Haptoglobin, a hemoglobin-binding plasma protein, is present in bony fish and mammals but not in frog and chicken

Krzysztof B. Wicher* and Erik Fries

Department of Medical Biochemistry and Microbiology, Uppsala University, 75123 Uppsala, Sweden

Edited by Morris Goodman, Wayne State University School of Medicine, Detroit, MI, and approved January 18, 2006 (received for review October 5, 2005)

Hemoglobin (Hb) released from erythrocytes may cause oxidation of lipids and proteins. Haptoglobin (Hp), which occurs in the plasma of all mammals, binds free Hb and inhibits its oxidative activity. It is not known whether this protective protein also exists in lower vertebrates. By analyzing available genomic sequences, we have found that bony fish, but not more primitive animals, have a gene coding for a protein homologous to mammalian Hp. Furthermore, we show that this protein is present in the plasma of Japanese puffin (Takifugu rubripes) and that it binds Hb. These results, together with a phylogenetic analysis, suggest that Hp evolved from a complement-associated protein (mannose-binding lectin-associated serine proteinase, MASp), with the emergence of fish. Surprisingly, we found that both chicken (Gallus gallus) and the Western clawed frog (Xenopus tropicalis) lack the Hp gene. In chicken plasma, however, we identified a different type of Hb-binding protein, PIT54, which has been reported to be a potent antioxidant. PIT54 is a soluble member of the family of scavenger receptor cysteine-rich proteins, and we found that its gene exists only in birds. We also show that the plasma of ostrich (Strutio camelus), a primitive bird, contains both PIT54 and Hp. Collectively, our data suggest that PIT54 has successively taken over the function of Hp during the evolution of the avian lineage and has completely replaced the latter protein in chicken.

Results

Hp in Invertebrates. To identify proteins similar to mammalian (human) Hp in different species, we analyzed genomic and EST sequences with the BLAST program. The predicted protein sequences with the highest scores were examined for the presence of the Hp hallmark: a serine proteinase (SP) domain lacking the essential serine and histidine residues of the active site and a “histidine-loop” cysteine pair (15). We investigated the genomes of the five invertebrates: Saccharomyces cerevisiae, Caenorhabditis elegans, Apis mellifera, Drosophila melanogaster, and Ciona intestinalis and found that none contained a gene coding for a protein similar to mammalian Hp. The closest relative found was a SP, MASp, in C. intestinalis (data not shown).

Hp Proteins in Bony Fish. The genomes of three bony (teleost) fishes have been fully sequenced to date: zebrafish (Danio rerio), spotted green pufferfish (Tetraodon nigroviridis), and Japanese pufferfish (Takifugu rubripes). In all three genomes, we found a gene encoding a protein with an amino acid sequence 32–34% identical and 49–52% similar to that of the β-chain of human Hp (Fig. 1). The Hp-like protein (HpL) encoded by this gene seems to comprise a single domain with similar structural characteristics to Hp described above. In addition, the cysteine residues of this domain are present at the positions corresponding to those of mammalian Hp.

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: Hp, haptoglobin; HpL, Hp-like protein; SP, serine proteinase; MASp, mannose-binding lectin-associated SP; CCP, complement control protein; DHODH, dihydroorotate dehydrogenase.

*To whom correspondence should be addressed. E-mail: krzysztof.wicher@imbim.uu.se.

© 2006 by The National Academy of Sciences of the USA
in the SP domain of mammalian Hps (the β-chain), with the exception of the residue that in mammalian Hps joins the β- and the α-chains. The SP domain of the fish HpLs differs from that of mammalian Hps in that it is encoded by two, rather than one, exon (data not shown). Each of these exons codes for the same region in all three proteins, while the length of the separating intron, as inferred from the ENSEMBL annotation, varies (97, 122, and 4,383 nt long for *T. nigroviridis*, *T. rubripes*, and *D. rerio*, respectively). The SP domain of these HpLs is preceded by a 16- to 22-residue extension and a putative signal peptide. The ORFs for the HpL genes of *T. nigroviridis* and *T. rubripes* code for an additional 90–100 aa upstream from the predicted signal peptides. However, analysis of the Kozak sequences around the possible translation initiation sites suggested that it is more likely that the second methionine codon is used for initiating (data not shown). The amino acid sequences of the extension differ considerably between the different species. However, the C termini of all of these extensions contain the consensus cleavage sequence for subtilisin-like proprotein convertases: R-X-(R/K)-R (16) (Fig. 1). Using the sequences of these three fish proteins, we identified ESTs coding for HpL from other teleost species. From their sequences, we were able to deduce the amino acid sequence of HpL from seven more bony fishes (Data Set 1, which is published as supporting information on the PNAS web site).

