Nucleotide sequence of the *Rhodospirillum rubrum* *atp* operon

Gunnar FALK,* Annie HAMPE† and John E. WALKER‡
M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

(Received 9 November 1984/18 January 1985; accepted 8 February 1985)

The nucleotide sequence was determined of a 8775-base-pair region of DNA cloned from the photosynthetic non-sulphur bacterium *Rhodospirillum rubrum*. It contains a cluster of five genes encoding F_1 -ATPase subunits. The genes are arranged in the same order as F_1 genes in the *Escherichia coli* *unc* operon. However, as in the related organism *Rhodopseudomonas blastica*, neither genes for components of F_0 , the membrane sector of ATP synthase, nor a homologue of the *E. coli* *uncI* gene are associated with this locus, as they are in *E. coli*.

Genes encoding the eight constituent polypeptides of *Escherichia coli* ATP synthase are grouped together in the *unc* or *atp* operon (Gibson, 1982; Walker *et al.*, 1984a). [The *unc* operon was so-named because mutants in some of its genes are uncoupled in oxidative phosphorylation. However, this is not true of *uncI*. So, as in *Rhodopseudomonas blastica*, we have preferred to use in *Rhodospirillum rubrum* the alternative name *atp* proposed by Hansen *et al.* (1981).] The operon also contains a ninth gene *uncI* (Gay & Walker, 1981); the protein it encodes is not a structural component of ATP synthase and its function is obscure (Gay, 1984). With the exception of *uncI*, the genes in the *unc* operon are in two sub-clusters (see Fig. 1). One encodes the three subunits of the intrinsic membrane sector, F_0 , and the other the five subunits of the F_1 assembly, which is attached to F_0 , but lies outside the membrane. A similar arrangement of F_1 genes is also found in one of the purple non-sulphur photosynthetic bacteria or Rhodospirillaceae, *Rhodopseudomonas blastica*; the F_1 genes are arranged in the same order as in *E. coli*, except that an extra gene is interposed between those for γ and β -subunits (Fig. 1). The function of this gene is unknown (Tybulewicz *et al.*, 1984). However, neither F_0 genes nor a homologue of *E. coli* *uncI* are associated with this locus. They are presumed to lie elsewhere in the bacterial genome.

As described in the present paper we have also studied ATP synthase genes in another member of the Rhodospirillaceae, *Rhodospirillum rubrum*. An advantage of this organism over *Rps. blastica* is that its ATP synthase has been characterized biochemically (see Baccarini-Melandri & Melandri, 1978). In particular, the F_0 subunit, the proteolipid, has been isolated and its protein sequence determined (Hoppe & Sebald, 1984), whereas no F_0 subunit in *Rps. blastica* has been studied. So, potentially, this sequence information could be used to isolate the corresponding gene. The studies described in this earlier paper and the studies of transcription of the genes described here (Falk & Walker, 1985) suggest that F_1 genes in *Rsp. rubrum* also form an operon and, as in *Rps. blastica*, are not associated with F_0 genes. The gene order is that of the F_1 cluster in *E. coli* (Walker *et al.*, 1984b).

Materials and methods

Cloning and preparation of DNA

Rsp. rubrum strain S1 was grown anaerobically in light at 30°C in the medium described by Bose *et al.* (1961). Cells were harvested towards the end of exponential growth and washed in buffer containing 0.15M-NaCl, 0.1M-EDTA and 0.05M-Tris/HCl, pH 8.0, and DNA was extracted by the method of Marmur (1961). DNA was partially digested with endodeoxyribonuclease *Sau*3A. Fragments in the size range 15–20kb were fractionated from the digest by electrophoresis in low-gelling-temperature agarose and recovered by phenol extraction and ethanol precipitation. Then they were cloned into bacteriophage λ 2001 (Karn *et al.*, 1983) as follows. Vector DNA was first

Abbreviations used: kb, kilobases; SSC, 150mM-NaCl/15mM-sodium citrate buffer, pH 8.5

* Present address: The Arrhenius Laboratory, Department of Biochemistry, University of Stockholm, Stockholm, Sweden.

† Present address: Centre Hayem, Hôpital St. Louis, Paris, France.

‡ To whom correspondence should be addressed.

Fig. 1. Order of genes in (a) *E. coli* *unc* operon and (b) *Rps. blastica* *atp* operon (Walker *et al.*, 1984b; Tybulewicz *et al.*, 1984)

With the exception of *E. coli* gene 1 the genes are denoted by letters of the subunits they encode. In the *unc* operon these correspond to previously designated names as follows: gene 1 = *uncI*; a = *uncB*; c = *uncE*; b = *uncF*; δ = *uncH*; α = *uncA*; γ = *uncG*; β = *uncD*; ϵ = *uncC*. The functions of the *E. coli* gene 1 and the *Rps. blastica* X gene are unknown. Transcription proceeds from left to right in both operons. F₁ and F₀ refer to the F₁ and F₀ complexes of ATP synthase.

Fig. 2. Hybridization of segments of α - and β -subunits of *Rps. blastica* ATPase with *Rsp. rubrum* DNA (a) Genomic digests of *R. rubrum* cut with (a) *Bam*HI, (b) *Eco*RI, (c) *Hind*III and (d) *Xho*I hybridized with probe from *Rps. blastica* β -subunit (see the Materials and methods section). (b) and (c) λ recombinant RRII of *Rsp. rubrum* digested with (e) and (k) *Sac*I, (f) and (l) *Sac*I and *Xho*I, (g) and (m) *Xho*I, (h) and (n) *Xho*I and *Bam*HI, (i) and (o) *Bam*HI, and (j) and (p) *Sac*I and *Bam*HI. (b) was hybridized with the *Rps. blastica* β -subunit probe and (c) with the *Rps. blastica* α -subunit probe. The positions of markers are indicated in kilobases.

cleaved with *Bam*HI and *Eco*RI and fragments of *Rsp. rubrum* DNA were ligated to it, packaged *in vitro* and transfected into *E. coli* Q358 (Karn *et al.*, 1980). The library contained about 50000 plaque-forming units. The bacteriophages were amplified in *E. coli* Q359 and stocks were stored over chloroform at 4°C in λ dil medium (10mM-Tris/HCl containing 20mM-MgSO₄, 0.2M-NaCl and 0.1% gelatin, final pH 7.4).

