Vol. 103, No. 2, 1981 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1981 Pages 604-612

NUCLEOTIDE SEQUENCE OF THE GENES CODING FOR α, β AND γ SUBUNITS OF THE PROTON-TRANSLOCATING ATPase of ESCHERICHIA COLI

Hiroshi Kanazawa, Toshiaki Kayano, Kazunori Mabuchi and Masamitsu Futai

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

Received September 14,1981

SUMMARY: A nucleotide sequence of 2328 base pairs comprising a portion of the gene cluster for the proton-translocating ATPase of E. coli was determined. The sequence covers most of the gene for α subunit, the entire gene for γ subunit and the amino terminal portion of the gene for β subunit, along with the flanking regions of these genes. The amino acid sequences of these subunits deduced from the DNA sequences indicate that the α and γ subunits have 513 and 287 amino acid residues, respectively. A possible secondary structure for each subunit was estimated from the inferred primary structure. The intercistronic regions between the genes for α and γ and between γ and β are 49 and 26 base pairs, respectively. The significance of codon usage in these genes is discussed in correlation with their expression.

Proton-translocating ATPase, F_0-F_1, including that of E. coli, catalyzes the synthesis of ATP utilizing the proton gradient established by the respiratory chain(1); the peripheral membrane portion, F_1, consists of 5 subunits(α,β,γ,δ and ε) and in vitro catalyzes ATP hydrolysis. The intrinsic membrane portion of the E. coli complex, F_0, consists of 3 subunits and forms a proton channel. All the structural genes for F_0-F_1 are clustered at 83 min region of E. coli linkage map (2). A restriction map of DNA segment has recently been constructed(3-6). We have recently reported the DNA sequences of the genes for the DCCD-bp(papH), δ subunit (papE) and the amino terminal portion of the gene for the α subunit(papA)(7,8).

We report here the DNA sequences of the major portion of the papA gene, the entire gene for γ subunit(papC) and the amino terminal portion of the gene for the β subunit(papB). The complete primary structure of the α and γ and partial

Abbreviations used: F_1 and F_0, periphral and integral membrane portions of proton-translocating ATPase, respectively; DCCD, N,N'-dicyclohexylcarbodiimide; DCCD-bp, DCCD-binding protein(a subunit of F_0).

0006-291X/81/220604-09$01.00/0

primary structure of the β subunit were deduced from the DNA sequence. The flanking sequences of these genes determined in this study may not be long enough to function as a regulatory site(s) in transcription, suggesting translational controls.

MATERIALS AND METHODS

Preparation of plasmid DNA and its fragments. Hybrid plasmid pMCR533(4) and pFT1501 were used in this study. pFT1501 was constructed in this study by ligating a DNA segment from λasn-5(3) into a segment of pBR533(9) digested by EcoRI and HindIII. Sequenced DNA fragments were prepared from these plasmids according to the sequencing strategy(Fig. 1). Restriction endonucleases were purchased from Takara Shuzo Co.(Japan), Boehringer Mannheim GmbH and Bethesda Research Lab.
Nucleotide sequencing. A DNA fragment recovered from a polyacrylamide gel was labeled with ^{32}P-γ-ATP by T4-polynucleotide kinase(Boehringer Mannheim GmbH) (7). The DNA sequence of the fragments was determined by the method of Maxam and Gilbert(10).
Determination of the amino acid sequence by Edman degradation. About 100μg of the γ subunit and 900μg of the β subunit purified from E. coli ML308-225(11) were subjected to manual Edman degradation(8). The phenylthiohydantoin derivatives from each cycle were analyzed by high performance liquid chromatography on a C18μ Bondapack column(Waters Associates, Inc.).

