Vol. 161, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
May 30, 1989 Pages 371-378

ISOLATION OF A SINGLE NUCLEAR GENE ENCODING HUMAN UBIQUINONE-BINDING PROTEIN IN COMPLEX III OF MITOCHONDRIAL RESPIRATORY CHAIN

Hiroshi Suzuki, Yoshitaka Hosokawa, Haruo Toda,
Morimitsu Nishikimi and Takayuki Ozawa

Department of Biomedical Chemistry, Faculty of Medicine,
University of Nagoya, Nagoya 466, Japan

Received April 11, 1989

Summary: We have isolated five genomic DNA clones which contain nucleotide sequences hybridizable to a cDNA for human ubiquinone-binding protein in Complex III (QP). Nucleotide sequence analysis revealed that two of them contained different types of pseudogenes suggesting molecular evolution of the gene, and that the other three clones contained the overlapping fragments from the same QP gene. The gene spans 4.5 to 5 kb in length. The sequences of exons in the gene were determined and found to be identical to the corresponding parts of the human QP cDNA. The exon-intron boundaries follow the GT/AG rule. Two CAAT boxes were found in the promoter region. It is concluded from these results that the isolated human QP gene is functional. Genomic Southern blot analysis showed that the gene is present in a single copy in the human genome.

The mitochondrial oxidative phosphorylation system in mammals consists of five multisubunit complexes. Since all the complexes but Complex II are composed of subunits encoded in both the mitochondrial and the nuclear genomes, there should be a continuously operating regulatory system that coordinates the expression of mitochondrial and nuclear genes under various conditions such as proliferation (1), differentiation (1), and malignant transformation (2) of the cells, and mitochondrial DNA mutations in mitochondrial cytopathies (3, 4). Most of the genetic information required for mitochondrial biogenesis is provided by the nuclear genome. Therefore, genetic and biochemical analyses of nuclear genes encoding mitochondrial proteins are essential to an understanding of the biogenesis and the informational interactions between physically separated nuclear and mitochondrial genomes. We have focused on the genetic information residing in the flanking regions of nuclear genes encoding mitochondrial respiratory components. We have recently isolated the human cytochrome c_1 gene (5) and characterized its 5' flanking region, and found two common sequence elements in the region in comparison with the 5' flanking regions of genes for human F_0F_1-ATPase β subunit (6) and chicken 5-aminolevulinate synthase (7). It is very interesting to examine whether the same sequence elements occur in

0006-291X/89 $1.50

nuclear genes for the other mammalian mitochondrial respiratory components and participate in the cooperative expression of many nuclear genes responsible for the respiration. Indeed, in yeast, two consensus sequence elements participating in the cooperative expression of nuclear genes for different subunits of Complex III have been reported (8). We have isolated a cDNA for human ubiquinone-binding protein of Complex III (QP) (9) and, in this study, isolated several genomic fragments which are hybridizable to the cDNA. Here we report the sequence of a single gene encoding human QP.

Materials and Methods

Materials: Enzymes and chemicals were purchased from the following sources: restriction enzymes and T4 DNA ligase from Toyobo Co., Ltd. (Osaka, Japan); Klenow fragment of DNA polymerase I, and a 7-deaza-sequencing kit from Takara Shuzo Co., Ltd. (Kyoto, Japan); calf intestinal alkaline phosphatase from Boehinger-Mannheim; and a multiprime DNA labeling system and [α-^{32}P]dCTP (3000 Ci/mmol) from Amersham International plc.. A human leukocyte EMBL3 genomic library was obtained from Clontech Laboratories, Inc. and GeneScreen *Plus* nylon membranes from Du Pont Company.

Isolation of phage clones containing the human QP gene: A genomic DNA library of about 1.6 x 10^6 EMBL3 phages containing a partial *Mbo*I digest of human leukocyte DNA was screened by plaque hybridization as previously described (5) using a human QP cDNA (9) as a probe.

Restriction endonuclease mapping and Southern blotting: The genomic DNA inserts in cloned phages were excised by digestion with *Sal*I for subsequent restriction mapping and subcloning. Southern blot analysis was performed as described (5). DNA fragments were labeled with [α-^{32}P]dCTP by multiple priming and used as hybridization probes.