Transfer to nitrocellulose and hybridization

Fractionated digests of DNA were transferred from agarose gels to nitrocellulose as described by Southern (1975). After transfer, the nitrocellulose filters were prehybridized with yeast RNA (50 μ g/ml) dissolved in 6 $\times$ SSC containing 0.2% bovine serum albumin, 0.02% Ficoll, 0.02% polyvinylpyrrolidone and 0.1% *N*-laurylsarcosine by

Vol. 228

V A A K H R S F A V P G M I G A F L E R L A A R R G E V T A R I V S A T A L T S
 66TCGCGGCGAAGCATCGTTCTTCCGCGTCCCCGCGATGATGGCGGCTTCTTGAACGCTCGCCGCCCGCGGGGCGAAGTACCGCCCGGATCGTACGCGCCACGGCCCTTACATC
 3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960
 CCAAGCGCGCTTCGTGCAAGGAAGCGGAGGGCGCTAGCCGCGCAAGGAACCTTGGAGAGCGGGCGGCCCGCTTCACTGGCGGCGCTAGCAGTCGCGGTGCGGGAATGTAG

A Q K S A L T T A L N K A T G N T V T I D A S V D P A L L G G M V V R V G S R M
 66CCAGAAAAGCGCCTGACGACGGCTCTCAATAAAGCAACGGGAATACGGTCACCATCGACGCTCGGTGGATCGGCGCCCTTCTGGCGGGATGGTCGCGGTGGATCGCGAT
 3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080
 CCGGGTCTTTTGGCGGACTGCTCGCGAGATTATTTGGTGCCCTTATGCAAGTGGTAGCTGCGGAGCCAACTAGGCGGGGAAGACCGCCCTACCAGCAGCGCGCAACCTAGCGCAT

α-subunit
 Bam HI
 V D S S L S T K L K R L Q L A M K G V G * H E I R A A E I S A I L K E Q I A N F
 66TCGACAGTTCCTCAGCAAGCTCAACCGTTCGACGTCGCGATGAGGGTGGTGGTGAATCCGCGCTGCGGATGATCTCCGCGATCTCAAGGAGCAGATCGCAATTTT
 4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
 CCAGCTGTCAAGGAGTGTGTTTCAGTTTGCCACGTTGAGCGGTACTTTCCCAACCCACTACCTTTAGGCGCGACGCTCTAGAGGCGCTAGGAGTTCCTGCTCTAGCGCTTAAG

Bgl II
 G T E A E S A E V G Q V L S V G G I A R A E V Y G L D N V Q A G E M V E F A N G V
 66AACCGAGCGAATCCGAGGTGGGTGAGTCTCGTGGCGAGCGGCATTCGCGGCTTACGGGCTGGCAACGTCAGGCGGGAATGGTCGAGTTCGCCAACGGCGTC
 4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320
 CTTGCTCGCGCTTAGCGGGTCCACCCAGTCAGGACAGCAGCGCGTCCGTAACGGGCGCAATGCCGACCTGTTGACGGTCCGGCCACTTTACCAGCTCAAGCGGTTGCCGCG

K G M A L N L E S D N V G I V I F G E D R G I K E G D V V K R T Q T I V D P V
 66GGCAGTTCCTTAACCTCGAAGCGCAACGTCGATCTCGGCGACGCGCATTCGCGGCTTACGGGCTGGCAACGTCAGGCGGGAATGGTCGAGTTCGCCAACGGCGTC
 4330 4340 4350 4360 4370 4380 4390 4400 4410 4420 4430 4440
 TTCCGCTACCGGAATTGGAGCTTTCGCTGTTGTCAGCCATAGCAAGCCGCTTCTAGCGCGTAGTTCCTTCGCTACAGCAGTTCGCGTAGTCTGCTAGCAGCTGACCGGCCAC

G K G L L G R V D G L G N P I D G K G G D L V D V E R K R A E V K A P G I I P R
 66GACGGCTGCTTGGCGGCTGGTGCATGCGCTGGGCAATCCGATCGACGGCAAGGCGATTTGGTCGACGTCGAGCGCAAGCGCGCGAGGTGACGCGCGGGGCAATCATCCGCGC
 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560
 CGGTTCGCGGACGAACCGCGCACCACTACCGGACCGGTTAGGCTAGCTGCCCTTCCCGCTAAACAGCTGACGCTCGCGTTCGCGGGCTCCAGTTCGCGGGCCGTAGTAGGGCGCG

K S V H E P V Q T G I K A I D S L I P I G R G Q R E L I I G D R Q T G K T A V I
 66AGTCGGTGACGAGCGGGTGCAGACCGGTCAAGGCCATCGACAGCTGATCCGATCGGCGGTGGCGGAGCTGATCATCGGTGACCGTCAGACGGCAAGACCGCGCTCATC
 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680
 TTCAGCCAGCTGCTCGGCGACGCTGCGCATAGTTCGCGACTAGGCTAGCGGCGACCGGTCGCGCTCGACTAGTACGCTGACGCTGCGGCTTTCGCGGGCCGTAGTAGGGCGAG

L D T I L N Q K A V N D K A K D D S E K L F C V Y V A V G Q K R S T V A Q V V K
 66CTGACACGATCCTCAATCAGAAGCGCTCAACGACAAGGGCAAGGACGATAGCGAGAAGCTGTTCTGCGCTATGTCGCGCTCGGTGAGAAGCGCTCGACGCTGGCCGAGGTGCTCAAG
 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800
 GAGCTGTGCTAGGAGTAGTCTTCGCGAGTTGCTGTTCCGCTTCTGCTATCGCTCTTCGACAGACGAGATACAGCGGACGAGCTTCTTCGAGCTGGCACCGGGCTCCAGCGATT

V L A D H G A L D Y T I V V A A T A S E P A P L Q F L A P Y T G C T M G E F F R
 66TCCTGGCGGATCAGCGCGCTGGATTACCATCGTCGTCGCGGCCACCGGCTCGAAGCCGCTCCCTTGCAGTTCTTGCTCCCTATACCGGCTGACGATGGCGGAGTTCTTCCGC
 4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920
 CAGGACCGCTAGTGCGCGGCACTAAATGTGGTAGCAGCGCGCGGCGGCTGGCGAGGCAAGCTCAAGGACCGAGGAGATGGCGGAGCTGCTACCGCTCAAGAAGGCGG

