

The human serum albumin gene: structure of a unique locus

(Human gene family; unique gene; recombinant phage; genomic library)

James W. Hawkins and Achilles Dugaiczky

Department of Cell Biology, Baylor College of Medicine, Texas Medical Center, Houston, TX 77030 (U.S.A.)

(Received February 27th, 1982)

(Accepted April 22nd, 1982)

SUMMARY

The entire gene for human serum albumin (HSA) has been isolated from a genomic DNA library, carried in the λ Charon 4A vector. Six independent isolates have been found to hybridize to a cloned HSA cDNA probe, and all six clones share restriction site sequence homology in the overlapping portion of their DNA. These results seem to indicate that the albumin gene is single-copy, or unique, within the human haploid genome. Measuring from the "CAP" site to the "poly(A)" addition site, the albumin gene comprises 16.5 kb of DNA.

INTRODUCTION

Serum albumin is a monomeric protein with an M_r -value of 65000, characterized by its high solubility in water and by lack of a carbohydrate moiety. It is a product of the liver and, as a secreted protein, serves in the plasma to maintain osmotic pressure and as a carrier of ions, fatty acids, metabolites, bilirubin, drugs and hormones (Rothschild and Oratz, 1976). The primary amino acid structure (Meloun et al., 1975; Brown, 1976) and nucleotide sequence of the messenger RNA (Dugaiczky et al., 1982) of human albumin are known. In the light of the accepted physiological functions of serum albumin, perhaps most intriguing is the existence of analbuminemic individuals

(Boman et al., 1976) with no apparent clinical disorders.

Recently, we have been engaged in defining the organization and primary DNA structure of the human chromosomal region containing the developmentally related genes for albumin and α -fetoprotein. This article reports the isolation of a continuous stretch of chromosomal DNA coding for the primary RNA transcript of the human serum albumin gene. The gene appears to represent a unique locus within the human genome.

MATERIALS AND METHODS

(a) Reagents and enzymes

All reagents were obtained from commercial sources at the highest purity desirable. [α -³²P]dATP (> 400 Ci/mmol), used in the nick-

Abbreviations: CAP site, position of 5'-terminal 7-methylguanosine modification of mRNA; HSA, human serum albumin; kb, kilobase pairs.

translation of probes, was obtained from Amersham.

(b) Hybridization probes

The two cDNA clones (Dugaiczky et al., 1982) depicted in Fig. 1 were used as screening and hybridization probes in these experiments. Clone pHA36, referred to as the "3'-probe", extends from the 3' poly(A) sequence of the albumin mRNA some 1.6 kb toward the 5' end of the mRNA. The second clone, pHA206, referred to as the "5'-probe", overlaps pHA36 by some 150 nucleotides at the 5' end of pHA36 and extends further about 550 residues to within 62 nucleotides of the putative CAP site of the albumin mRNA.

(c) Screening of a genomic library

The human gene library used in these experiments was generously provided to us by Dr. Tom Maniatis. This library was constructed in the Charon 4A λ phage vector using human DNA partially digested with the restriction endonucleases *Hae*III and *Alu*I (Lawn et al., 1978). The library was plated at a density of 10000 phage per plate and screened by the procedure of Benton and Davis (1977). The cDNA clone pHA36, labeled by nick-translation to 10^8 cpm/ μ g, was used in these hybridizations. Positive clones were plaque-purified.

(d) Analysis of phage DNA

Phage DNA was digested with the restriction endonuclease *Eco*RI and analyzed by 1% agarose or 5% acrylamide gel electrophoresis, followed by staining with 5 μ g/ml ethidium bromide. Restricted DNA fragments were transferred from agarose gels to nitrocellulose filters by the method of Southern (1975). Both cDNA clones, pHA36 and pHA206, were used in these hybridizations. Subsequently, also the genomic 2.2-kb DNA, isolated from clone λ HSA 243, and the 4.4-kb DNA, isolated from the same clone λ HSA 243 (Fig. 2), were used as probes in Southern hybridization to demonstrate sequence homology between the DNA end fragments of the six overlapping chromosomal segments carried in the phage clones.

