Volume 11 Number 17 1983 Nucleic Acids Research

Sequence of human haptoglobin cDNA: evidence that the α and β subunits are coded by the same mRNA

G.Raugei, G.Bensi, V.Colantuoni[1], V.Romano[2], C.Santoro[3], F.Costanzo[1] and R.Cortese

European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, FRG

Received 22 July 1983; Accepted 5 August 1983

ABSTRACT

We have isolated and sequenced a cDNA clone coding for human haptoglobin. Our sequence shows that haptoglobin is very likely synthesized as a single polypeptide chain which is then cleaved at an Arg residue to generate its two characteristic α and β subunits. Southern blot analysis suggests that there are at least two copies of the haptoglobin gene per haploid genome.

INTRODUCTION

Haptoglobin is synthesized in the liver and secreted in the plasma where normal levels are about 300 mg/100 ml. The molecule is composed of two subunits, α and β, which associate in an α_2-β_2 quaternary structure (1,2). Haptoglobin is found in a complex with hemoglobin. It is postulated that its function is to prevent loss of heme at the level of kidney tubules and to facilitate the reabsorbtion of hemoglobin in the liver. These functions are apparently dispensable because healthy anhaptoglobinemic individuals have been described (1). Haptoglobin behaves as an acute phase protein in a variety of pathological conditions and experimental systems (3).

Human haptoglobin has been the object of extensive genetic analysis. It is known that there is an interesting polymorphism at the level of the hpα subunits; more than 50% of the population in Europe has a partially duplicated α chain, called $hp\alpha^2$, whose sequence has been determined (4). A much rarer partial α chain triplication has been also described (5). Another polymorphism has been identified at the level of aminoacid 53, which can be a lysine (in about 30% of the population in Europe) or glutamic acid (in about 70% of the population in Europe). This results in electrophoretically distinguishable Fast (F) and Slow (S) migrating classes of α chains (6). Several rarer variants have also been observed on the basis of different electrophoresis patterns (2). Much less is known about the genetics of the hpβ chain, but rare variants have also been described in this case (2). Chung et al. have

recently presented indirect evidence that there might be at least two haptoglobin species, one synthesized predominantly during fetal life and the other characteristic of the adult (7).

In this paper we present the sequence of a human cDNA clone showing that the sequence coding for the hpα chain, in this case an hpα^2 type, is adjacent to the sequence coding for the hpβ chain on a single mRNA. These results concerning the mechanism of biosynthesis of haptoglobin are in agreement with the hypothesis and the experimental evidence of Chow et al. (8) and with the data of Haugen et al. (9). Southern blot analysis provides evidence for the presence of more than one haptoglobin gene per haploid genome and confirms that there is restriction site polymorphism at this locus.

RESULTS AND DISCUSSION

Cloning of human haptoglobin cDNA

We have previously reported the cloning of a short segment of human cDNA coding for part of the hpβ chain (10). We have used this segment as a probe to screen a human liver cDNA library. We have screened 4000 colonies and found 24 positives, which were called Hapto 1 to Hapto 24. A preliminary characterization revealed that clone Hapto 6 contained the largest insert. Restriction analysis revealed the presence of two PstI sites, as predicted by the cloning strategy, delimiting an insert of about 1500 bases. This corresponded satisfactorily to the size of haptoglobin mRNA determined by northern blot analysis (fig. 1). The Pst-Pst insert also contained a BamHI and a HindIII site. The segments Pst-Bam, Bam-Hind, Hind-Pst of 480,647 and 370 bases respectively, were subcloned in the vectors pEMBL8 and 9, and mp8 and 9. The DNA sequence of both strands was determined with the Sanger and Maxam-Gilbert methods and is shown in fig. 2. The analysis of the sequence reveals the presence of an open reading frame starting from the second base after the polyG tail and ending with a stop codon in position 1219. In position 1363 there is a typical polyadenylation site sequence (11), followed, 18 bases downstream, by a polyA sequence of 80 bases. The first 16 aminoacids of the open reading frame are identical with 16 residues of the rabbit hpα leader peptide reported by Chow et al. (8). The following 142 aminoacid residues are almost identical to the published sequence of the hpα^2 chain (4). From position 484 the deduced sequence is almost identical to the hpβ chain of Kurosky et al. (4). This fact gives a direct evidence in favour of the mechanism of haptoglobin biosynthesis proposed by Chow et al. (8) and Haugen et al. (9). The two chains are separated by an arginine residue which is