D N G M H A V I F Y D D L T K Q A V A Y R U M S L L L R R P P G R E A F P G D V
 66ACACGGCATGACCGCGGTGATCTTCTATGACGATCGACCAAGCAGGCGGCTGCGCTACCGTCAGATGTCGCTGCTGCTGCGCGCGCGCGGGCGCAAGCCTTCCCGGCGACGTG
 4930 4940 4950 4960 4970 4980 4990 5000 5010 5020 5030 5040
 CTGTTGCGCTAGTGCGCCACTAGAAGATCTGCTAGCTGGTTCGTCGCGGACCGGATGCGAGCTTACAGCGACGACGCGGGCGGGCGCGGCTTGGAGAGGGCGCGCTGAC

F Y L H S R L L E R A A K L N D D N G A G S L T A L P V I E T Q A N D V S A Y I
 66TTCTATCTCACTCCGTTTGTGTCGAAGCGCGGCAAGCTCAATGACGACACCGGCGGCTCCCTGACCGCCTGCGGTCATCGAGACCCAGGCGAAGCAGTGTGCGGCTATGTT
 5050 5060 5070 5080 5090 5100 5110 5120 5130 5140 5150 5160
 AAGATAGAGGTAGGCGCAACGAGCTTGGCGGCGGTTGAGTTACTGCTGTTGCGCGGCGGAGGAGTGGCGGAGCGGCGAGTAGCTCTGGGTCGCTGCTGACACCGGATATAA

P T N V I S I T D G Q I F L E T D L F F K G I R P A V N V G L S V S R V G S S A
 66CGACCAACGTGATTTGATCAGCAGCGGCGAGATCTTTTGGAAACGACCTGTTCTCAAGGGCATCGCGCGGCGGTAACGTCGCGCTGTGGTGTGCGCGGTTGCTCATCGGCC
 5170 5180 5190 5200 5210 5220 5230 5240 5250 5260 5270 5280
 GGTGGTTGCACTAAAGCTAGTGGCTGCGGCTAGAAAGACCTTTGGCTGGACAGAAAGTTCCTGAGCGGCGGCGCCTTCGAGCGGACAGCCACAGCGCGCAACCGAGTAGCGG

Q I K A M K Q V A G S I K L E L A Q Y R E M A A F A Q F A S D L D P A T Q K L L
 66CAGTCAAGGCGATGAAGCAGGTTGCGGCTCGATCAAGCTGGAACCTGCGCCAGTACCGTGAATGGCGGCTTCGCGCAGTTCGCTCCGACCTCGATCCCGCCACCCAGAAGCTTCTG
 5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400
 GTCTAGTTGCTGCTACTTCGTCACCGCGGAGCTAGTTGACCTTGAGCGGGTATGGCACTTTACCGCGGAAGCGGCTCAAGCGGAGGCTGGAGCTAGGCGGCTGGGCTTCTCAAGAC

A R G A R L T E L L K Q A Q Y S P L A V E E Q V C V I Y A G T R G Y L D K L K T
 66CGCGCGGCGCGGTTGACCAATGCTCAAGCAGGCGGAGTACTCGCGCTCGCGCTGAGGAAACAGGTTTGGTGTGATCTATGCGGCTACGCGGCGCTATCTGGACAGCTCAAGACC
 5410 5420 5430 5440 5450 5460 5470 5480 5490 5500 5510 5520
 CGGCGCGCGCGGCACTGGCTTAACGAGTTCTGTCGGGTATGAGCGGCGAGCGGCGAGCTCTTGTCCAAAGCAGTAGATACGGCATGCGCGCGATAGACCTGTTGAGTTCTGG

T D V V R Y E A S L L G A L R T S G A D L L E S I R T G K A L S K E I E Q K L V
 66ACGATGTTGGTTCGTTACGAGGCTTCGCTGCTGCGGCTTTCGCGACCTCGGGCGCGGCTTTCGAAATCGATCGGACCGGCAAGGCTGTCTCAAGGAAATCGAGCAGAGCTGCTG
 5530 5540 5550 5560 5570 5580 5590 5600 5610 5620 5630 5640
 TGGCTACACCGCAAGTCTCCGAAGCGACGAGCGCGAAGCGGTGGAGCCGCGGCTAGGAAGCTTAGCTAGGCGTGGCGGTTCCGAGACAGGTTCTTTAGCTGCTCTTCGACCA

γ-subunit
 M A S L K D L R S R I T
 K F L D D F G K K F A *
 66AGTTCCTGACGATTTTCGCAAGAAGTTTTCGTTGCTCAGGATTGACTCCAGAGGACGAGGAGGCTCCGCGAGCATGGCCAGGCTTAAAGACCTCAGAAGCGGATCA
 5650 5660 5670 5680 5690 5700 5710 5720 5730 5740 5750 5760
 TTCAAGGACCTGCTAAAGCGGTTCTTCAACGGACTAAGCGAGTGCTCACTAGGCTCTCCCTGGTCCCTTCCGAGGCTCTTGGTACCGGTCGGAATTTCTGGAGTCTTTCGCGCTAGT

S V K S T Q K I T S A M K M V A A S R L R R A Q D T A E A A R P Y T Q R M E R M
 66CTCGGTCAAAATCGACCCAGAAGTACCTCGGCGATGAAGATGGTCGCGGCTCGCGGCTGCGCGGTCGCGAGGATACCGCGAGGCGGCGGCGGCTTACCCAGCGCATGGAGCGGA
 5770 5780 5790 5800 5810 5820 5830 5840 5850 5860 5870 5880
 GGAGCCAGTTAGCTGGGCTCTTAGTGGAGCGCTACTTACCAGCGGCGGAGCGGCGGACGCGGCTCTATGGCGGCTCGCGCGGCGGCGGATAGGCTGCGTACCTCGCT

ATCATCTGAACATCTTGGCATGGAAGAGAGCTGCGAAGAGGACCGCTTGGTCGTGCGCGGCGCGCGCAAGATCQARGGCTTCTGTCTCCGACGGCTTCCATATGAGGCGAAGTGTTCACGGGG
7810 7820 7830 7840 7850 7860 7870 7880 7890 7900 7910 7920

TAGTAGCGGTAGGAACCGTACCTGCTCGACAGGCTTCTCTGGCAGACAGCAGCGGGCGGGCGGCTTCTAGGTGCGGGAAGGACAGGGTCGGGAAGGTACACGGGCTTCACAAGTGGCCG

Fig. 3. Nucleotide sequence of the *Rsp. rubrum atp operon* and adjacent DNA

Boxes before genes indicate ribosome-binding sites (Shine & Dalgarno sequences); proposed initiation codons are underlined. The *atp* mRNA start and termination sites are also indicated (Falk & Walker, 1985). Two *Bgl*II sites and a *Bam*HI site that were important in the cloning and sequencing experiments are marked by boxes.

incubation at 65°C for 3–4 h. The filters were then hybridized at the same temperature for 15–20 h in the presence of radioactive hybridization probe (10^4 – 10^5 c.p.m./ml) dissolved in the same solution as used for prehybridization. Then the filters were washed four times for 30 min at 65°C in a solution of the same composition except that $2 \times$ SSC replaced $6 \times$ SSC. Autoradiographs were developed with fluorescent screens for 1–3 days at -70°C .