To investigate whether fish HpL binds to Hb, we incubated serum from *T. rubripes* with gel beads to which Hb from koi carp (*Ciprinus carpio*) was coupled, eluted bound proteins with urea, and analyzed them by SDS-PAGE (Fig. 2). Two of the major bands, with apparent molecular masses of ~75 and 45 kDa (bands 1 and 2), were identified by MS as IgM heavy chain (μ-chain) and HpL of *T. rubripes*, respectively. The other two major bands of ~30 and 15 kDa, bands 3 and 4, respectively, were found to be globin, the subunit of Hb.

**Amphibian Hb-Binding Proteins.** We searched the *Xenopus tropicalis* genomic database for a gene coding for a protein homologous to mammalian Hb. The closest related proteins we could find were orthologs of C1r, C1s, and MASPs. However, the active sites of these fish proteins, we identified ESTs coding for HpL from other teleost species. From their sequences, we were able to deduce the amino acid sequence of HpL from seven more bony fishes (Data Set 1, which is published as supporting information on the PNAS web site).

To investigate whether fish HpL binds to Hb, we incubated serum from *T. rubripes* with gel beads to which Hb from koi carp (*Ciprinus carpio*) was coupled, eluted bound proteins with urea, and analyzed them by SDS-PAGE (Fig. 2). Two of the major bands, with apparent molecular masses of ~75 and 45 kDa (bands 1 and 2), were identified by MS as IgM heavy chain (μ-chain) and HpL of *T. rubripes*, respectively. The other two major bands of ~30 and 15 kDa, bands 3 and 4, respectively, were found to be globin, the subunit of Hb.

**Amphibian Hb-Binding Proteins.** We searched the *Xenopus tropicalis* genomic database for a gene coding for a protein homologous to mammalian Hb. The closest related proteins we could find were orthologs of C1r, C1s, and MASPs. However, the active sites of these fish proteins, we identified ESTs coding for HpL from other teleost species. From their sequences, we were able to deduce the amino acid sequence of HpL from seven more bony fishes (Data Set 1, which is published as supporting information on the PNAS web site).
sites of all of these proteins contain the amino acids residues essential for proteolytic activity. We therefore conclude that there is no Hp in X. tropicalis. To determine whether the serum of this type of amphibian contains any Hb-binding protein, we used affinity chromatography and SDS/PAGE, as described above; for practical reasons, Xenopus laevis was used for these experiments. The final MS analysis of the major electrophoretic band yielded MGGCG9066, a protein similar to the α-chain of IgM. This protein has not previously been assigned any function and was not studied further.

**Avian Hb-Binding Proteins.** We did not find any gene coding for a protein similar to mammalian Hp in the genome of chicken. The most similar proteins, as was the case for *X. tropicalis*, were orthologs of mammalian C1r, C1s, MASPs, and other related SPs. We conclude, therefore, that chicken also does not have Hp. However, the fact that a Hb-binding protein was previously isolated from chicken plasma prompted us to identify this protein. In agreement with a previous study (6), we found that affinity chromatography with immobilized chicken Hb followed by SDS/PAGE yielded a 69-kDa band. MS analysis showed that this band contained PIT54 (or 18-B), a soluble member of the family of scavenger receptor cysteine-rich proteins. Similarly, we isolated a Hb-binding protein of 75 kDa from plasma of domestic goose (*Anser anser*). Analysis by liquid chromatography tandem MS showed that is a was PIT54 homolog (data not shown).

