Daphnia pulex Didomain Hemoglobin: Structure and Evolution of Polymeric Hemoglobins and Their Coding Genes

S. Dewilde,* M.-L. Van Hauwaert,* K. Peeters,* J. Vanfleteren,† and L. Moens*†

*Department of Biochemistry, University of Antwerp, Antwerp, Belgium; and †Department of Biology, University of Ghent, Ghent, Belgium

The high-molecular-weight extracellular hemoglobin of Daphnia pulex is composed of at least three different didomain globin chains. The primary structure of one of these chains was determined at the protein and cDNA levels. Each globin domain of the polypeptide chain displays the standard structural characteristics. The first domain is preceded by a 30-residue extension containing an 18-residue unprecedented threonine-rich segment and a 12-residue preA segment which is homologous to the preA segments of other nonvertebrate globin chains. Both domains are linked together by a preA′ segment, which is homologous to other preA segments and lacks the threonine-rich segment. Dimerization of the globin chains by the formation of a disulfide bridge linking the unique cysteines near the amino-termini results in a covalent, vertebrate-like tetradomain structure. The flexible amino-terminal extension most likely facilitates dimerization. The gene coding for this globin chain is interrupted by six small introns. Each domain displays two intradomain introns at the conserved positions B12.2 and G7.0. A precoding intron occurs at position preA(-27.0) and a bridge intron at occurs preA′(-13.2). We propose a crossover event as the most likely mechanism for duplication. Arthropod globin trees reflect the added effects of gene diversification, gene duplication, and species evolution. The position of monodomain intracellular globins in the tree suggests that they resemble the ancestral globin more than the derived didomain extracellular globins do.

Introduction

Unlike vertebrates, in which hemoglobins (Hb’s) are mainly in erythrocytes, many invertebrates display extracellular Hb’s dissolved in the hemolymph. It is generally assumed that extracellular Hb’s must acquire high Mr values to minimize elimination from the hemolymph by excretory processes. High Mr values (10⁶ to 1.2 x 10⁷) are obtained by aggregation of many myoglobin-like low Mr chains (M_r ~17,000) into a functional molecule (e.g., in Annelida) or by the concatenation of globin domains into polymers of Mr 32,000–300,000 (Mollusca and Arthropoda; Vinogradov et al. 1993; Suzuki and Imai 1998). The only reported exceptions to this rule are the Hb’s of the chironomids (M_r ~17,000).

Didomain globins are well characterized in the extracellular high-affinity Hb’s of parasitic nematodes such as Ascaris suum and Pseudoterranova decipiens and in the intracellular high-affinity Hb’s of the clams Barbatia reeveana and Barbatia limacina, as well as in the Hb of Daphnia magna (De Baere et al. 1992; Dixon et al. 1992; Tokishita et al. 1997). The only polymeric Hb characterized until now is of the crustacean Artemia sp. (Manning, Trotman, and Tate 1990; Trotman et al. 1994; Jellie, Tate, and Trotman 1996).

Several models have been proposed to explain polypeptide gene duplications leading to the formation of a globin gene of the clam B. reeveana occurred by a crossover event involving the precoding introns and the 3′ noncoding regions of two identical or very similar genes coding for a single-domain globin. This resulted in the joining of both globin domains without creating an additional bridging sequence. (2) Manning, Trotman, and Tate (1990) and Jellie, Tate, and Trotman (1996) invoked a series of gene duplications or unequal crossovers and fusions to explain the nine-domain Hb of Artemia starting from an ancestral monodomain globin gene (monodomain → didomain → tridomain → nonadomain). Within the polymer, the domains are connected by an approximately 14-amino-acid-long linker sequence containing a consensus sequence of -Val-Asp-Pro-Val-Thr-Gly-Leu-, resembling the Petromyzon preA globin sequence (Trotman et al. 1994). (3) In contrast, the nematode didomain globins are devoid of any linker sequence, since residue H23 of the first domain immediately precedes residue A2 of the second domain, and several putative mechanisms were considered (De Baere et al. 1992; Sherman et al. 1992; Dixon et al. 1992).

Intron origin and function are still highly controversial (Trotman 1998), with two opposing hypotheses. The “intron early or exon” theory of genes claims that genes coding for some modern complex proteins evolved through the fusion of primordial minigens (exons) coding for compact protein domains. Introns represent the noncoding linkers between the original minigens. The “formative” mode of intron origin explains the essential conservation of the intron positions in homologous genes throughout the phyla (Go 1981; Gilbert 1987). According to the “intron late” hypothesis, introns are inserted during evolution into preexisting genes by a mechanism resembling transposition. The “insertional” nature of such introns suggests that their positions were originally random (Cavalier-Smith 1991; Palmer and Logsdon 1991). To test each hypothesis, large protein and gene data sets were analyzed for intron
patterns and insertion positions and for correlations between protein and gene structure. Results, however, were statistically marginal in both instances (Stoltzfus et al. 1994, 1997; Gilbert, Souza, and Long 1997).

