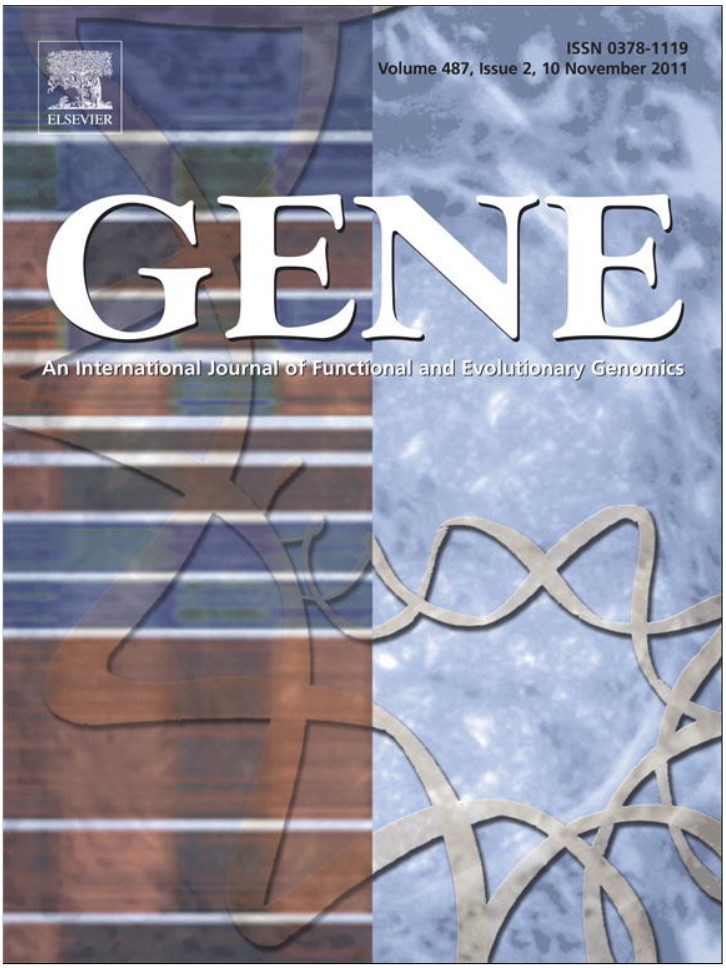


Gene 487 (2011) 118–128

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

The cDNA sequence of three hemocyanin subunits from the garden snail *Helix lucorum*

Lina De Smet [a,1], Ivan Dimitrov [b,1], Griet Debyser [a], Pavlina Dolashka-Angelova [b], Aleksandar Dolashki [b], Jozef Van Beeumen [a], Bart Devreese [a,*]

[a] Laboratory of Protein Biochemistry and Biomolecular Engineering, Department of Biochemistry and Physiology, Ghent University, Ledeganckstraat 35, 9000 Gent, Belgium
[b] Institute of Organic Chemistry, Bulgarian Academy of Sciences, Acad. G. Bonchev str bl.9, Sofia 1113, Bulgaria

ARTICLE INFO

Article history:
Accepted 31 July 2011
Available online 7 August 2011

Received by Richard Cordaux

Keywords:
Mollusks
Evolution
Mass spectrometry

ABSTRACT

Hemocyanins are blue copper containing respiratory proteins residing in the hemolymph of many molluscs and arthropods. They can have different molecular masses and quaternary structures. Moreover, several molluscan hemocyanins are isolated with one, two or three isoforms occurring as decameric, didecameric, multidecameric or tubule aggregates. We could recently isolate three different hemocyanin isopolypeptides from the hemolymph of the garden snail *Helix lucorum* (HlH). These three structural subunits were named α_D-HlH, α_N-HlH and β-HlH. We have cloned and sequenced their cDNA which is the first result ever reported for three isoforms of a molluscan hemocyanin. Whereas the complete gene sequence of α_D-HlH and β-HlH was obtained, including the 5′ and 3′ UTR, 180 bp of the 5′ end and around 900 bp at the 3′ end are missing for the third subunit. The subunits α_D-HlH and β-HlH comprise a signal sequence of 19 amino acids plus a polypeptide of 3409 and 3414 amino acids, respectively. We could determine 3031 residues of the α_N-HLH subunit. Sequence comparison with other molluscan hemocyanins shows that α_D-HlH is more related to *Aplysia californicum* hemocyanin than to each of its own isopolypeptides. The structural subunits comprise 8 different functional units (FUs: a, b, c, d, e, f, g, h) and each functional unit possesses a highly conserved copper-A and copper-B site for reversible oxygen binding. Potential N-glycosylation sites are present in all three structural subunits. We confirmed that all three different isoforms are effectively produced and secreted in the hemolymph of *H. lucorum* by analyzing a tryptic digest of the purified native hemocyanin by MALDI-TOF and LC–FTICR mass spectrometry.

1. Introduction

Three major classes of respiratory proteins exist within the animal kingdom: hemoglobins, hemerythrins, and hemocyanins. Iron atoms are present in the active sites of hemoglobins and hemerythrins, whereas copper atoms are characteristic for hemocyanins. Arthropodan hemocyanins belong to the hemocyanin/hexamerin/phenoloxidase superfamily, whereas mollusc hemocyanins probably evolved independently from a common ancestral non-hemocyanin protein (Burmester, 2001; Decker and Tuczek, 2000; Martin et al., 2007; van Holde et al., 2001).

The structural subunit of molluscan hemocyanin has a molecular mass of around 350–400 kDa and is arranged as a linear sequence of seven or eight different globular functional units (FU) (De Ioannes et al., 2004; Dolashka-Angelova et al., 2007; Harris et al., 2004; Harris and Markl, 2000). The mass of the functional units is approximately 50 kDa, each carrying an oxygen binding site that contains two copper ions complexed by six strictly conserved histidine residues. The so-called Cu_A and Cu_B sites of each FU together reversibly bind one dioxygen molecule, leading to a maximum of up to eight O_2 molecules per structural subunit. The eight functional units are structurally distinct and have been termed FUa – FUh, starting from the N-terminal part. The FUs are connected via short linker regions of 10 to 15 amino acids (Gatsogiannis and Markl, 2009). The number of structural subunits present in a species varies between one and three (Bergmann et al., 2006; Lieb et al., 2001; Velkova et al., 2010). In contrast to what is the case for the marine gastropods of the genera *Megathura*, *Haliotis*, *Aplysia*, *Nucula*, *Concholepas*, and *Rapana* where there are one or two structurally and functionally distinct hemocyanin isoforms, three different isoforms (β-HlH, α_D-HlH and, α_N-HlH) were isolated from the hemolymph of *Helix lucorum* (Velkova et al., 2010). Three

Abbreviations: α_D-HlH, α_N-HlH, alpha structural subunits of *Helix lucorum* hemocyanin; β-HlH, beta structural subunit of *Helix lucorum* hemocyanin; ACN, acetonitrile; ESI, electrospray ionization; FT-ICR, Fourier transform ion cyclotron resonance; FU, functional unit of hemocyanin; HPLC, high performance liquid chromatography; LC, liquid chromatography; MALDI, matrix assisted laser desorption/ionization; MS/MS, tandem mass spectrometry; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcriptase polymerase chain reaction; TFA, trifluoro acetic acid; UTR, untranslated region.

* Corresponding author at: Ghent University, L-PROBE, K.L. Ledeganckstraat 35, 9000 Gent, Belgium. Tel.: +32 9 264 5273; fax: +32 9 264 5338.
E-mail address: bart.devreese@ugent.be (B. Devreese).

[1] These authors contributed equally to the work.

0378-1119/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2011.07.030

hemocyanin isoforms were also reported to occur in *Helix pomatia* (Lambert et al., 1995; Lontie, 1983; Wood et al., 1985). The three isoforms of *H. lucorum* hemocyanin form didecamers of homogeneous decamers, as was likewise observed for *Megathura crenulata*, *Rapana venosa* and *Helix tuberculata* hemocyanin (Altenhein et al., 2002; Dolashka-Angelova et al., 2003; Harris et al., 2004; Velkova et al., 2010). However, after dissociation and reassociation, these proteins have a different behavior, as they reassociate into didecamers and tubules of different length: long tubules are observed in β-HlH, in contrast to α_D and α_N-HlH which reassociate into shorter tubules. Although the capacity for the formation of tubules is also observed in *Rapana* and *H. pomatia* hemocyanins (Dolashka-Angelova et al., 2003; Lambert et al., 1995), a striking different behavior for the β-HlH isoform was observed, in the sense that it forms multidecamers.

