Extraction, purification and structural characterisation of Hemocyanin from the invasive species *Crepidula fornicata* (Slipper limpet)

Emily Lewis

(Bsc) Biochemistry with a Professional Training Year

Dr Mark Young

School of Biosciences, Cardiff University

5793

1.1 Abstract

Since its introduction non-native invasive species *Crepidula fornicata* (slipper limpet) (Linnaeus 1758) is known for its ability to rapidly colonise coastal areas, its stacked appearance and its blue hemolymph which contains in abundance a freely dissolved, extracellular protein – hemocyanin. Hemocyanin subunits range in molecular mass from 300 kDa to 450 kDa with each subunit containing 7 to 8 oxygen binding functional units (FUs). During this study extraction, purification and structural characterisation of *Crepidula fornicata* hemocyanin has been undertaken using techniques such as SDS-PAGE, Gel Filtration Chromatography, Ammonium Sulphate Precipitation, Sequence analysis, Transmission Electron Microscopy and Cryo-EM. Results implicate the presence of two isoforms termed SLH-1 and SLH-2 which display similarities to that of KLH-1 and KLH-2 in both FU composition and overall structure. Interestingly, extraction has revealed a high purity sample in which purification does not seem essential. In conclusion, successful extraction, purification and a basis for structural characterisation have been achieved and findings have supported further research.

1.2 Introduction

Costal invasions by non-native invasive species have become increasingly common up and down the UK coasts, posing a key danger to biodiversity and ecosystems not only to the UK but worldwide (Groholz, 2002). Introduction of non-native invasive species such as *Crepidula fornicata* (slipper limpet) (Linnaeus 1758) is having detrimental effects on native species like *Ostrea edulis (European flat oyster)*(Linnaeus 1758)*(*Groholz, 2002). *C. fornicata* was first introduced within the 1880's to 1890's due to transportation of *Crassostrea virginica* (Eastern oyster) from North America to other areas of the world, more specifically, the first introduction to south-west Wales coasts was 1953 (Bohn, Richardson and Jenkins, 2015) (shown in NBN gateway at http://www.nbn.org.uk). Since its introduction *C. fornicata* is known for its ability to rapidly colonise coastal areas, its stacked appearance (**Figure 1**) and its blue hemolymph which contains in abundance a freely dissolved, extracellular protein – hemocyanin.

Figure 1. *Figure depicts a stack of Crepidula fornicata. From top to bottom slipper limpets are assumed to be from youngest to oldest, and the bottom being female and the rest male; this is due to slipper limpet's classification as sequential hermaphrodites. The stacked structure is typical of slipper limpets that are present up and down the coasts of the UK, muscular foot of one limpet is adhesively placed onto the outer shell of another*. (Hiscock, K. 2018)

Ecological Implications of *Crepidula fornicata* to coastal environments

Invasion of coastal areas within the UK is giving rise to the following three problems, i) trophic competition with other suspension feeders like native oysters, ii) spatial competition and iii) enhancement of silt and clay sedimentation within a specific habitat (Montaudouin, Audemard and Labourg, 1999). These problems are substantial and ultimately could cause a species-specific shift

consequently reducing the number of native species in a habitat. Introduction of a market for this species with possible immunological application would not only reduce the number of slipper limpets on UK coast but produce a solution to more than one problem and provide extensive opportunities for research and funding alike.

Hemocyanin – a Biomolecule of Immunological Relevance

Hemocyanin is the oxygen carrier in the blood of molluscs; where it carries out a similar function to that of haemoglobin but unlike haemoglobin it is free in the extracellular medium of the blood, not packaged into blood cells. Hemocyanin subunits range in molecular mass from 300 kDa to 450 kDa with each subunit containing 7 to 8 oxygen binding functional units (FUs). Upon oxygenation, deoxygenated Cu(I) changes to its oxygenated form Cu (II) and this is responsible for its shift in colour from a milky white solution to its distinctive blue colour (Bonaventura, Bonaventura and Sullivan, 1975). In addition to its role as an efficient oxygen transporter ongoing research has illustrated hemocyanin to stimulate and enhance immune response and more notably display anti-tumour activity. Anti-tumour activity has been reported in several different species of hemocyanin and most notably KLH as shown in Riggs et al, 2002 and this could theoretically be a possibility for slipper limpet hemocyanin and may provide the foundations for a market.

Structural Characteristics of Molluscan Hemocyanin

All hemocyanins share the same fundamental characteristics, they are symmetrical polymers of subunits with multiple oxygen-carrying functional groups containing dicuporic centres within the subunit structures. However, there are slight variations in their 3D-structure between different organisms especially between arthropod and mollusc species. As *Crepidula fornicata* is classified as mollusc, this section will focus on the characteristics of molluscan hemocyanins such as *Megathura crenulata* (Keyhole limpet abbreviated to KLH).

There are three structural fundamentals that are accepted and expected of a hemocyanin protein: i) it obtains a high molecular mass ranging from 330kDa to 450kDa, ii) in its monomeric form it is composed of 10 subunits giving an overall decamer structure (**Figure 2d**), iii) higher order assemblies are often favoured due to face-to-face interaction of monomers of hemocyanin resulting in didecamers, tridecamers and elongated, multimeric structures (Gatsogiannis and Markl, 2009). It is the structural architecture within the subunit structures that ultimately provides its function of being able to uptake oxygen, transport it, and release it. For each of the ten subunit structures that make up the decamer there are seven to eight globular functional units (FU) noted FU-a (present at the N-terminus) to FU-h (present at the C-terminus) (**Figure 2b and c**). Each FU contains a type 3 copper centre that can reversibly bind oxygen (**Figure 2a**) (Markl, 2013). Type 3 copper centres consist of two closely spaced binuclear copper ions CuA and CuB. Binuclear copper ions are directly coordinated by three conserved histidine residues, it's this spatial arrangement that facilitates its primary function as an oxygen transporter and causes a distinctive spectroscopic absorption at 340 nm (Bacci., et al. 1983).
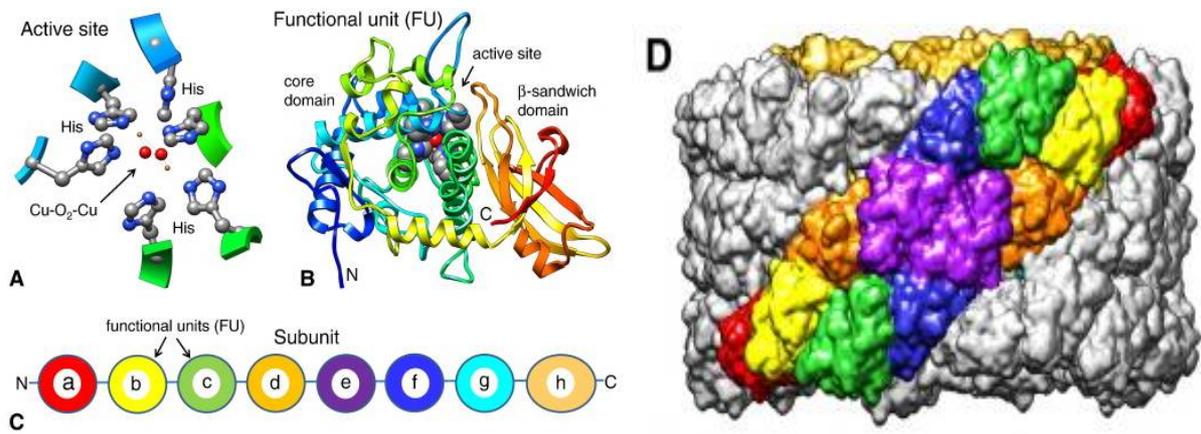
Figure 2. *Figure shows varying composition of images that contribute to a functional unit (FU) general structure, active site and molecular organisation in respect to other functional units. (A) Visual representation of the binuclear type 3 copper protein active site with two copper ions (orange), six closely coordinated histidine residues, precisely 3 to each singular copper atom. Histidine residues are directly bound to the dioxygen molecules (red spherical shaped atoms). (B) Standard molecular structure of FU-a to FU-f. The N-terminal core domain is primarily α-helical and occupies the active site (copper atoms displayed as red spherical shapes). The C-terminal is predominantly constructed of six-stranded β-sandwich (yellow, orange, red) this is used to shield the active site of hemocyanin. (C) Schematic image of the organisation of FU in respect to each other usually consisting of eight different functional units – this is typical of gastropod hemocyanin. FU-a to FU-f (red, yellow, green, orange, purple and dark blue) construct the wall of hemocyanin and FU-g and FU-h (cyan and peach) the collar D) General gastropod hemocyanin decamer shape created from a 7 A resolution of KLH-1, image is a side view of a decamer with FU's running antiparallel to each other. FU-a to FU-f (red, yellow, green, orange, purple and dark blue) construct the wall of hemocyanin and FU-g and FU-h (cyan and golden) (Markl, 2013).*

Fundamental Purification Methods of Hemocyanin

There are two main strategies of purification for hemocyanin across literature and these are Gel filtration chromatography which is relatively expensive process and ammonium sulphate precipitation which conversely is relatively inexpensive process and both techniques were used during this study.

Gel filtration chromatography is an interesting and well renowned technique which essentially separates proteins depending on molecular weight. The composition of a column is responsible for the unique elution pattern of gel filtration (**Figure 3**). Tightly packed aqueous beads are present within the columns structure and when a sample in the mobile phase is added bigger proteins stay within the mobile phase of the column and are eluted first, whereas smaller proteins will integrate into the mobile phases of the aqueous beads thus meaning they are eluted last (Hagel, 2001). This technique has been implemented into the purification of hemocyanin. Hemocyanin is generally in high abundance within the hemolymph and surrounded by much smaller salts and proteins. Drastic differences in molecular weight between hemocyanin and contaminants mean that size exclusion can efficiently purify hemocyanin in a short space of time as it is usually eluted first.
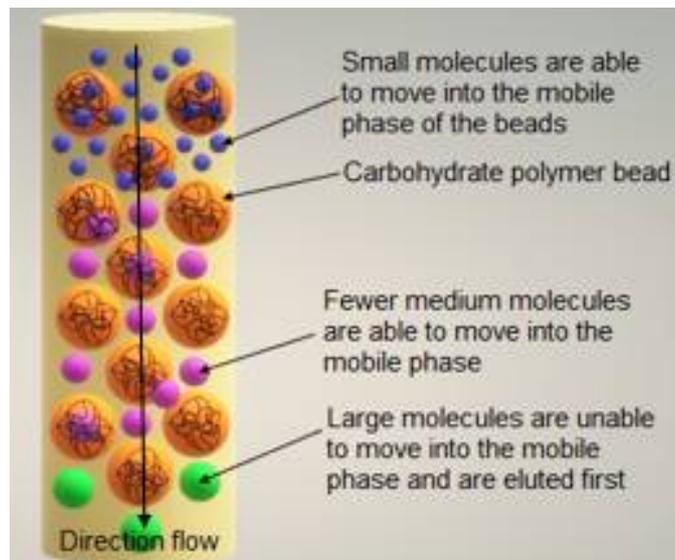
Figure 3. *Diagram illustrates inside of a Gel Filtration column and separation between different molecules of varying sizes and was created by 3D paint. Purple molecules represent "smaller" molecules with low molecular weight, this allows them to diffuse into the mobile phase of the carbohydrate porous, polymer beads. Pink molecules represent "medium" molecules with average molecular weight, some diffusion can still occur between the mobile phase of the polymer beads. Green molecules represent "large molecules", these are unable to diffuse into the mobile phase of the beads and are eluted first from the column which will be represented by an initial peak on a chromograph.*

Another form of purification method is precipitation by salt such as ammonium sulphate. The addition of any salt can shift the equilibrium of solubility of a protein by either increasing or decreasing its solubility. In terms of solubility of globular proteins, the addition of salt (<0.15 M) increases the solubility of the protein and this is termed as "salting in" (Wingfield, 1998). The converse process, "salting out", uses increasing salt concentration to increase both surface tension and the hydrophobic interactions between protein and water. This favour both self-association and association with salt, leading to the precipitation of the protein (Wingfield 1998). This method is an alternative to gel filtration chromatography as it is inexpensive and has been used to purify hemocyanin at 38% saturation of ammonium sulphate. (Idakieva et al., 2009)

Aims

The aim of this project is to successfully extract, purify and characterise the structure of hemocyanin from *Crepidula fornicata* using methodologies adapted from those used previously for other molluscan hemocyanin and with successful purification and basic characterisation this product could provide a solution to an ecological pest and hypothetically exhibit immunological properties similarly to KLH and provide a basis for further extensive research.

Key results, conclusions and outcomes

During this study successful extraction, purification at 38% ammonium sulphate saturation and basic characterisation has been successfully reported. Basic insight into the structural architecture of *C. fornicata* hemocyanin (SLH) has revealed similar characteristics to that of KLH. There are also unique characteristics individual to SLH that have seemingly not been reported in other marine species. SLH has potentially two isoforms, SLH-1 and SLH-2, similarly to most molluscan species such as keyhole limpet which also generally have two noted, KLH-1 and KLH-2. These results have potential significance structurally and immunologically and success of this study has formed the basis of three future projects, outside of the scope of this project, that will directly follow on from the

work done within this report: "*Treatment of Bone Malignancies – Proof of Principle Study*" "*Assessing the anticancer properties of Biocyanin SLH*" and an Innovate UK grant to "*Develop a scalable process to isolate pure, stable SLH*".

1.3 Materials and Methods

Extraction of hemolymph from the marine mollusc *Crepidula fornicata*

Living *C. fornicata* were collected from the shores of Pembroke Dock and stored within seawater. To collect hemolymph, a direct incision was made into the muscular foot of the slipper limpet and all liquid was collected by draining under gravity. Following collection hemolymph was centrifuged at 5000 *x g* for 30 minutes to ensure the removal of hemocytes and other contaminants within the solution. Supernatant was placed into sterile eppendorfs, kept on ice, and transported to Cardiff University the same day.

Isolation of *Crepidula fornicata* hemocyanin for resuspension

Hemocyanin from *C. fornicata* was isolated by precipitation method at 38% saturation with crystalline ammonium sulphate as described in Wingfield, 1998. In brief, a 38% (w/v) saturation was achieved by adding ammonium sulphate powder directly to hemolymph and incubating at 4°C for 12 hours. Samples were centrifuged at 10,000 x g for 30 minutes at room temperature and the hemocyanin-containing pellet was resuspended in 50 mM Tris-HCl, 150 mM NaCl, (pH 7.4). Resuspended hemocyanin placed into Slide-A-Lyze™ G2 Dialysis Cassette 20 kDa molecular weight cut off point (MWCO) 1 mL and 30 mL (**Thermo Scientific, 87737 and 87735**) to allow for dialysis against 50 mM Tris-HCl, 150 mM NaCl (pH 7.4) overnight to remove excess ammonium sulphate. After initial dialysis further dialysis against Stabilising Buffer (50 mM Tris-HCl, 150 mM NaCl, 5 mM $MgCl_2$ and 5 mM $CaCl_2$, pH 7.4) occurred overnight and after duration of time purified hemocyanin was removed from cassette and placed into sterile Eppendorfs.

Solution was clarified by centrifugation at 3,000 x g for 20 minutes at 4°C. Supernatants containing hemocyanin in a concentration range from 10-80 mg/mL (as determined by Bio-Rad Protein Assay according to manufacturers' instructions) were then stored at -20°C until further use.

Gel Filtration Chromatography

Gel filtration chromatography of intact SLH solutions was performed on an Akta Pure 25 instrument equipped with Superose® 6 10/300 GL column. Smaller V8 proteolytic fragments of SLH were separated using a Superdex 200 Increase 10/300 GL column, and all proteins were eluted in 50 mM Tris-HCl, 150 mM NaCl, 5 mM $MgCl_2$ and 5 mM $CaCl_2$ (1 mL fraction at a flow rate of 0.75 mL/min). Both protein (280 nm) and type 3 copper centre (350 nm) absorbance was monitored.

Bradford Assay

Levels of protein expression were measured using 1x Bradford Reagent and following Bradford, 1976 protocol. Bovine serum albumin (BSA) standards were created to produce a standard curve with the following concentrations, 0.5, 0.25, 0.125, 0.075, 0.05, 0.025 and 0 (mg/mL) from a 1 (mg/mL) stock solution. Additionally, dilutions were produced using either ultra-pure water or Stabilising Buffer (150mM NaCl, 50Mm Tris/HCl, 5mM MgCl2, 5mM CaCl2, pH 7.4) and absorbance values were obtained at a wavelength of 595nm using a CLARIOstar Microplate Reader.

SDS gel electrophoresis

SDS-PAGE preformed using protocol of Walker, 2002, 4-15% Tris/Glycine Gradient Gels used and Precision Plus Marker (**Biorad #1610374**) gave ten bands ranging from 10 kDa to 250 kDa.

Sequence Analysis

Sequence assembly was performed by Prof Pete Kille using previously deposited slipper limpet transcriptome data. Bioproject PRNJNA249058 (*Comparative population genomics in 76 metazoan species (popphyl project) containing 8 SRA samples, SRR1324873, SRR1324874, SRR1324875, SRR1324876, SRR1324877, SRR1324878, SRR1324879 and SRR1324880*) provided sequence data sufficient for analysis.

Transmission electron microscopy (TEM)

Purified hemocyanin samples (1mg/mL to 0.1mg/mL) were adsorbed on to 400 mesh carbon-coated copper grids (**Agar Scientific**). Grids were glow discharged for 1 minute before sample application. Excess liquid was blotted away using Whatman No. 3 filter paper, and grids were washed twice with ultrapure water. Excess water was blotted away, and samples were stained for approximately one minute with 2% (w/v) uranyl acetate. After the one-minute period excess uranyl acetate was blotted away and grids were stored at room temperature. Grids with protein sample were viewed under a JEOL JEM-2100 transmission electron microscope operating at 200 kV.

