Organization and sequences of genes for the subunits of ATP synthase in the thermophilic cyanobacterium *Synechococcus* 6716

Hendrika S. VAN WALRAVEN,*† René LUTTER*‡ and John E. WALKER*§

*Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.,

and †Biological Laboratory, Vrije Universiteit, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands

The sequences of the genes for the nine subunits of ATP synthase in the thermophilic cyanobacterium *Synechococcus* 6716 have been determined. The genes were identified by comparison of the encoded proteins with sequences of ATP synthase subunits in other species, and confirmed for subunits α , β , δ and ϵ , by determining their N-terminal sequences. They are arranged at three separate loci. Six of them are in one cluster in the order α : γ : δ : β : ϵ : α , and those for the β and ϵ subunits form a second and separate cluster. The gene for the γ -subunit is at a third site. As in other bacteria, the gene for subunit α is immediately preceded by a gene coding for a small hydrophobic protein of unknown function, known as *uncI* in *Escherichia coli*. The gene

orders in *Synechococcus* 6716 are related to the orders of ATP synthase genes in the plastid genomes of higher plants, and particularly of a red alga and a diatom. The sequences of the subunits are similar to those of chloroplast ATP synthase, the α , β and γ subunits being particularly well conserved. Differences in the primary structures of the *Synechococcus* 6716 and chloroplast γ subunits probably underlie different mechanisms of activation of ATP synthase. The nucleotide sequences that are presented also contain 12 other open reading frames. One of them encodes a protein sequence related to the *E. coli* DNA repair enzyme, photolyase, and another codes for a protein that contains internal repeats related to sequences in the myosin heavy chain.

INTRODUCTION

Oxidative and photosynthetic electron transport in bacteria, mitochondria and chloroplasts generate transmembrane electrochemical gradients of protons, which are used by the proton-translocating ATP synthase (H^+ -ATP synthase) to make ATP by phosphorylating ADP (Mitchell, 1966). The H^+ -ATP synthases consist of a hydrophilic catalytic domain (F_1), joined by a slender stem to a hydrophobic membrane sector (F_0), the site of a proton channel [see Senior (1988), Fillingame (1990) and Walker et al. (1990) for reviews]. F_1 is composed of five subunits, known as α , β , γ , δ and ϵ , assembled with a stoichiometry of 3:3:1:1:1 respectively. In *Escherichia coli* F_0 , the stoichiometry of its three constituent subunits, a , b and c , is 1:2:10–12 (Foster and Fillingame, 1982). In chloroplasts and photosynthetic bacteria, there are two related but different subunits, b and b' (in chloroplasts known as subunits I and II respectively), in addition to subunits a and c (subunits IV and III respectively in chloroplasts).

ATP synthase has been isolated from the thermophilic cyanobacterium *Synechococcus* 6716 (Lubberding et al., 1983) and has been reconstituted into proteoliposomes (Van Walraven et al., 1983). In many of its catalytic properties, the cyanobacterial enzyme resembles ATP synthase from spinach chloroplasts (Pick and Racker, 1979; Van Walraven, 1985), and ATP hydrolysis by F_1 -ATPase from both the cyanobacterium and chloroplasts is cryptic and can be activated by trypsin (Lubberding et al., 1981). However, there are also striking differences between the cyanobacterial enzyme and the enzyme from other sources. For example, in cyanobacterial ATP synthase (both isolated and membrane-bound) and in F_1 -ATPase, ATP hydrolysis cannot be activated by reduction with thiol agents, heating or exposure to

methanol (Lubberding et al., 1981; Van Walraven, 1985; Bakels et al., 1991), whereas these treatment methods activate chloroplast ATP synthase and CF_1 (see Strotman and Bickel-Sandkötter, 1984). Also, the *Synechococcus* 6716 ATP synthase can translocate considerably more than the usual 2–4 protons per ATP synthesized or hydrolysed (Van Walraven et al., 1986, 1990). In order to investigate the possibility that significant structural differences might underlie these differences in enzymic properties, as described below, we have determined the primary structures of the ATP synthase subunits from *Synechococcus* 6716 by sequencing the corresponding genes.

MATERIALS AND METHODS

Preparation of genomic DNA

Synechococcus strain PCC6716 (formerly known as *Synechococcus lividus*) was grown in semi-continuous anaerobic culture at 50 °C, its optimal growth temperature (Lubberding et al., 1981). DNA was isolated as described by Marmur (1961) from about 5 g of bacterial cells.

DNA hybridization

Fractionated digests of cyanobacterial DNA were transferred from agarose gels to nitrocellulose filters (Southern, 1975), and hybridized as described before (Cozens and Walker, 1987) with radiolabelled 'prime cut' probes (Farrell et al., 1983) containing sequences coding for the α , β , γ and a subunits of *Synechococcus* 6301 ATP synthase [see Cozens and Walker (1987) and the legend to Figure 1]. Autoradiographs were exposed with fluorescent screens for 1–72 h at –70 °C.

Abbreviation used: URF, unidentified reading frame.

§ To whom correspondence should be addressed.

† Present address: The Academic Medical Centre, University of Amsterdam, Department of Pulmonology, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands.

The sequences in this paper will appear in the EMBL database under the accession numbers K70431, K70432 and K70433.

Preparation and screening of a genomic DNA library

Restriction fragments about 15–20 kb in length, taken from a partial *Sau3A* digest of cyanobacterial DNA, were cloned into the *Bam*HI site of the vector λ 2001, as described before (Cozens and Walker, 1987), and packaged *in vitro* (Collins and Hohn, 1978). This library of *Synechococcus* 6716 DNA contained over 50 000 recombinants per ml. About 10 000 of these recombinants were screened with each of the probes described above. Fragments of interest, identified by Southern blotting of restriction digests of recombinant phage DNAs, were cloned into the plasmid vector pUC19.

DNA sequence analysis

A random sequencing strategy (Bankier and Barrell, 1983) was mainly followed. DNA fragments (300–500 bp in length), produced by sonication of larger fragments recovered from phage or plasmid DNA, were cloned into M13 mp18 (Yanisch-Perron et al., 1985). Sequences were determined nine times on average, and at least once on each DNA strand, by the modified chain termination method (Sanger et al., 1977; Biggin et al., 1983). The orientations of some inserts were reversed by transferring them from M13mp18 to M13mp19, and vice versa. 'Compressions' in GC-rich regions of DNA were resolved by replacement of dGTP by dTTP in the sequence reactions (Mills and Kramer, 1979). Some sequences were completed with the use of synthetic oligonucleotide primers 17 bases in length, made with an Applied Biosystems 380B oligonucleotide synthesizer.

Data were compiled with the aid of the computer programs DBAUTO and DBUTIL (Staden, 1982). Potential coding sequences in DNA sequences were detected by use of the positional base preference option in ANALYSEQ (Staden, 1985). Protein sequences were compared with sequences in the Protein Information Resource (PIR) database with FASTP (Lipman and Pearson, 1985).

Isolation and characterization of ATP synthase

The enzyme was isolated from about 5 g of cells of *Synechococcus* 6716 according to Lubberding et al. (1983), with the following modification. About 10 mg of the protein fraction precipitated with 35–50% (w/v) satd. $(\text{NH}_4)_2\text{SO}_4$ was redissolved in the column buffer, and the solution was passed through a column (80 cm $\times$ 1.8 cm internal diameter) of Sephacryl S-300 (Pharmacia, Uppsala, Sweden), packed in a buffer containing 10 mM Tricine/KOH (pH 7.8), 5 mM MgCl_2 , 500 mM NaCl, 4 mM octylthioglucoside, 1.6 mM sodium cholate, 1.5 mM dithiothreitol, 1 mM EGTA, 10% (v/v) glycerol and 10 mg/l phenylmethanesulphonyl fluoride. The ATPase-containing fractions were concentrated immediately by ultrafiltration.

ATP-hydrolytic activity was measured at 50 °C by either a linked enzyme assay (Stutterheim et al., 1980) or release of P_i (Fiske and SubbaRow, 1925). ATP synthase was activated by treatment with trypsin (25 $\mu\text{g}/\text{ml}$) for 10 min (Lubberding et al., 1981). Protein concentrations were determined as described by Bradford (1976). Polyacrylamide gradient gel electrophoresis (7.5–15% acrylamide) in the presence of SDS was performed in buffers as described by Laemmli (1970). Proteins were detected by either staining with Coomassie Blue dye or silver staining (Morrissey, 1981). Subunits were also transferred to poly(vinylidene difluoride) membranes (Matsudaira, 1987) and subjected to Edman degradation in a modified Applied Biosystems 470 gas phase sequencer.