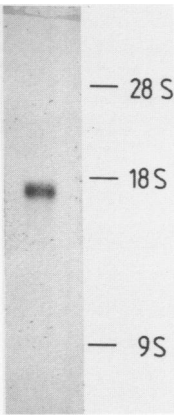

Fig. 1: Northern blot analysis: 10 µgr. of human liver polyA$^+$ were size-fractionated on a 1.5% formaldehyde-agarose gel and transferred to a nitrocellulose filter. 32P-labelled T40 cDNA was used as a probe.

probably excised from the α-β precursor to generate the two subunits. The amino acid sequence at the junction between the α and β chain, where the proteolytic cleavage is likely to occur, is very characteristic of several proteases at the position of cleavage required for activation (fig. 3). This conservation gives further support to the hypothesis that the haptoglobin gene has evolved from a common ancestor to the serine protease family (4). We found minor differences with the aminoacidic sequence determined by Kurosky et al. (4): an Asp instead of a Asn at position 52 of the hpα^2 chain; in the hpβ chain an inversion in the order of residues 219 and 220 which in our sequence are Ser-Cys rather than Cys-Ser and an Asp instead of an Asn in position 236.

The sequence of the hpα^2 chain, showing a Lis at position 53 and a Glu at posiition 112, suggested that the Hpα^2 gene might have been generated by an unequal crossover between an Hpα^{1F} and an Hpα^{1S} gene (15). In our sequence the findings of an Asp instead of an Asn in position 52 and the confirmation of an Asn in position 111 suggests the possibility of a linkage between the F or S phenotypes with the presence of an Asp or an Asn respectively in the position preceding the aminoacid responsible for the two phenotypes.

Southern blot analysis of haptoglobin genes

We have previously described the cloning of a 120 bp cDNA segment coding for part of the human hpβ (10). This segment, called T40, starts at the Bam site in the position 477 of the DNA sequence shown in fig. 1 and ends at an AluI site at position 680. P32-labelled T40 DNA was used as a probe for Southern blot analysis of EcoRI restricted human chromosomal DNA. The pattern obtained is shown in fig. 3 panel A. We observed three distinct segments, a, b and c of 9, 12.5 and 14 kb respectively. Band a is approximately twice as