Cryo-EM

Purified hemocyanin (6 µl) was applied to glow-discharged (20 seconds) gold Quantifoil grids (R2/2), blotted on both sides for 4 seconds and plunge-frozen into liquid ethane using an FEI Vitrobot. Grids were mounted in a Gatan 914 holder for imaging on the JEM-2100 microscope.

Partial cleavage of SLH by V8 protease from *Staphylococcus aureus*

As described within Gebauer and Harris, 1999**,** before partial digestion, aliquots of SLH were dialysed against a dissociation buffer made up of 130mM glycine/NaOH (pH 9.6), aliquots of SLH were in a range of concentration between 10-15 mg/mL. *Staphylococcus aureus (S. aureus)* (Rosenbach 1884) V8 protease type XVII (**Sigma-Aldrich P6181; dissolved within NH$_4$HCO$_3$ buffer**). 2% (w/w) of V8 protease was mixed with SLH and incubated at 37°C for a 5-hour period. V8 protease was additionally added every hour for the 5 hours. Treatment was terminated by freezing samples at -20°C.

1.4 Results

Basic Investigation into Hemocyanin and Purification Method

Following extraction the first initial experiments focused on basic analytical techniques just to display the protein in its natural state without any purification process. As displayed within **Figure 4,** from a concentrated hemolymph sample 5µg of protein was loaded onto a 4-15% Tris/Gylcine Gel. There are two apparent bands sitting just above the 250 kDa marker noted as "higher band" and "lower band" with no other banding patterns present. From analysing the SDS-PAGE estimated molecular weights were deduced for both these bands with the higher band giving molecular weight of 274 kDa (n=3) and the lower band 254 kDa (n=3).
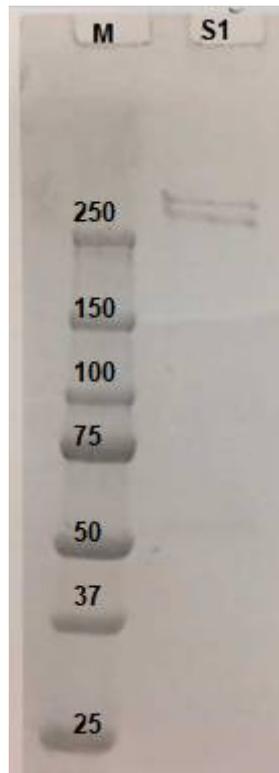
Figure 4. *4-15% Tris/Glycine Gradient Gel showing Sample 1 (S1). S1 contains 5µg of freshly extracted hemolymph which has been exposed to 10,000 xg for 30 minutes to remove any cellular debris. Two distinct high range bands estimated to be of 274.04 kDa and 254.29 kDa (n=3) as calculated by equation of the line y = -1.7384x +2.658, where y = molecular weight (log10 kDa) and x = Rf value, all raw data is displayed within* **APPENDIX 2** *and* **APPENDIX 3**.

With the same stock of sample 500µl of hemolymph which has not been purified was placed onto a Superose® 6 10/300 GL Gel Filtration column and eluted in 1mL fractions within Stabilising Buffer (150mM NaCl, 50mM Tris/HCl, 5mM MgCl2 and 5mM CaCl2, pH 7.4) at a flow rate of 1.5 mL/min, chromograph displayed in **Figure 5**. As illustrated by **Figure 5**, two measurements were taken at 280 nm a protein signal (blue) and 340 nm a signal typical of type 3 copper centres present within hemocyanin (red). There is a distinctive single peak created by both signals at the same retention time at 9.29 mLs with an estimated molecular weight of 72795.98 kDa with no other distinctive peaks. This peak is representative of an aggregated form of slipper limpet hemocyanin.
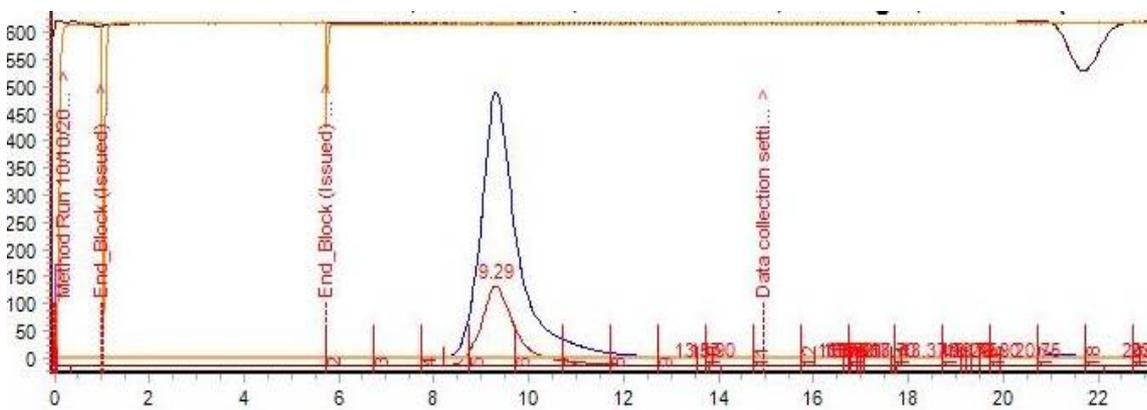
Figure 5. *500µl of Sample 1 placed onto Gel filtration (Superose® 6 10/300 GL) trace showing 280nm protein signal (blue) and 340nm signal typical of a type 3 copper protein (red). Single peak indicates high purity. Estimated weight is 72795.98 kDa, molecular weight calculated using calibration graph in* **APPENDIX 4**.

From **Figure 4** and **5** the hemolymph is displayed as relatively pure with no other outstanding protein contaminants, hence ammonium sulphate "purification" was utilised to remove hemocyanin from the hemolymph and resuspend into a buffer of choice. **Figure 6** shows successful removal of hemocyanin from a 10-15 mg/mL sample of Fresh Sample 1 (FS1) by ammonium sulphate purification as the exact same banding pattern is present in both Fresh Sample 1 (FS1) and Purified Sample 1 (PS1), this result occurred independently three times. A band of is 278 kDa is displayed in FS1 and a band of 279 kDa in PS1, molecular weights are of similar values to the molecular weight calculated in **Figure 4** therefore showing that hemocyanin can be successfully removed by 38% ammonium sulphate precipitation method without the loss of product.

Figure 6. *4-15% Tris/Glycine Gradient Gel showing Fresh Sample 1 (FS1) and a Purified Sample 1 (PS1). FS1 sample was subject to 38% ammonium sulphate "purification". Gel displays the same band at relatively the same intensities. It should be noted that the size of the band is probably due to unclear separation of the two usual bands that are displayed in* **Figure 4**. *Average molecular weight of band present in FS1 lane is 278 kDa (n=3) and 279 kDa (n=3) in PS1 as calculated by equation of the line y = -1.714x + 2.6145, where y = molecular weight (log10 kDa) and x = Rf value, repeated data for molecular weight estimation is displayed in* **APPENDIX 5** *with an example of a calibration graph used to obtain equation of the line displayed in* **APPENDIX 2**.

Partial Cleavage by V8 Protease from *Staphylococcus aureus* – Revealing the Subunit Architecture of Hemocyanin

With successful extraction, purification and analysis by SDS-PAGE analysis and Gel Filtration Chromatography, V8 Protease from *S. aureus was used to partially cleave SLH and reveal the subunit architecture of its structure.* V8 Protease from *S. aureus* (**Sigma-Aldrich P6181**) is an enzyme of 29 kDa and is classified as a serine Endoproteinase that can specifically hydrolyse peptide bonds at the C-terminus of glutamate and aspartate residues. Partial cleavage revealing specific FU's has been demonstrated of keyhole limpet hemocyanin (Gebauer, W. and Harris, J. 1999) and this has also been demonstrated in slipper limpet hemocyanin **Figure 7.**

Successful partial cleavage is displayed in **Figure 7** revealing a series of bands. 10-15 mg/mL samples were exposed to V8 protease for 5 hours at 37 °C. The presence of higher molecular weight bands between 250-75 kDa implicate intact subunit structures of hemocyanin and are not of interest. However, three distinctive low molecular bands are shown and were reproducible over

three independent experiments. Bands have the following average molecular weight (kDa) - Band 1 (yellow) has an estimated molecular weight of 63.38 kDa, Band 2 (blue) has an estimated molecular weight of 46.25 kDa and Band 3 (purple) has an estimated molecular weight of 24.34 kDa these are believed to be FU-h, FU-g and protease enzyme respectively.

Figure 7. *4-15% Tris/Glycine Gradient Gel. 10-15 mg/mL purified hemocyanin cleaved with V8 protease for 5 hours at 37 ℃ a clear "fingerprint" of cleavage is displayed. FU-g banding pattern (blue) average MW of 46.25 kDa (n=3). FU-h banding pattern (yellow) average MW of 63.38 kDa (n=3). Purple banding pattern is representative of enzyme used with an average MW = 24.34 kDa (n=3), molecular weight calculation and repeated data shown in **APPENDIX 6**.*

The cleaved sample shown in **Figure 7** was further analysed by Superdex 200 Increase 10/300 GL Gel Filtration column and revealed a similar array of cleavage fragments. There are 8 distinctive peaks observed over three independent experiments and measured at specific wavelengths - 280 nm protein signal (blue) and 340 nm typical type 3 copper protein signal (red) as shown in **Figure 8**. From Peak 1 to Peak 8 the respective estimated molecular weights are of the following, Peak 1 = 1211.87 kDa , Peak 2 = 873.56 kDa , Peak 3 = 487.02 kDa , Peak 4 = 282.39 , Peak 5 = 178 kDa and Peak 6 = 37.91 kDa. During this experiment a further two more peaks were displayed but were not reproduced in further repeats, Peak 7 = 11.03 kDa and Peak 8 = 1.96 kDa. Protein fragments were eluted in Stabilising Buffer at 0.75 m/min and were further analysed by SDS-PAGE.
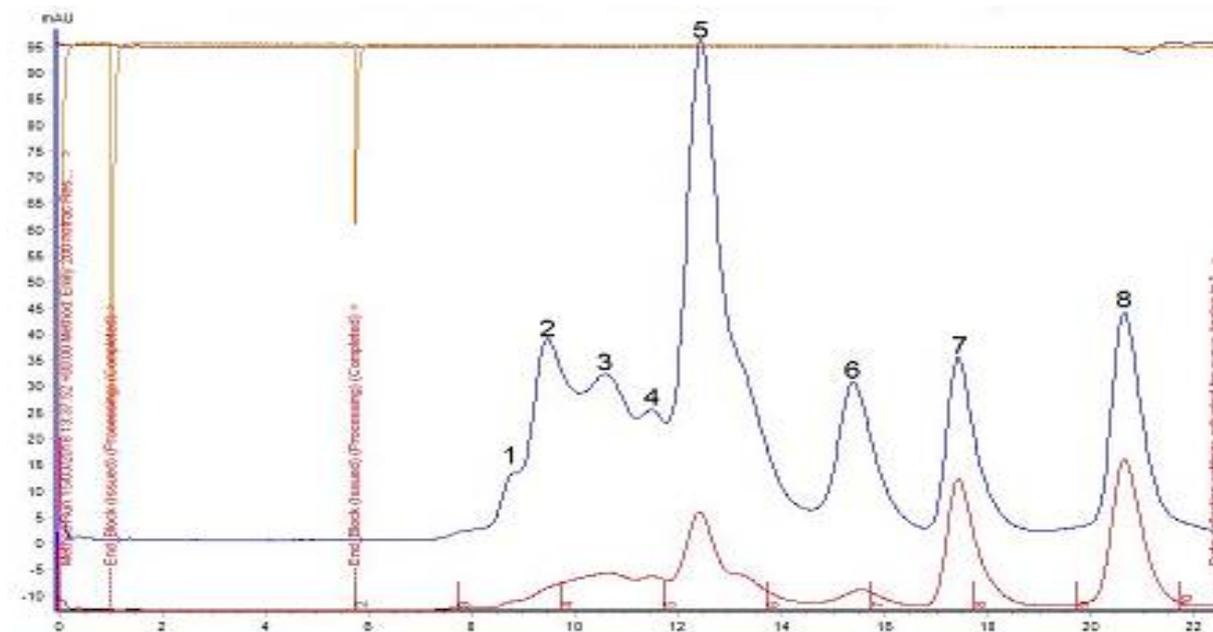
Figure 8. *100μl of 10-15mg/mL purified hemocyanin cleaved with V8 protease for 5 hours at 37 ℃ placed onto Gel Filtration (Superdex 200 Increase 10/300 GL). Trace showing 280nm protein signal (blue) and 340 nm signal typical of type 3 copper protein (red). Protein eluted in Stabilising Buffer in 2mL fractions. Peak 1 = 1211.87 kDa (n=3), Peak 2 = 873.56 kDa (n=3), Peak 3 = 487.02 kDa (n=3), Peak 4 = 282.39 (n=3), Peak 5 = 178 kDa (n=3), Peak 6 = 37.91 kDa (n=3), Peak 7 = 11.03 kDa (n=1) and Peak 8 = 1.96 kDa (n=1). Estimated molecular weight using equation of the line y= -0.2648x + 8.3795 created by Calibration Graph in* **APPENDIX 7**. *Data calculations and repeated data are illustrated in* **APPENDIX 8**.

Peaks 1 to 8 displayed in **Figure 8** that corresponded to specific fractions were taken (Fraction 3 to Fraction 9). Analysis of F3 to F6 revealed similar banding patterns in the three separate experiments but banding patterns illustrated in F7 And F8 were not reproducible. F3 and F4 contained higher order molecular weight assemblies of hemocyanin these are clearly intact / have occurred little to no cleavage and are not relevant to this study, so no estimated molecular weight was calculated. F5 contains three bands, higher, middle, and lower band of 140.61, 51.81 and 23.04 kDa (green). F6 contains a higher and lower band of 97.82 and 43.18 kDa (purple) – lower band previously seen on **Figure 7** and deduced FU-g. F8 and F7 contain a band with an average MW of 15.87 kDa (yellow). The general intensities of the fraction banding patterns are less than 1mg/mL BSA banding pattern which would suggest that concentration of protein in each fraction is less than 1mg/mL.
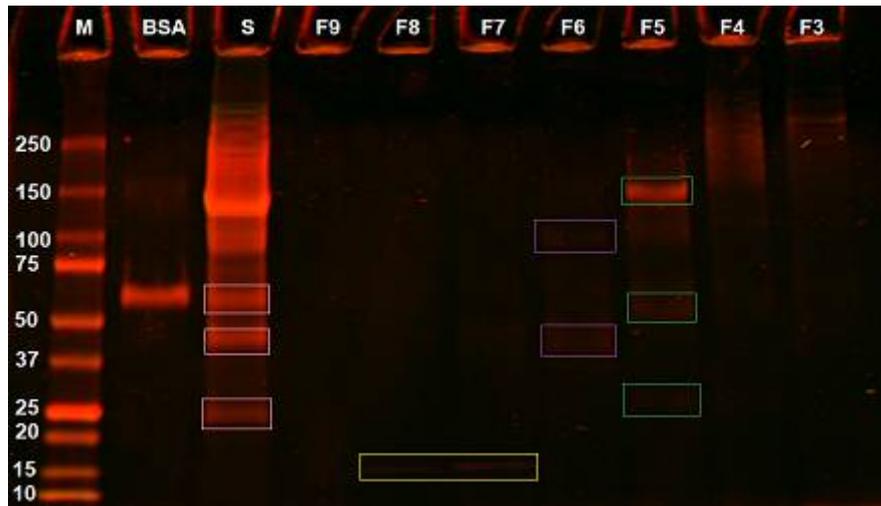
Figure 9. *4-15% Tris-Glycine Gradient Gel. Gel displays eluted fractions obtained from Gel Filtration in **Figure 7**. From left to right Marker (M), Sample cleaved with V8 Protease (S), Fraction 9 (F9), Fraction 8 (F8), Fraction 7 (F7), Fraction 6 (F6), Fraction 5 (F5), Fraction 4 (F4), Fraction 3 (F3). All fractions were eluted in 2mL Stabilising Buffer (150mM NaCl, 50mM Tris/HCl, 5mM MgCl2 and 5mM CaCl2 pH 7.4) from Superdex 200 at a flow rate of 0.75 mL/min. F8 and F7 contain a band with an average MW of 15.87 kDa (n=1, yellow). F6 contains a higher and lower band of 97.82 and 43.18 kDa (n=3, purple). F5 contains three bands, higher, middle, and lowers band of 140.61, 51.81 and 23.04 kDa (n=3, green). F4 and F3 contain higher order assemblies that have remained intact and are not of importance. All raw numerical data is displayed in **APPENDIX 9**.*

Sequence, EM and Cryo-EM Analysation of Hemocyanin – Insight into Structural Characteristics