Another reason to determine the gene sequence of *H. lucorum* hemocyanin is that it has immunological properties that make it a potential tumor vaccine carrier (Iliev et al., 2008; Yossifova et al., 2009; Dolashka et al., 2011), a property that has also been shown for other hemocyanins (Harris and Markl, 1999; Toshkova et al., 2006). The immunostimulatory properties are likely due to the high carbohydrate content and specific monosaccharide composition of these proteins (Dolashka-Angelova et al., 2008; Harris and Markl, 1999; Wuhrer et al., 2000, 2004). The, yet incompletely known, oligosaccharide structures of *H. lucorum*, *R. venosa* and *M. crenulata* hemocyanins appear to differ, which explains the different immunostimulatory properties of these proteins. To help unraveling the complete covalent structures of the hemocyanins, the knowledge of their amino acid sequence is essential to know the positions of the oligosaccharide linkage sites. In the present study, the complete cDNA sequences from two subunits and a partial cDNA sequence from the third subunit of *H. lucorum* are described and evaluated from an evolutionary point of view. The sequences and the expression of the three genes were confirmed by in-depth mass spectrometric analysis of tryptic digests from the purified proteins.

2. Materials and methods

2.1. Chemicals

All the solvents for mass spectrometric sample preparation were from Biosolve (Valkenswaard, The Netherlands).

2.2. Animals

Five snails were collected in the wild in Sofia (Bulgaria). The foot was isolated and stabilized in RNAlater (Sigma) and stored at −20 °C.

2.3. Preparation of RNA and RT-PCR, purification and sequencing of PCR fragments

RNA was extracted from 20 mg RNAlater-stabilized foot tissue of *H. lucorum* using the RNAeasy mini kit from Qiagen (Hilden, Germany). The tissue was disrupted and homogenized by means of glass beads. The spin-column preparation was performed according to the manufacturer's instructions. The RNA was eluted in 30 μl RNase-free water and stored at −20 °C.

cDNA was synthesized using the Transcriptor First strand cDNA synthesis kit starting from 1 μg total RNA (Roche Diagnostics, Mannheim, Germany). The first RT-PCR reactions were performed using degenerated primers (listed in Table 1) based on the conserved oxygen binding regions from different hemocyanin sequences. Degenerated primers were also designed based on the known N-terminal sequences on *H. lucorum* hemocyanins (Velkova et al., 2010). All the primers used in the PCR reactions were synthesized by Sigma. PCR reactions were performed with the long template PCR system (Roche Diagnostics, Mannheim, Germany) using buffer 2, with different combinations of the degenerated primers. A temperature gradient was set for the different reactions to determine the optimal PCR conditions. PCR reaction products were separated by agarose gel electrophoresis, and different fragments with the expected length were isolated using the gel extraction kit from Qiagen. The purified PCR products were cloned in the TOPO TA sequencing kit following the manufacturer's instruction manual (Invitrogen, Karlsruhe, Germany). The cDNA containing plasmids were sequenced from both ends using M13 forward/reverse primers (Cogenics, Hope End, Essex, UK). Primer walking was performed with gene-specific primers, purchased from Cogenics as well. Based on the determined sequence, gene-specific primers were designed. Different combinations of primer pairs were used trying to obtain fragments which could fill the gaps. The designed primer sequences were always controlled for their specificity. 5′ and 3′ RACE experiments were performed by means of the GeneRacer Kit from Invitrogen using the standard conditions supplied by the manufacturer.

Table 1
Degenerated primers used in this study.

Primers	Sequence
N-terminal β-subunit	TTRAAYGTICAYTTRYTIACITTRCA
Conserved stretch 1 forward	GGIWKGSCIRYITTYCCICAITGGCA
Conserved stretch 1 reverse	TACCAITGIGGRAAIRYIGSCMWICC
Conserved stretch 2 forward	CARVAITTYGARRTICARTWYGAGCT
Conserved stretch 2 reverse	AGCTCRWAYTGIAYYTCRAAITBYTG

2.4. Sequence data analyses, phylogenetic studies and molecular clock calculations

Data obtained from the sequencing reactions were analyzed with the software packet Vector NTI (Invitrogen). Overlapping regions were detected and assembled. The sequences are submitted to Genbank (accession numbers JF752343–JF752345). SIGNALP was used for the prediction of the signal peptide (Emanuelsson et al., 2007). Multiple sequence alignments and identity matrices were calculated by CLUSTALW2 (www.ebi.ac.uk). Phylogenetic analyses were performed involving the sequences of the *H. lucorum* hemocyanin isoforms with those of the following proteins deposited in GenBank: hemocyanins with one isoform: *Aplysia californica* (AJ556169) and *Nautilus pompilius* (AJ619741), hemocyanins with two isoforms: *Haliotis tuberculata* isoform 1 (AJ252741) and isoform 2 (AJ297475), *M. crenulata* isoform 1 (AJ698341) and isoform 2 (AJ698342), *Nucula nucleus* isoform 1 (AJ786639), *N. nucleus* isoform 2 (AJ786640), *Octopus dofleini* isoform 1 (AF338426, isoform 2 is 97% identical at the amino acid level and was not included in the phylogenetic analysis), and two FUs of *R. venosa* (RvH-b and RvH-c, Dolashka-Angelova et al., 2009). The assignment of the single functional unit (FU) for each clone was done with MEGA 5.05 (Tamura et al., 2007) using the implemented Neighbor Joining algorithm.

2.5. MS analyses

2.5.1. Trypsin digestion

Approximately 20 μg of different protein fractions (purified hemocyanin or purified structural subunits, in 50 mM Tris.HCl, pH 8.5) was reduced by means of 2 μl 50 mM dithiothreitol (Fluka, Buchs, Switzerland) at 55 °C for 45 min. After cooling to room temperature, 2 μl 100 mM iodoacetamide (Fluka, Buchs, Switzerland) was added and the mixture was placed in the dark for 30 min. An amount of 0.5 μg of modified trypsin (Promega, Madison, WI) was added and the protein was incubated overnight at 37 °C. Digestion was stopped by adding 0.5 μl formic acid.

2.5.2. LC–MALDI–MS/MS

Some 4.5 μg of digest material was separated on an Ultimate nanoHPLC system (Dionex, Amsterdam, The Netherlands). The sample was loaded onto a reverse phase trapping column (Pepmap C18,

800 μm × 5 mm, Dionex) using the SWITCHOS device. The loading pump was operated at 10 μl/min, and 2% acetonitrile and 0.05% TFA was used as mobile phase. After 8 min, the valve was switched and the sample was eluted onto the analytical separation column (PepMap C18, 75 μm × 150 mm, Dionex) in the back-flush mode, using the nanoLC pump operated at 250 nl/min. The mobile phases used were 5% ACN/0.1% TFA (v/v) for Buffer A and 80% ACN/0.1% TFA (v/v) for Buffer B. Peptides were resolved by gradient elution using a gradient of 0–50% Buffer B over 25 min, followed by a gradient of 50–100% Buffer B over 10 min. Column effluent was monitored using a 3 nl UV flow cell (214 nm). The sample was mixed on-line with matrix solution (4 mg/ml α-cyano-4-hydroxycinnamic acid in 70% ACN containing 0.1% TFA) and spotted directly onto a MALDI target at 30 s intervals, using the Probot device (Dionex). The result-dependent MS/MS experiments were performed on a 4800 Proteomics Analyzer (AB Sciex, Foster City, CA), using the job-wide interpretation method. In this strategy, full mass spectra were collected, and individual molecular ions were chosen for fragmentation from the spots in which their peak intensity was the highest. The spot to spot precursor selection was set to 200 ppm, and the S/N ratio of the parent ion was at least 200. MS/MS analysis was performed from the 8 highest peaks, from the highest to the least intense peak. For MS data acquisition, a total of 1000 shots were collected (50 sub-spectra accumulated from 20 laser shots each). All MS/MS data was acquired in the 1 keV MS/MS mode using air as the collision gas (1.2×10^{-7} Torr). A total of 4000 shots (50 sub-spectra accumulated from 80 laser shots each) were acquired and the timed-ion-selector window was set to resolution 400 (FWHM). The 4800 Proteomics Analyzer was externally calibrated prior to analysis, as outlined by the manufacturer.

The MS/MS data collected during an LC-MALDI run were submitted to the search software Mascot v2.1 (Matrix Science, London, UK) using the GPS explorer v3.6 software (AB Sciex). MS/MS spectra were searched against the translated sequence of the 3 subunits concatenated with a shuffled decoy database generated by the Decoy Database Builder software tool (Reidegeld et al., 2008). Carbamidomethylation (of Cys) and oxidation (of His, Met, Trp) were selected as variable modifications, and two missed cleavages were allowed with trypsin as the cleaving agent. Searches were done with a tolerance of 200 ppm for the precursor ion and 0.5 Da in the MS/MS mode. The results were filtered using the MS/MS ion score ≥ 20.