We also examined the presence of Hb-binding proteins in ostrich, a representative of paleognaths, a separate phylogenetic line of modern birds (reviewed in ref. 17). To this end, plasma from common ostrich (*Struthio camelus*) was incubated with gel beds coupled with ostrich Hb. After washing, bound proteins were eluted with solutions of increasing urea concentration (Fig. 3A). Analysis by SDS/PAGE under reducing conditions showed that a 75-kDa polypeptide (band 1) was eluted with 1.5–3.0 M urea, and three polypeptides of 35, 15, and 12 kDa (bands 2, 3, and 4, respectively) were eluted with 3.0–8.0 M urea. Band 4 was apparently globin because it was also eluted without prior incubation with plasma (data not shown). Under nonreducing conditions, the polypeptide eluted with 1.5–3.0 M urea had an apparent molecular mass of 60 kDa (band 5). Under the same conditions, bands 2 and 3 were absent, but a single band of ~100 kDa appeared (band 6), indicating that the polypeptides of the former bands were linked by disulfide bridges.

N-terminal sequencing of the 35-kDa polypeptide (band 2) yielded the sequence IGGGLLACGKGSFPWQ, which is 66–80% identical with that of the N terminus of the β-chain of mammalian Hps. No sequence was obtained when the same technique was applied to the 75-kDa polypeptide, probably because the α-amino group of the N-terminal residue was blocked. This polypeptide, therefore, was analyzed with liquid chromatography tandem MS. Five of the seven peptide sequences obtained with this technique could be aligned with a high degree of identity with the protein sequence of chicken PIT54 (Fig. 3B). Three of these peptides aligned at two to three places, presumably reflecting the fact that PIT54 consists of four repeated domains (18).

**Phylogenetic Analysis.** Given the fact that fish HpLs differ from mammalian Hps both with respect to domain and exon organization, we wanted to determine whether phylogenetic analysis indicates a common ancestor. To this end, we constructed a phylogenetic tree of the SP domains of fish HpLs, human Hp, and fish C1r/s, and MASp proteins. High bootstrap values were obtained with the maximum-parsimony method (Fig. 4) and the neighbor-joining method (Fig. 7, which is published as supporting information on the PNAS web site), 964 of 1,000 and 1,000 of 1,000 respectively, for the branch containing human Hp and fish HpLs, indicating that these proteins have a common ancestor. The two methods did not give a consistent bootstrap support for a hierarchy between the branch containing Hp and HpL and the other branches.

**Chromosomal Localization of Hp and HpL.** To further test the hypothesis that Hp and HpL have a common origin, we analyzed the chromosomal regions close to the respective genes. We found that both Hp in mammals and HpL in fish are located next to the gene for dihydroorotate dehydrogenase (*DHODH*) (Fig. 5). The fact that *DHODH* is a single-copy gene in both groups of animals suggests that Hp and HpL are orthologs.

The finding that the Hp gene is not present in frog and chicken prompted us to study the corresponding part of the genome of these animals. We found that, in both *Xenopus* and chicken, there is a gene coding for ATP-dependent RNA helicase DHX38 (*DHX38*) located next to *DHODH* (Fig. 5). In mammals, *DHX38*
is also located close to DHODH, separated only by Hp and TXNL4B. Taken together, these observations indicate that Hp has been deleted from the region between DHODH and DHX38 in both X. tropicalis and chicken.

Discussion

In this study, we have sought to ascertain whether animals lower than mammals express the plasma protein Hp. We show that, in bony fish, there is a HpL consisting of only one of the two domains of mammalian Hp, whereas, in both chicken and the African clawed frog, the protein is absent. On the other hand, in ostrich, which is a primitive type of bird, there is Hp with apparently the same domain structure as that of mammalian Hp.