The globin gene family usually displays two or three introns per globin domain. Two introns are highly conserved throughout the phyla at positions B12.2 (insertion after the second base of the codon of amino acid 12 of the B-helix) and G7.0 (insertion between the codons of the sixth and seventh amino acids of the G-helix) (Long, Rosenberg, and Gilbert 1995). These introns might be of formative origin. Deviations from these positions are explained variously by intron sliding (Gilbert, Souza, and Long 1997) or by independent intron loss followed by insertion at another position (Stoltzfus et al. 1997). The insertion position of the central, or E-helix, intron is quite variable, suggesting an insertional origin (Hankeln et al. 1997; Trotman 1998).

Arthropod globin genes may therefore contain both types of intron. Formative introns could be present in *Gastrophilus intestinalis* (G7.0) and *Artemia* (B12.2 and G7.0) globin genes, whereas the E-helix introns in chironomids and the D-helix intron in *G. intestinalis* could be of insertional origin (Kao, Trewitt, and Terwilliger 1994; Hankeln et al. 1997; Dewilde et al. 1998).

The hemolymph of *Daphnia pulex* contains several (1–4) isoforms of Hb in various concentrations depending on the oxygen tension of the environment. They all bind oxygen with moderate affinity (P50: 0.95–2.67 mm Hg) and cooperativity (n = 1.4 to 2.3) and a positive Bohr effect (Fox 1949; Wolf, Smet, and Declerq 1983). These Hb molecules have an apparent Mr of 420,000–460,000 (S20,w = 17.05) and are composed of 12 didomain globin chains of Mr 32,000–35,000 (table 1).

We describe the structure of one didomain globin chain and its coding gene of *D. pulex* (Crustacea, Branchiopoda, Cladocera) and discuss the phylogenetic and evolutionary interpretation of this structure.

Materials and Methods

Isolation of *Daphnia* Hemoglobin and Globin Chains

*Daphnia pulex* and *Daphnia carinata* were collected from a local pond. Contamination with other organisms was negligible. After extensive washing, live animals were homogenized in 50 mM Tris-HCl (pH 7.5), 100 mM NaCl in the presence of a protease-inhibiting cocktail (Complete, Boehringer). A crude Hb fraction was purified as previously described (Peeters et al. 1990). Sample purity was checked by one- and two-dimensional electrophoresis (Laemmli 1970; Görg et al. 1995). Globin chains were separated by reverse phase–high performance liquid chromatography (RP-HPLC) on a Bakerbond wide-pore C-4 column using a 0.1% TFA/CH3CN system.

Partial Sequence Determination of Chain β

Intact globin β was subjected to Edman degradation. The total Hb fraction was fragmented by CNBr and partial acid hydrolysis (Allen 1989). Resulting fragments were purified by sodiumdodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Swank and Munkres 1980), electroblotted onto polyvinylidene difluoride (PVDF), and sequenced in an ABI 471B sequencer as recommended by the manufacturer.

Derivation of Degenerate Oligonucleotides

Two degenerate primers were designed based on the obtained amino acid sequence. DaF1, 5'-CAYCCNGARTAYCARAARATG-3', a 21mer, corresponds to the sense strand predicted by the peptide fragment HPEYQKM. DaF2, 5'-GCNCCRAAYGTYTCRAACAT-3', a 20mer, corresponds to the antisense strand predicted by the peptide fragment of MFEQFGA. Both primers were situated in domain 1. cDNA Sequencing

*Daphnia pulex* were pulverized under liquid nitrogen, and total RNA was isolated using the Total RNA Isolation Kit from Stratagene. First-strand cDNA was synthesized using DaF2, and PCR was performed using the two degenerate primers. PCR was carried out for 35 cycles of 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min. A positive fragment was subsequently purified and cloned in a pGEM vector (Promega). The sequences of several clones were determined with cycle sequencing and the use of a 377 ABI sequencer. Based on this sequence, a specific primer, DaF3 (5'-TTGGCTAACCAAATCAACGC-3'), was designed, and a PCR of 30 cycles of 1 min at 94°C, 1 min at 50°C, and 2 min at 72°C was set up using the primers DaF3 and oligoT. Another PCR was then done by inserting a sterile tip into the gel at the site of the banded fragment and inoculating a new PCR mix (Hengen 1995). The band of interest was then excised out of the Separide Gel Matrix (GibcoBRL), purified, cloned, and sequenced as described above. To determine the 5’ end of the sequence, RACE was done as previously described (Dewilde et al. 1996). First-strand cDNA was synthesized using the specific primer DaF10 (5’-CA-
Genomic DNA Sequencing

