Nucleic Acids Research

The atp operon: nucleotide sequence of the region encoding the α-subunit of Escherichia coli ATP-synthase

N.J.Gay and J.E.Walker

The Laboratory of Molecular Biology, The MRC Centre, Hills Road, Cambridge CB2 2QH, UK

ABSTRACT

Part of the atp (or unc) operon encoding the α, β, γ, δ and ε subunits of Escherichia coli ATP-synthase has been cloned into the plasmid pACYC 184. The DNA coding for the largest of these proteins, the α subunit, has been sequenced by cloning into the bacteriophage M13 and sequencing with dideoxy nucleotide chain terminators. It comprises 1539 nucleotides corresponding to a protein of 513 amino acids.

INTRODUCTION

The structure of the membrane bound ATP synthase complex of Escherichia coli closely resembles that of mitochondria and chloroplasts [1,2]. It comprises an intrinsic membrane fraction (F_0) and an extrinsic portion, F_1, which can be solubilised intact. F_1 contains five different polypeptides designated α, β, γ, δ and ε for which a stoichiometry of 3:3:1:1:1 has been proposed [3], although this is not universally accepted [4]. The catalytic site of synthesis of ATP from ADP is found within F_1 [1,2]. Bacterial F_0 contains three polypeptides [5] and forms a proton channel coupling the vectorial movement of protons by energy transducing membranes to ATP synthesis [1]. The entire bacterial complex is encoded by the atp [6] (or unc [2]) operon located at about 83 minutes on the E. coli chromosomal map [7] close to the origin of replication, oriC. Genetic analysis suggests that the Fo polypeptides are clustered at the promoter proximal portion of the operon and are followed by the genes encoding F_1 subunits [2,6,8], but an unambiguous order of genes has not yet been established [6,8,9]. UncA and uncD genes have been assigned to the α and β subunits respectively [2]. In the present work we have precisely located the uncA gene and determined its coding sequence.

MATERIALS AND METHODS

Cloning and preparation of DNA

DNA from bacteriophage λAsn5 containing about 26 kilobases of the
E. coli chromosome including asn, oriC, atp (or unc), glmS and bglC,R was
prepared as described elsewhere [10]. We thank C. Lichtenstein and S. Brenner
for providing us with samples of phage DNA and E. coli KY7485 [9]. An EcoRI
digest of this DNA was ligated into the EcoRI site of plasmid pACYC184
located in the CamR gene [11] and transformed into E. coli-K12 HB101 [12].
Clones containing recombinant plasmids were identified by their antibiotic
resistance markers (Tc resistant, Cam sensitive) and then screened by the
alkaline SDS procedure [13] followed by EcoRI digestion and agarose gel
electrophoresis. Recombinant plasmids pN5R1 and pN5R2 containing R1 and R2
respectively were obtained by this procedure. Plasmid pN5R1 was then
amplified with chloroamphenicol and prepared by a scaled up version of this
procedure from a 2 ℓ bacterial culture. Fragment R1 was excised from plasmid
and purified by preparative agarose gel electrophoresis in a 1% gel according
to Sanger et al. [14].

Ligation into M13

The whole fragment, R1, was cloned into the EcoRI site and digests with
SalI and AccI similarly cloned into their corresponding sites of M13mo7 by
published methods [14-16]. TaqI and HpaII digests were cloned into the AccI
site and Sau3A digests into the BamI site of the same vector. Digests with
AluI, HaeIII and HpaI were blunt-end ligated into the HincII site of M13mp7,
pretreated with calf intestinal phosphatase (Worthington) as described else-
where [17].

Transformation and nucleotide sequencing

Transformation of E. coli JM101[Δ(lacpro)supE, Thi, F'tra D36 pro AB lac
1^qΔM15] [18] and purification of recombinant phages was carried out according
to [17]. Clones were sequenced by the dideoxy chain termination method [14,
19] using a synthetic primer 17 nucleotides in length complementary to a
region of M13mp7 immediately adjacent to the linker sequence [20]. Nucleotide
sequences were compiled and analysed with the aid of computer programs
described by Staden [21-23].