RESULTS AND DISCUSSION

DNA sequences of cyanobacterial genes for subunits of ATP synthase

Restriction digests of DNA from *Synechococcus* 6716 were hybridized with each of the four probes for ATP synthase genes. Similar patterns of hybridization were obtained with the probes for the α and β subunits (Figure 1), and therefore their genes are close to each other in the genome. These experiments, and the subsequent library-screening and sequencing experiments described below, strongly suggest that there is a single F_1F_0 -ATPase in the cyanobacterium, which is likely to participate in ATP synthesis linked to both photosynthesis and oxidative phosphorylation.

Both probes hybridized with a 7.6 kb *Eco*RI fragment present in digests of genomic DNA (lanes e in Figure 1) and a recombinant phage isolated from a library of *Synechococcus* 6716 DNA. This fragment was shown by sequencing to encode six subunits of ATP synthase, including subunits α and β , as anticipated (Figure 2a). Another phage hybridized with both the γ and α probes, but all of the hybridizing restriction fragments

Figure 1 Hybridization of *Synechococcus* 6716 genomic DNA with probes for genes for various ATP synthase subunits

The probes (about 300–500 bp in length) were derived from the genes for subunits α , β , γ and α in *Synechococcus* 6301, starting at nucleotides 5750 (α), 2580 (β), 7050 (γ) and 2400 (α) in the sequences described by Cozens and Walker (1987). The DNA from *Synechococcus* 6716 was digested with *Sac*I (lanes b), *Pst*I (lanes c), *Hind*III (lanes d), *Eco*RI (lanes e) and *Bam*HI (lanes f). In lanes a, the DNA was not digested. Marker and fragment sizes are indicated at the sides in kb.

(a) GAATTCGCCAGGCCAGTCTCTGCTGCCACGTGCGAATACTCGCACGGGCTTCATCACTACGGGCTGCTCGAGGGCTGCTTCAGTGGCGCCAGACGGTGCGAATGCCGATCACGCC
 20 40 60 80 100 120
 CTTAAGGGCGGTTCGGTCAAGGACGACGGTGACGGCTTATGAGCGTGCCGAAGTAGTGATGCCCGACGAGCTCCCGACGAAGTGACCGCCGGGTCTGCCAGCTTACGGTATGTCGGG
 F E R W A L E Q Q W T R I S A R A E D S R A H E L A A E S A A W V T R I G I V G

AAATTTTAAAGCAGGGCTTAATAGGCGAGTGCCGGGTCGCCGGGGTAGTTCGGCTGCTCACCCTATTCGTAATGGCGTGCTGGCTGAAGTGCCGAGCTGTTGGGCTGCTGCTGCTTC
 140 160 180 200 220 240
 TTTAAATTCGGTCCGGAATATCCCGTACGGCCCGAGGGGCCCCATCAAGCGACGAGTGCCATAAGCATTTACCGACGACGAGCTTCACGGCTCGAACCCGACGACGACGAGG
 F K L A P S L L A T G P R G P Y N R Q E G Y E Y I A H Q S F H A L Q Q A A A E

TCCTGGAGCTAAATCAACTCCCCGAACAGTCAAAGCCTAAGTCATGGGTGCTGGGCAAGGGGATGCTGCTATTTCCGTTGGCTCTAGGGGGCAAAGCTTGGTGGCGTGGCAGGG
 260 280 300 320 340 360
 AGGACCTCGATTTCAGTGGGGGCTGGTCACTTCGATTAGTACCCAGACCGGTTCCCTGACGAGCGATAAAGCCACCGGAGATCCCCCGTTTCAGACACCGCCACCGTGGCC
 G P A L I L E G S W D F G L D H T S P L P I S A I E T A E L P A F D P P P P V P

AGCGGCTTTGGCAGGGCTGACCAATTCGCCCAAAGGGGTATAGACAGTGTAGGGCTGTCCCTGTTGGGTTTGACACAGCACTGGGGGGTGCAATAGTATGCCACTGGCGGTGAAC
 380 400 420 440 460 480
 TCGCGAAACCGCTCCGACTGGTTAAGCGGGTTTCCCCATATCTGTACATCCGACAGGGACAACCCAAACGTGTGCTGACCCCCACGTTATCGACTAGGGTGACCGCCACTTG
 A A K P L A S W N R W F P T Y V T Y P Q G Q Q T Q V A S P P H L L Q D W Q R H V

GGCGATGCCAAGGTCATTTAGGGTGGCGGCAACGACGCGATCGGCTCTCTGCGGTAGGGTTCACGCTCTCTGTCGACGTGTACCGCTGTGGCTTTAAGGTGGTGGCCACTCTGGGAT
 500 520 540 560 580 600
 CCGCTACGGTTCAGTAATCCACCGCGGTTCTGCTGCGCTAGCGCGAGAGACCGCATCCCAAGGTGAGAGACCGGTACATGGCGACACCGGAATTCACACCGGTGGAGACCTA
 A I G L D N L T A A V V R D R E R A Y P E V D E H W H V A T A K L T T A V E P I

CACCTGCGCGGATCCCCGTCACCATGAGCAGACGCTCCCGCTGCTGGTACGCTTCCTTGAGCGCTGTAGGCACCCAGCAAGTAGGCCACCCGACTGCGGCCACATCTTGCC
 620 640 660 680 700 720
 GTGAACGGCGCTAGGGGAGCTGTACTGCTGCTGCGAGGGGCAACGACATCGCAAGGAACCTGCGGACATCCGTTGGGTCTTCATCCGGTGGCGGTGACCGCGGTGAGAGGGG
 V Q R P D G Q V M L L C G G R Q Q Y A E K L A Q L C G L L Y A V R V A A V D Q G

TTGCAGCAGTGTGGGTCAAAACAGAAGACCCCAACCAACCGTTGGGGTGTGCTGCCGCTGCTGCTACAAACCTAAGTTATCGCAAGCGCAGGTGCGCGGTTGCCAAACAGATGTAG
 740 760 780 800 820 840
 AACGTCGTACGACCCAGTTTGTCTCTGGGGTGGTGGCAACCCACAGCGCGGCGACGATGTTTGGATTCAATAGCGGTTCCCGTCCAGCGCCCGCGGTTTGTCTACATC
 Q L L A P D F C F V G V V T P T H Q R A A Y L G L N D A L R L D R R H W F L H L

GGGTAACATCATCAAACTCCGGCAACCCCGTTAGCAAGGGGAGGATGTCAAGGCATCTTTAATCTTAAAGATAGAGTAATGCTTTATGGGTGGCCGCTTCAGCTCTGGCGT
 860 880 900 920 940 960
 CCCCATTGAGTGGTTTGGAGCGGTTGGGCAATCTGCTCCCTCTACAGTTCGCTAGAAATTTGTAATTTTCTATCTCCATACGCTCAATACCCACCGCGGGAAGTGGAGACCGCA
 P Y S M * F I S T I C N H T A A K L E P E
 ← URF 1

L V G V K L T K L L I G H G F L E A V N D A G A V A I A A I A K A G R S L R R K
 TCCTCGTTGGGGTCAAGCTGACGAACTGCTCATAGGGCATGGTTCTTGGAGCGGTAAATGACGCAAGTGTCAGTAGCGATCGCCGATCGCCAAAGCGGCGAGTGTGCGCGCA
 980 1000 1020 1040 1060 1080
 AGGAGCAACCCAGTTCGACTGCTTTGACGAGTATCCGTAACCAAGAACCTCCGCGCAATTAAGTCCGCTCCAGTGTGCTGCTAGCGGCGCTAGCGGTTTCGGCGTCCAGCAACCGGCGT
 E N P D L Q R F Q E Y P M T E Q L R N I V C T C Y R D G R D G F G A P R Q A A F

L R F T V S H K G S N R F W W V S I G C I C Y S G Q R P K A *
 AACCTAGGTTACAGTCAACCATAAAGCAGCAACCGCTTCTGGTGGGTAAATTTGGGTGTAATTTGCTATAGTGGGAGCGCCCAAGGCATAGGCTGCTGGCTCTAGCAAGGCAGCA
 1100 1120 1140 1160 1180 1200
 TTGAATCAAGTGTGCTGCTGTTTCCGCTGTTGGCGAAGCCACCAATTCATAACCCACATAAAGCATATCACCGCTGCGGGGTTCGGTATCCGACGACCGAGGATGCTTCCGCTGT
 K P E C D A M
 ← URF 3