RESULTS AND DISCUSSION

DNA sequence of the gene for α subunit. Previously we determined 207 base pairs from the amino terminal end of papA(7). According to this determination and other results, we have estimated the loci of papC and papB on the physical map of the E. coli genome(Fig. 1)(8). A hybrid plasmid pMCR533 carries entire papA gene and a part of the papC gene. pFT1501 carries the rest of the papC and a part of the papB. DNA fragments were prepared from these plasmids(Fig. 1), and the 2328-nucleotide long sequence in the papA-papB region was determined(Fig. 2). An open reading frame coding for α subunit has 1539 base pairs, and the amino acid sequence of 513 residues(55,264 daltons) was deduced(Fig. 2). The amino acid composition was in good agreement with protein chemical data(Table 1)(12). While we were preparing this manuscript, Gay and Walker independently reported a DNA sequence for papA which lacked first 11 base pairs from the amino terminus(13). Our results are consistent with theirs except for the 4 amino acid residues indicated in Fig. 2.

DNA sequence of the gene for γ subunit. An open reading frame adjacent to the carboxyl terminus of papA starts 49 base pairs after the third letter of the ter-

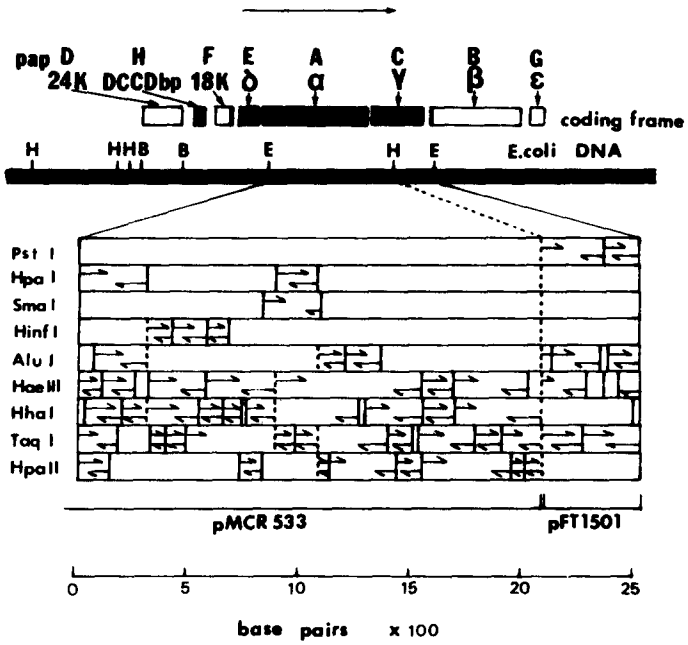

Fig. 1 Organization of the genes for F_0-F_1 and DNA sequencing strategy. The direction of transcription of the gene cluster is shown at the top of figure. The coding frame of each gene with its nomenclature(pap) and coding subunit is shown above the E. coli DNA. The coding frames shown by the solid bars indicate the region in which the DNA sequences were determined in the present or previous studies. Cleavage sites for endonucleases are as follows: E, EcoRI; H, HindIII; B, BamHI. The cleavage map with PstI, HpaI, SmaI, HinfI, AluI, HaeII, HhaI, TaqI, and HpaII are also shown. Arrows indicate the sequenced DNA segments with the direction and approximate length. Plasmids pMCR533 and pFT1501 cover the region as shown. The scale shown at the bottom corresponds to the numbers of nucleotide residues in Fig. 2. Several restriction sites in papA and papC determined previously by fragments analyses(4) were revised in the present study.

mination codon of papA. This open reading frame encodes a protein of 287 amino acid residues(31,387 daltons). This agreed well with data for the γ subunit (Table 1)(14). Residues 2-6 from the amino terminus were identical to residue 1-5 of the γ subunit determined by Edman degradation by S. D. Dunn(personal communication confirmed in this study). The predicted amino terminal methionine was presumably cleaved after synthesis. From these results we concluded that this open reading frame codes papC.