Screening the genomic library

The plaque hybridization technique of Benton & Davis (1977) was used to identify λ bacteriophages containing genes for the ATP synthase α and β subunits. Single-stranded probes from *Rps. blastica* clones in bacteriophage M13mp8 were made by the prime cut method (Farrell *et al.*, 1983). The clones were selected from regions of the genes of α - and β -subunits of ATP synthase that are well conserved in *E. coli*, *Rps. blastica* and bovine mitochondria. The α -subunit probe extended from nucleotide 6740 to 7096 and the β -subunit probe from nucleotide 9591 to 9978 (see Tybulewicz *et al.*, 1984). The hybridization conditions described

above were employed. Seven positively hybridizing plaques containing parts of the α -subunit and/or β -subunit genes were isolated. Three of these recombinants hybridized to both probes.

Sub-cloning for DNA sequencing

An appropriate restriction fragment (4.7 kb) derived from one of the λ recombinants, λ RR11, hybridizing to both α -subunit and β -subunit probes, was sub-cloned into the vectors pUC9 (Vieira & Messing, 1982) and M13mp8 (Messing & Vieira, 1982), as was an adjacent 4.0 kb *Bam*HI-digest fragment. The amplified fragments were recovered from plasmid pUC9 by restriction-enzyme digestion followed by electrophoresis through low-melting-point agarose (Wieslander, 1979). Random libraries of the sub-clones were generated by sonication and end repair (Deininger, 1983) followed by cloning into bacteriophage M13mp8.

DNA sequence analysis

DNA sequences were determined by the deoxy chain termination method (Sanger *et al.*, 1977) as modified by Biggin *et al.* (1983). Other

details concerning sequencing, preparation of DNA templates, restriction-enzyme digestion, computer analysis of DNA sequences and sources of biochemicals have been described in detail before (Walker & Gay, 1983; Walker *et al.*, 1984b; Tybulewicz *et al.*, 1984).

Results and discussion

Cloning the operon

Regions of the genes for the α - and β -subunits of *Rps. blastica* ATPase were employed as hybridization probes. The β -subunit probe corresponds to a region encoding amino acids 125–254 of the protein, and the α -subunit probe was the complement of the sequence encoding amino acids 258–376. Thus they both cover parts of the genes that encode protein sequences in α - and β -subunits that are conserved over a wide range of species (Tybulewicz *et al.*, 1984). By plaque hybridization a recombinant, λ RR11, that hybridized to both probes was isolated (Figs. 2b and 2c), the hybridizing region being in a 4.7kb *Bam*H1-digest fragment. A fragment of the same size was also detected in the *Bam*H1 digest of genomic DNA (Fig. 2a, track a). The *Bam*H1 digest of λ RR11 contained two other fragments of 4.0kb and about 12kb (results not shown). From restriction-enzyme-digestion analysis it seemed likely that the 4.0kb and 4.7kb fragments were adjacent. They were both cloned into bacteriophage M13mp8 and DNA sequences were determined at their ends. At one end of each fragment sequences homologous to the *E. coli* δ -subunit of ATPase were detected. So both fragments were cloned and amplified in plasmid pUC9 and then their DNA sequences were determined by using a random strategy. The sequences of these two *Bam*H1-digest fragments were overlapped by cloning a 3.9kb *Bgl*II-digest fragment (extending from nucleotides 272–4168 in Fig. 2) into bacteriophage M13mp8 and then sequencing from one end across the *Bam*H1-cleavage site (around nucleotide 4033 in Fig. 3). The complete sequence of 8775 base-pairs is presented in Fig. 3. On average each nucleotide was determined nine times and minimally once on each strand of the DNA. In the region sequenced the G+C content was 65.75% (T, 17.42%; C, 33.01%; G, 32.74%; A, 16.82%). This corresponds well to the value of 63.8–65.8% reported earlier (Pfennig & Trüper, 1974).

Identification of genes

A search for genes encoding ATP synthase subunits was made by translation of the DNA sequence and comparison of the encoded protein sequences with *E. coli* F_1 -ATPase sequences (Walker *et al.*, 1984b). With the aid of the com-

puter program DIAGON (Staden, 1982) homologies were detected with each of the five *E. coli* F_1 proteins (Fig. 4), but not with *E. coli* F_0 proteins nor with the *Rsp. rubrum* proteolipid sequence (Hoppe & Sebald, 1984). Moreover, further comparisons of the *Rsp. rubrum* protein sequences with the *E. coli* *unc*I protein and the *Rsp. blastica* X protein also failed to find any significant homologies. The M_r values for *Rsp. rubrum* F_1 -ATPase subunits calculated from the predicted protein sequences were in the general agreement with values obtained by gel electrophoresis (Johansson & Baltscheffsky, 1975) (see Table 1).

