The EMBO Journal Vol.2 No.6 pp.1003 – 1007, 1983

Molecular cloning of human haptoglobin cDNA: evidence for a single mRNA coding for α^2 and β chains

A. vander Straten, A. Herzog, P. Jacobs, T. Cabezón[1] and A. Bollen*

Laboratory of Genetics, University of Brussels, B-1640 Rhode-St-Genèse, and [1]Smith-Kline/RIT, B-1330 Rixensart, Belgium

Communicated by R. Thomas
Received on 23 March 1983

Human haptoglobin (Hp) is a plasma glycoprotein composed of α and β polypeptide chains that are covalently associated by disulfide bonds. It had been suggested that α and β polypeptides could be synthesized *via* a common precursor polypeptide. We report the molecular cloning of DNA complementary to human Hp mRNA. One of the clones, pULB1148, carries a full length copy coding for both α^2 and β polypeptides. *In vitro* translation of human liver mRNA hybridizing with this cDNA gives a protein mol. wt. of 49 000 daltons. The sequence of the $\alpha^2\beta$ cDNA shows the presence of a single Arg residue between Gln 142 of the α^2 chain and Ileu 1 of the β chain. With a few minor exceptions, the DNA sequence fits the previously published amino acid sequences. The differences are the presence of an Asp residue at position 52 of α^2 instead of Asn, the existence in β of only one Lys residue between Gly 65 and the following Gln, the presence of Ser and Cys at positions 218 – 219 instead of Cys-Ser, and of Asp residues at positions 205 and 235 instead of Asn.

Key words: haptoglobin/cloning/sequence/post-translational processing/chromosome 16

Introduction

Haptoglobin (Hp) is a serum α_2-glycoprotein present in humans and other mammals (Putnam, 1975). One of its functional roles is to protect the kidneys from tissue destruction by binding free hemoglobin (Hb) following hemolysis; indeed the binding of Hb to Hp, which occurs in the ratio of one Hb$\alpha\beta$ subunit per Hp$\alpha\beta$ subunit, forms an effectively irreversible complex rapidly taken up by liver cells and digested (Putnam, 1975; Higa *et al.*, 1981). Human Hp is polymorphic and is found in one of three major phenotypes: Hp1-1, Hp2-2 and Hp2-1. Inherited variations in the smaller subunit, the α chain, are responsible for this polymorphism; indeed the α chain occurs in two major allelic forms, α^1 with 83 residues and α^2, which is a partial gene duplication, with 142 residues. Common variants in the α family are referred to as α^F and α^S; they differ by Lys/Glu amino acid substitutions at positions 53 for α^1 or 53 and 112 for α^2 (Bowman and Kurosky, 1982). The variants of α chain have been mapped on chromosome 16 by Robson *et al.* (1969) and Magenis *et al.* (1970). On the other hand, the β chain is considered to contain 245 amino acids and only a few variants have been described (Bowman and Kurosky, 1982). The chromosomal localisation of the β chain and its linkage with the α gene are still under discussion.

The α polypeptides of Hp are homologous to the fifth kringle region of plasminogen (Kurosky *et al.*, 1980), the β polypeptide being homologous to the serine proteases family (Kurosky *et al.*, 1980). On the basis of these observations, Kurosky *et al.* (1980) suggested that the α and β polypeptides could be synthesized as a single chain subsequently processed. Further evidence supporting this hypothesis came from the observation by Chow *et al.* (1980), Haugen *et al.* (1981) and Costanzo *et al.* (1983) that the mRNA encoding the β chain in rats and humans was translated *in vitro* into a polypeptide of sufficient mol. wt. to contain both α and β sequences.

We report here the cloning of a cDNA sequence coding for both the α^2 and β chains in tandem. The DNA sequence indicates that processing of the $\alpha^2\beta$ precursor polypeptide is accompanied by the release of an isolated Arg residue.

Results

Human liver mRNA was prepared and fractionated according to the procedure described in Materials and methods. Aliquots of the individual mRNA fractions were used for *in vitro* translation experiments to identify those coding for proteins of ~50 000 daltons. mRNA from those fractions (of ~1.5 kb) was used to prepare double-stranded cDNA. The cDNA was fractionated by size, and material >1.1 kb was cloned into pBR322.

