J Mol Evol (1988) 27:102–108

Journal of
Molecular Evolution

© Springer-Verlag New York Inc. 1988

Structure of the Human Hemopexin Gene and Evidence for Intron-Mediated Evolution

F. Altruda, V. Poli, G. Restagno, and L. Silengo

Cattedra di Biologia Generale, Via Santena, 5 bis, 10126 Torino, Italy

Summary. The human hemopexin gene was isolated and its structure determined. The gene spans approximately 12 kb and is interrupted by nine introns. When the intron/exon pattern was examined with respect to the polypeptide segments they encode, a direct correspondence between exons and the 10 repeating units in the protein was observed. The introns are not randomly placed; they fall in the middle of the region of amino acid sequence homology in strikingly similar locations in 6 of the 10 units and in a symmetrical position in the two halves of the coding sequence. These features strongly support the hypothesis that the gene evolved through intron-mediated duplications of a primordial sequence to a five-exon cluster. A more recent gene duplication led to the present-day gene organization.

Key words: Hemopexin — Evolution — Gene structure — Recombination — Sequence homology

Introduction

Hemopexin is a heme-binding plasma glycoprotein composed of a single polypeptide chain 439 amino acids long (Altruda et al. 1985, Takahashi et al. 1985). It is produced only in the liver and it is secreted into the bloodstream where it plays an important role in heme disposal, facilitating the conservation of iron by preventing urinary excretion (Muller-Eberhard and Liem 1974). The internalized iron is stored in the hepatic cells and the first event involves the interaction of the hemopexin with a specific receptor on the hepatic cell surface (Smith and Morgan 1979). Hemopexin plasma levels increase during acute infections suggesting that its biosynthesis is subject to control mechanisms responsible for the acute phase reaction (Baumann et al. 1983). Moreover, there are developmental variations in the pattern of gene expression since hemopexin is made by the fetus and newborn, and its concentration does not approach the adult level until after the first year. For these reasons we have begun to clone the human hemopexin gene in order to understand the molecular mechanisms responsible for the regulation of its expression. We previously characterized the human hemopexin cDNA (Altruda et al. 1985) showing that the coding region is composed of 10 repeating units that are structurally homologous. On the basis of these data we suggested that the gene might have evolved through a series of duplications; in this paper we report the genomic organization of the hemopexin gene and present a model for the evolution of the present-day sequence arrangement.

Offprint requests to: F. Altruda

Materials and Methods

Bacterial Strains, Plasmids, and Phage Vectors. Escherichia coli K12 (strain 71/18) was used for transformation (Messing et al. 1977). The M13 derivatives mp 8, mp 9 (Messing and Vieira 1982), and pUC18-19 (Vieira and Messing 1982) were used as phage and plasmid vectors for subcloning and sequencing.

Enzymes and Chemicals. Restriction endonucleases, T4 DNA ligase, DNA polymerase I holoenzyme, and Klenow fragment were purchased from Boehringer. ^{32}P-labeled compounds were purchased from Amity-PG.

Screening of the Human Genomic Library. The human genomic library was kindly provided by G. Bensi (Bensi et al. 1985)

Fig. 1. Structural maps of the human hemopexin gene clones. Restriction sites: S = SalI; H = HindIII; P = PstI; R = EcoRV; B = BamHI; E = EcoRI. Solid bars indicate exons; solid lines represent introns (I–IX) and flanking DNA; open boxes are vector sequences. L.A. = left arm; R.A. = right arm.

and screened according to Frischauf (Frischauf et al. 1983). Nick-translated PstI cDNA of Hpx11 clone was used as a probe in the screening.

Blot Hybridization and Nucleotide Sequence Analysis. Restriction maps and Southern blot analysis were performed as previously described (Altruda et al. 1985). The strategy utilized to locate the desired regions for sequencing (intron–exon junctions and all exonic segments) was to use specific radiolabeled Hpx11 cDNA subfragments that represented the entire cDNA. By digesting with several different enzymes, small hybridizing fragments were identified as suitable for direct sequencing. Sequence analysis was carried out by the dideoxy method (Sanger et al. 1977).

S1 Mapping. S1 mapping was performed using, as a probe, a HindIII–EcoRV fragment 730 bp long consisting of 134 bp of the first intron, 83 bp of the first exon, and 513 bp of the 5′ flanking region of the gene. The coding strand of this fragment was cloned in the M13 mp19 phage vector, and the labeling was carried out by priming with the M13 sequencing primer and elongating the anticoding strand with the Klenow enzyme in the presence of dGTP, dTTP, ^{32}P-dATP, and ^{32}P-dCTP. Labeled probe (10^6 cpm) was hybridized to 15 μg of human liver total RNA. The experimental procedure for DNA–RNA hybridization and for S1 digestion with S1 endonuclease was according to Berk and Sharp (1977).