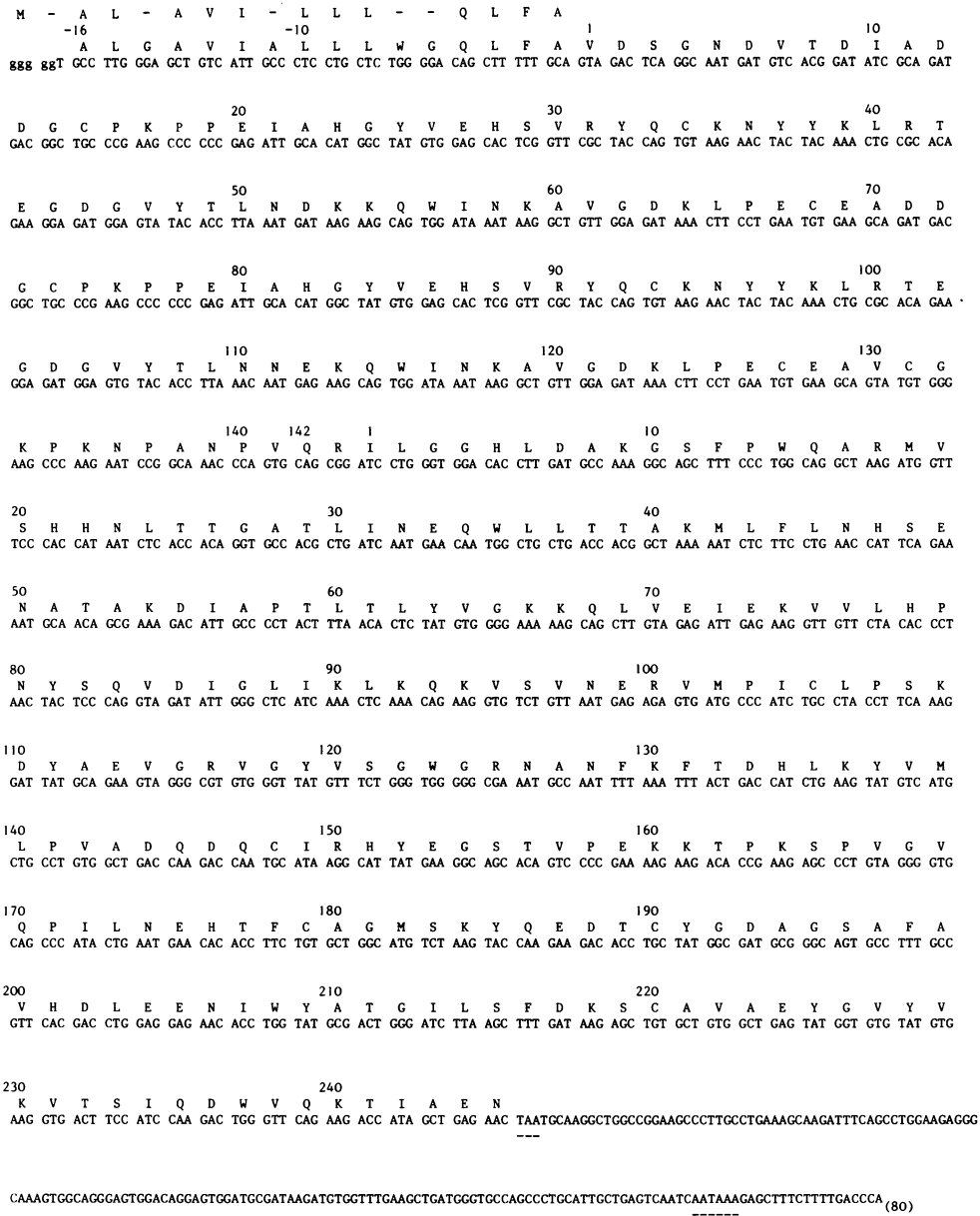

Fig. 2: DNA sequence of the clone hapto 6 (lower line). Upper line shows the deduced aminoacid sequence. The first line at the top of the figure shows the aminoacid sequence of the leader peptide of the rabbit hpα chain (8). Stop codon and polyadenylation side are underlined.

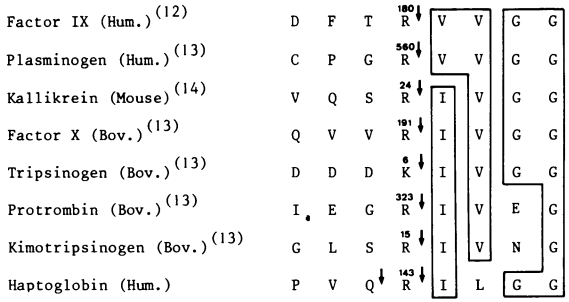

Fig. 3: Comparison of cleavage sites of several serine proteases with the presumptive cleavage site of the human haptoglobin. The cleavage sites are indicated by arrows.

intense as the others. As there are no EcoRI sites in T40 DNA, the three band pattern can be explained in two alternative ways. We can postulate the presence of an intron, containing an EcoRI site, in the genomic sequence corresponding to the T40 cDNA segment. In addition we have to postulate a polymorphism of the two parental alleles involving an EcoRI site. According to this hypothesis, band a represents the EcoRI segment common to both alleles; bands b and c represent EcoRI segments which, as a consequence of polymorphism, are different in the two alleles.

An alternative explanation is that there might be two (or more) copies of the haptoglobin gene per haploid genome. In order to explain the presence of three EcoRI segments in this case, we have to postulate a polymorphism at an EcoRI site. Both hypothesis imply the existence of polymorphism at this locus. This could be related to the several different phenotypes that are known to be derived from the haptoglobin locus in humans. In order to examine this question and to distinguish between the single gene and the multiple gene hypothesis, Southern blot analysis was carried out on DNA samples obtained from 4 individuals. Analysis of their haptoglobin subtype revealed that individual 1 has a Hp2-2 phenotype, individuals 2 and 3 have a Hp2-F phenotype and individual 4 has a HpF-S phenotype. We used 3 different probes corresponding to i) the entire hpα^2 chain (Pst-Bam fragment), ii) to part of the hpβ chain (Bam-Hind fragment) and iii) to the last 30 aminoacids of the hpβ chain followed by the 3' untranslated flanking region and the polyA (Hind-Pst fragment). If an intron, containing an EcoRI site, were present in the region covered by the T40 sequence, then the Pst-Bam, Bam-Hind and Hind-Pst fragments should not hybridize to the same genomic DNA fragments. On the contrary, if an identical pattern is found with the three fragments, then the