2.5.3. LC-ESI-FT-MS/MS

A 0.5 μg tryptic digest was diluted 10 times in 2% (v/v) ACN, 0.1% (v/v) formic acid and a volume of 5 μl of this dilution of each fraction was analyzed independently using a fully automated LC–MS/MS setup. Peptides were first separated on an Agilent 1200 chromatographic system (Agilent, Santa Clara, CA, USA) and on-line measured on an LTQ-FT Ultra mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). The samples were first loaded and desalted on a Zorbax 300SB-C18 trapping column (5 mm × 0.3 mm, Agilent) at a flow rate of 4 μl/min using a 2% (v/v) ACN, 0.1% formic acid buffer. They were then separated on a Zorbax 300SB-C18 analytical column (150 mm × 75 μm, Agilent), using a 50 min linear gradient ranging from 2% (v/v) to 80% (v/v) ACN, 0.1% formic acid, at a flow rate of 0.3 μl/min. The LC-effluent was directly coupled to a Triversa NanoMate ESI source (Advion, Ithaca, NY, USA), working in the nano-LC mode and equipped with a D-chip upon which a 1.55 kV voltage was supplied. During the LC-separation, the FT-ICR mass analyzer acquired MS scans at 100,000 resolution; the three most intense precursor ions for each MS scan were automatically selected and fragmented by the LTQ ion trap mass analyzer. Raw LC-ESI-FT-MS/MS data were analyzed with the Sequest database searching algorithm (Eng et al., 1994) implemented in the Bioworks v3.3.1 software (Thermo Fisher Scientific). MS/MS spectra were searched against the translated cDNA sequence. Mass tolerances were set to 10 ppm for the peptide mass and 0.5 amu for the fragment mass. Two missed cleavages were allowed with trypsin as the cleaving agent and only b- and y-ions were considered. The thresholds were set as follows: minimum Xcorr score +1, 1.5; +2, 2.0; +3, 2.5; minimum DeltaCN, 0.08; maximum peptide probability, 0.001 and maximum protein probability, 0.001. Raw LC-ESI-FT-MS/MS data were also analyzed using Mascot Distiller v2.3.2, with the spectra being processed as outlined by the manufacturer. The processed spectra were searched against the database described above. Carbamidomethylation (Cys) and oxidation (His, Met, Trp) were selected as variable modifications, and two missed cleavages were allowed with trypsin as the cleaving agent. Searches were done with a tolerance of 0.3 Da for the precursor ion and 0.3 Da in the MS/MS mode. The results were filtered using the MS/MS ionscore ≥ 20.

3. Results and discussion

Several molluscan hemocyanins with one or two isoforms were isolated and their gene sequences determined (Bergmann et al., 2006; Lieb et al., 2001). However, an additional third isoform was identified in the hemocyanins of *H. pomatia* and *H. lucorum* (Lambert et al., 1995; Velkova et al., 2010). No sequence information was hitherto available on those hemocyanins. In this study, we determined the gene sequence of all three isoforms (β-HlH, α_D-HlH, and α_N-HlH) from *H. lucorum* and compared them with the sequences from related organisms.

For the subunits β-HlH and α_D-HlH we obtained the complete sequence, whereas for the third subunit we obtained a partial sequence missing the 5′ and 3′ ends. The resulting gene sequences are shown in Supplementary Fig. S1, the translated sequences are shown in Fig. 1. cDNA was synthesized using oligo dT primers. Fragments of hemocyanin were generated during an RT-PCR reaction using degenerated primers. The available hemocyanin sequences were aligned to determine the conserved sequence regions. The first primer was situated around the conserved WHRL motif (conserved stretch 1) with His as one of the conserved copper ligands. The other primer contained another copper ligand within the sequence pattern FEVQFEVSHN (conserved stretch 2). The conserved stretches are boxed in Fig. 1. The primers were synthesized in both directions and different combinations were made in the PCR reactions. After amplification, the PCR products were analyzed by agarose gel electrophoresis and fragments with the expected length were isolated and cloned in the TOPO TA cloning kit. The hemocyanin is composed of 8 functional units which each are joined by a sequence linker, meaning that these conserved stretches are repeated several times in the same molecule. As the primers can bind several times within the same molecule, fragments of different lengths could be amplified. The inserts were sequenced in both directions using the standard vector primers. The resulting sequences were analyzed and assembled using Vector NTI. The obtained sequences were aligned against the known hemocyanin sequences to determine their relative positions and were used to design gene-specific primers. The design of these primers was extremely challenging due to the high similarity between the three subunits on the one hand and between the functional subunits on the other. Through alignment with the *A. californica* hemocyanin sequence, the relative position of the fragment-specific primers was determined which allowed us to combine them in order to amplify the lacking sequence fragments. Using this strategy we were able to assemble the different fragments into three different subunits. In all cases, the fragments were only assembled when the overlapping region was more than 200 bp, making sure that they represented a single mRNA. The 5′ and 3′ regions of the different genes were determined using the 5′ and 3′ RACE procedure, as described in Materials and methods.

The obtained sequences could be assembled into three cDNA sequencing, corresponding to the three subunits isolated from the hemolymph of *H. lucorum*, (Velkova et al., 2010). Based on the N-

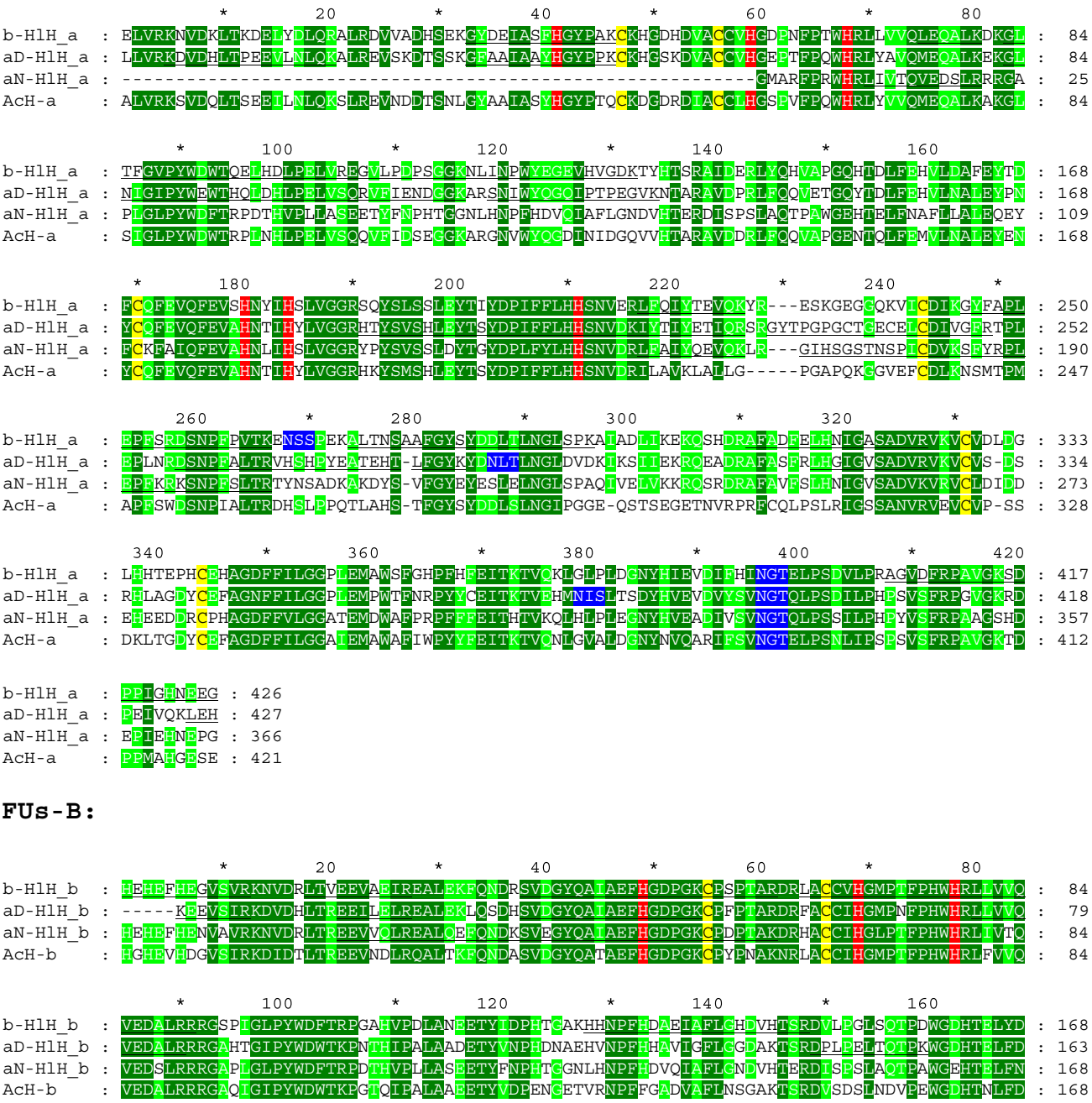
Signal sequences

β-HlH-A : MAKLWFALSLALLICLGGC : 21

α_D-HlH-A : MAPTIVWLAFTFMLVCSNA : 19

AcH-A : MVGYLGQALMALLLLALSNA : 20

FUs-A:



