Volume 9 Number 16 1981 Nucleic Acids Research

The atp operon: nucleotide sequence of the promoter and the genes for the membrane proteins, and the δ subunit of Escherichia coli ATP-synthase

Nicholas J.Gay and John E.Walker*

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

Received 23 June 1981

ABSTRACT

The nucleotide sequence of the promoter region and the first five genes of the atp (or unc) operon of Escherichia coli has been determined. The first proposed gene in the operon contains four AUA codons and may be poorly expressed; it encodes a basic but yet hydrophobic protein which could function as a pilot protein for assembly of ATP-synthase. The three genes that follow are structural genes for proteins comprising the proton channel of the enzyme. The fifth gene codes for the δ-subunit of F_1-ATPase.

INTRODUCTION

The atp [1] or unc [2] operon contains structural genes for the ATP-synthase complex of Escherichia coli. The bacterial enzyme is thought to be composed of eight distinct proteins [2,3]. Three of them, called a, b and c, are in the cytoplasmic membrane and comprise a proton translocating channel (F_o) which couples the proton potential of the membrane to ATP synthesis from ADP and P_i [3]. The remaining five proteins, called α, β, γ, δ and ε, comprise an extra-membrane assembly called F_1-ATPase. This is intimately associated with F_o [2,3] and protrudes into the cytoplasm; it contains the catalytic sites of the enzyme [1,2]. Recently, we have cloned and sequenced a region of the E. coli chromosome encompassing the genes for the α, β, γ and ε proteins, thereby establishing a gene order α: γ: β: ε [4 and unpublished results]. The cloning and sequence analysis of the remaining genes in the operon and the promoter are described here. This sequence of 2,785 nucleotides has been determined by cloning into bacteriophage M13 coupled with sequencing by the dideoxy chain termination method [5]. The sequence analysis has been speeded up by the use of a non-random strategy.

MATERIALS AND METHODS

Cloning and preparation of DNA

The preparation of Asn5 DNA was as described earlier [4].

M13mp3.NH4. A HindIII digest of λAsn5 DNA (50 µg) was fractionated on a 1% low melting agarose gel. A band of about 4 kb was recovered [5] and ligated to HindIII-cleaved M13mp3 replicative form (RF). After transfection into E. coli JM101 clones containing inserts were picked and single-stranded DNA prepared [4,5]. T-Track sequence analysis revealed two sequences, one adjacent to the HindIII site in fragment R1 (Fig. 1b, 1c) [1,2,6]. Single-stranded DNA from the appropriate plaque was transfected into E. coli JM101 and a plaque transferred to a 50 ml early log phase culture of the same organism. It was grown to OD_{590} 1.0 and RF-DNA prepared by the alkaline SDS method [6] yielding 300 µg of pure DNA. A fragment of 2.3 kb (H3R1) was prepared from an EcoRI digest of this DNA by gel electrophoresis [5]. It was cloned into M13mp7 and used to extend the sequence leftwards of fragment R1 (Fig. 1c). Cloning into M13 of subdigests of H3R1 with the HaeIII. HincII, PvuII and HpaII were carried out as described [4]. A subdigest of H3R1 with RsaI was ligated into the HincII site of M13mp7 [7].

mp7.NB4. A 4 kb fragment isolated by agarose gel electrophoresis of a BamHI digest of λAsn5 DNA was cloned into M13mp7 and RF prepared as above.

mp3.NH1.5. The 4 kb BamHI fragment was excised from mp7.NB4 with EcoRI and the HindIII digest of it cloned into HindIII-cleaved M13mp3. Clones of both orientations of a 199 bp HindIII fragment were identified from T-tracks. A clone of the adjacent 1.5 kb HindIII fragment was identified by reference to an extended sequence of mp7.NB4. The experiments are summarised in Figures 1 and 2.

Nucleotide sequencing

Procedures concerning sequence analysis and computer analysis of data have been described [4].

RESULTS

Cloning of genes in the atp operon

Our earlier studies have shown that the atp (unc) operon extends leftwards of fragment R1 (Fig. 1c) towards the origin of replication, oriC. Additionally, genetic studies had indicated that genes for membrane proteins of ATP-synthase would lie in this region [1,2]. So the appropriate primary fragments were cloned into bacteriophage M13 as shown in Figure 1c.

