Vol. 102, No. 1, 1981 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
September 16, 1981 Pages 172-179

NUCLEOTIDE SEQUENCE OF THE GENE CODING FOR THE δ SUBUNIT OF PROTON-TRANSLOCATING ATPase OF *Escherichia coli*

Kazunori Mabuchi, Hiroshi Kanazawa, Toshiaki Kayano and Masamitsu Futai

Department of Microbiology, Faculty of Pharmaceutical Sciences
Okayama University, Okayama 700, Japan

Received July 10, 1981

SUMMARY: The DNA sequence of a part of the gene cluster for the proton-translocating ATPase of *E. coli* was determined. Two reading frames were found between the genes for dicyclohexylcarbodiimide-binding protein and α subunit. The following evidence indicates that one frame codes the δ subunit; i) The primary structure deduced from DNA sequence agreed with amino acid composition of the protein. ii) Five residues from the amino terminal were the same as those deduced from DNA sequence. The α helix content of this protein (estimated from the primary structure) was 60.2% of the total residues with the longest helical domain of 50 residues. The intercistronic sequence between the genes for δ and α had 15 base pairs, suggesting that the synthesis of the α is not regulated transcriptionally. The organization of the gene cluster was also shown on the physical map.

The H^+-ATPase of *E. coli* catalyzes the synthesis of ATP utilizing a proton gradient established by the electron transfer chain. The ATPase consists of two portions, F_1 and F_0. F_1 has a catalytic function as an ATPase and has five different subunits, α, β, γ, δ, and ε. F_0 has three subunits (24K, 19K, and DCCD-binding protein) (4,5) and is intrinsic to membranes, while F_1 is extrinsic to membranes. All the structural genes for the ATPase are clustered in the 83 min region on the *E. coli* linkage map (6,7). A detailed physical map of this region was constructed, and structural genes for subunits were located on the map (8-11). Recently, we determined the nucleotide sequence of the gene cluster using a hybrid plasmid pMCR533 carrying most of the cluster and found sequences corresponding to the gene for DCCD-binding protein (*papH*) and the amino terminal portion of the α subunit gene (*papA*)

Abbreviations used: H^+-ATPase, proton-translocating ATPase; F_1, peripheral membrane component of H^+-ATPase; F_0, intrinsic membrane component of H^+-ATPase; DCCD, N, N'-dicyclohexylacrbodiimide; DCCD-binding protein, a subunit of F_0 capable of binding DCCD; α, β, γ, δ, and ε, five subunits of F_1. 24K, 18K, 19K and 14K indicate polypeptides with apparent molecular weights of 24,000, 18,000, 19,000 and 14,000 daltons, respectively. As discussed in the text 18K and 19K may be the same component in F_0, and the slight difference of the molecular weights may be within experimental errors in calibrating molecular weights.

0006-291X/81/170172-08$01.00/0

(12). From the exact loci of these two genes we located structural genes in the following order: *uncB* (F_0), *papH*, *papA*, *papC* (γ) and *papB* (β) (12).

We are attempting to determine the whole nucleotide sequence of the gene cluster to obtain information on the fine structure of each subunit and the mechanism of gene expression. In this study we found a nucleotide sequence that codes the δ subunit between *papH* and *papA*. The primary structure of the subunit showed that it has a high helical content as a possible secondary structure. The intercistronic region between genes coding for the δ and α subunits was only 15 base pairs, suggesting that synthesis of the α subunit is not regulated at a transcriptional level.

MATERIALS AND METHODS

Preparation of plasmid DNA and its fragments. A hybrid plasmid pMCR533 (8) was used in this study. The plasmid DNA was digested with appropriate restriction endonucleases according to the sequencing strategy shown in Fig.1. Restriction endonucleases were purchased from Boehringer Mannheim GmbH, New England Biolabs and Takara Shuzo Co., Japan.