FUs-B:

Fig. 1. Alignment of the amino acid sequence of the eight functional units of three subunits from *Helix lucorum* hemocyanin as predicted from their cDNA sequence. For each functional unit a multiple sequence alignment of the different subunits in comparison to *Aplysia californica* hemocyanin is shown. For FU-g the *Octopus dofleini* sequence is included in the alignment. N-glycosylation sites are highlighted, putative disulfide and thioether bridges are marked as given at the bottom of the legend. The alignment was performed using ClustalX and was edited using GeneDoc.

terminal sequence of the isolated β-subunit of HlH obtained by Edman degradation, we were able to conclusively identify one subunit as β-HlH. This subunit is 10 859 bp long (GenBank ID: JF752343). The potential start codon is situated at position 48, yielding a 5′ UTR (untranslated region) of 47 bp. The stop codon (TGA) is at position 10348, followed by a 3′ UTR of 510 bp. In the 3′ UTR, a typical polyadenylation signal (AATAAA) and a poly(A) tail is present. The second subunit is 10 461 bp and assigned as subunit α_D (GenBank ID: JF752344), also according to the Edman degradation results. However, we here correct the N-terminal sequence at two positions compared to the previously reported one (Velkova et al., 2010). The first starting codon is situated at position 53 bp, yielding a 5′ UTR of 52 bp. The stop

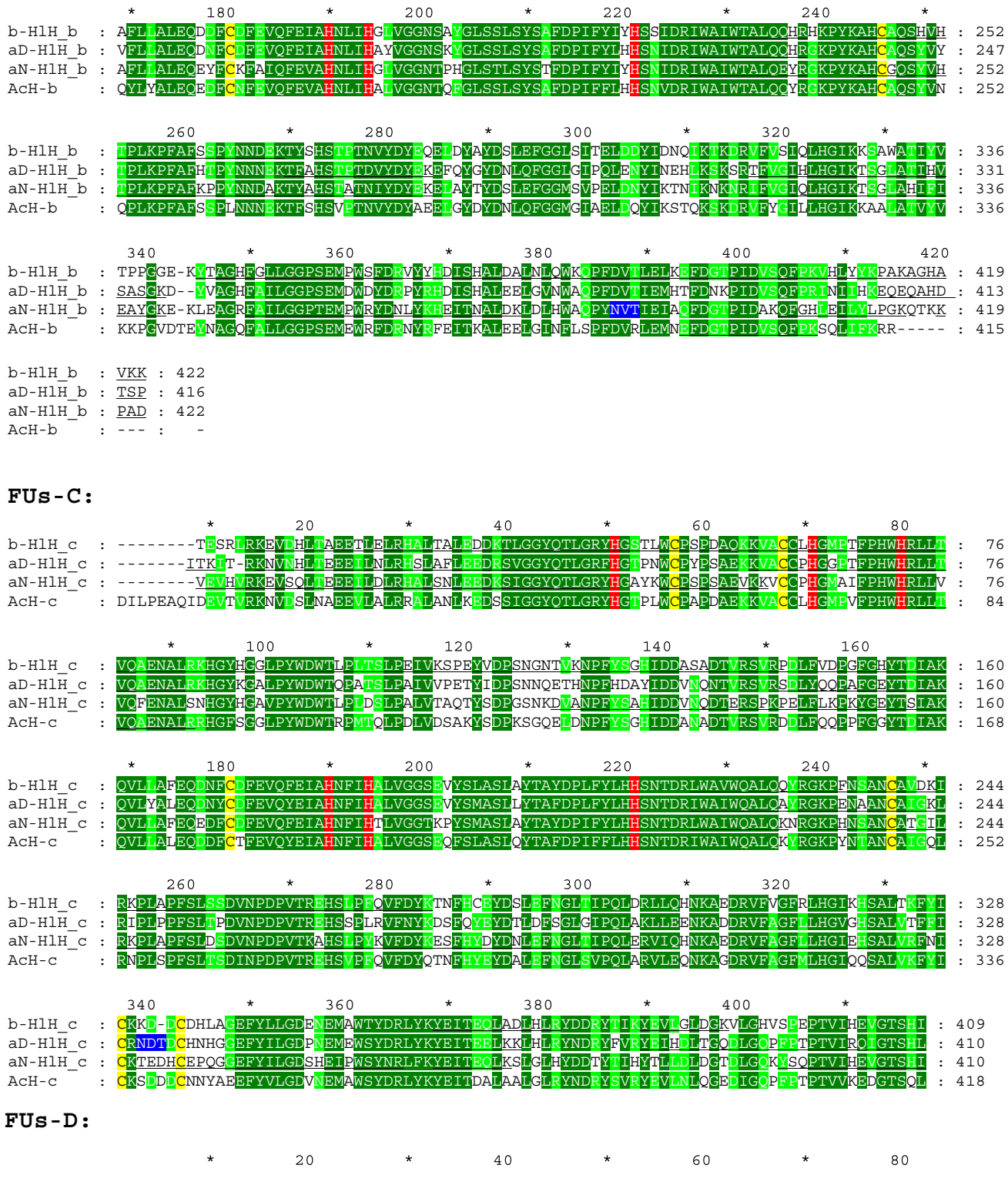
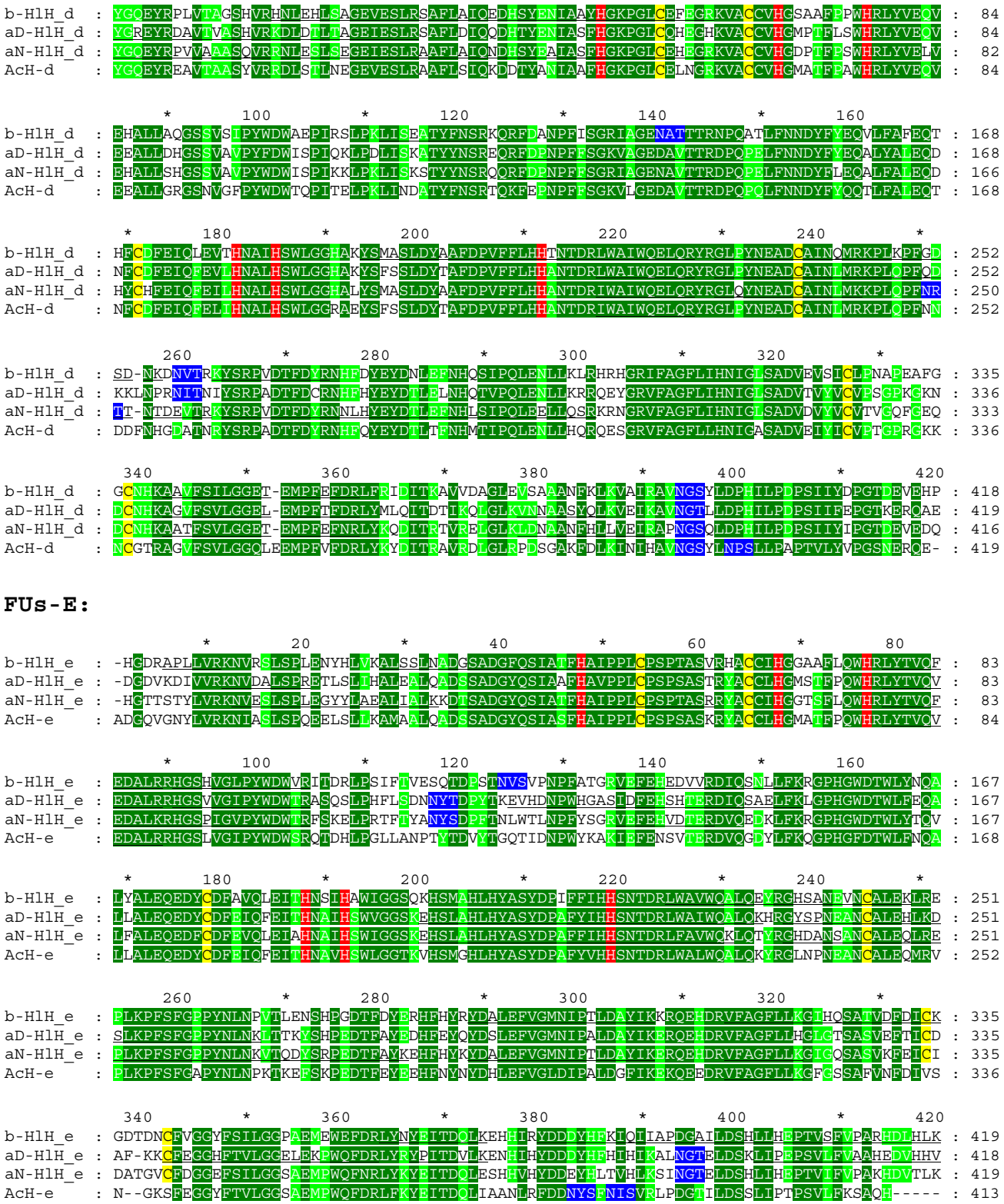


Fig. 1 (*continued*).

codon (TAG) is at position 10338 bp followed by a 3′ UTR of 121 bp. The 3′ UTR possesses the poly(A) tail. We were unable to determine the 5′ and 3′ end of the cDNA for the third subunit; nevertheless we identified 9095 bp and named it α_N (GenBank ID: JF752345).

