DNA sequence of a gene cluster coding for subunits of the F₀ membrane sector of ATP synthase in *Rhodospirillum rubrum*

Support for modular evolution of the F₁ and F₀ sectors

Gunnar FALK*† and John E. WALKER†‡

*Department of Biochemistry, The Arrhenius Laboratory, University of Stockholm, Stockholm S-106 91, Sweden, and †Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

A region was cloned from the genome of the purple non-sulphur photobacterium *Rhodospirillum rubrum* that contains genes coding for the membrane protein subunits of the F₀ sector of ATP synthase. The clone was identified by hybridization with a synthetic oligonucleotide designed on the basis of the known protein sequence of the dicyclohexylcarbodi-imide-reactive proteolipid, or subunit c. The complete nucleotide sequence of 4240 bp of this region was determined. It is separate from an operon described previously that encodes the five subunits of the extrinsic membrane sector of the enzyme, F₁-F₀-ATPase. It contains a cluster of structural genes encoding homologues of all three membrane subunits a, b and c of the *Escherichia coli* ATP synthase. The order of the genes in *Rsp. rubrum* is a-c-b'--b where b and b' are homologues. A similar gene arrangement for F₀ subunits has been found in two cyanobacteria, *Synechococcus* 6301 and *Synechococcus* 6716. This suggests that the ATP synthase complexes of all these photosynthetic bacteria contain nine different polypeptides rather than eight found in the *E. coli* enzyme; the chloroplast ATP synthase complex is probably similar to the photosynthetic bacterial enzymes in this respect. The *Rsp. rubrum* b subunit is modified after translation. As shown by N-terminal sequencing of the protein, the first seven amino acid residues are removed before or during assembly of the ATP synthase complex. The subunit-a gene is preceded by a gene coding for a small hydrophobic protein, as has been observed previously in the *atp* operons in *E. coli*, bacterium PS3 and cyanobacteria. A number of features suggest that the *Rsp. rubrum* cluster of F₀ genes is an operon. On its 5' side are found sequences resembling the -10 (Pribnow) and -35 boxes of *E. coli* promoters, and the gene cluster is followed by a sequence potentially able to form a stable stem–loop structure, suggesting that it acts as a rho-independent transcription terminator. These features and the small intergenic non-coding sequences suggest that the genes are cotranscribed, and so the name *atp*2 is proposed for this second operon coding for ATP synthase subunits in *Rsp. rubrum*. The finding that genes for the F₁ and F₀ sectors of the enzyme are in separate clusters supports the view that these represent evolutionary modules.

**INTRODUCTION**

The membrane-bound proton-translocating ATP synthase (F₁-F₀-ATPase) enzymes isolated from mitochondria, chloroplasts and bacteria have a common function and closely related structures (for reviews see Senior, 1979; Fillingame, 1981). They catalyse the synthesis of ATP from ADP and Pᵢ by utilizing the energy stored in the electrochemical membrane potential gradient for protons ΔµH₊, (Mitchell, 1961, 1974; Nicholls, 1982). This gradient is generated by the vectorial translocation of protons across the membrane during oxidative or photosynthetic electron transport. However, the mechanism by which the ATP synthase couples proton translocation to phosphorylation is unknown. In strictly anaerobic bacteria the membrane potential gradient for protons is maintained by hydrolysis by ATP synthase of ATP generated by fermentation. In these circumstances the ATP synthase works in reverse and it pumps protons out of the bacterium. Purple non-sulphur photosynthetic bacteria such as *Rhodospirillum rubrum* have two modes of growth. They can use either light-energy when grown anaerobically or substrate oxidation under aerobic conditions in the dark to produce ATP. The same ATP synthase is employed in both cases (Baccarini-Melandri & Melandri, 1978). The ATP synthase from *Rsp. rubrum* has been isolated (Oren & Gromet-Elhanan, 1977; Bengis-Garber & Gromet-Elhanan, 1979) and characterized biochemically (Oren et al., 1980). It is similar to other ATP synthases, being a multi-subunit enzyme consisting of two structurally and functionally distinct sectors called F₁ and F₀. The F₁ domain is an extramembrane assembly that is attached to the intrinsic membrane sector F₀. It contains the catalytic and regulatory sites of the enzyme and can be digested from the membrane as a soluble assembly, F₁-ATPase (Johansson et al., 1973; Johansson & Baltscheffsky, 1975; Lücke & Klemme, 1976). It consists of five different polypeptides, α, β, γ, δ and ε, assembled with a probable stoichiometry of 3:3:1:1:1 respectively. Genes encoding these subunits in *Rsp. rubrum* have been cloned and sequenced previously (Falk et al., 1985) and have been shown to be co-transcribed (Falk & Walker, 1985). The proteins are highly homologous to the