Nucleotide sequencing. A DNA fragment recovered from polyacrylamide gel was labeled with ^{32}P-γ-ATP with T4-polynucleotide kinase (Boehringer Mannheim GmbH), and the nucleotide sequence was determined by the method of Maxam and Gilbert (13).

Determination of the amino acid sequence by Edman degradation. About 100μg of the δ subunit from *E. coli* ML308-225 was subjected to manual Edman degradation (14). The phenylthiohydantoin derivatives of amino acids in each cycle were analyzed by high performance liquid chromatography on a C18μ Bondapack column (Waters Associates, Inc.).

RESULTS

The DNA sequences corresponding to the structural genes for DCCD-binding protein (*papH*) and α subunit (*papA*) were located on a hybrid plasmid pMCR533 (12). Two successive reading frames were found in the region between *papH* and *papA*. The first frame adjacent to *papH* encodes a protein of 156 amino acids, possibly 18K of F_0, which may be the same as the protein of 19,000 daltons observed previously in purified F_0 (5) or the 18,000 dalton protein of the in vitro translation product (11). However, more detailed sequencing and protein chemical studies are required to establish the complete structure of this putative protein. The first frame has one termination codon (TAA) which is only 13 base pairs apart from the start of the next reading frame. In this region we found a Shine-Dalgarno sequence (15), sug-

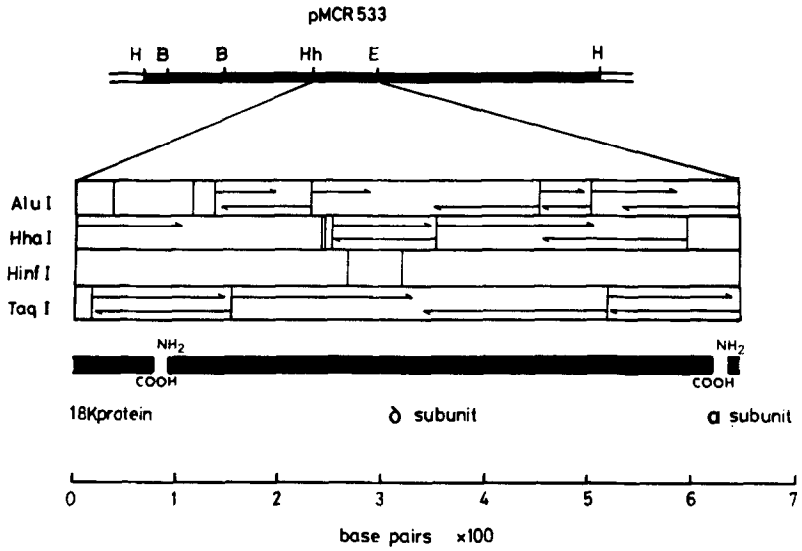

Fig. 1 Nucleotide sequencing strategy. *E. coli* DNA (about 4,300 base pairs) carried on hybrid plasmid pMCR533 is shown at the top of the figure. The cleavage sites by various restriction endonucleases are shown as follows: E, *EcoRI*; B, *BamHI*; H, *HindIII*; Hh, *HhaI*. The detailed cleavage maps with *AluI*, *HhaI*, *HinfI*, and *TaqI* are also shown. Arrows indicate the sequenced DNA segments with its direction and approximate length. The location of the DNA segment coding for the δ subunit is shown with amino ($-NH_2$) and carboxyl (-COOH) terminals.

gesting that this region is a preceding part of the next gene. We determined the nucleotide sequence of the second frame and the flanking region by analyzing both strands of the DNA segment according to the sequencing strategy shown in Fig.1. The primary structure of the putative protein (176 residues) was deduced from the nucleotide sequence (Fig. 2).

From the following evidence we conclude that this protein is the δ subunit of F_1: i) Five residues in the amino terminal were the same as those determined in this study by Edman degradation, namely Met-Ser-Glu-Phe-Ile. ii) The amino acid composition was in good agreement with protein chemical data (16), although the contents of arginine, isoleucine and proline were slightly different (Table I). The molecular weight of the protein obtained from the primary structure was 19,582, while that of δ determined by gel electrophoresis was 18,500 (16). iii) The secondary structure determined from the primary structure supports this notion as discussed below. The intercistronic region between the genes for δ and α subunits was 15 base pairs and included one termination codon (TAA).