During analysis of previously deposited sequence data (Bioproject PRNJNA249058) Dr Peter Kille constructed three sequences of hemocyanin, i) Sequence 1 – a complete sequence (SEQ_1), ii) Sequence 2 – lacked C-terminus and is therefore incomplete (SEQ_2) iii) Sequence 3 – unable to be reconstructed due to lack of alike sequence data (SEQ_3). Complete sequence 1 (SEQ_1) was aligned against KLH-1 (pdb4BED), KLH-1 and KLH-2 showed highly conserved residues at the N-terminus of individual FU's between all sequences. The conserved residues are of the following; valine, arginine and lysine (**VRK**), isoleucine, arginine and lysine (**IRK**), isoleucine and arginine (**IR**) followed by leucine (**L**) residue roughly four amino acids after these amino acids and this corresponds to previous data published by Gebauer, et al. 1999 (**APPENDIX 10**). This pattern of conserved residues at the start of each functional unit was utilised to propose sequences for FUs for SEQ_1 of slipper limpet data (APPENDIX). This revealed 8 functional units which is consistent with data collected during this report from SDS-PAGE analysis and V8 protease treatment due to the presence of FU-h, SEQ_1 is assumed to be SLH-1. To check accuracy of this method of proposal KLH-1 (PDB4BED) functional unit sequence data was extracted using the same technique and sequences were mapped back onto the 9A cryo-EM Structure and Molecular Model of KLH1 didecamer (pdb4BED) to test the fit of proposed functional unit data (**Figure 10**). There is a clear fit between FUs and correct composition in the wall suggests that the proposed data is of high accuracy and hence sequence data proposed for SLH-1 FUs are likely to be accurate.
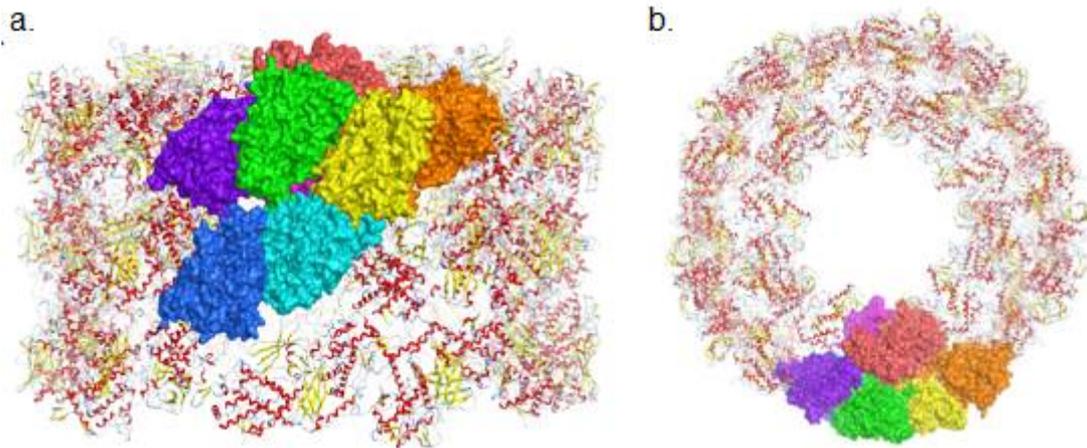
Figure 10. *Image depicts 9A cryoEM structure and molecular model of KLH1 didecamer (pdb4BED). Using N-terminal sequences functional units were proposed and mapped back onto the structure. Two images show (a) sideview of KLH1 and (b) collar view of KLH1. Functional units are represented in the following colours, FU-a (orange), FU-b (yellow), FU-c (green), FU-d (light blue), FU-e (dark blue), FU-f (purple), FU-g (fuchsia) and FU-h (salmon). Sequence alignment and conserved residue data used to create this image are present in **APPENDIX 10**.*

Electron Microscopy images reveal similar structure to that of other molluscan species. An array of decamer structures are seen ranging from decamer, didecamer to multimeric structures (**Figure 10a** and **Figure 10b**) and in particular within Figure10c, FU-g is seen due to the different contrast of intensities that reveal a arch-like structure typical of that of FU-g conformation.
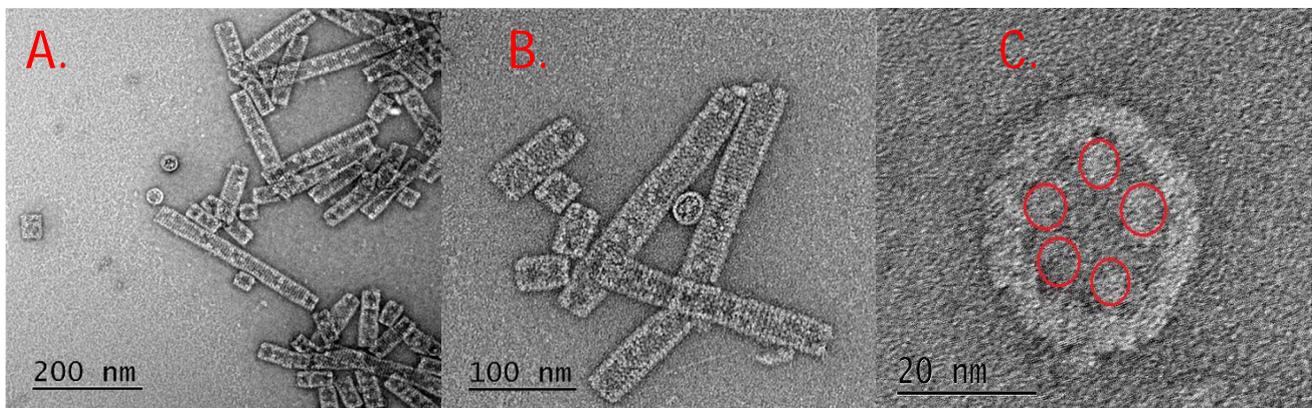
Figure 10. *Electron Microscopy of hemocyanin. All three images show an array of decamer structures of hemocyanin ranging from decamer, didecamers, tridecamers to multimeric. Stripe like intensities across the structure of hemocyanin representative of FU's running anti-parallel to each other creating the hollow-cylindrical like structure. A) B) C) 20 nm image of a top view orientation of hemocyanin clear outer and inner collar and five red circles map slight differences in intensities which are consistent with possible FU-g displayed at the face of hemocyanin. Average lengths of decamer structures were taken across the various EM images and a diameter and length of approximately 33 cm and 65 cm.*

3D reconstruction shows a hollow cylindrical like structure of didecamer of SLH with applied D5 point symmetry**.** Point group symmetry represents various points in which the structure of the protein doesn't change and looks the same. There are five symmetry elements which are

considered at each point, identity, point of symmetry inversion, line of symmetry rotational axis, plane of symmetry reflection, rotation-reflection "improper rotation". When the collective of symmetry elements are applied to a protein and it looks the same this is known as a point group. The "D" refers to the shape which is a dihedral group. Lack of intermediate data meant poor reconstruction of wall to collar, but this can be overcome by creating a 3D model from Cryo-EM data and successful sample preparation has been achieved of the same sample in Cryo-EM has displayed in **Figure 12abc**.
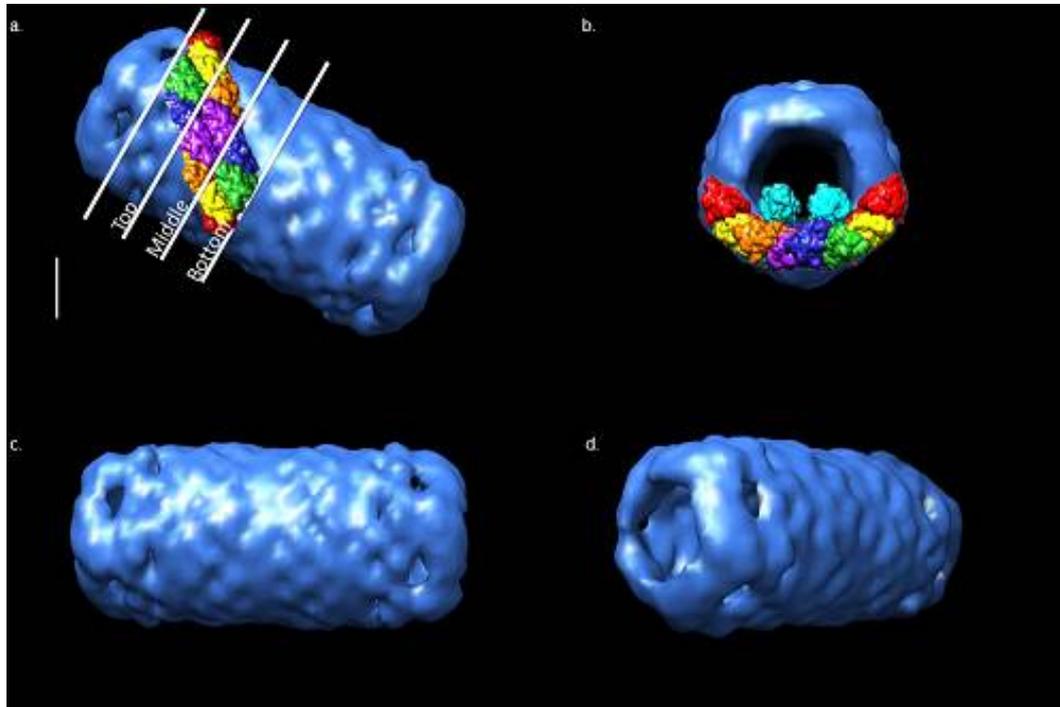
Figure 11. *3D Reconstruction of didecamer structures obtained from Electron Microscopy images from **Figure 10** using EMAN 2.2 to classify particles and reconstruct image with D5 symmetry. A low resolution of 41.2 Å was obtained (resolution data, class averages and stack projections displayed in **APPENDIX 12**) due to lack of intermediate data. **A**) Oblique image of didecamer 60.5 nm in height and 31.8 nm in width obtained by the scale bar which is 100 Å = 10 nm. As decamer should be of similar size half of this structure was mapped out to theorise structure. FU-a to FU-f form the decamer wall and run antiparallel to each other, each FU falls into a specific tier as displayed. Top tier incorporates FU-a (red), FU-b (yellow), FU-c (green) and FU-f (dark blue). Middle tier incorporates FU dimers constructed of FU-d (orange) and FU-e (purple). Lower tier incorporates FU-a (red), FU-b (yellow), FU-c (green) and FU-f (dark blue). Functional units were taken from 7 Å resolution atomistic model of KLH1 (Markl, 2013). **B**) Front view of "collar" of didecamer with a height and width of 343 nm. Functional units taken from 7 Å resolution atomistic model of KLH1 (Markl, 2013) and placed onto the collar once again antiparallel structure is displayed between a series of FU's. FU-a (red), FU-b (yellow), FU-d (orange), FU- f (dark blue), FU-c (green) and FU-g (cyan). Should be noted that the FU-g or FU-h can be present within the collar not just FU-g. **C** and D are further side views of reconstructed protein.*
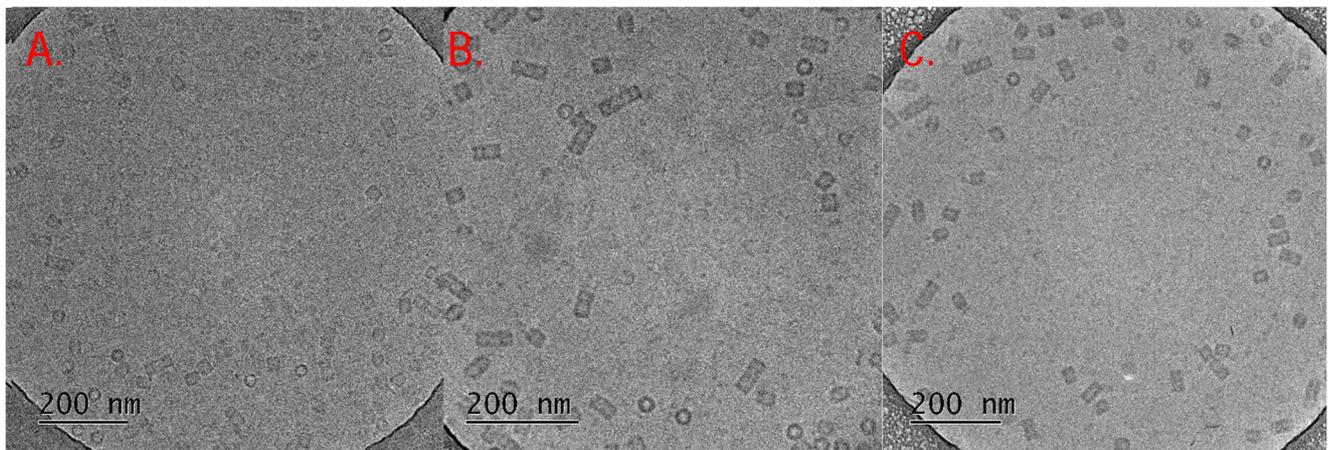
Figure 12. *Cryo-EM Images of purified hemocyanin sample at 1 mg/mL Grids were mounted in a Gatan 914 holder and imaged on the JEM-2100 microscope. A, B and C show successful sample preparation of hemocyanin with clear hemocyanin protein displayed within the ice. A, B, C show an array of structures from decamer, didecamer, tridecamers and multimeric structures. Future analysis and 3D reconstruction are needed.*

1.5 Discussion

Crepidula fornicata (slipper limpet) is a UK coast pest that has negative implications on coastal ecosystems and native species. However, as problematic as they are previous and current research has revealed that an abundant, extracellular protein – hemocyanin - may provide the basis of a market for these pests. Hemocyanin has several immunological applications as an adjuvant as well as an anti-tumour reagent but, without basic understanding of how to extract, purify and structurally characterise hemocyanin there is no basis for immunological research. Hence, the aim of this study was to extract, purify and structurally characterise hemocyanin from Crepidula fornicata (slipper limpet) and build a basic understanding of its structural architecture whilst providing a solution to possibly two problems – removing an invasive species and creating a potential product that could have future immunological applications.

During this study both analytical and biochemical techniques were used to extract, purify and structurally characterise hemocyanin – SDS-PAGE analysis, Gel Filtration Chromatography, Ammonium Sulphate Purification, Electron Microscopy and Cryo-EM are among the various techniques used.

Basic Investigation into Hemocyanin and Purification Method

During this section molecular weight and purification method were investigated. **Figure 3**, an SDS-PAGE, revealed two banding patterns of 274 kDa and 254. The average molecular weight in both cases of SLH sits slightly smaller in size in comparison to the usual range of 300-450 kDa in gastropod species, but the presence of a smaller, second band is a characteristic illustrated in *Megathura crenulata* (Keyhole Limpet) hemocyanin where the presence of a second isoform is seen. These are noted as KLH-1 (keyhole limpet hemocyanin one) and KLH-2 (keyhole limpet hemocyanin two). KLH-1 and KLH-2 are extremely similar in structure and contain 10 subunits which contribute to its decamer appearance; the difference comes within the subunit structure of the FU's. KLH-1 contains 8 functional units termed FU-a to FU-h whereas KLH-2 contains 7 FU's termed FU-a to FU-g (Jaenicke et al., 2011)(Markl, 2013). The addition of FU-h at the C-terminus of the FU chain means there is a slight molecular weight difference between the two isoforms as displayed in **Figure 3** and for this reason the 274 kDa band will be termed SLH-1 and the 254 kDa will be termed SLH-2.
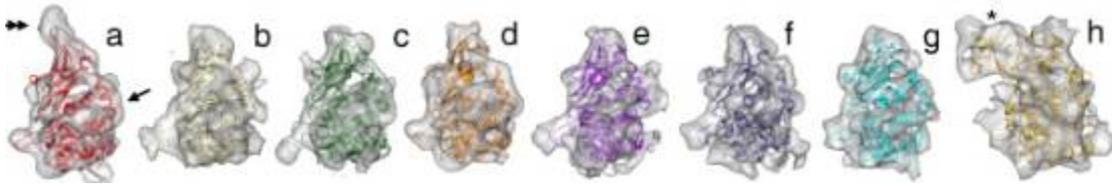
Figure 13. *Subunit structures were extracted from KLH1 9-Å cryoEM density map. Individual FUs with their specific molecular model. Note the similar composition of structure between FU-b to FU-g with slight differences in FU-h and FU-a. FU-h contains a C-terminal cupredoxin-like fold which has been noted by an * and this is responsible for FU-h being slightly larger than other FUs. FU-a contains an anchor like structure which is slightly elongated out from the top of its functional unit and this is noted by the arrows to show the position (Gatsogiannis and Markl, 2009).*

Purification by ammonium sulphate though is successful as displayed in **Figure 6** where hemocyanin is retained in both cases but lack of contaminants in the first place proposes the question of is this purification if the fresh SLH sample lacks contaminants? In this case the addition of ammonium sulphate facilitated the removal of SLH-1 and SLH-2 from hemolymph to a buffer of choice, rather than directly purified the sample.

Partial Cleavage by V8 Protease from Staphylococcus aureus – Revealing the Subunit Architecture of Hemocyanin

During partial cleavage by V8 protease as described by Gebuer and Harris, 1999, cleavage occurs in a way creating specific fragments of FUs. Initially a decamer of hemocyanin is cleaved into two fragments of FUs – *abc* and *defgh*. Further cleavage of fragment *defgh* results in any of the following cleavages - FU-*def*, FU-*defg*, FU-g or FU-h. Hence, it can be theorised that the singular banding patterns displayed in **Figure 7** at 40-45 kDa and 55 – 60 kDa ranges are representative of FU-g and FU-h fragments (FU-h is usually of a higher molecular weight because of its additional cupredoxin domain which is 100 amino acids long)(Gebauer, et al. 1999) and the presence of FU-h means SLH-1 has most likely been cleaved in comparison to that of SLH-2.

This data is further backed by a chromograph displayed in **Figure 8** with a series of distinctive peaks – Peak 4 is likely a SLH-1 decamer with a molecular weight (MW) of 282 kDa. Peak 5 represents fragment patterns of *abc, def, defg or defgh*. These cleavage patterns prove that SLH-1 is being cleaved in an extremely similar way to KLH-1 and therefore it is highly likely that the FU composition of KLH-1 and SLH-1 are extremely similar, allowing accurate assumptions to be made about SLH-1 about KLH-1 data. Fractions corresponding to each peak were separated by SDS-PAGE and further confirmed cleavage patterns with similar molecular weights displayed (**Figure 9**).