Nucleotide sequence

The sequence of the region was determined rapidly from these primary clones and subclones derived from them in a series of experiments summarised chronologically in Figure 1d. The non-random strategy employed makes use of

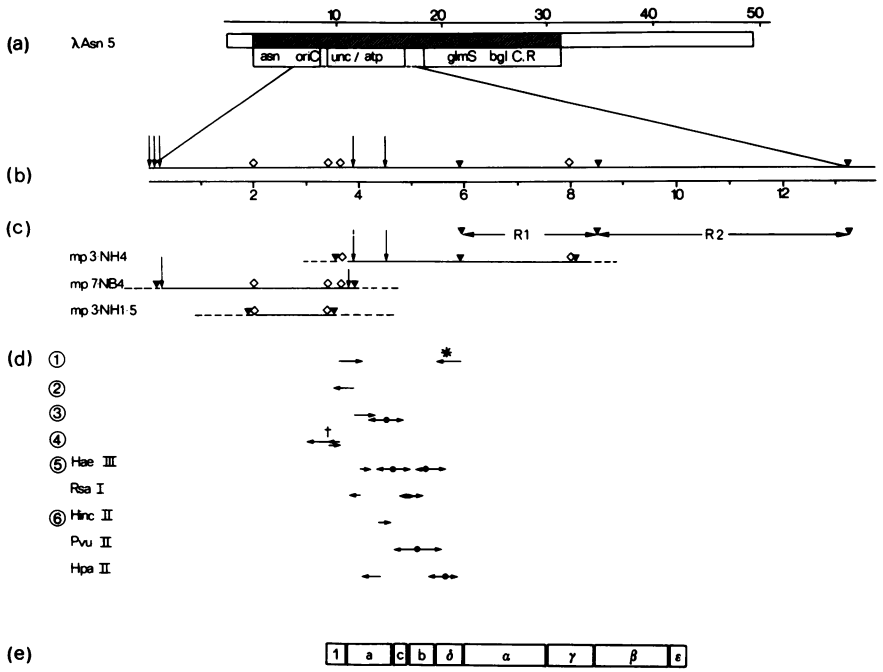

Figure 1. Clones used in establishing the DNA sequence of the promoter proximal region of the atp operon. They are aligned with genetic and restriction maps [1-3]. (a) The extent of the E. coli chromosome (hatched) in λAsn5 showing genetic markers [1-3]. The scale is in kilobases. (b) A restriction map with nucleotide numbers (kilobases) of part of region in (a) redrawn from [3]. Arrows denote restriction sites for BamHI, ▼ EcoRI and ◇ HindIII. (c) Alignment of EcoRI fragments R1 and R2 cloned previously [4] and the primary fragments used for sequence analysis in clones mp3.NH4, mp7.NB4 and mp3.NH1.5. (d) Chronology of sequencing experiments. Arrows show the ends of sequenced subclones. Lengths of arrows are proportional to lengths of sequences obtained from each subclone. ① 2.3 kb fragment H3R1 cloned into M13mp7. * This clone was recovered from a blue plaque (β-galactosidase-positive). ② Sequenced directly from M13mp7.NB4. ③ H3R1 was digested with BamHI. The digest cloned into M13mp7 (BamHI-cleaved). From this, these three clones were isolated and sequenced. ④ The 4.0 kb fragment, purified from an EcoRI digest of M13mp7.NB4, was digested with HindIII and cloned into M13mp3 (Methods). The sequences were determined from isolates from this cloning. † Overlap between the HindIII clones provided by extended sequencing of the clone in 2 . ⑤ Following computer analysis of data from ① to ④ HaeIII and RsaI digests were cloned and sequences of appropriate isolates (identified from T-tracks) established. ⑥ Further computer analysis of data from ① to ⑤ indicated that the sequence of the 2.3 kb fragment could be completed by appropriate PvuII, HincII and HpaII clones. These were identified as in ⑤. (e) Alignment of the nine genes of the operon with restriction map [this work, 4, and unpublished work].