The complete sequence of the *D. pulex* globin gene was determined by amplifying two overlapping fragments. Two PCRs were set up with, respectively, the primers DafF11 (5′-TTCCCCAGAACTTTG-3′) and DafR12 (5′-AAATGGGTAAGTAGCAGGACG-3′) and the primer of DafF19 (5′-GATATTTTCAGGACAATTGCAAC-3′) and DafR14 (5′-CTGGTGAGTGTAGCGCATGT-3′). Both PCRs were carried out for 35 cycles of 1 min at 94°C, 1 min at 55°C, and 3 min at 72°C. The amplified products were purified, cloned, and sequenced.

Alignment and Tree Construction

The alignment of globins in our database was improved manually according to the alignment of nonvertebrate globins with known tertiary structures (Bolognesi et al. 1997). Penalty scores for the nonvertebrate template were calculated for selected globin sequences (Moens et al. 1996). As the three-dimensional structure of the didomain Hb of *D. pulex* is unknown, it is impossible to define the borders of the helices with accuracy. Therefore, we defined the preA helix as the amino acid sequence preceding the A-helix identified by known tertiary structures of globins, and we defined the preA’ helix as the sequence between the H-helix of the first domain and the A-helix of the second domain (fig. 4, preA segment in bold).

Neighbor-Joining trees were constructed using the Poisson formula as implemented in TREECON without taking gaps into account. The confidence of the trees obtained was tested by bootstrap analysis with 1,000 replicates. Nodes were considered reliable if they had a bootstrap value of at least 70% (Hillis and Bull 1993; Van de Peer and De Wachter 1994).

Results and Discussion

Determination and Interpretation of the *D. pulex* Hemoglobin Primary Structure

Separation of *D. pulex* and *D. carinata* globins by one- and two-dimensional electrophoresis revealed three major globin chains, α, β, and γ. The α and β chains are involved in the formation of disulphide-linked dimers of the types α-α, α-β, and β-β, whereas the γ chain is always present as the monomer. The native Hb’s are ill-defined noncovalent aggregates of high *M*₅ (fig. 1211).
1A–C). Due to their similar physicochemical characteristics, the globin chains are very difficult to separate. Only enriched fractions of the β chains of *D. pulex* and *D. carinata* could be obtained by RP-HPLC. A partial amino-terminal sequence was determined for both β chains as well as of some peptides generated by partial acid hydrolysis and CNBr cleavage of a total *D. pulex* globin fraction (Peeters et al. 1990). The obtained sequence data, aligned with the final cDNA-derived amino acid sequence, are summarized in fig. 2. Oligonucleotide primers were designed based on conserved regions of this sequence, and the full cDNA sequence of *D. pulex* globin was determined as described in Materials and Methods.

The cDNA sequence encompasses the entire coding region of a globin chain. The initiation codon is preceded by a 33-base-long 5′ UTR, followed by 352 codons and a 96-base 3′ UTR, excluding the polyadenylation tail (fig. 3).

The cDNA-derived amino acid sequence encodes for a didomain globin chain with a 21-amino-acid-long leader sequence. It is three amino acids longer than the leader sequence of *D. magna* and contains an unusual serine stretch of five residues. Otherwise, the classical features of a signal peptide are present, including the presence of a positively charged residue around the N-terminus (Lys) and a hydrophobic stretch at the central region. The peptides sequenced at the protein level differ from the homologous regions in the final cDNA-derived sequence (fig. 2). This means that they originate from different globin chains which have almost identical sequences chosen for primer design (5′ of DafF1 differs).