Interestingly, Fraction 6 contains two bands of 97.82 and 43.18 kDa these are roughly double the size of each other and the lower band is consistent with FU-g and the higher band is twice this size. It is possible that 97.82 kDa band is a product of dimerisation between FUs. Dimerisation is common for FU-g and FU-h and is another characteristic of molluscan hemocyanin and is responsible for the overall structure of the "collar" (**Figure 15**) (Markl, J. 2013). There are conformers of both FU-g and FU-h which are generally noted as FU-g1 and FU-h1, the different spatial arrangement between FU-g and FU-g1 mean that parallel dimerization can occur which is what gives the collar structure its shape and allows the "face" of hemocyanin to stabilise itself ready for "face to face" interaction with another decamer structure. (Markl, J. 2013)
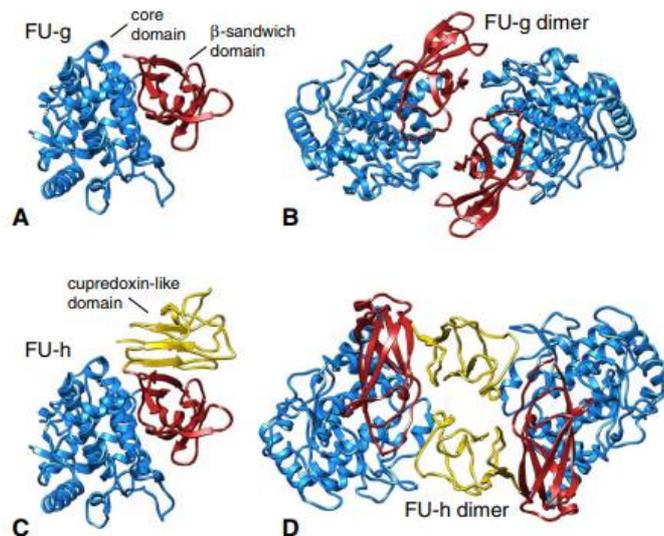
Figure 15. *Image depicts crystal structure of FU-g and FU-h which are both collar functional units within a decamer structure.* ***A)*** *FU-g has a core domain (blue) and a B-sandwich domain (red), crystal structure taken from (PDB-ID 1JS8) and* ***B)*** *displays dimerization of FU-g, units are displayed as conformers of each other.* ***C)*** *FU-h has a core domain (blue) and a B-sandwich domain (red) but has an extra cupredoxin C-terminal fold (yellow), crystal structure taken from (PDB-ID 3QJO).* ***D)*** *Dimerization can occur by two FU-h's forming conformers of each other as displayed by FU-g.* (Markl, J. 2013)

Sequence, EM and Cryo-EM Analysis of Hemocyanin – Insight into Structural Characteristic

Electron microscopy analysis and 3D reconstruction of hemocyanin reveal clear similarities to molluscan hemocyanin in appearance. Its hollow-cylinder shape is consistent of molluscan hemocyanin (**Figure 10a** and **b**), but the presence of long, multimeric structures may be yet another unique characteristic either produced by SLH-2 or by all isoforms (Markl, J. 2013). The ability to create multimeric structures is a direct correlation with the presence of FUs more specifically FU-g and FU-h. As previously stated, dimerization between pairs of either FU-g and FU-h can occur and result in different collar shapes at the "face" end of hemocyanin. If FU-g dimerization is seen there is a non-continuous, arch like pattern consistent of an "open face" as displayed in **Figure 16a** and in **Figure10c at 20 nm**. Conversely when FU-h dimerization is seen there is a continuous peripheral shape consistent of a "closed face" as displayed in **Figure 16b**.

Generally, molluscan hemocyanin's will create didecamers, tridecamers and multimeric structures by "open" face interaction to another hemocyanin (Decker et al. 2007) (Harris, et al. 2000*)*. However, in some gastropod species varying isoforms can either limit structures to decamer or didecamer level or even promote excessive multimeric formation. KLH-1 and KLH-2 are an example of this - KLH-1 is restricted to decamer and didecamer formation, despite similarities between the structures KLH-1 occupies a sugar chain at the "closed face" which prohibits "open face" interaction. (Markl, J. 2013) This could be the case here, SLH-1 could have a sugar chain extended from the FU-h which prohibits formation of larger structures and SLH-2 could be responsible for creating the multimeric structures. On the other hand, all isoforms SLH-1 and SLH-2 may favour multimeric formation as it could be more stable this way as demonstrated by Harris, et al. 1999 with *Haliotis tuberculata hemocyani*n (HtH1 and HtH2).
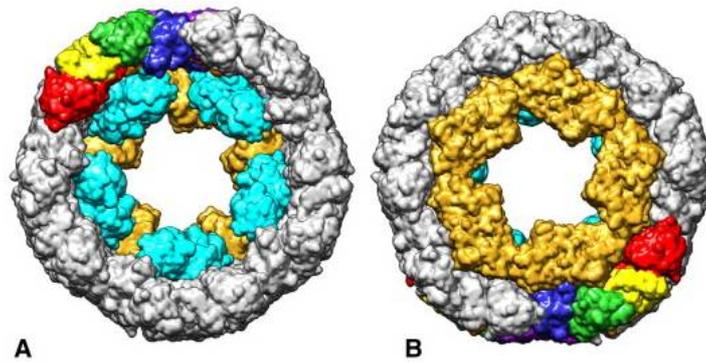
Figure 16. *Structure of an atomistic model of KLH-1 at 7A resolution. A) FU-a to FU-f (red, yellow, green, orange, purple and dark blue) construct the wall of hemocyanin and FU-g and FU-h (cyan and golden) the collar. This depicts an "open face" structure with ten FU-g (cyan) are dimerised to each other (Fu-g to Fu-g1) giving 5 pairs and give a non-continuous arch like collar. B) This depicts a "closed face" structure with ten Fu-h (golden) are dimerised to each other (Fu-h to Fu-h1) giving 5 pairs and give a continuous peripheral arch like collar.*
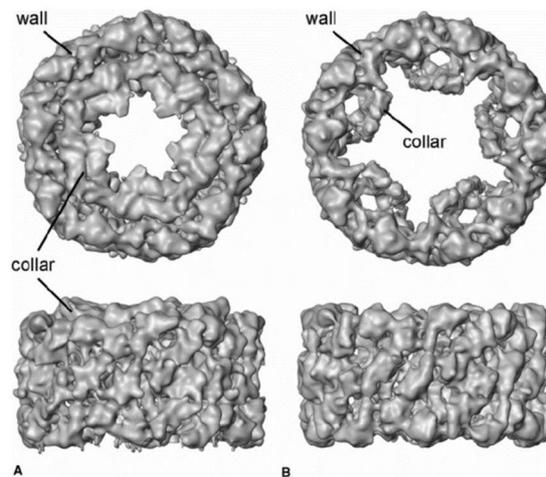
Figure 17. *Cryo-EM 3D reconstruction of molluscan decamer hemocyanin. A) Depicts top and side view of hemocyanin at 11 A resolution from Haliotis tuberculata (Green omer)(Linnaeus 1758) hemocyanin isoform 1 (HtH1) B) Depicts top and side view of hemocyanin at 11 A resolution from Nautilus pompilius (Chambered nautilus)(Linnaeus 1758) hemocyanin (HpH). Both have extremely similar structures with the same D5 poitngroup symmetry (Decker et al. 2007).*

3D Reconstruction of SLH obtained poor resolution of *41.2 Å* but similarities in the overall structure to other 3D reconstructions are displayed (**Figure 16 and Figure 17**). Evidently there was a clear problem with reconstruction of the collar region meaning the structure appears extremely hollow. Lack of data to reconstruct the collar at the "face" region of the hemocyanin presents another problem of what isoform is this model of? If a better resolution was achieved, shape of the collar could help identify an isoform. If there are non-continuous arch like shapes at either end of the 3D model typical of FU-g dimerization, then it can be assumed this is SLH-2 as it lacks FU-h. If there is a continuous peripheral shape at one end typical of FU-h and a non-continuous arch at the other, FU-g then it can be assumed that this is SLH-1.

Problems in reconstruction is due to electron microscopy only capturing either "side view" or "face view" with a lack of intermediate data which has resulted in lack of data within the class averages

and stacked data projections displayed in **APPENDIX 11**, and therefore certain areas of the model have not been reconstructed properly. This limitation can be solved by Cryo-EM, Figure 12 shows various angles of hemocyanin and this would provide the intermediate data that was previously lacked.

Limitations of Study

During this study there were numerous limitations which impacted the consistency of results. The main limitation of this study is that slipper limpets lacked consistency in terms of how "fresh" they were. This has a direct correlation in isoform production of hemocyanin and the presence of certain isoforms can be an indicator of initial degradation of the slipper limpet. This has been proven by Oakes et al 2004 during a study that monitored "*The effect of captivity and diet on KLH isoform ratios in Megathura crenulata*". Percentage of KLH-1 within the hemolymph drastically dropped when they were exposed to a food restriction and an increase in KLH-2 was seen. Though this was over a period of 4 months the reality of collecting limpets from the shore means that there is an unawareness of time frame and degradation may have begun. It begs the question are these "*unique*" results because these limpets have been on the shore for months and have a restricted diet? Before additional structural and purification methods are implemented it is imperative limpets are collected from the sea and monitored in terms of diet and certain captivity environments to ensure that this unique characteristic independent to SLH is a reproducible result in captivity and not just an interesting mishap that has shown itself because a selection of limpets have started to degrade.

Additionally, as previously discussed, 3D reconstruction of hemocyanin warrants a huge amount of improvement. Electron microscopy though produced detailed images did not produce the same standard of reconstruction and this is because of lack of intermediate data. It is imperative that Cryo-EM is used to achieve several rotational angle images of hemocyanin that can provide that missing data which means that reconstruction between the wall and collar can be achieved. This alone will give a massive implication as to what isoform has been reconstructed and with a clear reconstruction of SLH immunological experiments can begin. It would be possible to use SLH as an adjuvant and see its effect on diseases but, without fundamental modelling of its structure this could prove difficult.

Lastly, the whole process was subject to human error at any one point. The processing of hemocyanin occurred at Mikota Ltd laboratory and then samples were sent to Cardiff University the same day and further processed. This means that each batch of sample was a shared responsibility between at least 2-3 people. Everyone has their own techniques and experience, and this can create huge differences between batches of sample that are later sent to Cardiff University. This was seen a lot throughout this study some samples were not used for repeated experiments because they were not consistent with other samples that were previously received. Human error was also seen during the V8 protease experiment. Peak 7 and 8 in **Figure 8** were not reproducible and most likely a product of contamination. However, shared handling during the process of hemocyanin would account for small in regular inconsistencies as a strict protocol was being followed.

Following the information retained during this study there is definite basis for future immunological study as KLH is an upcoming immunotherapy for cancer following its potent anti-tumour effects described in Riggs et al, 2002 and this could be a potential for SLH-1 due to its similar structure. The unique characteristics and structural similarities to KLH reported in this study has already provided three research projects which have commenced as of this summer and has already proved promising - "*Treatment of Bone Malignancies: Proof of Principle Study*" "*Assessing the anticancer properties*

of Biocyanin SLH” and an Innovate UK grant to *“Develop a scalable process to isolate pure, stable SLH”*.

In conclusion, the aim of this study has been clearly met and SLH has been extracted, purified and structurally characterised. SLH has exhibited not only unique characteristics but similar characteristics to KLH-1 and KLH-2 and this is a good foundation for possible future market and will provide a platform for further immunological and structural research. With further study and the possibility of a market it is clear there could be a solution to this non-native invader that could solve more than one problem and eventually reduce their impact on coastal ecosystems worldwide.

Acknowledgements

Bibliography

Bacci, M., Baldecchi, M., Fabeni, P., Linari, R. and Pazzi, G. (1983). Emission spectra from copper proteins containing type-3 centres. Biophysical Chemistry, 17(2), pp.125-130.

Bohn, K., Richardson, C. and Jenkins, S. (2015). The distribution of the invasive non-native gastropod Crepidula fornicata in the Milford Haven Waterway, its northernmost population along the west coast of Britain. Helgoland Marine Research, 69(4), pp.313-325.

Bonaventura, J., Bonaventura, C. and Sullivan, B. (1975). Hemoglobins and hemocyanins: Comparative aspects of structure and function. Journal of Experimental Zoology, 194(1), pp.155-174.

Christy, C., Adams, G., Kuriyel, R., Bolton, G. and Seilly, A. (2002). High-performance tangential flow filtration: a highly selective membrane separation process. Desalination, 144(1-3), pp.133-136.

Decker, H., Hellmann, N., Jaenicke, E., Lieb, B., Meissner, U. and Markl, J. (2007). Minireview: Recent progress in hemocyanin research. Integrative and Comparative Biology, 47(4), pp.631-644.

Gatsogiannis, C. and 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. Journal of Molecular Biology, 385(3), pp.963-983

Gebauer, W. and Harris, J. (1999). Controlled cleavage of KLH1 and KLH2 by the V8 protease from Staphylococcus aureus. Reassociation, electrophoretic and transmission electron microscopy study of peptide fragments. European Journal of Biochemistry, 262(1), pp.166-175.

Gebauer, W., Harris, J., Geisthardt, G. and Markl, J. (1999). Keyhole Limpet Hemocyanin Type 2 (KLH2): Detection and Immunolocalization of a Labile Functional Unit h. Journal of Structural Biology, 128(3), pp.280-286.

Grosholz, E. (2002). Ecological and evolutionary consequences of coastal invasions. Trends in Ecology & Evolution, 17(1), pp.22-27.

Hagel, L. (2001). Current protocols in Molecular Biology. 1st ed. Sweden: Amersham Pharmacia Biotech AB, pp.Chapter 10, Unit 10.9

Harris, J., Scheffler, D., Gebauer, W., Lehnert, R. and Markl, J. (2000). Haliotis tuberculata hemocyanin (HtH): analysis of oligomeric stability of HtH1 and HtH2, and comparison with keyhole limpet hemocyanin KLH1 and KLH2. Micron, 31(6), pp.613-622.

Hiscock, K. (2018). MarLIN - The Marine Life Information Network - Slipper limpet (Crepidula fornicata). [online] Marlin.ac.uk. Available at: https://www.marlin.ac.uk/species/detail/1554 [Accessed 1 Sep. 2018].

Idakieva, K., Chakarska, I., Ivanova, P., Tchorbanov, A., Dobrovolov, I. and Doumanova, L. (2009). Purification of Hemocyanin from Marine GastropodRapana Thomasianausing Ammonium Sulfate Precipitation Method. Biotechnology & Biotechnological Equipment, 23(3), pp.1364-1367.

Jaenicke, E., Büchler, K., Decker, H., Markl, J. and Schröder, G. (2011). The refined structure of functional unit h of keyhole limpet hemocyanin (KLH1-h) reveals disulfide bridges. IUBMB Life, 63(3), pp.183-187.

Markl, J. (2013). Evolution of molluscan hemocyanin structures. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 1834(9), pp.1840-1852.

Montaudouin, X., Audemard, C. and Labourg, P. (1999). Does the slipper limpet (Crepidula fornicata, L.) impair oyster growth and zoobenthos biodiversity? A revisited hypothesis. Journal of Experimental Marine Biology and Ecology, 235(1), pp.105-124.

Oakes, F., McTee, S., McMullen, J., Culver, C. and Morse, D. (2004). The effect of captivity and diet on KLH isoform ratios in Megathura crenulata. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 138(2), pp.169-173.

Robin Harris, J., Gebauer, W., Söhngen, S. and Markl, J. (1995). Keyhole limpet haemocyanin (KLH): Purification of intact KLH1 through selective dissociation of KLH2. Micron, 26(3), pp.201-212.

Robin Harris, J., Gebauer, W., Söhngen, S. and Markl, J. (1995). Keyhole limpet haemocyanin (KLH): Purification of intact KLH1 through selective dissociation of KLH2. Micron, 26(3), pp.201-212.

van Reis, R., Gadam, S., Frautschy, L., Orlando, S., Goodrich, E., Saksena, S., Kuriyel, R., Simpson, C., Pearl, S. and Zydney, A. (1997). High performance tangential flow filtration. Biotechnology and Bioengineering, 56(1), pp.71-82.

Walker, J. (2002.). Gradient SDS Polyacrylamide Gel Electrophoresis. Proteins, pp.57-62.