The translated amino acid structures of the hemocyanin polypeptide of subunits β and α_D start with a signal sequence of 19 amino acids, as predicted using the SIGNALP 3.0 software (Fig. 1). Such a signal sequence is also present in other molluscan hemocyanins (Bergmann et al., 2006; Lieb et al., 2001). The signal sequence is used to secrete the hemocyanin via endoplasmic reticulum mediated exocytosis (Albrecht et al., 2001; Sminia and Vlugh-vanDallen, 1977). The native hemocyanins are 3414 and 3409 amino acids long for the β-subunit and the α_D-subunit respectively. The cDNA sequence for the α_N-subunit codes for a polypeptide chain of 3029 residues, but lacks the N- and C-termini. The structural subunits are clearly substructured into eight different functional units of around 400



FUs-E:

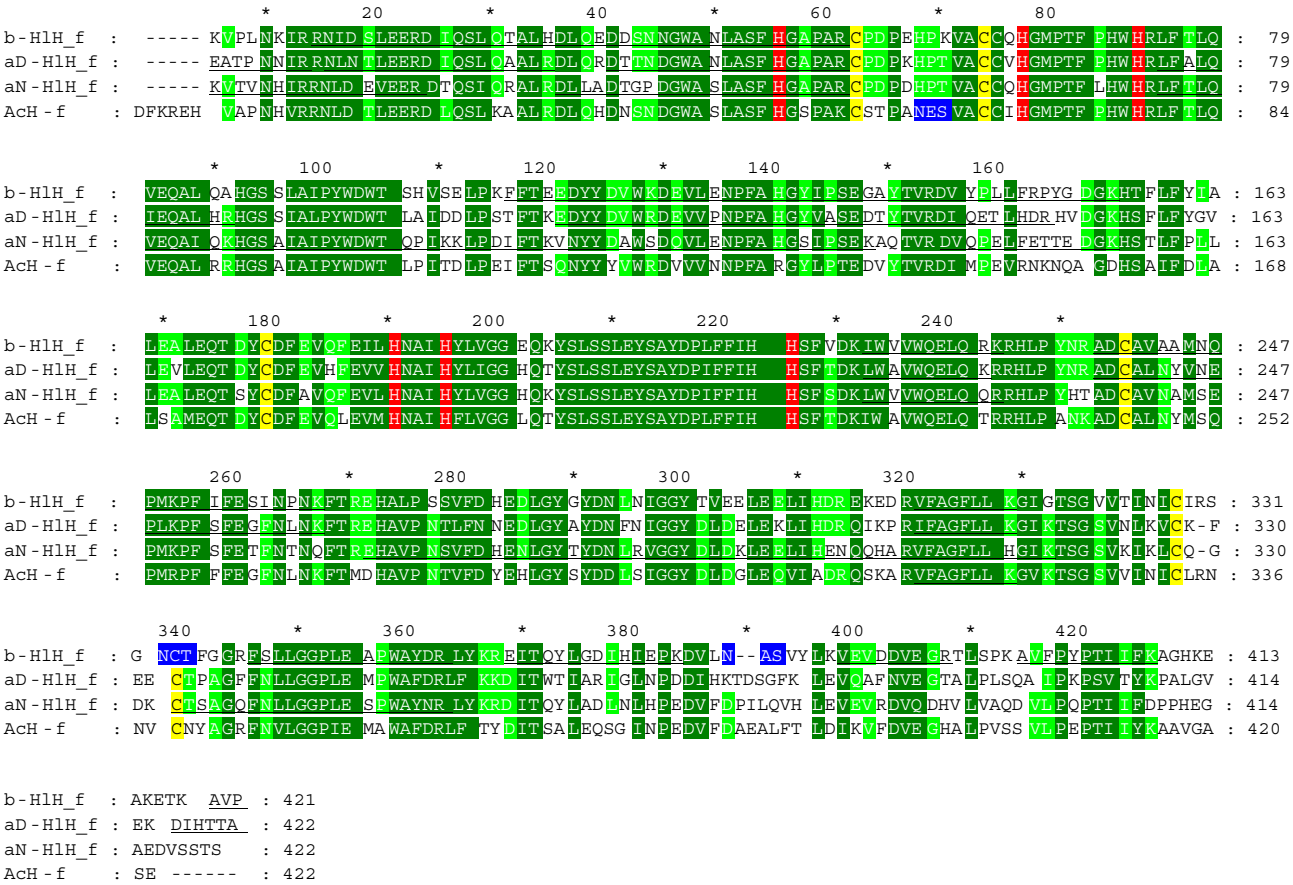
Fig. 1 (*continued*).

amino acids (Fig. 1), known as FU a, b, c, d, e, f, g and h. FUh carries a unique tail extension.

Hemocyanin sequences and properties of several FUs from Vetigastropoda (*Haliotis*), Caenogastropoda (*Rapana*), Heterobranchia (*Aplysia*) and from the cephalopod *Octopus dolfeini*, *N. pompilius and Sepia officinalis* are available (Altenhein et al., 2002; Bergmann et al., 2006; Dolashka-Angelova et al., 2003, 2007; Lieb et al., 2000, 2004). Multiple sequence alignments (Fig. S2) with all the known sequences from hemocyanins of other organisms show that the *H. lucorum* sequence are most closely related to the *A. californica* sequence, which is the reason why this sequence is used in the alignment shown in Fig. 1.

The primary structures of the *H. lucorum* hemocyanin polypeptide chains, as predicted from its cDNA sequence, show a variety of features and motifs which are typical for all the molluscan

FUs-F:



FUs-G:

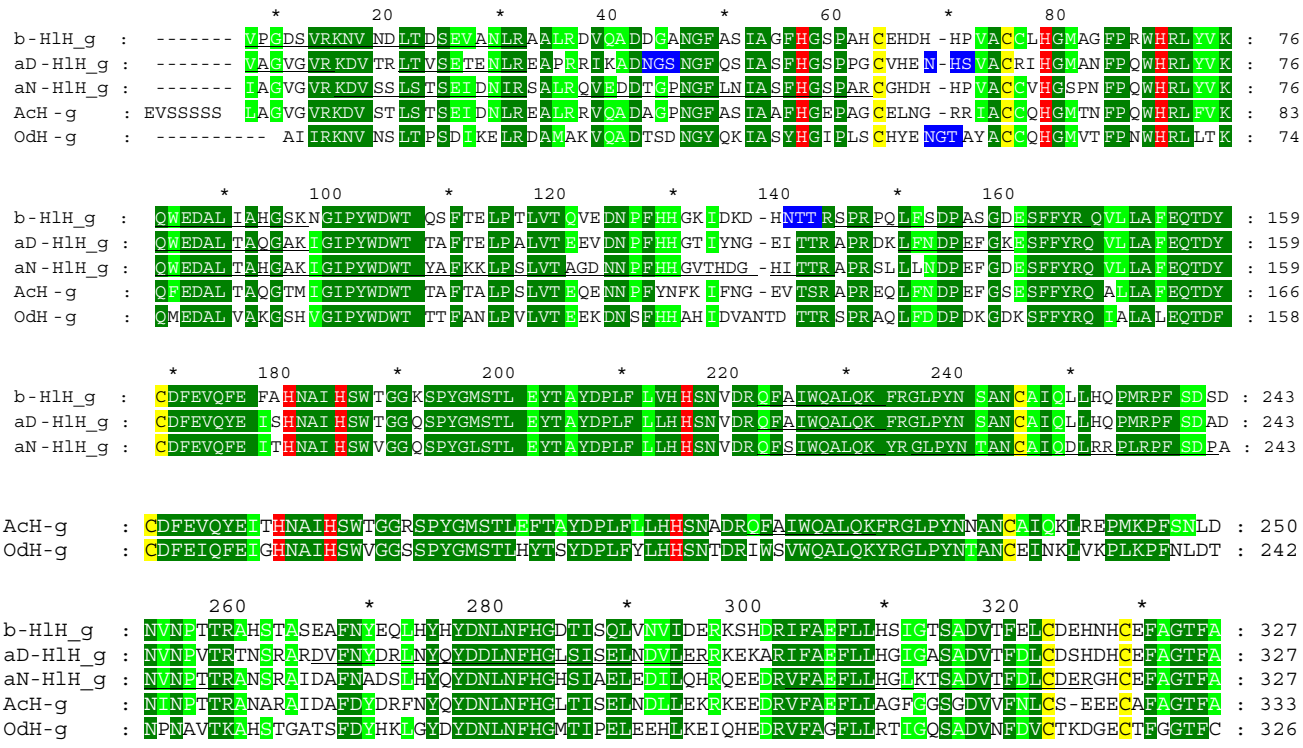


Fig. 1 (*continued*).