A second analysis of the DNA sequence was carried out with the computer program ANALY-SEQ (Staden, 1984). An option in this program calculates and plots the probability that a given reading frame in the DNA is coding for a protein by comparison of its codon usage with that in a 'standard' gene. The method assumes that neighbouring genes will have similar codon preferences. The α -subunit codon usage was used as standard in Fig. 5. As a further aid to defining genes, the program also indicates on the probability plots the distribution of potential 'start' and 'stop' codons (see Fig. 5). This analysis produced the gene map shown in Fig. 6, which also incorporates the gene assignments made for ATP synthase genes as described above. It was found that preceding the identified ATP synthase genes for α , β , γ - and ϵ -subunits there are short sequences resembling the ribosome-binding sites or 'Shine and Dalgarno' (S&D) sequences of *E. coli* DNA (Shine & Dalgarno, 1974). As shown in Fig. 7, these sequences are complementary to the 5'-end of 16S ribosomal RNA and are believed to be important for correct initiation of translation. The sequence preceding the proposed initiation codon of the δ -subunit does not contain a canonical S&D sequence. However, the sequence G-G-G-G found there could function as a ribosome-binding site requiring the formation of a G·U base-pair with the 16S ribosomal RNA. G-rich tracts that may

Table 1. M_r values of *Rsp. rubrum* F_1 -ATPase subunits

Subunit	M_r	
	From DNA	By gel electrophoresis (Johansson & Baltscheffsky, 1975)
α	55026	54000
β	50852	50000
γ	32437	32000
δ	19543	13000
ϵ	14307	7500

Fig. 4. Comparison of protein sequences of *E. coli* F_1 -ATPase with those of *Rsp. rubrum* ATPase subunits. The comparisons were made with the DIAGON program (Staden, 1982) with the use of a window length of 25. Each dot represents sequences 25 amino acids in length that have double-matching probability (McLachlan, 1971) of 4×10^{-5} or less. With random sequences of the same composition this score would be attained once every 25000 sequences.

Fig. 5. Gene predictions for *Rsp. rubrum* atp operon and surrounding DNA by the gene prediction method of Staden & McLachlan (1982)

This is a function of the ANALYSEQ program (Staden, 1984). The abscissae (in a and b) represent the DNA sequence in its two orientations. The probability of coding is plotted on the ordinate. For further details see Tybulewicz *et al.* (1984).

Fig. 6. Graphical representation of the ATPase genes and upstream unidentified reading frames in *Rsp. rubrum*. Genes drawn above the line are transcribed from left to right and below from right to left. The symbol † represents the site of termination of transcription. The *atp* transcriptional start site (Falk & Walker, 1985) is indicated by an arrow with an open circle.

α -Subunit	CATGAAAGGGGTTGGGTCATGGAAATCC
β -Subunit	AATTTCAAGGAGCTTTATTTCATGGCCAAG
γ -Subunit	AGGGGAAGGGCTCCGAGAACCATGGCCA
δ -Subunit	TATACTTGGGGCTATCGCACGTGTCATC
ϵ -Subunit	ATCCAAAGGCGTTTGCCATGGCCGAAAC
URF1	CGCCTTAGGAAAGCGATGGCGGCTCT
URF1a	CGCCATCGGGGCGATACCTGATGGGC
URF1b	ATCATCCGGAGGTTCCTGATGATGGCGAC
URF2	ATGGCGCGGAGGTTCGCCGATGGCCGT
URF2a	TTCCTAAGGCGCGCATGGATGTTGCCGC

E. coli S&D sequence

AAGGAG

Rsp. rubrum 16S RNA
sequence

3'HO-UUUCUCCACUA- 5'

Fig. 7. Postulated ribosome-binding sites for *Rsp. rubrum* genes

Boxes have been placed around sequences that match the *E. coli* Shine & Dalgarno sequence and initiation codons are underlined. Also shown is the 3'-end of 16S RNA from *Rsp. rubrum* (Gibson *et al.*, 1979).

function as ribosome-binding sites have also been observed in bacteriophage λ (Sanger *et al.*, 1982).

In addition to the five F_1 genes, other potential genes (open reading frames) were detected by using the ANALYSEQ program in the region upstream of the F_1 cluster. Open reading frames URF1 and URF2 overlap on opposite strands of the DNA and are both preceded by potential ribosome-binding sites. Within these open reading frames are to be found other potential ribosome-binding sites (1a, 1b and 2a in Fig. 6). In the absence of experimental evidence concerning the protein products of this region, it is not possible to know more precisely the arrangement of genes in this region. Overlapping reading frames on opposite strands of the DNA are in part a consequence of the high G+C content, as there are fewer opportunities to form potential stop codons.

In an attempt to discover more about the nature of these putative proteins their sequences were compared not only with F_1F_0 -ATP synthase sequence but also with a wide variety of other

protein sequences (Doolittle, 1981). However, no significant homologies were detected. Moreover, these putative proteins are not related to the proteins encoded in DNA surrounding the *Rps. blastica atp* operon (Tybulewicz *et al.*, 1984).

Possible secondary structures

The DNA sequence was searched for hairpin loops with eight or more bases in the stem and a minimum loop size of four bases. Twenty-two possible structures were identified. However, most of them lie within coding sequences and their significance is unclear. It is possible that they could be involved in regulation of gene expression. Hairpin loops have been noted in the *unc* operon. Brusilow *et al.* (1982) suggested that they could help to regulate the un-co-ordinate expression of the cistrons in a post-translational mechanism.

An extensive region of stem/loop structures corresponding to the inverted repeats in Fig. 3 can form in the region following the ϵ -subunit gene.

Two of them resemble *E. coli* rho-independent transcriptional termination signals, having a G+C-rich stem structure followed by a run of T residues. As shown by experimentation, transcription of the *atp* genes does terminate in the region of nucleotide 8738 (Falk & Walker, 1985). Stem/loop structures of this type are not evident in the region preceding the δ -subunit gene.

Evolutionary considerations

The order of genes in the *Rsp. rubrum atp* operon is identical with that found for the corresponding genes in the *E. coli unc* operon. Moreover, it is also identical with the arrangement found in the *Rps. blastica atp* operon except that there is no extra gene interposed between those for the γ - and β -subunits. In both members of the Rhodospirillaceae neither genes for F_0 subunits nor a homologue of *uncI* are associated with genes for F_1 subunits. So association of F_0 -subunit and F_1 -subunit genes is not an obligatory arrangement for expression and assembly of F_1F_0 -ATPase. As discussed previously, it is more likely that the F_1 cluster reflects the evolution of F_1 -subunit genes from a common ancestor (Tybulewicz *et al.*, 1984), a conclusion that is supported in part by the finding that α - and β -subunits are related in sequence (Walker *et al.*, 1982a).

Together with sequences of F_1 subunits of other species, the sequences of the *Rsp. rubrum* F_1 subunits (see Fig. 8 for alignments) help to illustrate the evolution of the subunits of the enzyme complex.

Binary comparisons (see also Fig. 9) illustrate that the α -subunits, and particularly the β -subunits, are highly conserved, the γ -subunits somewhat less so, and, in contrast, the sequences of ϵ - and δ -subunits are rather poorly maintained. However, it is evident in Figs. 4 and 8 that conserved residues in γ -, δ - and ϵ -subunits are clustered, suggesting the presence of conserved structural elements.