Identification of bacterial clones containing Hp cDNA sequences

Plasmid DNA from the cDNA library was prepared as described in Materials and methods. Pools of DNAs issued from groups of four clones were bound onto nitrocellulose and hybridized to total human liver mRNA. The specifically bound mRNA was translated *in vitro* and analysed by

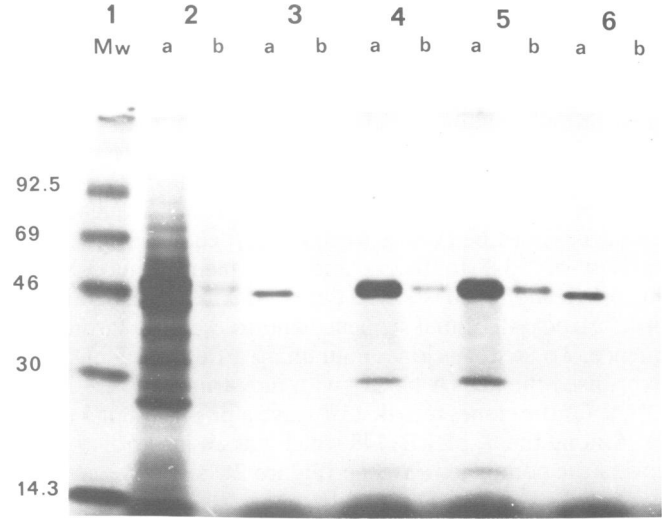

Fig. 1. Immunoprecipitation of the Hp $\alpha^2\beta$ precursor from the translation products of liver mRNA selected by hybridization to different cDNA clones. **Lane 1**: mol. wt. markers. Sizes are given in kd (92 500/69 000/46 000/30 000/14 300). **Lane 2, a** and **b**: total liver mRNA. **Lane 3, a** and **b**: mRNA selected by hybridization with F3-3. **Lane 4, a** and **b**: mRNA selected by hybridization with F3-1. **Lane 5, a** and **b**: mRNA selected by hybridization with F3-25. **Lane 6, a** and **b**: mRNA selected by hybridization with F3-29. **Lanes a**: 2 μl of translation products of the mRNA. **Lanes b**: immunoprecipitate from 10 μl of the translation mix.

*To whom reprint requests should be sent.

