Diversity of stonefly hexamerins and implication for the evolution of insect storage proteins

Silke Hagner-Holler\textsuperscript{a}, Christian Pick\textsuperscript{b}, Stefan Girgenrath\textsuperscript{c}, James H. Marden\textsuperscript{c}, Thorsten Burmester\textsuperscript{b,*}

\textsuperscript{a}Institute of Zoology, Johannes Gutenberg University of Mainz, D-55099 Mainz, Germany
\textsuperscript{b}Institute of Zoology and Zoological Museum, University of Hamburg, D-20146 Hamburg, Germany
\textsuperscript{c}Department of Biology, Pennsylvania State University, USA

Received 4 April 2007; received in revised form 30 May 2007; accepted 1 June 2007

Abstract

Hexamerins are large storage proteins of insects in the 500 kDa range that evolved from the copper-containing hemocyanins. Hexamerins have been found at high concentration in the hemolymph of many insect taxa, but have remained unstudied in relatively basal taxa. To obtain more detailed insight about early hexamerin evolution, we have studied hexamerins in stoneflies (Plecoptera). Stoneflies are also the only insects for which a functional hemocyanin is known to co-occur with hexamerins in the hemolymph. Here, we identified hexamerins in five plecopteran species and obtained partial cDNA sequences from \textit{Perla marginata} (Perlidae), \textit{Nemoura} sp. (Nemouridae), \textit{Taeniopteryx burksi} (Taeniopterygidae), \textit{Allocapnia viscipara} (Capniidae), and \textit{Diamphipnopsis samali} (Diamphipnoidae). At least four distinct hexamerins are present in \textit{P. marginata}. The full-length cDNA of one hexamerin subunit was obtained (PmaHex1) that measures 2475 bp and translates into a native polypeptide of 702 amino acids. Phylogenetic analyses showed that the plecopteran hexamerins are monophyletic and positioned at the base of the insect hexamerin tree, probably diverging about 360 million years ago. Within the Plecoptera, distinct hexamerin types evolved before the divergence of the families. Mapping amino acid compositions onto the phylogenetic tree shows that the accumulation of aromatic amino acids (and thus the evolution of “arylphorins”) commenced soon after the hexamerins diverged from hemocyanins, but also indicates that hexamerins with distinct amino acid compositions reflect secondary losses of aromatic amino acids.

Keywords: Hexamerin; Molecular clock; Plecoptera; Phylogeny; Stoneflies; Storage protein

1. Introduction

All insect species investigated so far harbor large proteins in their hemolymph that are referred to as hexamerins. Sequence comparisons have shown that these proteins belong to the superfamily of arthropod hemocyanins, phenoloxidases, crustacean pseudohemocyanins and dipteran hexamerin receptors (Beintema et al., 1994; Burmester and Scheller, 1996; Burmester, 2001, 2002). Like arthropod hemocyanins, hexamerins usually consist of six identical or similar subunits in the range of 80 kDa, giving rise to native molecules of about 500 kDa (Telfer and Kunkel, 1991; Burmester, 1999). However, in contrast to the respiratory hemocyanins, hexamerins do not bind oxygen because the copper-binding histidines residues have been replaced by other amino acids (Beintema et al., 1994; Burmester, 2001, 2002).

In many insect species, hexamerins accumulate in the hemolymph to extraordinarily high concentrations, reaching up to 50% of the total salt-extractable proteins of the animal in premolt developmental stages (Scheller et al., 1990; Telfer and Kunkel, 1991). Therefore, hexamerins are thought to act mainly as storage proteins, which are used as a source of energy and amino acids during non-feeding periods, such as pupal or nymphal stages (Telfer and Kunkel, 1991; Burmester, 1999). Hexamerins have also...
been identified as constituents of the sclerotizing system of the insect’s cuticle (Peter and Scheller, 1991). They may also serve as carriers for ecysteroids (Enderle et al., 1983), juvenile hormone (Braun and Wyatt, 1996) and other organic compounds such as riboflavin (Magee et al., 1994). More recent studies suggest that hexamerins function in caste differentiation of termites via regulation of the juvenile hormone levels (Zhou et al., 2006, 2007). There is also evidence that some hexamerins play a role in the insect’s humoral immune response (Phipps et al., 1994; Beresford et al., 1997). Molecular phylogenetic analyses have shown a complex pattern of hexamerin evolution in some insect orders, with several hexamerin subtypes, such as the highly aromatic arylphorins or methionine-rich hexamerins occurring in parallel in a single species (Burmester et al., 1998; Burmester, 1999, 2001). Hexamerin sequences have been successfully used as markers for the inference of insect phylogeny (Burmester et al., 1998; Burmester, 1999).

