Vol. 100, No. 1, 1981 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 15, 1981 Pages 219-225

NUCLEOTIDE SEQUENCE OF GENES CODING FOR DICYCLOHEXYLCARBODIIMIDE-BINDING PROTEIN AND THE α SUBUNIT OF PROTON-TRANSLOCATING ATPase of *ESCHERICHIA COLI*

Hiroshi Kanazawa, Kazunori Mabuchi, Toshiaki Kayano
Fumie Tamura and Masamitsu Futai

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

Received March 27,1981

SUMMARY: The recombinant plasmid pMCR533 carries part of the gene cluster coding for the proton-translocating ATPase (F_1-F_0) of *E. coli*. The DNA sequence of the structural gene for dicyclohexylcarbodiimide-binding protein was obtained by analyzing a DNA fragment from this plasmid. The DNA sequence corresponding to the amino terminus of the structural gene for the α subunit was also found on the plasmid. Both sequences were in good agreement with the amino acid sequences determined previously. From the exact locations of these two genes we have located genes coding for other components of F_1-F_0 on the defined portion of *E. coli* DNA. A proposal is made to use the nomenclature *pap* (proton-translocating ATPase subunit protein) for the structural gene cluster for F_1-F_0.

The proton-translocating ATPase (F_1-F_o) of *E. coli* has the same basic structure as those from mitochondria and chloroplasts (1-3). The ATPase is a complex membrane enzyme consisting of two portions: F_1 (5 different subunits) and F_o (at least 2 different subunits). Information is still limited regarding the primary structures of the subunits. So far, only the complete amino acid sequences of proteolipid subunits of F_o (DCCD-binding protein) from various origins have been determined (4). Determination of the primary structures of other subunits should provide important clues for understanding the function and assembly of this complex enzyme.

In a series of elegant genetic studies, Gibson and his colleagues identified 7 different genes (*unc* genes) coding for components of F_1-F_o of *E. coli* and mapped

---

Abbreviations used: F_1, the portion of the proton-translocating ATPase that is peripheral to the membrane; α, β and γ subunits, three major subunits of F_1; F_0, integral membrane portion of the proton-translocating ATPase; DCCD, dicyclohexylcarbodiimide; DCCD-binding protein, a subunit of F_0 functioning as a proton channel; M Dal, megadalton; bp, base pairs.

0006-291X/81/090219-07$01.00/0

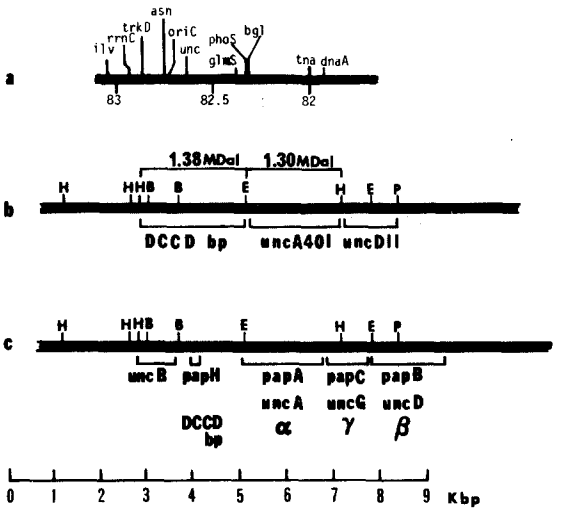

Fig. 1 Organization of the gene cluster (*unc*) coding for F_1-F_0 of *E. coli*.
(a) *E. coli* linkage map around the gene cluster (*unc*) coding for F_1-F_0 (23).
(b) Organization of the structural genes for F_1-F_0 on the physical map of *E. coli*. Location of a gene coding for DCCD-binding protein (DCCD bp) and mutant alleles of the α subunit (*uncA401*) and β subunit (*uncD11*) were determined by genetic complementation as reported previously (7,8). The cleavage sites of endonucleases *Hin*dIII (H), *Eco*RI (E), *Pst*I (P) and *Bam*HI (B) are shown.
(c) Organization of the structural genes for F_1-F_0 determined in the present study. The structural genes coding for DCCD-binding protein and the α subunit were located from determination of the DNA sequence. The locations of other genes were estimated as described in the text. Nomenclatures of structural genes proposed in this paper are shown.

them at about 82.5 min of the linkage map (5)(Fig. 1a). Recently we found that a transducing phage λ*asn*-5 carries a whole set of structural genes for F_1-F_o (6) and we localized them on a defined segment (4.5 M Dal molecular weight) of *E. coli* DNA carried by this phage (7)(Fig. 1b). The approximate location of the genes coding for DCCD-binding protein of F_o and the α and β subunits of F_1 were determined previously by genetic complementation (7,8)(Fig. 1b). On the basis of these findings, in the present study we obtained the nucleotide sequences corresponding to the complete amino acid sequence of DCCD-binding protein and the amino terminal sequence of the α subunit. From these results we located the genes coding for the DCCD-binding protein and α subunit exactly on the DNA segment and discussed the organization of the genes coding for other subunits of F_1-F_o, including the genes for β and γ. Now it will be easier to determine the DNA sequences of other subunits, and to deduce primary structures from the nucleotide sequences.