---

1 To whom correspondence should be addressed.

Vol. 254
equivalent subunits from other sources (Falk et al., 1985). The $F_o$ portion of the enzyme contains a transmembrane proton channel through which the electrochemical membrane potential gradient for protons is coupled to ATP synthesis (Hoppe & Sebald, 1984). In Escherichia coli this is an assembly of three subunits, $a$, $b$ and $c$, and their stoichiometry is proposed to be $1:2:9-11$ respectively (Foster & Fillingame, 1982). Analysis of mutant strains of E. coli lacking one of the three $F_o$ subunits (Friedl et al., 1983) and reconstitution experiments with isolated subunits (Schneider & Altendorf, 1985) indicate that all three proteins are necessary for a functional $F_o$ complex that can translocate protons and bind $F_1$. The proteolipid ($c$ subunit) of $F_o$ is very hydrophobic and reacts rapidly with $NN'$-dicyclohexyl-carbodi-imide, an inhibitor of ATP synthase. Its primary structure has been determined from a number of organisms, including Rsp. rubrum (Hoppe & Sebald, 1984).

In the work described in the present paper the protein sequence of Rsp. rubrum subunit $c$ has been used to design a synthetic oligonucleotide probe, which has been employed in hybridization experiments to isolate a clone containing the corresponding gene. DNA sequence analysis of this region of the Rsp. rubrum genome has revealed that the gene for subunit $c$ is part of a cluster of five genes. It is immediately preceded by a gene coding for a homologue of $a$ subunits of the $F_o$ sector of other ATP synthases and is followed by two genes that code for proteins that are both homologous to the $b$ subunit of E. coli ATP synthase. The fifth gene, the first in the cluster, precedes the gene for the $a$ subunit and also encodes a hydrophobic protein. It is not homologous to any known protein, but its location suggests that it may be analogous to the unc1 gene in the E. coli unc (or atp) operon (Gay & Walker, 1981) and to relatives discovered in a similar location in operons encoding ATP synthase subunits in cyanobacteria (Cozens & Walker, 1987; H. S. van Walraven & J. E. Walker, unpublished work) and in a thermophilic bacterium PS3 (Ohta et al., 1988). Therefore the gene order is the same as that found in clusters encoding the subunits of the $F_o$ sector of ATP synthases in other micro-organisms (Walker et al., 1984; Cozens & Walker, 1987; Ohta et al., 1988; H. S. van Walraven & J. E. Walker, unpublished work), but Rsp. rubrum is the only case where the genes for the $F_o$ and $F_1$ sectors are segmented precisely.

**MATERIALS AND METHODS**

**Chemicals and biochemicals**

DNA polymerase I (Klenow fragment) was obtained from Boehringer–Mannheim and T4 DNA ligase from New England Biolabs. Bacteriophage T7 DNA polymerase (Sequenase) was obtained from United States Biochemical Corp., Cleveland, OH, U.S.A. All other enzymes were from Pharmacia PL Biochemicals. Radiocchemicals were obtained from Amersham International. Other reagents were purchased from BDH Chemicals, Poole, Dorset, U.K., Sigma Chemical Co. and Merck except for deoxyribonucleotides and dideoxynucleotides, which were from Pharmacia PL.

**Oligonucleotide synthesis**

A mixture of eight oligonucleotides, 24 bases long containing deoxynosine in three positions, was purchased from Kabigen AB, Stockholm, Sweden. It has the sequence 3'TACCTRCGICTYCGICTTATC'S, which corresponds to the protein sequence Met-Asp-Ala-Glu-Ala-Ala-Lys-Met, residues 1–8 determined in the $c$ subunit (Hoppe & Sebald, 1984). Its minimum ‘melting’ temperature is 62 °C, as calculated by the ‘rule of thumb’ that each A·T base-pair contributes 2 °C and each G·C base-pair 4 °C (Suggs et al., 1981), the contribution to duplex stability of base-pairs involving inosine residues being ignored (Martin & Castro, 1985; Ohtsuka et al., 1985). Specific sequencing primers were synthesized with an Applied Biosystems 380 B automated oligonucleotide synthesizer.

