Gene Structure of the Human Mitochondrial Adenosine Triphosphate Synthase β Subunit*

(Received for publication, November 13, 1987)

Shigeo Ohta*, Hideaki Tomura, Kakuko Matsuda, and Yasuo Kagawa
From the Department of Biochemistry, Jichi Medical School, Minamikawachi-machi, Tochigi-ken, 329-04 Japan

The mitochondrial ATP synthase β subunit is encoded by a nuclear gene and assembled with the other subunits encoded by both mitochondrial and nuclear genes. As the next step in the analysis of the molecular mechanisms coordinating the two genetic systems, the gene for the human β subunit was cloned, and its structure was determined. The gene contains 10 exons, with the first exon corresponding to the noncoding region and most of the pressequence which targets this protein to the mitochondria. Eight Alu repeating sequences including inverted repeats were found in the 5′ upstream region and introns. An S1 nuclease protection experiment revealed two initiation sites for the transcription. A typical TATA box was not present at about 30 base pairs upstream from either initiation site. Three CAT boxes (CCCAAT) were found between the two initiation sites. In addition, one CAT box was found at 41 base pairs upstream from the first initiation site. Two GC boxes (potential Sp1 binding sites) were located in the 5′ upstream region, one of them linked to Alu repeating sequences. For determination of the promoter activity, fragments of various length from the 5′ upstream region were fused to a chloramphenicol acetyltransferase gene and transfected into cultured cells. This experiment showed the existence of an enhancing structure(s) for transcription between nucleotide −400 and −1100 in the upstream region.

**Mitochondrial ATP synthase (F₀F₁) catalyzes ATP formation, using the energy of proton flux through the inner membrane during oxidative phosphorylation (1, 2). Two subunits of mammalian F₀F₁ are encoded by a mitochondrial gene (3) and the other subunits (7–12 subunits) by a nuclear gene (4). The β subunit is encoded by the nuclear genome, synthesized in the cytosol, imported into mitochondria, and then assembled with the other subunits (5). The numbers of mitochondria per cell vary greatly depending on the developmental stage, cell activity, and type of tissue (6, 7). These facts suggest a functional interaction between the two genetic systems. However, the molecular mechanism for coordinating the two genetic systems is unknown. To understand the molecular basis of this coordination requires an analysis of the regulatory system for the subunit encoded on the nuclear genome. As the first step, we have cloned cDNA of the human β subunit (8).

All subunit structures of F₀F₁, reported to date, for either prokaryotes or eukaryotes, are very similar (1, 2). In particular, the primary structure of the β subunit is highly conserved between various species. In addition, it contains sequences very similar to those in parts of some nucleotide binding proteins, such as recA protein, adenylate kinase, and the ras gene product (2). Thus, it may be possible to correlate function of each region of the human β subunit with exon structure.

Here, we report the organization of the gene, some structures involved in its expression, and the existence of an enhancing structure for transcription in the 5′ upstream region.

**EXPERIMENTAL PROCEDURES**

**Gene Cloning and DNA Sequencing**—Genomic DNA encoding the β subunit was isolated from a human genomic library, using the β subunit cDNA (8) (EcoRI fragment; 1.8 kilobase pairs) as a hybridization probe. The library was a gift from Dr. Nojima (Department of Pharmacology, Jichi Medical School). Human genomic DNA was partially digested with HaeIII and AluI and ligated to the EcoRI sites of λ phage DNA, Charon 4A, with EcoRI linkers. Phage DNA was purified from positive plaques as described (9). DNA fragments were subcloned into plasmid pUC18 or pUC19 (Fig. 1). The nucleotide sequence was determined by the dyeoxy nucleotide chain termination method using M13 phage single-stranded DNA as template (10, 11). Most of the sequence was determined by the shot-gun method using sonication for shearing the fragments (12).

The DNA sequences were assembled and analyzed by GENETYX program (SDC Inc., Tokyo).

**S1 Nuclease Mapping**—S1 nuclease mapping was performed as described (13). The EcoRI-BamHI fragment (2.2 kilobase pairs) was subcloned into the plasmid pTZ18R (purchased from Pharmacia LKB Biotechnology Inc.). The HindIII fragment (nucleotide −416 to +273, Fig. 2; 689 bp) was isolated by agarose electrophoresis and treated with calf intestinal phosphatase. The fragment was labeled at the 5′ end with T4 polynucleotide kinase and [γ-32P]ATP. The labeled fragment was digested with Drai (at nucleotide −355). The fragments (61 and 618 bp) labeled at their 5′ ends were hybridized with poly(A)* RNA and digested with S1 nuclease. Poly(A)* RNA was prepared from HeLa cells (14).

