Using cloned human prealbumin cDNA as a probe, Southern blot hybridization of human genomic DNA revealed that the prealbumin gene consists of an unique, single-copy DNA. The nucleotide sequences of the entire human prealbumin gene, including both 581 base pairs of the 5'- and 95 base pairs of the 3'-flanking sequences, were determined. The gene spans about 7.0 kilobase pairs and consists of four exons and three introns. As in most eukaryotic genes, the consensus TATA and CAAT sequences are found 30 and 101 nucleotides, respectively, upstream from the putative cap site, and a polyadenylation signal sequence AAATAAA is found in the 3'-untranslated region. Unexpectedly, two independent open reading frames were found within the gene: one in the first intron and the other in the third intron.

Prealbumin is a protein which plays an important role in the plasma transport of vitamin A and is also involved in the transport of thyroid hormones. The human prealbumin molecule is composed of four identical subunits of 127 amino acid residues, and its amino acid sequence is known. The human prealbumin molecule is composed of four identical subunits of 127 amino acid residues, and its amino acid sequence is known. The human prealbumin molecule is composed of four identical subunits of 127 amino acid residues, and its amino acid sequence is known. The human prealbumin molecule is composed of four identical subunits of 127 amino acid residues, and its amino acid sequence is known. The human prealbumin molecule is composed of four identical subunits of 127 amino acid residues, and its amino acid sequence is known.

**RESULTS AND DISCUSSION**

Southern Blot Analysis of Total Human Genomic DNA—Southern blot analysis of DNAs extracted from human placenta was performed, using two kinds of 32P-labeled prealbumin cDNA probes: one a PstI/PvuII fragment derived from the pPAl cDNA insert and covers a whole prealbumin cDNA, and the second is a PstI/XbaI fragment covering only the 5' part of the cDNA. When the total probe, i.e. the PstI/PvuII fragment was used, the EcoRI or HindIII digests of human DNA yielded two bands and the EcoRI/HindIII digests yielded three bands on the autoradiogram (Fig. 1a). On the other hand, in the experiment using the 5' part of the pPAl cDNA insert as a probe, all of these three differently digested human DNAs yielded only one band (Fig. 1b). Because the prealbumin cDNA contains no EcoRI and HindIII sites, the result suggests that there is only one prealbumin gene which contains at least two introns. This idea was given support when isolating phage clones containing a whole human prealbumin gene (see below).

Isolation and Analyses of Phage Clones Carrying a Prealbumin Gene—Two different human gene libraries, one constructed by Lown et al. (12) and the other by Tsuzuki et al. (11), were screened by in situ plaque hybridization, using a 32P-labeled human prealbumin cDNA as a probe. Twelve positive clones were isolated from the first library and two clones from the second library.

Digestion of the cloned DNAs with restriction enzymes and Southern blot hybridization, using the prealbumin cDNA as a probe, indicated that all these cloned DNAs were covering overlapping portions of a chromosomal segment. One of the clones, named Lm PAM-5, contained three EcoRI/HindIII restriction fragments of 2.2, 2.0, and 1.05 kilobase pairs in length (Fig. 1d) and seemed to contain the entire prealbumin gene, as only these three bands were detected by Southern blot analysis of the human genomic DNA (Fig. 1a, lane E/H). We found that Lm PAE-7, which was isolated from the EcoRI partial library, is carrying the 9.0-kilobase pair EcoRI fragment hybridizable to the prealbumin cDNA probe (data not shown). Fig. 2a summarizes the restriction maps of the genomic DNA fragments cloned in Lm PAM-5 and Lm PAE-7.

**Nucleotide Sequence of the Human Prealbumin Gene**—The
Human Prealbumin: Gene Structure

DNA sequencing strategy is shown in Fig. 2b, and the entire nucleotide sequence of the human prealbumin gene region is summarized in Fig. 3. These sequences were determined for the cloned DNAs present in Lambda PAM-5 and Lambda PAB-7 and include exons, introns, and 5'- and 3'-flanking regions (581 base pairs of the 5'- and 95 base pairs of the 3'-flanking regions). As shown schematically in Fig. 2b, the human prealbumin gene consists of four exons and three introns. The sequences of four exons were identical with that of the human prealbumin cDNA (9), except for one base substitution in the nucleotide sequence of the human prealbumin gene region. Base pairs of the 5'- and 95 base pairs of the 3'-flanking regions. As shown schematically in Fig. 2b, the human prealbumin gene consists of four exons and three introns. The sequences of four exons were identical with that of the human prealbumin cDNA (9), except for one base substitution in the nucleotide sequence of the human prealbumin gene region. Base pairs of the 5'- and 95 base pairs of the 3'-flanking regions.

The transcription initiation site or the cap site was tentatively assigned to residue A located at position +1. This assignment is based on the finding that an A preceded by a C in most cases is the preferred cap site and on the assumption that our prealbumin cDNA clone (9) covers the full length of the human prealbumin mRNA. Furthermore, another independently isolated prealbumin cDNA clone has been reported that our prealbumin cDNA clone (9) covers the full length of the human prealbumin mRNA. montage (16).

