Vol. 103, No. 2, 1981 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1981 Pages 613-620

NUCLEOTIDE SEQUENCE OF THE GENES FOR F_0 COMPONENTS OF THE PROTON-TRANSLOCATING ATPase FROM ESCHERICHIA COLI: PREDICTION OF THE PRIMARY STRUCTURE OF F_0 SUBUNITS

Hiroshi Kanazawa, Kazunori Mabuchi, Toshiaki Kayano, Takato Noumi
Takao Sekiya* and Masamitsu Futai

Department of Microbiology, Faculty of Pharmaceutical Sciences
Okayama University, Okayama 700, and
*Department of Biology, National Cancer Center Research Institute
Tokyo 104, Japan

Received September 23,1981

SUMMARY: The 1763 nucleotide-long-DNA sequence of part of the gene cluster for the proton-translocating ATPase from E. coli was determined. The sequence covers the genes for the a and b subunits of F_0 along with the intercistronic regions. In the region preceding the gene for the a subunit, a reading frame encompassing 127 amino acids was found. The primary structure of the a and b subunits were deduced and the properties of these proteins were predicted. Analysis of codon usage in these genes was made.

The proton-translocating ATPase, F_0-F_1, of E. coli catalyzes the synthesis of ATP utilizing the proton gradient established by the respiratory chain(1-3). The membrane peripheral portion, F_1, has the catalytic site and 5 subunits($\alpha,\beta,\gamma,\delta$ and ε). The intrinsic membrane portion, F_0, was purified recently and found to have 3 subunits(a,b,c)(4,5). We have determined the DNA sequence of the genes for α, γ, δ and c subunits and part of the genes for β and b subunits(6-8). In the present study, we determined the 1763 nucleotide-long-DNA segment of the genes for the F_0. The primary structures of the a(30,258 daltons) and b(17,233 daltons) subunits were deduced from the DNA sequence. An open reading frame encompassing 127 amino acids was found before the gene for the a subunit, suggesting that the F_0-F_1 may have a ninth component. The significance of the intercistronic sequences in the gene cluster and codon usage in the genes for the a and b subunits are discussed extensively.

Abbreviations used: tRNA, transfer RNA; F_0-F_1, peripheral and integral portions of proton-translocating ATPase, respectively; a, b, and c subunits, the subunits of F_0 in the order of decreasing molecular weight. The subunits a, b and c were called 24K, 19K(or 18K) and 8K, respectively, in previous publications(4-8). The c subunit is also known as DCCD(dicyclohexylcarbodiimide)-binding protein(1,2).

0006-291X/81/220613-08$01.00/0

MATERIALS AND METHODS

Preparation of DNA fragments and DNA sequencing. Hybrid plasmids pMCR533(9), pAT1(10)and pKY159-16(constructed by K. Yamaguchi) were used in this study. Each plasmid covers a different portion of the gene cluster for F_0-F_1(Fig. 1). DNA fragments were prepared from these plasmids. They were labeled at the 5' end by T4-polynucleotide kinase with ^{32}P-γ-ATP. The DNA sequence was determined by the method of Maxam and Gilbert(11) All the reagents used and other methods were as described previously(6-8).

Genetic complementation. The E. coli strains used were obtained from B. J. Bachman(E. coli Genetic Stock Center): AN719, uncB402; AN1419, uncF469; AN936, uncE429(12). Transformation of these mutants with plasmids was carried out as described(9).

RESULTS AND DISCUSSION

Organization of the genes for F_0 components. Downie et al. showed that the uncB402, uncF469 and uncE429 mutations affected a, b, and c subunits of F_0, respectively, and suggested that genes were arranged in this order(12). However, the order of uncF and uncE was not demonstrated conclusively. We performed genetic complementation between the mutants and a hybrid plasmid pKY159-16 carrying the DNA segment of 1898 base pairs extending from the HindIII site to the HaeII (HhaI) site(Fig. 1). The uncB402 and uncE429 mutations were complemented by this plasmid, while uncF469 was not complemented, suggesting that the uncF locus exists downstream from uncB and uncE. The DNA sequence for the c and δ subunits were determined(6,7)(Fig. 1). Thus we could conclude the gene order as uncB(a), uncE(c) and uncF(b)(Fig. 1).