**Transfer of DNA to nitrocellulose and hybridization**

Fractionated digests of DNA were transferred from agarose gels to nitrocellulose filters as described by Southern (1975). When oligonucleotide probes were used, the baked filters were incubated at 42 °C for 4–16 h with sonicated salmon sperm DNA (0.1 mg/ml) dissolved in a solution containing $5 \times$ SSPE (1 $\times$ SSPE contains 180 mM-NaCl, 10 mM-NaH$_2$PO$_4$, pH 7.4, and 1 mM-EDTA), $5 \times$ Denhardt’s solution (1 $\times$ Denhardt’s solution contains 0.02% bovine serum albumin, 0.02% Ficoll and 0.02% polyvinylpyrrolidone) and 0.1% SDS. Hybridization was carried out by shaking filters slowly at 50 °C for 15–20 h with the oligonucleotide, which had been radioactively labelled to at least $10^5$–$10^6$ c.p.m./ml with $[\gamma-32P]ATP$ in the presence of kinase. The oligonucleotide concentration was approx. 10 ng/ml of hybridization solution. Free $[\gamma-32P]ATP$ was separated from labelled oligonucleotide by centrifugation of the sample through a column (8 cm $\times$ 0.5 cm internal diam.) of Sephadex G-25 (coarse grade) in 10 mM-Tris/HCl buffer, pH 8.0, containing 0.1 mM-EDTA. Washing of filters was performed twice in 6 $\times$ SSC (1 $\times$ SSC contains 150 mM-NaCl and 15 mM-sodium citrate) at room temperature for 10 min each and then twice at 55 °C for 30 min each in 6 $\times$ SSC. Autoradiography was carried out for 2–5 days at −70 °C with Fuji RX150 film in the presence of a fluorescent screen. Single-stranded probes were made by the ‘prime-cut’ method (Farrell et al., 1983). Hybridization with ‘prime-cut’ probes was carried out as described previously (Tybulewicz et al., 1984) except that the washes were carried out three times in the presence of 0.2 $\times$ SSC for 45 min at 65 °C.

**Preparation of genomic library**

The extraction of DNA from cells of Rsp. rubrum and the construction of a library in the vector λ2001 have been described previously (Falk et al., 1985).

**Screening of the genomic library with labelled oligonucleotide**

The library of Rsp. rubrum DNA in bacteriophage λ2001 was placed on E. coli Q358 grown on 82 mm-diameter agar plates. About 2000 plaque-forming units were put on each plate. The plaques were screened in accordance with Benton & Davis (1977). The probe and conditions used were as described above. Between two and four plaques hybridized on each plate. Ten recombinants were re-screened and four positively hybridizing recombinants, ARR1–ARR4, were chosen for further characterization. DNA from recombinant bacteriophages was prepared in accordance with Maniatis et al. (1982).
Sub-cloning for DNA sequencing

A fragment 1.4 kb in length present in XhoI digests of all four recombinant bacteriophages λRR1–λRR4 hybridized to the oligonucleotide probe. It was isolated and sub-cloned into the plasmid vector pSVL (Pharmacia PL) and into M13 mp19 (Yanisch-Perron et al., 1985). This permitted the fragment to be amplified and its DNA sequence to be determined from the ends of the fragment. Subsequently two adjacent SacI fragments, 2.1 kb and 2.3 kb respectively, overlapping the entire XhoI fragment were identified by hybridization and sub-cloned into the same vectors. Plasmid preparations were carried out by the alkaline lysis procedure as described in Maniatis et al. (1982). The amplified fragments were excised from the plasmids by restriction-endonuclease digestion and purified by electrophoresis in low-melting-point agarose (Wieslander, 1979). DNA fragments were recovered from melted agarose by extraction with phenol and precipitation with ethanol. Random libraries of the sub-cloned fragments were generated by breaking them up by sonication and end-repair of the resultant DNA segments, followed by cloning into M13 mp8 (Messing & Vieira, 1982) that previously had been digested with SmaI and treated with phosphatase (Deininger, 1983).

DNA sequencing

This was performed by the dideoxynucleotide chain-termination method (Sanger et al., 1977) as modified by Biggin et al. (1983). Most of the sequence was determined by the random strategy (Bankier & Barrell, 1983). It was completed by using unique synthetic oligonucleotide primers, 17 bases in length, and in some cases clones in M13 mp8 were turned around to determine complementary sequences (Bankier & Barrell, 1983). ‘Compressions’ in DNA sequences were resolved by substitution of dGTP by dITP in the sequence reactions (Mills & Kramer, 1979). Bacteriophage T7 DNA polymerase (Sequenase) was used instead of the Klenow fragment in some sequencing reactions (Tabor & Richardson, 1987). It was found to be useful in resolving one ‘pile-up’ that had arisen when Klenow fragment was employed. However, it gives rise to its own specific artifacts in DNA sequencing reactions.