RESULTS

(a) Isolation and preliminary analysis of phage clones

Approx. 3.5×10^6 phage clones were screened using the human albumin cDNA probe, pHA36, as described in MATERIALS AND METHODS. This screening yielded six positive clones. The DNA from each positive clone was digested with the restriction endonuclease *Eco*RI and the resulting fragments characterized by agarose gel electrophoresis (not shown). The analysis indicated the presence of six unique DNA clones: HSA10, HSA110, HSA191, HSA215, HSA243, and HSA322. Analyzing results of this experiment, a 10.6-kb and a 1.6-kb DNA fragment were found to be common to five of the clones. Since it appeared that there were no other common fragments, it was concluded that 1.6-kb and 10.6-kb DNA must be consecutive within the chromosome.

(b) Orientation of the overlapping clones

When *Eco*RI-generated DNA fragments were hybridized to the 3'-probe, and, separately, the 5'-probe, it was found that only the 1.6-kb fragment hybridized to both probes. It was concluded that this fragment must be homologous to the overlapping region of the two probes (Figs. 1 and 2). The remaining DNA fragments were placed relative to each other on the basis of their hybridization to either the 5'- or 3'-probe.

The position of the end fragments of the six clones was in addition confirmed by hybridization experiments showing sequence homology between these DNA end fragments. Thus, a nick-translated probe, generated from the 4.4-kb DNA of clone HSA243 (Fig. 2), hybridized to only one DNA fragment of each clone, i.e. to the 4.5-kb fragment of HSA322, 4.4 of HSA243, 1.9 of HSA215, 1.1 of HSA10, 1.3 of HSA191, and one of the two 7.5-kb fragments of HSA110. The other 7.5-kb fragment of HSA110 can be cleaved with *Hha*I endonuclease into two sub-fragments which do not hybridize to the 4.4-kb probe.

Similarly, sequence homology was established at the 5' end of the overlapping clones. A probe, generated from the 2.2-kb DNA of clone HSA243,

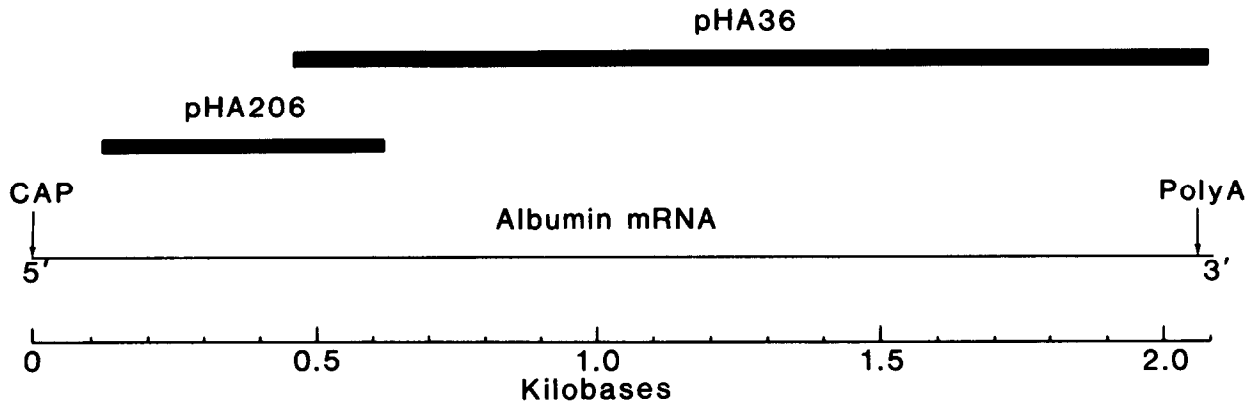


Fig. 1. Simplified map of the overlapping human albumin cDNA clones pHA36, the 3'-probe, and pHA206, the 5'-probe, used in initial phage screening and subsequent hybridization experiments. Clones are positioned above the albumin mRNA molecule with respect to structural homology. The mRNA "CAP" site and the point at which the mRNA is polyadenylated (Poly(A)) are indicated. Further details of these clones are presented in Dugaiczky et al. (1982).

hybridized again to only one fragment of each clone, i.e. to 2.2 of HSA243, 1.0 of HSA215, 3.3 of HSA10, and 3.3-kb of HSA191. Taken as a whole, the experiments allow the characterization of all phage-genomic *Eco*RI fragments in terms of relative overlap with each other, and a 5'-3' orientation with respect to the albumin mRNA coding sequence (Fig. 2).

