Volume 13 Number 11 1985 Nucleic Acids Research

The primary structure of human hemopexin deduced from cDNA sequence: evidence for internal, repeating homology

Fiorella Altruda, Valeria Poli, Gabriella Restagno, Patrick Argos[1], Riccardo Cortese[1] and Lorenzo Silengo

Cattedra di Biologia Generale,Via Santena 5 bis, 10126 Torino, Italy, and [1]European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, FRG

Received 12 April 1985; Accepted 9 May 1985

ABSTRACT

We have cloned and analyzed a cDNA containing the coding sequence for human hemopexin. We have first identified, by immunological screening of 30.000 colonies of a liver cDNA library in the expression vector pEX1, a clone carrying an insert 1170 base pairs long that shows 100% homology with a known human hemopexin peptide. The complete sequence coding for hemopexin was isolated from a liver cDNA library in the vector pAT218. The DNA insert of 1523 base pairs shows an open reading frame coding for 439 amino acids, a 3' noncoding region of 159 nucleotides long, followed by a poly(A) tail. The insert spans the entire coding region and from which the primary structure of the protein was deduced. By computer assisted analysis of the amino acid sequence, it was possible to recognize a core unit, of about 45 amino acids, which is repeated 8 or possibly even 10 fold along the polypeptide chain. This feature suggests that the gene might have evolved through a series of duplications. This characteristic, together with prediction of secondary structure, suggest a rough model for the tridimensional folding that allows some speculations on the function of hemopexin.

Blot hybridization of total RNA from human liver with nick translated hemopexin cDNA detected a message of about 1600 nucleotides. Southern blot experiments to identify the hemopexin gene (s) suggest that it is not a large multi-gene family, but that there is only one or at most a few genes in the human genome.

INTRODUCTION

Hemopexin is a serum β-glycoprotein consisting of a single polypeptide chain of 439 amino acids with Mr=600,000 which contains 20% carbohydrate (1,2). It binds heme in an equimolar ratio with high affinity, higher than albumin with which it competes for free heme in the plasma (1,3). Its function is the transportation of heme derived from methemoglobin to a specific receptor on the hepatic cell surface (4). The heme is subsequently catabolized to bile pigments and the iron is stored

into hepatic ferritin (5) while the apohemopexin returns to the circulation. Therefore, hemopexin plays an important role in heme disposal and facilitates the conservation of iron by preventing urinary excretion. Moreover this protein interacts with a wide variety of natural and synthetic metalloporphyrins.

Hemopexin is synthesized in the liver and has a normal plasma concentration of 50-100 mg/100 ml (1). It is made by the fetus and newborn, but the concentrations do not approach the adult level until after the first year. The protein levels fall dramatically in some intravascular hemolytic diseases such as sickle cell disease, thalassemia major, autoimmune hemolytic anemia. In contrast hemopexin levels were found significantly higher than normal in individuals shown to be Duchenne muscular distrophy carriers (6). It is also well established that hemopexin plasma levels increase during acute infections, suggesting that its biosynthesis is subjected to control mechanisms responsible for the acute phase reaction (7). It appears therefore that this protein participates in interesting physiological phenomena and is involved directly or indirectly in some important pathological states. For this reason we started a project aimed at cloning the corresponding gene.

Various sensitive and general techniques have been devised to clone genes by using antibodies as probes and to screen DNA libraries in bacterial expression vectors (8-12). We have taken this approach to search for hemopexin cDNA, using a liver cDNA library cloned in the expression vector pEX1 (12) and a rabbit antibody to human hemopexin with anti-rabbit IgG antibodies conjugated to the peroxidase as detection system.

In the present studies, we describe the isolation and characterization of two cDNAs coding for human hemopexin. The analysis of the amino acid sequence deduced from the cDNA clones suggests several repeating units that are structurally homologous.

MATERIALS AND METHODS

Screening of the cDNA Libraries

The immunological screening of the human liver cDNA library in the expression vector pEX1 (12) has been performed (13) with

rabbit anti-human hemopexin antibodies purchased from Nordic Immunology. The human liver cDNA library in the vector pAT218 (14) was screened using as a probe the nick translated DNA insert from clone HpxA (Fig. 1).

DNA Sequence Analysis

The nucleotide sequence was determined by the dideoxy chain-termination method (15) with double-stranded DNA from pUC8/18 or pUC9 recombinant clones and with single-stranded DNA from M13mp8 subclones (16). Sequences were determined on both strands.

RNA Blot Analysis

Ten μg of total RNA isolated from human adult liver with guanidine hydrochloride were denatured in 50% (vol/vol) formamide, 2M formaldehyde at 65°C for 15 min, and then subjected to electrophoresis on a 1.5% agarose gel containing 2.2M formaldehyde in 0.2M morpholinopropanesulfonate, pH 7.0 / 0.05M sodium acetate / 5mM EDTA (17,18) and transferred to nitrocellulose (19). RNA species hybridizing to ^{32}P-labelled hemopexin insert DNA were detected by autoradiography. PolyA$^-$ liver RNA, primarily 18S and 28S rRNA, was treated in the same manner and used as size markers.

DNA Blot Analysis

Human placenta DNA was extracted as previously described (20). DNA was also extracted from the buffy-coat obtained from 30ml of blood (21). Ten μg of DNA from each sample were digested overnight with restriction endonuclease (EcoRI, BamHI, HindIII, PstI; Boehringer, Mannheim), size fractionated on 0.7% agarose gel electrophoresis, and transferred to nitrocellulose filters according to Southern (22). Filters were hybridized with nick translated ^{32}P DNA probes and washed in stringent condition as previously described (23).