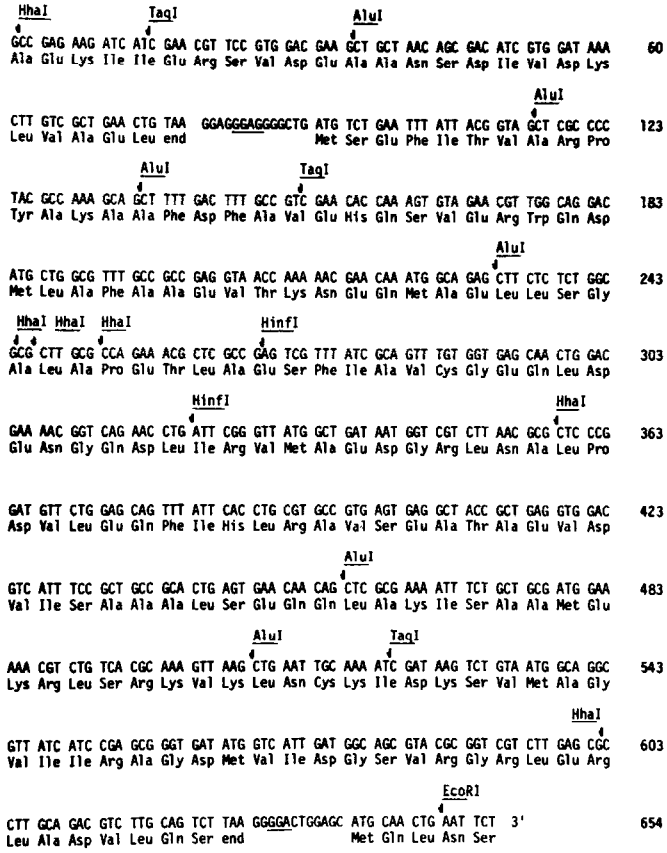

Fig. 2 Nucleotide sequence of the structural gene of the δ subunit and its flanking region. The nucleotide sequence in the antisense strand is shown with the deduced amino acid sequence and cleavage sites for restriction endonucleases. The DNA sequences coding for the carboxyl terminal of a putative protein (18K), and the amino terminal of the α subunit are also shown. The Shine-Dalgarno sequence (15) preceding each gene is underlined.

DISCUSSION

Secondary structure of the δ subunit. From the primary structure of the δ subunit, we estimated the possible secondary structure by the procedure of Chou and Fasman (17) (Fig.3). The structure was found to be highly helical: the helical content is 60.2% of the total residues. The longest domain of the helix consists of 50 residues (Fig.3). It should be noted that this is an exceptionally long helical domain in comparison with those of other proteins so far studied (18). This high helical content is in good agreement with previous results obtained by circular dichroism (55-70%) (16).

Table I
Amino Acid Composition of δ Subunit of E.coli F_1

Amino Acid	Predicted from DNA Sequence	Reported*
	moles/19,558 daltons	
Phe	6	5.89
Leu	18	17.1
Ile	10	8.37
Met	7	5.95
Val	16	14.5
Ser	12	13.0
Pro	3	4.32
Thr	4	4.28
Ala	27	26.3
Tyr	1	0.99
His	2	2.09
Gln + Glu	26	26.4
Asn + Asp	15	15.5
Lys	8	8.32
Cys	2	2.23
Trp	1	1.19
Arg	11	10.6
Gly	8	9.43

* Reported mol per cent values (16) were multipled by 177 residues for the δ subunit.