Results

Gene Structure

We isolated a complete coding sequence cDNA clone (Hpx11) for human hemopexin (Altruda et al. 1985). The results of Southern blot analysis of genomic DNA probed with Hpx11 cDNA indicate that there is only one copy of the hemopexin gene in the human haploid genome and it is entirely included within a single 12-kb PstI fragment.

About 3 × 10^5 plaques of a human DNA library in the EMBL3 phage vector (Bensi et al. 1985) were screened using the Hpx11 cDNA clone as a probe; two independent clones, λHpx3 and λHpx4, were isolated. Restriction maps of the two recombinant clones were determined (Fig. 1) and Southern blot analysis was carried out using various segments derived from the 5′, intermediate, and 3′ regions of the cDNA clone Hpx11 as probes. The results indicate that the two clones overlap but are not identical. λHpx3 does not contain the complete protein-coding region because no hybridization was detected with the 3′-specific probe (Fig. 1). On the other hand, the λHpx4 clone hybridized to both the 5′ and 3′ probe, indicating that the entire coding region is present and is contained in a 12-kb PstI fragment; this λHpx4 clone contains approximately 2 kb of the 5′ flanking DNA and 4 kb of the 3′ flanking DNA. Furthermore, in the region hybridizing to the cDNA probes, λHpx3 and λHpx4 show identical restriction patterns, supporting the hypothesis that there is a single gene. Coding regions were identified by restriction mapping and DNA sequencing of λHpx4 (Fig. 1). This analysis established the precise locations of nine introns ranging in size from 80 to 3000 bp that interrupt the Hpx gene. The following features were observed (Fig. 2):

1) There is a complete leader sequence 23 amino acids long contained in the first exon.
2) The optimal context for recognition of the ATG codon is contained in the sequence CAGCATGG: an adenine in position −3 (three nucleotides upstream of the ATG codon) and a guanine in position +4 (immediately following the ATG codon) are conserved for initiation sites in eukaryotes (Kozak 1986).
3) The nucleotide sequence of the 10 exons agrees with that determined previously for the hemopexin cDNA where their lengths range from 59 bp to 418 bp.
4) All donor and acceptor splice junctions conform to the GT/AG rule (Breathnach et al. 1978; Lerner et al. 1980).