RESULTS

Cloning of genes for the F1 polypeptides

The genes for the five F1 polypeptides have been mapped previously to a
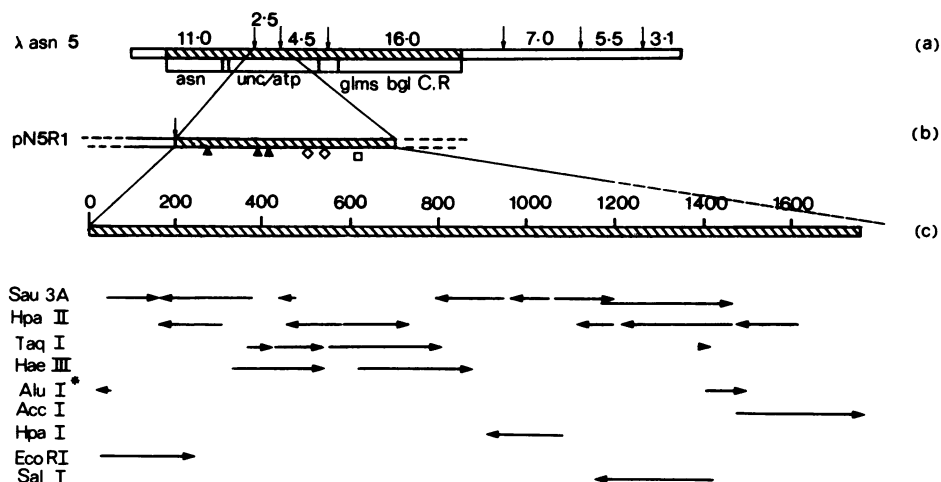region of the E. coli chromosome shown in detail in Figure 1(a) and appear to

Figure 1. The genetic and restriction map in the region of the atp operon.
(a) The extent of E. coli chromosomal DNA in transducing phage λAsn5 (hatched).
EcoRI restriction sites are shown by arrows, EcoRI fragment sizes are given in
kilobases [2,6]. (b) Plasmid pN5R1 containing the 2.5 kilobase EcoRI fragment,
R1 (hatched). Restriction sites are denoted by ▲ HpaI, ◊ SalI and □ HindIII.
(c) The portion of fragment R1, with nucleotide numbers, sequenced in the
present work. The arrows indicate the 5' ends of the various clones sequenced
by the M13 shotgun-dideoxy chain termination strategy [15]; the length of
arrows shows the length of the sequence determined. * The AluI clone containing
nucleotides 1-73 contained both ends of fragment R1 in which the EcoRI
ends had been religated.

be contained within two EcoRI fragments (R1 and R2) of approximately 2.5 and
4.5 kilobases respectively [8,9]. Accordingly these two fragments were cloned
into plasmid from an EcoRI digest of λAsn5 DNA (Figure 1b). Their identities
were confirmed by comparison of their mobilities on agarose gels with an
EcoRI digest of λAsn5 and by restriction digestion analysis with the enzyme
PstI in the case of R2 (A. Eberle, unpublished work).

Nucleotide sequence

Fragment R1 was cloned into the EcoRI site of M13mp7 and isolated in
both orientations. Then the sequence for about 200 bases was determined at
each end of the clone. This allowed the orientation of the clone to be
determined since sequence corresponding to the NH_2-terminal sequence of the
α-subunit could be identified [14]. This showed that an EcoRI site was
located between nucleotides 9 and 10 of the coding sequence and hence that
all of the coding sequence for α except for the first nine nucleotides has

been determined (see legend to Figure 2).   The sequence of R1 was
extended almost to completion by random sequencing of restriction digests by
the M13 cloning chain termination procedure [19]. A search for restriction
sites in the partial sequence revealed HpaI sites at nucleotides 910 and 1102
(as numbered in the final sequence). The fragment between these two sites
was cloned from an HpaI digest and on sequence analysis gave the final over-
lap. A summary of the clones used to establish the sequence is given in
Figure 1(c). Most regions of the sequence (95%) are covered by two or more
independent clones and 51% of the sequence has been determined on both
strands. The nucleotide sequence of the coding region of the ATP-synthase
α-subunit and the deduced protein sequence are shown in Figure 2.