DNA sequencing analysis: Genomic subfragments were subcloned into pUC18 plasmid. DNA sequencing was carried out by the dideoxy chain termination method (10) using 2'-deoxy-7-deazaguanosine triphosphate instead of dGTP for better resolution (11).

Computer analysis: Computer analysis was performed with a GENETYX program (SDC, Tokyo).

Results and Discussion

Molecular cloning: We previously obtained a human QP cDNA (9). Using the cDNA as a hybridization probe, a human leukocyte EMBL3 genomic library containing 1.6 x 10^6 phages was screened and 28 positive clones were isolated at high stringency. These clones were divided into five groups, designated HQP I to V, based on restriction maps (Fig. 1). Southern blot analysis showed that all but HQP III contained fragments hybridizable to both 5'- and 3'-specific probes of the cDNA, while HQP III contained a fragment hybridizable only to the 5'-specific probe.

Preliminary nucleotide sequence analysis: We previously isolated 17 clones for a human QP cDNA and reported the nucleotide sequence of one (clone #15) of them (9). Two other clones (#2 and #6) were further sequenced. As shown in Fig. 2, the nucleotide sequences were identical to the reported one (9) except for a two-base difference (CC and GG) at nucleotide positions -31 and

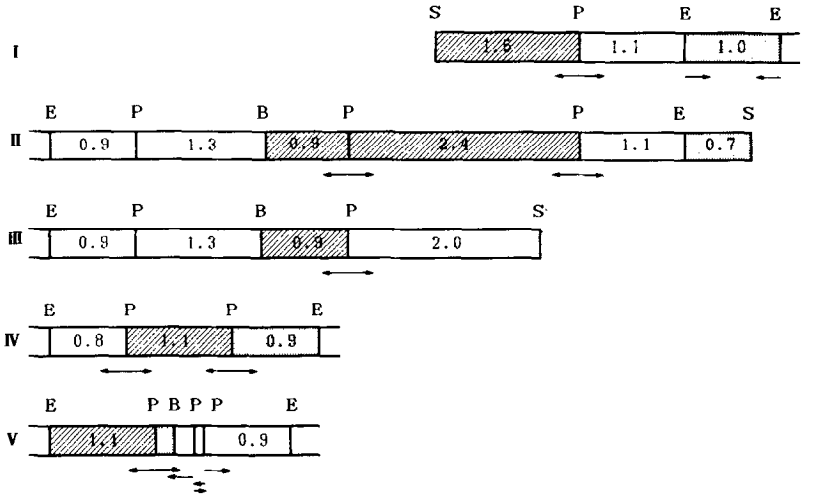
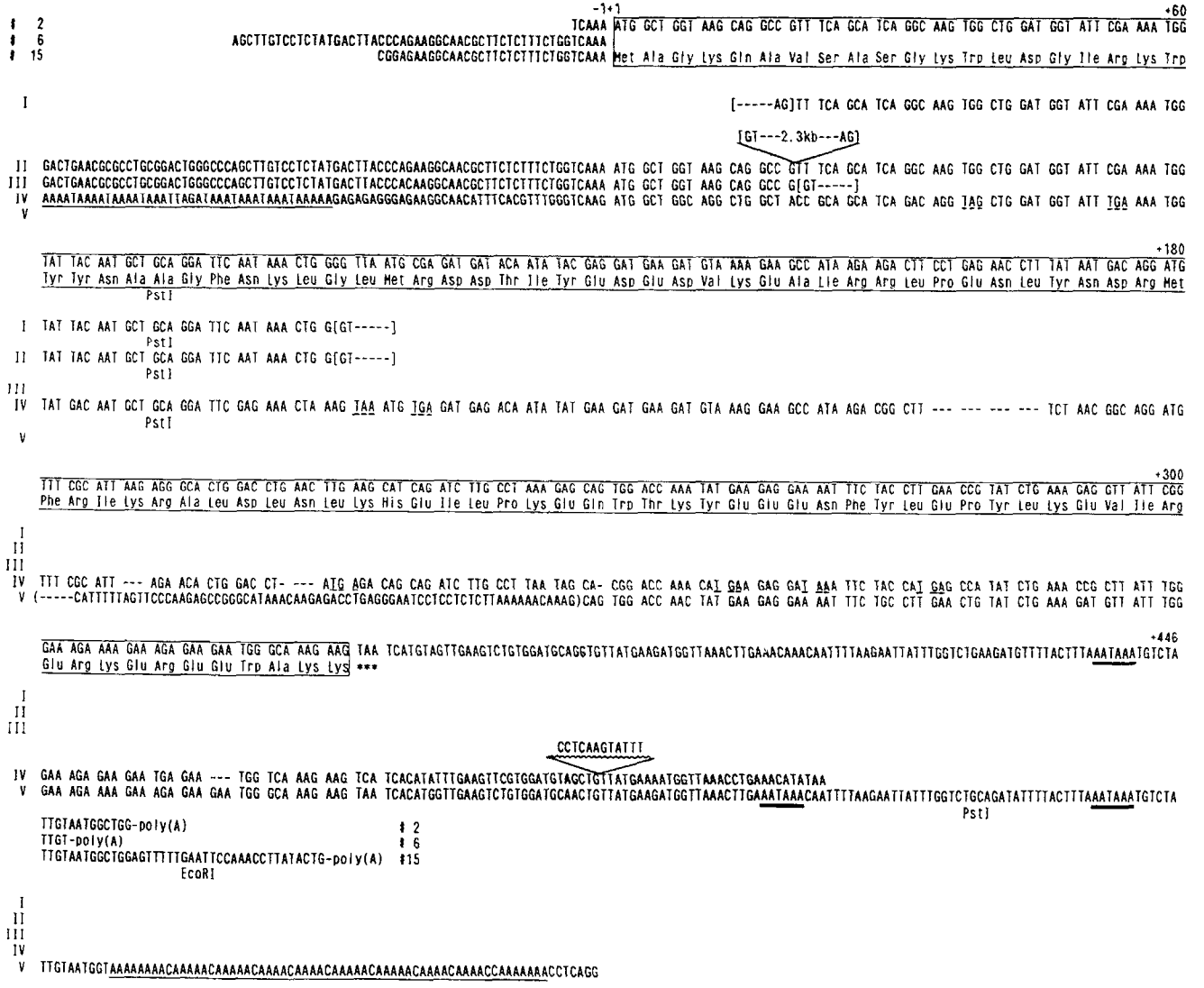