Flanking sequence. We found that there are only 15 base pairs between the carboxyl terminal end of the δ subunit gene and the amino terminal end of the α subunit gene. A Shine-Dalgarno sequence (15) was found just before the initiation codon of the gene for the α subunit. It was reported that the α and β subunits in *E. coli* F_1 have three copies each, while other subunits γ, δ and ε have only one copy each (19). Therefore there must be some regulation of the synthesis of these subunits to establish a suitable molar ratio of each subunit in F_1. The structural gene for H^+-ATPase may be transcribed as an operon (20). If there is transcriptional regulation, there should be a regulatory site(s) (internal promotor or operator site(s)) within the operon. However, the intercistronic sequence is too short to expect such a site(s) (21), suggesting that there is no transcriptional regulation between expression of the genes for the δ and α subunit. The situation seems to be the same in the intercistronic region (13 base pairs) between a putative protein (possibly 18K) and δ subunit genes.

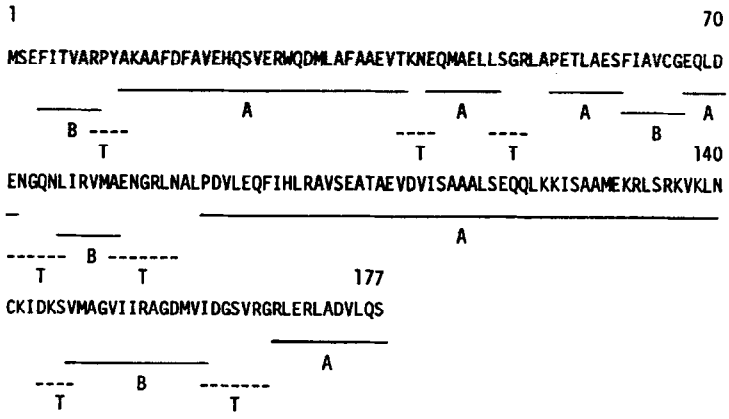

Fig. 3 Putative secondary structure of the δ subunit. The secondary structure of δ was estimated by procedure of Chou and Fasman (17). Amino acid residues are shown by single letter symbols (18). The α helix (A), β sheet (B) and turn (T) are indicated under the amino acid sequence. Each α helical domain shown here ends one amino acid residue before the residue known to be a helix breaker.

Codon usage. There are some differences in the usage of several codons in the δ subunit from those in other proteins: the usage of CTG in δ is 56% of the total codon used for Leu, whereas the usages of the same codon in ribosomal proteins (23) and the DCCD-binding protein (12) are 91% and 92%, respectively, of the total codon used for Leu. A similar type of difference was observed in usage of several other codons and may be interesting for understanding the regulation of the synthesis of this subunit.

Organization of the gene cluster for the H^+-ATPase subunits. We have proposed nomenclature (based on the gene products) for the structural genes coding for subunits of the H^+-ATPase (12). According to this system of nomenclature, we propose to call the gene coding for δ *papE*. Recently, Downie *et al.* showed that the genes for F_1F_0 were arranged in the order of *uncB* (24K), [*uncF* (18K), *uncE* (DCCD-binding protein)], [δ subunit gene, 14K gene] and *uncA* (α subunit) (11), although the order of genes coding for δ and 14K and *uncE* and *uncF* were uncertain. In our sequencing study, we could not find a structural gene corresponding to 14K. In this regard it must be noted that purified active F_1F_0 did not contain 14K as a subunit (5). Recently, Brusilow *et al.* showed 24K, 18K and DCCD-binding protein but no 14K in their in vitro translation product of F_0 (24). From the present