Delineation of Repeating Sequences

The hemopexin sequence was compared with itself to check for repeating, structurally homologous segments. Every possible pairwise comparison of sequence spans 15 residues in length was assessed by two scoring procedures: (a) the Dayhoff relatedness odds matrix (24,25,26) whose elements express relative weights with which amino acids substitute for one another in aligned

sequences in 71 protein families and (b) calculation of the mean correlation coefficient over six residue physical characteristics (27,28) important for protein folding (cf. 29). The characteristics (helix, sheet, and turn secondary structural conformational preferences; residue polarity; and two hydrophobicity measures) are listed by Argos et al. (27). The two scores were then scaled and combined. When the hemopexin search matrix was complete over all 15-residue matches, all coefficients were recalculated as a number of standard deviations (σ) above the mean matrix coefficient. These fractional standard deviations were then placed in the matrix corresponding to the beginning residue number of the two oligopeptides compared. The comparison method has been described in detail elsewhere (28).

In the hemopexin search matrix plots, peaks above 3.5 σ were plotted over the entire search window of 15 residues for a more obvious view of any sequence relationships and positioned in the matrix at the corresponding sequence number of the matched amino acids. The theoretical probability of a matrix value in the 3.5 to 5.5 σ range occurring randomly is between 10^{-4} and 10^{-6} (25). Repeats in the hemopexin sequence should show high search values along lines colinear with the main diagonal.

A more sensitive approach to delineation of the hemopexin repeat locations involved a summation of the matrix values along a given line at each hemopexin sequence position, excluding those of the exact self-comparison and including those above a certain threshold value (3.0 σ for the hemopexin case). The sums were then divided by the expected number of repeats minus the self-repeat (e.g. hemopexin, nine) to prevent a large, isolated value along a line from dominating the results. If the number of values above the threshold was greater than the number of expected peaks, the larger number was used as a divisor to prevent overlapping matrix values near repeat termini from influencing the repeat search. Finally a linear plot of the hemopexin sequence position versus the averaged sum of the standard deviation fractions was used to delineate the beginning residue of each repeat. The starting sequence position for the

repeating units would be expected to occur near points where the averaged sum increased dramatically. This latter plot was smoothed by a sliding averaging procedure (30) over three successive points and for 10 complete cycles for easier visual observation of any repeats.

Secondary Structure Prediction

Plots of the sequence number versus the conformational preference parameter (helix, β-strand, turn (31)) for a given amino acid were determined for each protein region using a least squares smoothing procedure. Every successive group of three points (i to i+2) were fit by a least squares line and the value at (i+1) was replaced by the one calculated from the line. The smoothing process was repeated for three cycles over each of the parametric plots. The smoothed curves for each potential were averaged over the aligned hemopexin sequence repeats, a procedure which should yield a better prediction than that from any one sequence (32). The structural type assigned at each aligned residue position corresponded to the largest of the three mean potentials that was greater than 1.0, the neutral preference value (33). Five successive values greater than 1.0 were required for helix initiation and three were used for strands or turns. If a proline or glycine residue occurred at the fifth or smaller position in a region predicted to be helical, the span was assigned to the coil conformation due to the rare appearance of such residues in the central or COOH-terminal parts of helices (34). For all other conditions (e.g. all mean potentials less than 1.0), the coil structure was also predicted.

RESULTS

Immunological Detection of the Hemopexin cDNA

A human liver cDNA library cloned into the PstI site of pEX1 plasmid expression vector (12) was screened for cDNAs coding for human hemopexin. In these studies polyclonal antibodies were used to detect bacterial colonies directing the synthesis of a fusion protein of β-galactosidase and hemopexin. One positive clone was isolated by screening 30.000 colonies and was purified and retested for positive signal. This clone,

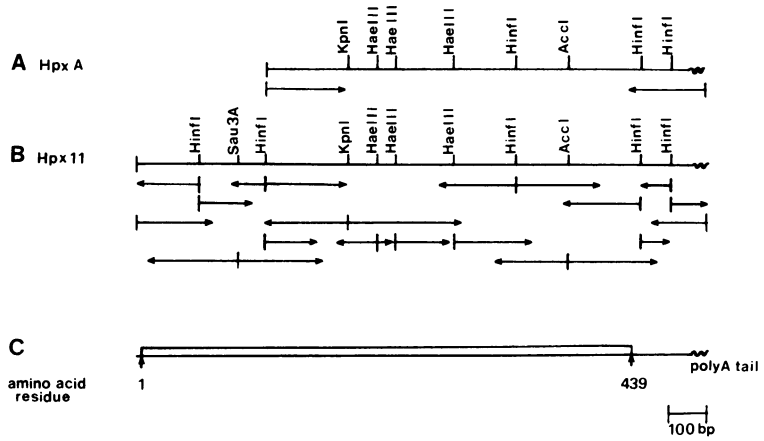

Fig. 1. Partial restriction map and sequencing strategy for the cDNA inserts in clones HpxA (A) and Hpx11 (B). The structure of the coding region and noncoding sequence are shown in (C). Horizontal arrows indicate the direction and length of DNA sequences obtained with the dideoxy chain-termination method.

called HpxA, gave a strong signal with the antibody probe and was found to contain an insert of approximately 1170 base pairs. The structure of the HpxA clone is shown in Figure 1. The DNA fragment was then cloned into the PstI site of pUC8 and the nucleotide sequence of both ends was determined by the dideoxy method.