Fig. 1. Restriction maps and sequence strategies of genomic fragments containing the human QP genes (HQP I to III) and pseudogenes (HQP IV and V). Hatched and stippled boxes represent the fragments which are hybridizable to 5'- and 3'-specific probes as specified blow, respectively. The extent and direction of sequence determinations are indicated by the horizontal arrows. Figures in boxes indicate the length of the fragments in kb. S, P, E, and B are *Sal*I, *Pst*I, *Eco*RI, and *Bam*HI recognition sites, respectively. The 120-bp *Pst*I fragment including a part of pcD vector and the 390-bp *Pst*I-*Eco*RI fragment of the human QP cDNA (clone #15) were used as 5'- and 3'-specific probes, respectively (see Fig. 2).

-30 in clones #6 and #15 and for different polyadenylation sites in all the three clones. The CC in clone #6 was found at the corresponding position in the sequence of the functional gene (HQP II and III), suggesting that the GG in clone #15 possibly resulted from cloning artifact.

As shown in Fig. 2, the human QP cDNAs have only one *Pst*I recognition site in the 5' side of the coding region. We sequenced the cloned genomic DNAs from the *Pst*I site to both sides to find the promoter region of the gene. HQP IV and V were found to contain different types of psuedogenes. HQP IV lacked an intron, and had an abnormal adenylic acid-rich sequence in the 5' untranslated region, and eight stop codons in the coding region with a change of the stop codon at the normal position from TAA to TCA. Deletions of 1 to 12 nucleotides at five positions in the coding region and an insertion of 12 nucleotides in the 3' untranslated region were also found. HQP IV is, therefore, a processed psuedogene for QP which is generated by the retroposition of a processed mRNA or a cDNA into the chromosome. HQP V had an intron-like sequence but it terminated with stretches of adenylic acid reminiscent of the poly (A) tail downstream from the second putative polyadenylation signal sequence. Five of 32 amino acids encoded in the coding region sequenced were substituted with different amino acids from those encoded in the human QP cDNA (data not shown). These two distinct types of