Although hexamerins are probably ubiquitously present in insects, most information derives from investigations of dipteran and lepidopteran species, whereas data from other taxa are scarce (for a review see Telfer and Kunkel, 1991; Burmester, 1999). Here, we investigate for the first time the hexamerins from Plecoptera (stoneflies). Stoneflies are thought to have arisen near the base of the winged insects and have retained several ancient morphological and behavioral features (Hennig, 1969, 1981). Recently, we have demonstrated that the stonefly *Perla marginata* harbors a functional hemocyanin in its hemolymph (Hagner-Holler et al., 2004). Thus, the Plecoptera are the only known insect order which possesses both types of hexameric proteins in their hemolymph. This observation prompted us to investigate the occurrence and evolution of hexamerins in selected stonefly species.

2. Material and methods

2.1. Analyses of hemolymph proteins

Hemolymph was withdrawn from the dorsal abdomen of living *P. marginata* nymphs (1–3 years) and adults by the aid of a syringe, and immediately diluted with an organic compounds such as riboflavin (Magee et al., 1994). Hemolymph was withdrawn from the dorsal abdomen of living *P. marginata* nymphs (1–3 years) and adults by the aid of a syringe, and immediately diluted with an approximately equal volume of 100 mM Tris–HCl, pH 7.5, 10 mM Ca²⁺, 10 mM Mg²⁺. Hemocytes and cell debris were removed by 10 min centrifugation at 10,000 × g at 4 °C. The total protein concentration in the hemolymph was determined according to the method of Bradford (1976). Denaturing SDS-PAGE was performed on 7.5% gels according to standard procedures. After electrophoresis, the gel was fixed in ethanol/acetic acid and stained with 0.1% Coomassie Brilliant Blue R-250.

2.2. RT-PCR cloning of stonefly hexamerins

Total RNA was extracted from adult and larval *P. marginata*, *Nemoura* sp., *Taeniapteryx burksi*, *Allocapnia vivipara* and *Diamphipnopsis samali* employing the guanidine-thiocyanate method (Chirgwin et al., 1979) or the urea procedure according to Holmes and Bonner (1973). We designed a set of degenerated oligonucleotide primers according to conserved amino acid sequences of insect hexamerins and hemocyanins (Burmester, 2001). The forward primers were: 5′-GAGGGNSAGTTCTGNTACGC-3′, 5′-CCNCCNCTAYGARRTCTACCC-3′; reverse primers: 5′-GAANGGYTTGTGGTNNAGRCG-3′, 5′-TCCG-TACTTGGGTCCNAGGAAGAC-3′. Reverse-transcription (RT)-PCR experiments were carried out applying the OneStep-kit according to the manufacturer’s instructions (Qiagen, Hilden, Germany). PCR-fragments of the expected size were cloned into the pCR4-TOPO®-TA or pGem-T Easy cloning vectors, and sequenced bidirectionally by a commercial service (GENEPRISE, Mainz, Germany) or on-site using an ABI Prism 377 DNA sequencer. Additional 3′ sequence of *T. burksi* and *D. samali* hexamerin was obtained by the 3′ rapid amplification of cDNA ends (RACE) method using RACE primers supplied with a kit (Roche, Indianapolis, IN, USA) in combination with gene specific PCR primers.

2.3. Molecular cloning of *P. marginata* hexamerin cDNA

A cDNA expression library from 3 years old *P. marginata* nymphs was constructed as previously described employing the STRATAGENE λ-ZAP system (Hagner-Holler et al., 2004). PCR-fragments of various *P. marginata* hexamerins were labeled with digoxigenin (Roche PCR labeling kit) and used as probes to screen the cDNA library. Positive phage clones were converted to pBK-CMV plasmid vectors using the material provided by STRATAGENE, and sequenced on both strands as described (GENEPRISE, Mainz).