The 5' end of HpxA contains DNA sequences corresponding to a known peptide of 15 amino acids in positions 154-168 from the NH_2-terminal of hemopexin (35), 115 base pairs downstream from the PstI site of insertion. This shows that HpxA is a cDNA coding for hemopexin, but is incomplete at the 5' end.

Isolation of the Complete Hemopexin cDNA Clone

The partially sequenced cDNA HpxA was used as a probe to screen the human liver cDNA library (14) for a clone containing the complete coding region of the hemopexin. We isolated nine positive clones out of 10^4 recombinant colonies using the Grunstein-Hogness colony screening procedure (36). The DNAs were purified and analyzed. All of them had inserts that strongly hybridized to the nick translated HpxA. The clone containing the largest insert (Hpx11, Fig. 1) was chosen for further

characterization by restriction endonuclease cleavage and DNA sequence analysis.

Sequence Analysis of the Hemopexin Clone

The 1523 base pair insert of Hpx11 contained cleavage sites for the enzymes HinfI, HaeIII, Sau3A, KpnI, AccI, but not for EcoRI, BamHI, HindIII and PstI as shown in the restriction endonuclease map in Figure 1. The complete nucleotide sequence was determined by subcloning in pUC8/9/18 or M13mp8 all the fragments according to the sequencing strategy shown in Figure 1.

The DNA analysis along with the predicted amino acid sequence is shown in Figure 2; it contained six nucleotides probably from the leader sequence (37), the complete hemopexin coding region that generates a polypeptide 439 amino acids long and the entire 3' untranslated region including a polyadenylate tract of 38 nucleotides.

The results are in complete agreement with the oligopeptides determined by protein sequence analysis by Takahashi et al. (35). The correct alignment with the DNA sequence is underlined in Figure 2. The 3' noncoding region is 159 nucleotides long and displays some interesting features. There is a duplication of a 70 base pair segment starting five base pairs upstream from the used stop codon and extending to the poly (A) tail. Each duplicated DNA portion contains a polyadenylation or processing sequence A-A-T-A-A-A but only the second one is used in both the isolated cDNAs. Moreover, the second portion of the sequence is flanked by direct repeats, six base pairs long, indicative of insertional events.

Internal Homology in the Hemopexin Amino Acid Sequence

The hemopexin self-search matrix is shown in Figure 3. Sequence stretches with a combined correlation of 3.5σ or greater are indicated by lines in the matrix. It is clear that an internal structural homology exists between the two halves of hemopexin roughly relating residues 40 to 170 with 250 to 380. However, there are high, seemingly spurious peaks within the matrix not generated by the internal duplication. As a result, another self matrix was calculated using all peaks 3.0σ or greater and a sum performed along each sequence position as

```
      1                                   10
ATTGCCACCCCTCTTCCTCCGACTAGTGCCCATGGGAATGTTGCTGAAGG   50
      T P L P P T S A H G N V A E G
            20                                  30
CGAGACCAAGCCAGACCCAGACGTGACTGAACGCTGCTCAGATGGCTGGA  100
 E T K P D P D V T E R C S D G W S
                        40
GCTTTGATGCTACCACCCTGGATGACAATGGAACCATGCTGTTTTTTAAA  150
  F D A T T L D D N G T M L F F K
  50                                60
GGGGAGTTTGTGTGGAAGAGTCACAAATGGGACCGGGAGTTAATCTCAGA  200
G E F V W K S H K W D R E L I S E
            70                                  80
GAGATGGAAGAATTTCCCCAGCCCTGTGGATGCTGCATTCCGTCAAGGTC  250
 R W K N F P S P V D A A F R Q G H
                        90
ACAACAGTGTCTTTCTGATCAAGGGGGACAAAGTCTGGGTATACCCTCCT  300
  N S V F L I K G D K V W V Y P P
 100                               110
GAAAAGAAGGAGAAAGGATACCCAAAGTTGCTCCAAGATGAATTTCCTGG  350
E K K E K G Y P K L L Q D E F P G
            120                                 130
AATCCCATCCCCACTGGATGCAGCTGTGGAATGTCACCGTGGAGAATGTC  400
 I P S P L D A A V E C H R G E C Q
                        140
AAGCTGAAGGCGTCCTCTTCTTCCAAGGTGACCGCGAGTGGTTCTGGGAC  450
  A E G V L F F Q G D R E W F W D
 150                               160
TTGGCTACGGGAACCATGAAGGAGCGTTCCTGGCCAGCTGTTGGGAACTG  500
L A T G T M K E R S W P A V G N C
            170                                 180
CTCCTCTGCCCTGAGATGGCTGGGCCGCTACTACTGCTTCCAGGGTAACC  550
 S S A L R W L G R Y Y C F Q G N Q
                        190
AATTCCTGCGCTTCGACCCTGTCAGGGGAGAGGTGCCTCCCAGGTACCCG  600
  F L R F D P V R G E V P P R Y P
 200                               210
CGGGATGTCCGAGACTACTTCATGCCCTGCCCTGGCAGAGGCCATGGACA  650
R D V R D Y F M P C P G R G H G H
            220                                 230
CAGGAATGGGACTGGCCATGGGAACAGTACCCACCATGGCCCTGAGTATA  700
 R N G T G H G N S T H H G P E Y M
                        240
TGCGCTGTAGCCCACATCTAGTCTTGTCTGCACTGACGTCTGACAACCAT  750
  R C S P H L V L S A L T S D N H
 250                               260
GGTGCCACCTATGCCTTCAGTGGGACCCACTACTGGCGTCTGGACACCAG  800
G A T Y A F S G T H Y W R L D T S
            270                                 280
CCGGGATGGCTGGCATAGCTGGCCCATTGCTCATCAGTGGCCCCAGGGTC  850
 R D G W H S W P I A H Q W P Q G P
                        290
CTTCAGCAGTGGATGCTGCCTTTTCCTGGGAAGAAAAACTCTATCTGGTC  900
  S A V D A A F S W E E K L Y L V
 300                               310
CAGGGCACCCAGGTATATGTCTTCCTGACAAAGGGAGGCTATACCCTAGT  950
Q G T Q V Y V F L T K G G Y T L V
            320                                 330
AAGCGGTTATCCGAAGCGGCTGGAGAAGGAAGTCGGGACCCCTCATGGGA 1000
 S G Y P K R L E K E V G T P H G I
                        340
TTATCCTGGACTCTGTGGATGCGGCCTTTATCTGCCCTGGGTCTTCTCGG 1050
  I L D S V D A A F I C P G S S R
 350                               360
CTCCATATCATGGCAGGACGGCGGCTGTGGTGGCTGGACCTGAAGTCAGG 1100
L H I M A G R R L W W L D L K S G
            370                                 380
AGCCCAAGCCACGTGGACAGAGCTTCCTTGGCCCCATGAGAAGGTAGACG 1150
 A Q A T W T E L P W P H E K V D G
                        390
GAGCCTTGTGTATGGAAAAGTCCCTTGGCCCTAACTCATGTTCCGCCAAT 1200
  A L C M E K S L G P N S C S A N
```