AAACGTTTACCGCAGGTCTCTTAAAGGGTATTTCTAATTTTGCCTAAAGGACTCTCTGCTGCAAAATCAAATTTTATTTAGGCAATGGACAAGACAACCGCTTTGCCTATACTTAG
 1220 1240 1260 1280 1300 1320
 TTTGCAAGTGGCGTCCAGAGGATTTTCCCATAAAGATTAAACGGGTTTCTGAGAGAGCAGCTTTAAGTTTAAAAATAACTCCGTTACCTGTCTGTTTGGCGAAACGGATATGAATC

CTATACTAGTTGCTCAACCGATCTCTAGCGCAGATATCTGGGGAACACGAAAGCAGGATCCAGTCTTGGGACGTGCCCAAGGGTTCTGATGTTTITAGAGCTTACAGATATTTAG
 1340 1360 1380 1400 1420 1440
 GATATGAATCAACGAGTTGGGCTAAGGATCGGCTCTTATAGACCCCTTGTGCTTCTGCTAGGTGAGAACCTGCACGGGTTCCCAAGACTACAAAAATCTCGGATGTCTATAAATC

CACAGGCTTCTGTTTCTGCTTGTGATTTTCTTAAGGAGGCTCCCAATTTCTGAACACATTTCCCTAACCAAGCTTACGGGGAACCGGTGAGGTGCTGATGACTACCAAGTTT
 1460 1480 1500 1520 1540 1560
 GTGTCCGAAGAACAAAAGACAGAACTCAATAAACGATTTCTCCGAGGGTAAAGACTTGTGAAAGGGATTTGGTTTCGAATGCCCTTGGGCACTCCACGCACTACTGATGTTCAAAA

CGGCTGACGGGCTCATGCTGCTTCTCGGAGTCTGCTTCCGCAAAAGTCAITACCAAGTCTTTGACGATGGCCAGCAACATTATCTTGACCTCTCGAAATTTGAATTTCTTGA
 1580 1600 1620 1640 1660 1680
 GCGCGACTGCCAGAGTAGCTACGGAAGAGCTCAGACGGAAGCGGTTTCACTAATGTTTACGAACTGCTACCGGGTCTGTTGTAATAGGAAGTGGAGAGCTTTTAACTTAAAGAACT

TAGTTCAAGCTTGGGAGTCTTGTCTCACTAGCAAAAAGGCACAGGACACCAAGGGCACTCCAGATTGTCAACATCCAGGGTGAACCAACCGTAAAAATGGTACGGCTCGAAAA
 1700 1720 1740 1760 1780 1800
 ATCAAGTCCGAACCTCAAGAACAGGTGATCGGTTTTCGGTGTCTGCTGGTTCGGGTGAGGTCTAACAGTGGTTAGGTGCCCACTGGGTTTGGCAITTTTACCAATGCCGAGCTTTT

Figure 2(a). For legend see page 247.

Figure 2(a) (contd.). For legend see page 247.

Figure 2(a) (contd.). For legend see page 247.

V E F E D G T V G I A L N L E E D N V G V V L M G A G L G I E E G S T V R A T G
 TAGTGGAAATTTGAAGACGGTACGGTGGGTATCGCCCTCAACCTCGAAGAAGATTAACCTTGGGGTGGTCTGATGGGGCTGGCCTAGGCATTCAGGAAGGCAGTACGGTTCGTCCACCG
 5780 5800 5820 5840 5860 5880
 ATCACCTTAAACTTTCGCATGCCACCCATAGCGGGAGTTGGAGCTTCTTCTATTGCAACCCACACGACTACCCCGACCGGATCCGTAACCTCTTCGCTCATGCCAAGCAGCGTGGC

K I A S V P V G E A V I G R V V D A L M R P I D G K G E I H A T A T R L L E S P
 GTAAATTCGCTCGGTTCAGTGGGTGAGCGCTGATGGCCGTGGTGGATGCTCTGATGGCTCCCATGACGGCAAGCGGAGATTCACGCTACGCAACCCGCTTCGTAGAGTCCC
 5900 5920 5940 5960 5980 6000
 CATTTTAAGCGGCAAGCTACCCACTCCGGCAGTAACCGGCACACACCTACGAGACTACGACGGGTAACCTGCGCTTTCGCTCTAAGTGGGATGACGTTGGGCGAAGCATCTCAGG

A P G I V Q R K S V C E P L Q T G I T A I D A M I P I G R G Q R E L I I G D R Q
 CCGCCCCGGGATGTGACGCGCAAGTCGGTCTCGCAACCCCTGCAAACTGGCAITACGCCATTCATGCGCATGATTCCTCATGGTCTGGCCAGCGGGATTAATCATTCGGCATCGCC
 6020 6040 6060 6080 6100 6120
 GCGGGGGCCCTAACACGTCGGCTTCAGCGAGACGCTTGGGGACGTTTACCGTAATGGCGTAACTACGGTACTAAGGGTAGCCAGCACCGGTCCCTCAATTAGTAACCGCTAGCGG

T G K T A V A I D T I L N Q K G Q D V I C V Y V A I G Q K A S S V A Q V V N V L
 AAATCGGAAACAGCCGTGGCAATCGATACGATCTGAAACAAAGGTACGAGCGGTGATCTGTCTATGTCGCTATTGGCCAAAGGCTCCAGCGTTCGCAAGTGGTAATGTTT
 6140 6160 6180 6200 6220 6240
 TTTGACCCCTTTTGTGGCGCGCTTAGCTATGCTAGGACTTGGTTTTCAGTCTGCACTAGACACAGATACAGCGATAACCGGTTTTCGGAGGTTCGCAACGGGTTTACCCTTACAGG

R E R G A L D Y T I V V A A N A S D P A A L Q Y L A P Y T G A S I A E Y F M Y Q
 TCGGGAGCGGGTGCCTTCGATACACCATTTGTTGTCGACGCAACCGCAGTACCTGCTGCTTCGAGTACTTAGCTCCCTACACCGGTGCTAGCATTCGTGAGTACTTATGATATC
 6260 6280 6300 6320 6340 6360
 ACGCCCTTGCACCGGACTGATGTGGTAACACACGCTCGGTTCGGTCACTGGGACGACGAGCTCATGAATCGAGGATGTGGCCACGATCGTAACGACTCATGAAATACATAG

G K H T L V I Y D D L S K Q A Q A Y R Q M S L L L R R P P G R E A Y P G D V F Y
 AAGCAAGCATACCTTAGTGATCTACGATGATCTCTGAAACAGCGCAGGCTTACCGGCAATGTCATTGTTACTGCGTCTGTCGCCCGGTGCGGAAGCCTATCCGGGGATGTGTTCT
 6380 6400 6420 6440 6460 6480
 TTCGTTGATGGGATCTAGATGCTACTAGAGAGCTTGTTCGCGTCCGAATGCGGCTTACAGTAACATGACGACGAGGGGGCCAGCGCTTCGGATAGCCCGCTACACAGA

L H S R L L E R A A K L N D A L G G G S M T A L P I V E T O A G D V S A Y I P T
 ACCCTCACTCTGCTTATTAGAGCGCGCCGCCAACTCAATAGCGCCCTAGGGGTGGCAGTATGACCGCACTACCAATTTGGAAGAACCAAGCGGGGACGCTGCTGCTATATATCCCA
 6500 6520 6540 6560 6580 6600
 TGGAGTGAGAGCAGATAATCTCGCGCGCGTTTGAAGTACTGCGGATCCCCACCGTCACTACTGGCGTATGTTAATACCTTTGGGTCGCGCGCTGCACAGACGGATATAGGGGT

N V I S I T D G Q I F L S S D L F N A G L R P A I N A G I S V S R V G S A A Q I
 CCAACGTGATTTCCATTACCGATGGCCAGATTTCTTGTGCTGACCTGTTTAAATCGGGTTTTCGACCTGGGATTAATCGGGTATTTCTGTGCTCGGGTGGTTCGGCAGCAGAA
 6620 6640 6660 6680 6700 6720
 GGTGCACTAAGGTAAATGGCTACCGGCTTAAAGAACAGCAGACTGGACAAATTCGCGCAACCGTGGACGCTAATTACGCCCATAAAGACAGAGCCACCAAGCGCTGCTGTTT

K A M K Q V A G K L K L E L A Q F D E L Q A F A Q F A S D L D K A T O N Q L A R
 TTAGGCAATGAAGCAGGTGGCTGGCAACTCAAACTAGAACTTGGCGAGTTTACCAACTCCAGGCTTTGCGCAATTTGCTCTGACTAGATAAAGCAACCGAGATCAACTGGCGC
 6740 6760 6780 6800 6820 6840
 AATTCCGTTACTTCGTCACCGACCGTTTGAAGTTGATCTTGAACCGCTCAAACTGTTGAGGTTCGGAACCGCTTAAACGGAGACTGAATCTATTTGTTGGGCTTATGTTGACCGG