Wingfield, P. (1998). Protein Precipitation Using Ammonium Sulfate. Current Protocols in Protein Science, pp.A.3F.1-A.3F.8.7

Appendix

APPENDIX 1. Numerical data and BSA Standard Curve at 595nm. Repeats of 5 BSA samples were recorded and a blank of just 1x Bradford Reagent was then used to remove any absorbance occurring because of solution hence "Average Absorbance (WB) = Average Absorbance (Without blank)", it's this average that was blotted accordingly on the graph. *BSA Standard Curve at 595nm. Concentration values of BSA (0.5, 0.25, 0.125, 0.075, 0.05, 0.025 and 0 mg/mL) obtained set absorbances at 595nm, numerical data can be found within APPENDIX 1. Equation of the line calculated as y = 1.479x + 0.0121 where y = absorbance at 595 nm and x = protein concentration (mg/mL). Standard deviation error bars plotted accordingly.*
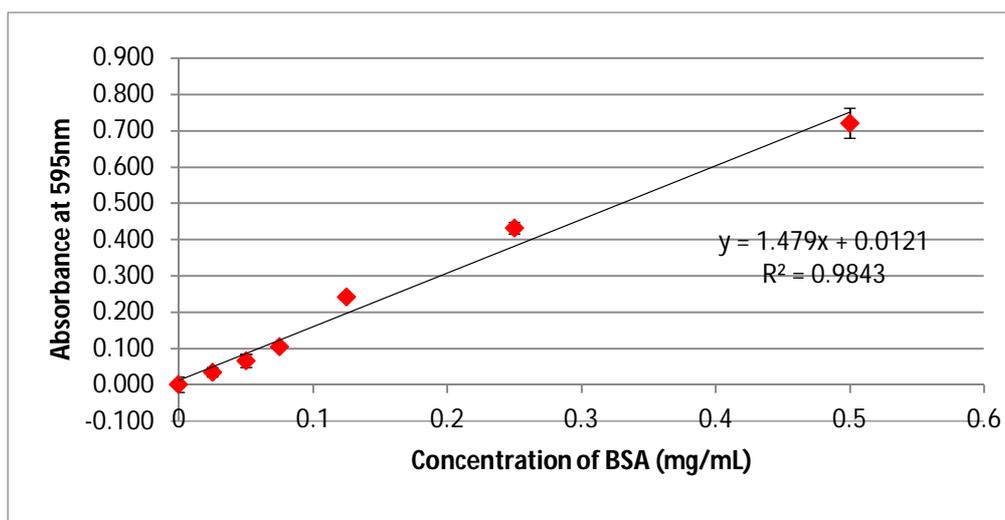
Absorbance at 595 nm								
Sample	1	2	3	4	5	Average Absorbance (WB)	Concentration (mg/mL)	Standard Deviation
1	1.013	1.008	0.998	1.049	0.936	0.721	0.5	0.041
2	0.691	0.708	0732	0.722	0.706	0.432	0.25	0.016
3	0.521	0.516	0.522	0.529	0.522	0.242	0.125	0.005
4	0.390	0.376	0.391	0.383	0.389	0.106	0.075	0.006
5	0.360	0.368	0.340	0.323	0.338	0.066	0.05	0.018
6	0.298	0.303	0.318	0.324	0.329	0.034	0.025	0.013
7	0.290	0.301	0.291	0.273	0.247	0.000	0	0.021

APPENDIX 2. Example SDS Calibration Graph for **Figure 4**, 4-15% Tris/Glycine Gradient Gel displaying 5µg of freshly extracted hemolymph.
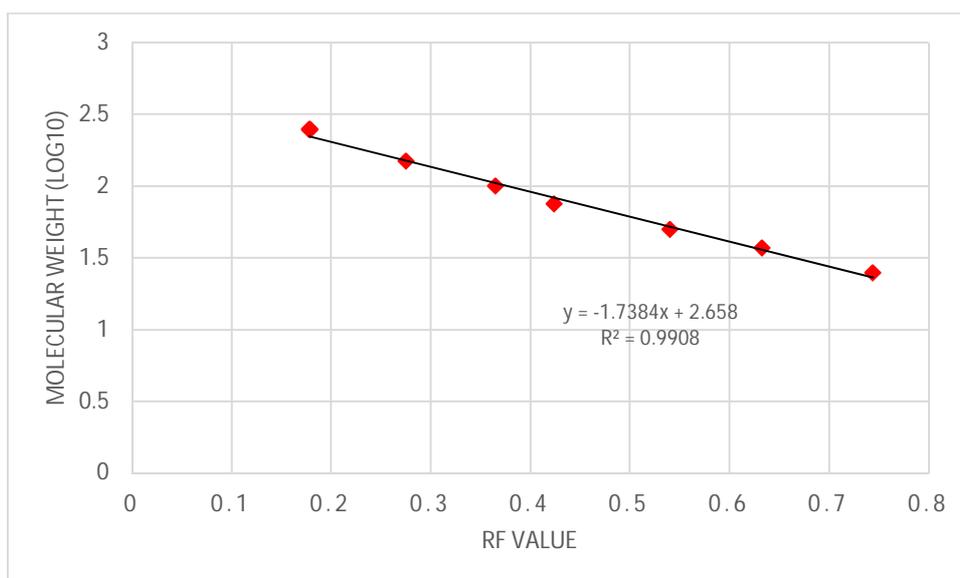
APPENDIX 3. *4-15% Tris/Glycine Gradient Gel displaying 5µg of freshly extracted hemolymph displaying higher and lower band in **Figure 4**. Table displays numerical data used to calculate average molecular weight (kDa) of both bands using equation of the line y= -1.7384x +2.658, where y= molecular weight (log10) and x = Rf value.*

Band	1	2	3	Average Length of Band (cm)	Rf Value	Molecular Weight (Log10 kDa)	Molecular Weight (kDa)
Higher Band n=1	62.514	59.135	60.033	60.56066667	0.126773399	2.43761712	273.92
Lower Band n=1	69.682	68.064	69.007	68.91766667	0.144267349	2.40598689	254.68
Higher Band n=2	61.008	59.034	59.734	59.92533333	0.125443437	2.43896073	274.76
Lower Band n=2	70.007	69.028	68.002	69.01233333	0.144465517	2.40563975	254.47
Higher Band n=3	61.008	60.128	59.576	60.23733333	0.126096555	2.43781666	274.04
Lower Band n=3	70.06	68.561	69.478	69.36633333	0.145206556	2.40434168	253.71

Band	Average MW from n=1	Average MW from n=2	Average MW from n=3	Average Molecular Weight (kDa)	Standard Deviation
Higher Band	273.92	274.76	274.04	274.24	0.458
Lower Band	254.68	254.47	253.71	254.29	0.508

APPENDIX 4. *Gel filtration Superose® 6 10/300 GL Molecular Weight Calibration Graph and data used to create graph, molecular weights, 670000, 158000, 44000, 1700, 1350 Da of set retention times plotted accordingly and equation of line created, y= -0.4013x + 11.595, where y= molecular weight (log10 Da) and x = retention volume (mL). Should be noted that markers were eluted in Stabilising Buffer (150 mM NaCl, 50mM Tris/HCl, 5mM MgCl2 and 5mM CaCl2) at a flow rate of 1.5 mL/min.*

Number of Peak	Molecular Weight Marker (Da)	Retention (mL)	Molecular Weight Marker (log10 Da)
1	670000	13.86	5.8260748
2	158000	16.91	5.19865709

3	44000	18.25	4.643545268
4	1700	19.25	3.23044892
5	1350	21.30	3.13033377

$y = -0.4013x + 11.595$
$R^2 = 0.8688$
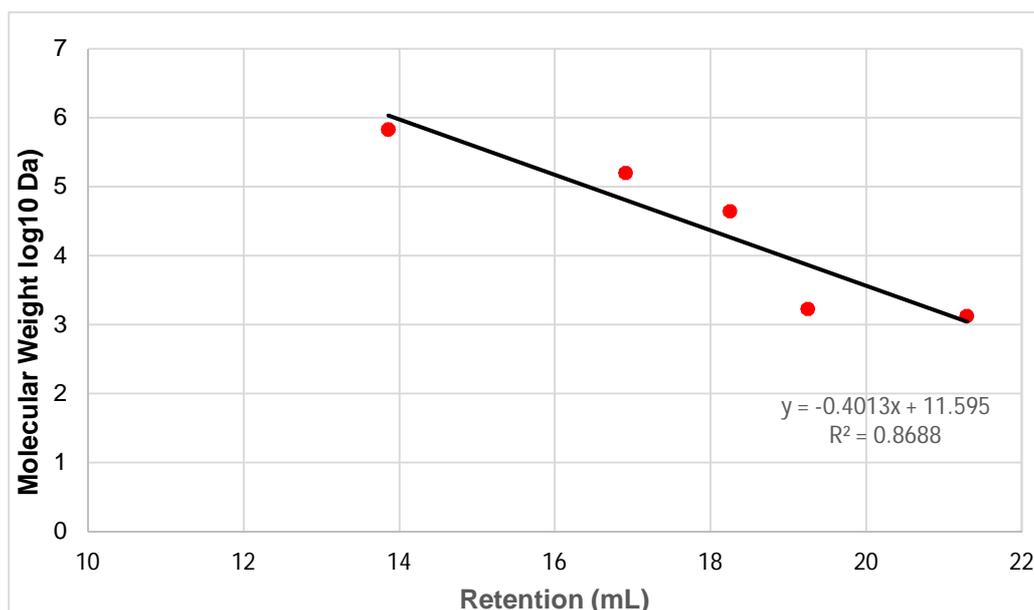
APPENDIX 5. *4-15% Tris/Glycine Gradient Gel displaying Fresh Sample 1 (FS1) and Purified Sample 1 (PS1) displaying one distinctive band in both lanes in Figure 6. Table displays numerical data used to calculate average molecular weight (kDa) of bands present in each lane using equation of the line y = -1.714x + 2.6145, where y = molecular weight (log10 kDa) and x = Rf value.*

Sample	Repeat Number (n= value)	Distance of Band	Rf Value	Molecular Weight (log10 kDa)	Molecular Weight (kDa)
FS1	1	0.582	0.1037618	2.436652255	273.31
FS1	1	0.56	0.0998395	2.443375022	277.57
FS1	1	0.556	0.0991264	2.444597344	278.35
FS1	2	0.574	0.1023355	2.439096898	274.85
FS1	2	0.556	0.0991264	2.444597344	278.35
FS1	2	0.53	0.094491	2.452542432	283.49
FS1	3	0.573	0.1021572	2.439402478	275.04
FS1	3	0.582	0.1037618	2.436652255	273.31
FS1	3	0.529	0.0943127	2.452848012	283.69
PS1	1	0.569	0.1014441	2.440624799	275.82
PS1	1	0.56	0.0998395	2.443375022	277.57
PS1	1	0.551	0.098235	2.446125245	279.33
PS1	2	0.547	0.0975218	2.447347566	280.12
PS1	2	0.56	0.0998395	2.443375022	277.57
PS1	2	0.541	0.0964521	2.449181048	281.31
PS1	3	0.575	0.1025138	2.438791318	274.66
PS1	3	0.529	0.0943127	2.452848012	283.69
PS1	3	0.539	0.0960956	2.449792209	281.70

Band	Average MW from n=1 (kDa)	Average MW from n=2 (kDa)	Average MW from n=3 (kDa)	Average Molecular Weight from n=1, n=2 and n=3 (kDa)	Standard Deviation
FS1	276.41	278.90	277.35	277.55	1.26
PS1	277.58	279.67	280.02	279.09	1.32

APPENDIX 6. Numerical data used to calculate average molecular weight (kDa) of both bands proposed to be FU-h and FU-g displayed in **Figure 7** on a 4-15% Tris/Glycine Gradient Gel. Equation of the line created from a calibration SDS-PAGE graph like graph displayed in **APPENDIX 2** – y=-1.8102x + 2.5755 where y = molecular weight (log10 kDa) and x = Rf value. Average molecular weight of 63.38 kDa for proposed FU-h and 46.25 kDa for FU-g.

Band	Repeat Number	Length of Band (cm)	Length of Gel (cm)	Rf Value	Molecular Weight (log10 kDa)	Molecular Weight (kDa)
FU-h	n=1	394.994	928.073	0.425606606	1.805066921	63.84
FU-h	n=1	398.723	928.073	0.429624609	1.797793532	62.78
FU-h	n=1	398.742	928.073	0.429645082	1.797756473	62.77
FU-h	n=2	398.623	928.073	0.429516859	1.797988582	62.80
FU-h	n=2	394.413	928.073	0.424980578	1.806200158	64.00
FU-h	n=2	396.188	928.073	0.426893143	1.802738032	63.49
FU-h	n=3	399.047	928.073	0.42997372	1.797161573	62.68
FU-h	n=3	390.129	928.073	0.420364562	1.81455607	65.25
FU-h	n=3	398.502	928.073	0.429386481	1.798224591	62.84
FU-g	n=1	465.394	928.073	0.501462708	1.667752206	46.53
FU-g	n=1	468.973	928.073	0.505319086	1.660771391	45.79
FU-g	n=1	464.744	928.073	0.500762332	1.669020026	46.67
FU-g	n=2	460.999	928.073	0.496727089	1.676324623	47.46
FU-g	n=2	468.436	928.073	0.504740468	1.661818806	45.90
FU-g	n=2	456.22	928.073	0.49157771	1.685646029	48.49
FU-g	n=3	474.713	928.073	0.511503944	1.64957556	44.62
FU-g	n=3	468.517	928.073	0.504827745	1.661660816	45.88
FU-g	n=3	473.25	928.073	0.50992756	1.652429132	44.92
Enzyme	n=1	609.892	928.073	0.657159512	1.385909851	24.32
Enzyme	n=1	611.444	928.073	0.658831794	1.382882686	24.15
Enzyme	n=1	614.833	928.073	0.662483447	1.376272464	23.78
Enzyme	n=2	6113.439	928.073	0.66098141	1.378991452	23.98
Enzyme	n=2	606.557	928.073	0.653566045	1.392414745	24.68
Enzyme	n=2	612.029	928.073	0.659462133	1.381741647	24.08
Enzyme	n=3	610.03	928.073	0.657308207	1.385640683	24.30
Enzyme	n=3	606.083	928.073	0.653055309	1.393339279	24.74
Enzyme	n=3	603.128	928.073	0.649871292	1.399102986	25.07

Band	Average MW from n=1 (kDa)	Average MW from n=2 (kDa)	Average MW from n=3 (kDa)	Average Molecular Weight (kDa)	Standard Deviation
FU-h	63.13	63.43	63.59	63.38	0.233
FU-g	46.33	47.28	45.18	46.25	1.052
Enzyme	24.08	24.23	24.70	24.34	0.323

APPENDIX 7. *Gel filtration Superdex 200 Increase 6 10/300 GL Molecular Weight Calibration Graph, molecular weights, 670000, 158000, 44000, 1700, 1350 Da of set retention times plotted accordingly and equation of line created, y= -0.2648x + 8.3795, where y= molecular weight (log10 Da) and x = retention volume (mL). Data used to create this graph is displayed below. Should be noted that markers were eluted in Stabilising Buffer (150 mM NaCl, 50mM Tris/HCl, 5mM MgCl2 and 5mM CaCl2) at a flow rate of 0.75 mL/min.*
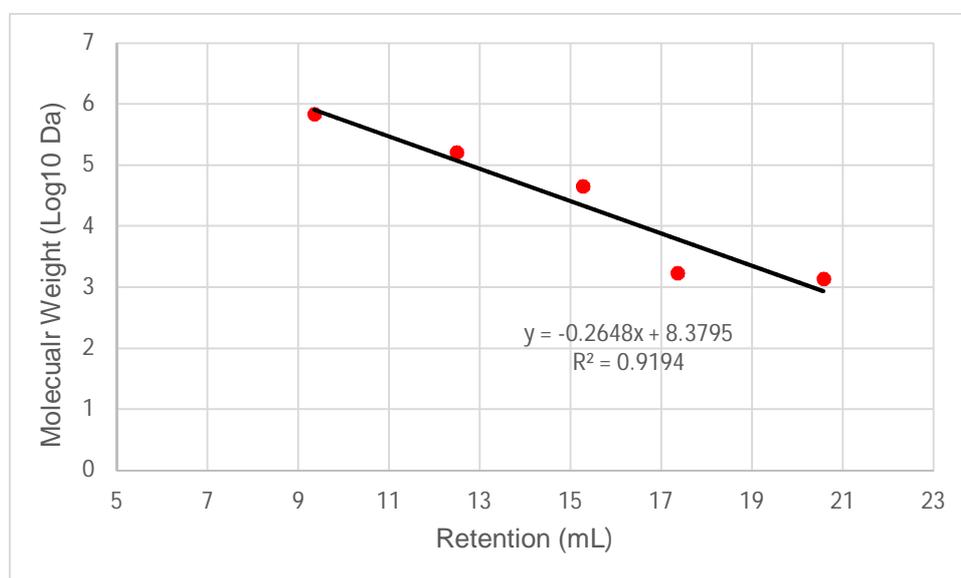
Number of Peak	Molecular Weight (Da)	Retention (mL)	Molecular Weight (log10 Da)
1	670000	9.35	5.826075
2	158000	12.49	5.198657
3	44000	15.27	4.643453
4	1700	17.36	3.230449
5	1350	20.57	3.130334

APPENDIX 8. *Molecular weight calculations from repeated data obtained from Gel Filtration (Superdex 200 Increase 10/300 GL) of 10-15mg/mL purified hemocyanin cleaved with V8 protease for 5 hours at 37 °C. Raw data for each repeat is displayed below and an average estimated molecular weight has been shown along with standard deviation. Example of chromograph obtained from this sample is displayed in **Figure 8**.*

Number of Peak	Repeated Sample (n value)	Retention (mL)	Molecular Weight (Log10 Da)	Molecular Weight (Da)	Molecular Weight (kDa)
1	n=1	8.845	6.07525	1189187	1189
1	n=2	8.76	6.0948	1243942	1244
1	n=3	8.824	6.08008	1202486	1202
2	n=1	9.46	5.9338	858618	859
2	n=2	9.386	5.95082	892935.3	893
2	n=3	9.437	5.93909	869140.5	869
3	n=1	10.578	5.67666	474963.2	475
3	n=2	10.489	5.69713	497886.1	498
3	n=3	10.526	5.68862	488225	488
4	n=1	11.485	5.46805	293798.8	294
4	n=2	11.55	5.4531	283857.3	284
4	n=3	11.648	5.43056	269500.8	270
5	n=1	12.438	5.24886	177361.8	177
5	n=2	12.379	5.26243	182991.1	183
5	n=3	12.478	5.23966	173644.1	174
6	n=1	15.389	4.57013	37164.65	37
6	n=2	15.298	4.59106	38999.59	39
6	n=3	15.369	4.57473	37560.38	38
7	n=1	17.425	4.10185	12643	13
8	n=1	20.641	3.36217	2302.343	2

Peak Number	Average Molecular Weight from n=1, n=2 and n=3 (kDa)	Standard Deviation
1	1211.87	28.75
2	873.56	17.47
3	487.02	11.53
4	282.39	12.06
5	178.00	4.58
6	37.91	1

APPENDIX 9. *Molecular weight calculations from repeated data obtained from SDS-PAGE analysis of 2mL fractions of hemocyanin cleaved with V8 protease for 5 hours at 37 °C separated by Superdex 200 Increase 10/300 GL column. Example of banding patterns displayed in **Figure 9**. Estimation of molecular weight deduced from* equation of the line created from a calibration SDS-PAGE graph like graph displayed in **APPENDIX 2** – y=-1.7622x + 2.6737 where y = molecular weight (log10 kDa) and x = Rf value.