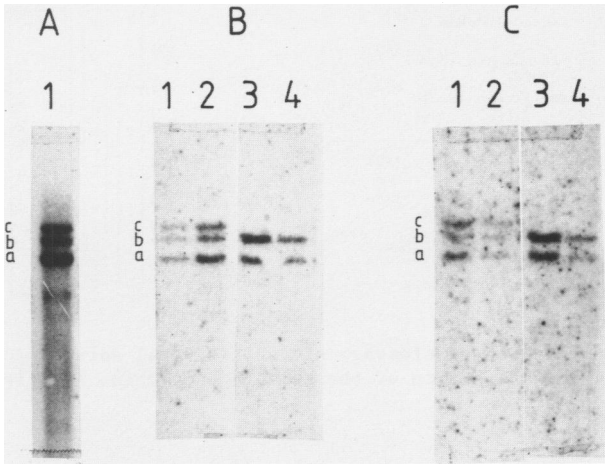

Fig. 4: Southern blot analysis. The hybridization of the 32P-labelled T40 DNA with the EcoRI digested DNA from individual 1 is shown in panel A. The hybridization pattern of the 32P-labelled PstI-Bam fragment with DNA from individuals 1 to 4 is shown in panel B. Panel C shows the hybridization pattern of the 32P-labelled Bam-Hind fragment with the DNA from individuals 1 to 4.

presence of an intron (containing an EcoRI site) is highly unlikely. The results obtained with the first two probes are shown in fig. 4, panels B and C (the Hind-Pst fragment gave an identical result, data not shown). The identity of the pattern obtained using three different probes indicates the absence of introns with EcoRI sites within the region covered by the probes. This fact strongly supports the hypothesis of the existence of at least two copies of the haptoglobin gene per haploid genome. The difference between the EcoRI patterns of individuals 1 and 2 (three bands) and 3 and 4 (two bands) is necessarily a consequence of polymorphism at one EcoRI site.

This polymorphism however, does not appear to be correlated to the haptoglobin phenotype since individuals with the same phenotype show different patterns (individuals 2 and 3) and individuals with different phenotypes show identical patterns. Classical genetic studies of the haptoglobin variants suggest the existence of only two alleles expressed in an individual. In this case we would postulate that some of our EcoRI segments contain copies of haptoglobin genes which are either not expressed (pseudogenes) or expressed at particular stages of development.

During the preparation of this manuscript we learned that Vander Straten et al. have independently cloned and sequenced a human haptoglobin cDNA (16).

The two sequences are almost identical: the only aminoacid difference is at residue 206 of the beta chain, where Vander Straten et al. find an Asp instead of an Asn.

MATERIALS AND METHODS

Bacterial strains, plasmids and phage vectors

Escherichia coli K12 (strain 71/18) was used for transformation (17). The M13 derivatives mp8 and mp9 (18) and plasmids pEMBL8 and pEMBL9 (19) were used as vectors. Transformation and preparation of double stranded DNA were as described (20,21).

Enzymes and chemicals

T4 DNA ligase and restriction endonucleases were gifts of Dr. V. Pirrotta. The 32P-labelled compounds were purchased from Amersham Buchler, Braunschweig. DNA Polymerase 1-Large fragment from E.coli was purchased from Biolabs.

Identification of human haptoglobin cDNA clones

The 126 bp fragment coding for hpβ chain was 32P-labelled as described by Hu and Messing (22) and used to screen a human liver cDNA library kindly provided by Dr. Woods (23).

DNA sequence

Sequence analysis was done using both the dideoxy (24) and Maxam and Gilbert methods (25).

Southern and northern blot analysis

DNA from peripheral blood was digested with EcoRI, size fractionated on a 1% agarose gel and transferred to nitrocellulose paper (26). Filters were hybridized to 32P-labelled fragments with a specific activity of 2x10/8 cpm/ µgr of DNA at 65°C for 48 hours in 10x Denhardt, 4x SET (20x SET: 3M NaCl, 0.6M Tris pH8, 40 mM EDTA), 0.1% SDS, 50 µgr/ml heat denaturated and sonicated salmon sperm DNA. After the hybridization filters were washed in the same solution at 65°C for 30 minutes, twice in 2x SET, 0.1% SDS and exposed at -80°C with preflashed Kodak films. RNA extraction and northern blot analysis were done as previously described (10).

Haptoglobin subtyping

Haptoglobin subtyping was carried out by polyacrylamide gel electrophoresis as described elsewhere (27).