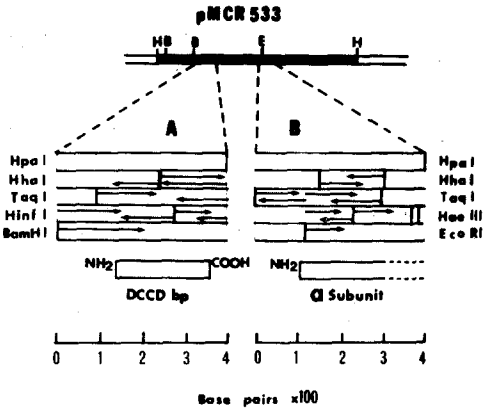

Fig. 2 Restriction fragments used for DNA sequencing.
The restriction maps of fragments A and B, carrying genes for DCCD-binding protein and the amino terminal portion of the α subunits, respectively, are shown. Restriction sites are shown by solid lines. The *arrows* indicate the direction and approximate length of the fragment sequenced. The approximate locations of DNA segments coding for DCCD-binding protein and the α subunit are shown with amino ($-NH_2$) and carboxyl (-COOH) terminals.

EXPERIMENTAL PROCEDURES

Preparation of plasmid DNA and its fragments. Covalently closed circular DNA of pMCR533 was prepared from strain KY7625 harboring this plasmid as described previously (7). The DNA fragments used for determination of nucleotide sequences were prepared by digesting the plasmid DNA with various restriction endonucleases according to the sequencing strategy. The restriction endonucleases used in the present study were purchased from Bethesda Research Lab., New England Biolab., and Takara Shuzo Co., Japan.

Nucleotide sequence determination. The 5'-end of DNA fragments was phosphorylated by T_4 polynucleotide kinase (Boehringer Mannheim GmbH) with γ-^{32}P-ATP prepared as described previously (9). Labeled DNA was cleaved chemically and its nucleotide sequence was determined by the method of Maxam and Gilbert (10).

RESULTS

Determination of the nucleotide sequence coding for DCCD-binding protein. The structural gene for DCCD-binding protein was mapped previously within the 1.38 M Dal DNA segment between the promoter proximal *Hin*dIII and *Eco*RI endonuclease sites (8)(Fig. 1b, 2). This DNA segment is carried by a hybrid plasmid pMCR533. In preliminary experiments, we sequenced most of the region of the DNA segment prepared from pMCR533, and found a sequence corresponding to the gene coding for DCCD-binding protein in the middle of the 1.38 M Dal fragment (about 1200 bp from the promoter proximal *Hin*dIII site)(Fig. 1c, 3). The amino acid sequence of the

```
                                                                    60
5'  ATG GAA AAC CTG AAT ATG GAT CTG CTG TAC ATG GCT GCC GCT GTG ATG ATG GGT CTG GCG
    Met Glu Asn Leu Asn Met Asp Leu Leu Tyr Met Ala Ala Ala Val Met Met Gly Leu Ala

                                                                HhaI
                                                                    120
    GCA ATC GGT GCT GCG ATC GGT ATC GGC ATC CTC GGG GGT AAA TTC CTG GAA GGC GCG GCG
    Ala Ile Gly Ala Ala Ile Gly Ile Gly Ile Leu Gly Gly Lys Phe Leu Glu Gly Ala Ala
              A                B
    HhaI              HinfI
                                                                    180
    CGT CAA CCT GAT CTG ATT CCT CTG CTG CGT ACT CAG TTC TTT ATC GTT ATG GGT CTG GTG
    Arg Gln Pro Asp Leu Ile Pro Leu Leu Arg Thr Gln Phe Phe Ile Val Met Gly Leu Val

                                                                    240
    GAT GCT ATC CCG ATG ATC GCT GTA GGT CTG GGT CTG TAC GTG ATG TTC GCT GTC GCG TAG
    Asp Ala Ile Pro Met Ile Ala Val Gly Leu Gly Leu Tyr Val Met Phe Ala Val Ala end


    TAA  3'
    end
```

Fig. 3 DNA sequence of the structural gene of DCCD-binding protein of *E. coli*. The sequence shown is that of the antisense strand. Nucleotides 1 and 243 correspond to the beginning and end of the gene. The amino acid sequence is based on the codons of the universal code, and is in complete agreement with the results of protein sequencing (4). Homologous amino acid residues in *E. coli* and *Saccharomyces* are underlined. The nucleotide sequence of the A and B region corresponding to those of *Saccharomyces* are GGA GCA and ATT GGT ATT, respectively (26,27).