Lane	Band	Repeat Number (n value)	Length of Band (cm)	Rf Value	Molecular Weight (log10 kDa)	Molecular Weight (kDa)

BSA	BSA 1mg/mL	n=1	131	0.476163	1.827377	67.20
BSA	BSA 1mg/mL	n=2	134.034	0.487191	1.80798	64.27
BSA	BSA 1mg/mL	n=3	132.095	0.480143	1.820377	66.13
S	FU-h	n=1	142	0.516146	1.757051	57.15
S	FU-h	n=2	150.03	0.545334	1.705713	50.78
S	FU-h	n=3	139.004	0.505256	1.776205	59.73
S	FU-g	n=1	161	0.585208	1.635578	43.21
S	FU-g	n=2	163.003	0.592488	1.622772	41.95
S	FU-g	n=3	162.012	0.588886	1.629108	42.57
S	Enzyme	n=1	206.002	0.748782	1.347867	22.28
S	Enzyme	n=2	208.002	0.756052	1.33508	21.63
S	Enzyme	n=3	207.002	0.752417	1.341473	21.95
F8	Band 1	n=1	23	0.083601	2.517854	329.50
F7	Band 1	n=1	23.008	0.08363	2.517803	329.46
F6	Higher Band	n=1	106	0.385292	1.98721	97.10
F6	Higher Band	n=2	105.005	0.381675	1.993571	98.53
F6	Higher Band	n=3	109.073	0.396462	1.967563	92.80
F6	Lower Band	n=1	162.111	0.589246	1.628475	42.51
F6	Lower Band	n=2	161.199	0.585931	1.634306	43.08
F6	Lower Band	n=3	163.077	0.592757	1.622299	41.91
F5	Higher Band	n=1	81.099	0.294781	2.14641	140.09
F5	Higher Band	n=2	82.024	0.298143	2.140496	138.20
F5	Higher Band	n=3	83.096	0.30204	2.133642	136.03
F5	Middle Band	n=1	151.013	0.548907	1.699428	50.05
F5	Middle Band	n=2	149.121	0.54203	1.711524	51.47
F5	Middle Band	n=3	149.164	0.542186	1.711249	51.43
F5	Lower Band	n=1	208.038	0.756183	1.33485	21.62
F5	Lower Band	n=2	201.062	0.730826	1.37945	23.96
F5	Lower Band	n=3	205.197	0.745856	1.353013	22.54

Lane	Band	Average Molecular Weight from n=1, n=2 and n=3 (kDa)	Standard Deviation

BSA	**BSA**		65.93	1.48
S	**FU-h**		56.81	4.61
S	**FU-g**		43.26	0.63
S	**Enzyme**		22.28	0.33
F6	**Higher Band**		97.82	2.98
F6	**Lower Band**		43.18	0.59
F5	**Higher Band**		140.61	2.03
F5	**Middle Band**		51.81	0.81
F5	**Lower Band**		23.04	1.18

APPENDIX 10. Sequence alignment data produced by Clostal O(1. 2. 4).

TABLE I
N-Terminal Amino Acid Sequences of KLH2 Functional Units

FU	Mass (kDa)	N-terminal sequence
KLH2-*a*	50 ± 5	VDTV**VRK**NVDS**L**SXD............
KLH2-*b*	50 ± 5	.......NLA**VRK**NINDL**L**TAN............
KLH2-*c*	55 ± 5	...DFGHSKK**IRK**NVHS**L**TAADDQ.........
KLH2-*d*	45 ± 5	...AVTSASH**IR**HNIRD**L**GEG............
KLH2-*e*	45 ± 5	.......VPX**IRK**NIK................
KLH2-*f*	50 ± 5	....HVGRNR**IR**MDLSD**L**TXXDLA.........
KLH2-*g*	40 ± 5	.....IAGSG**VRK**DV.................
KLH2-*h*	43 ± 5	......STGF**VRK**DIDS**L**SLDEANDLKNAL...