To determine the initiation site for transcription, an S1 mapping experiment was carried out (Fig. 3). The results indicate that the adenine located 28 bas-

```
agctgccgggaaaaaggagtcttggattcagatttctgtccagaccctgaccttatttgcagtgatgtaatcagcca
atattggcttagtcctgggagacagcacattcccagtagagttggaggtggggtggtgctgctgccaactctatat
agggagttcaactggtcacccagagctgtcctgtggcctctgcagctcagcATGGCTAGGGTACTGGGAGCACCCGT
                                                    M  A  R  V  L  G  A  P  V
TGCACTGGGGTTGTGGAGCCTATGCTGGTCTCTGGCCATTGCCACCCCTCTTCCTCCgtgagtaaagctgggactag
  A  Q  G  L  W  S  L  C  W  S  L  A  I  A  T  P  L  P  P
aagcgaaggattgagttctgggctagggtaaggtagggccagttttaggcctcggtcaaatttggggtcaggggct

atgggaaagggatcggtcccaatggatcaagatatctatttgttctccctagGACTAGTGCCCATGGGAATGTTGC
                                                        T  S  A  H  G  N  V  A
TGAAGGCGAGACCAAGCCAGACCCAGACGTGACTGgtgaggcctgactccctaagtctgtcttatctgtctggttgt
  E  G  E  T  K  P  D  P  D  V  T
gtctctgcattttatcaccttctggttttttttttttttttttttttactttgccatctccctacctccacccca

gAACGCTGCTCAGATGGCTGGAGCTTTGATGCTACCACCCTGGATGACAATGGAACCATGCTGTTTTTTAAAGgtag
 E  R  C  S  D  G  W  S  F  D  A  T  T  L  D  D  N  G  T  M  L  F  F  K
tgtagggacgaggttagggtgctttaggaccagactac---0.12kb---ggccccccagtgagtgtgggtttttc

agGGGAGTTTGTGTGGAAGAGTCACAAATGGGACCGGGAGTTAATCTCAGAGAGATGGAAGAATTTCCCCAGCCCTG
  G  E  F  V  W  K  S  H  K  W  D  R  E  L  I  S  E  R  W  K  N  F  P  S  P
TGGATGCTGCATTCCGTCAAGGTCACAACAGTGTCTTTCTGATCAAGgtactgctgggccaaaatcagggccaggct
V  D  A  A  F  R  Q  G  H  N  S  V  F  L  I  K
ggaaagggctggaatcgacactggggacccttcccccaaatggccttggcatggagcccatagcaataggtagcaga

tttctttcccttgtgccctcctttcctgtaaaagcttgggctaagggagtgtgcatgcgtgtgggcctggcaggtgc

accatccagtggctgttcttcagtcctagtcttagttctacaccgctctgctgtacctcacactgctggccatcctt

tttttctctggcaattgcttccttgccttcatgacctgtataagtc---1.15kb---ggcccacaggaggatatcg

tataagcacagtagataaaaaatgtgtgtaaatgcagagtggcagtatctggggatgcacagtcaaaaagagagtac

ttttgaatgcagGGGGACAAAGTCTGGGTATACCCTCCTGAAAAGAAGGAGAAAGGATACCCAAAGTTGCTCCAAGA
             G  D  K  V  W  V  Y  P  P  E  K  K  E  K  G  Y  P  K  L  L  Q  D
TGAATTTCCTGGAATCCCATCCCCACTGGATGCAGCTGTGGAATGTCACCGTGGAGAATGTCAAGCTGAAGGCGTCC
  E  F  P  G  I  P  S  P  L  D  A  A  V  E  C  H  R  G  E  C  Q  A  E  G  V
TCTTCTTCCAAGgttcagtccaggctgg---0.61kb---ggcccatcctggaacccagagaaagcacaaaggaggg
L  F  F  Q
cgaacccggtctcaccaaatgcctgttgattgattggacaaagGTGACCGCGAGTGGTTCTGGGACTTGGCTACGGG
                                            G  D  R  E  W  F  W  D  L  A  T  G
AACCATGAAGGAGCGTTCCTGGCCAGCTGTTGGGAACTGCTCCTCTGCCCTGAGATGGCTGGGCCGCTACTACTGCT
  T  M  K  E  R  S  W  P  A  V  G  N  C  S  S  A  L  R  W  L  G  R  Y  Y  C
TCCAGGGTAACCAATTCCTGCGCTTCGACCCTGTCAGGGGAGAGGTGCCTCCCAGGTACCCGCGGGATGTCCGAGAC
F  Q  G  N  Q  F  L  R  F  D  P  V  R  G  E  V  T  P  R  Y  P  R  D  V  R  D
```

Fig. 2. Nucleotide sequence of the human hemopexin gene. The entire coding reigon (uppercase letters) was identified, although only portions of some introns (lowercase letters) were sequenced. The approximate lengths of the unsequenced intronic regions are indicated in kb. The proposed cap site is indicated by an arrow. The presumptive TATAA sequence at position −56, and the CCAAT sequence at position −131 are boxed; the polyadenylation signal is underlined. The amino acid sequence is given below the corresponding codons. The leader sequence is underlined. The region of introns that were not sequenced is indicated by dotted lines.

es upstream from the methionine codon is the most probable cap site. This conclusion is supported by the presence of a TATA-box-like sequence located 28 bp upstream from this putative start site.

The nucleotide sequences of the flanking regions (about 200 bp upstream and 160 bp downstream of the coding region) were also determined and the following consensus sequences were shown:

1) A "TATA"-like sequence (Antonorakis et al. 1984) was found 56 bp upstream from the ATG codon and 28 bp from the proposed sites of initiation of transcription; a CCAAT consensus sequence (Benoist et al. 1980) was detected at position −131.
2) The position and sequence of an 11-bp stretch TTATTTGCAGT (−153, −143) are similar to the consensus sequences found in mouse and human light-chain promoters of immunoglobulin genes (Falkner and Zachau 1984; Parslow et al. 1984). In these systems the decamer element seems to be essential for the efficient use of immunoglobulin in B cells, probably mediating specific binding of a nuclear factor present in B cells and also in HeLa cells (Singh et al. 1986).