-1 +1
+60
2 TCAAA ATG GCT GGT AAG CAG GCC GTT TCA GCA TCA GGC AAG TGG CTG GAT GGT ATT CGA AAA TGG
6 AGCTTGTCCTCTATGACTTACCCAGAAGGCAACGCTTCTCTTTCTGGTCAAA
15 CGGAGAAGGCAACGCTTCTCTTTCTGGTCAAA Met Ala Gly Lys Gln Ala Val Ser Ala Ser Gly Lys Trp Leu Asp Gly Ile Arg Lys Trp
I [-----AG]TT TCA GCA TCA GGC AAG TGG CTG GAT GGT ATT CGA AAA TGG
[GT---2.3kb---AG]
II GACTGAACGCGCCTGCGGACTGGGCCCAGCTTGTCCTCTATGACTTACCCAGAAGGCAACGCTTCTCTTTCTGGTCAAA ATG GCT GGT AAG CAG GCC GTT TCA GCA TCA GGC AAG TGG CTG GAT GGT ATT CGA AAA TGG
III GACTGAACGCGCCTGCGGACTGGGCCCAGCTTGTCCTCTATGACTTACCCACAAGGCAACGCTTCTCTTTCTGGTCAAA ATG GCT GGT AAG CAG GCC G[GT-----]
IV AAAATAAAATAAAATAAATTAGATAAATAAATAAATAAAAAGAGAGAGGGAGAAGGCAACATTTCACGTTTGGGTCAAG ATG GCT GGC AGG CTG GCT ACC GCA GCA TCA GAC AGG TAG CTG GAT GGT ATT TGA AAA TGG
V
+180
TAT TAC AAT GCT GCA GGA TTC AAT AAA CTG GGG TTA ATG CGA GAT GAT ACA ATA TAC GAG GAT GAA GAT GTA AAA GAA GCC ATA AGA AGA CTT CCT GAG AAC CTT TAT AAT GAC AGG ATG
Tyr Tyr Asn Ala Ala Gly Phe Asn Lys Leu Gly Leu Met Arg Asp Asp Thr Ile Tyr Glu Asp Glu Asp Val Lys Glu Ala Ile Arg Arg Leu Pro Glu Asn Leu Tyr Asn Asp Arg Met
PstI
I TAT TAC AAT GCT GCA GGA TTC AAT AAA CTG G[GT-----]
PstI
II TAT TAC AAT GCT GCA GGA TTC AAT AAA CTG G[GT-----]
PstI
III
IV TAT GAC AAT GCT GCA GGA TTC GAG AAA CTA AAG TAA ATG TGA GAT GAG ACA ATA TAT GAA GAT GAA GAT GTA AAG GAA GCC ATA AGA CGG CTT --- --- --- --- TCT AAC GGC AGG ATG
PstI
V
+300
TTT CGC ATT AAG AGG GCA CTG GAC CTG AAC TTG AAG CAT CAG ATC TTG CCT AAA GAG CAG TGG ACC AAA TAT GAA GAG GAA AAT TTC TAC CTT GAA CCG TAT CTG AAA GAG GTT ATT CGG
Phe Arg Ile Lys Arg Ala Leu Asp Leu Asn Leu Lys His Glu Ile Leu Pro Lys Glu Gln Trp Thr Lys Tyr Glu Glu Glu Asn Phe Tyr Leu Glu Pro Tyr Leu Lys Glu Val Ile Arg
I
II
III
IV TTT CGC ATT --- AGA ACA CTG GAC CT- --- ATG AGA CAG CAG ATC TTG CCT TAA TAG CA- CGG ACC AAA CAT GAA GAG GAT AAA TTC TAC CAT GAG CCA TAT CTG AAA CCG CTT ATT TGG
V (-----CATTTTTAGTTCCCAAGAGCCGGGCATAAACAAGAGACCTGAGGGAATCCTCCTCTCTTAAAAAACAAAG)CAG TGG ACC AAC TAT GAA GAG GAA AAT TTC TGC CTT GAA CTG TAT CTG AAA GAT GTT ATT TGG
+446
GAA AGA AAA GAA AGA GAA GAA TGG GCA AAG AAG TAA TCATGTAGTTGAAGTCTGTGGATGCAGGTGTTATGAAGATGGTTAAACTTGAAAACAAACAATTTTAAGAATTATTTGGTCTGAAGATGTTTTACTTTAAATAAATGTCTA
Glu Arg Lys Glu Arg Glu Glu Trp Ala Lys Lys ***
I
II
III
CCTCAAGTATTT
IV GAA AGA GAA GAA TGA GAA --- TGG TCA AAG AAG TCA TCACATATTTGAAGTTCGTGGATGTAGCTGTTATGAAAATGGTTAAACCTGAAACATATAA
V GAA AGA AAA GAA AGA GAA GAA TGG GCA AAG AAG TAA TCACATGGTTGAAGTCTGTGGATGCAACTGTTATGAAGATGGTTAAACTTGAAATAAACAATTTTAAGAATTATTTGGTCTGCAGATATTTTACTTTAAATAAATGTCTA
PstI
TTGTAATGGCTGG-poly(A) # 2
TTGT-poly(A) # 6
TTGTAATGGCTGGAGTTTTTGAATTCCAAACCTTATACTG-poly(A) #15
EcoRI
I
II
III
IV
V TTGTAATGGTAAAAAAACAAAAACAAAAACAAAACAAAACAAAAACAAAAACAAAACAAAACCAAAAAAACCTCAGG