DISCUSSION
Comparison with the α-protein
     The nucleotide sequence described here agrees exactly with the indepen-
dently determined sequence of E. coli ATP-synthetase α-subunit from amino
acid residues 4-30 and with the C-terminal residue tryptophan determined by
hydrazinolysis of the protein [24]. The molecular weight of the protein
calculated from the sequence is 55,423 daltons, compared with values of 54-60
x 10^3 daltons determined by polyacrylamide gel electrophoresis under
denaturing conditions [2]. The amino acid composition of the α-subunit
derived from DNA sequence is compared in Table 1 with that determined by
analysis of hydrolysates of the α-subunit [24]. The two sets of values are
in good agreement, the only major discrepancy being between the values for
the number of alanine residues.
     Further support for the protein sequence described here has been
obtained by amino acid sequence analysis of peptides isolated from a cyanogen
bromide digest of the α-subunit of bovine mitochondrial ATP-synthase (N.J.G.,
M.J. Runswick, M. Saraste and J.E.W., unpublished results). These sequences
are homologous with regions of the E. coli α-subunit sequence as depicted in
Figure 2.
Codon usage
     It has already been observed that codon usage in other E. coli genes is
non-random with striking preferences for certain codons [5]. The codon usage
in the sequence described here (Table 2) resembles especially the codon
usages described for prolipoprotein, the most abundant protein in the E. coli
cell [26], and certain ribosomal proteins [27,18]. In these cases codon
usage reflects tRNA abundance: codons corresponding to abundant iso-accepting

Figure 2. The nucleotide sequence of the coding region of the α-subunit of E. coli ATP-synthetase. Nucleotides 1-8 are derived from the protein sequence [14]. This allows two ambiguities at the underlined position 6 (A or G) and 7 (T or C). Nucleotide 9 was deduced from the specificity of EcoRI. The remainder was sequenced experimentally as described. Boxed regions of the protein sequence correspond to the published NH$_2$-terminal sequences [24] and to homologous sequences of cyanogen bromide peptides of the bovine mitochondrial subunit (unpublished work). The carboxyl terminal residue tryptophan was determined by hydrazinolysis [24]

Table 1

Amino acid composition of the α-subunit of E. coli ATP-synthase

Amino acid	From DNA sequence	Recalculated from [24]
Asparagine	16 ⎤	43.4
Aspartic acid	28 ⎦	
Threonine	23	22.7
Serine	32	33.6
Glutamine	29 ⎤	60.7
Glutamic acid	31 ⎦	
Proline	18	19.3
Glycine	47	50.0
Alanine	54	62.1
Cysteine	3	3.6
Valine	41	44.4
Methionine	12	9.6
Isoleucine	40	37.6
Leucine	48	47.7
Tyrosine	15	14.7
Phenylalanine	14	13.7
Histidine	7	6.9
Lysine	24	23.6
Arginine	30	28.8
Tryptophan	1	1.0
TOTAL	513	-

tRNAs are preferred (for a summary of tRNA abundance in E. coli see supple-
ment to ref. 28). This may be because of the need of the cell for either
efficient or high fidelity translation of essential proteins [28]. Codon
usages in genes of abundant proteins differ strikingly from those observed in
genes expressed at low levels where rarer codons are used more frequently.
(For a quantitative assessmentof the role of codon usage in expression of
E. coli genes see ref. 25.)

    An interesting feature of the atp operon is that the proteins of the
ATPase complex are not present in stoichiometric amounts with (possibly)
three copies of α and β to each copy of γ, δ and ε [3]. Codon usage may
provide a mechanism for modulation of expression of the different cistrons in

Table 2

Codon usage in uncA. The numbers exclude the first three
codons which have not been unambiguously assigned.