protein deduced from the DNA sequence is in complete agreement with that obtained by a protein chemical procedure (4,11)(Fig. 3). The code for the DCCD-binding site is GAT (aspartic acid residue at position 61) and two successive termination codons were found (Fig. 3). The direction of the reading frame of the nucleotide sequence indicated that the gene cluster is transcribed from the *asn* side to the *glmS* side of the chromosome, confirming previous genetic studies (5,6). We could not find other copies of the gene within the 1000 bp from the amino and carboxyl terminal ends of the gene. These results suggest that this protein is coded for by a single copy of the structural gene, although it was shown that F_1-F_o has an oligomer(possibly a hexamer) of the proteins (12,13).

Determination of the DNA sequence coding for the amino terminal portion of the α subunit. The structural gene locus for the α subunit was also carried on pMCR533 and mapped on the 1.30 M Dal segment of DNA (Fig. 1b). We have sequenced the region around the *Eco*RI site in the middle of the DNA segment derived from *E. coli* and found the sequence corresponding to 69 amino acid residues of the amino terminal of the α subunit. First 26 residues except 23rd were in complete

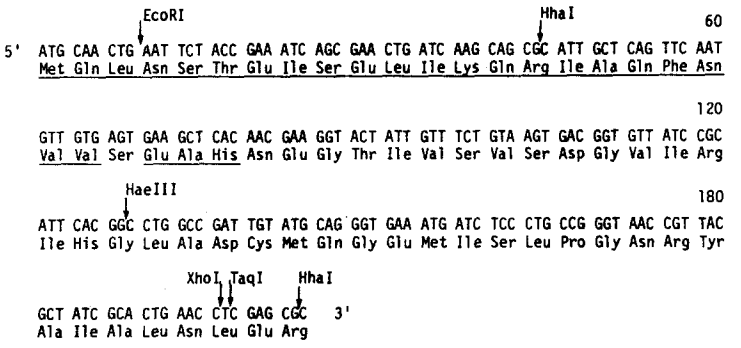

Fig. 4 DNA sequence of the structural gene of the amino terminal portion of the α subunit of *E. coli* F_1.
The sequence is that of the antisense strand. Nucleotides 1 and 312 correspond to the beginning and sequenced end of the gene. The amino acid sequence was deduced as described in the legend of Fig. 3. The 25 amino acid residues identified by Edman degradation (14) are underlined.

agreement with the amino terminal sequence reported by Dunn *et al.*(14)(Fig. 4). The 23rd amino acid, which they did not determine, was deduced to be serine from our result. The α subunit has about 539 amino acid residues, as estimated by amino acid analysis (15). Thus we estimated that the carboxyl end of the gene for this subunit is about 1617 (539 x 3) bp apart from the first letter of the initiation codon. The entire sequence of the gene, together with results of chemical analysis of the protein, will be published elsewhere.

DISCUSSION

Organization of the gene cluster coding for F_1-F_o. This is the first report of the DNA sequence of structural genes for bacterial F_1-F_o. Previously we found that the *uncB* gene and DCCD-binding protein gene are located within a DNA segment of 1.38 M Dal molecular weight (Fig. 1b)(7,8). Further evidence from other laboratories (16,17) indicated that the *uncB* locus is between *Hin*dIII and the second *Bam*HI site (Fig. 1c), suggesting that *uncB* does not code for DCCD-binding protein because the DNA sequence of this protein was found in an entirely different portion of DNA in this study (Fig. 1c). Downie *et al.* found another gene (*uncE*) between the *uncB* and *uncA* genes (18). The *uncE* gene may code for DCCD-binding protein, although there is as yet no definite evidence of this. Discrepant results were reported on the

subunit composition of *E. coli* F_o (19-21). The nucleotide sequences of the *uncB* and *E* genes and data on the protein chemistry of polypeptides in the F_o preparation will identify the functional subunits of F_o. Recently, Downie *et al.* showed by genetic analysis that the *uncG* gene is located between *uncA* and *uncD* giving the gene order *uncBEAGDC* (17). They also suggested that the *uncG* gene has a *Hind*III site within its DNA and that it appeared to code for the γ subunit. We located *uncD11* mutation (22) previously (7), and the carboxyl and amino terminal of the α subunit (*uncA*) on a defined segment of DNA as discussed above. Taking these results together, we summarized the organization of the genes for subunits of F_1-F_o. This information is valuable for determining the DNA sequences of subunits with unknown primary structures.

