ResearchOnline@JCU

JAMES COOK UNIVERSITY AUSTRALIA

This is the author-created version of the following work:

Mendoza-Porras, Omar, Kamath, Sandip, Harrison, James O., Colgrave, Michelle L., Huerlimann, Roger, Lopata, Andreas L., and Wade, Nicholas M. (2020) ***Resolving hemocyanin isoform complexity in haemolymph of black tiger shrimp Penaeus monodon - implications in aquaculture, medicine and food safety.*** **Journal of Proteomics, 218 .**

Access to this file is available from:

https://researchonline.jcu.edu.au/63530/

© 2020 Published by Elsevier B.V. All rights reserved.

Please refer to the original source for the final version of this work:

https://doi.org/10.1016/j.jprot.2020.103689

Title

Resolving hemocyanin isoform complexity in haemolymph of black tiger shrimp *Penaeus monodon* - Implications in aquaculture, medicine and food safety

Abstract

Hemocyanin (Hc) is a multifunctional macromolecule involved in oxygen transport and non-specific immunity in shrimp. Hc is crucial in physiology and nutrition linked with optimal performance in aquaculture production systems. In medicine, Hc has been approved for clinical use in humans as adjuvant and anticancer therapeutic. In contrast, Hc has also been identified as one of the proteins causing anaphylaxis following shrimp consumption. The role of individual Hc isoforms remains unknown due to a lack of resolved Hc isoforms. We successfully identified eleven different Penaeus monodon hemocyanin (PmoHc) γ isoforms including two truncated isoforms (50 and 20 kDa) and one PmoHc β isoform in haemolymph using proteomics informed by transcriptomics.

Amino acid sequence homology ranged from 24 to 97% between putative PmoHc gene isoforms. Hc isoforms showed specific patterns of transcript expression in shrimp larval stages and adult hepatopancreas. These findings enable isoform level investigations aiming to define molecular mechanisms underpinning Hc functionality in shrimp physiology and immunity, as well as their individual immunogenic role in human allergy. Our research demonstrates the power of proteomics informed by transcriptomics to resolve isoform complexity in non-model organisms and lay the foundations for improved performance within the aquaculture industry and advance allergenic applications in medicine.

Significance: The roles of hemocyanin (Hc) in shrimp homeostasis and immunity as well as in human allergy are not well understood because the complexity of Hc isoforms has remained unresolved. Our results have confirmed the existence of at least 12 individual Hc isoforms in shrimp haemolymph and validated putative Hc gene assemblies from transcriptomics. Our findings will enable monitoring the expression of specific Hc isoforms in shrimp haemolymph during different environmental, nutritional and pathogenic conditions, thus providing insights into isoform specific functional roles. In medicine, the potential allergenicity of each Hc isoform could be determined and advance allergenic applications. Lastly, since Hc comprises up to 95% of the total protein in haemolymph, these isoforms become ideal targets for prawn provenance, traceability and food contamination studies.

1. Introduction

Hemocyanin (Hc) is the name for extracellular, large, multimeric, copper-based respiratory proteins found in the haemolymph (Hm) of many arthropods, molluscs and metazoans [1-3]. Hc comprises up to 95% of the total protein content of crustacean haemolymph [4, 5]. Arthropod Hc forms large protein aggregates in excess of 450 kDa, combining a 75 kDa

structural subunit into hexamers, or multiples of hexamers depending on the species [1, 6]. Each 75kDa subunit has two central copper binding domains that perform the main oxygen transport function, and several key domains that influence quaternary structure [7]. In crustaceans, Hc is thought to be composed of three distinct subunits (α-type, β-type or γ-type subunits), although a single subunit is able to aggregate into hexameric structures [8, 9].

The crystal structure of a hexameric Hc has been resolved for the spiny lobster, *Panulirus interruptus* [7]. Within penaeid shrimp, Hc mRNA sequences have only been resolved for β-type and γ-type subunits [10]. In black tiger shrimp *Penaeus monodon*, only a single γ-type Hc subunit sequence has been identified, although three functional subunits (Pm1, Pm2 and Pm3) have been sequenced from purified haemolymph protein extracts [9]. More recently, one β-type and at least four γ-type subunit isoforms have been characterised from *Penaeus vannamei* [11, 12], as well as a number of potential splice variants [13]. In addition, proteomics has been applied to resolve some Hc isoforms from *P. vannamei* [11]. However, most of the γ-type subunits remain unresolved, as do any of the Hc subunits from *P. monodon* putatively identified using transcriptomics.

In addition to species differences, functional adaptations of crustacean Hc quaternary structure or subunit expression have been observed in response to respiratory stress from hypoxia and also hypercapnia [14]. Changes in these parameters also occur at different developmental stages and within the moult cycle [15, 16]. Crustacean Hc expression as well as protein abundance are known to be affected by viral and bacterial pathogens [12, 14, 16-18]. Hemocyanin and Hc-derived peptides function directly in many complementary innate immune responses, through anti-bacterial [19-21], anti-viral [17, 22], haemolytic [23], melanotic [24] and phenoloxidase-like activities [25], with each significantly contributing to the host response.

Hemocyanin has shown potential for medical applications as immune-stimulant [26, 27], adjuvant [28] and proposed as an alternate immunotherapeutic in different types of cancer including bladder [29], colon carcinoma [30] and melanoma [26]. The immunogenic properties of keyhole limpet hemocyanin (KLH) from the mollusc keyhole limpet *Megathura crenulata* have been extensively studied for over five decades, with applications as an adjuvant as well as an immune surveillance tool in cancer vaccines [31-33]. More importantly, KLH is one of the few naturally occurring bioactive proteins that has been approved for clinical

application for its remarkable immunostimulatory properties, reflected by the generation of high levels of antibodies, a strong cellular response to various antigens, and the ability to target an immune response to tumour-specific antigens [34]. Consequentially, Hc from abalone *Haliotis tuberculata* [28] and *L. vannamei* [35-37] have been targeted for similar applications in medicine.

However, arthropod Hc has also been identified and characterised as a causative protein for IgE antibody mediated allergic sensitisation. Previous studies have demonstrated patient serum IgE binding to Hc from cockroaches [38]. Studies in shrimp and crab have also demonstrated IgE reactivity with Hc establishing potential allergenicity [5, 39, 40]. The interaction of IgE with Hc has been established as a cause of shrimp allergy in children as well as adults, and potential role in clinical cross-reactivity between shrimps and house dust mites [39, 41-43]. Moreover, the isoform-specific IgE reactivity has not been yet well characterised, leading to a lack of knowledge on the immunological mechanisms governing the allergenicity imparted by shrimp Hc and the potential that isoform differences play a role in this activity.

At the transcript level, there has been substantial progress in understanding the complexity of shrimp haemocyanin and its functional response to external stimuli [10, 12, 13, 19]. However, only a single study aimed at resolving the total number of Hc isoforms present in shrimp by analytical chemistry has been reported to date [11]. Proteomics and transcriptomics are powerful complementary tools that have aided characterising novel proteins in non-model species [44, 45]. Here, proteomics informed by transcriptomics was used to characterise and resolve the complexity of Hc isoforms in one of the most farmed shrimp species, *P. monodon*.

2. Material and methods

2.1 Haemolymph collection

Sub-adult black tiger shrimp *P. monodon* were sourced from 100 L tanks with water flow of 1 L per min where temperature was maintained at 28ºC and salinity at 35 g/L. Shrimp were anaesthetised by immersion in a seawater-ice slurry and haemolymph (100 - 200 µL) collected from the pericardial sinus. To remove hemocytes and prevent coagulation, haemolymph was

centrifuged for 30 s at 3,000 x *g*. The hemocyte pellet was discarded, and haemolymph plasma stored at –80˚C until use. Identification of Hc isoforms using mass spectrometry (MS) was carried out using a pool of haemolymph plasma from seven shrimp. The raw haemolymph plasma was subjected to ultracentrifugation at 150, 000 x *g* at 4˚C for 30, 60 and 120 min in triplicate. The supernatant from each fraction was collected and prepared for MS analysis. Total protein content estimation was carried out using Bradford reagent using bovine serum albumin standard curve.