Data analysis

DNA sequences generated by the random strategy were compiled with the aid of the computer programs DBAUTO and DBUTIL (Staden, 1982b). ANALYSEQ (Staden, 1985) was used to predict protein coding regions in the nucleic acid sequence by a variety of statistical methods. Protein sequences were compared with known sequences of other ATP synthase subunits with DIAGON (Staden, 1982a), and further analysed for various features with ANALYSEP (R. Staden, unpublished work). Potential proteins encoded in unassigned reading frames were compared with the sequence database of the Protein Information Resource (PIR) by using FASTP (Wilbur & Lipman, 1983; Lipman & Pearson, 1985).

RESULTS AND DISCUSSION

Cloning and gene cluster and DNA sequence analysis

A specific 4.9 kb BamHI fragment in restriction-endonuclease digests of Rsp. rubrum DNA was found to hybridize to the mixed oligonucleotide probe under the conditions employed (see Fig. 1a). When the same probe was used to screen a library of Rsp. rubrum DNA in bacteriophage λ2001 four positively hybridizing bacteriophages, λRR1–λRR4, were isolated. They were purified and their DNA was extracted. Restriction-endonuclease digestion showed that the pairs bacteriophages ARR1 and λRR3 and bacteriophages ARR2 and λRR4 were identical. Subsequent hybridizations of digested DNA

![Fig. 1. DNA hybridizations in cloning the gene for the proteolipid subunit of Rsp. rubrum ATP synthase](image)

In (a) and (b) the synthetic oligonucleotide and in (c) the prime-cut probe extending from bases 1960 to 2179 were employed. (a) Digests of Rsp. rubrum DNA with HindIII (track 1), BamHI (track 2) and BglII (track 3). (b) Bacteriophages λRR1 (tracks 1–3) and λRR2 (tracks 4–6) digested with HindIII, BamHI and XhoI respectively. (c) Bacteriophages λRR1 (tracks 1 and 2) and λRR2 (tracks 3 and 4) digested with SacI and SacI plus XhoI respectively.
from bacteriophages λRR1 and λRR2 with the oligonucleotide probe showed that both contained an XhoI fragment of 1.4 kb that formed a stable duplex with the probe (see Fig. 1b). No significant hybridizing fragment was detected in the BamHI digests of these bacteriophages. This could be because the 4.9 kb BamHI fragmented detected in the digest of the genomic DNA was not present in the bacteriophage. Since the bacteriophage DNA was digested incompletely with BamHI, this has not been resolved. However, a BamHI site is present in the sequence at nucleotide 2179 (see Fig. 2). The XhoI fragment was sub-cloned from recombinant bacteriophage λRR1 and its DNA sequence was determined. This is nucleotide residues 1727–3191 in the sequence shown in Fig. 2. It encodes the C-terminal sequence of the α subunit, the complete sequences of the c and b′ subunits as well as the N-terminal sequence of the b subunit (see below).

In order to extend the sequence in both 5′ and 3′ directions, DNA from recombinant bacteriophages λRR1 and λRR2 was digested with a series of restriction enzymes all of which were known from the sequence to cut within the analysed XhoI fragment. By hybridization with a suitable ‘prime-cut’ probe derived from the XhoI fragment, two SacI fragments of 2.0 and 2.3 kb were detected in the bacteriophage λRR1 were chosen for sequence analysis (Fig. 1c).

The SacI fragments were both cloned into M13 mp19 and DNA sequences were determined in the flanking regions of the inserted DNA. This showed that the two SacI fragments were adjacent, in agreement with the hybridization data, and that the SacI site was overlapped by the XhoI fragment. Both SacI fragments were cloned in the plasmid vector pSVL, then amplified, and DNA sequences were determined by using the random sonication strategy as outlined in the Materials and methods section. The two SacI fragments correspond to nucleotide residues 1–2055 and 2056–4240 respectively in Fig. 2. On average each nucleotide of the total sequence of 4240 bp was determined 9.8 times, and, minimally, each nucleotide was determined at least once on each strand of the DNA. The contents of G + C residues in the region sequenced is 64.2% (T, 19.6%; C, 32.7%; G, 31.5%, A, 16.2%). This is in agreement with values of 63.8–65.8% reported previously for the Rsp. rubrum genome (Pfennig & Trüper, 1974). In the Rsp. rubrum atp1 operon and flanking regions the G + C content was 65.75% (Falk et al., 1985).

**Identification of genes**

The general DNA sequence analysis program ANALYSEQ (Staden, 1985) was used to translate the consensus DNA sequence in all six possible reading frames. Genes coding for the subunits of the F_{0} portion of ATP synthase in Rsp. rubrum were identified by comparisons with E. coli F_{0} subunit protein sequences (Walker et al., 1984) by using the computer program DIAGON (Staden, 1982a). The comparisons are summarized in Fig. 3 and are part of the basis for identification of F_{0} genes in the photosynthetic bacterium. The sequence of the c subunit determined by DNA sequence analysis agrees exactly with that determined by protein sequencing except that at position 34 in the protein sequence serine was identified (Hoppe & Sebald, 1984) whereas the DNA sequence codes for alanine. Residues 8–13 of b subunit correspond to the N-terminal sequence determined on a component of Rsp. rubrum ATP synthase complex (J. E. Walker, G. Falk & R. Lutter, unpublished work). So these experiments provide independent evidence for the identification of these two genes.