Correlation of Hpx Gene and Protein Structure

The primary structure of human hemopexin was previously deduced from the cDNA sequence (Altruda et al. 1985). There were no internally repetitive DNA sequences but, based on amino acid homology we discovered two regions of similarity corresponding to the two halves of the protein. Moreover, by computer-assisted analysis of the amino acid sequence, the complete polypeptide chain

```
TACTTCATGCCCTGCCCTGGCAGAGgtgagaaagccctagcacttgagacctgtcagaattcatccactttccctga
 Y  F  M  P  C  P  G  R
gcttgtggatctcacgtgtcctagctctcactttaactccgtgttgcgacaccttggcc---2.8kb---ctgggga

aaccccacacccactaaccccagccatctcttccaccttggacctcactctgacctctggcctccttctgtgttctc

ctcacccatttctctctccagGCCATGGACACAGGAATGGGACTGGCCATGGGAACAGTACCCACCATGGCCCTGAG
                     G  H  G  H  R  N  G  T  G  H  G  N  S  T  H  H  G  P  E
TATATGCGCTGTAGCCCACATCTAGTCTTGTCTGCACTGACGTCTGACAACCATGGTGCCACCTATGCCTTCAGTGg
 Y  M  R  C  S  P  H  L  V  L  S  A  L  T  S  D  N  H  G  A  T  Y  A  F  S
tgagagatggcccccaactttccccaatgtgctcctcaca---1.5kb---atattggcctctggatgcattctctc

cctctcatgactcatttctgcattcatcactagcctcttctctgcctgggcttctgccagcggccctagagcaacct

atggtattccacagGGACCCACTACTGGCGTCTGGACACCAGCCGGGATGGCTGGCATAGCTGGCCCATTGCTCATC
              G  T  H  Y  W  R  L  D  T  S  R  D  G  W  H  S  W  P  I  A  H
AGTGGCCCCAGGGTCCTTCAGCAGTGGATGCTGCCTTTTCCTGGGAAGAAAAACTCTATCTGGTCCAGgtgtgtatt
Q  W  P  Q  G  P  S  A  V  D  A  A  F  S  W  E  E  K  L  Y  L  V  Q
gggggagaggcttgaggtagagactgggacaagcatatccaactctgtatttattaccatcctttgtcctccagGGC
                                                                          G
ACCCAGGTATATGTCTTCCTGACAAAGGGAGGCTATACCCTAGTAAGCGGTTATCCGAAGCGGCTGGAGAAGGAAGT
 T  Q  V  Y  V  F  L  T  K  G  G  Y  T  L  V  S  G  Y  P  K  R  L  E  K  E  V
CGGGACCCCTCATGGGATTATCCTGGACTCTGTGGATGCGGCCTTTATCTGCCCTGGGTCTTCTCGGCTCCATATCA
  G  T  P  H  G  I  I  L  D  S  V  D  A  A  F  I  C  P  G  S  S  R  L  H  I
TGGCAGgtgaggggcttctgggggcttagagggcagcttgttctgctacctgtctgtggcatagatccccaccaggg
M  A
catgagaaggcc---0.13kb---ggatccccatgacatggaagccatgctatgtttggtgccttctccccagGACG
                                                                        G  R
GCGGCTGTGGTGGCTGGACCTGAAGTCAGGAGCCCAAGCCACGTGGACAGAGCTTCCTTGGCCCCATGAGAAGGTAG
  R  L  W  W  L  D  L  K  S  G  A  Q  A  T  W  T  E  L  P  W  P  H  E  K  V
ACGGAGCCTTGTGTATGGAAAAGTCCCTTGGCCCTAACTCATGTTCCGCCAATGGTCCCGGCTTGTACCTCATCCAT
D  G  A  L  C  M  E  K  S  L  G  P  N  S  C  S  A  N  G  P  G  L  Y  L  I  H
GGTCCCAATTTGTACTGCTACAGTGATGTGGAGAAACTGAATGCAGCCAAGGCCCTTCCGCAACCCCAGAATGTGAC
 G  P  N  L  Y  C  Y  S  D  V  E  K  L  N  A  A  K  A  L  P  Q  P  Q  N  V  T
CAGTCTCCTGGGCTGCACTCACTGAGGGGCCTTCTGACATGAGTCTGCCTGGCCCCACCTCCTAGTTCCTCATAATA
  S  L  L  G  C  T  H ***
AAGACAGATTGCTTCTTCGCTTCTCACTGAGGGGCCTTCTGACATGAGTCTGGCCTGGCCCCACCTCCCCAGTTTCT

CATAATAAAGACAGATTGCTTCTTCACTTGaatcaagggaccttggtcgtgaaacaatcttctttctttgagttgaa

aagttagcacttctcctttgagggtgtcgagctcaaacaaggctgtgagaaacaagggaggggagcactaaggggca

aacctatctctgcgcagatgattcttaggtccagatc
```