**CAT Assay**—An EcoRI-BamHI (nucleotide +452, Fig. 2) fragment containing the 5′ upstream region, the first exon, and part of the first intron of the β subunit gene was subcloned into plasmid pTZ18R. The plasmid was digested with EcoRI or Drai (nucleotide −355), shortened by treatment with Bal31 exonuclease, and then digested with BamHI. The ends of the digested fragments were filled in with T4-polymerase, ligated with HindIII linkers (dCAAGCTTG), digested with HindIII, and then fractionated by agarose electrophoresis. The isolated fragments were ligated to HindIII-digested pMLCAT (15, 16). pSV2CAT was used as a positive control (17). The plasmids were isolated by the alkaline sodium dodecyl sulfate method (9), digested with DNase-free RNase A, and purified by CaCl gradient centrifugation (9) followed by Sephacryl S 2000 column chromatography (Pharmacia LKB Biotechnology Inc.). These plasmids were RNA-free. The orientations and lengths of the inserts were deter-

---

* This research was supported by Grant 262617006 from the Ministry of Education, Science, and Culture of Japan. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03906.

† To whom correspondence should be addressed.

1 The abbreviation used is: bp, base pair(s).
FIG. 1. A schematic illustration of the restriction map and the sequencing strategy. The exons are shown by boxes, and the encoding regions are shown by closed boxes. The fragments were subcloned as shown by long arrows. The small fragments were determined by using synthetic oligonucleotide as a primer, respectively. The locations of the Alu repetitive sequences in the human F1-ATPase subunit are also shown by arrows. The direction and extent of sequence determinations are indicated by short arrows. →, •→, and ○→ indicate the nucleotide sequence obtained by sonication, with restriction enzymes, or by sequencing. Some parts were determined by using synthetic oligonucleotide (17 bases in length; synthesized with an automatic DNA synthesizer, Applied Biosystems Model 380B, Foster City, CA) as primers. The positive plaques were purified by successive rounds of plaque hybridization. The phage DNAs were isolated and subjected to "Southern" analysis using 5' (EcoRI-AatI1 fragment; 280 bp) and 3' (SmaI-EcoRI fragment; 866 bp) of the cloned cDNA (8) as hybridization probes. Three kinds of clones were isolated. DNA from the first clone hybridized with 5' and 3' regions of the cDNA. The second clone appeared to contain an allelic gene of the first one, because all but the KpnI fragments were identical. DNA in the third clone was shorter than that in the

mined by electrophoresis after digestion with selected restriction enzymes. Plasmid DNA (10 μg) was transfected into HeLa or mouse A9 cells (2 × 106 cells in 9-cm dishes) (14, 18). After 2 days, the cells were lysed, and the lysates (2 mg protein/ml) were incubated with acetyl-CoA and [14C]chloramphenicol for 1 h at 37 °C (14). The acetylated chloramphenicol was separated by thin layer chromatography on Merck high performance thin layer chromatography precoated plates of Silica Gel 60 with a concentrating zone. For quantitative analysis, the silica gel around the radioactive spot was scraped off and assayed for radioactivity in a scintillation spectrometer.

RESULTS

Molecular Cloning—A human genomic gene library constructed in Charon 4A was screened with 32P-labeled human F1 β subunit cDNA (8). Out of a total of 1,500,000 plaques, we identified five positives. The positive plaques were purified by electrophoresis after digestion with selected restriction enzymes. The phage DNAs were isolated and subjected to "Southern" analysis using 5' (EcoRI-AatI1 fragment; 280 bp) and 3' (SmaI-EcoRI fragment; 866 bp) of the cloned cDNA (8) as hybridization probes. Three kinds of clones were isolated. DNA from the first clone hybridized with 5' and 3' regions of the cDNA. The second clone appeared to contain an allelic gene of the first one, because all but the KpnI fragments were identical. DNA in the third clone was shorter than that in the

FIG. 2. Nucleotide sequence of the gene for the human ATP synthase β subunit and the deduced amino acid sequence of the exons. The sequences described in the text are noted with underlines or boxes. The nucleotides are numbered from the upstream initiation site for transcription. Since part of the gene was not sequenced, numbers are not accurate from the 5698th nucleotide. The undetermined segment was judged to be 2.0 kilobases in length based on an electrophoretic analysis.
others, so this gene may be a pseudo-gene. Therefore, the structure of the DNA in the first clone was determined in detail.

Organization of the human ATPase β subunit gene is shown in Fig. 1, and most of the nucleotide sequence is shown in Fig. 2. The intron/exon junctions were determined by direct comparison of the nucleotide sequences with the cloned cDNA. The sequence at the 5' end of the cDNA was not found in the gene. In addition, the sequence at the point of divergence of the cDNA and genomic sequences did not agree with the consensus sequence (AG/GT rule) at intron/exon junctions (19). An S1 mapping experiment showed that mRNA from HeLa cells hybridized with the fragment from the genomic gene (Fig. 4). Therefore, we conclude that 18 nucleotides encoding the presequence were rearranged, probably during construction of the cDNA library. The presequence should be MLGFVGRVAAAPASGALRRLTPSASLPPAQLLLRAVRRRSHPVREYAAQ instead of PSASLPPAQLLLRAVRRRSHPVREYAAQ as reported by Tomura et al. (20).