Primary structure of the b subunit deduced from the DNA sequence. We previously reported the DNA sequence of 78 base pairs coding the carboxyl terminal portion of b subunit(7). Here, the 1763 base pairs upstream from the carboxyl terminal of this subunit were determined including the sequence for the c subunit previously described(6)(Fig. 2). Three new open reading frames were found in the determined sequence. An open reading frame adjacent to the amino terminal end of the gene for δ subunit starts at position 1614 with the initiation codon GTG and ends at position 2084. This open frame encodes a protein of 156 amino acid residues(17,233 daltons). This agreed well with the protein chemical data of 19,000 dalton protein (4): residues 1-15 from the amino terminus were identical to the residues deter-

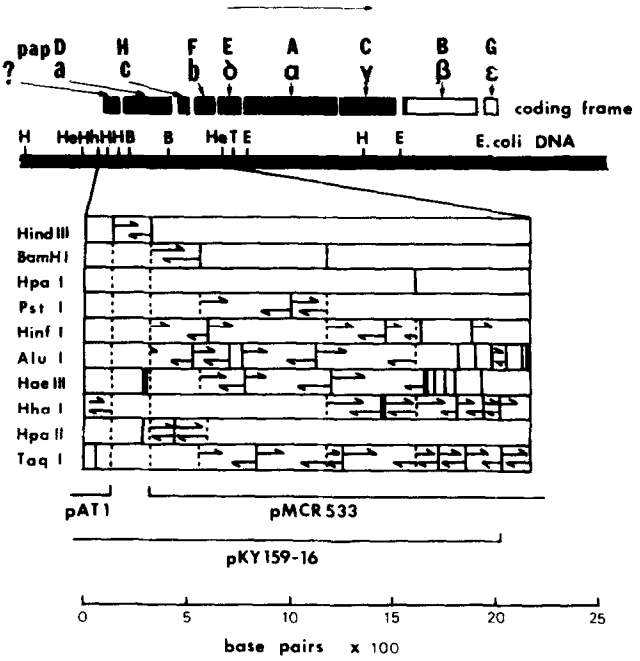

Fig. 1 Organization of the genes for F_0-F_1 and strategy of DNA sequencing. The direction of the transcription of the gene cluster is shown at top. The coding frame for each gene with its new nomenclature(pap)(6)(papD=uncB, papH=uncE, papF=uncF) and coding subunit is shown above the DNA. The coding frame shown by the solid bar indicates regions in which DNA sequence was determined in the present study or previously. Cleavage sites by endonucleases are shown as follows: E, EcoRI; H, HindIII; B, BamHI; Hh, HhaI; T, TaqI; He, HaeII. The cleavage maps with HpaI, PstI, HinfI, AluI, HaeIII, HhaI and HpaII are also shown. Arrows indicate the sequenced DNA segments with its direction and approximate length. DNA fragments were prepared from plasmid pMCR533, pAT1 and pKY159-16, which cover the regions shown. pKY159-16 is a derivative of pKY159 with insertion between the HaeII site(470 base pairs upstream from HindIII site(position 108)) and the HindIII site (position 108). pKY159 was constructed by ligating a E. coli DNA segment extending from the HaeII site(470 base pairs upstream from the HindIII site) to another HaeII site(2008) into HaeII site of pBR322(K. Yamaguchi, unpublished result).

mined by Edman degradation*. The content of polar residues in this protein(48.1%) is similar to that of soluble proteins(47 ± 6%)(11)(Table 1). However, it should be noted that hydrophobic domains were found near amino(11-33 residue) and carboxyl(124-132 residue) terminus(Fig. 3), suggesting that these portions are integrated into the membranes.

Primary structure of subunit a deduced from the DNA sequence. An open reading frame was found before the amino terminus of c subunit. Possible initiation codons were found at positions 451, 466, 535, 586 and 598(Fig. 2), while only one ter-

*Fillingame, R.H., Peters, L.K. and Hermodson, M., in preparation.