human pseudogenes may provide a molecular record of QP evolution in biological system.

In HQP I, one exon including the *Pst*I site and two introns adjacent to the exon were found. HQP II extended farther upstream and contained, in addition to the identical exon found in HQP I, one more exon encoding the N-terminal portion of human QP, the 5' untranslated region of its mRNA, and the 5' flanking region. HQP III contained the identical nucleotide sequence with the same structural organization as was found in HQP II but did not contain the exon found in HQP I. From these results together with comparison of the restriction maps (see Fig. 1), it is concluded that HQP I, II, and III contain overlapping fragments from the same QP gene.

Structural organization of the human QP gene: As shown in Fig. 3a, four exons and three introns were found in the QP gene by sequence analysis of two *Eco*RI fragments of the gene which were hybridizable to clone #15 cDNA. Exon I was localized to the 0.6-kb *Hind*III-*Pst*I fragment, and exon II to the region covering a 3' part of the 1.8-kb *Xba*I-*Pst*I fragment and a 5' part of the 0.8-kb *Pst*I-*Hind*III fragment. Other exons, III and IV, were localized to the 1-kb *Eco*RI fragment. The *Eco*RI site at the 3' end of the fragment corresponded to the *Eco*RI site in the 3' untranslated region of the human QP cDNA of clone #15. The isolated human QP gene, therefore, spans 4.5 to 5 kb in length. As shown in Fig. 3b, the nucleotide sequences of all the exons were identical to the corresponding parts of the QP cDNA (9). All exon-intron junctions follow the consensus splicing sequence of GT at the 5' donor site and AG at the 3' acceptor site (12). Existence of the *Alu* repeat sequence in introns 1 and 2 was suggested by Southern blot analysis (data not shown). A stop codon TGA was present in frame at nucleotide position -39 to -37 indicating that the codon ATG at nucleotide position +1 is the real translation initiation site. Two CAAT boxes but not a typical TATA box were found in the 5' flanking region. In the case of the human cytochrome c_1 gene (5), there are only GC boxes in the 5' flanking region. The 5' flanking

---

Fig. 2. Nucleotide sequence comparison of sequenced regions from the human genes (HQP I to III) and pseudogenes (HQP IV and V) with those of human QP cDNAs (clones #2, 6, and 15). The nucleotide sequence of clone #15 was reported previously (9). The nucleotide sequences of the human QP cDNAs and their deduced amino acid sequence are boxed. In HQP IV and V, adenylic acid-rich sequences and stop codons are shown by underlines and broken underlines, respectively, and deletions and an insertion of nucleotides by dashes and wavy underline, respectively. Introns in HQP I, II, and III, and an intron-like sequence in HQP V are closed by brackets and parentheses, respectively. Nucleotide residue +1 denotes the A of the translation initiation codon ATG and residues preceding it are indicated by negative numbers. Putative polyadenylation signal sequences are indicated by thick underlines. *Pst*I and *Eco*RI recognition sites are indicated under the nucleotide sequences.