Chemical modifications and other studies have shown that the β -subunit is involved in catalysis (summarized in Runswick & Walker, 1983), and both it and the α -subunit have nucleotide-binding sites (Harris, 1978). The roles of the γ -, δ - and ϵ -subunits are less well defined, but δ - and ϵ -subunits appear to be important for binding $\alpha_3\beta_3\gamma$ to the membrane (Futai *et al.*, 1974; Smith & Sternweis, 1977; Yoshida *et al.*, 1977; Sternweis, 1978). Given that F_0 subunits, as far as is known, are conserved rather poorly, it therefore seems to be likely that at least some of the conserved elements in α -, γ -, δ - and ϵ -subunits are those that form subunit interfaces, particularly with the β -subunit.

References

- Baccarini-Melandri, A. & Melandri, B. A. (1978) in *The Photosynthetic Bacteria* (Clayton, R. K. & Sistrom, W. R., eds.), pp. 615–628, Plenum Press, New York.
- Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
- Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965.
- Bose, S. K., Gest, H. & Ormerod, J. G. (1961) *J. Biol. Chem.* **236**, PC13–PC14.
- Brusilow, W. S. A., Klionsky, D. J. & Simoni, R. D. (1982) *J. Bacteriol.* **151**, 1363–1371.
- Deininger, P. L. (1983) *Anal. Biochem.* **129**, 216–222.
- Deno, H., Shinozaki, K. & Sugiura, M. (1983) *Nucleic Acids Res.* **11**, 2185–2191.
- Doolittle, R. F. (1981) *Science* **214**, 149–159.
- Falk, G. & Walker, J. E. (1985) *Biochem. J.* in the press.
- Farrell, P. J., Deininger, P. L., Bankier, A. & Barrell, B. G. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1565–1569.
- Futai, M., Sternweis, P. C. & Heppel, L. A. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 2725–2729.
- Gay, N. J. (1984) *J. Bacteriol.* **158**, 820–825.
- Gay, N. J. & Walker, J. E. (1981) *Nucleic Acids Res.* **9**, 3919–3926.
- Gibson, F. (1982) *Proc. R. Soc. London Ser. B* **215**, 1–18.
- Gibson, J., Stackebrandt, E., Zahler, L. B., Gupta, R. & Woese, C. R. (1979) *Curr. Microbiol.* **3**, 59–64.
- Hansen, F. G., Nielsen, J., Riise, E. & von Meyenburg, K. (1981) *Mol. Gen. Genet.* **183**, 463–472.
- Harris, D. A. (1978) *Biochim. Biophys. Acta* **463**, 245–273.
- Hoppe, J. & Sebald, W. (1984) *Biochim. Biophys. Acta* **768**, 1–27.
- Howe, C. J., Fearnley, I. M., Walker, J. E., Dyer, T. A. & Gray, J. C. (1985) *Plant Mol. Biol.* in the press.
- Johansson, B. C. & Baltscheffsky, M. (1975) *FEBS Lett.* **53**, 221–224.
- Karn, J., Brenner, S., Barnett, L. & Cesareni, G. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5172–5176.
- Karn, J., Brenner, S. & Barnett, L. (1983) *Methods Enzymol.* **101**, 3–19.
- Krebbers, E. T., Larrinua, I. M., McIntosh, L. & Bogorad, L. (1982) *Nucleic Acids Res.* **10**, 4985–5002.
- Marmur, J. (1961) *J. Mol. Biol.* **3**, 208–218.
- McLachlan, A. D. (1971) *J. Mol. Biol.* **61**, 409–424.
- Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
- Ovchinnikov, Y. A., Modyanov, N. N., Grinkevich, V. A., Aldanova, N. A., Trubetskaya, O. E., Nazimov, I. V., Hundal, T. & Ernster, L. (1984) *FEBS Lett.* **166**, 19–22.
- Pfennig, N. & Trüper, H. G. (1974) in *Bergey's Manual of Determinative Bacteriology*, 8th edn. (Buchanan, R. E. & Gibbons, N. E., eds.), pp. 24–75, Williams and Wilkins, Baltimore.
- Runswick, M. J. & Walker, J. E. (1983) *J. Biol. Chem.* **258**, 3081–3089.
- Saltzgeber-Müller, J., Kunapuli, S. P. & Douglas, M. G. (1983) *J. Biol. Chem.* **258**, 11465–11470.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467.
- Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F. & Petersen, G. B. (1982) *J. Mol. Biol.* **162**, 729–773.

-SUBUNIT
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

1 319
 MVTIRADEIS NIIRERIEQY NREVKIVNTG TVLQVGDGIA RIHGLDEVMA
 MATLRVDEIH KIVRELIEQY NRKVGIENIG RVVQVGDGIA RIIGLGEIMS
 .MQLNSTEIS ELIKQRIAQF NVVSZAHNEC TIVSVSDGVI RIHGLADCMQ
 .MGIQAAEIS AILKEQIKNF GKDAEVAEVC RVLSVGDGIA RVYGLDNVQA
 .MEIRAAEIS AILKEQIANF GTEAESAEVC QVLSVGDGIA RVYGLDNVQA

51 100
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

101 150
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

151 200
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

201 250
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

251 300
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

301 350
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

351 400
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

401 450
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

451 500
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

501 529
 Tobacco chloroplasts
 Wheat chloroplasts
E. coli
Rps. blastica
Rsp. rubrum

β -SUBUNIT

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blastica
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

1 50
MRINPTTSGS GVSTLEKKNP GRVVIIGFV LDVAFPPGKM PNIYNALVVQ
MRINPTTSDP GVSTLEKKNL GRIAIIGFV LNVAFPPGKM PNIYNALIVK
MRTNPTTSRP GVSTSEKST GRIDQIIGFV LDVTFPPGKL PNIYNALVVQ
MRTNPTTSPG GASTIEEKST GRIDQIIGFV LDVTFPPGKL PNIYNALVVQ
MRTNPTTSRP GISTIEEKSV GRIDQIIGFV LDITFPPGKL PNIYNALIVK
..... MAT..... GKIVQVIGAV VDVEFPQDAV PRVYDALEVQ
..... MT MATT...VSK GKVTQVIGAV VDVQF.EGVL PAILNALETT
..... MA....KNNL GTITQVTGAV VDVKF.EGEL PSILSALETD
....SPSPKA GATT..... GRIVAVIGAV VDVQFDEG.L PPIILNALEVQ