Hp in Bony Fish. In the three bony fishes whose genome has been fully sequenced, we found a gene coding for a protein with an amino acid sequence similar to that of the β-chain of mammalian Hp (Fig. 1). Using the sequences of these three HpLs as well as ESTs, we obtained sequences of HpLs from seven more bony fish (Data Set 1), indicating that this protein is widely expressed in this class of vertebrates. The predicted translational products of the genes coding for HpL consist of a putative signal peptide, a 16- to 22-residue extension with the consensus cleavage sequence for subtilisin-like proprotein convertases at its C-terminal end, and a SP domain (Fig. 1). These characteristics suggest that HpL is a secretory protein and that it is proteolytically cleaved near the SP domain while passing through the Golgi apparatus. Indeed, our preliminary experiments with HpL from T. rubripes expressed in COS-7 cells support this notion (our unpublished observation). In contrast, the primary translation product of mammalian Hp mRNA consists of a signal sequence, a CCP, and a SP domain. After the release of the signal sequence and while still in the endoplasmic reticulum, this polypeptide is apparently fully cleaved into the two domains by C1r-like protein (19). Whether the difference in cleavage mechanisms between HpL and Hp has any physiologic significance remains to be studied. We found that HpL of at least one bony fish occurs in plasma and binds Hb (Fig. 2), indicating that this protein has the same function as that of mammalian Hp: to inhibit the oxidative reactions of free Hb.

Our finding that fish HpL binds Hb is in agreement with earlier reports, showing that the binding of mammalian Hp to Hb is mediated by the β-chain (the SP domain) (20). The function of the α-chain of mammalian Hp is unknown. This part of the protein possibly mediates the binding of Hp to CD11b/CD18 (21), an αβ2 integrin present on monocytes, granulocytes, and natural killer cells; this type of integrin is not present in T. rubripes (22).

Hp in Birds. Among birds, only the genome of chicken (Gallus gallus) has been fully sequenced. In this genome, we could not find any gene coding for a protein similar to mammalian Hp. By biochemical means, however, we identified PIT54 (or 18-B), a protein homologous to the scavenger receptor cysteine-rich family of proteins, as the major Hb-binding protein in both chicken and goose plasma. Whether also the goose genome lacks protein homologous to the scavenger receptor cysteine-rich family of proteins is unknown. Interestingly, CD163, which has been shown to be the receptor for Hp–Hb complexes on human monocytes, is a member of the same protein family as PIT54 (23). Consistent with a role as a protection against free Hb, PIT54 has been found to be a potent antioxidant (18). Chicken and goose are representatives of two different neognath
orders. The absence of Hp in their plasma suggests that the Hp gene was lost at least by the common ancestor of these orders.

Ostriches represent a line of modern birds (the paleognaths) that seems to have separated 80 million to 90 million years ago from the neognathous birds (24). Using immobilized Hb, we isolated three polypeptides from ostrich plasma with the apparent molecular masses of 15, 36, and 75 kDa (Fig. 3). The 36-kDa polypeptide, which was found to be homologous to the β-chain of Hp, formed a disulfide-linked complex of 100 kDa with the 15-kDa polypeptide. This finding suggests that the complex is a heterotetramer as is the case for mammalian Hp (type 1). From the 75-kDa polypeptide, we obtained a partial amino acid sequence similar to that of chicken PIT54. A PIT54-like protein has also been found in emu, another paleognathous bird (25).

Based on the currently available genomic data, PIT54 seems to exist only in birds. It is possible that this protein first appeared in an ancestor of paleognathous and neognathous birds, which also possessed Hp. The paleognaths then retained both proteins whereas at least one of the neognaths lost Hp, making PIT54 the only Hb-binding protein in these animals. We are not aware of any other case where the function of one gene has been taken over by another, completely unrelated gene.

Evolution of Hp. On the basis of sequence similarity, mammalian Hps and fish HpLs seem to be related to the SPs that activate the complement system: C1r, C1s, MASP-1, MASP-2, and MASP-3. All of these proteinases consist of six domains: two C1r/C1s/Uegf/bone morphogenic protein 1 domains, one epidermal growth factor-like protein domain, two CCP domains, and one SP domain. C1r, C1s, and MASP-2 are encoded by separate genes, whereas MASP-1 and MASP-3 result from alternative splicing of the MASP-1/3 gene, which has two different regions, each coding for a SP domain. The SP domain of MASP-3 appears to have arisen by retrotransposition of a fragment of MASP1 mRNA. MASP-1 can be traced back to the urochordates and has been suggested to be the prototype of the family. This protein differs from Hp, C1r, C1s, MASP-2, and MASP-3 in that its SP domain is encoded by multiple exons and has a histidine-loop cysteine bridge. Based on these differences, it is likely that Hp, C1r, C1s, and MASP-2 have a common ancestor, whose gene appeared through a partial duplication of an ancestral MASP-1/3 (26). Indeed, our phylogenetic analysis of the human proteins placed Hp, MASP-2, MASP-3, C1r, and C1s on the same branch (Fig. 8, which is published as supporting information on the PNAS web site).

