Diplopod Hemocyanin Sequence and the Phylogenetic Position of the Myriapoda

Kristina Kusche and Thorsten Burmester
Institute of Zoology, University of Mainz, Mainz, Germany

Hemocyanins are copper-containing respiratory proteins of the Arthropoda that have so far been thoroughly investigated only in the Chelicerata and the Crustacea but have remained unstudied until now in the Myriapoda. Here we report the first sequence of a myriapod hemocyanin. The hemocyanin of Spirostreptus sp. (Diplopoda: Spirostreptidae) is composed of two distinct subunits that are arranged in a 6 × 6 native molecule. The cloned hemocyanin subunit cDNA codes for a polypeptide of 653 amino acids (75.5 kDa) that includes a signal peptide of 18 amino acids. The sequence closely resembles that of the chelicerate hemocyanins. Molecular phylogenetic analyses reject with high statistical confidence the integrity of the Tracheata (i.e., Myriapoda + Insecta) but give conflicting results on the position of the myriapod hemocyanin. While distance matrix and maximum-likelihood methods support a basal position of the Spirostreptus hemocyanin with respect to the other hemocyanins, parsimony analysis suggests a sister group relationship with the chelicerate hemocyanins. The latter topology is also supported by a unique shared deletion of an alpha-helix. A common ancestry of Myriapoda and Chelicerata should be seriously considered.

Introduction

The supply of cells with oxygen is an essential process that is accomplished in many multicellular organisms with the help of specialized transport proteins. Among these, the copper-based hemocyanins of the Mollusca and Arthropoda are unusual because they associate to quaternary structures up to several million daltons (Markl and Decker 1992; van Holde and Miller 1995). The hemocyanins of these two phyla differ in sequence and structure and are likely of independent evolutionary origins (van Holde and Miller 1995; Burmester 2001). Arthropod hemocyanins are composed of hexamers of six similar or identical subunits of about 75 kDa that may form quaternary structures containing up to 8 × 6 subunits (Markl 1986; Markl and Decker 1992). Each subunit carries one O₂ bound to two copper ions, which are coordinated by six histidines of the polypeptide chain (Linzen et al. 1985; van Holde and Miller 1995). Hemocyanins belong to a large arthropod-specific protein superfamily that includes the arthropod phenoloxidases, the nonrespiratory pseudohemocyanins (cryptocyanins) of some decapod Crustacea, the insect storage hexamerins, and the dipteran hexamerin receptors (Beintema et al. 1994; Burmester and Scheller 1996; Burmester et al. 1998; Burmester 1999, 2001; Terwilliger, Dangott, and Ryan 1999).

While hemocyanins have been studied in detail in many chelicerate and crustacean species (e.g., Markl 1986; Markl and Decker 1992; van Holde and Miller 1995), specialized oxygen transport proteins have been poorly known in the Myriapoda, and their existence is still ignored by the textbooks. Respiratory proteins have been considered unnecessary in this taxon because Chilopoda (centipedes) and Diplopoda (millipedes) possess, similar to the insects, a typical tracheal system that was thought to be sufficient to supply the internal organs with an adequate amount of oxygen (Brusca and Brusca 1990; Hilken 1998). Nevertheless, a 36mer (6 × 6) oxygen-carrying protein had been described in the genus Scutigera (Chilopoda) that closely resembled the other arthropod hemocyanins (Mangum et al. 1985; Boisset, Taveau, and Lamy 1990; Gebauer and Markl 1999). The exceptional presence of a hemocyanin in this taxon was attributed to the high activity of the Scutigeramorpha and their peculiar blind-ending tracheal system (Mangum and Godette 1986). However, it is now evident that hemocyanins also occur in at least one family of the Diplopoda, i.e., the Spirostreptidae, suggesting that such oxygen-carrying proteins are much more widespread within the Myriapoda than previously thought (Jaenicke et al. 1999). Here we report the cDNA-cloning and analysis of a hemocyanin subunit from the diplopod Spirostreptus sp., which is the first myriapod hemocyanin sequence, and discuss its implications for hemocyanin evolution and arthropod phylogeny.