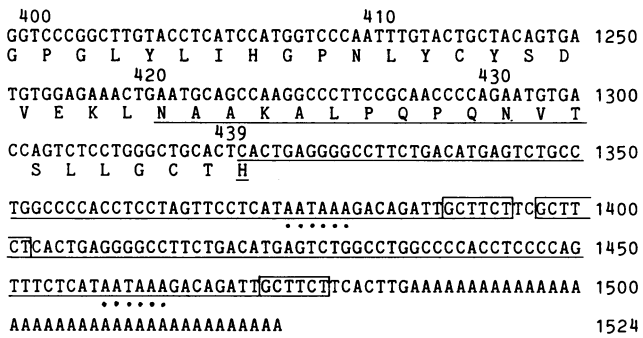

Fig. 2. Complete nucleotide sequence of the cDNA insert in Hpx11 that codes for human hemopexin. The predicted amino acid sequence is shown. The partial amino acid sequence previously published (35) is underlined. The direct repeats are shown in boxes and the duplicated sequence is underlined. Putative polyadenylation sites are dotted.

described under Methods. Since a total of 10 repeats were suspected (vide infra), nine was used as the sum divisor. The linear correlation plot is shown in Figure 4.

It is clear that at least eight repeats of about 40 residues in length are indicated (referred to as repeats 2 to 5 and 7 to 10). They could be aligned visually and are given in Figure 5. Table 1 lists the starting residue determined from the plot of Figure 4 and those used in the alignments; the agreement is excellent. Two residue patterns within each repeat are readily observable: in the N-terminal region, a Gly flanked on either side by three to four hydrophobic residues; and, on the C-terminal side, an Asp followed by three hydrophobic residues of which Ala is at least one and preceded by at least one hydrophobic amino acid which is often Val. The remaining residues not indicated as repeat participants spanned 1 to 37 and 210 to 244, each about 35 residues in length. Each of these displayed in an appropriate position a segment related to at least one of the previous repeats; namely, Phe-Asp-Ala at positions 33 to 35 as well as Val-Leu-Ser-Ala-Leu (related to similar regions in repeats 4 and 10) at 239 to 243. The two segments, referred to as repeats 1 and 6, are also shown in the alignments of Figure 5 and are likely candidates for a similar

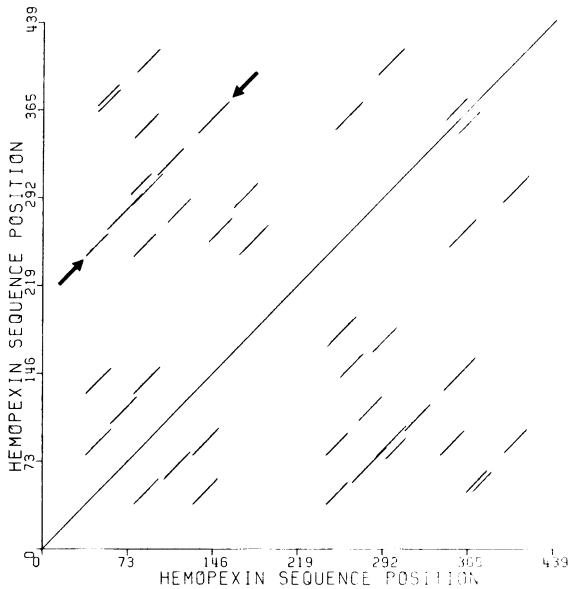

Fig. 3. Self-comparison homology search matrix for hemopexin based on residue structural characteristics. The search window length used was 15 residues. Lines are shown for all search values greater than 3.5σ and are plotted over the entire probe length. The self search is necessarily symmetric about the main diagonal which results from the self comparisons. The series of lines enclosed by arrows and colinear with the main diagonal suggest the structural relatedness between the two hemopexin halves. The other, seemingly spurious, high search values point to the repeating units of hemopexin.