Fig. 2 DNA sequence of the genes coding for the a, b and c subunits along with the flanking regions. The DNA sequence in the antisense strand is shown with the deduced amino acid sequence. The sequences 1313-1555 and 2007-2107 were reported previously(6,8). Sequences similar to a Pribnow box(17) are at A(position 227), B(1287) and C(1580). An open reading frame(59-442) coding for a putative protein is shown with the deduced amino acid sequence. A Shine-Dalgarno sequence(14) is underlined.

mination codon was found at position 1266. A ribosome binding site (Shine-Dalgarno sequence(14)) was found before ATG(position 451). We believe that the open reading (451-1266) codes for the subunit a. The molecular weight calculated from the primary structure(30,258) is somewhat higher than that obtained for the a subunit by gel electrophoresis(4,15). Protein chemical studies of subunit a are essential for understanding this discrepancy. The protein a seems to be highly hydrophobic

Table I

Amino Acid Composition of the a and b subunits of E. coli F_0

	Amino Acid	Subunit a Predicted Residues	Subunit a Mole Percent	Subunit b Predicted Residues	Subunit b Mole Percent
	Ser	18	6.64	5	3.21
	Thr	12	4.42	3	1.92
	His	7	2.58	1	0.64
P	Gln	6	2.21	10	6.41
	Glu	8	2.95	17	10.9
	Asn	13	4.80	5	3.21
	Asp	8	2.95	9	5.77
	Lys	10	3.69	15	9.62
	Arg	5	1.85	10	6.41
	Sub-total	87	32.1	75	48.1
	Phe	20	7.38	4	2.56
	Leu	44	16.2	14	8.97
	Ile	24	8.86	11	7.05
NP	Met	11	4.06	3	1.92
+	Val	23	8.49	10	6.41
IM	Pro	14	5.17	2	1.28
	Ala	17	6.27	31	19.1
	Tyr	6	2.21	1	0.64
	Cys	1	0.37	1	0.64
	Trp	7	2.58	1	0.64
	Gly	17	6.27	3	1.92
	Sub-total	184	67.9	81	51.9
	Total	271	100	156	100

Polar amino acids (P) are according to Capaldi and Vanderkooi (13). NP, Non-polar amino acids; IM, Intermediate amino acids.

(32.1% polar residues)(Table 1), and has striking similarity to that of F_6 protein of yeast F_0(16).

An open reading frame encompassing 127 amino acids. In the region preceding the gene for subunit a, we found an open reading frame(position 59-442) encompassing 127 residues(30% polar residues). A possible initiation codon was also found at position 71. In both cases, a Shine-Dalgarno sequence(14) was found within reasonable distance from the possible initiation codon(Fig. 2). It is too early to conclude whether this frame codes for a real subunit of F_0-F_1, because active F_0-F_1 does not contain a subunit of similar molecular weight(4,15).

Within this open reading frame a promoter-like sequence was found: Pribnow box (17) starts from position 227, and "-35 region" has a similar structure to the