Materials and Methods

Protein Biochemistry

Spirostreptus sp. was obtained from the Aquazoo in Düsseldorf, Germany, reared on mulh/sand with 12 h/12 h light/dark at 25°C, and fed on potatoes and fruits. Hemolymph samples were withdrawn from the dorsal intersegmental regions with a syringe and stored frozen at −20°C until use. Hemocyanin was purified by size-exclusion chromatography (Jaenicke et al. 1999). Antibodies against purified Spirostreptus hemocyanin were raised in guinea pigs. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and Western blotting were carried out as described (Jaenicke et al. 1999). For determination of the N-terminal ends, samples were separated by SDS-PAGE and transferred to a PDVF membrane. The hemocyanin bands were excised and submitted to Edman degradation.
Cloning of Spirostreptus Hemocyanin

RNA was extracted from the complete animal excluding the cuticle. Poly(A)⁺ RNA was purified from total RNA using the PolyATtract kit (Promega). Five milligrams of poly(A)⁺ RNA were used for the construction of a directionally cloned cDNA expression library applying the Lambda ZAP-cDNA synthesis kit (Stratagene). The library was amplified once and screened with the anti-Spirostreptus hemocyanin antibodies. Positive phage clones were converted to plasmid vectors using the material provided in the cDNA synthesis kit. The hemocyanin cDNAs inserted in the pBK-CMV vector the material provided in the cDNA synthesis kit. The incomplete clones were extended by RT-PCR using a set of specific oligonucleotide primers and a degenerate primer designed according to the highly conserved amino acid sequence of the arthropod hemocyanin CuA site. The missing 5’ end was obtained by two successive 5’ rapid amplification of cDNA ends (RACE) assays (Gibco-BRL kit) with a series of nested oligonucleotide primers according to the manufacturer’s instructions. The sequences were obtained after the cloning of the PCR products into the pGEMTeasy vector (Promega).

Sequence Data Analysis and Phylogenetic Studies

The Genetics Computer Group (GCG) Sequence Analysis Software Package 8.0 and the tools provided by the ExPASy Molecular Biology Server of the Swiss Institute of Bioinformatics (http://www.expasy.ch) were used for sequence analyses. Multiple-sequence alignments were carried out with the aid of GeneDoc 2.6 (Nicholas and Nicholas, 1997). The Spirostreptus hemocyanin DNA and amino acid sequences were added to previously published alignments of the hemocyanin superfamily (Burmeister, 2001) (see Supplementary Material). This alignment was constructed using CLUSTALX (Thompson et al., 1997) and corrected considering the crystallographic structure of Limulus polyphemus (Hazes et al., 1993) and Panulirus interruptus (Gaykema et al., 1984). Long gap regions and highly divergent regions were deleted from the final data set. The nucleotide alignment follows that of the proteins. Because saturation of silent sites was assumed, only the first and second codon positions of the DNA sequences were used. The program packages PHYLIP 3.6 (Felsenstein, 2000) and TREE-PUZZLE 5.0 (Strimmer and von Haeseler, 1996) were applied for phylogenetic inference. Gamma-corrected distances were calculated using the PAM (Dayhoff, Schwartz, and Orcutt, 1978) and JTT (Jones, Taylor, and Thornton, 1992) models with eight rate categories. Nucleotide distances were calculated according to the HKY model (Hasegawa, Kischno, and Yano, 1985). Tree constructions were performed by the neighbor-joining and maximum-parsimony methods available in the PHYLIP package. The reliability of the trees was tested by nonparametric bootstrap analysis (Felsenstein, 1985) with 100 repetitions using PUZZLEBOOT (shell script by M. Holder and A. Roger). Maximum-likelihood and likelihood mapping analyses were carried out using a reduced data set of 38 and 32 sequences, respectively, containing a maximum of 10 representatives of each protein family. TREE-PUZZLE 5.0 was used to test alternative topologies using the Dayhoff (Dayhoff, Schwartz, and Orcutt, 1978), JTT, and VT (Müller and Vinrong, 2000) models of amino acid substitution and the HKY model for nucleotide replacement, each under the assumption of a rate heterogeneity with eight gamma categories (Strimmer and von Haeseler, 1996, 1997). Likelihood ratio tests were carried out according to Kishino and Hasegawa (1989). Z-values of >3.29 are significant with \( P < 0.001 \).