2.2 SDS PAGE and in-gel digestion

Proteins (10 µg Bradford) from each fraction were resolved by molecular weight on SDS-PAGE gel (4 - 12%, Bis-Tris) as indicated by the manufacturer (Life Technologies, Carlsbad, CA). Following gel visualisation by Coomassie blue (PageBlue, Thermo Scientific) the gel bands corresponding to the molecular weight of Hc (~75 kDa) were excised and processed by in-gel digestion [46] for liquid chromatography (LC)-MS analysis. Briefly, gel pieces were washed twice with ultrapure water and twice with 100 mM ammonium bicarbonate followed by dehydration with 200 µL of 100% acetonitrile and 5 min incubation at room temperature. Gel pieces were rehydrated with 30 µL of 10 mM dithiothreitol and incubated for 30 min to reduce proteins. Excess solution was removed with a micropipette. To prevent protein refolding thiol groups were alkylated with 30 µL of 100 mM iodoacetamide. Dithiothreitol and iodoacetamide were prepared in 100 mM ammonium bicarbonate. Gel pieces were dehydrated again using acetonitrile as indicated before. Tryptic peptides were generated by addition of 50 µL of trypsin (0.01 µg/mL in 50 mM ammonium bicarbonate, Promega) and overnight incubation at 37˚C. Tryptic peptides were recovered dehydrating the gel pieces with buffer in a two-step process (A: 50% acetonitrile/5% formic acid/water as diluent; buffer B 80% acetonitrile/5% formic acid/water as diluent) followed by lyophilisation. Peptides were resuspended in 30 µL of 0.1% formic acid.

2.3 Filter aided sample preparation (FASP)

A gel-free discovery proteomics approach was employed to detect Hc isoforms from 100 µg of total protein (Bradford) from a haemolymph pool from seven shrimps collected as described above. A modified version of the filter-assisted sample preparation (FASP) protocol [47] was employed wherein plasma proteins were homogenised in 200 µL of urea buffer (8 M

urea, 100 mM Tris-HCl, pH 8.5) and transferred into a 3 kDa filter unit (Amicon) followed by centrifugation (14,000 x *g*, 15 min). Filters were washed twice using urea buffer and centrifugation (14,000 x *g*, 15 min). Proteins were reduced with 50 mM dithiothreitol and 30 min incubation at room temperature followed by centrifugation (14,000 x *g*, 15 min). Thiol groups were alkylated in the darkness with 100 mM iodoacetamide for 20 min at room temperature and centrifuged (14,000 x *g*, 10 min). Dithiothreitol and iodoacetamide were prepared in urea buffer. Excess of dithiothreitol and iodoacetamide were washed with two sequential centrifugation steps (14,000 x *g*, 15 min) using 200 µL of urea buffer. Filters and proteins were equilibrated with 200 µL of 50 mM ammonium bicarbonate (14,000 x *g*, 10 min) before proteolysis. Proteins were digested at 37˚C overnight using 200 µL of trypsin (0.01 µg/µL in 50 mM ammonium bicarbonate). Tryptic peptides were recovered by centrifugation (14,000 x *g*, 15 min) lyophilised and resuspended in 0.1% formic acid to an estimated final concentration of ~4 µg/µL for LC-MS/MS analysis.

2.4 LC-MS/MS analysis

Tryptic peptides (2 µL) from either in-gel or in-solution digestion were chromatographically separated using an Ekspert 415 nanoLC (Eksigent, Dublin, CA, USA) coupled to a TripleTOF 6600 (SCIEX, Redwood City, CA, USA). The peptides were directed onto a trap column (SGE ProtoCol, C18, 3 µm, 120 Å, 10 mm × 0.3 mm) for desalting for 5 min at a flow rate of 10 µL/min 0.1% FA. The peptides were separated in a ChromXP C18 column (3 µm, 120 Å, 150 mm × 0.3 mm) at flow rate of 5 µL/min. A stepped linear gradient from 3 to 25% solvent B over 38 min was employed followed by 25 − 32% B over 5 min, followed by 32 - 80% B over 2 min and a 3 min hold at 80% B, transitioning to 3% B over 1 min, and 9 min of re-equilibration. The linear gradient (3 to 25% solvent B) for raw haemolymph-FASP was extended to 68 min owed to its complexity, but the rest of chromatographic steps remained the same. The eluent from the HPLC was directly coupled to the DuoSpray source of the TripleTOF 6600 MS. The mass spectrometer parameters were set to 5500 V for ion spray voltage, 25 psi for curtain gas and ion gas sources 1 and 2 were set to 15 and 15 psi respectively (12 and 15 for in-solution digests). The heated interface was set to 200°C. Data acquisition was performed in information-dependent acquisition (IDA) mode comprising a high-resolution time-of-flight (TOF)-MS survey scan followed by 30 MS/MS, each with a 40 ms accumulation time. First stage MS analysis was performed within the mass range *m/z* 400−1250 with a 0.25 s

accumulation time with the instrument in positive mode. Product ion spectra were acquired for precursor ions over 200 counts/s with charge state 2−5. These spectra were acquired over the mass range of *m/z* 100−1500 using the manufacturer rolling collision energy (CE) based on the size and charge of the precursor ion. Dynamic ion exclusion was set to exclude precursor ions after one occurrence with an 8 s interval and a mass tolerance of 50 ppm. Isotopes within 6 Da of the precursor mass were excluded. Mass spectral datasets publicly available at institutional repository with identifier ##### (To be advised).

2.5 Protein identification

The protein database was constructed using 126,369 contigs from a publicly available *P. monodon* transcriptome [48]. The public transcriptome was assembled from nine different tissues and eight early life-history stages of *P. monodon* and can be accessed on NCBI BioProject: PRJNA421400 and TSA: GGLH00000000 [48]. Contigs from the transcript assembly were processed for open reading frames in all 6 frames using the transeq tool of the EMBOSS software suite [49]. For each contig the longest open reading frame was selected for inclusion in the database. A repository of adventitious proteins database (cRAP) was appended to the main database. The protein sequences in this repository are derived from common contaminants during proteomics sample preparation.

Identification of Hc isoforms was enabled by a combined database search using the Paragon™ algorithm [50, 51] version 5.0.1.0, 4874 embedded in the ProteinPilot™ software version 5.01. A total of 24 mass spectral datasets obtained by SDS-PAGE (n = 12) and FASP (n = 12) were used as input for the combined database search. Search parameters were defined as cysteine alkylation with iodoacetamide, trypsin as the digestion agent with no restrictions placed on taxonomy. Modifications were set to “generic workup” and “biological” modification as provided with this software package, which consisted of all biological modifications listed in Unimod, including acetylation, methylation and phosphorylation. The generic workup modifications set contains 59 potential chemical modifications that may occur as a result of sample handling, for example, oxidation, dehydration and deamidation. The criteria for positive protein identification were proteins with ≥95% confidence. Hc isoform complexity was resolved using the distinct peptide summary (1% global false discovery rate) wherein each Hc tryptic peptide fragment evidence was manually inspected and deemed unique or common according to its presence in one or more Hc isoforms. A general schematic

representation of the workflow used to resolve the complexity of Hc in haemolymph of black tiger *Penaeus monodon* is presented in Figure 1.

2.6 Transcriptome analysis, Hc alignments, phylogeny and isoform expression

Transcripts assembled from a recent *P. monodon* tissue-specific and larval stage data set [48] were interrogated by BLAST sequence similarity algorithms using the *P. monodon* Hc protein sequence (AEB77775). A list of 44 potential Hc isoforms were identified using sequence alignments to other known crustacean Hc genes, and analysed according to the known Hc subunit framework [10]. Protein alignments, phylogenetic trees and pairwise comparisons were performed using MEGA v10.0.5 [52]. Maximum likelihood trees from protein alignments were created using the Nearest-Neighbour-Interchange (NNI) method and Jones-Taylor-Thronton (JTT) protein substitution models over 100 bootstrap replicates. Relative isoform expression was quantified from transcriptomic data by mapping each individual Illumina paired end read to each contig using Salmon in RStudio V1.0.143 running RV3.4.1. Read counts were normalised across each sample by sequence read depth and contig length, calculated as TPM (transcripts per million), then shown as log10 TPM relative expression. Gene expression was quantified as the average from three replicate samples of each tissue (hepatopancreas, hemocytes, muscle, male gonad, female gonad, eyestalk, lymphoid organ, gill and stomach) and from single pools of approximately 400 individuals from each early life history stage (E – embryo; N – nauplii; Z – zoea; M – mysis; PL1,4,10,15 – post larvae day 1, 4, 10 or 15).