ACKNOWLEDGEMENTS

We thank Wendy Moses for typing this manuscript. G.R. and G.B. have a grant from Ente Farmacologico Italiano (EFI), Naples, Italy, V.C. was

supported by a long-term EMBO fellowship, C.S. by a grant from Technosystems, Naples, Italy, and F.C. is the recipient of a training contract GBI-047-D from the "Biomolecular Engineering Programme" of the Commission of the European Communities.

The work done in Italy was financed by CNR, Progetto Finalizzato Salute dell'Uomo, Rome, Italy and by a grant from Farmitalia Spa., Milan, Italy.

[1]On leave of absence from Istituto di Chimica Biologica, II Facolta di Medicina, University of Naples, Naples, Italy

[2]On leave of absence from Istituto di Biologia dello Sviluppo, Consiglio Nazionale delle Ricerche, Palermo, Italy

[3]On leave of absence from Istituto di Biologia, Facolta di Medicina, University of Turin, Turin, Italy

REFERENCES

1. Putnam,F.W. (1975) in Putnam,F.W. (ed.) The Plasma Proteins, vol. 2, Academic Press, N.Y. pp. 2-46.
2. Bowman,B.H. and Kurosky,A. (1982) Adv. Human Genet. 12, 189-454.
3. Pintera,J. (1971) Ser. Haematol. 4, No. 2, 1-183.
4. Kurosky,A., Barnett,D.R., Lee,T.H., Touchstone,B., Hay,R.E., Arnott,M.S., Bowman,B.H. and Fitch,W.M. (1980) Proc. Natl. Acad. Sci. USA 77, 3388-3392.
5. Giblett,E.R. and Brooks,L.E. (1963) Nature 197, 576-577.
6. Pastewka,J.V., Reed,R.A., Ness,A.T. and Peacock,A.C. (1973) Anal. Biochem. 51, 152-162.
7. Chung,J.H. and Shim,B.-S. (1976) Korean J. Biochem. 8, 19-27.
8. Chow,V., Murray,R.K., Dixon,J.D. and Kurosky,A. (1983) FEBS Lett. 153, 275-279.
9. Haugen,T.H., Hanley,J.M. and Heath,E.C. (1981) J. Biol. Chem. 256, 1055-1057.
10. Costanzo,F., Castagnoli,L., Dente,L., Arcari,P., Smith,M., Costanzo,P., Raugei,G., Izzo,P., Pietropaolo,T.C., Bougueleret,L., Cimino,F., Salvatore,F. and Cortese,R. (1983) EMBO J. 2, 57-61.
11. Proudfoot,N.J. and Brownlee,G.G. (1976) Nature 263, 211-214.
12. Kurachi,K. and Davie,E.W. (1982) Proc. Natl. Acad. Sci. USA 79, 6461-6464.
13. Dayhoff,M.O. (ed.) Atlas of Protein Sequence and Structure (1972-82), published by National Biomedical Research Foundation, Georgetown University Medical Center Washington,DC.
14. Mason,A.J., Evans,H.A., Cox,D.R., Shine,J. and Richards,R.I. (1983) Nature 303, 300-307.
15. Giblett,E.R. (1969) "Genetic Markers in Human Blood" Blackwell, Oxford, England.
16. Vander Straten,A., Herzog,A., Jacobs,P., Cabezon,T. and Bullen,A. (1983) EMBO J. 2, 1003-1007.
17. Gronenborn,B. and Messing,J. (1978) Nature 272, 375-377.
18. Messing,J. (1981) in Walton,A. (ed.) Third Cleveland Symposium on Macromolecules, Recombinant DNA, Elsevier 143-153.
19. Dente,L., Cesareni,G. and Cortese,R. (1983) Nucleic Acids Res. 6, 1645-1655.

20. Cortese,R., Melton,D.A., Tranquilla,T. and Smith,J.D. (1978) Nucleic Acids Res. 5, 4593-4611.
21. Cortese,R., Harland,R. and Melton,D.A. (1980) Proc. Natl. Acad. Sci. USA 77, 4147-4151.
22. Hu,N.T. and Messing,J. (1982) Gene 17, 271-277.
23. Woods,D.E., Markham,A.F., Ricker,A.T., Goldberg,G and Colten,H.R. (1982) Proc. Natl. Acad. Sci. USA 79, 5661-5665.
24. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
25. Maxam,A.M. and Gilbert,W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
26. Southern,E.M. (1975) J. Mol. Biol. 98, 5503-5517.
27. Santoro,C., Boccazzi,C., Carbonara,A.O. (1982) Hum. Genet. 60, 287-288.