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blastica
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

51 100
GRDSVGQPIN VACEVQQLLG NNRVRAIAMS ATEGLTRGME VIDTGAPISV
GRDTAGQPMN VTCEVQQLLG NNRVRVAVAMS ATDGLTRGME VIDTGAPLSV
SRDTADKQIN VTCEVQQLLG NNRVRVAVAMS ATDGLMRGME VIDTGAPLSV
SRDTDDKQIN VTCEVQQLLG NNRVRVAVAMS ATDGLMRGME VIDTGAPLSV
SRDTADKQIN VTCEVQQLLG NNRVRVAVAMS ATEGLMRGME VIDTGTPLSV
NGNE.....R LVLEVQQLLG GGVVRTIAMG SSDGLRRGLD VKDLEHPIEV
NNGK.....K LILEVAQHLLG ENTVRCIAMG ATEGLVRGAP VSDSGTPISV
NHGN.....R LVLEVAQHLLG ESVRTIAMG STEGLVRGQK VTSTGGPIITV
GRET.....R LVLEVAQHLLG ESTVRTIAMG GTEGLVRGQK VLDSGAPIRI

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blastica
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

101 150
PVGGATLGRI FNVLGEFVDN LGPVDTSSTS PIHRSAPAFI QLDTKLSIFE
PVGGPTLGRI FNVLGEFVDN LRPVDTRTTS PIHRSAPAFI QLDTKLSIFE
PVGGATLGRI FNVLGEFVDN LGPVDSSATF PIHRSAPAFI ELDTKLSIFE
PVGGATLGRI FNVLGEFVDN LGPVDSSATF PIHRSAPAFI ELDTKLSIFE
PVGGATLGRI FNVLGEFVDN LGPVDTSATF PIHRSAPAFI ELDTKLSIFE
PVGKATLGRI MNVLGEFVDM KGEIGEEERW AIHRAAPSYF ELSNSQELLE
PVGNATLGRI LNVVGEFIDE RGPATAEKRG AIHQKAPDFY SQSTESQILV
PVGPQVLGRI MNVLGEFVDE RGPVVTAQRY PIHQRAPTFA EQATETEILV
PVGPETLGRI MNVLGEFIDE RGPVKTKQFA AIHAEAPFV EMSVEQELLV

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blastica
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

151 200
TGIEVVDDLLA PYRRGGKIGL FGGAGVGKTV LIMELINNIA KAHGGVSVFG
TGIKVVNDLLA PYRRGGKIGL FGGAGVGKTV LIMELINNIA KAHGGVSVFG
TGIKVVDDLLA PYRRGGKIGL FGGAGVGKTV LIMELINNIA KAHGGVSVFG
TGIKVVDDLLA PYRRGGKIGL FGGAGVGKTV LIMELINNIA KAHGGVSVFG
TGIKVVDDLLA PYRRGGKIGL FGGAGVGKTV LIMELINNIA KAHGGVSVFG
TGIKVIDLMC PFAKGGKVL FGGAGVGKTV NMFELIRNIA IEHSGYSVFA
TGIKVIDLLA PYSKGGKIGL FGGAGVGKTV LIQELINNIA KVHSGYSVFA
TGIKVIDLIA PYTKGGKVL FGGAGVGKTV LIQELINNIA KGHGGYSVFA
TGDKVVDLLA PYAKGGKIGL FGGAGVGKTV FIMELINNIA KAHGGFSVFA
TGDKVVDLLA PYAKGGKIGL FGGAGVGKTV FIMELINNIA KAHGGYSVFA

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blastica
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

201 250
GVGERTREGN DLYMEMKESG VINEKNIAES KVALVYGQMN EPPGARMRVG
GVGERTREGN DLYMEMKESG VINEKNIAES KVALVYGQMN EPPGARMRVG
GVGERTREGN DLYMEMKESG VINEKNIEES KVALVYGQMN EPPGARMRVG
GVGERTREGN DLYMEMKESG VINEKNIEES KVALVYGQMN EPPGARMRVG
GVGERTREGN DLYMEMKESG VINEKNIEES KVALVYGQMN EPPGARMRVG
GVGERTREGN DLYHEMTDSN VID..... KVSIVYGQMN EPPGNRLRVA
GVGERTREGN DLYHEFIESG VINIDDLKS KVALVYGQMN EPPGARARVA
GVGERTREGN DLYHEMIDAG IIDLE.GDKS KVALVYGQMN EPPGARARVA
GVGERTREGN DLYHEMEDSG VINLE..GES KVALVYGQMN EPPGARARVA
GVGERTREGN DLYHEMIESG VINLKDA.TS KVALVYGQMN OPPGARARVA

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blastica
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

251 300
LTALTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTALTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTALTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTALTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTALTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTALTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP
LTGLTMAEYF RDVNEQDVLL FIDNIFRFVQ AGSEVSALLG RMP SAVGYQP

β -SUBUNIT

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blasticla
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

301
TLSTEMGSLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLSTEMGSLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLSTEMGSLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLSTEMGSLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLSTEMGSLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLAEEMGVLLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLATDMGQLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLATDMGQLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLATDMGQLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLATDMGQLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLATDMGQLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT
TLATDMGQLQ ERITSTKKEGS ITSISQAVYVP ADDLTDPAFA TTFAHLDAATT

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blasticla
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

351
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE
VLSRGLAAGK IYPAVDPLDS TSTMLQPRIV GEEHYETAQR VKQTLQRYKE

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blasticla
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

401
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG
LQDIIAILGL DELSEEDRLT VARARKIERF LSQPFVFAEV FTGSPGKYVG

Tobacco chloroplasts
Spinach chloroplasts
Barley chloroplasts
Wheat chloroplasts
Maize chloroplasts
E. coli
Rps. blasticla
Rsp. rubrum
Yeast mitochondria
Bovine mitochondria

451
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK
LAETIRGFQL ILSGELDGLP EQAFYLVGNI DEATAKAMNL EMESNLKK

 γ -SUBUNIT

E. coli
Rps. blasticla
Rsp. rubrum

1
MAGAKEIRSK IASVQNTQKI TKAMEMVAAS KMRKSDRMA ASRPYAEINR
MPSLKDKNR IGSVKNTRKI TKAMQMVAAA KLRRAQDSAE AARPYAERMC
MASLKDLSR IISVKSQTKI TSAMKMVAAS RLRRRAQTAE AARPYTQRM