G Q R L R E I L K Q P O Y S P I P V E Y Q V A T I Y A G T N G Y L D D I P V E A
 GGGTCAGCGGTACGGGAATCTCAAGCAGCCCGTACTCGCCCATTCGGTGGAGTACCAAGTTGCGACCATCTATGACGACCAATGGCTATTTGGACGATATTCCTGTTGAAG
 6860 6880 6900 6920 6940 6960
 CCCCAGTCCCAATGCGCTTTAGAGTTCTGCGGGTATGAGCGGGTAAGGCCACCTCATGGTTCAACGCTGGTAGATACGTCCTGGTTACCGATAAACCTGCTATAAGGGCAACTTC

V A K F V A G L R D Y L A T S K P Q Y G E A V R S S Q K L D E T A E A L L K E A
 CCGTGGCTAAGTTTGTGCTGGCTCGGGATTATTTGGCAACAGTAACACAGTACGCGTGAAGCGGTACGAGTAGCCAAAACCTGACGAAACCGCAGAAGCTCTGCTCAAGAGG
 6980 7000 7020 7040 7060 7080
 GGCACCGATTCAAAACAGCAGCGGCCCTAATAAACCGTTGTTTATTTGGTGTGATGCCACTTCGCCATGCGTCACTCGGTTTTTGAAGTCTTTGAGAGAGGTTTCTCC

I A E Y K A G F T A *
 CGATCGCGGAATATAAGCAGGCTTTACTGCTAGGTTCTCTCCACAATTAGCTGAGGGAGTGGGGGGTAGATACGCCCTGCATGTAGCCGTCGAGAGGCGAGTACGGGATACAGA
 7100 7120 7140 7160 7180 7200
 GCTAGCGGCTTATATTTCTGCTCCGAAATGACGGATCCAGGAGGTGTTAATCGACTCCTCAGCCCGCATCTATGCGGGGACGTGACATCGGCAGCTCTCCGTGATGATGCTATGCT

TACCCCGTGCCAAATTCGACAGTAGCAGCAAGACGACACCGCACGCCAACTAGCGCGGTTTACTCAACCGGGGTGGCTTTCTTGTAGTGGCCAGCTCCTGTAGGGTCACTTGGC
 7220 7240 7260 7280 7300 7320
 ATGGGGGACCGGTTAAGTCTGATGCTGCTGCTTGTCTGTGGGGTGGGTTGATCCGCGCAATGAGTTGTGCCCCCACCAGAAAGAACATCACCAGTTCAGGACATCCAGTGAACGG

CTTAACACTGCGGCGGTCTACTGCGGGGACGTCGCGAAGACCCCTTGGGGGGGATAGCAGGGTTTGGAGATCACTACGCTAGCAATGGCAATTAAGAAGTAACCAATGCTTGT
 7340 7360 7380 7400 7420 7440
 GAATTTGACGCGCGCACGATGACGCGCCGTCAGGGCTTCTTGGGAATCCCGCTATCTGCTCCCAAACTCTAGTGTATGCGATCGTTACCGTAAATCTTTCATTGGTTTACGAACA

TGAAGTTTTTGGGTGAGACTCGTGTCCCGAGTAGCTTCTAGCAAGGAGCCAGGGGCTCGGCAAGGGTTAATGTGACGCTAGCGGTGGATCGAGTGTACAGTTCCTAAGTAGCCC
 7460 7480 7500 7520 7540 7560
 ACTTCAAAAACCCACTCTGAGCGACAGGGTCAATGAGGATCGTCTCGGTCCCGACGCGCTTCCCAATTACAGTGCAGTCCGACCTAGCTCACAATGTCAAGGATTCATCGGG

CATACCCAGCCACAGAATT
 7580
 GTATGGGTGCGTCTTAAG

Figure 2(a) (contd.). For legend see page 247.

Figure 2(b). For legend see page 247.

(c) CTCGAGTTCGCGCTACTACCCCGCTCCCAATGATGCGCGATCTAGTTCGCGCGGTTCTAGGCGAGTGGTCACTGTTGGCGATCTTGGCTTAGAAATGCCACTGATGA
 20 40 60 80 100 120
 GAGCTCAAGAGCGGATGATGGGCGGAAGGTGTACTACCGGCTAGATCAAGCCGCCCAAGATCCCGTACACAGTTCGACACCGCTAAGAACCGAAATCTTTACGGGTGACTACT
 E L E R A V V G A E V I I R R D L E A A T E L A T T L Q Q R N K A K S I G V S S

GCCAAAGTTGGCGGCTTGACAAAGCAGGGGTAGCCCAATTCGGTTTCGATGCGATGCGACACCTGACGATAGCGGCGAGGATCCGACCAAGTTCTGAGCGGTGATGCGACATACTG
 140 160 180 200 220 240
 CGGTTCCAAACCGCGAAGTCTTTCGTCCTCCATCGGTTAAGCCAAAGCTACGCTAGCGTGTGAGCTGCTATCGCGTCTAGGCTGGTTTCAAGACTCGCGACTACGGGTGATGAC
 G L N A P K V F C P Y G L E T E I R D C V Q R Y R C P D S W L E S R S I A V Y Q

CACCTGCGGTAGCCACCGCTCCAAAAACGTTTTCATTAAAAATTTTATCCATGCCAATGCGGAGGCGCCACCCAGACCCACGTTAGGAGCTGATCATCTTCCACAGTCTCTG
 260 280 300 320 340 360
 GTGAACGCCATCGGGTGGCGAGGTTTTCGCAAAAGTAAATTTTAAATAGGTACGGTTACCGCTCCGGCGGTGGGTCGGGTGATCCCTGCAACGTAGTAAGGTGTGACGAAAC
 V Q P L G V A G F F T K M L I K D M G I A S A A V G S G V Y P R Q M M E L L G Q

AATTGTGCCATCTTACCCTGTGGGCGGTGGAGAATGGGGAACCAACATCAATGGTTTCTGCGCACTGCCGAAATGACACACGAGTATGATTATAGAGAAGTGCTGCGCACTGAC
 380 400 420 440 460 480
 TTAACACGTAGAAAGTGGCAACCCCGCACCTCTTACCCCTTGGTTTGTAGTTACCAAGACGCTGACGCGCTTACTGCTGCTAATAAGTCTCTTCCACGACCGGTGACTG
 I T G D E G N P G H L I P F W V D I T E A V A P F S W L S Q N D L S P A P A S V

GGTTTCGTTACACGCCAACGACCATCTTTTGAATATAAATGGGACCCCTCATAGCGATGCCATTGAGAGGATCGGCAAGGCTTGGCAATTGCCCGCCGACATACACCGAGAC
 500 520 540 560 580 600
 CCAAGGCCATGTGGCGGTGTGCTAGAAAACTTATATTTACCGGTGGAGTATGCTAGCGGTAACTCTCTAGCGGTTTCCGAACCCGTTAAGCGGCGGCTATGTGGCTCTG
 T E T C R W R G D K Q I Y I P V V E Y R D G N L P D A F A Q A I A A A S V V S V

CTCGTCTCCCGGAGCGACCGCCAAATAGTAACCCACCGAGACGCCATAATGTAACCTGTGTTAACCGGACCGTTTCTTGGTACGATTTTCAAGGATTATACAACTTTGCG
 620 640 660 680 700 720
 GAGCAGGAGGCGCTGCTGCGGTTTATCATTGGGTTGGCGTCTGCTGGTATTACATTTTGGCACAATTGCGCTGGCAAGAACCATGTAAAGTTCCTAATATGTTTGAACCCCT
 E H E R S R G G F L L G V R L R G M

URF 9 → ← URF 8

M D V G Q K V R V C R I R D R V A Q E V L K K L G Q V G Q V T G F K M T D G
 740 760 780 800 820 840
 CAACGATACGTGACCCCGTTTTCACGCGCAGACGGCTAAGCGTAGCGCAGCGGTTCTTCAAGGTTTTCAGCCAGTTCAGTCCCGGAAATCTACTGACTGCCG

S G V G V V V T F D D R S S T W F F E D E I E A V G *
 860 880 900 920 940 960
 TCACCTCAGCCCCACCCAGTGGAACTACTAGCAGGTGATGTAACAAAACTTCTGCTTTAGCTTCGGCAACCGATCGGGGACACCGGGTAATCACCATGCAACATACACAC