Structure Related to Transcription—Three CAT boxes (CCAAT) (19, 23, 24) were found between the first and second initiation sites, one at -41. Two GC boxes (CCGCC) (23, 25) were found, one of them linked to Alu repeating sequences. Various lengths of the 5' flanking region of the β gene were inserted into the DNA of a plasmid which harbors the gene for chloramphenicol acetyltransferase. These plasmids were transfected into HeLa cells or A9 cells by calcium phosphate co-precipitation (14, 17).

The proximal 200-bp segment of 5' flanking DNA region was sufficient to elicit measurable levels of expression (pHB3CAT). The fragment between 400 and 1100 bp (pHBICAT) enhanced the expression 9-fold over the level obtained with pHB3CAT (Fig. 5). Similar results were obtained on transfection into mouse A9 cells (data not shown). These results suggest the existence of an enhancing structure upstream from about −400.

FIG. 2—continued

FIG. 3. Comparison of the human gene and tobacco (20) genes. A, the length of the exons and the introns are illustrated. The intron/exon junctions are connected with their counterparts with lines. Only the encoding regions are shown by boxes. B, the deduced amino acid sequences. The identical amino acid residues are boxed. The intron points are shown by arrows.
**DISCUSSION**

Organization of the Gene—The $F_oF_1$ complex is widely distributed in most organisms. The $\beta$ subunit is a catalytic center of $F_oF_1$, and its structure is highly conserved (2). Thus, the organization of the gene should provide evidence of evolutionary relationships.

Boutry and Chua (20) described organization of the gene for the $\beta$ subunit of Nicotiana plumbaginifolia. The length and number of introns and exons differ from those of the human gene. Four intron/exon junctions of the human gene were close to their counterparts in tobacco, but shifted three or four codons in a 3' direction. The shifting may be explained by mutations that block a splice junction, unmasking a cryptic splice site that adds or deletes small amounts of protein (26). However, it is difficult to explain why the splice junctions were shifted in the same direction in each case.

Most mitochondrial proteins encoded in the nuclear genome have transient presequences (5). These peptides, which direct proteins into the mitochondria, are sometimes embedded in a cytosolic protein (27). Since such targeting sequences are often found even in Escherichia coli, (28), they may have been fused to an existing protein during evolution making it into a mitochondrial protein. According to the "exon-shuffling" model (29), the addition of such an additional function should be provided in a separate exon. In fact, the first exon in the gene for the human ATP synthase $\beta$ subunit corresponds to the presequence. According to the "module hypothesis" by Go (30), an exon is a structural unit which makes a compact peptide. This suggests the compact structure of the presequence peptide.

Nucleotide binding proteins have regions of sequence similarity which constitute nucleotide binding sites. In the $\beta$ subunit, such regions are found in exons 5 and 6 (8). The Alu repeating sequences are located in introns 3 and 7. Thus, the exons that code for the nucleotide binding sites may have been rearranged at the sites of the repetitive sequences during evolution. Our finding of a pseudo-gene lacking exons 4-7 is consistent with this hypothesis. The cluster of acidic and basic amino acid residues in the $\beta$ subunit is proposed to

---

Footnote: H. Tomura and S. Ohta, unpublished observation.
include the conformational change by a proton flux (31). The region for the acid-base cluster was found in exon 9, also located between two Alu sequences.

Expression of the Gene—ATP synthase and some enzymes related in the respiratory chain are complexes of the products of both nuclear and mitochondrial genes (4, 32). Mitochondria vary in number and morphology depending on intercellular and extracellular conditions. However, the fundamental molecular mechanism for the coordination of the two genetic systems is unknown. Mammalian cells in culture may be useful for experimental systems to study this problem for several reasons. Mammalian mitochondrial DNA has been studied extensively (33–35) as the yeast one. Mammalian DNA fragments can be transfected into cultured cells and promoter activity easily estimated (18). Several transcriptional factors have been purified from mammalian cells. In addition, it may be possible to study nuclear-mitochondrial interactions using some abnormal mitochondria from human patients with mitochondrial myopathy (36). Therefore, we used a human cell line for our analysis.

S1 mapping revealed two initiation sites for the transcription. The role of the two start sites in the regulation of transcription is unknown. Three CAT boxes (CCAT) are located between the first and second initiation sites. If a transcriptional factor is bound to one of those CAT boxes, the sequence (GT)4 is widely distributed in eukaryotic genes (21). The relations of the left-handed helix of DNA with the potential effect on transcription were discussed (21). The role of the alternating GT structure in the expression of the subunit gene is under investigation.