Results

Protein Purification and N-Terminal Sequencing

The 36meric hemocyanin of Spirostreptus has a high molecular mass of about 2,600 kDa and thus could easily be purified from the hemolymph by one-step size-exclusion chromatography (Jaenicke et al., 1999). After separation by SDS-PAGE, the hemocyanin fraction showed two distinct bands with apparent molecular weights of about 75 kDa each. These polypeptides (named Hc1 and Hc2) most likely represent different subunits of the Spirostreptus hemocyanin (fig. 1). Antibodies raised against this fraction recognized both Hc1 and Hc2 but no other protein of the hemolymph. Microsequencing of the amino acids of the N-termini demonstrated that the two putative hemocyanin subunits were distinct polypeptides. However, the sequence data were poor and not sufficient to allow the assignment of these proteins to the hemocyanin superfamily or any other protein class.

cDNA Cloning and Sequencing of Spirostreptus Hemocyanin

Because the site of hemocyanin synthesis in the myriapods is unknown, a cDNA library of Spirostreptus was constructed from poly(A)⁺ RNA of the complete animal without the cuticle. This is the first myriapod
The missing alpha-helix 1.2 in the hemocyanins of Spirostreptus sp. shows. The amino acid sequence of the Spirostreptus hemocyanin is derived from the crystallographic analysis of Limulus polyphemus hemocyanin subunit II (Hazes et al. 1993) as shown. The missing alpha-helix 1.2 in the hemocyanins of Spirostreptus and the Chelicerata is in brackets. The nomenclature follows the standard convention for hemocyanin structure (Linzen et al. 1985); α = α-helix; β = beta-sheet. The signal peptides are underlined, and the potential N-glycosylation sites in the Spirostreptus hemocyanin are in boldface. Abbreviations: SpiHc1, Spirostreptus sp. hemocyanin subunit 1 (accession number AJ297738); LpHc1, L. polyphemus hemocyanin subunit II (P04253); PinHcA, Panulirus interruptus hemocyanin subunit a (P04254); SanHc, Schistocerca americana hemocyanin (AF038569).

cDNA library available. The library was screened with the anti-hemocyanin antibodies, and 37 positive clones were identified. Although sequencing reveals that they represent at least three independent clone types that only differ in the lengths of the 5′ ends by a few base pairs, all contain cDNA of only about 600 bp from the 3′ end of the putative hemocyanin subunit. Extensive screening did not reveal any longer hemocyanin cDNA in the library. The hemocyanin cDNA was extended by RT-PCR with the aid of a degenerated primer corresponding to the conserved arthropod hemocyanin CuA site (fig. 2) and eventually by 5′ RACE. The complete Spirostreptus cDNA sequence encompasses 2,082 bp (plus a polyA tail of 18 bp) that includes an open reading frame of 1,959 bp beginning with a methionine ATG at bp 66. The presence of three purines upstream of the putative initiator ATG fulfills the minimum criteria for a eukaryotic translation start site (Kozak 1984). The polyadenylation signal AATAAA is located 21 bp upstream of the polyA tail. Conceptual translation of the open reading frame results in a protein of 653 amino acids with a deduced molecular mass of 75.5 kDa. The N-terminal amino acid sequences allow the assignment of this protein to the Hc1 polypeptide (fig. 1) and show the presence of a typical signal peptide of 17 amino acids (fig. 2), as also predicted by the computer programs. Thus, the native secreted hemocyanin subunit comprises 636 amino acids, with a calculated molecular mass of 73.7 kDa, which is in good agreement with that estimated by SDS-PAGE (fig. 1). Three potential N-glycosylation sites (NXT/S) are present in the sequence (fig. 2), but none of these sites are conserved in the other arthropod hemocyanins, and their significance remains uncertain. The six copper-coordinating histidines, which are strictly conserved in all arthropod hemocyanins (Linzen et al. 1985; Burmester 2001), are present in the myriapod protein as well.