Fig. 2. Continued

was shown to be composed of 10 repeats of a core unit of 45 amino acids (Fig. 4). All 10 repeats can be aligned according to the amino acid similarity, although their structural relationship is different: units 2, 3, and 4 are more related to units 7, 8, and 9, respectively. Repeats 1 and 6 show weaker similarity to the core unit. Based on these peculiar characteristics, we proposed that the hemopexin gene probably originated from a single sequence followed by amplification of this sequence to a region composed of five repeats. A subsequent duplication of these five repeats would explain the final genomic structure. This model of stepwise evolution would also explain the different degrees of similarity between the repeats, particularly units 1 and 6, which are more divergent and are likely candidates for the ancestral repeat unit. Our determination of the genomic organization of the coding region supports this hypothesis. When we correlated the primary structure of hemopexin with the intron/exon pattern of the gene, a striking pattern emerged: the splice junctions are each located in the same relative positions in the six repeat units that are more conserved, namely units 2, 3, 4, 7, 8, and 9. In these repeats the introns interrupt the reading frame precisely in the codon for the aligned glycine (Fig. 4). In addition, the introns tend to fall between predicted secondary structural elements, namely between two contiguous regions of β-sheet units. Repeats 1 and 6 also contain introns, but these are not in corresponding position with the others: repeat 1 is interrupted by two introns between codons for proline and threonine and threonine and glutamic acid, while the intron in repeat 6 falls in a codon for a nonaligned glycine. Repeats 5 and 10 do not carry any introns. Thus, except for repeats 5 and 10 the intron

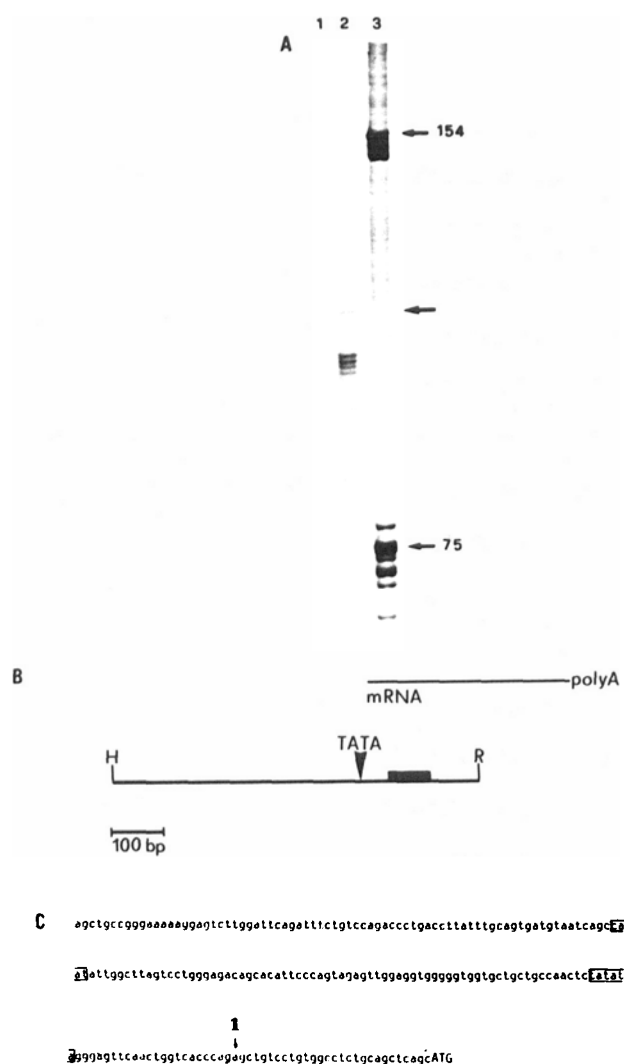

Fig. 3. A S1 mapping of total human liver RNA (8% denaturing polyacrylamide gel). 1: 15 μg tRNA; 2: 15 μg liver RNA; 3: molecular weight marker. **B** The HindIII–EcoRV clone used as probe. The first exon is in the solid block. The "TATA" box is indicated. H: HindIII; R: EcoRV. **C** 5′ flanking sequence. The arrow points to the putative start site of transcription (number 1). The TATA and CAAT sequences are boxed.

location has been preserved during the evolution of this gene; however, neither the size nor the sequences of the introns have been conserved within the gene.