CTTTTGGGCGACAGCGTCACTTCCATTAGGTTTGAATCGTCCGCTCTCATGTGTAACCTTTTCGCTACAGATGATGGCTGAAGCGATCGCCAGTTACAAAAATGAACATTACAA
 980 1000 1020 1040 1060 1080
 GAAACGCCCGTGTGCAAGAGTAAATCCAACTAGCAGCGAGAGTACACACTTGGGAAGCGAGTGTCTACTACCGGACTTCGCTAGCGGTTCAATGTTTTACTTGAATGTT
 * T Q D D A R M H S G K R E C I I A Q L S R W N C F I F K G L

β Subunit →

G F I T Q V I G P V I D I E F P S G K M P A I Y N
 1100 1120 1140 1160 1180 1200
 CTGTATCCGCTCAGGTACAGTGTGTGCTTCTGTTTACATCCGAATATGGGTTCAATGCCCGGTGAGTAACTGTAACTTAAAGGTCACCGTTTACGGGCGGTAAATAT
 C Y A S D M T V V A S R V L T P K I V M T I P G T M S M S N G L P L I G A M

← URF 10

A L R I Q G K N A A G L D V A V T C E V Q Q L L G D N R V R A V A M S S T D G L
 1220 1240 1260 1280 1300 1320
 ATGCCCTCCGATCAAGGAAACGCGCGGATTTGACGTTGCGCTCACTGCGGAGTGGCAAGTGTCTTGGGATATCGTGTGCGTGGCGTTCGCAATGAGCAGCACCGATGGC
 TACGGGCGTAAAGTTCCTTTTCGCTGCGCTAACCTGCAACGGCAGTGACGCTTCAGCTGTTGAGGAACCGCTATTAGCAGGCAACCGGTAAGTCTGCTGCTGCTACCG

V R G M E A V D T G A P I S V P V G T A T L G R I F N V L G E P V D E K G E V N
 1340 1360 1380 1400 1420 1440
 TTGTACGAGCATGGAAGCACTGATACCGGTGCCCCATTAGCGTTCGCTTGGCAGCGCACCTTGGCGCATCTTAAATGTGTGGGGGAACCCGCTGATGAAAGAGCGCGGCA
 AACATGCTCCGTACCTTGTGACGATAGGCCACGGGGTAATCGCAAGGCAACCGTGGCGGTGGGAACCGGCTAGAAATACACAACCCCTTGGGCGAGTACTTTTCCCTCCAGT

I S E T L P I H R P A P S F T E L E T K P S V F E T G I K V I D L L T P Y R R G
 1460 1480 1500 1520 1540 1560
 ACATCAGTGAACCTTCCGATTCACCGCTTCCCTAGCTTTACAGAACTGGAAACGAAGCCCTCGGCTTTGAAACCGGATCAAGGTGATGATCTGCTGACTCCCTATCCCGGTG
 TGTAGTCACTTTGGAACCGCTAAGTGGCAGGACGGGATCGAAATGCTTGAACCTTGTCTGGGAGCCAGAACTTTGGCGTAGTTCCACTAGCTAGACGACTGAGGATAGCGGCAC

G K I G L F G G A G V G K T V I M M E L I N N I A T Q H G G V S V F A G V G E R
 1580 1600 1620 1640 1660 1680
 GCGTAAAAATTGGCTGTTTGGCGGTGCGGGTGTGGCAAACTGTATCATGATGGAACCTGATCAACAACATTGCCACCAACACGCGCGGTTTCAGTGTTCGGGGGTGGGAGGAC
 CGCCATTTTAAACCGGACAAACGCCACGCCACACCGCTTTGACAGTAGTACTACCTTGAAGTGTGTTGTAACCGTGGGTTGTGCGCGCGCAAGTACAAACGCCGCCACCGCTCG

T R E G N D L Y N E M I E S G V I D K D D P S K S K I A L V Y G Q M N E P P G A
 1700 1720 1740 1760 1780 1800
 GCACCCGCGAAGGCAACGACCTTACAAAGAAATGATTGAATCGGAGTCAATTGATAAAGACGATCCGAGCAATCGAAATTCGCTTGGTCTATGCGGAGTGAACGAGCCGCCCGGG
 CTGCGGCGCTTCCGTTGCTGAGATGTTGCTTTACTAATCTAGCCCTCAGTAATATTTCTGCTAGGCTGTTTGTGTTTAAACGGAACAGATACCGGTCTACTTGTGCGGCGGCGCC

R M R V G L S G L T M A E Y F R D V N K Q D V L L F V D N I F R F V Q A G S E V
 1820 1840 1860 1880 1900 1920
 CACGATGCGGTTGGGTTGACCATGCGGGAATCTTCCGATGTGAACAGCAGATGTGCTGCTTTGTTGATAATTTTCCGCTTGTCCAGCTGAGGCTCTATTACCTCCATCC
 GTGCCATCGCCACCGAATAGACCACTGGTACCGCTTATGAAGGCGCTACACTTGTGCTCTACACGACGAGAAACATATTATAAAGGGAACAGGTTTCGACCGAGCTTC

S A L L G R M P S A V G Y Q P T L G T D V G A L Q E R I T S T M E G S I T S I O
 1940 1960 1980 2000 2020 2040
 TGTACGCTCTTGGTGGATGCCCTCGGCGAGTACGCTACCTAGGCAACGATGTGGGGGCGCTCCAGGAGCGGATACCTCGACCATGGAAGGCTCTATTACCTCCATCC
 ACAGTCGGGAGAACCGAGCTACCGGAGCGCTACCCCTAGTCCGATGGGATCCGCTGCTACACCCCGCGAGGCTCTCGCTAATGGAGCTGCTACCTTCCGAGATATGAGGAGTAGG

Figure 2(c). For legend see opposite.

Figure 2 Nucleotide sequences of three fragments of *Synechococcus* 6716 DNA encoding the subunits of ATP synthase

(a) Both strands of the DNA of a 7.6 kb *EcoRI* fragment containing the genes for subunits α , c , b , b' , δ and α ; (b) the non-coding strand of a 2.8 kb *EcoRI* fragment encoding the γ subunit; (c) both strands of the DNA of a 3.2 kb fragment encoding the β and ϵ subunits. The boxed sequences are potential ribosome-binding sites. Amino acid sequences determined by N-terminal sequencing of ATP synthase subunits α , β , δ and ϵ are shaded. URF, unidentified reading frame.

were larger than 10 kb, and so the two genes appeared to be at least 10 kb apart. The γ gene was localized by sequencing a 2.8 kb *EcoRI* fragment from a phage that hybridized with the γ probe only (Figure 2b). Other hybridization experiments indicated that the gene for the β subunit was distant from those for subunits γ , α and a . Part of its sequence was identified by sequencing a 2.8 kb *XhoI*–*Sau3A* fragment (nucleotides 1–2807 in Figure 2c) from a positively hybridizing phage, and this sequence was extended to a total length of 3.2 kb by sequencing an overlapping 4.0 kb *BamHI* fragment (nucleotides 2677–3234 in Figure 2c) from another recombinant.

The G+C contents of the 7.6, 2.8 and 3.2 kb segments of DNA sequence (see Figure 2) were calculated to be 53.3%, 55% and 54.1% respectively, in agreement with the value reported by Herdman et al. (1979), and lower than the values in sequences determined in *Synechococcus* 6301 (Cozens and Walker, 1987).

Identification of genes

By comparison of the protein sequences encoded in the three DNA fragments with the sequences of the subunits of ATP synthase from *E. coli*, it was found that the 3.2 kb fragment

contained the genes for β and ϵ , that the gene for γ was in the 2.8 kb fragment, and that the 7.6 kb fragment encoded ATP synthase subunits, α , b , c , δ and α . The coding sequence of each gene starts with the initiation codon ATG, except for the gene for the γ subunit, where GTG is used. Each of these codons is preceded by sequences resembling *E. coli* ribosomal binding sites (see Figure 2), which are complementary to sequences near to the 3' end of 16S rRNA in *Synechococcus* 6301 (Cozens and Walker, 1987).