Despite the differences at both the DNA and protein level, fish HpLs and mammalian Hps seem to be orthologs. This notion is supported by two observations. First, our phylogenetic analysis indicated that their SP domain originates from a common ancestor (Figs. 4 and 7). Second, we found that, in all analyzed fish and mammals, the gene coding for the respective protein is located close to the DHODH gene (Fig. 5), indicating conserved synteny.

The evolution of fish HpL and mammalian Hp could be explained by the following sequence of events (Fig. 6). Through partial duplication (step I) of the gene coding for the common ancestor of Hp, C1r, C1s, and MASP-2 [anc(MASP-2/C1r/C1s/Hp)], the Hp prototype appeared. It had a domain organization similar to that of mammalian Hp (consisting of both a CPP and a SP domain) but was still an inactive protease. Subsequently, the essential serine (S) and histidine (H) residues in the active site of its SP domain were substituted with alanine (A) and either lysine (K) or arginine (R) residues (step II), resulting in a proteolytically inactive protein. All of these events occurred in a common ancestor of bony fish and mammals. In the fish lineage, the CCP domain of Hp was then deleted (step III), and an intron was inserted in the exon coding for the SP domain (step IV), giving rise to the present HpL gene.

In mammalian genomes, the Hp gene is flanked by DHODH and DHX38 (Fig. 5). In chicken and X. tropicalis, which lack Hp, the latter two genes are located closely together, indicating that the Hp gene has been deleted. Given the fact that amphibians emerged before the divergence of reptiles/birds and mammals, this deletion must have occurred independently in the amphibian and neognathous lineages. Alternatively, Hp has been lost before the emergence of amphibians and has then reappeared independently in mammals and neognathous. This possibility is less likely in view of the fact that we found a Hb-binding protein similar to the β-chain of mammalian Hp in serum of Chinese soft-shell turtle (data not shown); turtles seem to represent a separate reptilian class (27).

Conclusions

Our results show that the Hp gene appeared early in vertebrate evolution, between the emergence of urochordates and bony fish. During this period, blood cells with high concentrations of Hb also seem to have emerged (28), supporting the idea that the function of early Hps, like its modern counterpart, was to protect against the oxidative damage caused by released Hb. Lamprey, a jawless fish, is the most primitive extant vertebrate with erythrocytes containing a high Hb concentration. When its genome sequence becomes available, it will be interesting to determine whether it has a gene coding for a HpL. Our results also show that in two classes of vertebrates (amphibians and neognathous birds) at least one species has lost the Hp gene. The conditions that induced this change remain to be elucidated.
Materials and Methods

Blood and Plasma Samples. Chicken (Gallus domesticus) blood and koi carp (C. carpio) blood was from the National Veterinary Institute (Uppsala, Sweden). Common ostrich (Struthio camelus) blood was from a local ostrich farm, and African clawed frog (X. laevis) blood was a gift from O. Wrabne (Karolinska Institute, Stockholm). Plasma from domesticated goose (Anser anser) and Chinese soft-shell turtle (Trionyx sinensis) were from Sigma. Plasma of domesticated pufferfish (T. rubripes) was a gift from T. Miyadai (Fukui Prefectural University, Fukui, Japan).