ancestral origin as the more obvious repeats though their structural relationship with the other units is weaker. It has also been suggested that correlations of residue characteristics for aligned sequences are significant when they fall in the neighbourhood of 0.2 (27,38). Table 2 lists the highest mean correlations over the six characteristics discussed here between each repeating unit and one of the remaining repeats. It is clear that the average correlations are all above 0.2. Furthermore, mean characteristic correlations amongst aligned sequences of various dehydrogenase nucleotide binding domains, whose known tertiary folds can be spatially superimposed with resultant aligned amino acids, are in the range of 0.12 to 0.26 (28). The correlations of Table 2 are almost all above this

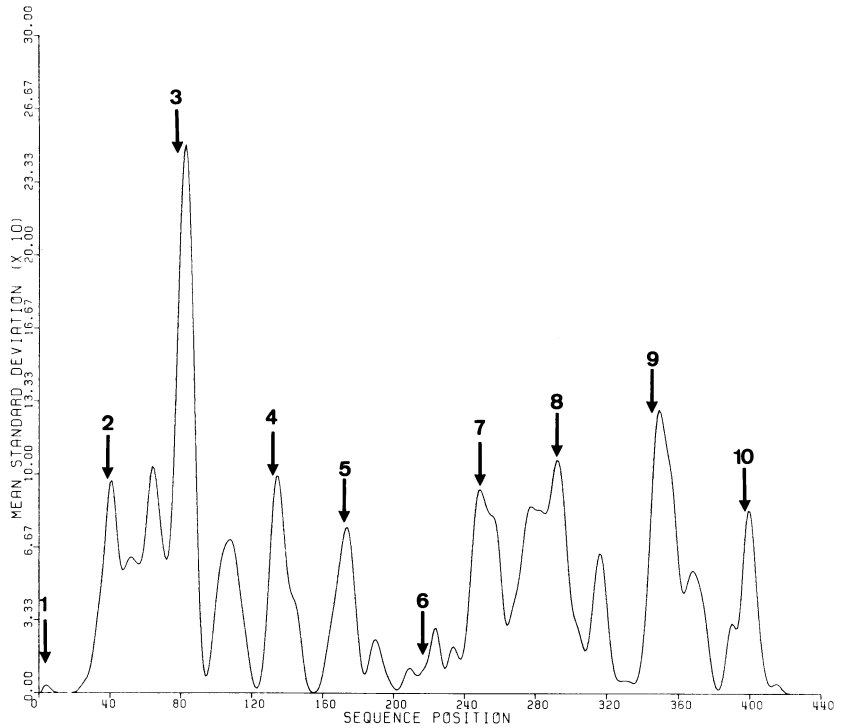

Fig. 4. Plot of the hemopexin sequence position versus the averaged sum of the search matrix elements (with value 3.0σ or greater) along a particular sequence position. Arrows indicate the position of the starting residue of each of the hemopexin repeats as aligned in Figure 5. The value above the arrow refers to the designated repeat number.

latter range. During the preparation of this manuscript we learned that Takahashi et al. (39) have independently determined the primary structure of human hemopexin by amino acid sequence analysis. These authors have also noticed an internal homology, limited however to two repeated regions corresponding to the two halves of hemopexin.

An averaged secondary structural prediction was generated over the 10 repeats (Fig. 6); the predicted structures (helix, β-strand, turn, coil) are shown in Figure 5. All the insertions and deletions occur in predicted turn, coil, or helical segments which have been generally observed in known tertiary structures

```
Repeat      1        10        20        30        40        50        60  64
  1   (1)     TPLPPTSAHGNVAEGETKPDPDV----------TERC-SDGWSFDATT----------
  2   (38)  LDDNGTMLFFKGEFVWKS-------HKWDREL-----ISERWKNFPSPVDAFF----------
  3   (79)  RQGHNSVFLIKGDKVWVYPPEKKEK---GYPKLLQ----DEFPGEPSPLDAAVECHR----GEC
  4   (132) QAEG--VLFFQGDREWFWD------LATGTM-------ERSWPAVG-NCSSAL----------
  5   (170) RWLGR-YYCFQGNQFLRFDPVRGEVPP----------RYPR--DVRDYFMPCP--------
  6   (210) ---GRGHGHRNGTGHGNST-----HHGPEYN---------RC-SPH-LVLSALT----------
  7   (245) SDNHGATYAFSGTHYWRLDTSRDGWHSWPIAH--------QWPQGPSAVDAAF----------
  8   (290) SWEEK-LYLVQGTQVYVFLTKGGYTLVSGYPKRLEKEVGTPHGEELDSVDAAFICP--------
  9   (345) -GSSR-LHIMAGRRLWWLDLKSGAQATW----------TELPWPHEKVDGALCMEKSLGPNSC
 10   (396) SANGPGLYLEHGPNLYCYSDVEK--LNAAKA----------LPQP-QNVTSLLGCTH
                +++++ *  ++ +.      +              ++      +* *+
           ttttttßßßßttttßßßßttttttttttttaaaaaaaattttttttttaaaaaccttttttcc
```

Fig. 5. Alignment of the 10 repeat sequences in hemopexin. The residue matches were determined by visual inspection and with the aid of the linear plot of Figure 4. 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) were taken from the data of Figure 6. Glycosylated asparagines (N) and the observed cleavage point (Lys101) are underlined. The sequence number of the first residue in each repeat (designated 1 to 10) is given in parentheses. Conservation at a given alignment position is indicated by a (*) if 7 or more of the residues were identical or by a (+) if 7 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 match position numbers used in Figure 6 are also given above the aligned sequences.

to be exposed on the protein surface and to be the likely place for easy structural accomodation of insertions or deletions. Each repeat predicts as a pair of anti-parallel β-strands connected by a short four-residue reverse turn, followed by a

Table 1. Starting hemopexin repeat sequence position indicated by the correlation plot of Figure 4 and actually used in the sequence alignments of Figure 5. A (-) refers to the lack of a start indication.