DNA sequence of a part of the gene for β subunit. An open reading frame adjacent to the carboxyl terminus of papC starts 26 base pairs after the third letter of the termination codon of papC. We determined 51 base pairs of this reading frame

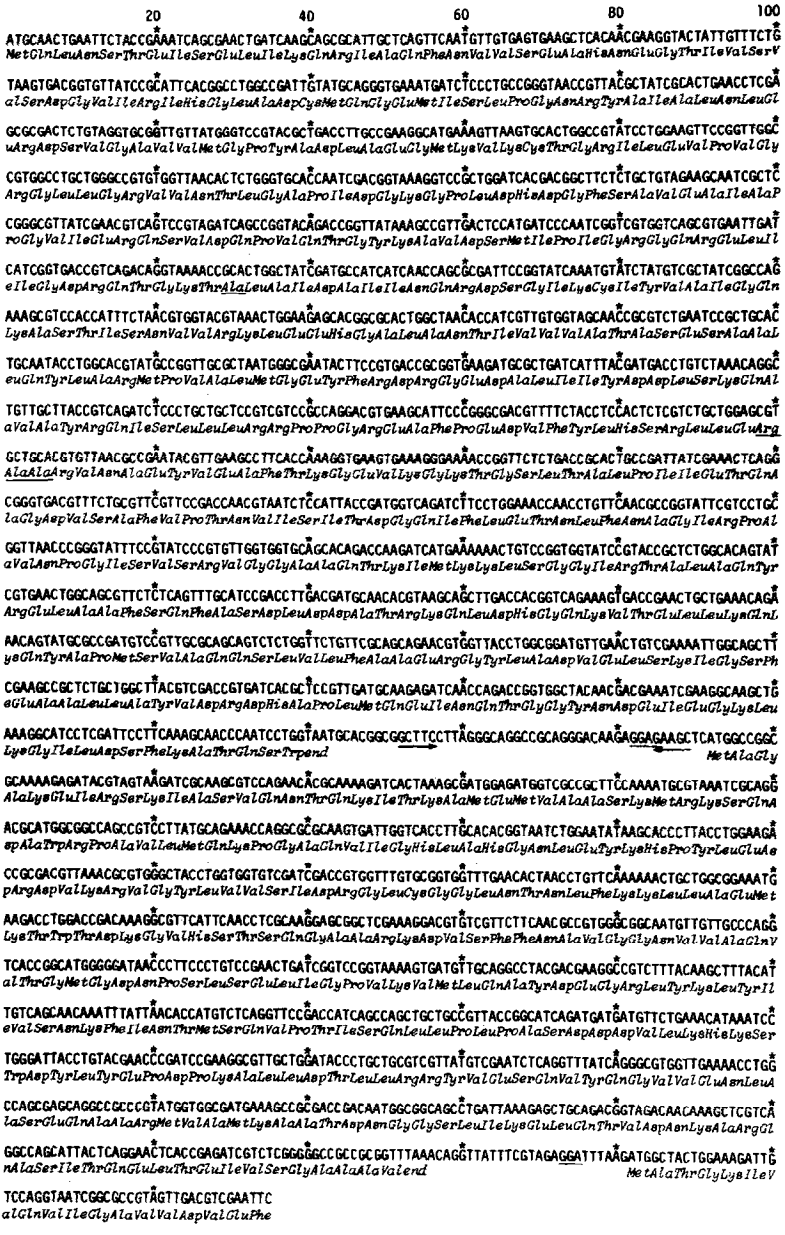

Fig. 2 DNA sequence of the genes papA, papC and papB. The DNA sequence in the antisense strand is shown with the deduced amino acid sequence. 207 base pairs from the first letter in papA was previously determined(7). Deduced amino acid residues different from those by Gay and Walker(13) and the Shine-Dalgarno sequence are underlined. An inverted repeat sequence between papA and papC is shown by arrows.

from amino terminus. Residue 2-9 from the amino terminus deduced from the DNA sequence were in complete agreement with residue 1-8 from Edman degradation.

Table I

Amino Acid Composition of the α and γ Subunits of E.coli F_1

Amino Acid	α subunit		γ subunit	
	Predicted from DNA Sequence	Reported*	Predicted from DNA sequence	Reported*
	moles/55,264		moles/31,387	
Asn + Asp	44	43.2	28	29.1
Thr	23	22.6	14	12.0
Ser	32	33.4	20	18.7
Gln + Glu	59	60.3	30	29.3
Pro	18	19.2	10	9.8
Gly	47	49.7	20	20.1
Ala	57	61.7	29	27.3
Cys	3	3.6	1	2.3
Val	41	44.1	27	27.4
Met	11	9.5	11	11.6
Ile	40	37.3	12	13.5
Leu	47	47.4	28	29.6
Tyr	15	14.7	10	10.5
Phe	14	13.6	4	4.2
His	7	6.8	5	4.4
Lys	24	23.4	23	23.0
Trp	1	1.0	3	2.1
Arg	30	28.6	12	12.9

* Reported mole per cent values for the α (12) and γ (14) were multipled by 513 residues for the α and 287 residues for the γ, respectively.