E. coli
Rps. blasticla
Rsp. rubrum

51
KVIGHLAHGN ..LEYKHPYL EDRDVKRVGY LUV.STDRGL CGGLNINLFK
AVIASLASGQ .G.AGAPRL AGNGRDQIHL LUVMTSERGL CGGFNSTIVR
RMLGNLAEST AOMAGASPL GGTGKDNVHL IVALTANRGL CGGFNGSIIR

E. coli
Rps. blasticla
Rsp. rubrum

101
KLLAEMKTWT DKGVQCDLAM IGSKGVNFF. .NSVGGNVVA QVTGMGDNPS
LARQRANELV AOKTKVKILT VCKKGREQLK RDWASAFVGH VDLSDVRRLG
ATRTLVRLE AOKTKVKLLC ICKKGRDLK REFPPKQIIG IADQSSKAIG

E. coli
Rps. blasticla
Rsp. rubrum

151
LSELIGPVKV MQAYDEGRL DKLYIVSNKF INTMSQVPTI SOLLPLPAS.
YSNAQGIARE VLAFAFEAGEA DVVTIFYNRF QSVISQVPTA QOVIPA....
FSDADRFSRL ILDMFQAGEF DVCTLVYNRF QSAISQVMT RQQILFFAVPT

E. coli
Rps. blasticla
Rsp. rubrum

201
.....DDDD LKHKSWDYLY EPDPKALLDT LLRRYVESQV YQGVVENLAS
.....KFE AETNALYDY EPSEEAIALD LLPRGVATQI FTALLENAS
TVAAGNDNDR TAGPKAIYEX EPSEEAIALD LLPKNMAIQV FRGMLSEFAS

E. coli
Rps. blasticla
Rsp. rubrum

251
EQAARMVANK AATDNCGSLI KELQLVYNKA RQASITQELT EIVSGAANV
EQGARMAMND NATRNAGDMI NKLTIQYNRS RQAATIKELI EIISCAEAL
EQGARMAMND NATRNAGDMI NKLSLTYNRT RQAATIKELI EIISCAEAL

Fig. 8. Alignment of protein sequences of F_1 -ATPase subunits from various sources

Identities are boxed. Dots denote gaps. Sequences are shown for all five subunits of *Rsp. rubrum*, *Rsp. blastica* (Tybulewicz *et al.*, 1984) and *E. coli* (Walker *et al.*, 1984a, b) for tobacco chloroplast α -subunit (Deno *et al.*, 1983) and β - and ϵ -subunits (Shinozaki *et al.*, 1983), spinach chloroplast β - and ϵ -subunits (Zurawski *et al.*, 1982), maize chloroplast β - and ϵ -subunits (Krebbes *et al.*, 1982) wheat chloroplast α -, β - and ϵ -subunits (Howe *et al.*, 1985) and bovine mitochondrial β -subunit (Runswick & Walker, 1983). The partial sequence of the yeast mitochondrial β -subunit has also been included (Saltzgaber-Muller *et al.*, 1983). Bovine oligomycin-sensitivity-conferral protein (OSCP) (Ovchinnikov *et al.*, 1984) is the counterpart of bacterial δ -subunits (Walker *et al.*, 1982b).

Shine, J. & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342-1346
Shinozaki, K., Deno, H., Kato, A. & Sugiura, M. (1983) *Gene*, **24**, 147-155
Smith, J. B. & Sternweis, P. C. (1977) *Biochemistry* **16**, 306-311
Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503-517

Staden, R. (1982) *Nucleic Acids Res.* **10**, 2951-2961
Staden, R. (1984) *Nucleic Acids Res.* **12**, 521-538
Staden, R. & McLachlan, A. D. (1982) *Nucleic Acids Res.* **10**, 141-156
Sternweis, P. C. (1978) *Biochemistry* **253**, 3123-3128
Tybulewicz, V. L. J., Falk, G. & Walker, J. E. (1984) *J. Mol. Biol.* **179**, 185-214

α -Subunit	1	2	3	4		
1. <i>E. coli</i>	100					
2. <i>Rps. blastica</i>	53.4	100				
3. <i>Rps. rubrum</i>	55.9	74.5	100			
4. Tobacco chloroplasts	53.3	57.9	60.2	100		
β -Subunit	1	2	3	4	5	
1. <i>E. coli</i>	100					
2. <i>Rps. blastica</i>	69.8	100				
3. <i>Rps. rubrum</i>	72.0	76.8	100			
4. Bovine mitochondria	71.8	76.4	76.2	100		
5. Tobacco chloroplasts	67.5	67.0	66.9	69.0	100	
γ -Subunit	1	2	3			
1. <i>E. coli</i>	100					
2. <i>Rps. blastica</i>	37.3	100				
3. <i>Rps. rubrum</i>	37.2	60.0	100			
δ -Subunit	1	2	3	4		
1. <i>E. coli</i>	100					
2. <i>Rps. blastica</i>	26.4	100				
3. <i>Rps. rubrum</i>	19.6	43.3	100			
4. Bovine mitochondria	26.4	30.5	28.9	100		
ϵ -Subunit	1	2	3	4		
1. <i>E. coli</i>	100					
2. <i>Rps. blastica</i>	32.1	100				
3. <i>Rps. rubrum</i>	32.6	32.0	100			
4. Tobacco chloroplasts	26.9	22.9	23.9	100		

Fig. 9. Binary comparison matrices of F_1 -ATPase subunits

Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268
 Walker, J. E. & Gay, N. J. (1983) *Methods Enzymol.* **97**, 195–218
 Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. (1982a) *EMBO J.* **1**, 945–951
 Walker, J. E., Runswick, M. J. & Saraste, M. (1982b) *FEBS Lett.* **146**, 393–396
 Walker, J. E., Saraste, M. & Gay, N. J. (1984a) *Biochim. Biophys. Acta* **768**, 164–200

Walker, J. E., Gay, N. J., Saraste, M. & Eberle, A. N. (1984b) *Biochem. J.* **224**, 799–815
 Wieslander, L. (1979) *Anal. Biochem.* **98**, 305–309
 Yoshida, M., Okamoto, H., Sone, N., Hirata, H. & Kagawa, Y. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 936–940
 Zurawski, G., Bottomley, W. & Whitfield, P. R. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6260–6264