2.4. Sequence analyses

Sequences were assembled with the Vector NTI 10 program (Invitrogen) or with the software package Lasergene™ (DNA STAR, Madison, WI, USA). The tools provided by the ExPASy Molecular Biology Server of the Swiss Institute of Bioinformatics (http://www.expasy.org) were used for the analyses of DNA and amino acid sequences. Signal peptides were predicted using the online version of SignalP V1.1 (Nielsen et al., 1997). An alignment of the stonefly hexamerins was constructed by hand using GeneDoc 2.6 (Nicholas and Nicholas, 1997). Statistical analysis was carried out with ANOVA, as implemented in the Microsoft EXCEL 2003 spreadsheet program. The Plecoptera sequences were added to an alignment of insect hexamerins and hemocyanins, as it has been used in previous studies (Burmester et al., 1998; Burmester, 1999). More recently available sequences have been added using ClustalX (Thompson et al., 1997) and manually adjusted with GeneDoc. Except for the plecopteran hexamerins, incomplete sequences were excluded from the analyses. A list of sequences used in this study is
provided in Supplemental Table 1. Amino acid compositions were deduced from the translated cDNA sequences.

2.5. Molecular phylogeny

The program packages PHYLIP 3.6b (Felsenstein, 2004), TREE-PUZZLE 5.2 (Strimmer and von Haeseler, 1997) and MrBayes 3.1 (Huelsenbeck and Ronquist, 2001) were used for the phylogenetic tree reconstructions. Distance matrices were calculated with TREE-PUZZLE using the WAG (Whelan and Goldman, 2001) model of amino acid evolution, assuming gamma distributions of substitution rates with eight categories. Neighbor-joining trees were inferred with the program NEIGHBOR from the PHYLIP package. The reliability of the branching pattern was tested by bootstrap analysis (Felsenstein, 1985) with 100 replications, employing the PUZZLEBOOT shell script (M. Holder and A. Roger). Bayesian phylogenetic analyses were performed by MrBayes, using the WAG model and assuming a gamma distribution of substitution rates. Prior probabilities for all trees were equal. Metropolis-coupled Markov chain Monte Carlo (MCMCMC) sampling was performed with one cold and three heated chains that were run for 1,000,000 generations. Starting trees were random, trees were sampled every 100th generation and posterior probabilities were estimated on the final 2000 trees (burnin ¼ 8000).

2.6. Molecular clock estimates

Linearized trees were calculated under the assumption of a relaxed molecular clock employing the program R8S 1.71 (Sanderson, 1997, 2002). We applied the Langley–Fitch maximum likelihood method (Langley and Fitch, 1974) with the Powell algorithm to estimate divergence times. The tree was calibrated with fossils constraints (Kukalova-Peck, 1991; Ross and Jarzembowski, 1993). Stratigraphic information was obtained from http://www.fossilrecord.net (Benton and Donoghue, 2007); numerical ages derived from the “International Stratigraphic Chart” (Gradstein et al., 2004). We assumed that brachyceran and nemato-
ceran Diptera diverged between 238.5 and 295.4 million years ago (MYA), Diptera and Lepidoptera between 281.5 and 307.2 MYA, Hymenoptera and Mecopteria (i.e., Diptera + Lepidoptera) more than 299.0 MYA, Isoptera and Blattaria more than 145.5 MYA, and that the first Hexapoda appeared at least 407.0 MYA.

3. Results

3.1. Plecopteran hexamerins

Hemolymph samples from 1- to 3-year-old larval and adult specimens of P. marginata were withdrawn and analyzed by SDS–PAGE. In all four stages, we observed several prominent bands in the range of 75–80 kDa, which is the typical mass of arthropod hemocyanins and insect hexamerins (Telfer and Kunkel, 1991; Burmester, 1999, 2001) (Fig. 1). The two lower bands at about 75 kDa were previously identified as hemocyanins by using specific antibodies in Western blotting (Hagner-Holler et al., 2004). In 3-years-old larvae and in adult stoneflies, multiple (about four) additional bands with higher molecular masses (about 80 kDa) appear, which we suspect to be hexamerins. Due to the lack of a specific antibody, we cannot conclusively evaluate this assumption. The molecular evidence presented below, however, demonstrates the presence of multiple hexamerin chains.