3. Results

3.1 Protein separation and SDS-PAGE and FASP MS analysis

Separation of raw haemolymph plasma by SDS-PAGE showed a predominant protein conglomerate around the expected molecular weight of Hc (~75 kDa) (Figure 2A, lanes A-C). After haemolymph plasma ultracentrifugation (at 30, 60 or 120 min) a higher diversity of protein bands across a wider molecular weight range were visible (Figure 2A, lanes D-L). Both, raw and differentially centrifuged haemolymph were used to elucidate Hc isoforms from

227 excised Hc gel fragments or directly after FASP employing trypsin combined with shotgun
228 proteomic analysis.
229 Following LC-MS/MS of SDS-PAGE and FASP digests, a combined protein database search
230 facilitated identification of 238 non-redundant Hc tryptic peptides in shrimp raw and
231 ultracentrifuged haemolymph. From these, 45 tryptic peptides were uniquely matched to 12
232 Hc protein isoforms (Table 1). Fragmentation evidence of these 45 peptides and sequence
233 coverage for the 12 Hc isoforms are provided in Supplementary Files 1 and 2. The remaining
234 193 Hc tryptic peptides were conserved across a varying number of Hc transcriptomic
235 assemblies and therefore deemed unresolved. An example of such conservation was peptide
236 WNAIELDK that was present in 33 Hc transcriptomic assemblies (Supplementary File 3)
237 including most gamma isoforms reported in Table 1.
238 The inherent purification capacity of SDS-PAGE enabled the identification of seven unique Hc
239 tryptic peptides that were not present after FASP (Figure 2B). Consequently, PmoHc γ7 was
240 exclusively identified in SDS-PAGE through the peptides IRDAIAHGYIADR and
241 GINVLGDIIESSLYSPNVQYYGALHNTAHIVLGR that were not discoverable in FASP (Table 1,
242 Figure 2C). FASP elucidated two Hc tryptic peptides associated to isoforms PmoHc γ 1 and
243 PmoHc γ 8 (Figure 2B); however, other peptides associated to those isoforms were also
244 identified by SDS-PAGE (Table 1). The remaining 11 Hc isoforms were identified by both
245 approaches with the use of 36 peptides identified by both SDS-PAGE and FASP. Hc isoforms
246 PmoHc β and PmoHc γ1 were identified through 14 peptides while the remaining 28unique
247 peptides were distributed between the remaining 10 Hc isoforms (Table 1). Sequence
248 coverage ranged from 39-80%.

Table 1. Hemocyanin isoforms resolved in haemolymph of black tiger shrimp *Penaeus monodon* using proteomics informed by transcriptomics

	Protein							Peptide				
Name	Accession[1]	Sequence coverage[2]	Length[3]	kDa[4]	p*I*[5]	Common peptides	**Identified peptides**	Da[6]	*z*[7]	Times detected	dMass[8]	Origin[9]
PmoHc β	GGLH01016677.1	39	667	77.16	5.64	5	AGENHIVR	894.47	2	14	-0.001	
							ELATTWNPR	1086.54	2	30	-0.002	
							FDSNGMQIPFDNNR	1670.69	2	3	-0.014	
							HDDSVTVR	927.44	2	15	0.001	
							ILTELEQGR	1057.58	2	27	0.000	
							LNHKPFTFDIYANAK	1778.89	3	1	-0.005	Gel
							MLQTPGR	817.40	2	14	-0.011	
							MPPGVMEHFETATR	1633.72	3	2	0.000	Gel
							MTFTGTLK	897.49	2	1	0.026	
							QWFSLFNPR	1193.60	2	13	0.003	
							TLIQQADEAVASNTDLPHDIDR	2421.17	3	14	-0.008	
							VLGAQSDPLGK	1083.59	2	25	0.000	
							VNEPLLSSFTDLK	1461.77	2	43	-0.002	
							YGGEFPTRPDNLVFEDVAGVAR	2408.16	3	11	-0.012	
PmoHc γ1	GGLH01020339.1	57	533	61.88	5.33	19	AQQIQTPGK	969.53	2	30	0.002	
							DAHDSAVTVPDVPSFHSLFQMTEK	2673.23	4	4	-0.008	Gel
							DMIIIDSR	961.49	2	3	-0.001	
							EGTNHIVR	924.48	2	14	0.004	
							HFFSLFNAR	1137.58	2	11	0.009	
							HREEALMLFDVLIHCK	2026.97	3	2	-0.024	Gel
							IYNHGEHIHKE	1375.66	3	7	-0.007	
							LFEELDNFK	1153.57	2	20	0.008	
							LSNHLDPVEELDWNKPIHHGFAPHTTYK	3296.64	5	13	0.049	FASP
							SGLEEFVSATGLPNR	1575.79	2	27	0.004	
							SHGYPLDR	951.45	2	41	-0.002	
							TFISNAAYFR	1188.60	2	26	0.009	

							YGGQFPSRPDNVNFEDVDDVAR	2497.13	3	1	0.018	
							IRDMIIIDSR	1230.67	3	6	-0.005	
PmoHc γ2	GGLH01021097.1	69	678	77.45	5.53	49	ADSFDPEANLSHYSDGGEAIQK	2393.03	3	12	0.001	
PmoHc γ3	GGLH01022336.1	81	688	78.11	5.53		VFNHGEHIQ	1079.52	2	149	0.005	
							VFNHGEHIQK	1207.62	3	36	0.009	
PmoHc γ4	GGLH01022418.1	70	678	77.69	5.37	46	CNDWDTFVSNAAYFR	1864.78	2	11	-0.005	
PmoHc γ5	GGLH01023547.1	77	677	77.24	5.37	49	DALSGADSGLTDFESATGIPNR	2406.11	3	1	-0.035	
							TKDALSGADSGLTDFESATGIPNR	2464.15	3	9	-0.026	
PmoHc γ6	GGLH01024761.1	54	678	77.56	5.59	30	HRHEALMLFDVLIHCK	2059.02	3	4	-0.008	Gel
							VFEELPNFGHIQVK	1655.88	3	15	0.011	
PmoHc γ7	GGLH01024786.1	58	678	77.34	5.45	32	GINVLGDIIESSLYSPNVQYYGALHNTAHIVLGR	3683.88	4	3	-0.013	Gel
							IRDAIAHGYIADR	1469.80	3	1	0.022	Gel
PmoHc γ8	GGLH01024971.1	71	678	77.47	5.50	41	DSLAPYSK	879.43	2	23	-0.001	
							EHKDSLAPYSK	1273.63	3	14	-0.004	FASP
							TEAALGGADSGLTEFESATGIPNR	2420.17	3	3	0.022	
PmoHc γ9	GGLH01045235.1	80	446	50.79	5.06	43	QIDISNEK	987.49	2	4	0.000	
PmoHc γ10	GGLH01088644.1	67	183	20.42	5.31	15	AKEALGGADSGLEGFESATGIPNR	2388.15	3	1	-0.010	
							EALGGADSGLEGFESATGIPNR	2188.98	2	11	-0.048	
PmoHc γ11	GGLH01235518.1	63	578	65.90	5.14	33	QGDPHGKYDLPPGVLEHFETATR	2564.26	4	3	0.031	

1. Genbank accession numbers as part of transcriptome shotgun assembly TSA:PRJNA421400. 2. Sequence coverage in % derived from ProteinPilot calculated including common peptides. 3. Length: protein amino acid length. 4. kDa, protein theoretical molecular weight. 5. p*I*, protein theoretical isoelectric point. 6. Da, observed molecular weight of peptide in Daltons. 7. *z*, observed peptide charge state. 8. dMass, Peptide Delta mass. 9. Origin, indicates whether peptide was identified in FASP or SDS-PAGE, otherwise, peptide was identified in both approaches if space is empty.