The identities of the genes for the α , β , δ and ϵ subunits were confirmed by N-terminal protein sequence analysis of the subunits of the partially purified enzyme (see Figures 2 and 3). These analyses showed that the initiator formyl-methionine is not present in either the mature α , δ or ϵ subunits, and that it is present but deformed in the β subunit. No N-terminal protein sequences were determined by analysis of the other subunits. In the case of the γ subunit, this is at least partly because the protein transfers poorly to poly(vinylidene difluoride) membranes, and in some cases the subunits may be modified. Therefore at present these latter subunits have not been identified with certainty in the polyacrylamide gel (Figure 3), although the values for molecular mass calculated from the sequences and estimated from their gel

Figure 3 Subunits of ATP synthase from *Synechococcus* 6716

The subunits were separated by SDS/PAGE. (a) Three identical samples of a gel stained with Coomassie Blue dye. (b) The same preparation of the cyanobacterial enzyme (right-hand lane) after silver staining compared with ATP synthase from bovine heart mitochondria (Walker et al., 1991), in which the positions of the F_1 subunits are indicated. Subunits α , β , δ and ϵ of the former preparation were identified by N-terminal protein sequence analysis (see the Materials and methods section and Figure 2).

Table 1 Molecular masses of subunits of ATP synthase in *Synechococcus* 6716

Subunit	Molecular mass (kDa)	
	From sequence	From gels
α	53.904*	54
β	51.762†	51
γ	34.911‡	Not known
δ	20.479*	23
ϵ	14.752*	17
a	27.998‡	Not known
b	19.260‡	Not known
b'	17.920‡	Not known
c	8.180‡	Not known

* Assumes removal of initiator formyl-methionine.

† Assumes removal of N-terminal formyl group.

‡ Assumes the presence of N-terminal formyl group and no post-translational processing, although this may take place in subunits b and b' (see the text).

positions are concordant (Table 1). Usually nine different subunits can be detected by Coomassie Blue staining of preparations of *Synechococcus* 6716 ATP synthase fractionated by PAGE, but sometimes a tenth band can be detected below the β subunit, as in the sample in Figure 3. Its N-terminal sequence is not related to that of any known subunit of an ATP synthase, and it is considered to be a contaminant. It is clear from the silver-stained gel that other contaminants are also present in the cyanobacterial preparation.

Between the c and δ genes were found genes for subunits b and b', homologues of *E. coli* subunit b. Similar findings have been made in *Synechococcus* 6301 (Cozens and Walker, 1987), *Anabaena* 7120 (McCarn et al., 1988), *Rhodospirillum rubrum* (Falk and Walker, 1988) and *Synechocystis* 6803 (Lill and Nelson, 1991). As expected, the hydrophobic profiles (not shown) of the b and b' subunits from *Synechococcus* 6716 are related, and indicate that their N-terminal regions consist of a stretch of about 20 hydrophobic amino acids. In *Synechococcus* 6301, it

was proposed that the equivalent sequences were removed by post-translational processing to produce mature proteins similar in length and hydrophobic profiles to the homologous subunit I in chloroplast ATP synthase (Cozens and Walker, 1987). No complete sequence of a chloroplast subunit II is yet available, but the N-terminal sequence of this subunit from spinach chloroplasts (Berzborn et al., 1990) is related to subunit b' of *Synechococcus* ATP synthase, and the N-terminus of the chloroplast sequence corresponds to residue 12 of the cyanobacterial sequence. The N-termini of the mature thermophilic b and b' subunits are not known. They may be processed in a similar way to the chloroplast subunit I, but the functions of the processed regions remain obscure.

A second stretch of 25–30 hydrophobic amino acids following this N-terminal stretch could form a transmembrane α -helix in the mature b and b' subunits, thereby anchoring their otherwise hydrophilic sequences to the bacterial membrane, as proposed (Walker et al., 1982) and established (Dunn, 1992) in the *E. coli* b subunit. Although the overall homology between different b subunits is low (see Table 2), their structures and functions seem to be similar, and it has been reported that the *E. coli* b subunit can be replaced in the ATP synthase complex by chloroplast subunit I, without impairing the enzyme's function (Schmidt et al., 1990).

It was first observed in the *E. coli* *unc* operon (Gay and Walker, 1981), and subsequently in other related bacterial operons, that a potential gene encoding a hydrophobic protein of unknown function lies between the transcriptional promoter and the gene for subunit a. The *E. coli* gene is known as *uncI*, and the protein has been shown to be expressed, although it does not form part of the ATP synthase complex (Schnepp et al., 1990). A potential gene is found in the equivalent location in *Synechococcus* 6716, and the hydrophobic profile of the encoded protein is similar to those of *uncI* products in other species. Consensus sequences typical of many bacterial promoters were not evident in the sequences to the 5' side of the *uncI* homologue, and the position of none of the transcriptional promoters of the *Synechococcus* 6716 ATP synthase genes has been determined experimentally. It appears that the sequence immediately upstream of *uncI* in *Synechococcus* 6716 could contain another

Table 2 Conservation of sequences of subunits of ATP synthase in *Synechococcus* 6716 and other species

Subunit	Sequence identity (%) with homologues from:				
	<i>Synechococcus</i> 6301 ^a	Chloroplast ^{b,c}	<i>E. coli</i> ^d	PS3 ^e	<i>Rsp. rubrum</i> ^f
α	81	76	57	66	62
β	83	80	65	67	68
γ	78	55	37	34	33
δ	52	24	19	22	28
ϵ	66	43	26	28	26
a	73	60	17	13	16
b	41	26	20	14	7
b'	46	—	—	—	17
c	89	88	31	40	27

References: ^aCozens and Walker (1987); ^bHowe et al. (1982, 1985) for subunits α , β , ϵ and c from wheat; Bird et al. (1985) for subunit b from wheat; Cozens et al. (1986) for subunit a from pea; Hermans et al. (1988) for subunit δ from spinach; Miki et al. (1988) for subunit γ from spinach; ^cthe only available sequence information about a chloroplast subunit b' (subunit II) is the N-terminal sequence from spinach (Berzborn et al., 1990); this is not included in the Table. ^dWalker et al. 1984; ^eOhta et al. 1988; ^fFalk et al. (1985); Falk and Walker (1988).

Figure 4 Arrangement of genes in three fragments of DNA sequence from *Synechococcus* 6716

The three segments correspond to parts (a)–(c) in Figure 2. The genes encoding nine subunits of ATP synthase are marked according to the subunit (a, c, b', b, δ , α , γ , β and ϵ). The positions of gene 1 (*E. coli* *uncI* equivalent) and 12 other potential genes, URFs 1–12 (U1–U12), are also shown. The proteins encoded by URFs 1 and 6 respectively are related to *E. coli* photolyase and myosin heavy chain. The scale is in kb.

potential gene (URF 4 in Figures 2a and 4), and therefore in this respect the gene organization may differ from those found in other cyanobacteria.

URF 4 is one of 12 other potential genes (URFs 1–12 in Figure 2) that were detected in the three segments of DNA sequence. URF 1, an incomplete reading frame in the 7.6 kb fragment, codes for a partial protein sequence of 284 amino acids, which is identical in 28 % of its amino acids with the *E. coli* DNA repair enzyme, deoxyribodipyrimidine photolyase (Sancar et al., 1984). URF 6 in the 2.8 kb fragment encodes a protein which contains many short internal repeats that are related to short repeats in the tail region of myosin heavy chains. None of the remaining URFs is related to known protein sequences, including proteins that are encoded in URFs associated with ATP synthase genes in *Synechococcus* 6301 (Cozens and Walker, 1987) and *Rsp. rubrum* (Falk et al., 1985; Falk and Walker, 1988).

The arrangement of genes in the three segments of DNA sequence are summarized in Figure 4.

Organization of genes for ATP synthase subunits

The organization of genes for ATP synthase in *Synechococcus* 6716 is unique among the bacteria that have been investigated. It is most similar to the arrangement common to three other cyanobacteria, *Synechococcus* 6301, *Anabaena* 7120 and *Synechocystis* 6803 (Cozens and Walker, 1987; Curtis, 1987; McCarn et al., 1988; Lill and Nelson, 1991; see Figure 5), and

differs from these arrangements only in that the gene for the γ subunit is at a separate locus and does not follow the α gene. Attention has been drawn previously to the similarity between the cyanobacterial gene orders and the orders of genes for ATP synthase subunits in the chloroplast genomes of higher plants. Even greater similarity is found between the cyanobacterial gene orders, particularly those from *Synechococcus* 6716, and the gene orders found in the plastid genomes of the red alga, *Antithamnion* (Kostrzewa and Zetsche, 1992), and of the diatom, *Odontella sinensis* (Pancic et al., 1992). These similarities in gene organization are clearly of significance in providing support for endosymbiotic origins of the plastids from precursors of the present-day cyanobacteria.