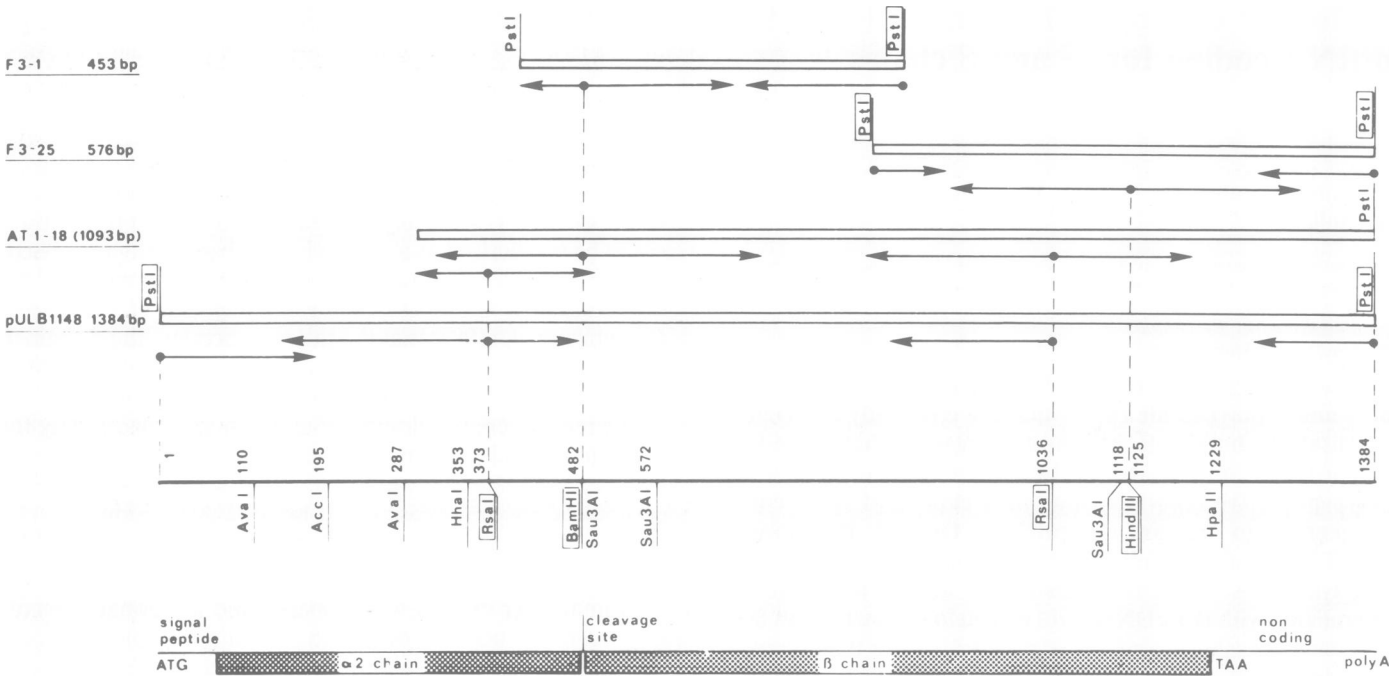

Fig. 2. Sequencing strategy of Hp cDNA clones. *PstI* inserts of clones F3-1, F3-25, AT1-18 and pULB1148 used for sequencing are represented. The restriction sites used for labeling are boxed. the arrows indicate the sequenced regions. The mRNA structure is shown. Base number 1 is the adenine of the initiation codon. The left side *PstI* site of clone AT1-18 has not been regenerated.

immunoprecipitation and SDS-polyacrylamide gel electrophoresis. The same procedure was followed with DNAs of individual clones from pools giving a positive signal (Figure 1). As expected, the mol. wt. of the products was ~50 000 daltons. Systematic screening of the proteins synthesized by the hybridized mRNAs was performed with several antisera raised against human proteins. Among them was a commercial (Immuno) anti-antithrombin III serum which gave positive results with the translation product of mRNA hybridized with DNA from clones F3-1 and F3-25 (Figure 1, lanes 4 and 5). Restriction maps of cDNA inserts of F3-1 and F3-25 were established (Figure 2) and their base sequences, determined by the Maxam and Gilbert procedure, were translated in all possible frames. Comparison with the antithrombin III amino acid sequence revealed no homology at all; we thus decided to compare the translated sequences of F3-1 and F3-25 to a local collection of protein sequences and found F3-25 to be coding for the Hp β chain. Further comparison of F3-1 to the α and β sequences published by Kurosky *et al.* (1980) showed that F3-1 was homologous to α^1 or α^2 carboxy-terminal sequence and to β amino-terminal sequence. To identify clones containing a full length cDNA we hybridized the cDNA library with nick-translated F3-1 insert DNA. Of the clones tested, 15% gave a positive signal (Figure 3). Among them, pULB1148 which has an insert of ~1.4 kb covers the complete $\alpha^2\beta$ gene (Figure 2).

Sequence of the complete $\alpha^2\beta$ cDNA

The complete sequence of the $\alpha^2\beta$ cDNA deduced from the sequences of overlapping fragments (Figures 2 and 4) shows that α^2 and β sequences are contiguous. The open reading frame starts with an initiation triplet (ATG) 18 codons upstream from the triplet encoding the first amino acid of the mature α^2 protein. The first stop signal (TAA) is found at the 3′ end of the β encoding sequence. The polyadenylation signal AATAAA (Proudfoot and Brownlee, 1976) is found in the cDNA sequence at position −23 from the poly(A) tract. Moreover, the α^2 sequence is separated from the β sequence by three bases coding for an Arg residue; both sequences are thus in the same reading frame.