3.2. Molecular cloning and analyses of P. marginata hexamerins

We used an alignment of insect hexamerins and hemocyanins to deduce degenerate oligonucleotide primers. These primers were applied in reverse transcription-PCR on total RNA from 3-years-old nymphs and adults of P. marginata, which resulted in several cDNA products of about 1000 bp lengths. Cloning and sequencing of the cDNAs revealed that these fragments represent four different hexamerin sequences, which we named PmaHex1–4.
A cDNA library had been constructed from poly(A)+ RNA isolated from 3 years old \textit{P. marginata} nymphs (Hagner-Holler et al., 2004). By screening about 600,000 pfu with four different probes that are derived from these sequences, we obtained one full-length clone (PmaHex1) and two 5' incomplete clones (PmaHex3 and PmaHex4). PmaHex2 could not be identified in the library.

The cDNA sequence of PmaHex1 comprises a total of 2475 bp and covers an open reading frame of 2109 bp, beginning with a ATG at bp 13 (Supplemental Fig 1). The translation stop codon (TAA) is present at position bp 2119 and is followed by a polyadenylation signal (AATAAA) at bp 2441 and poly(A) tail of 18 bp. The incomplete sequences of PmaHex3 and PmaHex4 comprise 1899 and 1847 bp, respectively. By comparison with PmaHex1, we inferred that 408 and 308 bp of the 5' coding ends are missing in PmaHex3 and Hex4 (Supplemental Fig 2). The partial PmaHex2 sequence covers 1086 bp of the middle region, and misses ~450 bp of the 5' and ~600 bp of the 3' coding regions. Within the overlapping regions, the \textit{P. marginata} sequences are 65.8–87.6% identical on the DNA level and 54.0–83.1% identical on the amino acid level (Table 1).

The deduced primary structure of PmaHex1 yields a polypeptide of 702 amino acids with a calculated molecular mass of 82.52 kDa. Computer analysis of the sequence revealed the presence of a typical 16 amino acid signal peptide necessary for transmembrane transport and export into the hemolymph (von Heijne, 1986). Therefore, the native secreted PmaHex1 subunit comprises 686 amino acids with a mass of 80.95 kDa (pI 5.68), which is in excellent agreement with the data obtained from SDS-PAGE (Fig. 1). The six copper-binding sites essential for the function in O$_2$ binding are not conserved in the middle region, and misses 450 bp of the 5' and 600 bp of the 3' coding regions. Within the overlapping regions, the \textit{P. marginata} sequences are 65.8–87.6% identical on the DNA level and 54.0–83.1% identical on the amino acid level (Table 1).

The deduced primary structure of PmaHex1 yields a polypeptide of 702 amino acids with a calculated molecular mass of 82.52 kDa. Computer analysis of the sequence revealed the presence of a typical 16 amino acid signal peptide necessary for transmembrane transport and export into the hemolymph (von Heijne, 1986). Therefore, the native secreted PmaHex1 subunit comprises 686 amino acids with a mass of 80.95 kDa (pI 5.68), which is in excellent agreement with the data obtained from SDS-PAGE (Fig. 1). The six copper-binding sites essential for the function in O$_2$ binding are not conserved in the middle region, and misses 450 bp of the 5' and 600 bp of the 3' coding regions. Within the overlapping regions, the \textit{P. marginata} sequences are 65.8–87.6% identical on the DNA level and 54.0–83.1% identical on the amino acid level (Table 1).

The deduced primary structure of PmaHex1 yields a polypeptide of 702 amino acids with a calculated molecular mass of 82.52 kDa. Computer analysis of the sequence revealed the presence of a typical 16 amino acid signal peptide necessary for transmembrane transport and export into the hemolymph (von Heijne, 1986). Therefore, the native secreted PmaHex1 subunit comprises 686 amino acids with a mass of 80.95 kDa (pI 5.68), which is in excellent agreement with the data obtained from SDS-PAGE (Fig. 1). The six copper-binding sites essential for the function in O$_2$ binding are not conserved in the middle region, and misses 450 bp of the 5' and 600 bp of the 3' coding regions. Within the overlapping regions, the \textit{P. marginata} sequences are 65.8–87.6% identical on the DNA level and 54.0–83.1% identical on the amino acid level (Table 1).