One option in ANALYSEQ, the positional base preference method (Staden & McLachlan, 1982; Staden, 1984), was used to calculate the probability of coding and predict potential genes in the DNA sequence (see Fig. 4). This method takes into account that a protein has an average amino acid composition, and that there are preferences for certain bases to occupy particular positions in codons, and that these are different in each of the three reading frames. A modified version of the program was used where the preferences for C or G in the third position of codons observed in Rsp. rubrum genes were taken into account by using the known gene coding for the c subunit in the cluster as a ‘standard
Fig. 3. Comparison of protein sequences of *E. coli* ATP synthase *F*<sub>0</sub> subunits with those coded in the *Rsp. rubrum* gene cluster

The comparisons were made with the computer program DIAGON (Staden, 1982a) by using a window length of 25 amino acid residues and a score of 280.

Fig. 4. Gene predictions for the *Rsp. rubrum* *F*<sub>0</sub> subunits of ATP synthase

This was performed with ANALYSEQ (Staden, 1985) by using the modified positional base-preference option with subunit *c* as a standard gene. The abscissae represent the DNA sequence in the direction of the sense strand and the three boxes represent the three phases of the DNA. The relative probability of coding is plotted on the ordinate and potential start and stop codons are shown on the abscissae and median lines respectively. The complementary strand is not predicted to contain potential genes over this region (calculation not shown).
Fig. 5. Postulated ribosome-binding sites for genes for F, subunits of Rsp. rubrum ATP synthase and for two potential genes

Boxes have been placed around the proposed Shine & Dalgarno sequences and initiation codons are underlined. Also shown is the sequence of the 3' end of 16 S rRNA in Rsp. rubrum (Gibson et al., 1979).

gene'. The program also plots in the same diagram the distribution of potential start and stop codons (see legend to Fig. 4). The possibility that GTG acts as an initiation codon, as has been observed previously for the Rsp. rubrum ATP synthase subunit δ (Falk et al., 1985), was also considered. Two open reading frames URF3 and URF4 were predicted as in the same phase of the DNA as the b'-subunit gene (see Fig. 4). No potential reading frame was evident on the complementary strand of the segment of DNA presented in Fig. 2.

The presence of potential ribosome-binding sequences (Shine & Dalgarno, 1974) on the 5' side of initiation codons was used as an additional criterion for identification of potential protein-coding regions. These sequences are complementary to the sequence 3'UUUCCUCCACUA5' near to the 3' end of 16 S rRNA in Rsp. rubrum (Gibson et al., 1979). The Shine & Dalgarno sequences for the proposed genes in this region of the Rsp. rubrum genome are summarized in Fig. 5. The arrangement of genes detected by these criteria is shown in Fig. 6 and the justification for their proposed identities is presented in the following sections.

(i) ATP synthase subunits. A somewhat surprising finding was that the Rsp. rubrum gene cluster contains two adjacent genes that we have called b' and b, which both encode proteins that are related to the b subunit of E. coli ATP synthase (Fig. 3). A similar duplication has been found previously in two cyanobacteria, in the mesophile Synechococcus 6301 (Cozens & Walker, 1987) and in the thermophile Synechococcus 6716 (H. S. van Walraven & J. E. Walker, unpublished work).

Pair-wise comparisons of the Rsp. rubrum b' and b subunits with E. coli b subunit and bacterium PS3b subunit (Ohta et al., 1988) (Fig. 7) show that the Rsp. rubrum b' protein has a stronger homology in the terminal regions of the protein whereas the Rsp. rubrum b protein is more closely related to the E. coli homologue in the central part. Comparison of the Rsp. rubrum b protein with itself revealed a duplicated and diverged internal repeated region, as also was found in E. coli (Walker et al., 1982). A similar repeat is not evidently present in the Rsp. rubrum b' protein. Sequence homology between the Rsp. rubrum homologues themselves is also apparent when the two sequences are aligned as shown in Fig. 8.

Comparison of the hydrophobic profiles of these proteins indicates that both the b and b' genes could code for proteins similar in structure to other b subunits, giving further support to their identification (Fig. 9). In common with other known b subunits, the Rsp. rubrum proteins would have a hydrophobic N-terminal domain, the rest of the protein being hydrophilic. In E. coli this structure of the b subunit has been interpreted as indicating that the protein is anchored in the membrane by its N-terminus in association with subunits a and c, and that the remaining portion is proposed to lie outside the lipid bilayer and interact with F, subunits (Walker et al., 1982). This model for the b subunit has been supported subsequently by experimental evidence (Hoppe et al., 1983a,b).