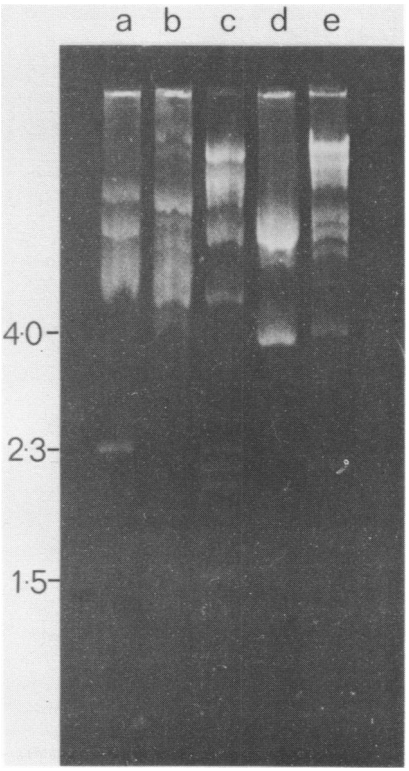

Figure 2. Digests of clones containing primary fragments used for sequence analysis. Samples of the following digests were analysed on a 1% agarose gel [5]. a, b, and d: EcoRI digests of mp3.NH4, mp3.NH1.5 and mp7.NB4 showing release of fragments containing a 2.3 kb HindIII-EcoRI fragment, a 1.5 kb HindIII fragment and a 4.0 kb BamHI fragment, respectively. Each fragment contains short M13 linker sequences at its ends. c and e: λAsn5 DNA digested with HindIII and BamHI, respectively.

known cleavage sites of enzymes which cut rarely to prepare clones covering the region of interest. These are sequenced, at least in part, revealing new suitable restriction sites. These are used to prepare a further simple mixture of clones from which clones appropriate for extension of the existing DNA sequence are identified. Thus, the sequence is built up in an ordered manner. The major advantage of this approach over a random strategy [5] is that the isolation of many duplicate clones covering known sequence is avoided.

The DNA sequence and an interpretation of it showing the proposed promoter and genes is shown in Figure 3. The sequence extends from the promoter region to the beginning of the uncA gene encoding the α-subunit of ATP-synthase [4]. It appears to contain five genes as discussed below. Also it resolves two ambiguities in the DNA sequence in the beginning of the uncA gene [4].