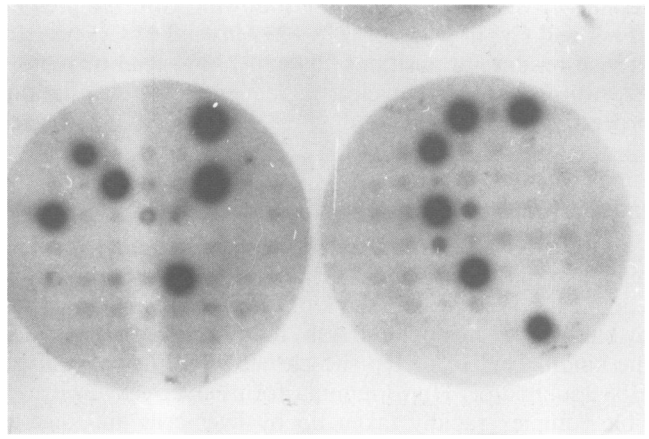

Fig. 3. Colony hybridization with ^{32}P-labeled cDNA of clone F3-1. 15% of the clones tested gave a positive signal.

Comparison of our translated sequence with published amino acid sequences (Kurosky *et al.*, 1980; Bowman and Kurosky, 1982) revealed a number of differences; indeed we found residues 52 of α^2 to be Asp instead of Asn; residues 218 and 219 of β to be, respectively, Ser and Cys instead of Cys-Ser, and residues 205 and 235 to be Asp instead of Asn. There is only one Lys residue between Gly 65 and the following Gln instead of two, so that the β chain seems to contain only 244 amino acids. Moreover, as the amino acids 53 and 112 of α^2 are respectively lysine and glutamate, the cDNA we have cloned codes for the $\alpha^{2(FS)}$ variant.

-18 -1

MetSerAlaLeuGlyAlaValIleAlaLeuLeuLeuTrpGlyGlnLeuPheAla

GGGGGGATGAGTGCCTTGGGAGCTGTCATTGCCCTCCTGCTCTGGGGACAGCTTTTTGCA

1 30

ValAspSerGlyAsnAspValThrAspIleAlaAspAspGlyCysProLysProProGluIleAlaHisGlyTyrValGluHisSerVal

GTGGACTCAGGCAATGATGTCACGGATATCGCAGATGACGGCTGCCCGAAGCCCCCCGAGATTGCACATGGCTATGTGGAGCACTCGGTT

31 60

ArgTyrGlnCysLysAsnTyrTyrLysLeuArgThrGluGlyAspGlyValTyrThrLeuAsnAspLysLysGlnTrpIleAsnLysAla

CGCTACCAGTGTAAGAACTACTACAAACTGCGCACAGAAGGAGATGGAGTATACACCTTAAATGATAAGAAGCAGTGGATAAATAAGGCT

61 90

ValGlyAspLysLeuProGluCysGluAlaAspAspGlyCysProLysProProGluIleAlaHisGlyTyrValGluHisSerValArg

GTTGGAGATAAACTTCCTGAATGTGAAGCAGATGAGGGCTGCCCGAAGCCCCCCGAGATTGCACATGGCTATGTGGAGCACTCGGTTCGC

91 120

TyrGlnCysLysAsnTyrTyrLysLeuArgThrGluGlyAspGlyValTyrThrLeuAsnAsnGluLysGlnTrpIleAsnLysAlaVal

TACCAGTGTAAGAACTACTACAAACTGCGCACAGAAGGAGATGGAGTGTACACCTTAAACAATGAGAAGCAGTGGATAAATAAGGCTGTT

121 142 1 7

GlyAspLysLeuProGluCysGluAlaValCysGlyLysProLysAsnProAlaAsnProValGlnArgIleLeuGly-X-HisLeuAsp

GGAGATAAACTTCCTGAATGTGAAGCAGTATGTGGGAAGCCCAAGAATCCGGCAAACCCAGTGCAGCGGATCCTGGGTGNACACTTGGAT

8 37

AlaLysGlySerPheProTrpGlnAlaLysMetValSerHisHisAsnLeuThrThrGlyAlaThrLeuIleAsnGluGlnTrpLeuLeu

GCCAAAGGCAGCTTTCCTTGGCAGGCTAAGATGGTTTCCCACCATAATCTCACCACAGGTGCCAGCCTGATCAATGAACAATGGCTGCTG

38 67

ThrThrAlaLysAsnLeuPheLeuAsnHisSerGluAsnAlaThrAlaLysAspIleAlaProThrLeuThrLeuTyrValGlyLysGln