In the wheat chloroplast b protein it is known that 17 amino acid residues are removed from the N-terminus by post-translational processing to produce the mature b subunit (subunit l) that is assembled in the ATP synthase complex (Bird et al., 1985). The biological function of this extension is obscure since there is no apparent need for a leader sequence to direct this subunit through the membrane, such as is required by nuclear-encoded chloroplast proteins (Robinson & Austen, 1987). N-Terminal protein sequencing of the Rsp. rubrum b protein isolated from chromatophores of photosynthetically grown cells has shown that seven amino acid residues are removed by proteolysis before or during assembly of the ATP synthase complex (J. E. Walker, G. Falk & R. Lutter, unpublished work). The alignment of b and b' subunits from Synechococcus 6301 also showed that these proteins have N-terminal extensions relative to the E. coli b protein. This, together with a close homology to
**Rhodospirillum rubrum** F₀ gene cluster

Fig. 7. Pairwise comparisons of sequences of b subunits of ATP synthases from various species

The calculations were performed with DIAGON by using the parameters described in the legend to Fig. 3. One division represents ten amino acid residues. For sequences of b subunits from *E. coli* and bacterium PS3 see Gay & Walker (1981) and Ohta et al. (1988) respectively.

---

Fig. 8. Alignment of the sequences of subunits b and b' from the F₀ sector of *Rsp. rubrum* ATP synthase

Identical residues are boxed. The arrow indicates the start of the mature b subunit (G. Falk & J. E. Walker, unpublished work). The start of the mature b' subunit is not yet known.
the chloroplast protein, indicates that also the *Synechococcus* protein is processed after translation. One possible function for these N-terminal extensions in b subunits found in these photosynthetic organisms is that they are needed for correct insertion of the protein into specialized photosynthetic membranes.

The presence in *Rsp. rubrum* of two related genes for b and b' subunits implies that the ATP synthase complex in this organism may contain one subunit per complex of each of the homologues, rather than two identical b subunits as is found in the *E. coli* enzyme. This has also previously been suggested for the *Synechococcus* 6301 enzyme (Cozens & Walker, 1987) and is in agreement with present information on the chloroplast ATP synthase. Gel analysis in the presence of urea of the ATP synthase isolated from spinach and bean chloroplasts showed that the complexes probably contain nine subunits; five of them can be identified as components of CF$_1$ (Pick & Racker, 1979; Westhoff et al., 1985; Süss, 1986).

The *Rsp. rubrum* F$_1$F$_0$-ATPase complex was purified previously (Oren & Gromet-Elhanan, 1977; Bengis-Garber & Gromet-Elhanan, 1979) but no suggestion was made that the complex consists of nine subunits rather than eight. The best-defined cyanobacterial ATP synthase complex is from the thermophilic *Synechococcus* 6716; this has been partially purified (Lubberding *et al*., 1983; van Walraven, 1985) but the subunit composition has not been determined.

(ii) An analogue of *E. coli* unc$\text{I}$. Immediately upstream of the *Rsp. rubrum* gene for ATP synthase subunit a is an open reading frame that overlaps the proposed initiation codon of the $a$ gene by 8 bp. The predicted gene encodes a small and hydrophobic protein of 123 amino acid residues with an $M_r$ of 12460. The amino acid sequence of this protein shows no obvious homology to the *E. coli* unc$\text{I}$ protein. Its hydrophobic profile suggests that the protein contains two hydrophobic segments of about 25 amino acid residues in length, which are compatible with the presence in the protein of two transmembrane $\alpha$-helical segments (Fig. 10). The predicted protein has a net charge of $-3$. The *E. coli* unc$\text{I}$-gene product has a net charge of $+9$ and those from cyanobacterial genes have net charges of $+6$ (*Synechococcus* 6301) and $+5$ (*Synechococcus* 6716) respectively. (The presence of fMet in these proteins and that histidine residues contribute +0.5 are assumed in calculating these values.)