Further MS analysis was performed on additional gel bands (Figure 2, lanes H and L, gel bands 1-4) to ensure that other Hc isoforms were not missed. Multiple protein identifications that included Hc proteins and non-Hc proteins occurred in those gel bands, but no new isoforms different from those identified in Table 1 were found (Supplementary File 4). In addition to this, the protein conglomerate corresponding to molecular weight of Hc appeared to have separated into two bands (Figure 1) which suggested the separation of Hc proteins into two groups. These bands were further characterised by SDS-PAGE and the top and bottom bands excised and digested in triplicate and analysed by MS as described above. Both bands contained the same Hc and no new isoforms were identified (Supplementary File 5A). Further MS analysis of haemolymph plasma prepared on 3, 10 and 30 kDa (Supplementary File 5B) filters as well as pelletised Hc (after 120 min ultracentrifugation) prepared on a 10 kDa filter did not yield any new Hc identifications (Supplementary File 5C).

3.2 Phylogeny and relative gene expression

Transcriptomic assemblies from *P. monodon* yielded a total of 44 contigs that were annotated or shared significant homology with the amino acid sequence of the publicly available *P. monodon* Hc (AEB77775). There was a high degree of sequence similarity between putative *PmoHc* gene isoforms, with protein alignments of all isoforms (Supplementary File 6, supplied as FASTA file) showed an average of 83% amino acid identity across isoforms (range 24.3% to 97.4%). This high degree of sequence similarity meant that proteomics detected many shared peptides across all isoforms, but only those that contained unique peptides were considered for further analysis. Phylogenetic analysis of these 12 Hc isoforms clustered mainly within the γ-subunit, but also revealed the presence of a single β-subunit that was most similar to the β-subunit of *P. vannamei* (Figure 3, sequence 41). The two Penaeid Hc β-subunits shared 92.6% protein homology across 667 amino acids. Of the 12 isoforms detected by proteomics, sequence Pmon Hc γ5 (GGLH01021772.1) was the full-length sequence most similar to the published full-length *P. monodon* Hc (Figure 3, sequence 10), with 97.8% identity across 677 amino acids (Figure 4A). As expected, isoforms that shared most homology with one another clustered together, such as Pmon Hc γ3 and γ9 (98.4% homology, Figure 3) or Pmon Hc γ4 and γ8 (96.3% homology, Figure 3). Interestingly, proteomics also confirmed the presence of two truncated isoforms, Pmon Hc γ9 and Pmon Hc γ10, (~50 and ~20 kDa, respectively) that could be uniquely identified from all other Hc isoforms.

Relative gene expression using read counts showed that the majority of uniquely identified isoforms were only expressed in the hepatopancreas (Figure 4B), but also had specific expression patterns within shrimp larval stages (Figure 4C). For example, Pmon Hc γ3 and γ9 isoforms were both highly expressed in the hepatopancreas. Meanwhile, the Pmon Hc γ6 and γ7 isoforms shared a similarly high expression pattern throughout most larval stages from zoea stage onwards (Figure 4C) In addition, Pmon Hc γ5 and γ10 were highly expressed and clustered with the only other known Hc sequences from *P. monodon*. These two isoforms also shared a similar expression pattern that increased at later PL stages (Figure 4C). On the other hand, the expression of *PmoHc γ1* or *γ6* could not be detected in adult hepatopancreas tissue, but their expression was comparatively high during larval development, particularly during late mysis and throughout PL stages.

4. Discussion

4.1 Proteomics informed by transcriptomics in novel hemocyanin isoform discovery

High throughput genomic sequencing technologies produce vast amounts of data and reveal the potential breadth of sequence diversity that exists in different species, but few studies attempt to resolve the complexity, redundancy or incorrect annotations of those sequences. In this study 44 transcriptomic putative Hc isoforms were identified and using an independent proteomics approach 12 unique Hc isoforms within the circulating haemolymph of shrimp were resolved, irrespective of the shrimp used to produce the transcriptome. This included defining the presence of a β-subunit for the first time in *P. monodon* identified through 14 unique peptides. Further confirmation of putative transcriptomic Hc isoforms may be resolved using other complementary enzymes in addition to trypsin to discriminate unique peptides, or by subjecting the haemolymph to additional purification steps before LC-MS/MS analysis. Herein, the benefit of combining transcriptomics with mass spectrometry as a powerful approach to detect and resolve isoform complexity in a non-model organism was highlighted.

Unfractionated shrimp haemolymph represents a significant technical challenge for proteomic studies due to the abundance of Hc, similar to the well-known challenges that albumin, haemoglobin and immunoglobulins create for investigating vertebrate plasma. Although ultracentrifugation has been commonly used in other non-crustacean species, only

317 recently has haemolymph been analysed from shrimp using shotgun proteomics and
318 ultracentrifugation [11, 53]. In this study ultracentrifugation and the resolving capacity of SDS-
319 PAGE were sufficient to identify all reported Hc isoforms. This could be highly relevant when
320 sample availability is a constraint as FASP, at different molecular weight cut-offs, utilised ten
321 times more protein without a yield increase in Hc isoform identification.
322
323 *4.2 Implications for shrimp biology*
324 Our research resolved a total of 11 unique γ-type subunits and a single β-type subunit. Many
325 of the sequences matched the 20-25 amino acid N-terminal protein sequences derived in the
326 study by Stoeva [9] with the exception of one or two substitutions that could be either protein
327 sequencing errors or population variants. In *P. monodon*, only a single γ-type Hc subunit has
328 been identified, despite evidence for several distinct proteins in haemolymph extracts [9].
329 Several β- and γ-type subunits have been resolved from two Penaeid shrimp species using
330 transcriptomic evidence [10]. At protein level several γ-type subunits have been resolved
331 including a β-subunit in *P. vannamei* [11]. All other reported sequences from other penaeids
332 resolved within the γ-type Hc clade (Figure 3).
333 The two peptides VLGAQSDPLGK and YGGEFPTRPDNLVFEDVAGVAR used to uniquely identify
334 the β-subunit in *P. vannamei* [11] were amongst the 14 unique peptides used to identified a
335 β-subunit in *P. monodon* in our study. The sequence conservation of these peptides in *P.*
336 *vannamei* and *P. monodon* would be particularly delicate in proteomics-based studies of
337 traceability, provenance or food contamination where neither peptide could be used for
338 species authentication or food certification [54, 55].
339 Each of the Hc sequences reported in the public databases for *P. monodon* clustered as one
340 γ-type Hc clade with the Pmon Hc γ5 and γ10 isoforms identified by proteomics (Figure 3,
341 numbers 6-8). However, sequence diversity and the resulting bootstrap scores from the
342 maximum likelihood phylogeny supported the presence of a further 9 different γ isoform. This
343 enabled absolute confirmation of the putative transcriptome Hc sequence diversity within
344 the haemolymph of the black tiger *P. monodon*. A potential small Hc (HcS) γ-subclade may be
345 emerging containing the Pmon Hc γ3 and γ9 subunits, as well as the Pvan HcS and several
346 other Penaeid species. The Pvan HcS subunit was reported as the most abundant isomer
347 within the multimeric Hc protein [11], consistent with the Pmon Hc γ3 and γ9 subunits
348 hepatopancreas expression levels. However, there was no simple relationship with size or

theoretical isoelectric point, as unique Pmon Hc γ1 and γ11 subunits were also truncated (533 and 578 amino acids, respectively) but clearly separated from any putative HcS clade.