DNA sequence analysis, not described here,

places the 5' end of the mRNA at the position designated "CAP" in Fig. 2 and the 3' terminus of the mRNA at the position designated "poly(A)". These data indicate that the albumin primary RNA transcript is encoded in a stretch of DNA of approx. 16.5 kb long, a figure similar to that reported for the mouse albumin gene, about 17.5 kb (Kioussis et al., 1981) and that for the rat albumin gene, about 15.0 kb (Sargent et al., 1981).

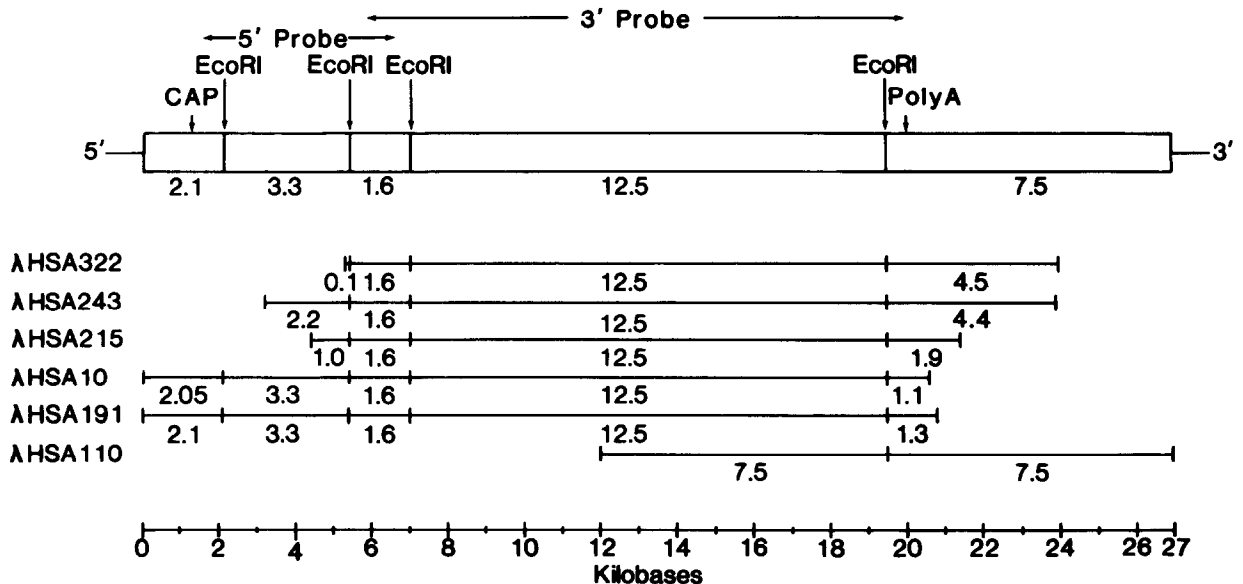


Fig. 2. Overlapping *Eco*RI DNA fragments of human genomic inserts derived from phage clones containing the human albumin gene. Single lines represent the designated phage inserts. The open box represents the continuous stretch of genomic DNA spanned by all phage clones. Above the open box, the genomic distance over which the 3'-probe, pHA36, and the 5'-probe, pHA206, hybridize is indicated. The correct size of the 12.5 kb DNA fragment is 10.6 kb.

DISCUSSION

In the course of evolution genes develop into families, the exact number of their members can be only determined by structure analysis of chromosomal DNA. Although some genes become non-functional pseudogenes, they can still be recognized as belonging to a given family by a residual sequence homology. During the same time of evolution, some genes evolve into a multi-gene family while others remain as single-copy, or unique, within the same species.