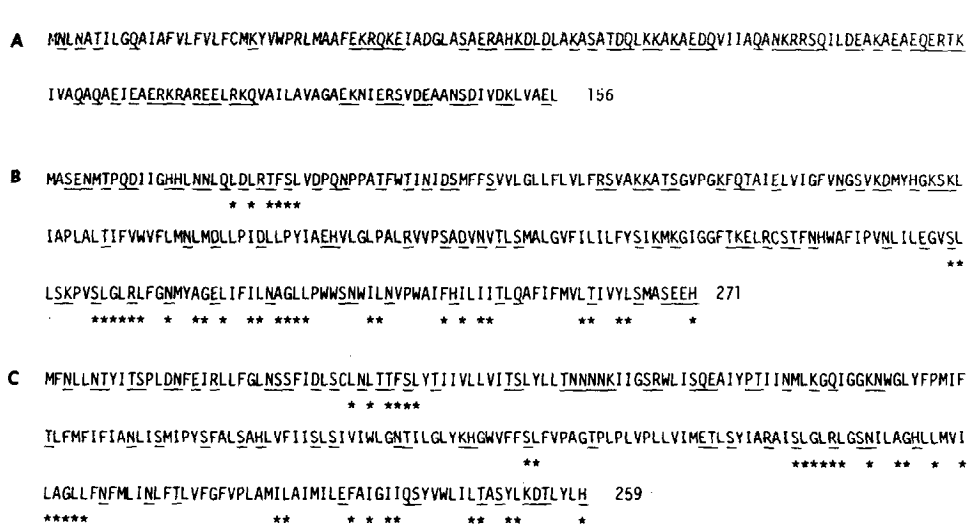

Fig. 3 Distribution of polar amino acids in the primary structures of the a and b subunits of F_0. Amino acid residues are shown by single letter symbols(7). Polar amino acids are defined according to Capaldi and Vanderkooi(13) and underlined. Stars indicate the homologous residues in the subunit a of E. coli F_0(B) and F_6 of yeast mitochondrial F_0(C)(16). Homologous residues are also observed in F_6 of human mitochondria(20), but the extent of homology is smaller(data not shown). A long hydrophobic domain appears at the carboxyl terminal region in all three cases. A, Subunit b of E. coli. B, Subunit a of E. coli. C, F_6 of yeast mitochondria(16).

typical promoter(CTG instead of TTG, position 201-203). In the present study we found no promoter sequence within 58 base pairs upstream from the putative initiation codon at position 59. Thus, the real promoter of the gene cluster is not certain at present.

Intercistronic sequence and codon usage. The intercistronic sequences between the genes for the a and c subunits and the c and b subunits are 46 and 59 base pairs, respectively. Promoter-like sequences were observed in these regions(Fig. 2). However, they were not identical to previously reported promoter sequences(17).

Usages of codons in the a and b subunits were shown along with those of highly and weakly expressed bacterial messenger RNAs(18). The codon used for the a and b subunits are similar and had a pattern intermediate between those of highly and weakly expressed type. For instance, codons used for ala and ser are similar to those in the weakly expressed type, whereas the codons for leu are similar to those in the highly expressed type. We analyzed the correlation between the codon usage and the frequency of cognate tRNA as described by Ikemura(19). The correlation

Table 2

Codon Usage in the Genes for a and b subunits of *E. coli* F_0

AA	codon	a	b	high*	low**	AA	codon	a	b	high*	low**
Arg	CGA	1	1	0	-	Gly	GGA	1	0	2	7
	CGC	1	2	17	24		GGC	6	3	32	27
	CGG	0	0	0	8		GGG	1	0	1	11
	CGT	3	7	44	18		GGT	9	0	45	19
	AGA	0	0	1	9	Val	GTA	6	2	32	11
	AGG	0	0	0	4		GTC	3	2	7	12
Leu	CTA	0	0	0	5		GTG	9	4	14	19
	CTC	1	1	2	12		GTT	5	3	34	20
	CTG	34	7	58	39	Lys	AAA	7	10	62	31
	CTT	2	5	3	15		AAG	3	5	19	15
	TTA	3	1	3	14	Asn	AAC	6	4	39	15
	TTG	4	0	3	10		AAT	7	1	3	18
Ser	TCA	3	0	1	13	Gln	CAA	2	2	8	18
	TCC	2	2	16	10		CAG	4	8	28	29
	TCG	3	1	3	14	His	CAC	4	0	8	11
	TCT	3	0	28	9		CAT	3	1	10	20
	AGC	7	2	10	9	Glu	GAA	5	13	39	33
	AGT	0	0	1	13		GAG	3	4	10	23
Thr	ACA	2	1	4	6	Asp	GAC	5	8	36	22
	ACC	3	1	22	20		GAT	3	1	16	33
	ACG	7	0	1	16	Tyr	TAC	4	1	14	12
	ACT	0	1	32	11		TAT	2	0	5	19
Pro	CCA	4	1	5	9	Cys	TGC	1	1	5	7
	CCC	1	0	1	8		TGT	0	0	1	8
	CCG	6	1	22	15	Phe	TTC	17	3	15	19
	CCT	3	0	5	6		TTT	3	1	5	29
Ala	GCA	2	6	43	21	Ile	ATA	1	0	2	8
	GCC	6	7	13	27		ATC	13	8	42	22
	GCG	5	9	21	26		ATT	10	3	17	30
	GCT	4	9	65	16	Met	ATG	11	2	19	25
						Trp	TGG	7	1	5	16