For the 12 Hc isoforms uniquely identified by proteomics, each showed significant sequence diversity and distinctive tissue specific expression patterns (Figure 4), including several that were only present in early larval stages. The Pmon Hc γ3 was constitutively expressed across all larval stages and the only one identified in eggs and was also the most abundantly expressed isoform in the hepatopancreas of adults. Isoforms of Hc have been described between moult stage of shrimp [16] and in juvenile crabs, where certain isoforms change in relative abundance during larval development, while others remain constitutively expressed [56]. Interestingly, differential expression of Hc isoforms has been reported between shrimp subjected to different conditions and challenges including rearing sites [10], hypoxic and cold stress [14, 57], inter-moult and pre-moult [16] and bacterial and viral infection [18, 53, 58]. This shift in expression has been linked with oxygen binding affinity and haemolymph magnesium levels, potentially as an environmentally adaptive mechanism [59]. As we begin to understand the abundance of Hc diversity, the ability to accurately discriminate between these isoforms is key to understanding their biological function.

Proteomic studies employing 2D gels have identified several truncated Hc isoforms that were not only up-regulated in response to bacterial infection, but also showed *in vitro* anti-bacterial, anti-viral or agglutination activities [19, 60, 61]. Recombinantly produced versions of these truncated isoforms either reduced mortality from *Vibrio* infection after injection [19], or reduced white spot syndrome virus (WSSV) transcription *in vitro* by binding to the WSSV envelope protein VP28 [60]. Both the C- and N-terminal sequence of hemocyanin have been shown to be highly diverse and possess bacterial agglutination and hemolytic activities [20, 61]. These functional truncations may be produced from cleavage of full length Hc protein, known in insects to induce phenoloxidase activity after N-terminal Hc cleavage [62]. The sequence of the 165 aa fragment (493-697 aa) LvaHcL4 (Figure 3, sequence 2 AHY86474) isoform [19] corresponded to the coding region of the 183 aa fragment (501-683 aa) from the PmoHc γ10 isoform. This isoform showed only 90% similarity to any other PmoHc sequence, and clustered independently with other Hc isoforms including the truncated PvaHcL4 sequence (Figure 3, sequences 1-6). Most importantly, this truncated isoform could be identified from all other Hc isoforms using two unique peptide sequences (Table 1). This

strongly suggested this truncated sequence is an independent gene within penaeid shrimp, with specific expression patterns, regulation and functions in pathogen response.

Hc diversity has been suggested as a basis for mounting an innate immune response against a range of pathogens in shrimp [20, 61, 63]. Other benefits of such Hc gene duplications can be the ability to transcribe huge amounts of mRNA with a subsequent high concentration of Hc proteins stored in the haemolymph [6, 15]. These provide the potential to immediately respond to various challenging conditions, caused by either environment or disease. The existence and sequence diversity of different genes, together with their flexible expression, constitutes the genetic basis for the inter- and intra-specific polymorphism and provides an efficient intrinsic mechanism to adapt to dynamic and changing conditions [15, 64].

4.3 Medical implications

Recent studies have revealed shrimp Hc in causing anaphylaxis; the most severe form of an allergic reaction [65] but only a few reported potential epitopes derived from Hc [5, 42]. For example, five Hc peptides from lobster and shrimp bearing the QHDVN motif were reported as allergens in sera of patients with shrimp allergy [42]. In our study, the peptide **QHDVN**FLLHK (containing the same motif) was identified conserved across several of the resolved gamma isoforms reported here (Supplementary File 3). It then becomes complex to attribute allergenicity to a particular isoform or to all when the only evidence is an allergen derived from a conserved peptide. An additional study in shrimp and lobster identified the Hc peptides FNMPPGVMEHFETATR and HWFSLFNER [5] that share homology with the PmoHcβ peptides MPPGVMEHFETATR and QWFSLFNPR identified here (Table 1). Further Hc allergen peptides included QREEALMLFTVLNQCK and EEALMLFDVLMHCK [5] that are similar to the PmoHc γ1 peptide HREEALMLFDVLIHCK identified here (Table 1). The IgE antibody binding property of shrimp Hc also puts into perspective the risk of allergic cross-reactivity to other invertebrates including insects and house dust mites as Hc is also present in their haemolymph [43, 66]. In this regard, it is essential to generate knowledge on shrimp Hc isoform allergenicity to provide insights into the molecular mechanisms underpinning shellfish allergy. The purified or recombinant expressed individual Hc isoforms could be included in *in vitro* diagnostics to test serum of patients with suspected crustacean allergies and further IgE antibody epitope mapping and discovery. In this way allergic reactions to

ingested shrimp can be assessed in the context of the abundance of specific Hc isoforms or to the presence of specific conserved epitopes or a combination of both.

Hc is also known for its immunostimulatory and antibody production [26, 28, 33] as well as anti-proliferative and anti-tumour properties [30, 33, 36]. Hc from the keyhole limpet *M. crenulata* is the most advanced of the Hc analogues in biomedicine and is currently used as immunogenic adjuvant in clinical trials. However, there is a lack of knowledge on the isoform-specific immunogenicity of Hc. Resolving the complexity of Hc isoforms is key for determining their specific molecular functions and in finding new applications for this very abundant oxygen transport protein.

5. Conclusions

In this study the complexity of Hc isoforms in shrimp haemolymph was resolved revealing 12 isoforms including ten of approximately 75 kDa and two truncated forms of approximately 50 and 20 kDa. The identification and purification of Hc isoforms sets the foundation for investigations using specific isoforms to explore the potential allergenicity and thus impacting food safety in relation to shellfish consumption. Future studies will seek to explore the isoform expression in response to biotic or abiotic stresses and define markers indicative of high performing animals. These would be applied with the aim to increase aquaculture sustainability through the development of improved feeds and rearing conditions that enhance animal welfare and performance.

References

[1] T Burmester, Origin and evolution of arthropod hemocyanins and related proteins, J Comp Physiol B **172** (2002) 95-107.

[2] CJ Coates, H Decker, Immunological properties of oxygen-transport proteins: hemoglobin, hemocyanin and hemerythrin, Cell Mol Life Sci **74** (2017) 293-317.

[3] EM Costa-Paiva, CG Schrago, CJ Coates, KM Halanych, Discovery of novel hemocyanin-like genes in Metazoans, The Biological Bulletin **235** (2018) 134-151.

[4] J-C Chen, S-y Cheng, Studies on haemocyanin and haemolymph protein levels of *Penaeus japonicus* based on sex, size and moulting cycle, Comp Biochem Physiol Part B Biochem **106** (1993) 293-296.

[5] S Piboonpocanun, O Jirapongsananuruk, T Tipayanon, S Boonchoo, RE Goodman, Identification of hemocyanin as a novel non-cross-reactive allergen from the giant freshwater shrimp *Macrobrachium rosenbergii*, Mol Nutr Food Res **55** (2011) 1492-1498.

[6] H Decker, N Hellmann, E Jaenicke, B Lieb, U Meissner, J Markl, Minireview: Recent progress in hemocyanin research, Integr Comp Biol **47** (2007) 631-644.

[7] A Volbeda, WGJ Hol, Crystal structure of hexameric haemocyanin from *Panulirus interruptus* refined at 3.2 Å resolution, J Mol Biol **209** (1989) 249-279.

[8] J Markl, Evolution and function of structurally diverse subunits in the respiratory protein hemocyanin from arthropods, Biol Bull **171** (1986) 90-115.

[9] S Stoeva, K Idakieva, N Georgieva Dessislava, W Voelter, N Genov, *Penaeus monodon (tiger shrimp) Hemocyanin: subunit composition and thermostability*, in *Z Naturforsch C*. 2001. p. 416.

[10] JG Johnson, LE Burnett, KG Burnett, Uncovering Hemocyanin Subunit Heterogeneity in Penaeid Shrimp using RNA-Seq, Integrative and Comparative Biology **56** (2016) 1080-1091.

[11] J Wang, MG Janech, KG Burnett, Protein-level evidence of novel β-type hemocyanin and heterogeneous subunit usage in the Pacific whiteleg shrimp, *Litopenaeus vannamei*, Front Mar Sci **6** (2019).

[12] JX Xu, LW Ruan, Z Li, XM Yu, SD Li, H Shi, X Xu, Characterization of four hemocyanin isoforms in *Litopenaeus vannamei*, Acta Oceanol Sin **34** (2015) 36-44.