Table 1

<table>
<thead>
<tr>
<th></th>
<th>PmaHex1</th>
<th>PmaHex2</th>
<th>PmaHex3</th>
<th>PmaHex4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PmaHex1</td>
<td>100.0</td>
<td>73.9</td>
<td>73.9</td>
<td>73.9</td>
</tr>
<tr>
<td>PmaHex2</td>
<td>83.1</td>
<td>100.0</td>
<td>95.3</td>
<td>95.3</td>
</tr>
<tr>
<td>PmaHex3</td>
<td>60.4</td>
<td>65.8</td>
<td>100.0</td>
<td>95.3</td>
</tr>
<tr>
<td>PmaHex4</td>
<td>67.4</td>
<td>67.7</td>
<td>71.9</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Percent identities between hexamerins were calculated from nucleotide (above diagonal) and amino acid sequences (below).

### 3.3. Identification of hexamerins in other plecopteran species

Using various combinations of degenerate primers, we obtained partial sequences from four additional plecopteran species: \textit{Nemoura} sp. (Nemouridae), \textit{T. burksi} (Taeoniopterygidae), \textit{A. vivipara} (Capniidae), and \textit{D. samali} (Diamphipnidae). They represent two other major superfamilies of the Plecoptera. While \textit{P. marginata} belongs to the Perloidea, \textit{Nemoura} sp., \textit{T. burksi} and \textit{A. vivipara} are members of the Nemuroidea, and \textit{D. samali} belongs to the Eusthenioidea. Three distinct hexamerins were obtained from \textit{Nemoura} sp., and one for each \textit{T. burksi}, \textit{A. vivipara}, and \textit{D. samali} (Supplemental Fig 2). The cDNA sequences measured 841–1931 bp and translate into polypeptide fragments from 279 to 567 amino acids. Comparisons of the amino acid sequences showed a high degree of identity. The closest relatives among hexamerins from different species were the proteins from \textit{T. burksi} (TbuHex1) and \textit{A. vivipara} (AviHex1), which display 96.2% identity on the nucleotide and 95.3% on the amino acid level. Within species, hexamers 1 and 3 from \textit{Nemoura} sp. (NspHex1 and 3) were found closely related, with 97.4% identity on the nucleotide and 97.5% on the amino acid level.

### 3.4. Phylogeny of insect hexamerins

The hexamerin sequences of the Plecoptera were included in an alignment of insect hemocyanins and selected other hexamerins (Supplemental Table 1). We omitted hexamerin sequences carrying long insertions or for which only partial database entries were available. We further excluded several sequences of the dipteran LSP-1 and LSP-2 types, for which multiple similar entries from closely related species were available. The final amino acid alignment includes four crustacean hemocyanins, five insect hemocyanins, 10 plecopteran and 67 other hexamerins. Phylogenetic trees were constructed employing Bayesian algorithm. Hexamerin 2 (RflHex2; accession no. \textit{AY572859}) from the termite \textit{Reticulitermes flavipes} (Schaf et al., 2005) was found to be the closest relative of PmaHex1 (Fig. 2). Both hexamerins were clearly distinct from crustacean or insect hemocyanins, as exemplified by the absence of the copper-binding histidines.
The basal position of the PmaHex1 was recovered, but this topology does not reach the significance levels (65% bootstrap support in NJ analysis; Fig. 4). The other hexamerins display an essentially identical tree topology as those published before (Burmester, 1999, 2001): the hexamerins from Diptera (LSP1 and LSP2) were monophyletic, as well as the methionine-rich and aromatic hexamerins (arylphorins) from the Lepidoptera. The position of the lepidopteran riboflavin (Rb)-binding hexamerins was uncertain: they were found either in sistergroup position to the other hexamerins from Holometabola (NJ; Fig. 4) or form a common clade with the other lepidopteran and the dipteran hexamerins (Bayesian analyses; data not shown). Coleopteran and hymenopteran hexamerins form a reasonably well-supported monophylum, which is basal to the mecopterian hexamerins. The hemipteran hexamerins were sister to all hexamerins from Holometabola. The recently identified hexamerins of the termites (Isoptera) (Scharf et al., 2005) join the cockroach (Blattaria) hexamerins, but were not monophyletic.