The polyadenylation site was inferred by comparing the nucleotide sequence of the 3'-untranslated region of the gene with that of the prealbumin cDNA (9). A polyadenylation signal sequence, AATAAA, was located 23 base pairs upstream from the polyadenylation site. Benoist et al. (18) identified another consensus sequence, TTTCACCTC, near the polyadenylation site in several mRNAs, and we found a similar sequence, TTTCACCTC, 4 nucleotides upstream from the AATAAA hexanucleotide in the human prealbumin gene. In the second and third introns, we found 300-nucleotide sequences strikingly homologous to the human Alu-type repeat elements (19). The characteristic features of the Alu-type repeats were the presence of short poly(A) tracts and a 9-nucleotide or a 15-nucleotide direct repeat (Fig. 3). Similar Alu sequences have been found in introns and exons of other genes (20, 21).

Presence of Two Open Reading Frames in the Introns—One of the characteristic features of this gene is the presence of two independent open reading frames, in the first and third introns. Both of these two open reading frames are provided, in their 5'- and 3'-flanking regions, with the consensus regulatory sequences for transcription.

There are two putative initiation codons for the first unidentified reading frame: one is located between nucleotide positions 684 and 686 and the other between positions 688 and 690. When the translation starts from the first ATG codon, the frame codes for 60 amino acids, and when it starts from the second codon, there is an open reading frame for 37 amino acids. We found a TATAAA sequence 106 nucleotides upstream from the first putative initiation codon and a CAAT sequence 55 nucleotides upstream from the TATA sequence. The polyadenylation signal sequence AATAAA is located 94 nucleotides downstream from the termination codon for the first open reading frame (nucleotide positions 684-666).

The second unidentified open reading frame also has two putative initiation codons: one is located at nucleotide positions 6061 and 6063 and the other at positions 6128 and 6130. When the translation starts from the first ATG codon, the frame codes for 49 amino acids, and when it starts from the second, there is an open reading frame for 69 amino acids. A possible TATA equivalent sequence, TATATAT, is located 60 nucleotides upstream from the first putative initiation codon, and the CAAT sequence is located 44 nucleotides upstream from the TATA sequence. In this case, we found two possible poly(A) addition signals: one located 45 nucleotides and the other 243 nucleotides, respectively, downstream from the termination codon for the second open reading frame (nucleotide positions 6559-6577).

The presence of two unidentified reading frames, one in the first intron and the other in the third intron, is an unexpected structure of the human prealbumin gene. "One gene's intron is another gene's exon" (22) was first found in the yeast mitochondrial cytochrome b gene (23), and this open reading frame is proved to code for a maturase, a protein which acts in the process of splicing the mitochondrial cytochrome b transcript. Further work will reveal if the open reading frames in introns 1 and 3 are expressed in vivo.

The characterized DNA segment of the human prealbumin gene described herein should facilitate not only elucidation of the control mechanism of prealbumin gene expression, but also studies on the genetic basis for accumulation of a variant prealbumin in FAP patients.

Acknowledgments.—We thank Dr. T. Maniatis for kindly providing the human genomic library and M. Ohara of Kyushu University for comments on the manuscript.

REFERENCES

1. Kanda, Y., Goodman, D. S., Canfield, R. E., and Morgan, P. J. (1974) J. Biol. Chem. 249, 6799-6804.
2. Robinson, J., and Rall, J. L. (1980) Physiol. Rev. 60, 415-449.
3. Kamai, M., Rez, A., and Goodman, D. S. (1986) J. Clin. Invest. 47, 2025- 2034.
4. Anzilote, C. (1952) Brain 75, 405-427.
5. Araki, S., Kawanishi, S., Sato, K., Helmer, T., Jay, G., and Agarwal, K. (1984) J. Clin. Invest. 74, 104-119.
6. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1985) Nature 310, 415-489.
7. Schmid, C. W., and Jelinek, W. (1982) Science 216, 1065-1070.
8. Kioussis, D., Efferman, F. V., de Rijk, P., Gorin, M. B., Ingram, R. S., and Tilghman, S. M. (1985) Cell 32, 399-406.
9. Ito, R., Suto, K., Helmer, T., Jay, G., and Agarwal, K. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 4662-4666.
10. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 157-171.
11. Taka, T., Nishimura, Y., Seto, A., and Ishida, T. (1981) Nucleic Acids Res. 8, 249-257.
12. Schmid, C. W., and Jelinek, W. (1982) Science 216, 1065-1070.
13. Kioussis, D., Efferman, F. V., de Rijk, P., Gorin, M. B., Ingram, R. S., and Tilghman, S. M. (1985) Cell 32, 399-406.
14. Ito, R., Suto, K., Helmer, T., Jay, G., and Agarwal, K. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 4662-4666.
15. Benoist, C., O'Hare, K., Breathnach, R., and Chambon, P. (1980) Nucleic Acids Res. 8, 157-171.
16. Schmid, C. W., and Jelinek, W. (1982) Science 216, 1065-1070.
17. Kioussis, D., Efferman, F. V., de Rijk, P., Gorin, M. B., Ingram, R. S., and Tilghman, S. M. (1985) Cell 32, 399-406.
18. Ito, R., Suto, K., Helmer, T., Jay, G., and Agarwal, K. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 4662-4666.
19. Schmid, C. W., and Jelinek, W. (1982) Science 216, 1065-1070.
20. Kioussis, D., Efferman, F. V., de Rijk, P., Gorin, M. B., Ingram, R. S., and Tilghman, S. M. (1985) Cell 32, 399-406.
21. Ito, R., Suto, K., Helmer, T., Jay, G., and Agarwal, K. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 4662-4666.
22. Benoist, C., and Gravel, L. (1981) Nature 289, 439-440.
23. Lazowska, J., Jacq, C., and Soinimaki, P. P. (1980) Cell 22, 333-348.