Note. The KLH2-*a* to KLH2-*g* data are from Gebauer *et al.* (1994) and Söhngen *et al.* (1997).

```
CLUSTAL O(1.2.4) multiple sequence alignment


slipper_Hc1_-_CDS_translation      MVCLPCQTMTTFCGPLALLVLWCQLLLASGSLIRKNVDSLTESEVLTLQTVLRQLEDDKT      60
tr|Q1MVA1|Q1MVA1_MEGCR             --------MWTILALLTAT-LLFEGAFSVDTVVRKNVDSLSSDEVLALEKALDDLQQDDS      51
tr|Q6KC55|Q6KC55_MEGCR             --------MWTILALLTAT-LLFEGAFSVDTVVRKNVDSLSSDEVLALEKALDDLQQDDS      51
tr|Q53IP9|Q53IP9_MEGCR             --------MLSV----RLL-IVVLALANAENLVRKSVEHLTQEETLDLQAALRELQMDSS      47
4BED:A-B|PDBID|CHAIN|SEQUENCE      ---------------------------ENLVRKSVEHLTQEETLDLQAALRELQMDSS      31
                                                              .::**.*: *:..*.* *: .* :*: *.:

slipper_Hc1_-_CDS_translation      ATGFQAIAAYHGEPSQCQTKDGTAIACCVHGEATFPQWHRLYVVQFEQALVSKGLANIGV     120
tr|Q1MVA1|Q1MVA1_MEGCR             NQGYQAIAGYHGVPTMCVDKHEKNVACCLHGMPSFPLWHRLYVVQLERALIRKK-ATISI     110
tr|Q6KC55|Q6KC55_MEGCR             NQGYQAIAGYHGVPTMCVDKHEKNVACCLHGMPSFPLWHRLYVVQLERALIRKK-ATISI     110
tr|Q53IP9|Q53IP9_MEGCR             SIGFQKIAAAHGAPASCVHKD-TSIACCIHGMPTFPHWHRAYVVHMERALQTKR-RTSGL     105
4BED:A-B|PDBID|CHAIN|SEQUENCE      SIGFQKIAAAHGAPASCVHKD-TSIACCIHGMPTFPHWHRAYVVHMERALQTKR-RTSGL      89
                                   *:* **. ** *: * *. . :***:** :** *** ***::*:** *   . .:

slipper_Hc1_-_CDS_translation      PYWDWTKPIKDLPALVKNQIFRDPKGKLGQKNAWFSAEIDHDNVHTQTARAIDDRLYESV     180
tr|Q1MVA1|Q1MVA1_MEGCR             PYWDWT3ELTHLPELV3HPLFVGTEGGKAHDN3WYRADITFL--NKKT3RAVDDRLFEKV     168
tr|Q6KC55|Q6KC55_MEGCR             PYWDWT3ELTHLPELV3HPLFVGTEGGKAHDN3WYRADITFL--NKKT3RAVDDRLFEKV     168
tr|Q53IP9|Q53IP9_MEGCR             PYWDWTEPITQLPSLAADPVYIDSQGGKAHTNYWYRGNIDFL--DKKTNRAVDDRLFEKV     163
4BED:A-B|PDBID|CHAIN|SEQUENCE      PYWDWTEPITQLPSLAADPVYIDSQGGKAHTNYWYRGNIDFL--DKKTNRAVDDRLFEKV     147
                                   ******. ::..** *. . :: . :* .: * *: .:* .    ..:* **:****:*.*

slipper_Hc1_-_CDS_translation      SPGGHTTLFNMVLDALEQDDYCQFEAQFEVAHNHIHYLVGGRHHYSMSTLEYTSFDPLFF     240
tr|Q1MVA1|Q1MVA1_MEGCR             QPGHHTRLMEGILDALEQDEFCKFEIQFELAHNAIHYLVGGRHTYSMSHLEYT3YDPLFF     228
tr|Q6KC55|Q6KC55_MEGCR             QPGHHTRLMEGILDALEQDEFCKFEIQFELAHNAIHYLVGGRHTYSMSHLEYT3YDPLFF     228
tr|Q53IP9|Q53IP9_MEGCR             KPGQHTHLME3VLDALEQDEFCKFEIQFELAHNAIHYLVGGKHDYSMANLEYTAYDPIFF     223
4BED:A-B|PDBID|CHAIN|SEQUENCE      KPGQHTHLME3VLDALEQDEFCKFEIQFELAHNAIHYLVGGKHDYSMANLEYTAYDPIFF     207
                                   .** ** *.:: :*********::*:** ***:*** *******:* ***: ****::**:**

slipper_Hc1_-_CDS_translation      LHH3NVDRIFAIWQELQKKRGKPYDHADCSVELFNKDLLPFARDSNPIQLTRTFSKPKDL     300
tr|Q1MVA1|Q1MVA1_MEGCR             LHH3NTDRIFAIWQRLQQLRGKDPNSADCAHNLIHTPMEPFDRDTNPLDLTREHAKPADS     288
tr|Q6KC55|Q6KC55_MEGCR             LHH3NTDRIFAIWQRLQQLRGKDPNSADCAHNLIHTPMEPFDRDTNPLDLTREHAKPADS     288
tr|Q53IP9|Q53IP9_MEGCR             LHH3NVDRIFAIWQRLQELRNKDPKAMDCAQELLHQKMEPFSWEDNDIPLTN--------     275
4BED:A-B|PDBID|CHAIN|SEQUENCE      LHH3NVDRIFAIWQRLQELRNKDPKAMDCAQELLHQKMEPFSWEDNDIPLTNEHSTPADL     267
                                   *****.********.**: *.*   .  **: :*:: : ** : * : **.

slipper_Hc1_-_CDS_translation      FSYMQLGYTYDDLTLNDMSIDQLYSFLKERHARERSFAVFSLHGMGFSANVKVSVCNVEE     360
tr|Q1MVA1|Q1MVA1_MEGCR             FDYGRLGYQYDDL3LNGMSPEELNVYLGERAAKERTFASFILSGFGGSANVVVVVVCRPAH     348
tr|Q6KC55|Q6KC55_MEGCR             FDYGRLGYQYDDL3LNGMSPEELNVYLGERAAKERTFASFILSGFGGSANVVVVVVCRPAH     348
tr|Q53IP9|Q53IP9_MEGCR             --------DYDTLNLNGMTPEELKTYLDERSSRARAFASFRLKGFGGSANVFVYVCIPDD     327
4BED:A-B|PDBID|CHAIN|SEQUENCE      FDYCELHYDYDTLNLNGMTPEELKTYLDERSSRARAFASFRLKGFGGSANVFVYVCIPDD     327
                                           ** *.**.*: ::* :* ** :: *:** * * *:* **** * **   .

slipper_Hc1_-_CDS_translation      EE--TTNCEFAGDIFILGGANEMAWEFTVPYYYDISDAVRKLGHQQDFNFGVKAEVYSVN     418
tr|Q1MVA1|Q1MVA1_MEGCR             DEISDDQCIKAGDFFLLGGPTEMKWGFYRAYHFDVTDSVASIDDDGHGHYYVK3ELFSVN     408
tr|Q6KC55|Q6KC55_MEGCR             DEISDDQCIKAGDFFLLGGPTEMKWGFYRAYHFDVTDSVASIDDDGHGHYYVK3ELFSVN     408
tr|Q53IP9|Q53IP9_MEGCR             NDRNDDHCEKAGDFFVLGGPSEMKWQFYRPYLFDLSDTVHHMGMKLDGHYTVKAELFSVN     387
4BED:A-B|PDBID|CHAIN|SEQUENCE      NDRNDDHCEKAGDFFVLGGPSEMKWQFYRPYLFDLSDTVHHMGMKLDGHYTVKAELFSVN     387
                                   ::   :* ***:*:*** .** * *  * :*::*:* .: .  . .  :: **:*::***

slipper_Hc1_-_CDS_translation      GTHLGSDVLSRPYASHRNPRGYTDPVIRDSRKPSNVVRKDIELLTDEEVYALRQAMHRFQ     478
tr|Q1MVA1|Q1MVA1_MEGCR             GSALSNDILRQPTLVHRPAKGHFDKPPVPVAQANLAVRKNINDLTAEETYSLRKAMERFQ     468
tr|Q6KC55|Q6KC55_MEGCR             GSALSNDILRQPTLVHRPAKGHFDKPPVPVAQANLAVRKNINDLTAEETYSLRKAMERFQ     468
tr|Q53IP9|Q53IP9_MEGCR             GTALPDDLLPHFVVVHHPEKGFTDPPVKHHQSANLLVRKNINDLTREEVLNLREAFHKFQ     447
4BED:A-B|PDBID|CHAIN|SEQUENCE      GTALPDDLLPHFVVVHHPEKGFTDPPVKHHQSANLLVRKNINDLTREEVLNLREAFHKFQ     447
                                   *:* * .*:* :*  *: .:*. *      .   *** :*: ** **. **:*:.:**

slipper_Hc1_-_CDS_translation      NDTSIDGYQAVAEFHGLPAKCPHPDAAVRYACCVHGMATFPHWHRLFVVEVEDELRSRGL     538
tr|Q1MVA1|Q1MVA1_MEGCR             NDK3VDGYQATVEFHALPARCPRPDAKDRFACCVHGMATFPHWHRLFVTQVEDALLRRG3     528
tr|Q6KC55|Q6KC55_MEGCR             NDK3VDGYQATVEFHALPARCPRPDAKDRFACCVHGMATFPHWHRLFVTQVEDALLRRG3     528
tr|Q53IP9|Q53IP9_MEGCR             EDRSVDGYQATAEYHGLPARCPRPDAKDRYACCVHGMPIFPHWHRLFVTQVEDALVGRGA     507
4BED:A-B|PDBID|CHAIN|SEQUENCE      EDRSVDGYQATAEYHGLPARCPRPDAKDRYACCVHGMPIFPHWHRLFVTQVEDALVGRGA     507
                                   :* *:*****..*:*.* ***:**:*** *:****** ********.:*** * **

slipper_Hc1_-_CDS_translation      EIGI-----PYWDWTRPNAHIPALANDVSYEDANTHQQLHNPFHDAPIAFLGEK--TTRQ     591
tr|Q1MVA1|Q1MVA1_MEGCR             TIGL-----PNWDWTMPMDHLPELAT3ETYLDPVTGETKNNPFHHAQVAFENGV--TSRN     581
tr|Q6KC55|Q6KC55_MEGCR             TIGL-----PNWDWTMPMDHLPELAT3ETYLDPVTGETKNNPFHHAQVAFENGV--TSRN     581
tr|Q53IP9|Q53IP9_MEGCR             TIGIPYWGLPYWDWTEPMTHIPGLAGNKTYVDS-HGASHTNPFHSSVIAFEENAPHTKRQ     566
4BED:A-B|PDBID|CHAIN|SEQUENCE      TIGIP-----YWDWTEPMTHIPGLAGNKTYVDS-HGASHTNPFHSSVIAFEENAPHTKRQ     561
                                    **:      **** *  *:* ** . :* *        **** : :** .   *.*:

slipper_Hc1_-_CDS_translation      VQDDLSESPEFGDHTSLFDGILLAFEQTDFCDFEVQFEVVHNAPHFLVGGFGAYTLATLH     651
tr|Q1MVA1|Q1MVA1_MEGCR             PDAKLFMKPTYGDHTYLFD3MIYAFEQEDFCDFEVQYELTHNAIHAWVGGSEKYSM3SLH     641
tr|Q6KC55|Q6KC55_MEGCR             PDAKLFMKPTYGDHTYLFD3MIYAFEQEDFCDFEVQYELTHNAIHAWVGGSEKYSM3SLH     641
tr|Q53IP9|Q53IP9_MEGCR             IDQRLFKPATFGHHTDLFNQILYAFEQEDYCDFEVQFEITHNTIHAWTGGSEHFSM3SLH     626
4BED:A-B|PDBID|CHAIN|SEQUENCE      IDQRLFKPATFGHHTDLFNQILYAFEQEDYCDFEVQFEITHNTIHAWTGGSEHFSM3SLH     621
                                    : *   :*.** **: :: **** *:******:*:.**: *  .** :  :::::**

slipper_Hc1_-_CDS_translation      YSAFDPIFYLHH3NVDRLWAIWQQLQIKRGLPYKAHCASSLTQEVLKPFGFPPPWNNDHK     711
tr|Q1MVA1|Q1MVA1_MEGCR             YTAFDPIFYLHH3NVDRLWAIWQALQIRRGKSYKAHCASSQEREPLKPFAFSSPLNNNEK     701
tr|Q6KC55|Q6KC55_MEGCR             YTAFDPIFYLHH3NVDRLWAIWQALQIRRGKSYKAHCASSQEREPLKPFAFSSPLNNNEK     701
tr|Q53IP9|Q53IP9_MEGCR             YTAFDPLFYFHH3NVDRLWAVWQALQMRRHKPYRAHCAISLEHMHLKPFAFSSPLNNNEK     686
4BED:A-B|PDBID|CHAIN|SEQUENCE      YTAFDPLFYFHH3NVDRLWAVWQALQMRRHKPYRAHCAISLEHMHLKPFAFSSPLNNNEK     681
                                   *:****:**:**:**********:** **::*   *:**** * :    **** .*  * **:.*

slipper_Hc1_-_CDS_translation      TFENARPTNIYDYEAVLDYTYDSLQFGGMTVDQLEHYLHERKTQNRDFVGIALHNIGVSA     771
tr|Q1MVA1|Q1MVA1_MEGCR             TYHN3VPTNVYDYVGVLHYRYDDLQFGGMTM3ELEEYIHKQTQHDRTFAGFFL3YIGTSA     761
tr|Q6KC55|Q6KC55_MEGCR             TYHN3VPTNVYDYVGVLHYRYDDLQFGGMTM3ELEEYIHKQTQHDRTFAGFFL3YIGTSA     761
tr|Q53IP9|Q53IP9_MEGCR             THANAMPNKIYDYENVLHYTYEDLTFGGISLENIEKMIHENQQEDRIYAGFLLAGIRTSA     746
4BED:A-B|PDBID|CHAIN|SEQUENCE      THANAMPNKIYDYENVLHYTYEDLTFGGISLENIEKMIHENQQEDRIYAGFLLAGIRTSA     741
                                   *. *: *.::*** **.* *:.* ***:::.::*. :*:.  .:* :.*: * * .**
```

```
slipper_Hc1_-_CDS_translation   WVTLSLQVEGGEPYQVGKIAVLGGEKEMPWRFDRLFKVEITSALQKLGLSYDDDFNVHLD    831
tr|Q1MVA1|Q1MVA1_MEGCR          SVDIFINREGHDKYKVGSFVVLGGSKEMKWGFDRMYKYEITEALKTLNVAVDDGFSITVE    821
tr|Q6KC55|Q6KC55_MEGCR          SVDIFINREGHDKYKVGSFVVLGGSKEMKWGFDRMYKYEITEALKTLNVAVDDGFSITVE    821
tr|Q53IP9|Q53IP9_MEGCR          NVDIFIKTTDSVQHKAGTFAVLGGSKEMKWGFDRVFKFDITHVLKDLDLTADGDFEVTVD    806
4BED:A-B|PDBID|CHAIN|SEQUENCE   NVDIFIKTTDSVQHKAGTFAVLGGSKEMKWGFDRVFKFDITHVLKDLDLTADGDFEVTVD    801
                                *  :  ::   _    ::.*.:.****.*** * ***::* :** .*: *.:: *.*..: ::


slipper_Hc1_-_CDS_translation   ITDVTGKGKWAEDTFHHNTIIHVPGEGHEDPG-----EDDDHVRNDIVTLTAEQEQNVRDA    886
tr|Q1MVA1|Q1MVA1_MEGCR          ITDVDGSPPSADLIPPPAIIFERADA-------KDFGHSRKIRKAVDSLTVEEQTSLRRA    874
tr|Q6KC55|Q6KC55_MEGCR          ITDVDGSPPSADLIPPPAIIFERADA-------KDFGHSRKIRKAVDSLTVEEQTSLRRA    874
tr|Q53IP9|Q53IP9_MEGCR          ITEVDGTKLASSLIPHASVIREHARGKLNRVKFDKVPRSRLIRKNVDRLSPEEMNELRKA    866
4BED:A-B|PDBID|CHAIN|SEQUENCE   ITEVDGTKLASSLIPHASVIREHAR-----VKFDKVPRSRLIRKNVDRLSPEEMNELRKA    856
                                **:* *. : : .  ::*   _      . :*: :  *: *:  ::* *


slipper_Hc1_-_CDS_translation   FQGLKNDKSVAGYNQIAAFHGQPNWCPSTQAERKYACCTHGMAVFPHWHRLLTVQAENGL    946
tr|Q1MVA1|Q1MVA1_MEGCR          MADLQDDKTSGGFQQIAAFHGEPKWCPSPEAEKKFACCVHGMAVFPHWHRLLTVQGENAL    934
tr|Q6KC55|Q6KC55_MEGCR          MADLQDDKTSGGFQQIAAFHGEPKWCPSPEAEKKFACCVHGMAVFPHWHRLLTVQGENAL    934
tr|Q53IP9|Q53IP9_MEGCR          LALLKEDKSAGGFQQLGAFHGEPKWCPSPEASKKFACCVHGMSVFPHWHRLLTVQSENAL    926
4BED:A-B|PDBID|CHAIN|SEQUENCE   LALLKEDKSAGGFQQLGAFHGEPKWCPSPEASKKFACCVHGMSVFPHWHRLLTVQSENAL    916
                                :  *::**: .*::*.:****:*:**** :*.:*:***.***:**********:.**.*


slipper_Hc1_-_CDS_translation   IANGLHSGLPYWDWTLPMTSLPSIVLNATYVHPKTGESVANPLYSGQVD--GHDTTRSVR    1004
tr|Q1MVA1|Q1MVA1_MEGCR          RKHGFTGGLPYWDWTRSMSALPHFVADPTYNDAISSQEEDNPWHHGHIDSVGHDTTRDVR    994
tr|Q6KC55|Q6KC55_MEGCR          RKHGFTGGLPYWDWTRSMSALPHFVADPTYNDAISSQEEDNPWHHGHIDSVGHDTTRDVR    994
tr|Q53IP9|Q53IP9_MEGCR          RRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNPWFDGHIDTVDKTTTRSVQ    986
4BED:A-B|PDBID|CHAIN|SEQUENCE   RRHGYDGALPYWDWTSPLNHLPELADHEKYVDPEDGVEKHNPWFDGHIDTVDKTTTRSVQ    976
                                 :*  ..*******   :. **  :  . .*    . **  *.::*  .: ***.*:


slipper_Hc1_-_CDS_translation   QELFEQPKFGHMTKIAEKVMLAFEQDNFCDFEIQYEIAHNYIHALVGGNETYSMASLRYT    1064
tr|Q1MVA1|Q1MVA1_MEGCR          DDLYQSPGFGHYTDIAKQVLLAFEQDDFCDFEVQFEIAHNFIHALVGGNEPYSMSSLRYT    1054
tr|Q6KC55|Q6KC55_MEGCR          DDLYQSPGFGHYTDIAKQVLLAFEQDDFCDFEVQFEIAHNFIHALVGGNEPYSMSSLRYT    1054
tr|Q53IP9|Q53IP9_MEGCR          NKLFEQPEFGHYTSIAKQVLLALEQDNFCDFEIQYEIAHNYIHALVGGAQPYGMASLRYT    1046
4BED:A-B|PDBID|CHAIN|SEQUENCE   NKLFEQPEFGHYTSIAKQVLLALEQDNFCDFEIQYEIAHNYIHALVGGAQPYGMASLRYT    1036
                                ::.*::.* *** *.**::*:**:***:*****:*:***** :*****:*****.:  * .*.:*****


slipper_Hc1_-_CDS_translation   AYDPIFFLHHSNTDRIWAIWQALQKYRGKPYNSANCAIAEMRKPLQPFAQTSLTNPDYVT    1124
tr|Q1MVA1|Q1MVA1_MEGCR          TYDPIFFLHRSNTDRLWAIWQALQKYRGKPYNTANCAIASMRKPLQPFGLDSVINPDDET    1114
tr|Q6KC55|Q6KC55_MEGCR          TYDPIFFLHRSNTDRLWAIWQALQKYRGKPYNTANCAIASMRKPLQPFGLDSVINPDDET    1114
tr|Q53IP9|Q53IP9_MEGCR          AFDPLFYLHHSNTDRIWAIWQALQKYRGKPYNVANCAVTSMREPLQPFGLSANINTDHVT    1106
4BED:A-B|PDBID|CHAIN|SEQUENCE   AFDPLFYLHHSNTDRIWAIWQALQKYRGKPYNVANCAVTSMREPLQPFGLSANINTDHVT    1096
                                ::**:*:**:*****:*************** ****::.**:*****. :   * * *


slipper_Hc1_-_CDS_translation   RDHSTPFDVFNYHSSFHYHYDNLDFNGMSVAQLQREVIRRRGLERAFAGFMLHGVKKSCL    1184
tr|Q1MVA1|Q1MVA1_MEGCR          REHSVPFRVFDYKNNFDYEYESLAFNGLSIAQLDRELQRRKSHDRVFAGFLLHEIGQSAL    1174
tr|Q6KC55|Q6KC55_MEGCR          REHSVPFRVFDYKNNFDYEYESLAFNGLSIAQLDRELQRRKSHDRVFAGFLLHEIGQSAL    1174
tr|Q53IP9|Q53IP9_MEGCR          KEHSVPFNVFDYKTNFNYEYDTLEFNGLSISQLNKKLEAIKSQDRFFAGFLLSGFKKSSL    1166
4BED:A-B|PDBID|CHAIN|SEQUENCE   KEHSVPFNVFDYKTNFNYEYDTLEFNGLSISQLNKKLEAIKSQDRFFAGFLLSGFKKSSL    1156
                                ::**.** **:*:.  *_*.*.*: * ***.*::*#:::    :. :* ****:*   :*.*


slipper_Hc1_-_CDS_translation   VVFDICKPDG--TCTKSGEFYLLGDENELPWEYDRLYKYEITHELEDMGLEPQDRFDVQY    1242
tr|Q1MVA1|Q1MVA1_MEGCR          VKFYVCKHNVSDCDHYAGEFYILGDEAEMPWRYDRVYKYEITQQLHDLDLHVGDNFFLKY    1234
tr|Q6KC55|Q6KC55_MEGCR          VKFYVCKHNVSDCDHYAGEFYILGDEAEMPWRYDRVYKYEITQQLHDLDLHVGDNFFLKY    1234
tr|Q53IP9|Q53IP9_MEGCR          VKFNICTDSS--NCHPAGEFYLLGDENEMPWAYDRVFKYDITEKLHDLKLHAEDHFYIDY    1224
4BED:A-B|PDBID|CHAIN|SEQUENCE   VKFNICTDSS--NCHPAGEFYLLGDENEMPWAYDRVFKYDITEKLHDLKLHAEDHFYIDY    1214
                                * * :*.  _      :****:**** *:** ***:::**::**.:*.*: *.  *.* :.*


slipper_Hc1_-_CDS_translation   HVYDLDNTDLGDDVFGKAVIIYSAGQGHQKGHEEDYLEEVQASSHVRRNLEDLTTGEAES    1302
tr|Q1MVA1|Q1MVA1_MEGCR          EAFDLNGGSLGGSIFSQPSVIFEPAAGSHQA--DEYREAVTSASHIRKNIRDLSEGEIES    1292