Three inverted Alu sequences were found in the 5′ upstream region, one of them linked with a GC box. The role of the repetitive sequence in regulating gene expression is not clear, but the CAT assay showed that the fragment (400 to −1100) containing the Alu repeats exhibited enhanced transcription. Further analysis is required to determine the roles of these structures. In any event, our data demonstrate that full expression of the gene required more than 400 bp of upstream sequence in the gene of a "house-keeping" enzyme. It will be important to isolate the trans-acting transcriptional factors involved. Analysis of the gene structure may allow us to elucidate the molecular basis of the regulation and of coordination of the synthesis of the ATP synthase complex in cultured cell systems.

Acknowledgments—We thank Dr. H. Nojima (Jichi Medical School, Department of Pharmacology) for providing the human genomic library and for valuable discussions. We thank Drs. A. Fujisawa-Sehara and Y. Fuji-Kuriyama (Cancer Institute, Japanese Foundation for Cancer Research) for gifts of pML-CAT and pSV2-CAT.

References

1. Kagawa, Y. (1984) in Bioenergetics (Ernster, L., ed) pp. 149–186, Elsevier Scientific Publishing Co., Inc., Amsterdam
2. Amzel, L. M., and Pedersen, P. L. (1983) Annu. Rev. Biochem. 52, 801–824
3. Fearnley, I. M., and Walker, J. E. (1986) EMBO J. 5, 2003–2008
4. Haceci, Y. (1985) Annu. Rev. Biochem. 54, 1015–1069
5. Schatz, G., and Butow, R. A. (1983) Cell 32, 316–318
6. Williams, R. S. (1986) J. Biol. Chem. 261, 12398–12394
7. Williams, R. S., Garcia-Moll, M., Mellor, J., Salmon, S., and Harlan, W. (1987) J. Biol. Chem. 262, 2764–2767
8. Ohu, S., and Kagawa, Y. (1986) J. Biochem. (Tokyo) 99, 135–141
9. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
10. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
11. Messing, J., Crea, R., and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309–321
12. Bankier, A. T., and Barrell, B. G. (1983) in Nucleic Acid Biochemistry, B508 ‘Flavell, R. A., ed.’ pp. 1–34. Elsevier Scientific Publishing Co., Inc., Amsterdam
13. Berk, A., and Sharp, P. A. (1977) Cell 12, 721–732
14. Davis, L. G., Dibner, M. D., and Battey, J. F. (1986) Methods in Molecular Biology, Elsevier Scientific Publishing Co., Inc., Amsterdam
15. Yusko, L., and Botchan, M. (1981) Nature 293, 79–81
16. Fujisawa-Sehara, A., Sogawa, K., Nishii, C., and Fuji-Kuriyama, Y. (1986) Nucleic Acids Res. 14, 1465–1477
17. Laimins, L., Khoury, G., Gorman, C., Howard, B. H., and Gruss, P. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 6453–6457
18. Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982) Mol. Cell. Biol. 2, 1044–1051
19. Breathnach, R., and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349–383
20. Boutry, M., and Chua, N-H. (1985) EMBO J. 4, 2159–2165
21. Rich, A., Nordheim, A., and Wang, A. H-J. (1984) Annu. Rev. Biochem. 53, 791–846
22. Schmid, C. W., and Jelink, W. R. (1982) Science 216, 1065–1070
23. Dynan, W. S., and Tjian, R. (1986) Nature 316, 774–778
24. Kelly, T. J., Jones, K. A., Kodonaga, J. T., Rosenfeld, P. J., and Tjian, R. (1987) Cell 48, 79–89
25. Gidoni, D., Dynan, W. S., and Tjian, R. (1984) Nature 312, 409–413
26. Gilbert, W., Marchionni, M., and McKnight, G. (1986) Cell 46, 151–154
27. Hurt, E. C., and Schatz, G. (1987) Nature 325, 499–503
28. Baker, A., and Schatz, G. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 3117–3121
29. Gilbert, W. (1978) Nature 271, 501
30. Go, M. (1981) Nature 289, 80–92
31. Kagawa, Y. (1984) J. Biochem. (Tokyo) 95, 295–298
32. Schatz, G., and Mason, T. L. (1974) Annu. Rev. Biochem. 43, 51–87
33. Anderson, S., Bankier, A. T., 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., and Young, I. G. (1981) Nature 290, 457–465
34. Clayton, D. A. (1982) Cell 28, 693–705
35. Clayton, D. A. (1984) Annu. Rev. Biochem. 53, 573–594
36. DiMauro, S., Bonilla, E., Zeviani, M., Nakagawa, M., and DeVivo, D. C. (1985) Annu. Neurol. 17, 521–538