The hydrophobic profiles of the potential proteins encoded by gene I in *E. coli*, *Synechococcus* 6301, *Synechococcus* 6716 and bacterium PS3 are strikingly alike (Fig. 10). This suggests that these proteins have similar secondary structures. They all have four hydrophobic segments, which are compatible with the presence of four transmembranous $\alpha$-helical segments. The function of *E. coli* unc$\text{I}$ (gene 1) is unknown. A large segment of the gene can be deleted without apparent effect or the expression of ATP synthase genes (Gay, 1984). Small basic hydrophobic proteins are associated with ATP synthase in yeast (*Velours et al*., 1984) and bovine mitochondria (*Anderson et al*., 1982) (Fig. 10). The product of the yeast aap$\text{I}$ gene is required for assembly of the ATP synthase (*Macreadie et al*., 1983), but the function of the bovine A6L protein is unknown.
The calculations were made as described in the legend to Fig. 9. The sequences of proteins are found in the references indicated: *E. coli* unc (Gay & Walker, 1981); *Synechococcus* 6301 (Cozens & Walker, 1987); bacterium PS3 (Ohta et al., 1988); bovine mitochondria (Anderson et al., 1982; Fearnley & Walker, 1986); yeast mitochondria (Velours et al., 1984).

(iii) A DNA-binding protein. In addition to the genes coding for ATP synthase subunits and to the *Rsp. rubrum* gene 1, one complete potential gene URF4 has been sequenced. The protein sequence predicted from this gene was compared against the Protein Information Resource (PIR) database by using the rapid screening program FASTP of Lipman & Pearson (1985). In this way URF4 was found to encode a protein with sequences in common with several DNA-binding proteins. The region of homology includes the proposed DNA-binding motif (Pabo & Sauer, 1984). Pairwise comparison by using DIAGON (Staden, 1982a) confirmed the relationship with the bacteriophage P22 repressor protein C2 and showed that the homology is confined to the N-terminal region of both sequences (results not shown). The predicted URF4 product is most closely related to the two DNA-binding proteins P22 Rep C2 and 434 Cro (see the alignment in Fig. 11). These proteins are thought to interact with DNA in a similar manner. The protein secondary structure involved is believed to be two a-helices that are linked by a tight turn (Sauer et al., 1982; Pabo & Sauer, 1984).

(iv) Unidentified potential genes. The incomplete reading frame, URF3, preceding the *atp2* gene cluster from nucleotide residues 1 to 524 in Fig. 2 did not show any significant homology with any of the sequences in the PIR database, nor was any relationship found with other predicted reading frames surrounding *atp* operons in *Rhodopseudomonas blastoica* (Tybulewicz et al., 1984) or *Synechococcus* 6301 (Cozens & Walker, 1987).

Transcriptional and translational signals

The genes encoding ATP synthase *F*₆ subunits in *Rsp. rubrum* are grouped in one cluster. The 5' end of the cluster contains sequences resembling both the −35 consensus and the −10 (or Pribnow) box (Pribnow, 1975; 1979) of *E. coli* (Fig. 12). These sequences have been shown to act as promoters for *E. coli* RNA polymerase. The consensus promoter sequences are based on homologies among 112 well-defined promoters in *E. coli* (Hawley & McClure, 1983). A similar possible promoter sequence is also present on the 5' side of URF4 (Figs. 2 and 12). The site of initiation of transcription for *Rsp. rubrum* *F*₆-ATPase genes has been determined.
Fig. 11. Alignment of *Rsp. rubrum* URF4 protein with DNA-binding proteins

For references to the bacteriophage P22 repressor protein C2 and the bacteriophage 434 Cro protein see Sauer et al. (1982) and Pabo & Sauer (1984). The numbers refer to the amino acid residues in the sequence. The underlined region corresponds to the DNA-binding motif.

---

Fig. 12. Potential transcriptional promoters in *Rsp. rubrum*

The boxed *E. coli* consensus promoter sequence is from Hawley & McClure (1983). The dashed double arrow and the numbers between boxes indicated the number of intervening nucleotide residues. The initiation of transcription of the *Rsp. rubrum* *atp1* operon at the underlined guanosine residue has been established experimentally previously (Falk & Walker, 1985).

---

Fig. 13. Possible stem–loop structures of mRNA on the 3’ end of *Rsp. rubrum* *atp* clusters that resemble *E. coli* transcriptional terminators

See Rosenberg & Court (1979). (a) *atp1* operon, including the transcriptional termination base, which is arrowed (Falk & Walker, 1985); (b) potential transcriptional terminator for the *atp2* gene cluster.

---

previously. Sequences found upstream of this site were also related to *E. coli* promoters (Falk & Walker, 1985) (Fig. 12).

The 3’ non-coding region following the gene clusters contain sequences with the potential to form a stem–loop structure followed by a run of T residues (Fig. 13). This structure resembles the *E. coli* rho-independent transcription terminators and appears to be widely conserved in eubacteria (Rosenberg & Court, 1979). A similar secondary structure was also found upstream of the transcriptional termination base in the *Rsp. rubrum* *F*₅-ATPase operon (Falk & Walker, 1985) (Fig. 13). These features makes it likely that also the gene cluster coding for ATP synthase *F*₅ subunits in *Rsp. rubrum* form an operon for which we suggest the name *atp2*, although further experimentation is required to ascertain the significance of these proposed signals in the transcription of the ATP synthase *F*₅ genes.