Discussion

The similarity of the intron/exon pattern in structurally similar regions of the protein strongly suggests that the genomic organization of the hemopexin gene was derived from successive duplications of an ancestral genetic domain. Furthermore, the location of the nine introns does not apepar to be distributed randomly, but rather, is strikingly conserved. They fall at sites that correlate with significant features of protein structure and in the case of repeats 2, 3, 4, 7, 8, and 9, precisely in the region of maximum amino acid sequence similarity (Fig. 4). Except for intron I, which separates the signal peptide exon from the rest of the gene—a feature found in many secreted proteins, all the other introns separate two β-strands. Intron VI falls exactly at one end of a β-strand; in no case is a β-strand or an α-helix divided. These facts do not support an insertional model for the origin of these introns but argue instead for the intron-dependent evolution of the human hemopexin gene. As proposed by Gilbert (1978), amplification of genetic domains is facilitated by the presence of preexisting intervening sequences that serve to increase the probability of useful recombination during unequal crossing-over.

The events leading from a single exon to the sequence organization of the gene, composed of 10 exons of quite different sizes, must have included amplification of that sequence (of as yet undetermined length) giving rise to multiple adjacent copies separated by intervening sequences. Individual exons could have originated by unequal crossovers during recombination followed by their divergence. Such an intron-mediated duplication probably occurred in the assembly of the three domains of ovomucoid (Stein et al. 1980), the numerous 18-amino acid domains of collagen (Ohkubo et al. 1980), and the three domains of α-fetoprotein (Tilghman 1982).

As noted previously, pairwise comparisons between the amino acid sequences of the 10 repeats can potentially be used to trace the order of the successive duplications of the primordial structure. The sequence similarities between these regions reflect, in part, the evolutionary time between successive domain duplications, the most similar corresponding to the most recent duplications. However, evolutionary pressures exerted at both the nucleotide and amino acid levels and involving extensive deletions or insertions in the protein-coding region may not have been equally distributed among the 10 repeating units. Thus, unequal rates of divergence of the domains may obscure the details of a model of divergence based on time. However, given these constraints, a possible evolutionary model for the human hemopexin gene is presented (Fig. 5): the earliest ancestor of the hemopexin gene probably corresponds to the exon encoding the most diverged repeat, number 6; a successive duplication occurring by recombination within flanking regions would lead to two repeats of this unit separated by an intron. Exons 8, 9, and 10 may represent the result of unequal crossover between exons 6 and 7. Subsequent duplication of this five-exon cluster produced a gene encoding a protein made up of 10 repeats. It is possible that the position of the introns within codons has shifted (intron I in repeat 1 and