Since the number of recombinant phage required to obtain a particular genomic sequence from the library we have used was 8×10^5 , and since 3.5×10^6 phage were screened, one would theoretically expect to obtain 4 to 5 albumin-containing phage if the albumin gene were present in the haploid genome as a single copy. This predicted number of phage is in close agreement with the six albumin-containing clones actually obtained. The close agreement of these predicted and observed figures also argues against the existence of closely related albumin pseudogenes. This suggestion is further supported by the observed identity of restriction sites in the overlapping DNA clones, and the fact that when restricted genomic DNA blots were probed under stringent conditions with human albumin cDNA, only DNA fragments of predicted size were observed (not shown).

Having isolated the complete albumin gene, we are now engaged in its fine structure analysis with particular emphasis upon inferring a possible mechanism of the gene's internal evolution and its relationship to the α -fetoprotein gene, evidently the only other member of this family (Jagodzinski et al., 1981; Gorin et al., 1981; Law and Dugaiczky, 1981).

A detailed structure of the human albumin gene will also serve as a baseline onto which results from disordered conditions, such as analbuminemia, can be projected.

ACKNOWLEDGEMENTS

We thank Drs. William Schrader and Stanley Glasser for critical reading of the manuscript. We

extend our appreciation to Dr. Tom Maniatis for his generous gift of the human genomic library. This work was supported by the Kleberg Foundation.

REFERENCES

- Benton, W.D. and Davis, R.W.: Screening of λ gt recombinant clones by hybridization to single plaques in situ. *Science* 196 (1977) 180-182.
- Boman, H., Hermanson, M., Hammond, C.A. and Motulsky, A.G.: Analbuminemia in an American Indian girl. *Clin. Genet.* 9 (1976) 513-526.
- Brown, J.R.: Structural origins of mammalian albumin. *Fed. Proc.* 35 (1976) 2141-2144.
- Dugaiczky, A., Law, S.W. and Dennison, O.E.: Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. *Proc. Natl. Acad. Sci. USA* 79 (1982) 71-75.
- Gorin, M.B., Cooper, D.L., Eiferman, F., Van de Rijn, P. and Tilghman, S.: The evolution of α -fetoprotein and albumin. *J. Biol. Chem.* 256 (1981) 1954-1959.
- Jagodzinski, L.L., Sargent, T.D., Yang, M., Glackin, C. and Bonner, J.: Sequence homology between RNAs encoding rat α -fetoprotein and rat serum albumin. *Proc. Natl. Acad. Sci. USA* 78 (1981) 3521-3525.
- Kioussis, D., Eiferman, F., Van de Rijn, P., Gorin, M.B., Ingram, R.S. and Tilghman, S.M.: The evolution of α -fetoprotein and albumin. *J. Biol. Chem.* 256 (1981) 1960-1967.
- Law, S.W. and Dugaiczky, A.: Homology between the primary structure of α -fetoprotein, deduced from a complete cDNA sequence, and serum albumin. *Nature* 291 (1981) 201-205.
- Lawn, R.M., Fritsch, E.F., Parker, R.C., Blake, G. and Maniatis, T.: The isolation and characterization of linked α - and β -globin genes from a cloned library of human DNA. *Cell* 15 (1978) 1157-1174.
- Meloun, B., Moravek, L. and Kostka, V.: Complete amino acid sequence of human serum albumin. *FEBS Lett.* 58 (1975) 134-137.
- Rothschild, M.A. and Oratz, M.: Albumin synthesis and degradation, in A.C. Allison (Ed.), *Structure and Function of Plasma Proteins*, Vol. 2. Plenum, New York, 1976, pp. 79-105.
- Sargent, T.D., Jagodzinski, L.L., Yang, M. and Bonner, J.: Fine structure and evolution of the rat serum albumin gene. *Mol. Cell Biol.* 1 (1981) 871-883.
- Southern, E.M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98 (1975) 503-517.

Communicated by A.D. Riggs.