The *D. pulex* globin sequence can be aligned unambiguously with selected arthropod globins using conserved key residues of globin structure, resulting in a low penalty score against globin templates (fig. 4 and table 2). The following important structural or functional positions can be mentioned: (1) at B6 and B7 of the first domain, a Pro occurs, resulting in a bent B-helix and a looser interaction between the B- and E-helices; (2) at the distal position, E7, a Gln occurs in both daphnids, in contrast to other arthropod globin sequences in which a His is found; (3) in domains 1 and 2, Ile and Val occupy position E11, respectively, which is in accordance with most other nonvertebrate globins; (4) Trp occurs at position H8 of both domains, as usual; and (5) Leu and Phe occupy position B10 in domains 1 and 2, respectively, possibly influencing the ligand-binding affinity; indeed, the nature of the residue at position B10 can have a fundamental effect on oxygen affinity. A Tyr residue in this position protrudes in the heme cavity and is potentially able to form a hydrogen bridge with the bound oxygen, resulting in a low k_{off} value as observed in nematode Hb’s (De Baere et al. 1994; Yang et al. 1995). Formation of this hydrogen bond also depends on the conformation of the side chain, however.

The globin sequences of *D. magna* and *D. pulex* reveal a very high degree of similarity when corresponding domains are compared (domain 1: 80.6%; domain 2: 81.2%; table 3). The majority of substitutions are at external positions and are physicochemically conserved with respect to the hydrophobicity, size, and charge of the side chains (Lesk and Chothia 1980). There are exceptions. In domain 1, a B15 Lys and an H13 Ala in *D. magna* are, respectively, substituted by a Val and an Arg in *D. pulex*. In domain 2, a C6 Phe and an H22 Ile in *D. magna* are, respectively, substituted by a His and a Lys in *D. pulex*. The effects of these substitutions cannot yet be predicted.

The most striking characteristics of both daphnid globin sequences (*D. pulex*, *D. carinata*, and *D. magna*), however, are the presence of a ~30-amino-acid-long, Thr-rich amino-terminal extension and the occurrence of a Cys at the amino-terminal position. Both characteristics are unprecedented in other globin chains, and a Cys at the amino-terminus is rare in proteins in general. The amino-terminal extension can be divided into a threonine-rich segment (*D. pulex*: residues (~13)–(~30)) and a preA segment (*D. pulex*: (~1)–(~12) by analogy with Scapharca inaequivalvis, where “preA” defines the extended NA segment of nonvertebrate globin sequences that precedes the onset of the A-helix. We define the amino acid preceding the A1 helical position as (~1).

The threonine-rich segment seems to be a unique feature of the *Daphnia* globin sequences (Tokishita et al. 1997), whereas the preA segment is homologous to the preA segments of many other nonvertebrate globin chains (table 4). Thr-rich stretches do occur in proteins and are mainly sites of extensive glycosylation. However, *D. pulex* and *D. magna* Hb’s are most likely not glycosylated, as they leave a ConA Sepharose column un retarded (Peeters et al. 1990). Also, no glycosylation sites (NXT/S) can be found in the globin sequences (Tokishita et al. 1997). The secondary structure of this threonine-rich segment is doubtless not helical but most likely a random coil or beta sheet. The preA is clearly homologous to the preA segment. This confirms that no supplementary sequence is used to link both domains but that this function is taken over by the preA segment of the second domain.

Cysteines are not uncommon in the amino-terminal segments of globin chains (40/173 nonvertebrate globin sequences in our database; unpublished data). They occur mainly at position A8 (nematodes) and either one (preA(-1); annelids) or two (preA(-2); annelids and mol luscs) positions before the start of the A-helix. They are frequently involved in the formation of intermolecular disulphide bridges, resulting in dimers (annelids, molluscs, and some nematodes) or tetrarimers (some nematodes) essential to build high *M* aggregates in order to avoid elimination from the hemolymph by excretory processes (Vinogradov et al. 1993; Kapp et al. 1995). Daphnid Hb contains disulphide-linked homo- and heterodimers of *M*, 70,000–78,000 (fig. 1 and table 4). These dimers arise by the coupling of monomers through the unique cysteine in the first position of the globin sequence, resulting in a covalent tetradomain complex. This complex is a building block of the high-*M* extracellular Hb (*M*; ~420,000–460,000) and most likely has a stable quaternary structure and potential cooperativity. The observed Thr-rich segment may serve
Fig. 2.—Partial amino acid sequences of *Daphnia pulex* and *Daphnia carinata* generated by protein sequencing. The obtained fragments are aligned with the *D. pulex* globin sequence translated from globin cDNA (fig. 3; leader sequence not shown). The sequence used for primer synthesis is in bold and underlined. N-term = fragment obtained by N-terminal sequencing of *D. pulex* and *D. carinata* chain β; AH = fragment obtained by partial acid hydrolysis; CNBr = fragment obtained by CNBr cleavage. Da. car = *D. carinata*; Da. pd1, 2 = *D. pulex* domains 1, 2; Da. pul = *D. pulex*.