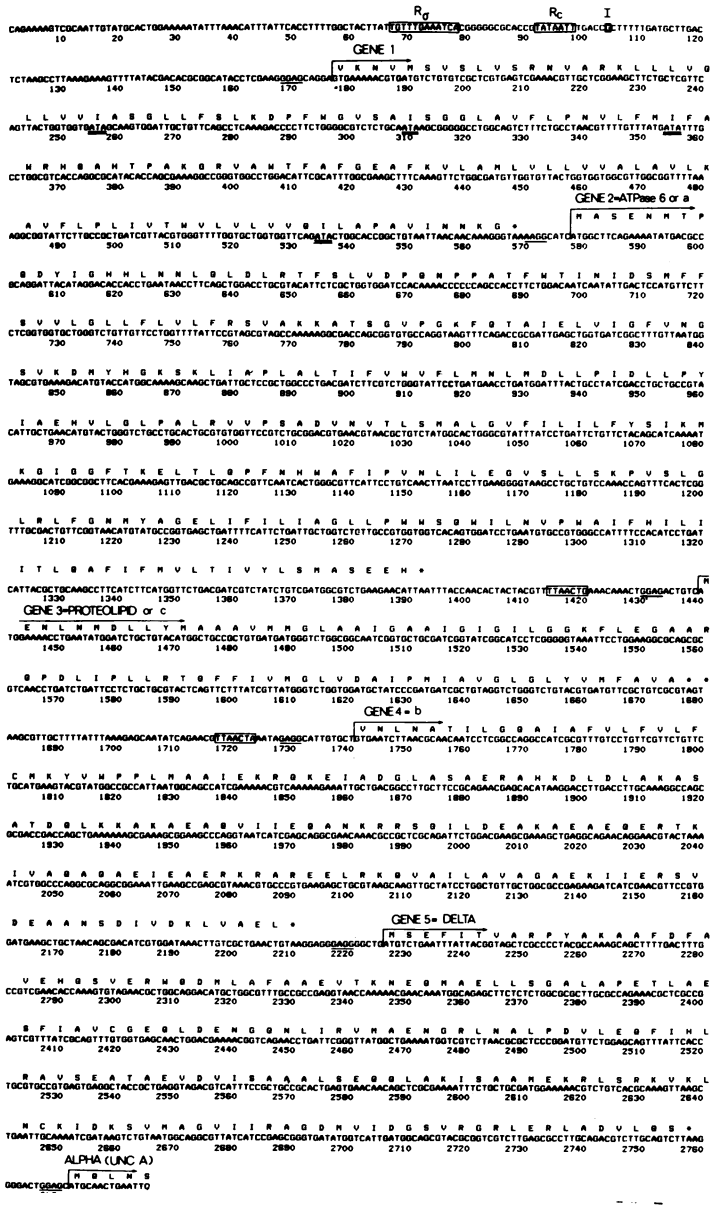

Figure 3. Nucleotide sequence in the promoter proximal region of the unc operon. R_σ and R_C are σ binding and core binding sequences of the proposed promoter. I is the proposed point of initiation of transcription. Underlined sequences are potential ribosome binding sites. The boxes between genes 2 and 3 and 3 and 4 contain sequences similar to each other and to the R_C promoter sequence of trp [8]. Their significance is obscure. ATA codons in gene 1 are doubly underlined.

DISCUSSION

The section of sequence of the atp or unc operon summarised in Figure 3 contains the promoter and five reading frames which could be genes for proteins. The proposed promoter resembles the canonical promoter sequences for σ and core binding sites for RNA polymerase derived by comparison of all known promoters [8] (Fig. 4). Transcription studies with E. coli RNA polymerase have confirmed the presence of an active promoter in this region of the DNA sequence (unpublished results).

The reading frames shown in Figure 3, called genes 1-5, are characterised by an initiation codon ATG or GTG preceded by a sequence complementary to the 3' end of 16S rRNA which would serve as a ribosome binding site [9] and followed by a coding sequence ending with a stop codon. With the exception of gene 3, which has two consecutive stop codons, TAG and TAA, all genes in this section of the operon terminate with stop codon TAA which is presumed to be the major stop codon in E. coli [10].

Gene 1 has been tentatively assigned to the first reading frame following the promoter. The putative gene product would be a hydrophobic protein of molecular weight 14,183 daltons with a net positive charge of 11 (assuming an unblocked amino terminus). These parameters do not correspond to any of the proteins characterised as constituents of the ATP-synthase complex nor to any of the abundant proteins produced by expression studies of the unc operon in vivo [11-13] or in vitro [13]. However, gene 1 contains four of the six AUA codons to be found in the entire operon [4 and unpublished results]. This codon is infrequently used in E. coli [14], the corresponding tRNA has a low abundance [15], and the presence of four AUA codons may be indicative of expression at low levels [14]. A possible function for this protein consistent with its hydrophobicity, positive charge and expression would be to function as a pilot protein to guide assembly of the membrane sector of the enzyme complex.

Gene 2. This codes for a hydrophobic protein of molecular weight 30,267

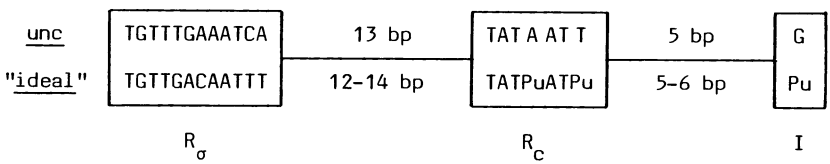

Figure 4. Comparison of the proposed unc promoter with "ideal" promoter sequences [8].

daltons. The predicted protein sequence is homologous in the carboxy terminal region with ATPase-6 gene product of human mitochondria [16] as shown in Figure 5, which has been identified from its homology with the corresponding yeast protein [17]. These properties suggest that gene 2 codes for protein a by gel electrophoresis in the presence of dodecyl sulphate. However, the molecular weights of other hydrophobic proteins (e.g. bacteriorhodopsin [18] and cytochrome oxidase subunit I [16]) have been similarly underestimated.

Gene 3. The predicted protein sequence is identical to the sequence of the proteolipid or protein c described by Hoppe, Schairer and Sebald [19].

Gene 4. This encodes a protein of molecular weight 17,212 daltons. Its sequence is very striking: residues 1-32 contain only one charged amino acid, lysine 23, and are otherwise hydrophobic; in contrast the rest of the protein is highly charged. This suggests that the amino terminal region of the protein (approximately 1-30) is buried in the lipid bilayer. The rest of the sequence which is predicted to be almost entirely α-helical (unpublished work) would protrude from the membrane. It would be envisaged that this protein might play a central role in the enzyme providing a central charged domain anchored to the membrane, around with the F_1 proteins could assemble. It would also play a role in proton translocation from the membrane proton channel to the catalytic sites.