Repeat	Correlation Position	Alignment Position
1	-	1
2	39	38
3	76	79
4	132	132
5	170	170
6	-	210
7	246	245
8	290	290
9	345	345
10	397	396

Table 2. Highest mean physical characteristic correlation between aligned sequences of each repeating unit and one of the remaining repeats.

Repeats compared	Mean Correlation
1-2	0.29
2-3	0.47
3-4	0.53
4-3	0.53
5-4	0.39
6-10	0.26
1-2	0.44
8-3	0.45
9-4	0.45
10-4	0.47

relatively long sequence segment that is largely turn in predicted configuration. The Asn residues known to be glycosylated (35) are underlined in Figure 5; all predicted in exposed turn regions which would easily accomodate the sugars. The known hemopexin cleavage site, (Lys101 (35)), would also be accessible for trypsin according to the predictions.

The strand-turn-strand configuration about the universally conserved glycine was predicted individually in repeats 2-5 and 7-10 while in repeats 1 and 6, the corresponding region was strongly predicted in a turn conformation. Thus, a possible low resolution model for hemopexin might be envisaged as two globular domains connected by a flexible hinge region (N-terminal portion of repeat 6) with eight anti-parallel strands forming a hydrophobic core in each domain (N-terminal portions of repeats 2-5 and repeats 7-10) and with 4 to 5 exposed clusters of residues surrounding the central β-structure in each domain. Repeats 1 and 6, though ancestrally related to the other units, may have diverged to form the hinge function or an N-terminal turn structure.

Table 3 shows the number of amino acid identities for all the pairwise repeat comparisons. It is clear that units 2 and 7 and units 3 and 8 are most closely related; they also show strong conservation in the Val-Asp-Ala-Ala-Phe region. It is possible that these segments form cellular receptor sites with the Asp residue being central for recognition. There is no known

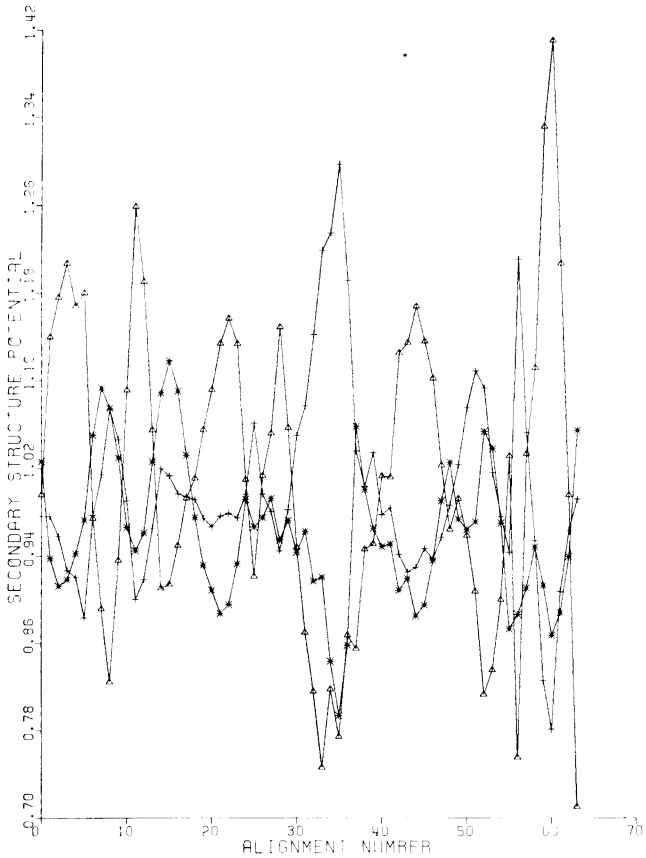

Fig. 6. The mean predicted secondary structure plots for the aligned sequences of Figure 5. The alignment number refers to the match position annotated in Figure 5. The plot symbols indicate the α-helix (+), β-strand (*), and turn (△) potential curves. The procedures used for prediction are given by Zalkin et al. (28).

case of an Asp being involved in protoporphyrin binding, the other important function of hemopexin. There are also other conserved aspartates in some of the other repeats; they too could be important for cell surface recognition though they lack the strong conservation of flanking residues displayed by units 2, 3, 7, and 8. The number of amino acid identities amongst the repeats do not appear to suggest any consistent evolutionary scheme of gene duplication, allowing a model for construction of

Table 3. Number of identical amino acids (upper right portion) in all pairwise comparisons of aligned sequences for hemopexin repeating (R) units 1 to 10 as identified in Figure 5. The lower left matrix portion shows the number of aligned residues for each comparison.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
R1	X	5	3	2	5	5	3	4	2	2
R2	30	X	10	7	4	6	15	7	6	2
R3	36	36	X	7	7	3	9	14	7	6
R4	28	36	33	X	7	6	7	6	7	8
R5	32	30	38	29	X	4	8	11	6	6
R6	28	32	32	31	28	X	5	5	7	7
R7	34	37	41	34	36	33	X	10	9	8
R8	36	40	48	37	40	34	43	X	8	9
R9	35	33	43	32	39	31	38	43	X	8
R10	31	35	40	33	37	32	40	42	39	X

the present-day hemopexin except for the apparently most recent duplication responsible for the two molecular halves.