Again, the predicted amino terminal methionine was absent. This result indicates that the open reading frame codes papB.

Secondary structure of the α and γ subunits. A possible secondary structure of the α and γ subunits was estimated from the deduced primary structure by the method of Chou and Fasman(15)(Fig. 3). The contents of α helix and β sheet are 46.4% and 19.1%, respectively, of total residues in the α subunit and 49.0% and 17.8%, respectively, in γ subunit. These values are similar to average proteins(α helix, 35%; β sheet, 15%)(16), and are in sharp contrast with those of the δ subunit (α helix, 61.0%) reported previously(8).

Codon usage in the genes for F_0-F_1. Uses of codons in genes for α, γ, DCCD-bp, and δ are summarized in Table 2. Recently Grantham et al.(17) classified bacterial messenger RNAs into two groups, those highly expressed and those weakly expressed for protein synthesis. They showed that each type has a different codon usage

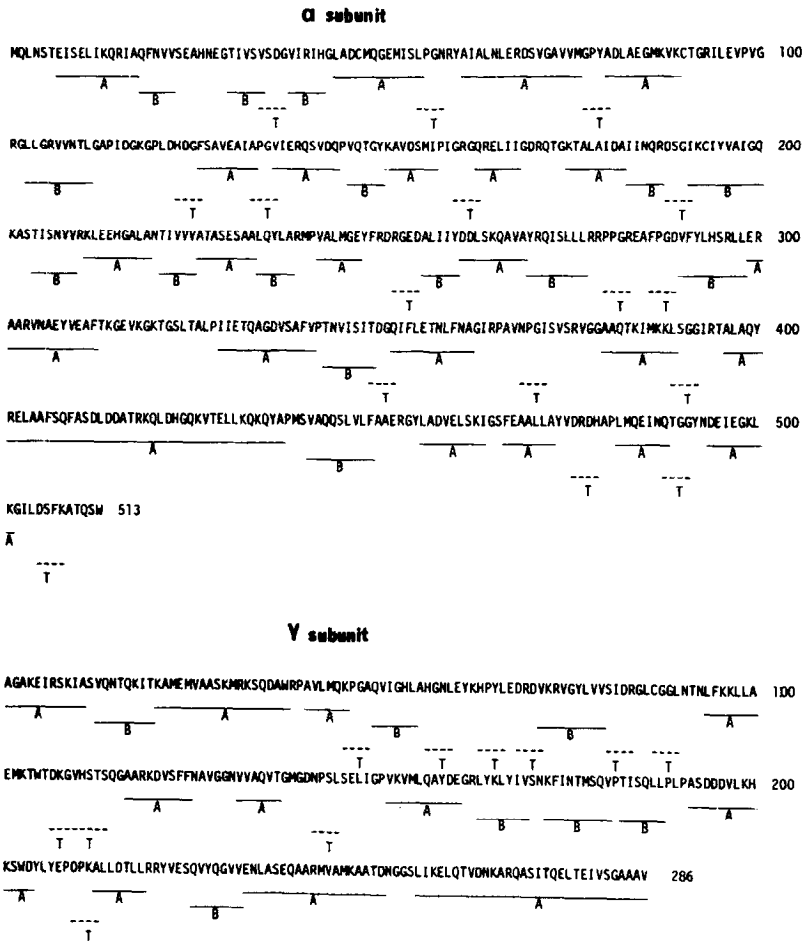

Fig. 3 Putative secondary structure of the α and γ subunits. The secondary structure of α and γ was estimated by the procedure of Chou and Fasman(15). Amino acid residues are shown by single letter symbols(16). The α helix(A), β sheet(B) and turn(T) are indicated under the amino acid sequence. Each α helical domain ends one amino acid residue before the residue known to be a helix breaker.