	U			C			A			G		
U	UUU	Phe	1	UCU	Ser	11	UAU	Tyr	4	UGU	Cys	3
	UUC	Phe	13	UCC	Ser	16	UAC	Tyr	11	UGC	Cys	0
	UUA	Leu	0	UCA	Ser	0	UAA	Ochre	1	UGA	Opal	0
	UUG	Leu	2	UCG	Ser	1	UAG	Amber	0	UGG	Trp	1
C	CUU	Leu	3	CCU	Pro	1	CAU	His	0	CGU	Arg	25
	CUC	Leu	4	CCC	Pro	0	CAC	His	7	CGC	Arg	5
	CUA	Leu	1	CCA	Pro	3	CAA	Gln	4	CGA	Arg	1
	CUG	Leu	38	CCG	Pro	14	CAG	Gln	24	CGG	Arg	0
A	AUU	Ile	10	ACU	Thr	4	AAU	Asn	2	AGU	Ser	2
	AUC	Ile	30	ACC	Thr	17	AAC	Asn	14	AGC	Ser	2
	AUA	Ile	0	ACA	Thr	2	AAA	Lys	19	AGA	Arg	0
	AUG	Met	11	ACG	Thr	0	AAG	Lys	5	AGG	Arg	0
G	GUU	Val	24	GCU	Ala	16	GAU	Asp	12	GGU	Gly	28
	GUC	Val	2	GCC	Ala	8	GAC	Asp	18	GGC	Gly	16
	GUA	Val	9	GCA	Ala	18	GAA	Glu	27	GGA	Gly	1
	GUG	Val	6	GCG	Ala	12	GAG	Glu	4	GGG	Gly	1

the atp operon. Analysis of other genes in this operon should test this
idea.

REFERENCES

1. Futai, M. and Kanazawa, H. (1980) Curr. Topics Bioenerg. 10, 181-215.
2. Downie, J.A., Gibson, F. and Cox, G.B. (1979) Ann. Rev. Biochem. 48,
   103-131.
3. Bragg, P.D. and Hou, C. (1975) Arch. Biochem. Biophys. 167, 311-321.
4. Vogel, G. and Steinhardt, R. (1976) Biochemistry 15, 208-216.

5. Foster, D.L., Mosher, M.E., Futai, M. and Fillingame, R.H. (1980) J. Biol. Chem. 255, 12037-12041.
6. von Meyenberg, K. and Hansen, F.G. (1980) In: "Mechanistic Studies of DNA Replication and Genetic Recombination" ICH-UCLA Symp. Vol. XIX, Alberts, B. and Fox, C.F., Eds. Academic Press, New York, in the press.
7. Bachman, B.J. and Low, B.K. (1980) Microbiol. Rev. 44, 1-56.
8. Downie, J.A., Langmans, L., Cox, G.B., Yanofsky, C. and Gibson, F. (1980) J. Bact. 143, 8-17.
9. Kanazawa, H., Tamura, F., Mabuchi, K., Miki, T. and Futai, M. (1980) Proc. Nat. Acad. Sci. USA 77, 7005-7009.
10. Lichtenstein, C. (1980) Ph.D. thesis, Cambridge.
11. Chang, A.C.Y. and Cohen, S.N. (1978) J. Bact. 134, 1141-1156.
12. Boyer, H.W. and Roulland-Dussiox, D. (1969) J. Mol. Biol. 41, 459-472.
13. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1529.
14. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
15. Winter, G. and Fields, S. (1980) Nucleic Acids Res. 8, 1965-1974.
16. Messing, J., Gronenborn, B., Müller-Hill, B. and Hofschneider, P.H. (1977) Proc. Nat. Acad. Sci. USA 74, 3642-3646.
17. Winter, G., Fields, S. and Gait, M.J. (1981) Nucleic Acids Res. 9, 237-245.
18. Messing, J. (1979) Recombinant DNA Technical Bulletin 2, 43-48.
19. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA 74, 5463-5467.
20. Duckworth, M.L., Gait, M.J., Goelet, P., Hong, G.F., Singh, M. and Titmas, R.C. (1981) Nucleic Acids Res., submitted for publication.
21. Staden, R. (1977) Nucleic Acids Res. 4, 4037-4051.
22. Staden, R. (1978) Nucleic Acdis Res. 5, 1013-1015.
23. Staden, R. (1979) Nucleic Acids Res. 6, 2601-2610.
24. Dunn, S.D. (1980) J. Biol. Chem. 255, 11857-11860.
25. Grantham, R., Gautier, C., Gony, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. 9, r43-r74.
26. Nakumura, K., Pirtle, R.M., Pirtle, I.L., Takeishi, K. and Inouye, M. (1980) J. Biol. Chem. 255, 210-216.
27. Post, L.E., Strycharz, G.D., Nomura, M., Lewis, H. and Dennis, P.P. (1979) Proc. Nat. Acad. Sci. USA 76, 16970-1701.
28. Post, L.E. and Nomura, M. (1980) J. Biol. Chem. 255, 4660-4666.