![Fig. 2](image)

Fig. 4.—Alignment of selected arthropod globin sequences. *Phys* = *Physeter catodon* Mb; CTT III = *Chironomus thummi thummi* globin III; Tok V, VII = *Tokunagayusurika* globin V, VII; Gas1 = *Gastrophilus intestinalis* globin 1; Arte T3, 4, 6 = *Artemia* sp. Hb domains T3, 4, 6; Da. md1, 2 = *Daphnia magna* domains 1, 2; Da. pd1, 2 = *Daphnia pulex* domains 1, 2. The preA and preA’ segments are in bold.
as a kind of extended linker that provides the flexibility to allow "tetramerization" of both didomain molecules.

Phylogenetic Analysis

Figure 5 shows a Neighbor-Joining tree relating 38 arthropod globin sequences (TREECON; Van de Peer and De Wachter 1994). This tree reflects species phylogeny as well as the evolution of the genes, complicating correct interpretation of the relationships. Indeed, globins constitute a multigene family, and tracking of orthologous genes may be difficult. The rapid radiation of metazoan taxa would also result in poor bootstrap support for the deeper-branching points. Given these constraints, it is clear that the intracellular globin of *G. intestinalis* branches off before the extracellular arthropod globins. This reflects that the monomeric intracellular globins resemble the ancestral globin structure more than the more diverged extracellular types do. The ancestral nature of this globin is also reflected in the structure of the encoding gene. The *G. intestinalis* Hb1
gene is interrupted by two introns at positions D7.0 and G7.0, whereas most chironomid globin genes have lost all introns (Dewilde et al. 1998). Most domains of the *Artemia* globin genes have two introns inserted at the ancestral positions B12.2 and G7.0. Thus, we assume that the *G. intestinalis* globin gene retained one ancestral intron and secondarily acquired a novel intron in D7.0. The extracellular globins form two clusters as expected, one leading to the di- or multidomain crustacean globins, and the other leading to the monodomain insect globins. As also observed by Goodman et al. (1983) for the *Chironomus* lineage, the genes encoding for monomeric Hb’s (*Chironomus thummi* I, IA, III, IIIA, IV) branch off first, with the exception of *Chironomus thummi* thummi HbX, which can be either monomeric or dimeric. The genes encoding the dimeric *Chironomus* Hb’s (*C. thummi* IIB, VI, VIIA, VIIIB, VIII, IX) form a distinct cluster. This dichotomy correlates with the clustering of genes encoding monomeric Hb in one chromosomal region, whereas those encoding dimeric Hb are clustered on another chromosomal region. Phylogenetic radiation is reflected by the clustering of the *Tokunagayusurika* and chironomid globins and the globins from *Daphnia* and *Artemia*.

A Neighbor-joining tree constructed using the sequences aligned in figure 4 and all known multidomain globin sequences clearly demonstrates that similarity between domains can fail to coincide with taxon delinea-
tion (table 3 and fig. 6). Our phylogenetic analyses confirm the results of Tokishita et al. (1997). Both *Daphnia* globins are clearly built up from two different domain types (table 3). The most parsimonious explanation is that the events leading to the formation of didomain globin chains preceded speciation. This is especially clear for *D. pulex* and *D. magna*. For *Barbatia*, the similarity of the two globin domains is much greater, suggesting a more recent duplication event (Suzuki and Imai 1998), although still preceding species divergence.

Globin Gene Structure and the Evolution of Didomain Globins

Comparison of the globin gDNA with the cDNA reveals that the coding sequence is interrupted by six introns (fig. 3). All introns display the standard splicing donor and acceptor sites. Their sizes are similar to those of the dipteran globin introns but are much smaller than those of *Artemia* (table 5). This agrees well with the hypothesis that the intron insertion position, rather than the intron length, has structural and evolutionary importance (Dewilde et al. 1998). There is, however, a slightly different splicing mechanism depending on the size of the intron and the presence of splice site consensus sequences (e.g., 5’ and 3’ splice sites, polypyrimidine tract). The small introns of *Drosophila*, for example, are too small to be functional in a vertebrate system (Talerico and Berget 1994).