It is appealing to suggest that the hinge region of unit 6 provides residues for interaction with the porphyrin. Five histidines, which are frequently observed ligands for heme iron, are contained within this span. An alternative hypothesis is suggested by the results of homology searches made between hemopexin and other sequences of proteins known to bind heme as cytochrome $\underline{c}$, $\underline{b}$, P450, $\underline{b}_{562}$, $\underline{c}_2$, $\underline{c}'$, $\underline{c}_3$ as well as globin, catalase, cytochrome peroxidase, and the like. No overall homologies could be found with hemopexin though several local correlations of about 20 to 30 residues in length were displayed within the 4.0 to 5.5σ range. Only one of these showed conservation between residues known to be essential for heme binding; namely, those of a cytochrome $\underline{c}_2$ segment from

```
Cytochrome c2  (1)   ADAPPPAFNQCKACHSIDAGKNGV
Hemopexin      (113) FPGEPSPLDAAVECHRGECQAEGV
                       +   *  ++ +  **  ++  +**
```

Fig. 7. A speculative homology between a hemopexin sequence segment and one from Rhodospirillum molischianium cytochrome $\underline{c}_2$ (40) which is composed of 100 amino acids. The beginning sequence position of each segment is shown in parentheses while residue conservation is indicated as explained in Figure 5. The identically conserved Cys-His are respectively used in cytochrome $\underline{c}_2$ for covalent linkage of heme to the protein and liganding to the porphyrin iron.

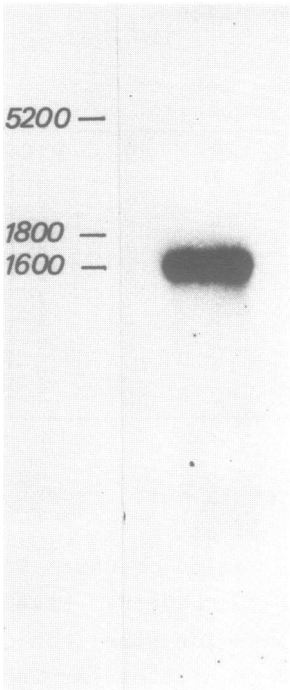

Fig. 8. Blot hybridization of RNA. Ten µg of total human liver RNA were analyzed for hemopexin mRNA by electrophoretic separation on 1.5% agarose gel, transfer to a nitrocellulose filter and hybridization of nick translated Hpx11 (^{32}P) DNA. The hybridizing mRNA molecules were detected by autoradiography. The size of the mRNA, in bases, is indicated; 18S and 28S rRNA from human liver RNA were used as size markers.

Rhodospirillum molischianium (Fig. 7). The conserved Cys-His provide respectively covalent binding of the heme to the protein and iron liganding in the c_2 structure. The supposedly homologous region in hemopexin is found in an added external loop on repeat 3 close to the Val-Asp-Ala-Ala-Phe sequence. This speculative homolgy would imply such functions for the corresponding Cys-His residues in hemopexin though it must be emphasized that a second cytochrome c_2 cysteine (three residues N-terminal to the histidine) also involved in linking the heme is not conserved in hemopexin.

Analysis of the Transcriptional Product

The sequenced cDNA was used as a probe to explore the number and size of mRNAs coding for hemopexin-related protein; the result of a RNA gel-blot analysis is shown in Figure 8. Hybridization of a radiolabelled probe for human hemopexin to total cytoplasmic RNA from adult liver detected one specific transcript of about 1600 nucleotides. With this type of

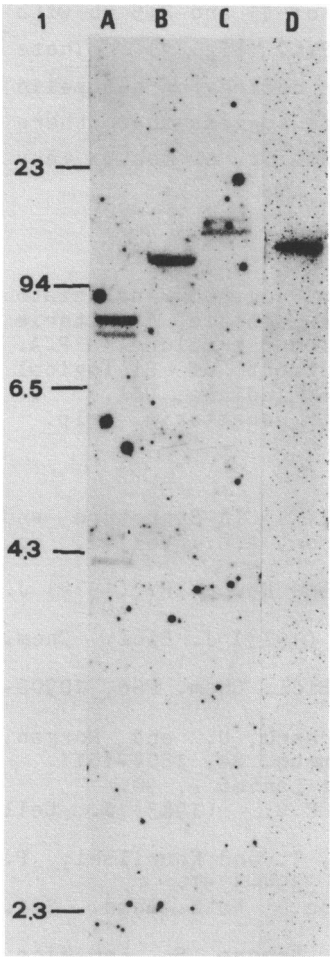

Fig. 9. Southern hybridization of the Hpx11 probe to human placenta DNA. Ten µg of DNA, digested with the restriction enzymes BamHI (lane A), HindIII (lane B), EcoRI (lane C) and PstI (lane D), were electrophoresed on a 0.7% agarose gel and transferred to a nitrocellulose filter. Fragments derived from HindIII digestion of λ DNA were used as size markers. Numbers are represented in Kbp (lane 1).

analysis, however, the resolution would not be sufficient to distinguish between two RNA species differing by 80 bases and generated by the use of both polyadenylation sites indicated in Figure 2.