Gene 5. The predicted protein has a molecular weight of 19,310 daltons and a net charge of -7.5. These properties correspond to the δ-protein which has been genetically mapped to a position upstream of uncA [2].

CONCLUSION

This sequence taken with the sequence of the genes for the α, β, γ and ε subunits [4 and unpublished work] shows that the atp or unc operon probably

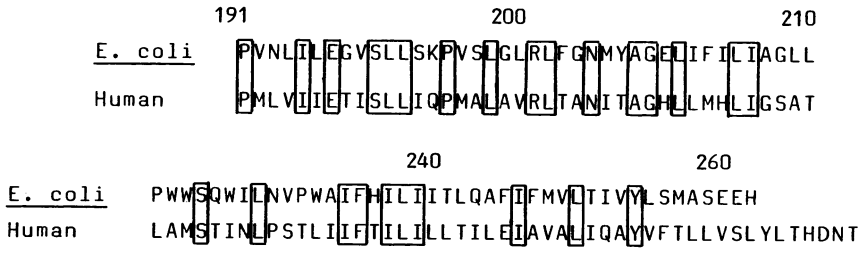

Figure 5. Comparison of the sequences of gene 2 protein with ATPase-6 of human mitochondria in the carboxyl terminal region of the proteins. The numbering is based on the E. coli protein.

contains nine genes in the order genes 1-4: δ, α, γ, β, ε. Eight are structural genes for the proteins of ATP-synthase complex, a ninth may be a pilot to direct membrane assembly.

ACKNOWLEDGEMENTS

We thank Dr. S. Brenner for a gift of λAsn5. (The phage was originally donated to him by Dr. Yura.) Also Dr. C. Lichtenstein for λAsn5 DNA, Drs. M. Saraste and A.A. Travers for discussions and Dr. F. Sanger for his advice and encouragement. N.J.G. is supported by an MRC Studentship.

*To whom correspondence should be sent

REFERENCES

1. Von Meyenberg, K. & Hansen, F.G. (1980) in Mechanistic Studies of DNA replication and Genetic Recombination. ICN-UCLA Symp. Vol. XIX, Alberts, B. and Fox, C.F., eds., Academic Press, New York. In the press.
2. Downie, J.A., Gibson, F. & Cox, G.B. (1979) Ann. Rev. Biochem. 48, 103-131.
3. Futai, M. & Kanazawa, H. (1980) Curr. Topics Bioenerg. 10, 181-215.
4. Gay, N.J. & Walker, J.E. (1981) Nucleic Acids Res. 9, 2187-2194.
5. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. & Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
6. Birnboim, H.C. & Daly, J. (1979) Nucleic Acids Res. 7, 1513-1529.
7. Winter, G., Fields, S. & Gait, M.J. (1981) Nucleic Acids Res. 9, 237-245.
8. Pribnow, D. (1978) in "Biological Regulation and Development" Col. I "Gene Expression". R.F. Goldberger, ed., Plenum Press, pp. 219-277.
9. Shine, J. & Dalgarno, L. (1974) Proc. Nat. Acad. Sci. USA 71, 1342-1346.
10. Brenner, S., Stretton, A.O. & Kaplan, S. (1965) Nature 206, 994-998.
11. Downie, J.E., Cox, C.B., Langman, L., Ash, G., Becker, M. & Gibson, F. (1981) J. Bacteriol. 145, 200-210.
12. Foster, D.L., Mosher, M.E., Futai, M. &Fillingame, R.H. (1980) J. Biol. Chem. 255, 12037-12041.
13. Brusilow, W.S.A., Gunsalus, R.P., Hardeman, E.C., Decker, K.P. & Simoni, R.D. (1981) J. Biol. Chem. 256, 3141-3144.
14. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981) Nucleic Acids Res. 9, r43-r74.
15. Ikemura, T. (1981) J. Mol. Biol. 146, 1-21.
16. Anderson, S., Bankier, A., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. & Young, I.G. (1981) Nature 290, 457-465.
17. Macino, G. & Tzagoloff, A. (1980) Cell 20, 507-517.
18. Bridgen, J. & Walker, I.D. (1976) Biochemistry 15, 792-797.
19. Hoppe, J., Schairer, H.U. & Sebald, W. (1980) Eur. J. Biochem. 112, 17-24.