Sequence Comparison and Hemocyanin Phylogeny

The amino acid sequence of the Spirostreptus hemocyanin was added to a previously published alignment of 93 other hemocyanins, phenoloxidases, pseudohemocyanins, and hexamerins (Burmester 2001). Pairwise comparison showed that Spirostreptus hemocyanin displayed the highest degree of sequence similarity with the chelicerate hemocyanins (37.5%–42.1% identity) and the phenoloxidases of Crustacea and insects (33.0%–39.0%), while lower scores were observed with the crustacean hemocyanins (≤32.8%), pseudohemocyanins (≤29.9%), and the insect hemocyanin (36.1%) and hexamerins (≤29.1%). Like in the other hemocyanins, most sequence variations are present in the first and third domains, while the second domain, which forms the core of the hemocyanin subunit and includes the copper-binding sites, is strikingly conserved between the Spirostreptus and other hemocyanins. It is noteworthy that in Spirostreptus hemocyanin, the second alpha-helix of the first domain (1.2) is missing, a feature that is otherwise shared only by the chelicerate hemocyanins (fig. 2).

Phylogenetic analyses were carried out with maximum-parsimony, distance matrix, and maximum-likelihood methods using both amino acid and nucleotide sequence alignments. Highly diverged regions were excluded from the calculation. The arthropod phenoloxidases most likely constitute the most ancient branch of the arthropod hemocyanin superfamily and thus were considered as the outgroup (Burmester and Scheller 1996; Sánchez et al. 1998; Burmester 2001). Except for the position of the Spirostreptus hemocyanin, the general topologies of the resulting trees are identical (fig. 3). While the parsimony methods consistently support a sister group position of the myriapod and the chelicerate hemocyanins with high bootstrap values (99% and 90%, respectively; table 1), distance matrix methods consider, with lower support values (65% and 78%; table 1), the myriapod hemocyanin as basal to all members of the hemocyanin superfamily except the phenoloxidases.

A maximum-likelihood approach was used to estimate the quality of the possible relationships of the myriapod hemocyanin (table 1). While the clustering of the Spirostreptus hemocyanin and the insect (or crustacean) proteins can be rejected with high statistical confidence (P < 0.001), this method does not resolve the
relative positions of the chelicerate, myriapod, and crustacean + insect clades. Likelihood mapping (Strimmer and von Haeseler 1997) again strongly supports the close relationship of the crustacean and insect proteins in 100% and 98.8%, respectively, of all quartets (fig. 4A). However, the position of the myriapod hemocyanin remains essentially unresolved (fig. 4B).

Table 1
Support for Five different Topologies of Arthropod Relationships According to the Hemocyanin Superfamily

<table>
<thead>
<tr>
<th>Topology</th>
<th>$\Delta \log L$ (PAM)</th>
<th>$\Delta \log L$ (JTT)</th>
<th>$\Delta \log L$ (VT)</th>
<th>$\Delta \log L$ (HKY)</th>
<th>MP (aa)</th>
<th>NJ (aa)</th>
<th>MP (DNA)</th>
<th>NJ (DNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (My, (Ch, (Cr, In))) . . . . . .</td>
<td>$-32,970.05$</td>
<td>$-0.06$</td>
<td>$-32,893.05$</td>
<td>$-0.46$</td>
<td>0</td>
<td>78</td>
<td>3</td>
<td>65</td>
</tr>
<tr>
<td>2. ((Ch, My), (Cr, In)) . . . . . .</td>
<td>$-0.48$</td>
<td>$-32,848.88$</td>
<td>$-0.38$</td>
<td>$-29,974.36$</td>
<td>99</td>
<td>9</td>
<td>90</td>
<td>32</td>
</tr>
<tr>
<td>3. (Ch, (My, (Cr, In))) . . . . . .</td>
<td>$-0.18$</td>
<td>$\pm 0$</td>
<td>$-0.40$</td>
<td>$-0.35$</td>
<td>1</td>
<td>13</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>4. (Ch, (Cr, (My, In))) . . . . . .</td>
<td>$-70.55^*$</td>
<td>$-72.39^*$</td>
<td>$-73.94^*$</td>
<td>$-50.58^*$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5. (Ch, (In, (My, Cr))) . . . . . .</td>
<td>$-68.69^*$</td>
<td>$-70.97^*$</td>
<td>$-72.85^*$</td>
<td>$-49.91^*$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note.—My = Myriapoda; Ch = Chelicerata; Cr = Crustacea; In = Insecta. Maximum-likelihood amino acid substitutions models: PAM, Dayhoff’s PAM matrix (Dayhoff, Schwartz and Orcutt 1978); JTT (Jones, Taylor, and Thornton 1992); VT (Müller and Vingron 2000). Maximum-likelihood nucleotide substitution model: HKY (Hasegawa, Kishino, and Yano 1985). MP = maximum parsimony; NJ = neighbor joining using the PAM model.