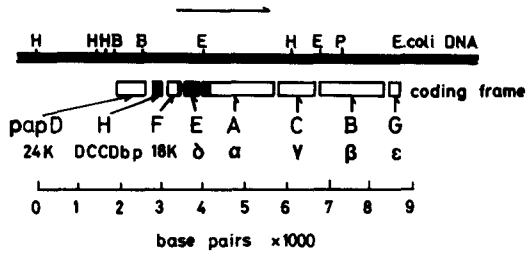

Fig. 4 Organization of the gene cluster for H^+-ATPase. On the basis of the nucleotide sequence determined in the present study together with previous results (11,12), the organization of the structural genes for H^+-ATPase is shown on the physical map of *E. coli* DNA. Coding frames shown by solid bars indicate regions in which nucleotide sequences were determined in the present study or previously (12). Coding frames shown by open bars were estimated from the results in this study and previous observations (11,12). The locus of the gene for 24K (*papD*) is according to Downie *et al.* (11). The loci for *papC*, *B*, and *G* were discussed previously (12). Cleavage sites by *Pst*I (P) and other endonucleases (legend to Fig. 1) are shown in abbreviated forms. The arrow at the top of the figure indicates the direction of transcription.

results and previous ones (11,12), we summarize the order of structural genes coding for F_1F_0 as follows: *papD* (24K), *papH* (DCCD-binding protein), *papF* (18K), *papE* (δ), *papA* (α), *papC* (γ), *papB* (β), and *papG* (ε) (Fig.4).

Acknowledgement. We thank Dr. K. Tamagake for providing a computer program for determining coding frames. We also thank Ms. Keiko Takeda for technical assistance in amino acid sequencing. This study was supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan and by a grant from the Japan-US Cooperative Science Program.

REFERENCES

1. Racker, E. (1976) "A New Look at Mechanisms in Bioenergetics", Academic Press, New York.
2. Kagawa, Y. (1978) Biochim. Biophys. Acta, 505, 45-93.
3. Futai, M., and Kanazawa, H. (1980) Curr. Top. Bioenerg. 10, 181-215.
4. Fillingame, R. H. (1980) Ann. Rev. Biochem. 49, 1079-1113.
5. Foster, D. L., and Fillingame, R. H. (1979) J. Biol. Chem. 254, 8230-8236.
6. Downie, J. A., Gibson, F., and Cox, G. B. (1979) Ann. Rev. Biochem. 48, 103-131.
7. Kanazawa, H., Miki, T., Tamura, F., Yura, T., and Futai, M. (1979) Proc. Natl. Sci. U.S.A. 76, 1126-1130.
8. Kanazawa, H., Tamura, F., Mabuchi, K., Miki, T., and Futai, M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 7005-7009.
9. Tamura, F., Kanazawa, H., Tsuchiya, T., and Futai, M. (1981) FEBS Lett. 127, 48-52.
10. Downie, J. A., Langman, G., Cox, G. B., Yanofsky, C., and Gibson, F. (1980) J. Bacteriol. 143, 8-17.
11. Downie, J. A., Cox, G. B., Langman, L., Ash, G., Becker, M., and Gibson, F. (1981) J. Bacteriol. 145, 200-210.
12. Kanazawa, H., Mabuchi, K., Kayano, T., Tamura, F., and Futai, M. (1981) Biochem. Biophys. Res. Comm. 100, 219-225.

13. Maxam, A. W., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
14. Tarr, G. E. (1977) Methods in Enzymol. 47, 335-357.
15. Shine, J., and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1342-1346.
16. Sternweis, P. C., and Smith, J. B. (1977) Biochemistry, 16, 4020-4025.
17. Chou, P Y., and Fasman, G. D. (1974) Biochemistry, 13, 222-245.
18. Shultz, G. E., and Shirmer, R. H. (1979) "Principles of Protein Structure", Springer-Verlag, Heidelberg.
19. Bragg, P. D., and Hou, C. (1975) Arch. Biochem. Biophys. 167, 311-321.
20. Gibson, F., Downie, J. A., Cox, G. B., and Radik, J. (1978) J. Bacteriol. 134, 728-736.
21. Rosenberg, M., and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
22. Nicols, B. P., and Yanofsky, C. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 5244-5248.
23. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H., and Dennis, P. P. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1697-1701.
24. Brusilow, W. S. A., Gunsalus, R. P., Hardeman, E. C., Decker, K. R., and Simoni, R. D. (1981) J. Biol. Chem. 256, 3141-3144.