The basal position of the PmaHex1 was recovered, but this topology does not reach the significance levels (65% bootstrap support in NJ analysis; Fig. 4). The other hexamerins display an essentially identical tree topology as those published before (Burmester, 1999, 2001): the hexamerins from Diptera (LSP1 and LSP2) were monophyletic, as well as the methionine-rich and aromatic hexamerins (arylphorins) from the Lepidoptera. The position of the lepidopteran riboflavin (Rb)-binding hexamerins was uncertain: they were found either in sistergroup position to the other hexamerins from Holometabola (NJ; Fig. 4) or form a common clade with the other lepidopteran and the dipteran hexamerins (Bayesian analyses; data not shown). Coleopteran and hymenopteran hexamerins form a reasonably well-supported monophylum, which is basal to the mecopterian hexamerins. The hemipteran hexamerins were sister to all hexamerins from Holometabola. The recently identified hexamerins of the termites (Isoptera) (Scharf et al., 2005) join the cockroach (Blattaria) hexamerins, but were not monophyletic. R. flavipes hexamerin 1 (RflHex1) joins the cockroach (Blattaria) hexamerins, but were not monophyletic. R. flavipes hexamerin 2 (RflHex2) joins the Blaberus discoidalis hexamerin (BdiHex).

3.5. Analyses of amino acid compositions

The proportions of the aryl-groups (Phe and Tyr) and of Met were deduced from the cDNA sequences for each of the insect hemocyanins and hexamerins (Fig. 5). The copper-binding histidines are shaded in black, other strictly conserved residues are shaded in gray. The signal peptides are underlined, the secondary structure of PinHcA is given in the lower row.
is a moderately aromatic hexamerin with 6.8% phenylalanine and 8.8% tyrosine (sum = 15.7%); the King and Jukes average (King and Jukes, 1969) is 4.0% and 3.3%, respectively (sum = 7.3%). As far as sequences are known, the other plecopteran hexamerins display similar amino acid compositions (mean = 16.5% aromatic amino acids).

The average of content in aryl groups of the insect hemocyanins is ~10.8%, which is significantly lower than that of the plecopteran hexamerins (or PmaHex1 alone) and other arylphorin-like hexamerins of the insects (P < 0.001; ANOVA). The content of Met was not significantly different. Insect hemocyanins do not significantly differ in their proportion of aromatic amino acids from the crustacean hemocyanins (data not shown).

Most other hexamerins have similar high contents in aromatic amino acids, with three notable exceptions: the orthopteran hexamerins and lepidopteran methionine-rich hexamerins have ~10% Phe and Tyr, a content that is not significantly different from that of hemocyanins. Some hexamerins are also enriched in methionine: while the King and Jukes average of Met is 1.8%, a Met-content of up to ~11% was observed (BmoSSP1). The plecopteran hexamerins have slightly more Met compared to the Jukes–Cantor average, but were not significantly different from the insect hemocyanins (P = 0.17; ANOVA).

3.6. Molecular clock estimates

The phylogenetic tree of insect hexamerins and insect plus crustacean hemocyanins was linearized assuming a molecular clock. Divergence dates known from the fossil record (Kukalova-Peck, 1991; Ross and Jarzembowski, 1993; Benton and Donoghue, 2007) were employed to calibrate the tree using a Monte Carlo approach as implemented in the program r8s (Sanderson, 1997, 2002). Applying a relaxed clock with five distinct rates, the time of divergence of Hexapoda and malacostracan Crustacea was estimated to have occurred ~490 MYA in the late Cambrian or early Ordovician period (Fig. 6). Within the insects, hexamerins evolved from the hemocyanins ~468 MYA. The Plecoptera and other winged insects split ~360 MYA in the Carboniferous period. The hexamerins of the Eumetabola (Hemiptera and Holometabola) diverged ~340 MYA, followed by the split of mecopteran and hymenopteran + coleopteran hexamerins ~317 MYA. The hexamerins from Blattaria and Isoptera diverged between 180 and 217 MYA, Hymenoptera vs. Coleoptera 297 MYA, Lepidoptera vs. Diptera ~307 MYA, and Brachycera vs. Nematocera ~251 MYA. Dipteran LSP1 and LSP2 hexamerins diverged ~271 MYA, lepidopteran Met-rich hexamerins and arylphorins separated ~291 MYA. The plecopteran hexamerins commenced to diversify into distinct subunit types around ~253 MYA (data not shown).