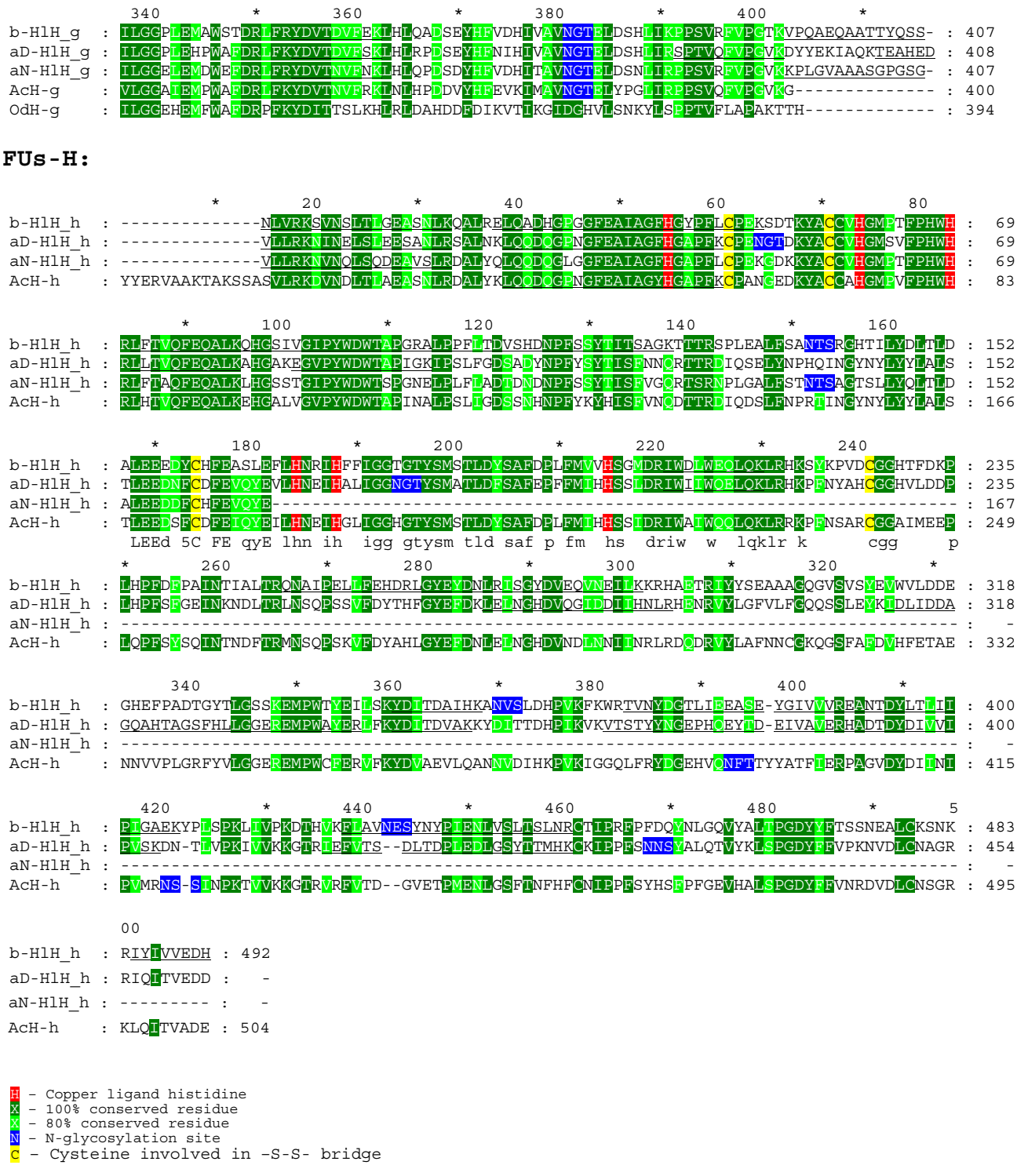


Fig. 1 (*continued*).

hemocyanins sequenced to-date. The most conserved regions surround the eight active sites, each showing a typical copper A and copper B site with six copper-binding histidines. Among the strictly conserved features are two disulfide bridges (indicated in yellow in Fig. 1), which likely stabilize the core domain, as it was shown that oxygen-binding was completely lost after disulfide bond reduction (Topham et al., 1999). The third disulfide bridge (e.g. between Cys326 and Cys339 in FUs A) is not conserved in functional unit B, as in all studied molluscan hemocyanins.

It is known that molluscan hemocyanins are powerful immunogens, probably due to their high carbohydrate content and specific monosaccharide composition (Lommerse et al., 1997). Structural studies of *H. lucorum* hemocyanin showed that it is very heterogeneously glycosylated, carrying mainly methylated high mannose-type glycans (Dolashka-Angelova et al., 2008). Analysis of the sequence of the different subunits revealed that there are 11 potential N-glycosylation sites with the motifs NXT or NXS in the β subunit, 14 in the α_D subunit, and 7 in de α_N subunit (Fig. 1). Most molluscan hemocyanins lack N-

Table 2
Properties of the three subunits of *Helix lucorum* and its functional units as predicted from the cDNA sequence.

Structure	# Amino acids	pI	Molecular mass	# N-linked glycosylation sites	# Amino acids	pI	Molecular mass	# N-linked glycosylation sites	# Amino acids	pI	Molecular mass	# N-linked glycosylation sites
	β-subunit				α_D-subunit				α_N-subunit			
	3412	5.52	388904.1	11	3374	5.60	386783.9	13	3031	5.58	346030.2	3
ss	21				19				ND			
FU-a	424	5.26	47980.5	2	427	5.96	48683.8	3	366	5.67	41700.6	0
FU-b	422	5.59	48079.5	0	416	5.72	47644.1	0	422	5.99	48311.3	0
FU-c	409	5.62	46847.6	0	410	5.62	47185.8	1	412	5.56	47164.8	0
FU-d	418	5.67	47595.3	3	419	5.60	48148.1	2	420	5.57	48232.0	1
FU-e	419	5.83	48159.9	1	418	5.42	48031.3	2	420	5.78	48552.4	1
FU-f	421	5.31	48334.4	0	422	5.61	48453.5	0	420	5.23	47906.6	0
FU-g	407	5.39	46090.9	2	418	5.67	47912.5	2	404	5.55	45605.6	1
FU-h	492	5.46	55944.9	3	489	5.28	56025.7	3	167	4.86	18682.8	0

glycosylation sites on FU-c, but FU-c of subunit α_D in *H. lucorum* does contain such a site. In the three subunits of *H. lucorum* hemocyanin, there are no N-glycosylation sites on FU-B, whereas there are two conserved sites, one in functional unit A and one in functional unit D. The availability of the sequence will be of great help in the detailed study of the glycosylation pattern of the protein, which is of importance because glycan-structures are responsible for the immunostimulatory properties of hemocyanins and for their application as potential tumor vaccine carriers (Wuhrer et al., 2000). Some characteristics for each of the functional units are reported in Table 2, being calculated from the sequence.