Genes for a bacterial ATP synthase were first characterized in *E. coli*, where they are organized at a single locus, the *atp* (*unc*) operon. It was remarked that the gene order in the operon reflects the structure of the complex; genes for the F_0 subunits form a promoter proximal cluster a: c: b, and are followed by a second cluster of genes for F_1 subunits in the order δ : α : γ : β : ϵ (Walker et al., 1984). The same arrangement has been found subsequently for genes of H^+ -ATP synthases in bacterium PS3 (Ohta et al., 1988), *Bacillus megaterium* (Brusilow et al., 1989) and *Bacillus firmus* (Ivey and Krulwich, 1991), and for a Na^+ -motive ATP synthase in *Vibrio alginolyticus* (Krumholz et al., 1989). In the purple non-sulphur bacterium, *Rsp. rubrum*, the F_0 cluster a: c: b: b' and the F_1 cluster δ : α : γ : β : ϵ are at two separately transcribed loci (Falk and Walker, 1985, 1988; Falk et

Figure 5 Organization of genes encoding ATP synthase subunits in various bacterial and plastid genomes

The scale is in kb, and the dashed lines indicate that the gene clusters are 10 kb apart or more. The ATP synthase subunits encoded in the genes are indicated by a, c, b', b, δ , α , γ , β , δ and ϵ . In plants, the genes are called *atpI*, H, G, F, D, A, C, B and E respectively, and subunits a, c, b and b' are more usually known as subunits IV, III, I and II. The remaining chloroplast subunits (γ , δ and ϵ) are encoded in the plant nucleus, and are imported into the organelle (Hennig and Herrmann, 1986). In *Antithamnion* and *O. sinensis*, the locations have not been determined of the genes for subunits γ , β and ϵ , and for subunit γ respectively. The function of the protein encoded by the gene marked I is unknown. The *E. coli* gene order (Walker et al., 1984) is also found in bacterium PS3 (Ohta et al., 1988), *B. megaterium* (Brusilow et al., 1989), *B. firmus* (Ivey and Krulwich, 1991) and *V. alginolyticus* (Krumholz et al., 1989). The *Synechococcus* 6301 gene order (Cozens and Walker, 1987) is found in *Anabaena* 7120 (Curtis, 1987; McCarn et al., 1988) and in *Synechocystis* 6803 (Lill and Nelson, 1991).

al., 1985). It is possible that these clusters represent protein modules that have evolved independently before coming together in the ATP synthase (Walker and Cozens, 1986; Walker, 1992). In the cyanobacteria, further rearrangements of genes may then have occurred.

Activation of cyanobacterial ATP synthase

Chloroplast ATP synthase is inactive in the dark and its physiological activation by light is brought about by the reduction by thioredoxin of a disulphide bond between γ -subunit cysteines 204 and 210 [spinach numbering from sequence alignment in Werner et al. (1990) used in Figure 6]. A similar activation can be produced *in vitro* by reduced thiols (Ketcham et al., 1984; Nalin and McCarty, 1984; Schumann et al., 1985). Upon reduction, the enzyme shifts to a conformational state that needs less energy for activation (Junesch and Gräber, 1987). Although the sequences of cyanobacterial and chloroplast γ subunits are well conserved overall (see Table 2), alignment of their sequences (Figure 6) reveals that residues 202 to 210 are present only in the chloroplast sequence, and that the cysteines required for formation of the disulphide bond in the chloroplast enzyme are absent from the cyanobacterial subunit. Therefore the chloroplast and cyanobacterial γ subunits differ in a region that is important in the regulation of the activity of the chloroplast enzyme. This same region is also involved in the trypsin activation of the CF₁. The unmasking of the latent ATPase activity of spinach CF₁ by trypsin is brought about by the sequential cleavage of the peptide bonds following Arg-220 and Lys-209 in the γ subunit [tryptic cleavage sites I and II in Figure 6 (Werner et al., 1990)]. In the cyanobacterial γ subunit, tryptic cleavage site I, but not site II, is maintained, and this may be sufficient to retain the trypsin activation (Lubberding et al., 1981; Bakels et al., 1991). The ATP synthase of *Synechococcus* 6716 corresponds to the reduced form of CF₁F₀ and needs a low proton gradient for activation (Bakels et al., 1991). This could be of physiological importance in oxidative phosphorylation in the cyanobacterium.

The proton-pumping mechanism

ATP synthase isolated from *Synechococcus* 6716 and reconstituted into artificial membranes of native lipids can translocate approximately 9 H⁺ per ATP synthesized or hydrolysed (Van Walraven et al., 1986). Since this number is so

	191	201	211	221
			↓	↓
Spinach chloroplast	IHTLLPLSPK	GEICDINGKC	VDAAEDELFR	LITKEGKLTIV
<i>Synechococcus</i> 6716	VQTLPLDPQ	G	LEAADEIFR	LITRASHLEV
<i>Synechococcus</i> 6301	VQTLPLDPQ	G	LAASDEIFR	LITRGGSFIV
<i>Synechocystis</i> 6803	VQTLPLSPQ	G	LEAPDEIFR	LITRGGFQV
<i>Anabaena</i> 7120	IQTLLPLDTQ	G	LEAADEIFR	LITRGGQFQV
<i>E. coli</i>	ISQLLPLP	—	ASDDDLK	HKSND
<i>Rsp. rubrum</i>	RQTLIPFVP	T	TVAGNDNR	TAG
<i>Rps. blastica</i>	AQVPAK	—	FEAEINAL	—
Bacterium PS3	ERKLPLT	—	DLAENKQRT	—

Figure 6 Alignment of a region of the γ subunit of chloroplast ATP synthase involved in the activation of the enzyme with related sequences in bacterial γ subunits

The sites of sequential cleavage in trypsin activation are indicated by arrows I and II. The numbers relate to the complete sequences of the polypeptides. The sequences are taken from the following sources: spinach chloroplasts (Miki et al., 1988); *Synechococcus* 6301 (Cozens and Walker, 1987); *Synechocystis* 6803 (Werner et al., 1990); *Anabaena* 7120 (McCarn et al., 1988); *E. coli* (Walker et al., 1984); *Rsp. rubrum* (Falk et al., 1985); *Rhodospseudomonas blastica* (Tybulewicz et al., 1984); and bacterium PS3 (Ohta et al., 1988). The amino acid sequences are numbered as described by Werner et al. (1990).

much higher than the generally observed 2–4 protons translocated for each ATP molecule synthesized (see Nicholls and Ferguson, 1992), *a priori* it was possible that the cyanobacterial enzyme might have significant changes in its membrane sector subunits. The sequences indicate that the cyanobacterial membrane sector is likely to contain four subunits, a, b, b' and c, as in the chloroplast enzyme. Also, the a and c subunits, which are intimately involved in proton translocation, are highly conserved in the enzymes from chloroplasts and *Synechococcus* 6716 (see Table 2). They include amino acids that are thought to be essential for the proton channel such as the equivalent residues of *E. coli* Glu-219 and Arg-210 in the a subunit of *E. coli* ATP synthase, and Asp-61 from subunit c (Lightowers et al., 1988). Recently, it was found that the proton translocation stoichiometry for both cyanobacterial and chloroplast ATP synthase depends on the lipid composition of proteoliposomes in which they are reconstituted, and is also changed when other proteins are present (Van Walraven et al., 1990). In addition, in the intact bacteria it is influenced by the growth temperature and light intensity (H. S. Van Walraven, A. P. N. de Boer, M. J. C. Scholts and R. H. A. Bakels, unpublished work) and by pH in membrane vesicles (Krenn et al., 1993). These effects are discussed in terms of the sequences of the F₀ subunits by Krenn et al. (1993).

Note added in proof (received 21 June 1993)

The sequence of subunit II of CF₀ from spinach chloroplasts is described by Herrmann et al. (1993).

H. S. V. W. was supported by a long-term EMBO fellowship. We thank I. M. Fearnley, F. D. Northrop and J. M. Skehel for their help with the protein sequencing experiments, M. J. C. Scholts for assistance with protein purification, and M. J. Runswick for help in the preparation of Figures.