4. Discussion

4.1. Early evolution of hexamerins within insects

Most previous studies that identify hexamerins in insects have focused on the evolutionarily more derived taxa (Telfer and Kunkel, 1991; Burmester, 1999). In fact, there are only a handful of studies on hexamerins from “lower” insects (e.g., Braun and Wyatt, 1996; Jamroz et al., 1996; Wu et al., 1996; Zhou et al., 2006). Here we present—to the best of our knowledge—the first data on hexamerins from the stoneflies, a taxon that is thought to have arose near the base of the clade of winged insects. Like other hexamerins (Telfer and Kunkel, 1991; Burmester, 1999), they commenced to accumulate in late larval (nymphal) stages (Fig. 1), and were also present in adults. The copper-binding sites typical for hemocyanins are absent (Fig. 2). It is therefore reasonable to assume that the hexamerins in Plecoptera have functions similar to those in other insects, and most likely store amino acids and energy for non-feeding periods (Telfer and Kunkel, 1991; Burmester, 1999).

Hexamerins emerged from hemocyanins probably early in insect evolution. The molecular clock calculations suggest that this event took place around 468 MYA (Fig. 6). Therefore, hexamerins evolved already in the insect stemline and it can be expected that distinct hemocyanins and hexamerins may be found in other “lower” Insecta and the endognathan Hexapoda (Collembola, Protura, Diplura). Stoneflies are the first known insects that harbor hemocyanin along with the hexamerins in their hemolymph. This fact clearly hints to distinct functions of these proteins.

The clade leading to the plecopteran hexamerins diverged from that of other insects around 360 MYA (Fig. 6), while the diversification of different plecopteran hexamerins types commenced around 253 MYA (not shown). This timing and
the paraphyly noted above (Fig. 3) suggests that the separation of (pre-)plecopteran hexamerins into distinct types preceded the emergence of the plecopteran families. The diversification of hexamerins by gene duplication, resulting in multiple paralogs within distinct species, is a common phenomenon within the hexamerins (Burmester, 1999). In some cases, this phenomenon may simply reflect the need of additional gene copies for mass expression, in other cases there may be distinct roles for different hexamerins.

4.2. Early accumulation of aromatic amino acids in hexamerins

In response to particular needs during postlarval development, many hexamerins are enriched in either aromatic amino acids (Phe + Tyr), Met, Glx or Asx (Telfer and Kunkel, 1991). While aromatic and Met-rich hexamerins are widespread in different insect orders, Glx- and Asx-rich hexamerins are apparently restricted to Hymenoptera (Wheeler and Buck, 1995; Martinez et al., 2000). To refer to its particularly high content of aromatic amino acids, the term “arylphorin” has been introduced for a hexamerin of *Hyalophora cecropia* (Lepidoptera) (Telfer et al., 1983). This designation has been adapted to hexamerins of other insect orders with a similar amino acid composition (e.g. Scheller et al., 1990; de Kort and Koopmanschap, 1994; Jamroz et al., 1996). However, the “arylphorins” of different insect orders are not particularly related (Fig. 4; Burmester, 1999), i.e., arylphorins may well be more closely related to hexamerins of their own order.
with lower content in aromatic amino acids than to arylphorins of other orders.

To address the question whether aromatic hexamerins evolved multiple times independently or whether this is a plesiomorphic character of the hexamerins, we mapped the relative content in Phe+Tyr and Met on the phylogenetic tree (Fig. 5). The plecopteran hexamerins were found to have high content in aromatic amino acids. Thus, the accumulation of aromatic amino acids, probably reflecting the need of building blocks for the formation of the cuticle (Scheller et al., 1990; Burmester, 1999), appears to have occurred in a very early stage of the evolution of hexamerins. Therefore, the enrichment in aromatic amino acids should be considered as plesiomorphic character of these proteins.

Hexamerins with lower content in Phe+Tyr most likely reflect other requirements, e.g., in the higher extant Lepidoptera, two distinct types of Met-rich hexamerins (averages 4.2% and 7.5% Met) have been identified (Ryan et al., 1985; Tojo and Yoshiga, 1994). Met-rich hexamerins are more abundant in the female than in the male, and it has been assumed that they support female reproduction and egg development by enhancing the pool of sulphur-containing amino acids at the time of vitellogenesis (Pan and Telfer, 1996). Likewise, the juvenile-hormone binding hexamerins of Orthoptera (cf. Braun and Wyatt, 1996) have a comparatively low content of Phe+Tyr (10.2%), but are rich in Met (4.5%). The phylogenetic tree implies that these proteins have evolved independently from highly aromatic hexamerins. Thus, the losses of aromatic amino acids most likely reflect secondary evolutionary events (Fig. 5).