ACCACGGCTAAAAATCTCTTCCTGAACCATTCAGAAAATGCAACGGCGAAAGACATTGCCCCCACTTTAACACTCTATGTGGGAAAGCAG

68 97

LeuValGluIleGluLysValValLeuHisProAsnTyrSerGlnValAspIleGly-X-IleLysLeuLysGlnLysValSerValAsn

CTTGTAGAGATTGAGAAGGTTGTTCTACACCCTAACTACTCCCAAGTAGATATTGGGNTCATCAAACTCAAACAGAAGGTGTCTGTTAAT

98 127

GluArgValMet-X-Ile-X-LeuProSerLysAspTyrAlaGluValGlyArgValGlyTyrValSerGlyTrpGlyArgAsnAlaAsn

GAGAGAGTGATGNCCATCTGNCTACCATCCAAGGATTATGCAGAAGTAGGGCGTGTGGGTTATGTTTCTGGGTGGGGGCGAAATGCCAAT

128 157

PheLysPheThrAspHisLeuLysTyrValMetLeuProValAlaAspGlnAspGlnCysIleArgHisTyrGluGlySerThrValPro

TTTAAATTTACTGACCATCTGAAGTATGTCATGCTGCCTGTGGCTGACCAAGACCAATGCATAAGGCATTATGAAGGCAGCACAGTCCCC

158 187

GluLysLysThrProLysSerProValGlyValGlnProIleLeuAsnGluHisThrPheCysAlaGlyMetSerLysTyrGlnGluAsp

GAAAAGAAGACACCGAAGAGCCCTGTAGGGGTGCAGCCCATACTGAATGAACACACCTTCTGTGCTGGCATGTCTAAGTACCAAGAAGAC

188 217

ThrCysTyrGlyAspAlaGlySerAlaPheAlaValHisAspLeuGluGluAspThrTrpTyrAlaThrGlyIleLeuSerPheAspLys

ACCTGCTATGGCGATGCGGGCAGTGCCTTTGCCGTTCACGACCTGGAGGAGGACACCTGGTATGAGACTGGGATCTTAAGCTTTGATAAG

218 244

SerCysAlaValAlaGluTyrGlyValTyrValLysValThrSerIleGlnAspTrpValGlnLysThrIleAlaGluAsnSTOP

AGCTGTGCTGTGGCTGAGTATGGTGTGTATGTGAAGGTGACTTCCATCTAGGACTGGGTTCAGAAGACCATAGCTGAGAACTAATGCAAG

GCTGGCCGGAAGCCCTTGCCTGAAAGCAAGATTTCAGCTGGAAGAGGGCAAAGTGGACGGGAGTGGACAGGAGTGGATGCGATAAGATGT

GGTTTGAAGCTGATGGGTGCCAGCCCTGCATTGCTGAGTCAATCAATAAAGAGCTTTCTTTTGACCCpoly(A)

Fig. 4. Complete nucleotide sequence of $\alpha^2\beta$cDNA. The *Pst*I inserts of F3-1, F3-25, AT1-18 and pULB1148 were sequenced by the method of Maxam and Gilbert (1977). The Arg residue and the peptide signal released during the proteolytic maturation of the precursor polypeptide are underlined. The amino acid sequence is derived from our DNA sequence. Numbering of the amino acids is identical to the published one (Kurosky *et al.*, 1980) until residue 65 of β and −1 after that residue. This results from the existence of only one Lys residue, in our sequence, between Gly 65 and the following Gln.

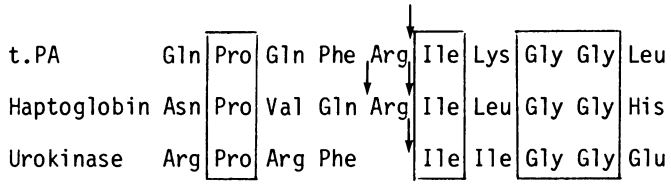

Fig. 5. Comparision of the cleavage regions of tissue plasminogen activator (t.PA), Hp and urokinase. The homologous regions are boxed. The cleavage site is indicated.