Codon usage. The overall usage of codons in genes sequenced in this study is basically similar to that of other genes in *E. coli* (24,25). However, codon usage in DCCD-binding protein of *E. coli* is different from that of *Saccharomyces cerevisiae* (26,27) even in the region coding for homologous amino acid sequences, as indicated in Fig. 3(legend). The protein of *E. coli* also has homology with those from other sources, such as beef heart mitochondria and spinach chloroplasts (28). Thus determination of the nucleotide sequences of the genes from various origins will be of interest for consideration of molecular evolution.

A proposal for nomenclature of the structural gene for F_1-F_o and its subunits. Gibson and his colleagues isolated mutants with uncoupled oxidative phosphorylation and identified *unc* genes (5). The product of the gene cluster has been assigned as F_1-F_o (5,7) and subunits coded by some of the structural genes have been assigned. Thus we propose to change the nomenclatures of structural genes to those related to their products. Our proposal is to call the structural genes *papA, B, C, D, E, F, G* and *H*. The name *pap* is an abbreviation of proton-translocating ATPase subunit protein and *A* through *H* indicate subunits in order of decreasing molecular weight (Fig. 1c). Genes for the α subunit, β subunit, and DCCD-binding protein are referred to as *papA*, *papB*, and *papH*, respectively, by the new nomenclature. How-

ever, for other genes the old nomenclature should be used till their products have been assigned.

Acknowledgements: We are deeply indebted to Dr. T. Sekiya for kind instruction in procedures for nucleotide sequencing. We also thank Drs. T. Miki and T. Tsuchiya for suggestions and discussion. This study was supported by a Grant-in-Aid from the Ministry of Education, Science and Culture of Japan and a grant from the Japan-United States 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) Current Topics in Bioenergetics(ed. Sanadi, D.R.) 10, 181-215, Academic Press, New York.
4. Sebald, W., Sebald-Althaus, M. and Wachter, E. (1977) in "Energy Conservation in Biological Membranes" 29th Mösbacher Colloq., 228-236, Springer-verlag Berlin.
5. Downie, J.A., Gibson, F. and Cox, G.B. (1978) Ann.Rev. Biochem. 48, 103-131.
6. Kanazawa, H., Miki, T., Tamura, F., Yura, T. and Futai, M. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1126-1130.
7. Kanazawa, H., Tamura, F., Mabuchi, K., Miki, T. and Futai, M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 7005-7009.
8. Tamura, F., Kanazawa, H., Tsuchiya, T. and Futai, M. (1981) FEBS Lett.,in press.
9. Walseth, A.M. and Johnson, R.A. (1979) Biochim. Biophys. Acta 526, 11-31.
10. Maxam, A.W. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
11. Hoppe, J., Schairer, H.U. and Sebald, W. (1980) FEBS Lett. 109, 107-111.
12. Fillingame, R.H. (1976) J. Biol. Chem. 251, 6630-6637.
13. Altendorf, K. (1977) FEBS Lett. 73, 271-275.
14. Dunn, S.D., Heppel, L.A. and Fullmer, C.S. (1980) J. Biol. Chem. 255, 6891-6896.
15. Dunn, S.D. (1980) J. Biol. Chem. 255, 11857-11860.
16. Ogura, T. and Hiraga, S. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3993-3997.
17. Downie, J.A., Langman, L., Yanofsky, C. and Gibson, F. (1980) J. Bacteriol. 143, 8-17.
18. Downie, J.A., Senior, A.E., Gibson, F. and Cox, G.B. (1979) J. Bacteriol. 137, 711-718.
19. Foster, D.L. and Fillingame, R.H. (1979) J. Biol. Chem. 254, 8230-8236.
20. Friedl, P., Friedl, C. and Schairer, H.U. (1979) Eur. J. Biochem. 100, 175-180.
21. Schneider, E. and Altendorf, D. (1980) FEBS Lett. 116, 173-176.
22. Kanazawa, H., Horiuchi, Y., Takagi, M., Ishino, Y. and Futai, M. (1980) J. Biochem (Tokyo) 88, 695-703.
23. Miki, T., Kimura, M., Hiraga, S., Nagata, T. and Yura, T. (1979) J. Bacteriol. 140, 817-824.
24. 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.
25. Nicols, B.P. and Yanofsky, C. (1979) Proc. Natl. Acad. Sci. U.S.A. 76 5244-5248.
26. Hensgens, L.A.M., Grivell, L.A., Borst, P. and Bos, J.L. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1663-1667.
27. Marcino, G. and Tzagoloff, A. (1979) J. Biol. Chem. 254, 4617-4623.
28. Sebald, W., Hoppe, J. and Wachter, E. (1979) in "Functions and Molecular Aspects of Biomembrane Transport"(Quagliariello, E. et al. eds) pp. 63-74, Elsevier/North-Holland, Amsterdam, New York.