Continued on next page.
SUPPLEMENTARY MATERIAL TO Structure of the Human Prealbumin Gene

Teruhisa Tsukui, Shuji Mita, Shuichiro Maeda, Shukuro Araki, and Kazuoori Shiozawa.

EXPERIMENTAL PROCEDURES

Materials. Restriction enzymes were purchased from Takara Shuzo Co., New England Biolabs, Toyobo Co., and Nippon Gene Co.. Calf intestinal alkaline phosphatase was from Boehringer Mannheim. E. coli DNA polymerase I was obtained from New England Biolabs and T4 polynucleotide kinase from Takara Shuzo Co. [+P]ATP (1,000 Ci/mmol) and [-32P]ATP (1,000 Ci/mmol) were purchased from Amersham and New England Nuclear, respectively.

Isolation and Characterization of Genomic Clones. Two independently constructed human genomic libraries were screened for the clones carrying the prealbumin gene. One was constructed from EcoRI partial digests of human placental DNA (11) and the other, from AluI/HindIII partial digests of human prealbumin gene. One was constructed from EcoRI partial digests of human and Stafford (10).

DNA Sequencing. DNA sequencing was done using the chemical method of Maxam and Gilbert (15) with analysis on 6 or 20 polyacrylamide gels.

DNA Sequencing. DNA sequencing was done using the chemical method of Maxam and Gilbert (15) with analysis on 6 or 20 polyacrylamide gels.

Fig. 1. Southern blot analysis of human DNA and Lambda DNA. 0.5 or 20% polyacrylamide gels.

Fig. 2a. Transformation of E. coli HB101 with E. coli plus HindIII-cut plasmid DNA. One was constructed from EcoRI partial digests of human placental DNA (11) and the other, from AluI/HindIII partial digests of human fetal liver DNA (12). Approximately 10 or 100 plages of each library were screened for the clones carrying the prealbumin gene.

The filters were hybridized with the [-32P]labeled pPAl-DNA insert (a) or Lambda DNAs cleaved with HindIII. Positive clones were plaque-purified and the purified clones grown in 400 ml cultures for preparation of DNAs. Each DNA extracted from these clones was digested with EcoRI, HindIII or EcoRI plus HindIII. DNA sequencing was done using the chemical method of Maxam and Gilbert (15) with analysis on 6 or 20 polyacrylamide gels.

The general procedures for thin translation and Southern blot hybridization were performed as described (11).

All of the cloning procedures were carried out in accordance with the guidelines for the use of recombinant DNA molecules issued by the Ministry of Education, Science and Culture of Japan.

DNA Sequencing. DNA sequencing was done using the chemical method of Maxam and Gilbert (15) with analysis on 6 or 20 polyacrylamide gels.

Fig. 1. Southern blot analysis of human DNA and Lambda DNA. 0.5 or 20% polyacrylamide gels.

Fig. 2a. Transformation of E. coli HB101 with E. coli plus HindIII-cut plasmid DNA. One was constructed from EcoRI partial digests of human placental DNA (11) and the other, from AluI/HindIII partial digests of human fetal liver DNA (12). Approximately 10 or 100 plages of each library were screened for the clones carrying the prealbumin gene.

The filters were hybridized with the [-32P]labeled pPAl-DNA insert (a) or Lambda DNAs cleaved with HindIII. Positive clones were plaque-purified and the purified clones grown in 400 ml cultures for preparation of DNAs. Each DNA extracted from these clones was digested with EcoRI, HindIII or EcoRI plus HindIII. DNA sequencing was done using the chemical method of Maxam and Gilbert (15) with analysis on 6 or 20 polyacrylamide gels.

The general procedures for thin translation and Southern blot hybridization were performed as described (11).

All of the cloning procedures were carried out in accordance with the guidelines for the use of recombinant DNA molecules issued by the Ministry of Education, Science and Culture of Japan.

DNA Sequencing. DNA sequencing was done using the chemical method of Maxam and Gilbert (15) with analysis on 6 or 20 polyacrylamide gels.

DNase I digestion...