As a complementary part of this gene sequence analysis, we carried out peptide mapping of the native hemocyanin in order to verify whether the three subunits are effectively part of the complex, and to exclude the possibility that one of the genes is a pseudogene. The peptide mixture obtained after trypsin digestion of the reduced and alkylated protein was analyzed using LC–MALDI–MS/MS and LC-EST-FT-MS/MS. We were able to identify peptides from the three subunits; for subunit β we traced more

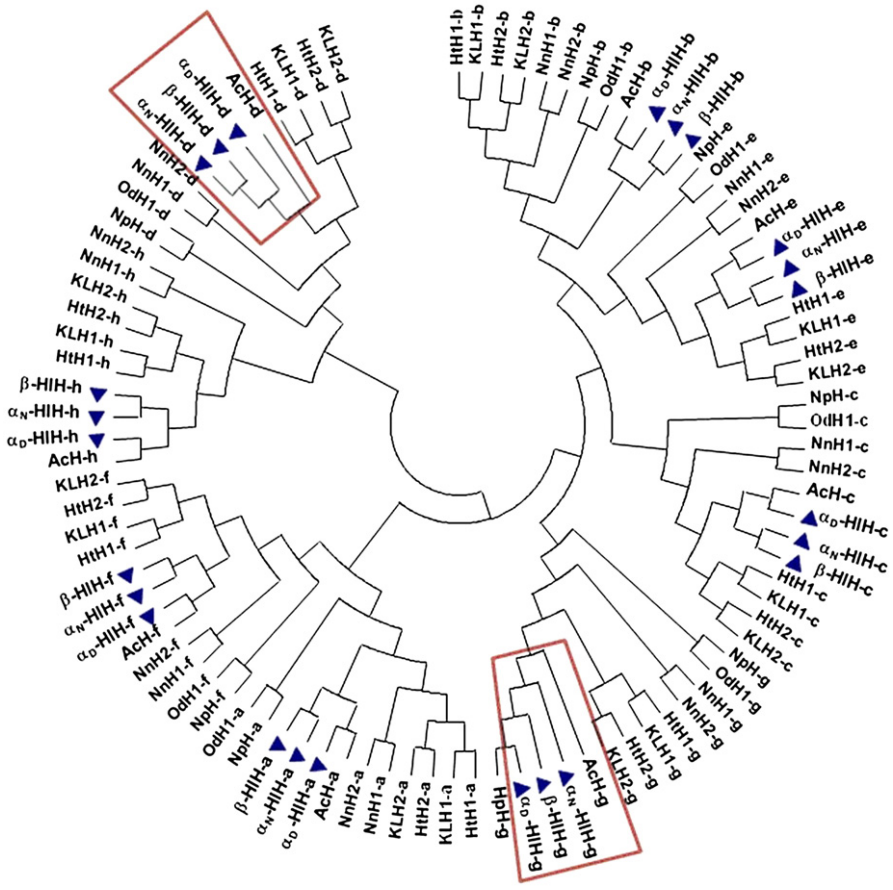


Fig. 2. Phylogenetic analysis of molluscan hemocyanins functional units. A radial phylogenetic tree of sequences of molluscan hemocyanins with one isoform: *Aplysia californica* (AcH)(AJ556169), and *Nautilus pompilius* (NpH)(AJ619741), with two isoforms: *Haliotis tuberculata* (HtH1)(AJ252741) and (HtH2)(AJ297475), *Megathura crenulata* (KLH1)(AJ698341) and (KLH2)(AJ698342), *Nucula nucleus* (NnH1)(AJ786639) and (NnH2)(AJ786640), *Octopus dofleini* isoform 1 (OdH1)(AF338426), the two FUs of *Rapana venosa* (RvH-b and RvH-c, Dolashka-Angelova et al., 2009), and the three structural subunits from *Helix lucorum* hemocyanin (β-HLH, α_D-HlH and α_N-HLH). The assignment of the single functional unit (FU) for each clone was done with MEGA 5.05 (Tamura et al., 2007) using the implemented Neighbor Joining algorithm. Topologically related FUs get clustered, forming eight distinct branches. Note that OdH2 is identical to OdH1 at the protein level.

than 30% sequence coverage, for subunit α_D this was 28%, and 30% for α_N. Supplementary Table S1 provides a summary of the mass spectral results. The identified peptides are underlined in Fig. 1. Supplementary Table S2 gives an overview of the unique peptides found with the different techniques.

The molluscan hemocyanins are extremely large polypeptides and are therefore rich in structural/sequence information useful for phylogenetic studies. Since polypeptides from different gastropod (clades: Vetigastropoda, Caenogastropoda and Heterobranchia) and cephalopod (*O. dolfeini*, *N. pompilius and S. officinalis*) molluscan hemocyanins are known, the data allow the reconstruction of a robust phylogenetic tree from multiple sequence alignments. Such an effort has recently been made with the aim to assess divergence time between Sepia and Spirula within the clan of the Cephalopoda (Warnke et al., 2011). So far, the evolutionary relationship between the Pulmonata (Helix) and the Opisthobranchia (Aplysia) within the Heterobranchia is controversially debated. The obtained sequences of Helix will stabilize this region of the tree.

Using the program Mega 5.05, we constructed an unrooted radial phylogenetic tree based on the three subunits from *H. lucorum* and all other known molluscan hemocyanin sequences. Molluscan hemocyanins are phylogenetically related to tyrosinase and catechol oxidase (Klabunde et al., 1998; van Gelder et al., 1997), but because of their overall close primary structure relationship the use of the latter enzymes as outgroup is inefficient. We therefore decided to construct an unrooted tree (Fig. 2). The tree reveals eight stable branches that represent the eight functional units, which emphasizes their very early divergence from a single ancestral FU (Altenhein et al., 2002; Lieb et al., 2001). Note that a phylogenetic tree constructed with all molluscan hemocyanin amino acid sequences known so far also generates eight branches (results not shown), in support of the existence of eight functional units of three isoforms in *H. lucorum* hemocyanin. It is remarkable that for 6 out of the 8 subunits, α_N-HlHg and α_N-HlHd are more related to the subunit of the Heterobranchian species *A. californica* (AcH-g) whereas for the 2 other subunits the homology is higher with the same subunits of the 2 other isoforms α_D and β. Our results also confirm that the eight FUs are phylogenetically older than the divergence between the Cephalopoda and the Gastropoda.

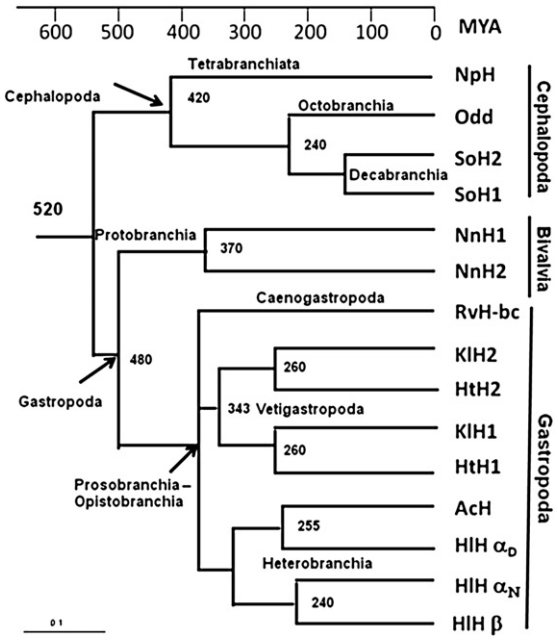


Fig. 3. Timescale of the evolution of molluscan hemocyanins. The molecular clock is calculated from the distance values of the various hemocyanin sequences mentioned in the legend of Fig. 2. The splitting point of gastropods, bivalvia and cephalopods was used for calibration (520 million years ago). MYA, million years ago.

We also attempted to quantify the molecular evolution of the molluscan hemocyanins as a whole. The tree shown in Fig. 3 was calculated from a difference matrix using all the completely known sequences. As in the phylogenetic trees previously published (Bergmann et al., 2007; Lieb et al., 2004; Warnke et al., 2011), there is a very stable separation into gastropod and cephalopod hemocyanin, and this splitting point of 520 million years ago was therefore used for calibration (Benton, 1993). In this phylogeny, a number of supraorders of the class of the Gastropoda are distinguished: clade Vetigastropoda (*M. crenulata* and *Haliotis tuberculata*); Caenogastropoda (*R. venosa*) and Heterobranchia (*A. californica* and *H. lucorum*). With respect to the Caenogastropoda, only two partial FU sequences are available, namely, from RvH-b and RvH-c of *R. venosa* (Dolashka-Angelova et al., 2009). As is the case for 6 out of the 8 subunits of *H. lucorum* individually, isoform α_D, rather clusters with the hemocyanin of *A. californica* suggesting that they evolved from the same ancestor.

In conclusion, since the structure, evolution, and diversity of hemocyanins are of general scientific interest and their immunotherapeutic effects of biomedical concern, we consider the present communication on the cDNA sequence of the three subunits of *H. lucorum* to be an essential contribution to unravel the biochemical properties of these respiratory proteins.

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2011.07.030.

Acknowledgments

The authors are indebted to FWO-Vlaanderen for supporting an exchange program with the Bulgarian Academy of Sciences (grant VS.016.009N) and a research grant by DFG-01/2008.

References

Albrecht, U., Keller, H., Gebauer, W., Markl, J., 2001. Rhogocytes (pore cells) as the site of hemocyanin biosynthesis in the marine gastropod *Haliotis tuberculata*. Cell Tissue Res. 304, 455–462.

Altenhein, B., Markl, J., Lieb, B., 2002. Gene structure and hemocyanin isoform HtH2 from the mollusc *Haliotis tuberculata* indicate early and late intron hot spots. Gene 301, 53–60.

Benton, M.J. (Ed.), 1993. The Fossil Record 2. Chapman and Hall, London, pp. 125–270.

Bergmann, S., Lieb, B., Ruth, P., Markl, J., 2006. The hemocyanin from a living fossil, the cephalopod *Nautilus pompilius*: protein structure, gene organization, and evolution. J. Mol. Evol. 62, 362–374.