The numbers following the indicated amino acids (AA) are the number of times the codons are used in the sequence shown in Fig. 2. high* and low** indicate the highly expressed type and weakly expressed type, respectively, in the translation of bacterial messenger RNA classified by Grantham *et al*(18).

could be expressed by $y = a + bx$(the amount of tRNA, x(19); the frequency of tRNA usage, y(Table 2)). Correlation coefficients given by the analysis were 0.62 and 0.75 for a and b subunits, respectively. These values were lower than those for α(0.93) and γ(0.86)(8). The values of b are 5.40 and 6.63 for the a and b subunits, respectively. The values of a are +0.60 and -0.49 for the a and b subunits, respectively. These values of a and b compared with those for α(a=-1.15, b=8.07) and c(-0.61, 7.68)(8) suggest that the a and b subunits are more weakly translated that the α and c subunits. The values of a and b for the a subunit are close to those for γ(+0.50, 5.20), suggesting that the a subunit may be translated with

efficiency similar to that of γ. The values for the b subunit suggested that this protein may be translated more efficiently than subunit a.

Acknowledgements: We are grateful to Dr. Yamaguchi for the plasmid(pKY159-16). We thank Dr. R. H. Fillingame for the amino terminal sequence and amino acid composition of protein b. We also thank Dr. Barry P. Rosen for critical reading of the manuscript. This study was supported by grants from the Ministry of Education, Science and Culture of Japan, the Japan-U.S. Cooperative Science Program and Yamada Science Foundation.

REFERENCES

1. Downie, J.A., Gibson, F. and Cox, G.B. (1978) Ann. Rev. Biochem. 48, 103-131.
2. Futai, M. and Kanazawa, H. (1980) Curr. Top. Bioenerg. 10, 181-215.
3. Fillingame, R.H. (1981) Curr. Top. Bioenerg. 11 35-106.
4. Foster, D.L. and Fillingame, R.H. (1979) J. Biol. Chem. 254, 3230-3236.
5. Foster, D.L., Moscher, M.E., Futai, M. and Fillingame, R.H. (1980) J. Biol. Chem. 255, 12037-12041.
6. Kanazawa, H. Mabuchi, K., Kayano, T., Tamura, F. and Futai, M. (1981) Biochem. Biophys. Res. Commun. 100, 219-225.
7. Mabuchi, K., Kanazawa, H., Kayano, T. and Futai, M. (1981) Biochem. Biophys. Res. Commun., in press.
8. Kanazawa, H., Kayano, T., Mabuchi, K. and Futai, M., in press.
9. Kanazawa, H., Tamura, F., Mabuchi, K., Miki, T. and Futai, M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 7005-7009.
10. Tamura, F., Kanazawa, H., Tsuchiya, T. and Futai, M. (1981) FEBS Lett. 127, 48-52.
11. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
12. Downie, J.A., Cox, G.B., Langman, L., Ash, G., Becker, M. and Gibson, F. (1981) J. Bacteriol. 143, 8-17.
13. Capaldi, R.A. and Vanderkooi, G. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 930-932.
14. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1342-1346.
15. Friedl, P., Friedl, C. and Schairer, H.U. (1979) Eur. J. Biochem. 100, 175-180.
16. Macino, G. and Tzagoloff, A. (1980) Cell 24, 507-517.
17. Rosenberg, M. and Court, D.(1979) Ann. Rev. Genet. 13, 319-353.
18. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acid Res. 9, 43-74.
19. Ikemura, T. (1981) J. Mol. Biol. 146, 1-21.
20. Anderson, S., Bankier, I.C., Barrel, B.G., de Bruijin, M.H.L., Coulson, A.R., Broudin, J., Eperon, I.C., Nielich, D.O., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, T.G. (1981) Nature 290, 457-465.