Discussion

Here, we show that both α^2 and β Hp chains are synthesized initially as a single polypeptide encoded by a single mRNA. This conclusion is based on the cloning and characterization of a cDNA which carries, in a single continuous open reading frame, the information coding for a signal peptide (18 amino acids) and for both α^2 and β Hp chains. The open reading frame starts with an ATG codon and stops with a TAA nonsense triplet. *In vitro* translation of the mRNA which hybridizes specifically with Hp cDNA yields a polypeptide of mol. wt. ~49 000 daltons. The length of this polypeptide is sufficient to encompass both α^2 and β chains. This polypeptide is immunoprecipitable by either rabbit (Behring) or goat (Immuno) antithrombin III antisera. This could be due to contamination of antithrombin III preparations by Hp or, more probably, to cross-reaction with Hp of antibodies effectively raised against antithrombin III. The demonstration that α^2 and β polypeptides are encoded by a single mRNA, taken together with the location of the α-coding sequence on chromosome 16 (Robson *et al.*, 1969; Magenis *et al.*, 1970), strongly suggests the existence of a single Hp gene coding for both polypeptides located on chromosome 16. Definite proof of this assumption could be obtained from hybridization experiments on mouse/human somatic cell hybrids that segregate human chromosomes. From the comparison of the translated sequence with the α^2 and β amino acid sequences published by Bowman and Kurosky (1982), we suggest the inversion of Ser and Cys at positions 218 and 219 of β and the presence of Asp residues at positions 52 of α^2 and 205 and 235 of β instead of Asn. Moreover, we find only one Lys residue between Gly 65 and Gln 67. As the ATG initiation codon of pULB1148 is contiguous to the oligo(dG) tail of the clone, it could be artefactual and the signal peptide could be longer. Cleavage between Gln 142 of α^2 and Ileu 1 of β is accompanied by the release of an isolated Arg residue located between them in the precursor. Comparison of the cleavage region of Hp precursor with those of tissue plasminogen activator (Pennica *et al.*, 1983) and urokinase (Steffens *et al.*, 1982; Günzler *et al.*, 1982) (Figure 5) shows a high degree of homology and suggests that the urokinase precursor could have a single Arg residue between the 18-K and 33-K polypeptides. Cleavage of the precursor of serine proteases activates the enzyme; while Hp lost any proteolytic function and gained a new one, it has conserved a high degree of homology with serine proteases not only in its structure but also in its processing mechanism. It is tempting then to speculate whether cleavage of the precursor of Hp also regulates its activity and, hence, whether this precursor would be able to bind Hb.

Materials and methods

RNA preparation

Human liver samples were obtained from kidney transplant donors and immediately frozen in liquid nitrogen. Poly(A)$^+$ RNAs have been prepared by the guanidium chloride method (Chirgwin *et al.*, 1979) followed by oligo(dT) cellulose chromatography. The liver poly(A)$^+$ RNAs were fractionated by centrifugation through linear sucrose gradients (10–30% sucrose in Tris-HCl 10 mM pH 7.5; EDTA 5 mM and SDS 1%) 17 h at 27 000 r.p.m. and 20°C in a Beckman SW41 rotor.

Synthesis and cloning of cDNA

Reverse transcription mixtures containing 6 μg RNA were primed with oligo(dT)$_{12-18}$ (Boehringer). Double-stranded cDNA was prepared, enriched in large size molecules (>1.1 kb) by sucrose gradient centrifugation (10–30% sucrose in Tris-HCl 10 mM pH 7.5; NaCl 10 mM; 20 h at 41 000 r.p.m. in a SW41 Beckman rotor), oligo(dC) tailed, inserted into *Pst*I-cleaved dG-tailed pBR322 and transformed into *Escherichia coli* MM294 as described in Lawn *et al.* (1981).

Hybridization selection of Hp mRNA

Human liver mRNA was hybridized to plasmid DNAs purified by CsCl gradient centrifugation and bound to nitrocellulose filters as described by Forde *et al.* (1981). Specifically bound mRNA (at a stringency of 52°C in 65% formamide, 10 mM PIPES pH 6.4 and 0.4 M NaCl) was eluted with boiling water and translated in the reticulocyte lysate system (NEN) using [^{35}S]methionine as tracer. Translation products were analysed by immunoprecipitation and SDS-polyacrylamide gel electrophoresis (Laemmli, 1970).