All the open reading frames presented here except the proposed subunit-a gene start with an ATG initiation codon. The a-subunit reading frame has an alternative initiation point from the one suggested at the ATG codon at nucleotide 997 in Fig. 2. The proposed gene start that we have preferred is primary based on the probability of codon usage (Fig. 4) and on the hydropplot for the predicted protein. Protein coding genes in *E. coli* only rarely use GTG (Lodish, 1976) or TTG (Young et al., 1981) as the initiation codon, and, for example, the genes encoding the δ subunits of ATP synthase in *E. coli* (Gay & Walker, 1981) and *Rsp. rubrum* (Falk et al., 1985) start with GTG. In both cases N-terminal protein sequencing of the δ-subunit has confirmed the proposed gene starts (Walker et al., 1982; J. E. Walker, G. Falk & A. Strid, unpublished work).

Conservation of Shine & Dalgarno sequences has been noted previously in various non-sulphur photosynthetic bacteria, including *Rhodopseudomonas capsulata* (Youvan et al., 1984), *Rhodopseudomonas blastic* (Tybulewicz et al., 1984) and *Rhodopseudomonas sphaeroides* (Gabellini & Sebold, 1986).
Fig. 14. Organization of genes encoding ATP synthase subunits in bacteria and plant chloroplasts

The letters a, b, c, α, β, γ, δ and e indicate the ATP synthase subunit encoded in the gene. The letters I and X indicate genes of unknown function (Walker et al., 1984; Tybulewicz et al., 1984; Cozens & Walker, 1987; H. S. van Walraven & J. E. Walker, unpublished results). Chloroplast c subunit is also known as subunit III and chloroplast b as subunit I; b contains an intron (Bird et al., 1985). b' is a duplicated and diverged form of b. The dashed line signifies that the gene clusters are at least 15 kb apart and are separately transcribed.

Unco-ordinacy

The F₀ subunits of E. coli ATP synthase are assembled with a stoichiometry of a₃c₅b₂.β₂.γ.δ estimated on the basis of radioactivity-incorporation studies (Foster & Fillinger, 1982). The E. coli atp (or unc) operon is transcribed as a single mRNA molecule containing all of the structural genes. This implies that widely differing amounts of the various proteins are produced from the same transcript, a phenomenon referred to as ‘unco-ordinacy’ (Goldberger, 1979). A variety of possible explanations has been advanced (Futai & Kanazawa, 1983; Walker et al., 1984), but at present the one with experimental support is that regulation occurs during initiation of translation (Brusilow et al., 1983; Walker et al., 1984; McCarthy et al., 1985; McCarthy, 1988) and that interactions between the secondary structure in the mRNA and the rRNA of the translocating ribosomes enhances translation initiation for highly expressed genes (McCarthy et al., 1985; McCarthy, 1988). In E. coli a pyrimidine (T)-rich sequence upstream of the gene for subunit c has been shown to enhance translation initiation efficiency when inserted upstream of other genes (McCarthy et al., 1985; McCarthy, 1988). Similar sequences have been found in the translational initiation regions at other efficiently translated genes in E. coli and bacteriophage λ (McCarthy et al., 1985). A pyrimidine (T)-rich region is also present immediately upstream of the proposed Shine & Dalgarno sequence for the Rsp. rubrum subunit c gene, which may indicate a similar translational enhancement. An alternative suggestion is that translational initiation of the less-expressed genes is inhibited by such interactions (Brusilow et al., 1983). A search for potential stem-loop structures in the Rsp. rubrum F₀ gene cluster finds 13 hairpin loops with at least 8 bp in the stem. However, potential secondary structures have not been conserved in similar positions in the various bacterial atp operons.

Gene order and evolution of ATP synthase

The arrangements of genes for ATP synthase subunits in bacteria and also chloroplasts (summarized in Fig. 14) are related to each other and illustrate that there has been a strong tendency to conserve these orders during evolution. Moreover, the finding that the F₀ and F₁ segments of the Rsp. rubrum enzyme are encoded by two separately transcribed gene clusters supports the notion examined elsewhere (Walker & Cozens, 1986) that the two sectors of the enzyme evolved as separate modules.

We thank Professor H. Baltscheffsky for his continued interest in this work, and him and also Dr. S. Nordlund for their encouragement and support. G. F. has been supported by a grant from the Swedish Ministry of Education and by a grant from the Swedish Science Research Council.

REFERENCES