Table 2
Penalty Scores According to the Nonvertebrate Template (Moens et al. 1996)

<table>
<thead>
<tr>
<th>Species</th>
<th>A Pattern</th>
<th>BC Pattern</th>
<th>E Pattern</th>
<th>FG Pattern</th>
<th>H Pattern</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Chironomus thummi</em></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>Tokunagayusurika</em></td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td><em>Tokunagayusurika</em> VII</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><em>Gastrophilus intestinalis</em></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.7</td>
<td>2.7</td>
</tr>
<tr>
<td><em>Artemia</em> sp. T3</td>
<td>0</td>
<td>0</td>
<td>1.2</td>
<td>2.2</td>
<td>0</td>
<td>3.4</td>
</tr>
<tr>
<td><em>Artemia</em> sp. T4</td>
<td>0</td>
<td>0</td>
<td>1.9</td>
<td>0.7</td>
<td>0.7</td>
<td>3.3</td>
</tr>
<tr>
<td><em>Artemia</em> sp. T6</td>
<td>0</td>
<td>1</td>
<td>0.9</td>
<td>1.7</td>
<td>0</td>
<td>3.6</td>
</tr>
<tr>
<td><em>Daphnia magna</em> Hb domain 1</td>
<td>0</td>
<td>0.2</td>
<td>1</td>
<td>0.7</td>
<td>0</td>
<td>1.9</td>
</tr>
<tr>
<td><em>D. magna</em> Hb domain 2</td>
<td>0</td>
<td>0</td>
<td>2.6</td>
<td>1</td>
<td>0</td>
<td>3.6</td>
</tr>
<tr>
<td><em>Daphnia pulex</em> Hb domain 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.9</td>
<td>0</td>
<td>1.9</td>
</tr>
<tr>
<td><em>D. pulex</em> Hb domain 2</td>
<td>0</td>
<td>0</td>
<td>2.1</td>
<td>0.7</td>
<td>0</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Table 3
Percentages of Similarity Between the Didomain Globins

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Dmd1</th>
<th>Dmd2</th>
<th>Dpd1</th>
<th>Dpd2</th>
<th>Asd1</th>
<th>Asd2</th>
<th>Pdd1</th>
<th>Pdd2</th>
<th>Brd1</th>
<th>Brd2</th>
<th>Bld1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dmd1</td>
<td>23.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dpd2</td>
<td>80.6</td>
<td>22.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asd1</td>
<td>23.6</td>
<td>81.2</td>
<td>22.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asd2</td>
<td>18.4</td>
<td>16.4</td>
<td>19.8</td>
<td>12.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pdd1</td>
<td>11.9</td>
<td>15.9</td>
<td>11.9</td>
<td>14.4</td>
<td>63.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pdd2</td>
<td>16.2</td>
<td>17.2</td>
<td>16.9</td>
<td>13.5</td>
<td>74.0</td>
<td>61.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brd1</td>
<td>16.2</td>
<td>17.9</td>
<td>15.4</td>
<td>14.8</td>
<td>52.7</td>
<td>59.2</td>
<td>53.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brd2</td>
<td>17.6</td>
<td>13.6</td>
<td>16.9</td>
<td>13.1</td>
<td>13.2</td>
<td>12.6</td>
<td>12.5</td>
<td>12.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bld1</td>
<td>19.2</td>
<td>15.3</td>
<td>15.8</td>
<td>14.6</td>
<td>15.4</td>
<td>13.2</td>
<td>13.2</td>
<td>15.9</td>
<td>79.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bld2</td>
<td>17.8</td>
<td>13.1</td>
<td>17.8</td>
<td>12.5</td>
<td>15.4</td>
<td>14.7</td>
<td>13.2</td>
<td>14.5</td>
<td>88.8</td>
<td>78.8</td>
<td></td>
</tr>
<tr>
<td>Dmd2</td>
<td>17.6</td>
<td>13.7</td>
<td>14.9</td>
<td>13.0</td>
<td>14.6</td>
<td>13.9</td>
<td>16.0</td>
<td>16.6</td>
<td>74.5</td>
<td>87.5</td>
<td>75.2</td>
</tr>
</tbody>
</table>

*Note.*—Dmd1 = *Daphnia magna* domain 1; Dmd2 = *D. magna* domain 2; Dpd1 = *Daphnia pulex* domain 1; Dpd2 = *D. pulex* domain 2; Asd1 = *Ascaris suum* domain 1; Asd2 = *A. suum* domain 2; Pdd1 = *Pseudoterranova* decipiens domain 1; Pdd2 = *P. decipiens* domain 2; Brd1 = *Barbatia reeveana* domain 1; Brd2 = *B. reeveana* domain 2; Bld1 = *Barbatia limacina* domain 1; Bld2 = *B. limacina* domain 2.
Table 4
Alignment of the PreA/Linker Regions of Multidomain Globins