4.3. Implications for insect phylogeny

Hexamerins have been successfully applied to estimate evolutionary patterns and divergence times of insects (Burmester et al., 1998; Burmester, 2001). Assuming a relaxed molecular clock with multiple substitution rates and fossil constraints, we calculated that Hexapoda and malacostracan Crustacea diverged around 490 MYA. This is in good agreement with previous estimates using
hemocyanins and hexamerins, and supports the hypothesis that Hexapoda are actually closely related to Crustacea (Burmester, 2001). Notably, the molecular clock estimates show that the evolution and radiation of the winged insects (Pterygota) commenced around 360 MYA and mainly occurred in the Carboniferous period until 300 MYA (Fig. 6). These calculations are in excellent agreement with the available fossil record (Hennig, 1969; Kukalova-Peck, 1991; Ross and Jarzembowski, 1993) and previous molecular clock estimates (Gaunt and Miles, 2002).

The position of the Plecoptera within the Neoptera is still uncertain. While Hennig (1969, 1981) considered the Plecoptera as a basal taxon within the Neoptera, Boudreaux (1979) placed the stoneflies as sistergroup of the embiids (web-spinners). On the basis of 18S rRNA sequences, Flook and Rowell (1998) suggested that the Plecoptera form a common clade with the Grylloblattodea and Dermaptera. Employing “total evidence” analyses of molecular and morphological data, Wheeler et al. (2001) concluded that the Plecoptera were sistergroup to the Embiida, but were not able to resolve the exact affinity of that clade. The poor phylogenetic signal is the most likely reason for the uncertain placement of the Plecoptera which was also observed with hexamerins. NJ analyses placed plecopteran hexamerins at the base of the neopteran tree, thus corroborating the view of Hennig (1969, 1981). However, bootstrap analyses show only weak support in NJ trees and this topology was not supported in the Bayesian trees, which left the basal Neoptera (“Polyneoptera”) largely unresolved.

The lack of clear morphological or molecular signals that could resolve plecopteran affinities and relationships among lower Neoptera can most likely be explained by the rapid diversification of most modern orders of winged insect in the Carboniferous period 300–360 MYA (Hennig, 1969; Kukalova-Peck, 1991).

Another interesting finding of the present study is the close relationship of cockroach and termite hexamerins. The monophyly of cockroaches is disputed, and some authors have concluded that the Blattaria are paraphyletic with respect to the Isoptera (e.g., Hennig 1981; Lo et al., 2000). Lo et al. (2000) presented molecular evidence that Isoptera may derive from subsocial, woodfeeding cockroaches. In fact, the known blattarian and isopteran hexamerins form a well-supported common clade. Moreover, they are mutually paraphyletic, showing that the hexamerins diversified into at least two classes before the Blattaria and Isoptera split. Molecular clock estimates suggest that the clades leading to the hexamerins of B. germanica and P. americana on one hand, and of R. flavipes on the other hand, diverged about 200 MYA. Although fossil data of termites are poor, there is in fact evidence for the presence of Isoptera in lower Cretaceous deposits (Ross and Jarzembowski, 1993), confirming the molecular clock estimates.

Thus, our molecular clock estimates are in excellent agreement with the fossil record and show the usefulness of hexamerins to infer the times of insect diversification. However, it should also be pointed out that (1) the
orthology of various hexamerins is not always thoroughly established and that (2) molecular clock estimates usually have large standard errors. Therefore, all our calculations should only be considered as rough estimates rather than as exact dates.

Acknowledgments

We thank Klaus Enting and Rainer Rupprecht (Mainz) for P. marginata specimens, and Falko Roeding and Robert Schöpflin for the help with the implementation of the r8s program. This work has been supported by grants of the Deutsche Forschungsgemeinschaft (Bu956/5; Bu956/9) and by NSF grant IBN-9722196. The nucleotide sequences reported in this paper have been submitted to the EMBL/GenBank™ Databases under the accession nos. AM690365 to AM690371, EF620538, EF617597 and EF617598.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ibmb.2007.06.001.

References