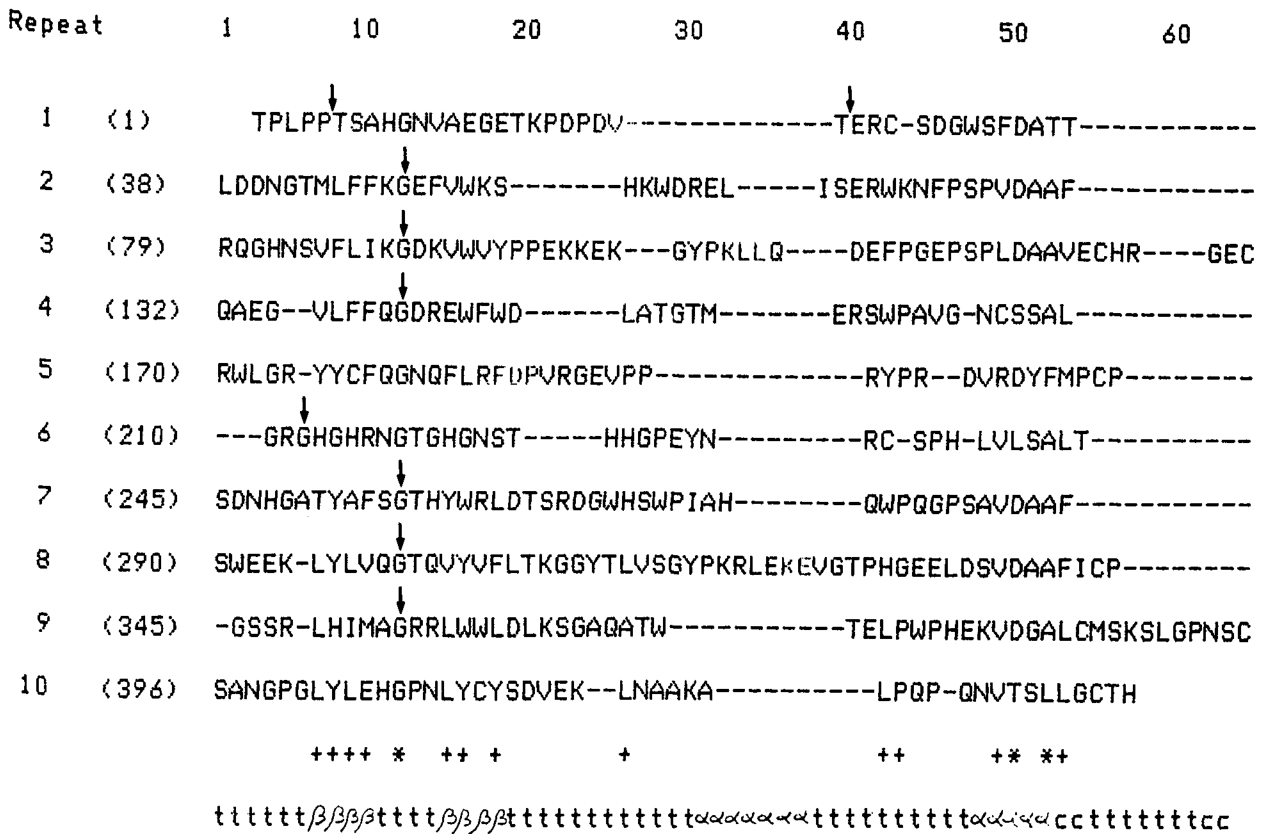

Fig. 4. Alignment of the 10 repeat sequences in hemopexin and the corresponding location of the introns in the gene. The two terminal regions, where few insertions and deletions are indicated, could be easily aligned, though the middle portion represents the best attempt to preserve the hydrophobic character of the amino acids. The secondary structure predictions (α, helix; β, strand; t, turn; c, coil) is shown. The sequence number of the first residue in each repeat (designated 1–10) is given in parentheses. Conservation at a given alignment position is indicated by an * if seven or more of the residues were identical or by a + if seven or more of the residues displayed conservation according to the following scheme: (P, G); (T, S); (K, R); (D, E, Q, N); (A, V, I, L, M, C, H, F, Y, W). The arrows indicate the corresponding locations of the introns in the gene.

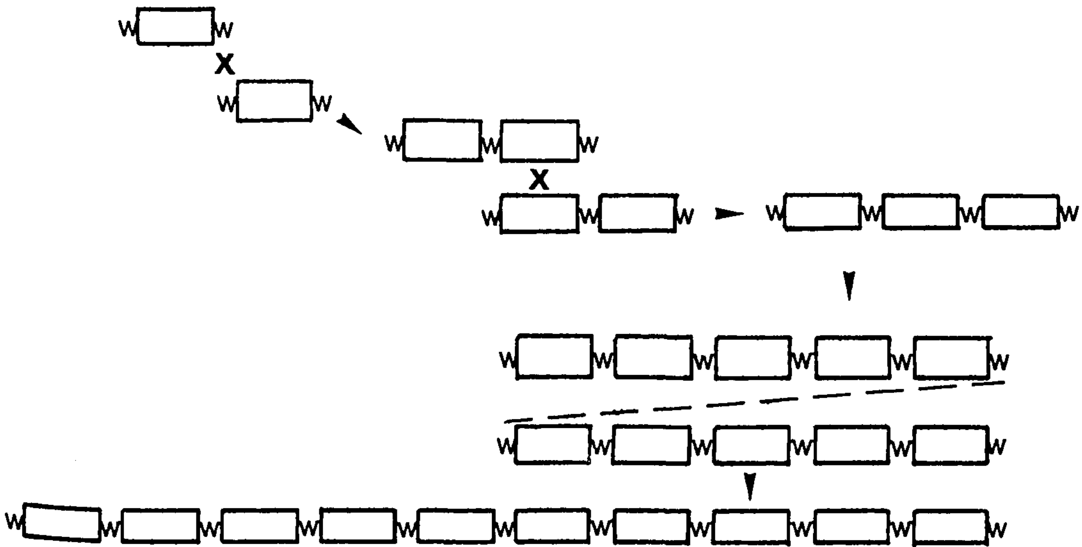

Fig. 5. Schematic model of the evolution of the human hemopexin gene. The boxes represent exons and the connecting wavy lines represent introns (not drawn to scale). See text for details.

intron VI in repeat 6) and that an intron originally present was subsequently lost as in the case of repeats 5 and 10. On the other hand, the location of the introns splitting the glycine codon in repeats 2, 3, 4, 7, 8, and 9 may have been conserved because of some regulatory significance which they impart on the gene.