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HHHHHHHHHHHHHH-CCC-</td>
<td>N</td>
<td>*NKTRELMKSL</td>
<td>*SKTRELMKSL</td>
<td>*SVSAKLDVTP-QPANKNLIRSTW</td>
<td>*SVEDKIEEVT-QPANKNLIRSTW</td>
<td>*CSQAPGTITTTVTVTVTVSADDSEAGLL-SAHERSLRKTV</td>
<td>*CSQAPGTITTTVTVTVTVADEDSGSSL-SSHDRS1IRKYG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asc.D1</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pse.D1</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bar.D1</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal.D1</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Da.mD1</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Da.pD1</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asc.D1/D2</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pse.D1/D2</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bar.D1/D2</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bal.D1/D2</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Da.pD1/D2</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Art.C3/C4</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Art.C4/C5</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Art.C6/C7</td>
<td>. . . . . . . . . . . . . . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*NOTE.*—The codon in which the procoding/bridge intron is inserted is in bold, and the amino-terminus is indicated with an asterisk.

**Fig. 5.**—Neighbor-Joining tree of most known arthropod globin sequences.
Based on the alignment of figure 4, it can be concluded that in both domains of the \textit{D. pulex} Hb gene, the intradomain introns are inserted into the conserved positions B12.2 and G7.0. In addition, a precoding (preA) intron and a bridge intron (preA') are present at positions preA(-27.0) and preA'(-13.2) (table 5). Unlike in some chironomid globin genes, no central intron is present. This might strengthen Hankeln’s and Kao’s conclusion that the chironomid central intron is of insertional nature and that the ancestral globin gene might have displayed a three-exon/two-intron pattern (Kao, Trewitt, and Bergstrom 1994; Hankeln et al. 1997; Trotman 1998).

The precoding and bridge introns were probably acquired, respectively, when the globin chain became extracellular and the gene duplicated (Hunt, Burr, and Blaxter, personal communication). Precoding and bridge introns have been found in didomain globins of several organisms. In the mollusc \textit{Barbatia reeveana}, the 5’ part of the bridge intron is derived from the 3’ noncoding region of the gene for the ancestral single domain chain. An unequal crossover event would explain these results (Naito et al. 1991). In \textit{D. pulex}, no sequence similarity can be seen between the introns and the 3’ UTR. However, table 3 shows that the similarity between the two domains of the same molecule in \textit{Daphnia} is much smaller than in \textit{Barbatia}, suggesting that gene duplication in \textit{Daphnia} has a longer history. Considering the instability of intron sequences during evolution, a polyadenylation signal, e.g., such as the one still seen in the bridge intron in \textit{Barbatia}, may well have been lost by mutation in \textit{Daphnia}. Thus, we assume that globin gene duplication in \textit{D. pulex} was also created by unequal crossover. Table 4 shows the alignment of the preA' region of several multidomain globins, with indication of the bridge intron insertion positions. In \textit{B. reeveana}, \textit{B. limacina}, \textit{Artemia} C4/C5, and the \textit{D. pulex} didomain globin, the intron is inserted in the codon of a Lys residue. In \textit{B. reeveana} and \textit{D. pulex}, a second Lys residue follows immediately. With a single mutation, the codon for Lys can form either a stop codon or a start codon, which would mark the two separate domains. In \textit{B. limacina}, a Met residue, potentially the relic of the initiation codon, is still present.

The didomain globins of the nematodes \textit{A. suum} and \textit{P. decipiens} also contain precoding and bridge in-