[13] S Zhao, X Lu, YL Zhang, XL Zhao, MQ Zhong, SK Li, JS Lun, Identification of a novel alternative splicing variant of hemocyanin from shrimp *Litopenaeus vannamei*, Immunol Lett **154** (2013) 1-6.

[14] JG Johnson, MR Paul, CD Kniffin, PE Anderson, LE Burnett, KG Burnett, High CO2 alters the hypoxia response of the Pacific whiteleg shrimp (Litopenaeus vannamei) transcriptome including known and novel hemocyanin isoforms, Physiological Genomics **47** (2015) 548-558.

[15] NB Terwilliger, Functional adaptations of oxygen-transport proteins, J Exp Biol **201** (1998) 1085-1098.

[16] AV Kuballa, TA Holton, B Paterson, A Elizur, Moult cycle specific differential gene expression profiling of the crab *Portunus pelagicus.*, BMC Genomics **12** (2011) 147.

[17] W Liu, D Qian, XJ Yan, Proteomic analysis of differentially expressed proteins in hemolymph of *Scylla serrata* response to white spot syndrome virus infection, Aquaculture **314** (2011) 53-57.

[18] V Chaikeeratisak, K Somboonwiwat, HC Wang, CF Lo, A Tassanakajon, Proteomic analysis of differentially expressed proteins in the lymphoid organ of *Vibrio harveyi*-infected *Penaeus monodon*, Mol Biol Rep **39** (2012) 6367-6377.

[19] Y Wen, SX Zhan, H Huang, MQ Zhong, JH Chen, CH You, F Wang, YL Zhang, Identification and characterization of an 18.4kDa antimicrobial truncation from shrimp *Litopenaeus vannamei* hemocyanin upon *Vibrio parahaemolyticus* infection, Fish Shellfish Immunol **56** (2016) 450-458.

[20] J Fan, X Li, H Lu, R Lin, JJ Aweya, Y Zhang, N-terminal diversity of *Litopenaeus vannamei* hemocyanin and immunity, Mol Immunol **112** (2019) 360-368.
[21] J Zhuang, CJ Coates, H Zhu, P Zhu, Z Wu, L Xie, Identification of candidate antimicrobial peptides derived from abalone hemocyanin, Dev Comp Immunol **49** (2015) 96-102.
[22] A Hernandez-Perez, JA Zamora-Briseno, E Ruiz-May, A Pereira-Santana, JM Elizalde-Contreras, S Pozos-Gonzalez, E Torres-Irineo, J Hernandez-Lopez, MG Gaxiola-Cortes, R Rodriguez-Canul, Proteomic profiling of the white shrimp *Litopenaeus vannamei* (Boone, 1931) hemocytes infected with white spot syndrome virus reveals the induction of allergy-related proteins, Dev Comp Immunol **91** (2019) 37-49.
[23] R Ishwarya, B Vaseeharan, A Iswarya, S Karthikeyan, Haemolytic and antibiofilm properties of haemocyanin purified from the haemolymph of Indian white shrimp Fenneropenaeus indicus, Fish Shellfish Immunol **59** (2016) 447-455.
[24] C Le Bris, B Cudennec, P Dhulster, D Drider, G Duflos, T Grard, Melanosis in *Penaeus monodon*: Involvement of the Laccase-like Activity of Hemocyanin, J Agric Food Chem **64** (2016) 663-670.
[25] CJ Coates, J Talbot, Hemocyanin-derived phenoloxidase reaction products display anti-infective properties, Dev Comp Immunol **86** (2018) 47-51.
[26] F Helling, A Shang, M Calves, S Zhang, S Ren, RK Yu, HF Oettgen, PO Livingston, GD3 vaccines for melanoma: superior immunogenicity of keyhole limpet hemocyanin conjugate vaccines, Cancer Res **54** (1994) 197-203.
[27] S Arancibia, C Espinoza, F Salazar, M Del Campo, R Tampe, T-Y Zhong, P De Ioannes, B Moltedo, J Ferreira, EC Lavelle, A Manubens, AE De Ioannes, MI Becker, A novel immunomodulatory hemocyanin from the limpet *Fissurella latimarginata* promotes potent anti-tumor activity in melanoma, PLoS One **9** (2014) e87240-e87240.
[28] J Markl, B Lieb, W Gebauer, B Altenhein, U Meissner, JR Harris, Marine tumor vaccine carriers: structure of the molluscan hemocyanins KLH and HtH, J Cancer Res Clin Oncol **127** (2001) R3-R9.
[29] O Boyanova, P Dolashka, D Toncheva, H-G Rammensee, S StevanoviĆ, In vitro effect of molluscan hemocyanins on CAL-29 and T-24 bladder cancer cell lines, Biomedical Reports **1** (2013) 235-238.
[30] V Gesheva, S Chausheva, N Mihaylova, I Manoylov, L Doumanova, K Idakieva, A Tchorbanov, Anti-cancer properties of gastropodan hemocyanins in murine model of colon carcinoma, BMC Immunol **15** (2014) 34.
[31] JR Harris, J Markl, Keyhole limpet hemocyanin (KLH): a biomedical review, Micron **30** (1999) 597-623.
[32] MA Swanson, RS Schwartz, Immunosuppressive therapy, N Engl J Med **277** (1967) 163-170.
[33] F Wimmers, N de Haas, A Scholzen, G Schreibelt, E Simonetti, MJ Eleveld, HMLM Brouwers, M Beldhuis-Valkis, I Joosten, MI de Jonge, WR Gerritsen, IJM de Vries, DA Diavatopoulos, JFM Jacobs, Monitoring of dynamic changes in Keyhole Limpet Hemocyanin (KLH)-specific B cells in KLH-vaccinated cancer patients, Sci Rep **7** (2017) 43486.
[34] J Pizarro-Bauerle, I Maldonado, E Sosoniuk-Roche, G Vallejos, MN López, F Salazar-Onfray, L Aguilar-Guzmán, C Valck, A Ferreira, MI Becker, Molluskan hemocyanins activate the classical pathway of the human complement system through natural antibodies, Front Immunol **8** (2017).
[35] J Liu, C Chen, C Ling, H Hu, J Cao, Y Gao, The effects of hemocyanin on T cells cultured in vitro, Oncol Lett **15** (2018) 2655-2660.
[36] S Liu, L Zheng, JJ Aweya, Z Zheng, M Zhong, J Chen, F Wang, Y Zhang, *Litopenaeus vannamei* hemocyanin exhibits antitumor activity in S180 mouse model in vivo, PLoS One **12** (2017) e0183783.