(cited in Table 2). In comparison with these data, we found that the usages in the papH and papA are similar to the highly expressed type and the usages in papC and papE are similar to the weakly expressed type. The typical differences are observed in case of ile, ala, ser, and leu. We also analyzed the relationship between the frequency of the codon usage in these genes and the abundance of the cognate transfer RNA according to the method of Ikemura(18)(Table 3). The analysis indicated that papA and papH have pattern similar to the highly expressed

Table 2

Codon Usage in the Genes for α, γ, δ and DCCD bp of the Proton-translocating ATPase

	DCCD bp	α	γ	δ	high*	low**		DCCD bp	α	γ	δ	high*	low**
Arg CGA	0	0	1	1	0	-	Gly GGA	0	1	1	0	2	7
CGC	0	5	2	5	17	24	GGC	2	16	11	3	32	27
CGG	0	0	1	1	0	8	GGG	1	1	2	0	1	11
CGT	2	25	8	4	44	18	GGT	7	29	6	5	45	19
AGA	0	0	0	0	1	9	Val GTA	1	9	2	5	32	11
AGG	0	0	0	0	0	4	GTC	1	2	7	4	7	12
Leu CTA	0	1	0	0	0	5	GTG	2	6	9	2	14	19
CTC	1	4	1	4	2	12	GTT	1	24	9	5	34	20
CTG	12	37	18	8	58	39	Lys AAA	1	19	16	6	62	31
CTT	0	3	4	5	3	15	AAG	0	5	7	2	19	15
TTA	0	0	1	0	3	14	Asn AAC	1	14	9	4	39	15
TTG	0	2	4	1	3	10	AAT	1	2	3	2	3	18
Ser TCA	0	0	2	1	1	13	Gln CAA	1	4	3	4	8	18
TCC	0	16	4	1	16	10	CAG	1	24	13	5	28	29
TCG	0	1	5	1	3	14	His CAC	0	7	3	2	8	11
TCT	0	11	2	5	28	9	CAT	0	0	2	0	10	20
AGC	0	2	6	1	10	9	Glu GAA	2	27	9	8	39	33
AGT	0	2	1	3	1	13	GAG	0	4	5	7	10	23
Thr ACA	0	2	0	0	4	6	Asp GAC	0	16	9	5	36	22
ACC	0	17	9	2	22	20	GAT	3	12	7	6	16	33
ACG	0	0	2	2	1	16	Tyr TAC	2	11	7	1	14	12
ACT	1	4	3	0	32	11	TAT	0	4	3	0	5	19
Pro CCA	0	3	2	1	5	9	Cys TGC	0	0	1	1	5	7
CCC	0	0	1	1	1	8	TGT	0	3	0	1	1	8
CCG	1	14	5	1	22	15	Phe TTC	3	13	3	0	15	19
CCT	3	1	2	0	5	6	TTT	1	1	1	6	5	29
Ala GCA	1	20	5	6	43	21	Ile ATA	0	0	1	0	2	8
GCC	1	8	13	7	13	27	ATC	7	30	6	4	42	22
GCG	5	12	8	7	21	26	ATT	1	10	5	6	17	30
GCT	5	17	3	7	65	16	Met ATG	8	11	11	7	19	25
							Trp TGG	0	1	3	1	5	16

The numbers following the codons for the indicated amino acids are the number of times the codons are used in the sequence shown in Fig. 2 and reported previously(7,8). High* and low* indicate codon usages of highly and weakly expressed types, respectively, of bacterial genes reported by Grantham et al.(17). For details, see the text and the reference(17).

type, while the papC and papE are similar to weakly expressed type. This analysis suggested how a single polycistronic message can be translated into a variable number of subunits, although such translational control may not be sufficient to account for the exact stoichiometry of α:γ:δ:DCCD-bp=3:1:1:6(1).