In situ hybridization of bacterial colonies

E. coli transformants were screened for the presence of Hp-encoding sequences by hybridization of colonies transferred onto nitrocellulose fiters (Sartorius type SM50) with [^{32}P]nick-translated DNA as described by Grunstein and Hogness (1975).

DNA sequencing

DNA sequence determination was performed by the method of Maxam and Gilbert (1977).

Acknowledgements

We thank Dr. M. De Wilde (Smith-Kline/RIT) for helpful discussions and generous hospitality in his laboratory, Dr. C. Wuilmart who gave us access to his protein sequences collection, F. Fievet for introducing us to the Maxam and Gilbert sequencing techniques and C. De Buyl for typing the manuscript. A. vd .S. is an IRSIA grantee and A.B. is Maître de Recherches at the Fonds National de la Recherche Scientifique. This work was financially supported by Smith-Kline/RIT.

References

Bowman,B.H. and Kurosky,A. (1982) *Adv. Human Genet.*, **12**, 189-454.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.
Chow,V., Kurosky,A. and Murray,R.K. (1980) *J. Cell Biol.*, **87**, 296.
Costanzo,F., Castagnoli,L., Dente,L., Arcari,P., Smith,M., Costanzo,P., Rangli,G., Izza,P., Pietropaolo,T.C., Bougueleret,L., Cimino,F., Salvatore,F. and Cortese,R. (1983) *EMBO J.*, **2**, 57-61.
Forde,B.G., Kreis,M., Bahramian,M.B., Matthews,J.A., Miflin,B.J., Thompson,R.D., Bartels,D. and Flavell,R.B. (1981) *Nucleic Acids Res.*, **9**, 6689-6707.
Grunstein,M. and Hogness,D. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 3961-3965.
Günzler,W.A., Steffens,G.J., Otting,F., Kim,S.M.A., Frankus,E. and Flohé,L. (1982) *Hoppe-Seyler's Z. Physiol. Chem.*, **363**, 1155-1165.
Haugen,T.M., Hamley,J.M. and Heath,E.C. (1981) *J. Biol. Chem.*, **256**, 1055-1057.
Higa,Y., Oshiro,S., Kino,K., Tsunoo,H. and Nakajuma,H. (1981) *J. Biol. Chem.*, **256**, 12322-12328.
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.
Laemmli,U.K. (1970) *Nature*, **227**, 680-685.
Lawn,R.M., Adelman,J., Bock,S.C., Franke,A.E., Houck,C.M., Najarian, R.C., Seeburg,P.H. and Wion,K.L. (1981) *Nucleic Acids Res.*, **9**, 6103-6114.
Magenis,R.E., Hecht,F. and Lovrin,E.W. (1970) *Science (Wash.)*, **170**, 85-87.

Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 560-564.
Pennica,D., Holmes,W.E., Kohr,W.J., Harkins,R.N., Veheer,G.A., Ward, C.A., Bennett,W.F., Yelverton,E., Seeburg,P.M., Heyneker,M.L. and Goeddel,D.V. (1983) *Nature,* **301**, 214-221.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature,* **263**, 211-214.
Putnam,F.W. (1975) in Putnam,F.W. (ed.), *The Plasma Proteins,* Academic Press, NY, vol. **2**, pp 1-50.
Robson,E.B., Palani,P.E., Dart,S.J., Jacobs,P.A. and Renwick,J.H. (1969) *Nature,* **223**, 1163-1165.
Steffens,G.J., Günzler,W.A., Otting,F., Frankus,E. and Flohé,L. (1982) *Hoppe-Seyler's Z. Physiol. Chem.,* **363**, 1043-1058.

Note added in proof

The existence of only one lysine residue between Gly 65 and the following Gln, observed in clones F 3-1 and F 3-25, has not been confirmed in the corresponding region of clone pULB1148, region which had not yet been sequenced (Figure 2) when the manuscript was submitted.