[37] L Zheng, X Zhao, P Zhang, C Chen, S Liu, R Huang, M Zhong, C Wei, Y Zhang, Hemocyanin from shrimp *Litopenaeus vannamei* has antiproliferative effect against HeLa cell *In Vitro*, PLoS One **11** (2016) e0151801.
[38] T Khurana, M Collison, FT Chew, JE Slater, Bla g 3: a novel allergen of German cockroach identified using cockroach-specific avian single-chain variable fragment antibody, Ann Allergy Asthma Immunol **112** (2014) 140-145.e1.
[39] S Khanaruksombat, C Srisomsap, D Chokchaichamnankit, P Punyarit, P Phiriyangkul, Identification of a novel allergen from muscle and various organs in banana shrimp (*Fenneropenaeus merguiensis*), Ann Allergy Asthma Immunol **113** (2014) 301-306.
[40] Y Zhang, L Zhu, S Li, J Zhang, T She, J Yan, Y Bian, H Li, Identification of the major allergenic epitopes of *Eriocheir sinensis* roe hemocyanin: A novel tool for food allergy diagnoses, Mol Immunol **74** (2016) 125-132.
[41] M Pascal, G Grishina, AC Yang, S Sánchez-García, J Lin, D Towle, MD Ibañez, J Sastre, HA Sampson, R Ayuso, Molecular diagnosis of shrimp allergy: efficiency of several allergens to predict clinical reactivity, Journal Allergy Clin Immunol: In Practice **3** (2015) 521-529.e10.
[42] MG Giuffrida, D Villalta, G Mistrello, S Amato, R Asero, Shrimp allergy beyond Ttopomyosin in Italy: clinical relevance of arginine kinase, sarcoplasmic calcium binding protein and hemocyanin, Eur Ann Allergy Clin Immunol **46** (2014) 172-7.
[43] SD Kamath, EB Johnston, S Iyer, PM Schaeffer, J Koplin, K Allen, AL Lopata, IgE reactivity to shrimp allergens in infants and their cross-reactivity to house dust mite, Pediatr Allergy Immunol **28** (2017) 703-707.
[44] CM Modahl, S Frietze, SP Mackessy, Transcriptome-facilitated proteomic characterization of rear-fanged snake venoms reveal abundant metalloproteinases with enhanced activity, J Proteomics **187** (2018) 223-234.
[45] H Liang, G Jiang, T Wang, J Zhang, W Liu, Z Xu, J Zhang, L Xiao, An integrated transcriptomic and proteomic analysis reveals toxin arsenal of a novel Antarctic jellyfish *Cyanea sp*, J Proteomics **208** (2019) 103483.
[46] P Jeno, T Mini, S Moes, E Hintermann, M Horst, Internal Sequences from Proteins Digested in Polyacrylamide Gels, Anal Biochem **224** (1995) 75-82.
[47] JR Wiśniewski, A Zougman, N Nagaraj, M Mann, Universal sample preparation method for proteome analysis, Nature Methods **6** (2009) 359.
[48] R Huerlimann, NM Wade, L Gordon, JD Montenegro, J Goodall, S McWilliam, M Tinning, K Siemering, E Giardina, D Donovan, MJ Sellars, JA Cowley, K Condon, GJ Coman, MS Khatkar, HW Raadsma, GE Maes, KR Zenger, DR Jerry, De novo assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome, Sci Rep **8** (2018) 13553.
[49] P Rice, I Longden, A Bleasby, EMBOSS: The European Molecular Biology Open Software Suite, Trends Genet **16** (2000) 276-277.
[50] IV Shilov, SL Seymour, AA Patel, A Loboda, WH Tang, SP Keating, CL Hunter, LM Nuwaysir, DA Schaeffer, The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra, Mol Cell Proteomics **6** (2007) 1638-1655.
[51] WH Tang, IV Shilov, SL Seymour, Nonlinear fitting method for determining local false discovery rates from decoy database searches, J Proteome Res **7** (2008) 3661-3667.
[52] S Kumar, G Stecher, K Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets, Mol Biol Evol **33** (2016) 1870-1874.
[53] MY Tao, HX Zhou, KW Luo, J Lu, YL Zhang, F Wang, Quantitative serum proteomics analyses reveal shrimp responses against WSSV infection, Dev Comp Immunol **93** (2019) 89-92.
[54] I Ortea, G O'Connor, A Maquet, Review on proteomics for food authentication, J Proteomics **147** (2016) 212-225.

[55] MF Mazzeo, RA Siciliano, Proteomics for the authentication of fish species, J Proteomics **147** (2016) 119-124.
[56] G Durstewitz, NB Terwilliger, Developmental changes in hemocyanin expression in the Dungeness crab, *Cancer magister*, J Biol Chem **272** (1997) 4347-4350.
[57] LF Fan, AL Wang, YT Miao, SA Liao, CX Ye, QC Lin, Comparative proteomic identification of the hepatopancreas response to cold stress in white shrimp, *Litopenaeus vannamei*, Aquaculture **454** (2016) 27-34.
[58] K Lei, F Li, M Zhang, H Yang, T Luo, X Xu, Difference between hemocyanin subunits from shrimp *Penaeus japonicus* in anti-WSSV defense, Dev Comp Immunol **32** (2008) 808-13.
[59] AC Brown, NB Terwilliger, Ontogeny of hemocyanin function in the dungeness crab *Cancer magister*: Hemolymph modulation of hemocyanin oxygen-binding, J Exp Biol **201** (1998) 819-826.
[60] SX Zhan, JJ Aweya, F Wang, DF Yao, MQ Zhong, JH Chen, SK Li, YL Zhang, *Litopenaeus vannamei* attenuates white spot syndrome virus replication by specific antiviral peptides generated from hemocyanin, Dev Comp Immunol **91** (2019) 50-61.
[61] Y-L Zhang, B Peng, H Li, F Yan, H-K Wu, X-L Zhao, X-M Lin, S-Y Min, Y-Y Gao, S-Y Wang, Y-Y Li, X-X Peng, C-Terminal domain of hemocyanin, a major antimicrobial protein from *Litopenaeus vannamei*: structural homology with immunoglobulins and molecular diversity, Front Immunol **8** (2017).
[62] H Decker, T Rimke, Tarantula hemocyanin shows phenoloxidase activity, Journal of Biological Chemistry **273** (1998) 25889-25892.
[63] CJ Coates, J Nairn, Diverse immune functions of hemocyanins, Dev Comp Immunol **45** (2014) 43-55.
[64] F Giomi, M Beltramini, The molecular heterogeneity of hemocyanin: Its role in the adaptive plasticity of Crustacea, Gene **398** (2007) 192-201.
[65] D Guillen, A Fiandor, V del Pozo, M Pedrosa, E Phillips-Angles, T Caballero, S Quirce, Anaphylaxis caused by hemocyanin contained in shrimp cephalothorax, Ann Allergy Asthma Immunol **113** (2014) 674-675.e2.
[66] R Ayuso, G Grishina, M Pascal, S Sanchez-Garcia, D Towle, C Smith, M Ibáñez, HA Sampson, Hemocyanin, troponin C and fatty acid-binding protein (FABP) may be cross-reactive allergens between crustaceans, cockroach and dust mites, J Allergy Clin Immunol **127** (2011) AB235.

Figures

Figure 1. Schematic representation of the workflow utilised to resolve the complexity of hemocyanin proteins in haemolymph of black tiger *Penaeus monodon* using proteomics informed by transcriptomics

Figure 2. Identification of Hc isoforms in haemolymph of shrimp *Penaeus* monodon. Panel A, Electrophoretic protein separation of ultracentrifuged haemolymph plasma. Coomassie blue stained SDS-GEL. Lanes: A-C, haemolymph plasma; D-F, 30 min ultracentrifuged haemolymph plasma; G-I, 60 min ultracentrifuged haemolymph plasma; J-L, 120 min ultracentrifuged haemolymph plasma. Large rectangle boxed area around hemocyanin molecular weight (~75 kDa) encompasses the twelve lanes that were excised for mass spectrometry analysis. Non-Hc protein were identified by MS in gel spots 1-4 on lane H and L (Supplementary File 4), further Hc identifications did not occur on these gel spots. Panel B, distribution of unique peptides identified by mass spectrometry in haemolymph of *P. monodon* using SDS-PAGE (10 µg protein) and FASP (100 µg protein). Panel C, distribution of hemocyanin isoforms identified in SDS-PAGE and FASP.

Figure 3. Resolving hemocyanin complexity in haemolymph plasma of black tiger shrimp *Penaeus monodon*. Maximum likelihood phylogeny of Hc isoforms detected as unique proteins by proteomics in this study (blue text) among other Hc isoforms reported sequence databases from other crustacean species (numbered). Numbers correspond to isoforms identified in species as follows: 1-6, 11-13, 15-18, 21-23, 41: Penaeus vannamei; 7, 19: Penaeus japonicus; 8-10: Penaeus monodon; 14: Penaeus chinensis; 20: Penaeus merguiensis; 25-27: other shrimp species; 24, 28-36, 42: spiny and clawed lobsters; 37-40: stomatopods; 43-46: crabs. Protein alignments provided in FASTA format in Supplementary File 6 and full sequence information and accession numbers provided in Supplementary File 7.

Figure 4 Pairwise comparison of unique Pmon Hc isoforms with sequences from the public database for *P. monodon* **(A)**. Relative expression and sequence diversity of unique *PmoHc* isoforms. Relative expression of each isoform is shown as: average log10 transcripts per million (TPM) ± SE within hepatopancreas tissue **(B)** and across prawn larval development **(C)**. E – embryo; N – nauplii; Z – zoea; M – mysis; PL1,4,10,15 – post larvae day 1, 4, 10 or 15.