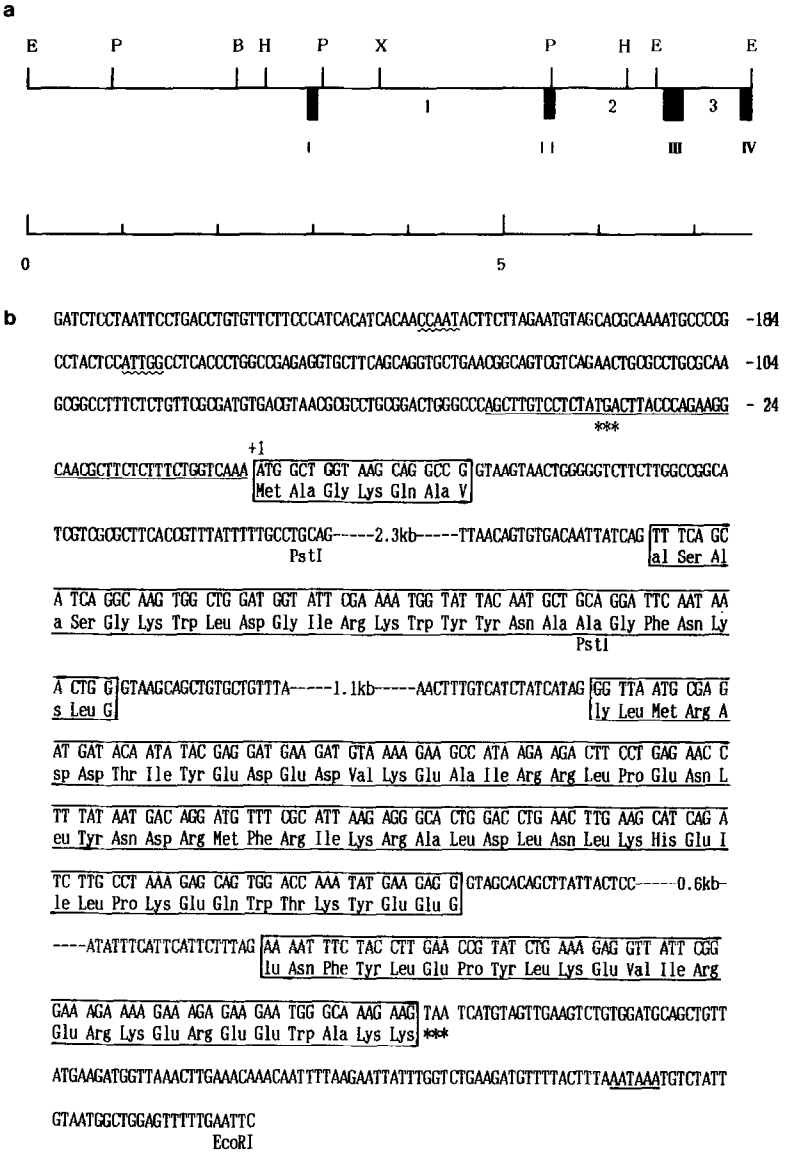


Fig. 3. Structural organization (a) and partial nucleotide sequence (b) of the human QP gene. a: Exons (I to IV) were represented by closed boxes and introns by figures 1, 2, and 3. E, P, B, H, and X are *Eco*RI, *Pst*I, *Bam*HI, *Hind*III, and *Xba*I recognition sites, respectively. The scale in kb is shown under the map. b: The nucleotide sequences of the coding regions and the deduced amino acid sequences are boxed. Nucleotide numbers denote the same as described in Fig. 2. CAAT boxes are indicated by wavy underlines. A sequence corresponding to the 5' untranslated region of the cDNA of clone #6 is underlined. The stop codon TGA present in frame at nucleotide position -39 to -37 is indicated by ***. A putative polyadenylation signal sequence is indicated by thick underline.

regions of genes for human F_0F_1-ATPase β subunit (6) and rat somatic cytochrome *c* (13) contain GC and CAAT boxes. It is very likely that nuclear genes encoding mammalian mitochondrial proteins participating in oxidative phosphorylation have commonly no typical TATA boxes and therefore, they are housekeeping type genes.