The active sites in the hemopexin protein (the heme-binding site and the hepatocyte receptor-binding site) are poorly characterized and whether these two functions reside in the same or in separate domains is not known. The demonstration that the gene is composed of genetic elements or exons encoding specific functional or structural domains of the protein would explain the selective advantage of such an evolutionary model. In addition, hemopexin shows strong amino acid sequence similarity with human vitronectin (Stanley 1986), suggesting that the genes coding for these two proteins are descended from a common ancestor, though their functional correlation is not obvious. The position of the introns within the human vitronectin gene may shed light on their common evolutionary origin.

Acknowledgments. We thank Larry Kedes for valuable comments and Mary Gaeta for the preparation of the manuscript. V.P. is a recipient of an Anna Villa Rusconi fellowship. The work has been supported by a grant from P.F. Oncologia, CNR, Rome, Italy.

References

Altruda F, Poli V, Restagno G, Argos P, Cortese R, Silengo L (1985) The primary structure of human hemopexin deduced from cDNA sequence: evidence for internal, repeating homology. Nucleic Acids Res 13:3841–3859

Antonorakis ES, Irkin SH, Cheng T, Scott AF, Sexton JP, Trusko SP, Charache S, Kazazian HH Jr (1984) β-thalassemia in American blacks: novel mutations in the "TATA" box and an acceptor splice site. Proc Natl Acad Sci USA 81:1154–1158

Baumann H, Jahreis GP, Gaines KC (1983) Synthesis and regulation of acute phase plasma proteins in primary cultures of mouse hepatocytes. J Cell Biol 97:866–876

Benoist C, O'Hare K, Breathnach R, Chambon P (1980) The ovalbumin gene-sequence of putative control regions. Nucleic Acids Res 8:127–142

Bensi G, Raugei G, Klefenz H, Cortese R (1985) Structure and expression of the human haptoglobin locus. EMBO J 4:119–126

Berk A, Sharp PA (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 17:721–732

Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon–intron boundaries. Proc Natl Acad Sci USA 75:4853–4857

Falkner FG, Zachau HG (1984) Correct transcription of an immunoglobulin k gene requires an upstream fragment containing conserved sequence elements. Nature 310:71–74

Frischauf AM, Lehrach H, Poustka AN, Murray N (1983) Lambda replacement vectors carrying polylinker sequences. J Mol Biol 170:827–842

Gilbert W (1978) Why genes in pieces? Nature 271:501

Kozak M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eucaryotic ribosomes. Cell 44:283–292

Lerner MR, Boyle JA, Mount SM, Wolen SL, Steitz JA (1980) Are snRNPs involved in splicing? Nature 283:220–224

Messing J, Gronenborn B, Muller-Hill B, Hopshneider PH (1977) Filamentous coliphage M13 as a cloning vehicle: insertion of a Hind III fragment of the lac regulatory region in M13 replicative form in vitro. Proc Natl Acad Sci USA 74:3642–3646

Messing J, Vieira J (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19:269–276

Muller-Eberhard U, Liem HH (1974) In: Allison AC (ed) structure and function of plasma proteins, vol. 1. Plenum, London, p 35

Ohkubo H, Vogeli G, Mudryj M, Avvedimento VE, Sullivan M, Pastan I, de Crombrugge B (1980) Isolation and characterization of overlapping genomic clones covering the chicken 2 (type 1) collagen gene. Proc Natl Acad Sci USA 77:7059–7063

Parslow TG, Blair DL, Murphy WJ, Granner DK (1984) Structure of the 5′ ends of immunoglobulin genes: a novel conserved sequence. Proc Natl Acad Sci USA 81:2650–2654

Sanger F, Nickeln S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Singh H, Sen R, Baltimore D, Sharp P (1986) A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature 319:154–158

Smith A, Morgan WT (1979) Haem transport to the liver by hemopexin. Biochem J 182:47–54

Stanley KK (1986) Homology with hemopexin suggests a possible scavenging function for S-protein/vitronectin. FEBS Lett 199:249–253

Stein JP, Catterall JF, Kristo P, Means AR, O'Malley BW (1980) Ovomucoid intervening sequences specify functional domains and generate protein polymorphism. Cell 21:681–687

Takahashi N, Takahashi Y, Putnam FW (1985) Complete amino acid sequence of human hemopexin, the heme-binding protein of serum. Proc Natl Acad Sci USA 82:73–77

Tilghman SM (1982) In: 27th Scient Rep Inst Cancer Res 125, Fox Chase Cancer Centre

Vieira J, Messing J (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene 19:259–268

Received January 27, 1987/Accepted September 1, 1987