Flanking sequences and gene expression. No usual promotor-like sequence(19) was observed in the non-coding regions between papA and papC and between papC and papB, suggesting that synthesis of the γ and β subunits may not be regulated transcriptionally. However, the intercistronic sequence between papA and papC(49 base pairs) is longer than those between papF and papE(14 base pairs), papE and papA

Table 3

Relationship between the Codon Usage and Abundance of Cognate Transfer RNA in the Genes for α, γ, δ and DCCD-bp: Analysis with Linear Regression.

Gene	Correlation coefficient	Slope	y Intercept
papH (DCCD-bp)	0.57	7.68	-0.61
papA (α)	0.93	8.07	-1.51
papC (γ)	0.86	5.20	+0.50
papE (δ)	0.72	5.08	+0.38
high*	0.95	8.10	-1.26
low**	0.81	3.34	+1.43

The frequency of the codon usage in the genes for α, γ, δ and DCCD-bp along with the weakly expressed type(low*) and highly expressed type(high**)(Table 2) and the transfer RNA abundance were plotted according to the procedure of Ikemura(18). The relationship of frequency and abundance could be expressed by $y = a + bx$(the amount of transfer RNA=x, the frequency of the RNA usage=y; x values are from reference(18) and y values are given from Table 2). The values a and b correspond to the y intercept and slope, respectively. Large value for the slope and negative value for the y intercept indicate that a gene uses an abundant type RNA more frequently. The correlation coefficient for papH is lower than that for other genes consistent with the large number of DCCD-bp's found in F_0-F_1 relative to other subunits.

(12 base pairs)(8) and papC and papB(26 base pairs), suggesting that unknown regulatory site(s) may exist between papA and papC. In this connection, it should be noted that an inverted repeat sequence exist in this region(Fig. 2).

Acknowledgements: We thank Dr. S. D. Dunn for allowing us to cite the unpublished results. We also thank Dr. Barry P. Rosen for critical reading of the manuscript and Ms. K. Takeda for technical assistance. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan, Yamada Science Foundation and the Japan-U.S.A. Cooperative Program.

REFERENCES

1. Futai, M., and Kanazawa, H. (1980) Curr. Top. Bioenerg. 10, 181-215.
2. Downie, J.A., Gibson, F., Cox, G.B. (1979) Ann. Rev. Biochem. 48, 103-131.
3. Kanazawa, H., Miki, T., Tamura, F., Yura, T., and Futai, M. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 1126-1130.
4. Kanazawa, H., Tamura, F., Mabuchi, K., Miki, T., and Futai, M. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 7005-7009.
5. Tamura, F., Kanazawa, H., Tsuchiya, T., and Futai, M. (1981) FEBS Lett. 127, 48-52.
6. Downie, J.A., Cox, G.B., Langman, L., Ash, G., Becker, M., and Gibson, F. (1981) J. Bacteriol. 145, 200-210.
7. Kanazawa, H., Mabuchi, K., Kayano, T., Tamura, F., and Futai, M. (1981) Biochem. Biophys. Res. Commun. 100, 219-225.
8. Mabuchi, K., Kanazawa, H., Kayano, T. and Futai, M. (1981) Biochem. Biophys. Res. Commun. 102, 172-179

9. Sutcliffe, J.G. (1979) Cold Spring Harbor Symposia on Quantitative Biology 43, 77-90.
10. Maxam, A.W., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 560-564.
11. Dunn, S.D., and Futai, M. (1980) J. Biol. Chem. 255, 113-118.
12. Dunn, S.D. (1980) J. Biol. Chem. 255, 11857-11860.
13. Gay, N.J., and Walker, J.E. (1981) Nucleic Acid Res. 9, 2187-2194.
14. Dunn, S.D. personal communication.
15. Chou, P.Y., and Fasman, G.D. (1978) Advances in Enzymol. 47, 45-148.
16. Shultz, G.E., and Shirmer, R.H. (1979) "Principles of Protein Structure" Springer-Verlag, Heidelberg.
17. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., and Mercier, R. (1981) Nucleic Acid Res. 9, 43-74.
18. Ikemura, T. (1981) J. Mol. Biol. 146, 1-21.
19. Rosenberg, M., and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
20. Shine, J., and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1342-1346.