359.
10. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl Acad. Sci. USA 74*, 5463–5467.
11. Farrel, P. J., Deininger, P. L., Bankier, A. & Barrel, B. G. (1983) *Proc. Natl Acad. Sci. USA 80*, 1565–1569.
12. Berk, A. J. & Sharp, P. A. (1978) *Proc. Natl Acad. Sci. USA 75*, 1274–1278.
13. Wachter, E., Machleidt, W., Hofner, H. & Otto, J. (1973) *FEBS Lett. 35*, 97–102.
14. Hoppe, J., Gatti, D., Weber, H. & Sebald, W. (1986) *Eur. J. Biochem.*, in the press.
15. Von Heijne, G. (1981) *Eur. J. Biochem. 120*, 275–278.
16. Gold, L., Pribnow, D., Schneider, T., Shinedling, S., Singer, B. S. & Stormo, G. (1981) *Annu. Rev. Microbiol. 35*, 365–403.
17. Gibson, J., Stackebrandt, E., Zablen, L., Gupta, R. & Woese, C. R. (1979) *Curr. Microbiol. 3*, 59–64.
18. Shine, J. & Dalgarno, K. (1975) *Nature (Lond.) 254*, 34–38.
19. Harnisch, U., Weiss, H. & Sebald, W. (1985) *Eur. J. Biochem. 149*, 95–99.
20. Gabellini, N. & Hauska, G. (1983) *FEBS Lett. 154*, 171–173.
21. Tybulewicz, V. L. J., Falk, G. & Walker, J. E. (1984) *J. Mol. Biol. 179*, 185–214.
22. Pribnow, D. (1975) *Proc. Natl Acad. Sci. USA 72*, 784–788.
23. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet. 13*, 319–353.
24. Platt, T. (1981) *Cell 24*, 10–23.
25. Crofts, A. R., Meinhardt, S. W., Jones, K. R. & Snozzi, M. (1983) *Biochim. Biophys. Acta 273*, 202–218.
26. Mitchell, P. (1976) *J. Theor. Biol. 62*, 327–367.
27. Yu, C. & Yu, L. (1981) *Biochim. Biophys. Acta 639*, 99–128.

28. Wakabayashi, S., Takao, T., Shimonishi, Y., Kuramitsu, S., Matsubara, H., Wang, T., Zang, Z. & King, T. E. (1985) *J. Biol. Chem. 260*, 337–343.
29. Widger, W. R., Cramer, W. A., Herrmann, R. G. & Trebst, A. (1984) *Proc. Natl Acad. Sci. USA 81*, 674–678.
30. Saraste, M. (1984) *FEBS Lett. 166*, 367–372.
31. Nobrega, F. G. & Tzagoloff, A. (1980) *J. Biol. Chem. 255*, 9828–9837.
32. Rieske, J. S., MacLennan, D. H. & Coleman, R. (1964) *Biochem. Biophys. Res. Commun. 15*, 338–344.
33. Von Heijne, G. (1984) *EMBO J. 3*, 2315–2318.
34. Sadler, I., Suda, K., Schatz, G., Kaudewitz, F. & Haid, A. (1984) *EMBO J. 3*, 2137–2143.
35. Willey, D. L., Auffret, A. D. & Grey, J. C. (1984) *Cell 36*, 555–562.
36. Alt, J. & Herrmann, R. G. (1984) *Curr. Genet. 8*, 551–557.
37. Wakabayaschi, S., Matsubara, H., Kim, C. H., Kawai, K. & King, T. E. (1980) *Biochem. Biophys. Res. Commun. 97*, 1548–1554.
38. Kaminsky, L. S., Chiang, Y. L. & King, T. E. (1975) *J. Biol. Chem. 250*, 7280–7287.
39. Stonehuerner, J., O'Brien, P., Geren, L., Millett, F., Steidl, J., Yu, L. & Yu, C. (1985) *J. Biol. Chem. 9*, 5392–5398.
40. Broger, C., Salardi, S. & Azzi, A. (1983) *Eur. J. Biochem. 131*, 349–352.
41. Papanicolau, C., Gouy, M. & Ninio, J. (1984) *Nucleic Acids Res. 12*, 31–44.