REFERENCES

- Bakels, R. H. A., Van Walraven, H. S., Scholts, M. J. C., Krab, K. and Kraayenhof, R. (1991) *Biochem. Biophys. Acta* **1058**, 225–234
- Bankier, A. T. and Barrell, B. G. (1983) in *Techniques in Nucleic Acid Biochemistry* (Flavell, R. A., ed.), pp. 1–34. Elsevier, County Clare, Ireland and New York
- Berzborn, R. J., Klein-Hitpass, L., Otto, J., Schünemann, S., Oworah-Nkruma, R. and Meyer, H. E. (1990) *Z. Naturforsch.* **45c**, 772–784
- Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965
- Bird, C. R., Koller, B., Auffret, A. D., Huttley, A. K., Howe, C. J., Dyer, T. A. and Gray, J. C. (1985) *EMBO J.* **4**, 1381–1388
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254
- Brusilow, W. S. A., Scarpetta, M. A., Hawthorne, C. A. and Clark, W. P. (1989) *J. Biol. Chem.* **264**, 1528–1533
- Collins, J. and Hohn, B. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4242–4256
- Cozens, A. L. and Walker, J. E. (1987) *J. Mol. Biol.* **194**, 359–383
- Cozens, A. L., Walker, J. E., Phillips, A. L., Huttly, A. K. and Gray, J. C. (1986) *EMBO J.* **5**, 217–222
- Curtis, S. E. (1987) *J. Bacteriol.* **169**, 80–86
- Dunn, S. D. (1992) *J. Biol. Chem.* **267**, 7630–7636
- Falk, G. and Walker, J. E. (1985) *Biochem. J.* **229**, 663–668
- Falk, G. and Walker, J. E. (1988) *Biochem. J.* **254**, 109–122
- Falk, G., Hampe, A. and Walker, J. E. (1985) *Biochem. J.* **228**, 391–407
- Farrell, P. J., Deininger, P. L., Bankier, A. and Barrell, B. G. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1565–1569
- Fillingame, R. H. (1990) *The Bacteria* **12**, 345–391
- Fiske, C. H. and Subbarow, Y. (1925) *J. Biol. Chem.* **66**, 375–400
- Foster, D. L. and Fillingame, R. H. (1982) *J. Biol. Chem.* **257**, 2009–2015
- Gay, N. J. and Walker, J. E. (1981) *Nucleic Acids Res.* **9**, 3919–3926
- Hennig, J. and Herrmann, R. G. (1986) *Mol. Gen. Genet.* **203**, 117–128
- Herdman, M., Janvier, M., Waterbury, J. B., Rippka, R. and Stanier, R. Y. (1979) *J. Gen. Microbiol.* **111**, 63–71
- Hermans, J., Rother, C., Bichler, J., Steppuhn, J. and Herrmann, R. G. (1988) *Plant Mol. Biol.* **10**, 323–330
- Herrmann, R. G., Steppuhn, J., Herrmann, G. S. and Nelson, N. (1993) *FEBS Lett.*, in the press

- Howe, C. J., Auffret, A. D., Doherty, A., Bowman, C. M., Dyer, T. A. and Gray, J. C. (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6903–6907
- Howe, C. J., Fearnley, I. M., Walker, J. E., Dyer, T. A. and Gray, J. C. (1985) *Plant Mol. Biol.* **4**, 333–345
- Ivey, D. M. and Krulwich, T. A. (1991) *Mol. Gen. Genet.* **229**, 292–300
- Junesch, U. and Gräber, P. (1987) *Biochim. Biophys. Acta* **893**, 275–288
- Ketcham, S. R., Davenport, J. W., Warncke, K. and McCarty, R. E. (1984) *J. Biol. Chem.* **259**, 7286–7293
- Kostrzewa, M. and Zetsche, K. (1992) *J. Mol. Biol.* **227**, 961–970
- Krenn, B. E., Van Walraven, H. S., Scholts, M. J. C. and Kraayenhof, R. (1993) *Biochem. J.*, in the press
- Krumholz, L. R., Esser, U. and Simoni, R. D. (1989) *Nucleic Acids Res.* **17**, 7993
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lightowlers, R. N., Howitt, S. M., Hatch, L., Gibson, F. and Cox, G. (1988) *Biochim. Biophys. Acta* **933**, 241–248
- Lill, H. and Nelson, N. (1991) *Plant Mol. Biol.* **17**, 641–652
- Lipman, D. J. and Pearson, W. R. (1985) *Science* **227**, 1435–1441
- Lubberding, H. J., Offerijns, F., Vel, W. A. C. and De Vries, P. J. R. (1981) in *Photosynthesis II. Electron Transport and Photophosphorylation* (Akoyunoglou, G., ed.), pp. 779–788, Balaban International Science Services, Philadelphia
- Lubberding, H. J., Zimmer, G., van Walraven, H. S., Schrickx, J. and Kraayenhof, R. (1983) *Eur. J. Biochem.* **137**, 95–99
- Marmur, J. (1961) *J. Mol. Biol.* **3**, 208–218
- Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035–10038
- McCarn, D. F., Whitaker, R. A., Alam, J., Vrba, J. M. and Curtis, S. E. (1988) *J. Bacteriol.* **170**, 3448–3458
- Miki, J., Maeda, M., Mukohata, Y. and Futai, M. (1988) *FEBS Lett.* **232**, 221–226
- Mills, D. R. and Kramer, F. R. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 2232–2235
- Mitchell, P. (1966) *Biol. Rev.* **41**, 455–502
- Morrissey, J. H. (1981) *Anal. Biochem.* **117**, 307–310
- Nalin, C. M. and McCarty, R. E. (1984) *J. Biol. Chem.* **259**, 7275–7280
- Nicholls, D. G. and Ferguson, S. J. (1992) in *Bioenergetics 2*, Academic Press, New York and London
- Ohta, S., Yohda, M., Ishizuka, M., Hirata, H., Hamamoto, T., Otawara-Hamamoto, Y., Matsuda, K. and Kagawa, Y. (1988) *Biochim. Biophys. Acta* **933**, 141–155
- Pancic, P. G., Strotmann, H. and Kowallik, K. V. (1992) *J. Mol. Biol.* **224**, 529–536
- Pick, U. and Racker, E. (1979) *J. Biol. Chem.* **254**, 2793–2799
- Sancar, G. G., Smith, F. W., Lorence, M. C., Rupert, C. S. and Sancar, A. (1984) *J. Biol. Chem.* **259**, 6033–6038
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Schmidt, G., Rodgers, A. J. W., Howitt, S. M., Munn, A. L., Hudson, G. S., Holten, T. A., Whitfield, P. R., Bottomley, W., Gibson, F. and Cox, G. B. (1990) *Biochim. Biophys. Acta* **1015**, 195–199
- Schneppe, B., Deckers-Heberstreit, G. and Altendorf, K. (1990) *J. Biol. Chem.* **265**, 389–395
- Schumann, J., Richter, M. L. and McCarty, R. E. (1985) *J. Biol. Chem.* **260**, 11817–11823
- Senior, A. E. (1988) *Physiol. Rev.* **68**, 177–231
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517
- Staden, R. (1982) *Nucleic Acids Res.* **10**, 4731–4751
- Staden, R. (1985) in *Genetic Engineering: Principles and Methods* (Setlow, J. K. and Hollaender, A., eds), pp. 67–114, Plenum Publishing Corporation, New York and London
- Strotmann, H. and Bickel-Sandkötter, S. (1984) *Annu. Rev. Plant Physiol.* **35**, 97–120
- Stutterheim, E., Henneke, M. A. C. and Berden, J. A. (1980) *Biochim. Biophys. Acta* **252**, 415–430
- Tybulewicz, V. L. J., Falk, G. and Walker, J. E. (1984) *J. Mol. Biol.* **179**, 185–214
- Van Walraven, H. S. (1985) Ph.D. Thesis, Vrije University, Amsterdam
- Van Walraven, H. S., Lubberding, H. J., Marvin, H. J. P. and Kraayenhof, R. (1983) *Eur. J. Biochem.* **137**, 101–106
- Van Walraven, H. S., Haak, N. P., Krab, K. and Kraayenhof, R. (1986) *FEBS Lett.* **208**, 138–142
- Van Walraven, H. S., Scholts, M. J. C., Koppenaal, F., Bakels, R. H. A. and Krab, K. (1990) *Biochim. Biophys. Acta* **1015**, 425–434
- Walker, J. E. (1992) *Q. Rev. Biophys.* **25**, 253–324
- Walker, J. E. and Cozens, A. L. (1986) *Chem. Scripta* **26B**, 263–272
- Walker, J. E., Saraste, M. and Gay, N. J. (1982) *Nature (London)* **298**, 867–869
- Walker, J. E., Saraste, M. and Gay, N. J. (1984) *Biochim. Biophys. Acta* **768**, 164–200
- Walker, J. E., Fearnley, I. M., Lutter, R., Todd, R. J. and Runswick, M. J. (1990) *Philos. Trans. R. Soc. London Ser. B* **326**, 367–378
- Walker, J. E., Lutter, R., Dupuis, A. and Runswick, M. J. (1991) *Biochemistry* **30**, 5369–5378
- Werner, S., Schumann, J. and Strotmann, H. (1990) *FEBS Lett.* **261**, 204–208
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* **33**, 103–119