* Significant at the $P < 0.001$ level (likelihood ratio test; Kishino and Hasegawa 1989).
the Myriapoda, the occurrence of a hemocyanin in the Scutigeramorpha was attributed to the peculiar tracheal system and postulated high oxygen demand of this taxon (Mangum and Godette 1986). However, we demonstrated that hemocyanins are also present in the Spirostreptus, a family of the diplopod Myriapoda (Jaenicke et al. 1999). Although these species have normal trachea, a large amount of hemocyanin was observed in the hemolymph of *Spirostreptus* sp. and two related species. The present data imply that, in contrast to previous assumptions, hemocyanins may be widespread within the Myriapoda, although information from the Symphyta and Pauropoda is still missing. Thus, contrary to the insects, the evolution of a tracheal system in the Myriapoda was not accompanied by the loss of respiratory pigments in the hemolymph.

**Spirostreptus Hemocyanin**

The Spirostreptus hemocyanin sequence contains all amino acids that are required for the function of a true oxygen carrier, particularly the six conserved copper-binding histidines. It is clearly distinct in sequence, structure, and biochemical properties from the phenoloxidases and the various nonrespiratory hemocyanin-like proteins. While the latter proteins are most likely absent in the Myriapoda, the phenoloxidase of Spirostreptus is a distinct protein with different antigenic and physiological characteristics (unpublished data). The Spirostreptus hemocyanin includes a typical N-terminal signal peptide necessary for the intracellular transfer into the endoplasmic reticulum and subsequent release into the hemolymph. While the chelicerate (arachnid) hemocyanins do not contain any signal sequences (Voit et al. 2000) and are released from cyanocytes by cell rupture (Markl and Decker 1992), the crustacean and insect hemocyanins have such signal peptides and are synthesized by the hepatopancreas or other organs (Sellos, Lemoine, and Van Wormhoudt 1997; Sánchez et al. 1998; Kusche and Burmester 2001). Thus, the mode of hemocyanin synthesis in the Myriapoda resembles that of the Crustacea and insects.

**Hemocyanin Evolution**

Hemocyanins are present in all euarthropod subphyla, including the Myriapoda and at least one insect. While the role of the insect protein is not well understood (Sánchez et al. 1998), the myriapod hemocyanins closely resemble their crustacean and chelicerate cognates in physiological and structural properties. The phylogenetic analyses including the Spirostreptus sequence show a monophyletic origin of all arthropod hemocyanins (fig. 3), most likely from an enzymatic phenoloxidase-like ancestor. Thus, the hemocyanins must have evolved in the arthropod stemline before the radiation of the major extant subphyla more than half a billion years ago. Their emergence might be correlated with the increase in body size of the animal and the formation of a hard cuticle. Both events were crucial in the evolution of the Arthropoda but made simple diffusion inefficient to supply the internal organs with sufficient oxygen. The hemocyanins fulfilled the upcoming need for an efficient oxygen carrier (cf. Burmester 2001).