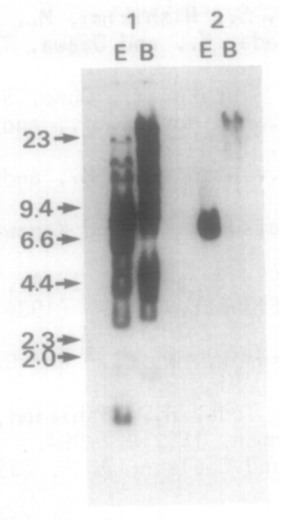


Fig. 4. Southern hybridization of fragments of the human QP gene to total human genomic DNAs. Human placenta DNAs (5μg each) were digested with *Eco*RI (E) or *Bam*HI (B), electrophoresed through a 0.6% agarose gel, and transfered to a nylon membrane. The 0.3-kb *Bam*HI-*Hind*III (1) or the 0.6-kb *Pst*I-*Xba*I (2) fragment of HQP III corresponding to a part of the 5' flanking region or a part of intron 1, respectively, of the gene was used as a probe. Autoradiography of the membrane was performed at -80°C with an intensifying screen for 48 h. Marker size (left) in kb.

Single gene for human QP: Southern blot analysis of human placenta DNA was performed to determine the copy number of the QP gene in the human genome. As shown in Fig. 4, using the 0.3-kb *Bam*HI-*Hind*III fragment corresponding to the 5' flanking region of the gene, multiple hybridization bands were observed. However, a single about 7-kb or over 23-kb hybridization band was observed in *Eco*RI or *Bam*HI digest of human genomic DNA, when the 0.6-kb *Pst*I-*Xba*I fragment which corresponds to a part of intron 1 was used as a probe. The size of 7 kb was essentially the same as that expected from the restriction map (see Fig. 3a). We therefore conclude that the isolated human QP gene occurs in a single copy in the human genome. The single gene and the identical nucleotide sequences of the three cDNAs rule out the occurrence of tissue-specific human QP.

Acknowledgments

This work was supported in part by the Grants-in-Aids for Scientific Research on Priority Areas (62617002) from the Ministry of Education, Science and Culture of Japan. Experiments using radioisotope were performed in Radioisotope Center Medical Division, Nagoya University.

References

1. Battini, R., Ferrari, S., Kaczmarek, L., Calabretta, B., Chen, S-T., and Baserga, R. (1987) J. Biol. Chem. 262, 4355-4359.
2. Glaichenhaus, N., Leopold, P., and Cuzin, F. (1986) EMBO J. 5, 1261-1265.

3. Tanaka, M., Miyabayashi, S., Nishikimi, M., Suzuki, H., Shimomura, Y., Ito, K., Narisawa, K., Tada, K., and Ozawa, T. (1988) Pediatric Res. 24, 447-454.
4. Ozawa, T., Yoneda, M., Tanaka, M., Ohno, K., Sato, W., Suzuki, H., Nishikimi, M., Yamamoto, M., Nonaka, I., and Horai, S. (1988) Biochem. Biophys. Res. Commun. 154, 1240-1247.
5. Suzuki, H., Hosokawa, Y., Nishikimi, M., and Ozawa, T. (1989) J. Biol. Chem. 264, 1368-1374.
6. Ohta, S., Tomura, H., Matsuda, K., and Kagawa, Y. (1988)J. Biol. Chem. 263, 11257-11262.
7. Maguire, D. J., Day, A. R., Borthwick, I. A., Srivastava, G., Wigley, P. L., May, B. K., and Elliott, W. H. (1986) Nucleic Acids Res. 14, 1379-1391.
8. Dorsman, J. C., van Heeswijk, W. C., and Grivell, L. A. (1988) Nucleic Acids Res. 16, 7287-7301.
9. Suzuki, H., Hosokawa, Y., Toda, H., Nishikimi, M., and Ozawa, T. (1988) Biochem. Biophys. Res. Commun. 156, 987-994.
10. Sanger, F., Nicklen, S.,and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
11. Mizusawa, S., Nishimura, S., and Seela, F. (1986) Nucleic Acids Res. 14, 1319-1324.
12. Breathnach, R., and Chambon, P. (1981) Ann. Rev. Biochem. 50, 349-383.
13. Evans, M. J., and Scarpulla, R. C. (1988) Mol. Cell. Biol. 8, 35-41.