Supporting documents

Supplementary Files 1. Peptide fragmentation evidence of unique peptides identified in specific Hc isoforms

Supplementary Files 2. Sequence coverage of Hc isoforms bearing unique peptides

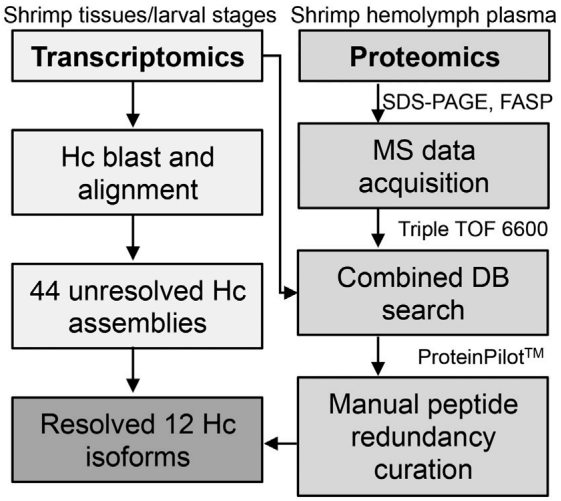
Supplementary File 3. Common Hc peptides detected by proteomics that are shared amongst transcriptomic Hc assemblies

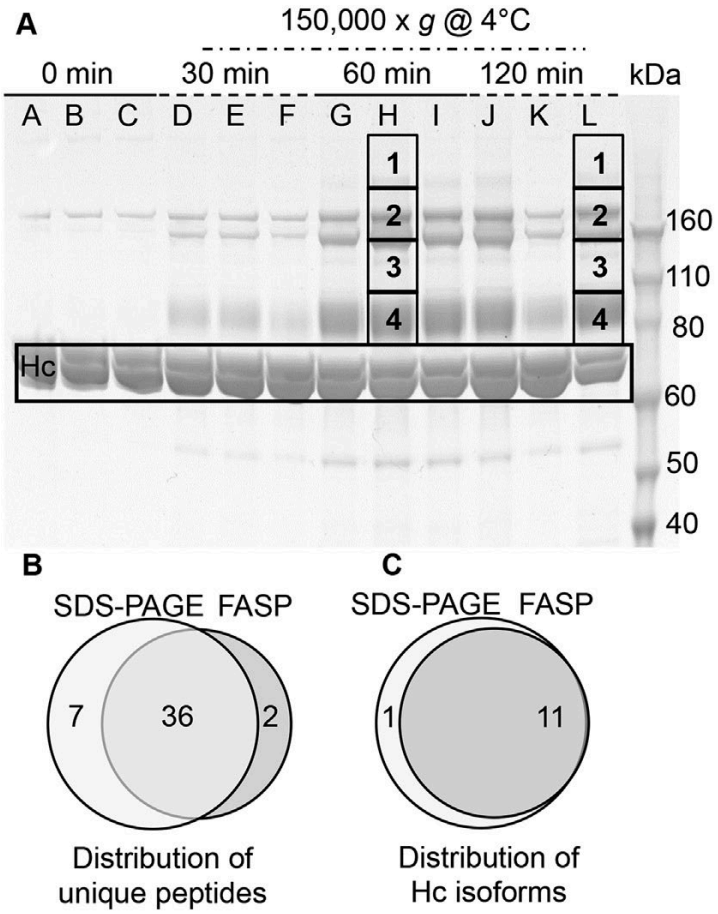
Supplementary File 4. Proteins identified in major non-hemocyanin gel spots (gel-spots 1, 2, 3 & 4, Figure 2)

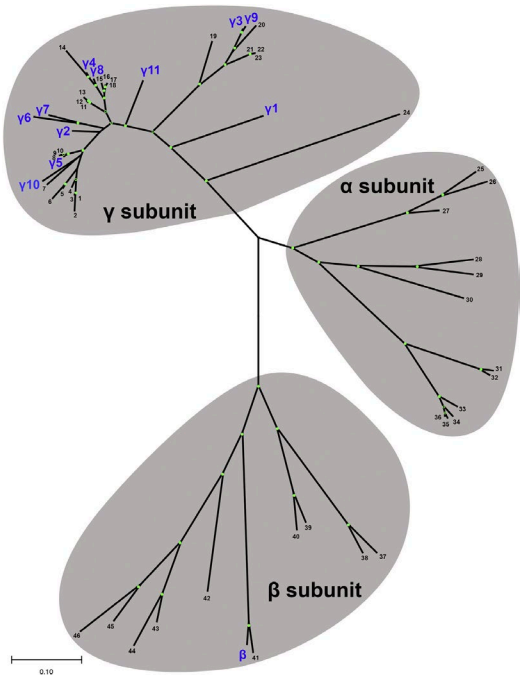
Supplementary Files 5. Information regarding MS identification of: **A)** Hc isoforms in top and bottom gel band corresponding Hc molecular weight (~75 kDa). **B)** Hc isoforms in MWCO filters of 3, 10 and 30 kDa. **C)** Hc isoforms in a pellet of hemolymph ultracentrifuged for 120 minutes at 150,000 x g at 4°C

Supplementary File 6. Fasta file for alignment of Hc isoforms.

Supplementary File 7. Hemocyanin sequences from several shrimp species used to build the phylogenetic tree in Figure 3.


Shrimp tissues/larval stages
Shrimp hemolymph plasma
Transcriptomics
Proteomics
SDS-PAGE, FASP
Hc blast and alignment
MS data acquisition
Triple TOF 6600
44 unresolved Hc assemblies
Combined DB search
ProteinPilot™
Resolved 12 Hc isoforms
Manual peptide redundancy curation


A
150,000 x g @ 4°C
0 min
30 min
60 min
120 min
kDa
A B C D E F G H I J K L
1
2
3
4
Hc
160
110
80
60
50
40
B
SDS-PAGE FASP
7
36
2
Distribution of unique peptides
C
SDS-PAGE FASP
1
11
Distribution of Hc isoforms



668

669

670

671

	Length	Pmon Hc β	Pmon Hc γ1	Pmon Hc γ2	Pmon Hc γ3	Pmon Hc γ4	Pmon Hc γ5	Pmon Hc γ6	Pmon Hc γ7	Pmon Hc γ8	Pmon Hc γ9	Pmon Hc γ10	Pmon Hc γ11	PmoHcVn AEB77775	PmoHcVn AAL27460	PmoHc AE92687
Pmon Hc β	667															
Pmon Hc γ1	533	57.0														
Pmon Hc γ2	678	55.3	63.7													
Pmon Hc γ3	663	55.2	63.8	79.9												
Pmon Hc γ4	678	55.2	63.1	92.0	77.9											
Pmon Hc γ5	677	54.9	63.4	91.9	79.3	86.9										
Pmon Hc γ6	678	53.8	63.9	86.7	76.1	87.5	86.1									
Pmon Hc γ7	678	54.1	63.6	90.1	77.1	87.6	90.1	89.2								
Pmon Hc γ8	678	54.7	63.0	92.9	78.5	96.3	87.8	86.7	88.8							
Pmon Hc γ9	446	55.8	57.8	79.4	98.4	80.5	81.4	78.3	78.9	80.3						
Pmon Hc γ10	183	41.0	18.0	91.3	77.0	84.2	90.7	79.8	80.9	83.6	74.9					
Pmon Hc γ11	578	58.1	61.1	89.1	81.5	90.7	88.1	87.2	86.9	90.3	63.1	26.6				
PmoHcVn AEB77775	684	54.8	63.0	91.2	78.7	85.8	97.4	84.9	89.0	87.0	53.1	24.3	74.7			
PmoHcVn AAL27460	386	58.8	79.3	95.6	79.3	88.3	98.2	85.8	92.7	90.4	34.7	0.0	66.6	100.0		
PmoHc AE92687	450	54.9	55.8	89.1	80.7	86.0	97.6	87.1	87.1	86.2	79.8	37.1	88.4	98.4	33.1	

672

673

674

675