\begin{table}[h]
\caption{Exon/Intron Structures of Arthropod Globin Genes}
\begin{tabular}{|l|c|c|c|c|c|}
\hline
Species & Precoding/Bridge & B-Helix & Central & F/G/\(\text{GH}\)- & Reference				\
 & Intron & & Intron & Helix-Intron & \\
\hline
\textit{Chironomus thummi} HbII* & \ldots & \ldots & \ldots & \ldots & \ldots & Antoine and Niessing (1984) \\
\textit{C. thummi} HbX* & \ldots & \ldots & \ldots & E9.1: 68 & \ldots & Kao, Trewitt, and Bergstrom (1994) \\
\textit{Chironomus melanotus} HbII* & \ldots & \ldots & \ldots & E15.0: 55 & \ldots & Hankeln et al. (1997) \\
\textit{Glyptotendipes barbibipes} Hb* & \ldots & \ldots & \ldots & E12.2 & \ldots & Hankeln, personal communication \\
\textit{Chironomus tentans} Hb* & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
\textit{Gastrophilus intestinalis} Hb* & \ldots & \ldots & \ldots & D7.0: 71 & G7.0: 88 & Dewilde et al. (1998) \\
\textit{Artemia} sp. Hb T1 & N-term; 1,400 & B12.2: 1,500 & \ldots & G7.0: 4,300 & Jellie, Tate and Trotman (1996) \\
\textit{Artemia} sp. Hb T2 & \ldots & B12.2: 2,000 & \ldots & G7.0: 2,300 & Jellie, Tate and Trotman (1996) \\
\textit{Artemia} sp. Hb T3 & \ldots & B12.2: 1,300 & \ldots & F3.1; 1,100 & Jellie, Tate and Trotman (1996) \\
\textit{Artemia} sp. Hb T4 & \ldots; +; 1,100 & \ldots & \ldots & G6.2; 1,300 & Matthews and Trotman (1998) \\
\textit{Artemia} sp. Hb T5 & \ldots; +; 2,300 & B12.2: 1,500 & \ldots & G7.0: 2,000 & Matthews and Trotman (1998) \\
\textit{Artemia} sp. Hb T6 & \ldots; +; 2,300 & B12.2: 2,700 & \ldots & F3.0: 60 & Jellie, Tate and Trotman (1996) \\
\textit{Artemia} sp. Hb T7 & \ldots; +; 1,200 & B12.2: 2,700 & \ldots & G7.0: 1,400 & Matthews and Trotman (1998) \\
\textit{Artemia} sp. Hb T8 & \ldots; +; 1,200 & B12.2: 2,700 & \ldots & G7.0: 2,300 & Jellie, Tate and Trotman (1996) \\
\textit{Artemia} sp. Hb T9 & C-term; 1,600 & B12.2: 1,000 & \ldots & G7.0: 3,800 & Matthews and Trotman (1998) \\
\textit{Daphnia pulex} d1 & \ldots; PreA(-27.0); 89 & B12.2: 77 & \ldots & G7.0: 85 & This paper \\
\textit{D. pulex} d2 & \ldots; PreA’(-13.2); 107 & B12.2: 66 & \ldots & G7.0: 71 & This paper \\
\hline
\end{tabular}
\footnote{NOTE.—Insertion position and phase are given, along with intron length in base pairs. * Included as an example of the different intron insertions in chironomid globins.}
trons but lack additional preA helix sequences and lysine residues. Several models were advanced to explain the head-to-tail duplication of the nematode didomain globin genes (Dixon et al. 1992). The two domains in the globin of *P. decipiens* differ in length by 13 amino acids, and Dixon et al. (1992) suggested that a stretch of 13 amino acids was removed by an unequal-crossover event. Because of the similarity between the 5′ end of the first domain and the 5′ and 3′ ends of the second domain, these regions may have been involved in the unequal crossover. This would mean that the bridge intron was inserted later. This would also be the case if the duplication occurred by recombination events involving genomic DNA. A third possibility is that a similar crossover to that seen in *B. reeveana* has occurred. However, time eroded any sequence similarity between the preceding intron and the 3′ end of the original gene (the bridge intron). A fourth possibility is that the gene duplicated during a mispaired gene conversion event such as that outlined by Slightom, Blechl, and Smithies (1980), which is capable of producing genes that are a fusion of two original genes. In *Artemia*, two globin chains (T and C), each containing nine domains separated by linker sequences, are present. This multidomain evolved by multiplication and fusion of a monomeric globin gene ancestor, possibly through a tridomain intermediate. In addition to the intradomain introns, the *Artemia* globin genes contain five additional introns. Three are bridge introns inserted within the linkers of domains 3/4, 4/5, and 6/7. The other two are inserted in the N- and C-terminal extensions (table 5, Jellie, Tate, and Trotman 1996; Matthews and Trotman 1998). Thus, the duplication events are facilitated by the presence of bridge (interdomain) introns (Matthews and Trotman 1998).

We conclude that there are different mechanisms that can lead to the formation of di- or multidomain globin chains. In the case of *D. pulex*, the duplication of the globin has taken place according to the mechanism that was proposed for the *Barbatia* didomain globins (Naito et al. 1991).

**Acknowledgments**

The Fund for Scientific Research–Flanders (FWO) is gratefully acknowledged for grants to J.V. (a Research Director with the FWO) and L.M. A. Vierstraete is acknowledged for DNA sequencing. The anonymous referees are acknowledged for valuable suggestions. The sequence reported in this paper has been submitted to the GenBank database with accession number AF074722.

**LITERATURE CITED**


Antony Dean, reviewing editor

Accepted May 25, 1999