The basic structure of a hexamer is conserved among all hemocyanins; therefore, a $1 \times 6$ molecule was the most likely design of the last common ancestor of the arthropod hemocyanins. However, the number and arrangement of the hexamers essentially differ among the crustacean, chelicerate, and myriapod hemocyanins (Markl and Decker 1992; van Holde and Miller 1995), and the formation of these multimers clearly occurred independently in these subphyla. This multimerization is most likely correlated with an enhanced oxygen transport capacity with less osmotic impact. While among the Crustacea and Chelicerata the structures of these multimers vary, ranging from $1 \times 6$ to $8 \times 6$ subunits, the $36$ mer hemocyanin appears to be unique to the myriapod hemocyanins. This structure was apparently conserved since the separation of the Diplopoda and Chilopoda, at least 400 MYA if not much earlier (Robinson 1990; Friedrich and Tautz 1995; Shear 1997). The unique hemocyanin structure can be considered an additional synapomorphy that favors a monophyly of the Myriapoda.

**Implications for Arthropod Evolution**

One of the most intensively debated issues in animal systematics focuses on the relative relationships within the arthropod phylum, i.e., how the Chelicerata, Crustacea, Myriapoda, and Hexapoda are related (e.g., Fortey and Thomas 1997; Giribet and Ribera 2000) (fig. 5). In the textbooks, the Myriapoda have long been combined with the insects in a taxon named “Tracheata,” “Antennata,” or “Atelocerata” (e.g., Snodgrass 1938; Brusca and Brusca 1990; fig. 5A). In recent years, evidence from both molecular and comparative developmental studies has suggested that the “Tracheata” do not exist, but rather that the Hexapoda are allied with...
the Crustacea (Turberville et al. 1991; Averof and Akam 1995; Friedrich and Tautz 1995; Dohle 1997; Boore, Lavrov, and Brown 1998; Giribet and Ribera 2000; Shultz and Regier 2000) or are even nested within a “pancrustacean” taxon (e.g., Zrzavý, Hypša, and Vlásková 1997; García-Machado et al. 1999; Wilson et al. 2000; Burmester 2001). However, in none of the molecular studies could the relationship of the Myriapoda to the other arthropods be resolved with sufficient confidence. Phylogenetic analyses of hemocyanins and related proteins had demonstrated a remarkably good resolution of the arthropod trees, at least at a higher taxonomic level (Beintema et al. 1994; Burmester and Scheller 1996; Burmester et al. 1998; Burmester 2001). Thus, we hoped to resolve the relative position of the Myriapoda with the help of the Spirostreptus hemocyanin. In fact, all phylogenetic approaches strongly reject the integrity of the Tracheata (Fig. 4A and table 1). The significance levels are higher than those obtained from ribosomal DNA (Turberville et al. 1991; Friedrich and Tautz 1995; Giribet and Ribera 2000) or other molecular phylogenetic study (Regier and Shultz 1997; Shultz and Regier 2000). This demonstrates the general usefulness of the hemocyanin superfamly in the reconstruction of arthropod phylogeny (Burmester 2001). Nevertheless, the available methods still give conflicting results (Figs. 3 and 4 and table 1). Distance matrix–based methods support a basal position of the myriapod hemocyanin with respect to all other arthropod hemocyanins (Figs. 3A and 5B) and agree with the studies using arthropod mitochondrial DNA (Ballard et al. 1992). Most studies using 18S and 28S ribosomal DNA tend to ally the Myriapoda with the Chelicerata (Turberville et al. 1991; Friedrich and Tautz 1995; Giribet et al. 1996; Giribet and Ribera 2000) and mitochondrial protein sequences (M. Friedrich, personal communication), a common origin of the Chelicerata and the Myriapoda should be seriously considered, although there are no striking morphological synapomorphies of these taxa.

Supplementary Material
The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL databases with the accession number AJ297738. Alignments of the protein and cDNA sequences are available at http://www.molbiolevo.org.

Acknowledgments
We wish to thank S. Löser for supplying the animals, W. Xylander for his advice, H. Enghoff for the determination of the species, E. Jaenicke for his help with protein purification, S. Pöttsch for protein sequencing, H. Decker for many discussions, J. Markl for the excellent working facilities and his generous support, and Ken van Holde for critical reading of the manuscript. This work is supported by the Deutsche Forschungsgemeinschaft (Bu956/3).

Literature Cited


RICHARD H. THOMAS, reviewing editor

Accepted April 23, 2001