Genes and Genomes. The genome sequences of baker’s yeast (S. cerevisiae), a nematode (C. elegans), honey bee (A. mellifera), fruit fly (D. melanogaster), sea squint (C. intestinalis), spotted green pufferfish (T. nigroviridis), Japanese pufferfish (T. rubripes), zebrafish (D. rerio), chicken (red jungle fowl, G. gallus), and human were analyzed with tools provided by the ENSEMBL project of the European Molecular Biology Laboratory–European Bioinformatic Institute and the Wellcome Trust Sanger Institute (www.ensembl.org). The genome of the African clawed frog (X. tropicalis) was studied by using tools of the Department of Energy Joint Genome Institute portal (http://genome.jgi-psf.org). Peptide and DNA databases were searched by using BLASTP and TBLASTN tools, respectively, with standard settings and the sequence of human HP1 (gi:184316) as the query. The complete Hg genes of D. rerio and T. rubripes were predicted from their genomic sequences by using the WISE program (www.ebi.ac.uk/Wise2). Protein sequences of previously known mammalian Hps were obtained from the UniProt database (Universal Protein Resource, www.uniprot.org). Additional full-length fish Hp sequences were obtained from the EMBL project of the European Molecular Biology Laboratory–European Bioinformatic Institute and the Wellcome Trust Sanger Institute (www.ensembl.org). The genome of the African clawed frog (X. tropicalis) was studied by using tools of the Department of Energy Joint Genome Institute portal (http://genome.jgi-psf.org). Peptide and DNA databases were searched by using BLASTP and TBLASTN tools, respectively, with standard settings and the sequence of human HP1 (gi:184316) as the query. The complete Hp genes of D. rerio and T. rubripes were predicted from their genomic sequences by using the WISE program (www.ebi.ac.uk/Wise2). Protein sequences of previously known mammalian Hps were obtained from the UniProt database (Universal Protein Resource, www.uniprot.org). Additional full-length fish Hp sequences were obtained from nucleotide sequences obtained from GenBank (www.ncbi.nih.gov) or the ENSEMBL project (Data Set 1).

Protein Sequence Analysis and Phylogeny. Signal peptides were predicted by using SIGNALP 3.0 (www.cbs.dtu.dk/services/SIGNALP) (29). Alignment of amino acid sequences was performed by using the CLUSTALW program (30) and then manually corrected. Unrooted phylogenetic trees were constructed from the aligned sequences by using the DRAWTREE program from the same package.

Isolation of Hb-Binding Proteins. Freshly isolated erythrocytes obtained from various species, as indicated, were washed five times with PBS and suspended in 5 vol of ice-cold 10 mM Tris-HCl, pH 8.0. The lysate was then centrifuged for 20 min at 40,000 × g at 4°C. The supernatant was collected, and Hb was isolated by gel filtration on a Superose 12 column. Fractons containing the protein were pooled and stored at −20°C. Purified Hb was coupled to CNBr-Sepharose (Amersham Pharmacia Biosciences) according to the manufacturer’s protocol. For the isolation of Hb-binding proteins, serum from various species, as indicated, was diluted three times with PBS and allowed to flow through the Hb-containing gel twice. The gel was then washed with 25 vol of PBS, and bound proteins were eluted with urea.

Identification of Proteins. Proteins were separated by SDS/PAGE in 12.5% gels and stained with colloidal Coomassie brilliant blue, and selected bands were excised. After trypsin treatment, peptide masses were determined with a MALDI-TOF mass spectrometer and analyzed with the Mascot Search Engine of Matrix Science, London (www.matrixscience.com). MALDI-TOF analyses were performed at the Expression Proteomics Laboratory in Uppsala. Liquid chromatography tandem MS analyses were performed at the Protein Analysis Center (Karolinska Institute, Stockholm). N-terminal protein sequence analysis was performed by BioCentrum (Krakow, Poland).

Cell Cultures and Transfections. Human embryonic kidney cells (HEK293) and green monkey kidney endothelial cells (COS-7) were grown in complete DMEM supplemented with 10% FBS, glutamine, and antibiotics. Cells were transfected with 2 μg DNA by using Lipofectamine (Invitrogen). We thank T. Miyadai, J. Harldig (National Veterinary Institute, Uppsala, Sweden), O. Wrabne, and P. Stenius (Stigtomta Struts, Stigtomta, Sweden) for providing plasma and blood samples and H. Schiot, O. Zetterqvist, and H. Ronne for critical comments on the manuscript. This work was supported by the Swedish Research Council and the E. O. and Edla Johansson Scientific Foundation. MALDI-TOF analyses performed at the Expression Proteomics Laboratory in Uppsala were supported by Wallenberg Consortium North Grant KAW 2000.0013 from the Knut and Alice Wallenberg Foundation.