Bergmann, S., Markl, J., Lieb, B., 2007. The first complete cDNA sequence of the hemocyanin of a bivalve, the protobranch *Nucula nucleus*. J. Mol. Evol. 64, 500–510.

Burmester, T., 2001. Molecular evolution of the arthropod hemocyanin superfamily. Mol. Biol. Evol. 18, 184–195.

De Ioannes, P., Moltedo, B., Oliva, H., Pacheco, R., Faunes, F., De Ioannes, A.E., et al., 2004. Hemocyanin of the molluscan *Concholepas concholepas* exhibits an unusual heterodecameric array of subunits. J. Biol. Chem. 279, 26134–26142.

Decker, H., Tuczek, F., 2000. Tyrosinase/catecholoxidase activity of hemocyanins: structural basis and molecular mechanism. Trends Biochem. Sci. 25, 392–397.

Dolashka, P., Velkova, L., Iliev, I., Beck, A., Dolashki, A., Yossifova, L., et al., 2011. Antitumor activity of glycosylated molluscan hemocyanins via guerin ascites tumor. Immunol. Invest. 40, 130–149.

Dolashka-Angelova, P., Schwarz, H., Dolashki, A., Stevanovic, S., Fecker, M., Saeed, M., et al., 2003. Oligomeric stability of *Rapana venosa* hemocyanin (RvH) and its structural subunits. Biochim. Biophys. Acta 1646, 77–85.

Dolashka-Angelova, P., Stevanovic, S., Dolashki, A., Devreese, B., Tzvetkova, B., Voelter, W., et al., 2007. A challenging insight on the structural unit 1 of molluscan *Rapana venosa* hemocyanin. Arch. Biochem. Biophys. 459, 50–58.

Dolashka-Angelova, P., Stefanova, T., Livaniou, E., Velkova, L., Klimentzou, P., Stevanovic, S., et al., 2008. Immunological potential of *Helix vulgaris* and *Rapana venosa* hemocyanins. Immunol. Invest. 37, 822–840.

Dolashka-Angelova, P., Lieb, B., Velkova, L., Heilen, N., Sandra, K., Nikolaeva-Glomb, L., et al., 2009. Identification of glycosylated sites in Rapana hemocyanin by mass spectrometry and gene sequence, and their antiviral effect. Bioconjug. Chem. 20 (7), 1315–1322.

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. Nat. Protoc. 2, 953–971.

Eng, J.K., McCormack, A.L., Yates III, J.R., 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J. Am. Soc. Mass Spectrom. 5, 976–989.

Gatsogiannis, Ch., Markl, J., 2009. Keyhole limpet hemocyanin: 9-Å CryoEM structure and molecular model of the KLH1 didecamer reveal the interfaces and intricate topology of the 160 functional units. J. Mol. Biol. 385, 963–983.

Harris, J.R., Markl, J., 1999. Keyhole limpet hemocyanin (KLH): a biomedical review. Micron 30, 597–623.
Harris, J.R., Markl, J., 2000. Keyhole limpet hemocyanin: molecular structure of a potent marine immunoactivator. J. Eur. Urol. 37, 24–33.
Harris, J.R., Meissner, U., Gebauer, W., Markl, J., 2004. 3D reconstruction of the hemocyanin subunit dimer from the chiton *Acanthochiton fascicularis*. Micron 35, 23–26.
Iliev, I., Toshkova, R., Dolashka-Angelova, P., Yossifova, L., Hristova, R., Yaneva, J., et al., 2008. Haemocyanins from *Rapana venosa* and *Helix lucorum* display an antitumour activity via specific activation of spleen lymphocytes. C. R. Acad. Bulg. Sci. 61, 203–210.
Klabunde, T., Eicken, C., Sacchettini, J.C., Krebs, B., 1998. Crystal structure of a plant catechol oxidase containing a dicopper center. Nat. Struct. Biol. 5, 1084–1090.
Lambert, O., Boisset, N., Taveau, J.C., Préaux, G., Lamy, J.N., 1995. Three-dimensional reconstruction of the alpha D and beta C-hemocyanins of *Helix pomatia* from frozen-hydrated specimens. J. Mol. Biol. 248, 431–448.
Lieb, B., Altenhein, B., Markl, J., 2000. The sequence of a gastropod hemocyanin (HtH1 from *Haliotis tuberculata*). J. Biol. Chem. 275, 5675–5681.
Lieb, B., Altenhein, B., Markl, J., Vincent, A., van Olden, E., van Holde, K.E., et al., 2001. Structures of two molluscan hemocyanin genes: significance for gene evolution. Proc. Natl. Acad. Sci. U. S. A. 98, 4546–4551.
Lieb, B., Boisguérin, V., Gebauer, W., Markl, J., 2004. cDNA sequence, protein structure, and evolution of the single hemocyanin from *Aplysia californica*, an opisthobranch gastropod. J. Mol. Evol. 59, 536–545.
Lommerse, J.P., Thomas-Oates, J.E., Gielens, C., Préaux, G., Kamerling, J.P., Vliegenthart, J.F., 1997. Primary structure of 21 novel monoantennary and diantennary N-linked carbohydrate chains from alphaD-hemocyanin of *Helix pomatia*. Eur. J. Biochem. 249, 195–222.
Lontie, R., 1983. Components, functional units, and active sites of *Helix pomatia* hemocyanin. Life Chem. Rep. S1, 109–120.
Martin, A., Depoix, F., Stohr, M., Meissner, U., Hagner-Holler, S., Hammouti, K., et al., 2007. *Limulus polyphemus* hemocyanin: 10 Å structure, sequence analysis, molecular modelling and rigid-body fitting reveal the interfaces between the eight hexamers. J. Mol. Biol. 366, 1332–1350.
Reidegeld, K.A., Eisenacher, M., Kohl, M., Chamrad, D., Körting, G., Blüggel, M., et al., 2008. An easy-to-use Decoy Database Builder software tool, implementing different decoy strategies for false discovery rate calculation in automated MS/MS protein identifications. Proteomics 8, 1129–1137.
Sminia, T., Vlugh-vanDallen, J.E., 1977. Haemocyanin synthesis in pore cells of the terrestrial snail *Helix aspersa*. Cell Tissue Res. 183, 299–301.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary. Genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Topham, R., Tesh, S., Westcott, A., Cole, G., Mercatante, D., Kaufman, G., et al., 1999. Disulfide bond reduction: a powerful, chemical probe for the study of structure–function relationships in the hemocyanins. Arch. Biochem. Biophys. 369, 261–266.
Toshkova, R., Ivanova, E., Nastke, M.D., Stevanovic, S., Velkova, L., Voelter, W., et al., 2006. Hemocyanins as immunostimulators. IDOSY global. J. Mol. Sci. 1, 22–32.
Van Gelder, C.W., Flurkey, W.H., Wichers, H.J., 1997. Sequence and structural features of plant and fungal tyrosinases. Phytochemistry 45, 1309–1323.
Van Holde, K., Miller, K., Decker, H., 2001. Hemocyanins and invertebrate evolution. J. Biol. Chem. 276, 15563–15566.
Velkova, L., Dimitrov, I., Schwarz, H., Stevanovic, S., Voelter, W., Salvato, B., et al., 2010. Structure of hemocyanin from garden snail *Helix lucorum*. Comp. Biochem. Physiol. B 157, 16–25.
Warnke, K.M., Meyer, A., Ebner, B., Lieb, B., 2011. Assessing divergence time of Spirulida and Sepiida (Cephalopoda) based on hemocyanin sequences. Mol. Phylogenet. Evol. 58, 390–394.
Wood, E., Chaplin, M., Gielens, C., De Saedeleer, J., Préaux, G., Lontie, R., 1985. Relative molecular mass of the polypeptide chain of bc-haemocyanin of *Helix pomatia* and carbohydrate composition of the functional units. Comp. Biochem. Physiol. B 82, 179–186.
Wuhrer, M., Dennis, R.D., Doenhoff, M.J., Geyer, R., 2000. A fucose-containing epitope is shared by keyhole limpet haemocyanin and *Schistosoma mansoni* glycosphingolipids. Mol. Biochem. Parasitol. 110, 237–246.
Wuhrer, M., Robijn, M.L.M., Koeleman, C.A.M., Balog, C.I.A., Geyer, R., Deelder, A.M., et al., 2004. A novel Gal(b1-4)Gal(b1-4)Fuc(a1-6)-core modification attached to the proximal *N*-acetylglucosamine of keyhole limpet haemocyanin (KLH) N-glycans. Biochem. J. 378, 625–632.
Yossifova, L., Petkova, S., Dolashka-Angelova, P., Mihov, L., Zacharieva, S., 2009. Immunological research on the protective properties of a conjugate of total larval antigen with hemocyanin derived from *Helix lucorum* against infection with *Trichinella spiralis*. Biotech. Equip. 23, 597–600.