Hemopexin Genomic Organization

To analyze the genomic pattern of human hemopexin, we carried out a Southern blot experiment of human genomic DNA probed with a nick translated insert of Hpx11 (Fig. 9). Placental or pheripheral blood cellular DNA revealed three bands of 8.4, 7.6, and 3.8 Kb after digestion with BamHI; two bands of

about 14 and 12.5 Kb with EcoRI; two bands of 11 and 2.5 Kb with HindIII; and one band of about 12 Kb with PstI (Fig. 4). These results show that the organization of genes coding for hemopexin is not complex: The intensity of the bands suggest that there is probably only one gene per haploid genome or, at most, very few.

ACKNOWLEDGEMENTS

This work was supported by grants from Progetto Finalizzato Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie of the CNR, and from Ministero della Pubblica Istruzione. P.A. is on sabbatical leave from the Department of Biological Sciences, Purdue University, West Lafayette, Indiana, USA. We are grateful to H. Seifert and I. Benner for secretarial help.

REFERENCES

1. Muller-Eberhard, U. and Liem, H.H. (1974) In Structure and Function of Plasma Proteins, Ed. Allison, A.C., Vol. 1, pp. 35-53, Plenum, London.
2. Uhlenbruck, G., Reese, I., Vaith, P. and Haupt, H. (1979) J. Clin. Chem. Clin. Biochem. 17, 29-34.
3. Seery, V.L. and Muller-Eberhard, U. (1973) J. Biol. Chem. 248, 3796-3800.
4. Smith, A. and Morgan, W.T. (1981) J. Biol. Chem. 256, 10902-10909.
5. Davies, D.M., Smith, A., Muller-Eberhard, U. and Morgan, W.T. (1979) Biochem. Biophys. Res. Commun. 84, 1504-1511.
6. Danieli, G.A., Angelini, C. (1976) The Lancet 2, 90.
7. Baumann, H., Jahreis, G.P., Gaines, K.C. (1983) J. Cell Biol. 97, 866-876.
8. Sanzey, B., Mercereau, O., Ternynck, T. and Kourilski, P. (1976) Proc. Natl. Acad. Sci. USA 73, 3394-3397.
9. Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA 75, 2746-2749.
10. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S.P., Chick, W.L. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA 75, 3727-3731.
11. Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80, 1194-1198.
12. Stanley, K.K. and Luzio, J.P. (1984) EMBO J. 6, 1429-1434.
13. Stanley, K.K. (1983) Nucl. Acids Res. 3, 4077-4092.
14. Woods, D.E., Markam, A.F., Ricker, A.T., Goldberg, G. and Colten, H.R. (1982) Proc. Natl. Acad. Sci. USA 79, 5661-5665.
15. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
16. Messing J. (1983) Methods in Enzymol. 101, 28-78.
17. Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) Biochemistry 16, 4743-4751.
18. Golberg, D.A. (1980) Proc. Natl. Acad. Sci. USA 77, 5794-5798.

19. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.
20. Gross-Bellard, M., Oudet, P. and Chambon, P. (1973) Eur. J. Biochem. 36, 32-38.
21. Steffen, D. and Weinberg, R.A. (1978) Cell 15, 1003-1010.
22. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
23. Migone, N., Feder, J., Cann, H., Van West, B., Hwang, J., Takahashi, N., Nakai, S. and Honjo, T. (1980) Nucl. Acids Res. 8, 5983-5991.
24. Barker, W.C., Ketcham, L.K., Dayhoff, M.O. (1978) in "Atlas of Protein Sequence and Structure" (Dayhoff, M.O., ed.) Vol. 5, Suppl. 3, pp. 359-362, National Biomedical Research Foundation, Washington, D.C., USA.
25. McLachlan, A.D. (1971) J. Mol. Biol. 61, 409-425.
26. Staden, R. (1982) Nucl. Acids Res. 10, 2951-2961.
27. Argos, P., Hanei, M., Wilson, J.M., Kelley, W.N. (1983) J. Biol. Chem. 258, 6450-6457.
28. Zalkin, H., Argos, P., Narayana, S.V.L., Tiedeman, A.A., Smith, J.M. (1985) J. Biol. Chem., in press.
29. Creighton, T.E. (1978) Biophys. Mol. Biol. 33, 231-297.
30. Kyte, J., Doolittle, R.F. (1982) J. Mol. Biol. 157, 105-132.
31. Palau, J., Argos, P., Puigdomenech, P. (1982) Intl. J. Protein Peptide Res. 19, 394-401.
32. Argos, P., Schwarz, J., Schwarz, J. (1976) Biochim. Biophys. Acta 439, 261-273.
33. Chow, P.Y., Fasman, G.D. (1974) Biochemistry 13, 211-245.
34. Argos, P., Palau, J. (1982) Intl. J. Protein Peptide Res. 19, 380-393.
35. Takahashi, N., Takahashi, Y., Putnam, F.W. (1984) Proc. Natl. Acad. Sci. USA 81, 2021-2025.
36. Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.
37. Blobel, G., Walter, P., Chang, C.N., Goldman, B.M., Erikson, A.H. and Lingappa, R. (1979) In Secretory Mechanisms: Society for Experimental Biology Symposium, Eds, Hopkins, C.R. and Duncan, C.J. Vol. 33, pp. 9-36, (Cambridge, Univ. Press, London).
38. Pongor, S. Szalay, A.A. (1985) Proc. Natl. Acad. Sci. USA 82, 366-370.
39. Takahashi, N., Takahashi, Y., Putnam, F.W. (1985) Proc. Natl. Acad. Sci. USA 82, 73-77.
40. Ambler, R.P. (1978) in "Evolution of Protein Molecules" (Matsubara, H. and Yamanaka, T., eds.), pp. 311-322, Japan Scientific Societies Press, Tokyo, Japan.