tr|Q6KC55|Q6KC55_MEGCR          EAFDLNGGSLGGSIFSQPSVIFEPAAGSHQA--DEYREAVTSASHIRKNIRDLSEGEIES    1292
tr|Q53IP9|Q53IP9_MEGCR          EVFDLKPASLGKDLFKQPSVIHEPRIGHHEG--EVYQAEVTSANRIRKNIENLSLGELES    1282
4BED:A-B|PDBID|CHAIN|SEQUENCE   EVFDLKPASLGKDLFKQPSVIHEPRIGHHEG--EVYQAEVTSANRIRKNIENLSLGELES    1272
                                 ..:**.  .** .:* :  :*..   * ::   : *   * :::*:*:::*: ** **


slipper_Hc1_-_CDS_translation   LRSALRDMEEDGSFDAIARFHGYPGLCEHDGHKAACCVHGSPAFPHWHRLYVEQVENSLL    1362
tr|Q1MVA1|Q1MVA1_MEGCR          IRSAFLQIQKEGIYENIAKFHGKPGLCEHDGHPVACCVHGMPTFPHWHRLYVLQVENALL    1352
tr|Q6KC55|Q6KC55_MEGCR          IRSAFLQIQKEGIYENIAKFHGKPGLCEHDGHPVACCVHGMPTFPHWHRLYVLQVENALL    1352
tr|Q53IP9|Q53IP9_MEGCR          LRAAFLEIENDGTYESIAKFHGSPGLCQLNGNPISCCVHGMPTFPHWHRLYVVVVENALL    1342
4BED:A-B|PDBID|CHAIN|SEQUENCE   LRAAFLEIENDGTYESIAKFHGSPGLCQLNGNPISCCVHGMPTFPHWHRLYVVVVENALL    1332
                                :*:*: :::::* :: **:*** ****: _*:  :***** *:******** .**:**


slipper_Hc1_-_CDS_translation   SHGSAVSVPYWDWTQPIKKLPALINDATYYDSRAHAKLENPFFRWKIPGTDMYTSRDPRP    1422
tr|Q1MVA1|Q1MVA1_MEGCR          ERGSAVAVPYWDWTEKADSLPSLINDATYFNSRSQTFDPNPFFRGHIAFENAVTSRDPQP    1412
tr|Q6KC55|Q6KC55_MEGCR          ERGSAVAVPYWDWTEKADSLPSLINDATYFNSRSQTFDPNPFFRGHIAFENAVTSRDPQP    1412
tr|Q53IP9|Q53IP9_MEGCR          KKGSSVAVPYWDWTKRIEHLPHLISDATYYNSRQHHYETNPFHHGKITHENEITTRDPKD    1402
4BED:A-B|PDBID|CHAIN|SEQUENCE   KKGSSVAVPYWDWTKRIEHLPHLISDATYYNSRQHHYETNPFHHGKITHENEITTRDPKD    1392
                                .:**:**:*******:  _ ** **.****::** :*****  *:  :*   : *:***:


slipper_Hc1_-_CDS_translation   DLFDSDYFLNNALLALEQTSYCDFEVQFEILHNALHSFLGGRGKYSLSSLDYSAFDPVFF    1482
tr|Q1MVA1|Q1MVA1_MEGCR          ELWDNKDFYENVMLALEQDNFCDFEIQLELIHNALHSRLGGRAKYSLSSLDYTAFDPVFF    1472
tr|Q6KC55|Q6KC55_MEGCR          ELWDNKDFYENVMLALEQDNFCDFEIQLELIHNALHSRLGGRAKYSLSSLDYTAFDPVFF    1472
tr|Q53IP9|Q53IP9_MEGCR          SLFHSDYFYEQVLYALEQDNFCDFEIQLEILHNALHSLLGGWGKYSMSNLDYAAFDPVFF    1462
4BED:A-B|PDBID|CHAIN|SEQUENCE   SLFHSDYFYEQVLYALEQDNFCDFEIQLEILHNALHSLLGGWGKYSMSNLDYAAFDPVFF    1452
                                .*:... * ::: **** .:****:*:*:.::***** *** *****:.**.* ***:******


slipper_Hc1_-_CDS_translation   LHHANVDRIWAIWQALQKIRGLPFDESDCALNIMGTPLHPFDDKEENQFDLTNKYSRPID    1542
tr|Q1MVA1|Q1MVA1_MEGCR          LHHANVDRIWAIWQDLQRYRKKPYNEADCAVNEMRKPLQPFNNPELNSDSMTLKHNLPQD    1532
tr|Q6KC55|Q6KC55_MEGCR          LHHANVDRIWAIWQDLQRYRKKPYNEADCAVNEMRKPLQPFNNPELNSDSMTLKHNLPQD    1532
tr|Q53IP9|Q53IP9_MEGCR          LHHATTDRIWAIWQDLQRFRKRPYREANCAIQLMHTPLQPFDKSDNN-DEATKTHATPHD    1521
4BED:A-B|PDBID|CHAIN|SEQUENCE   LHHATTDRIWAIWQDLQRFRKRPYREANCAIQLMHTPLQPFDKSDNN-DEATKTHATPHD    1511
                                ****._******** **:  *  *: *::**::  * .**:**:.  : *  .  *  .: * *
```

```
slipper_Hc1_-_CDS_translation   AFDYSNHFDYHYDTLKFNSWTIPQLEQVLEKQRSRDRVFAGFLLHNIGTSADVEIDVCVA    1602
tr|Q1MVA1|Q1MVA1_MEGCR          SFDYQNRFRYQYDNLQFNHFSIQKLDQTIQARKQHDRVFAGFILHNIGTSAVVDIYICVE    1592
tr|Q6KC55|Q6KC55_MEGCR          SFDYQNRFRYQYDNLQFNHFSIQKLDQTIQARKQHDRVFAGFILHNIGTSAVVDIYICVE    1592
tr|Q53IP9|Q53IP9_MEGCR          GFEYQNSFGYAYDNLELNHYSIPQLDHMLQERKRHDRVFAGFLLHNIGTSADGHVFVCLP    1581
4BED:A-B|PDBID|CHAIN|SEQUENCE   GFEYQNSFGYAYDNLELNHYSIPQLDHMLQERKRHDRVFAGFLLHNIGTSADGHVFVCLP    1571
                                .*:*.* * * **.*::* ::* :*:: :: :: :*******:********   .: :*:

slipper_Hc1_-_CDS_translation   TGNGAKSCNHPAGKFAILGGEYEMPFTFDRLYKYDISDTVRKLGLRLDSAADFDVQIKIF    1662
tr|Q1MVA1|Q1MVA1_MEGCR          QGG-EQNCKTKAGSFTILGGETEMPFHFDRLYKFDITSALHKLGVPLDG-HGFDIKVDVR    1650
tr|Q6KC55|Q6KC55_MEGCR          QGG-EQNCKTKAGSFTILGGETEMPFHFDRLYKFDITSALHKLGVPLDG-HGFDIKVDVR    1650
tr|Q53IP9|Q53IP9_MEGCR          TGEHTKDCSHEAGMFSILGGQTEMSFVFDRLYKLDITKALKKNGVHLQG--DFDLEIEIT    1639
4BED:A-B|PDBID|CHAIN|SEQUENCE   TGEHTKDCSHEAGMFSILGGQTEMSFVFDRLYKLDITKALKKNGVHLQG--DFDLEIEIT    1629
                                *   :.*.  ** *:***: ** * ****** **:::*  *: *:.  :**:::.:

slipper_Hc1_-_CDS_translation   AYNGSYIDASLLHRPTIIFEPGQGQTQDDVGHVERRLVRKSVLSLSPAERRSLVLAMRSL    1722
tr|Q1MVA1|Q1MVA1_MEGCR          AVNGSHLDQHILNEPSLLFVPGERKNIYYDGLSQHNLVRKEVSSLTTLEKHFLRKALKNM    1710
tr|Q6KC55|Q6KC55_MEGCR          AVNGSHLDQHILNEPSLLFVPGERKNIYYDGLSQHNLVRKEVSSLTTLEKHFLRKALKNM    1710
tr|Q53IP9|Q53IP9_MEGCR          AVNGSHLDSHVIHSPTILFEAGTDSAHTDDGHTEPVMIRKDITQLDKRQQLSLVKALESM    1699
4BED:A-B|PDBID|CHAIN|SEQUENCE   AVNGSHLDSHVIHSPTILFEAGTDSAHTDDGHTEPVMIRKDITQLDKRQQLSLVKALESM    1689



                                * ***::*  ::: *:::*  *  _        *  :  ::**.: .*   :: * *:..:

slipper_Hc1_-_CDS_translation   QEDSSADGFQSLASFHALPPLCPYPEATKRFACCIHGMATFPHWHRLYTVQFEDALRRHG    1782
tr|Q1MVA1|Q1MVA1_MEGCR          QADDSPDGYQAIASFHALPPLCPSPSAAHRHACCLHGMATFPQWHRLYTVQFEDSLKRHG    1770
tr|Q6KC55|Q6KC55_MEGCR          QADDSPDGYQAIASFHALPPLCPSPSAAHRHACCLHGMATFPQWHRLYTVQFEDSLKRHG    1770
tr|Q53IP9|Q53IP9_MEGCR          KADHSSDGFQAIASFHALPPLCPSPAASKRFACCVHGMATFPQWHRLYTVQFQDSLRKHG    1759
4BED:A-B|PDBID|CHAIN|SEQUENCE   KADHSSDGFQAIASFHALPPLCPSPAASKRFACCVHGMATFPQWHRLYTVQFQDSLRKHG    1749
                                : * **:*::********** * *::*.*** :******:*:********.*:*::**

slipper_Hc1_-_CDS_translation   ALVGIPYWDTVVPNSRLPDFIAESVWDDPLFHANFSNPWAGADIEFDNSAV--VRDVNMD    1840
tr|Q1MVA1|Q1MVA1_MEGCR          SIVGLPYWDWLKPQSALPDLVTQETYEHLFSHKTFPNPFLKANIEFEGEGVTTERDVDAE    1830
tr|Q6KC55|Q6KC55_MEGCR          SIVGLPYWDWLKPQSALPDLVTQETYEHLFSHKTFPNPFLKANIEFEGEGVTTERDVDAE    1830
tr|Q53IP9|Q53IP9_MEGCR          AVVGLPYWDWTLPRSELPELLTVSTIHDPETGRDIPNPFIGSKIEFEGENVHTKRDINRD    1819
4BED:A-B|PDBID|CHAIN|SEQUENCE   AVVGLPYWDWTLPRSELPELLTVSTIHDPETGRDIPNPFIGSKIEFEGENVHTKRDINRD    1809
                                ::**:****  *.* **::: ..  ..       : **: :.:***::.. *   **:: :

slipper_Hc1_-_CDS_translation   RISQKGPKGYDTWSWKQYLYALEQENFCDFEVQFEIAHNALHAWMGGSEVHSMAHLHYAS    1900
tr|Q1MVA1|Q1MVA1_MEGCR          HLFAKGNLVYNNWFCNQALYALEQENYCDFEIQFEILHNGIHSWVGGSKTHSIGHLHYAS    1890
tr|Q6KC55|Q6KC55_MEGCR          HLFAKGNLVYNNWFCNQALYALEQENYCDFEIQFEILHNGIHSWVGGSKTHSIGHLHYAS    1890
tr|Q53IP9|Q53IP9_MEGCR          RLFQGSTKTHHNWFIEQALLALEQTNYCDFEVQFEIMHNGVHTWVGGKEPYGIGHLHYAS    1879
4BED:A-B|PDBID|CHAIN|SEQUENCE   RLFQGSTKTHHNWFIEQALLALEQTNYCDFEVQFEIMHNGVHTWVGGKEPYGIGHLHYAS    1869
                                ::   . ::.*  :* * **** *:****:**** **.:*:*:**.:  ::.******

slipper_Hc1_-_CDS_translation   YDPVFILHHSNTDRIFALWQELQKYRGRDANEVNCALELMKEPLKPFSFGSPYNLNPTTT    1960
tr|Q1MVA1|Q1MVA1_MEGCR          YDPLFYIHHSQTDRIWAIWQALQEHRGLSGKEAHCALEQMKDPLKPFSFGSPYNLNKRTQ    1950
tr|Q6KC55|Q6KC55_MEGCR          YDPLFYIHHSQTDRIWAIWQALQEHRGLSGKEAHCALEQMKDPLKPFSFGSPYNLNKRTQ    1950
tr|Q53IP9|Q53IP9_MEGCR          YDPLFYIHHSQTDRIWAIWQSLQRFRGLSGSEANCAVNLMKTPLKPFSFGAPYNLNDHTH    1939
4BED:A-B|PDBID|CHAIN|SEQUENCE   YDPLFYIHHSQTDRIWAIWQSLQRFRGLSGSEANCAVNLMKTPLKPFSFGAPYNLNDHTH    1929
                                ***:* :***:****:*:** **_.** ...*_:.**:: ** *******:*****  *

slipper_Hc1_-_CDS_translation   LYSKPEDAFDYKGHFKYEYDTLELQGLDVQRLQDYINKEKEHDRVFAGFLLSGIGQSAHA    2020
tr|Q1MVA1|Q1MVA1_MEGCR          EFSKPEDTFDYH-RFGYEYDSLEFVGMSVSSLHNYIKQQQEADRVFAGFLLKGFGQSASV    2009
tr|Q6KC55|Q6KC55_MEGCR          EFSKPEDTFDYH-RFGYEYDSLEFVGMSVSSLHNYIKQQQEADRVFAGFLLKGFGQSASV    2009
tr|Q53IP9|Q53IP9_MEGCR          DFSKPEDTFDYQ-KFGYIYDTLEFAGWSIRGIDHIVRNRQEHSRVFAGFLLEGFGTSATV    1998
4BED:A-B|PDBID|CHAIN|SEQUENCE   DFSKPEDTFDYQ-KFGYIYDTLEFAGWSIRGIDHIVRNRQEHSRVFAGFLLEGFGTSATV    1988
                                :*****:***: :* * **:**: *  .:  :.. :.:.:* .*********.*:* ** .

slipper_Hc1_-_CDS_translation   SFSVCKANGECTAAGDFDILGGSAEMPWRFDRLYKYEITDVLEKKGLDVHDSFNITVSLT    2080
tr|Q1MVA1|Q1MVA1_MEGCR          SFDICRPDQSCQEAGYFSVLGGSSEMPWQFDRLYKYDITKTLKDMKLRYDDTFTIKVHIK    2069
tr|Q6KC55|Q6KC55_MEGCR          SFDICRPDQSCQEAGYFSVLGGSSEMPWQFDRLYKYDITKTLKDMKLRYDDTFTIKVHIK    2069
tr|Q53IP9|Q53IP9_MEGCR          DFQVCRTAGDCEDAGYFTVLGGEKEMPWAFDRLYKYDITETLDKMNLRHDEIFQIEVTIT    2058
4BED:A-B|PDBID|CHAIN|SEQUENCE   DFQVCRTAGDCEDAGYFTVLGGEKEMPWAFDRLYKYDITETLDKMNLRHDEIFQIEVTIT    2048
                                .*.:*:   .* ** *:***_ **** ******:**_.*.. *  .: * * *. :.

slipper_Hc1_-_CDS_translation   ALDGSALSSSLLPTPSVIFEPKSRTTELHEVAPNRIRHDLTHLSERDIRSLKSSIRDLQL    2140
tr|Q1MVA1|Q1MVA1_MEGCR          DIAGAELDSDLIPTPSVLLEEGKHGINVRHVGRNRIRMELSELTERDLASLKSAMRSLQA    2129
tr|Q6KC55|Q6KC55_MEGCR          DIAGAELDSDLIPTPSVLLEEGKHGINVRHVGRNRIRMELSELTERDLASLKSAMRSLQA    2129
tr|Q53IP9|Q53IP9_MEGCR          SYDGTVLDSGLIPTPSIIYDPAHHDISSHHLSLNKVRHDLSTLSERDIGSLKYALSSLQA    2118
4BED:A-B|PDBID|CHAIN|SEQUENCE   SYDGTVLDSGLIPTPSIIYDPAHHDISSHHLSLNKVRHDLSTLSERDIGSLKYALSSLQA    2108
                                *: *.*.*:****:: :   :   . :.:. *::* :*: *:*** *** :: .**

slipper_Hc1_-_CDS_translation   DDSNDGYQNIASFHGAPALCPSPEAAEYACCVHGMPTFPHWHRLYTVEMEDAMVRHGSSV    2200
tr|Q1MVA1|Q1MVA1_MEGCR          DDGVNGYQAIASFHGLPASCHDDEGHEIACCIHGMPVFPHWHRLYTLQMDMALLSHGSAV    2189
tr|Q6KC55|Q6KC55_MEGCR          DDGVNGYQAIASFHGLPASCHDDEGHEIACCIHGMPVFPHWHRLYTLQMDMALLSHGSAV    2189
tr|Q53IP9|Q53IP9_MEGCR          DTSADGFAAIASFHGLPAKCNDSHNNEVACCIHGMPTFPHWHRLYTLQFEQALRRHGSSV    2178
4BED:A-B|PDBID|CHAIN|SEQUENCE   DTSADGFAAIASFHGLPAKCNDSHNNEVACCIHGMPTFPHWHRLYTLQFEQALRRHGSSV    2168
                                *  . :*: ****** ** *  .  * ***:**** .********::: *:  ***:*

slipper_Hc1_-_CDS_translation   ALPYWDWTMPITQLPDLFTSETYYDAWRDEVVANPFARAYIKVAGGYTVRDPQAQLKQLS    2260
tr|Q1MVA1|Q1MVA1_MEGCR          AIPYWDWTKPISKLPDLFTSPEYYDPWRDAVVNNPFAKGYIKSEDAYTVRDPQDILYHLQ    2249
tr|Q6KC55|Q6KC55_MEGCR          AIPYWDWTKPISKLPDLFTSPEYYDPWRDAVVNNPFAKGYIKSEDAYTVRDPQDILYHLQ    2249
tr|Q53IP9|Q53IP9_MEGCR          AVPYWDWTKPIHNIPHLFTDKEYYDVWRNKVMPNPFARGYVPSHDTYTVRDVQEGLFHLT    2238
4BED:A-B|PDBID|CHAIN|SEQUENCE   AVPYWDWTKPIHNIPHLFTDKEYYDVWRNKVMPNPFARGYVPSHDTYTVRDVQEGLFHLT    2228
                                *:****** ** ::*.***.  *** **:*: *: ****:*:   . ***** * * :*
```

```
slipper_Hc1_-_CDS_translation   RDGQHSALFDLVLLVLEQTDYCDFEVQFEVVHNAIHYLVGGRQLYSMSSLEYTSYDPIFF   2320
tr|Q1MVA1|Q1MVA1_MEGCR          DETGTSVLLDQTLLALEQTDFCDFEVQFEVVHNAIHYLVGGRQVYALSSQHYASYDPAFF   2309
tr|Q6KC55|Q6KC55_MEGCR          DETGTSVLLDQTLLALEQTDFCDFEVQFEVVHNAIHYLVGGRQVYALSSQHYASYDPAFF   2309
tr|Q53IP9|Q53IP9_MEGCR          STGEHSALLNQALLALEQHDYCDFAVQFEVMHNTIHYLVGGPQVYSLSSLHYASYDPIFF   2298
4BED:A-B|PDBID|CHAIN|SEQUENCE   STGEHSALLNQALLALEQHDYCDFAVQFEVMHNTIHYLVGGPQVYSLSSLHYASYDPIFF   2288
                                 *.*:: .**.*** *:*** *****:**:****** *:*::** _*:**** **


slipper_Hc1_-_CDS_translation   VHHSFVDKIWAVWQELQKRRGMNYDRADCAVNFMNERMHPFDWEALNPDVRTREHSLPQS   2380
tr|Q1MVA1|Q1MVA1_MEGCR          IHHSFVDKIWAVWQALQKKRKRPYHKADCALNMMTKPMRPFAHDF-NHNGFTKMHAVPNT   2368
tr|Q6KC55|Q6KC55_MEGCR          IHHSFVDKIWAVWQALQKKRKRPYHKADCALNMMTKPMRPFAHDF-NHNGFTKMHAVPNT   2368
tr|Q53IP9|Q53IP9_MEGCR          IHHSFVDKVWAVWQALQEKRGLPSDRADCAVSLMTQNMRPFHYEI-NHNQFTKKHAVPND   2357
4BED:A-B|PDBID|CHAIN|SEQUENCE   IHHSFVDKVWAVWQALQEKRGLPSDRADCAVSLMTQNMRPFHYEI-NHNQFTKKHAVPND   2347
                                :*******:***** **::*    .:****:.:*.: *:**  :  *  :  *: *::*:


slipper_Hc1_-_CDS_translation   VWKYEDLGYHYDNYQIGGKSIEELEELIKKQQSHPRVFAGFQLHGLGTSADVELSVCKSQ   2440
tr|Q1MVA1|Q1MVA1_MEGCR          LFDFQDLFYTYDNLEIAGMNVNQLEAEINRRKSQTRVFAGFLLHGIGRSADVRFWICKTA   2428
tr|Q6KC55|Q6KC55_MEGCR          LFDFQDLFYTYDNLEIAGMNVNQLEAEINRRKSQTRVFAGFLLHGIGRSADVRFWICKTA   2428
tr|Q53IP9|Q53IP9_MEGCR          VFKYELLGYRYDNLEIGGMNLHEIEKEIKDKQHHVRVFAGFLLHGIRTSADVQFQICKTS   2417
4BED:A-B|PDBID|CHAIN|SEQUENCE   VFKYELLGYRYDNLEIGGMNLHEIEKEIKDKQHHVRVFAGFLLHGIRTSADVQFQICKTS   2407
                                ::.:: * * *** :*.* .:.::*  *: :: : ****** ***:  ****.: :**:


slipper_Hc1_-_CDS_translation   NSCVSAGVIFILGGKLEMPWAFDRLFKLDITDTLHDMGIEPEDVFDTQAPFFLSYEVHAV   2500
tr|Q1MVA1|Q1MVA1_MEGCR          DDCHASGMIFILGGSKEMHWAYDRNFKYDITQALKAQSIHPEDVFDTDAPFFIKVEVHGV   2488
tr|Q6KC55|Q6KC55_MEGCR          DDCHASGMIFILGGSKEMHWAYDRNFKYDITQALKAQSIHPEDVFDTDAPFFIKVEVHGV   2488
tr|Q53IP9|Q53IP9_MEGCR          EDCHHGGQIFVLGGTKEMAWAYNRLFKYDITHALHDAHITPEDVFHPSEPFFIKVSVTAV   2477
4BED:A-B|PDBID|CHAIN|SEQUENCE   EDCHHGGQIFVLGGTKEMAWAYNRLFKYDITHALHDAHITPEDVFHPSEPFFIKVSVTAV   2467
                                :.* .* **:***. ** **::* ** ***.:*:   * *****. . ***:. .*.*


slipper_Hc1_-_CDS_translation   NGTTLPLSTISPPTLVFQPAEGAAAEHSSYSIAGVGVRKDINTLTAAEMENLRDALGRVQ   2560
tr|Q1MVA1|Q1MVA1_MEGCR          NKTALPSSAIPAPTIIYSAGEGHTDDHGSDHIAGSGVRKDVTSLTASEIENLRHALQSVM   2548
tr|Q6KC55|Q6KC55_MEGCR          NKTALPSSAIPAPTIIYSAGEGHTDDHGSDHIAGSGVRKDVTSLTASEIENLRHALQSVM   2548
tr|Q53IP9|Q53IP9_MEGCR          NGTVLPASILHAPTIIYEPGLDHHEDHHSSSMAGHGVRKEINTLTTAEVDNLKDAMRAVM   2537
4BED:A-B|PDBID|CHAIN|SEQUENCE   NGTVLPASILHAPTIIYEPGLDHHEDHHSSSMAGHGVRKEINTLTTAEVDNLKDAMRAVM   2527
                                * *.** * :  **::: . _   :* *  :** ****::.:**::*::**:.*:  *


slipper_Hc1_-_CDS_translation   AGTGRLTYDNVVSAHGYPPQCT-HDGHKVACCQHGMASFPMWHRLFTRQMEVALSWEGAK   2619
tr|Q1MVA1|Q1MVA1_MEGCR          DDDGPNGFQAIAAYHGSPPMCHMXDGRDVACCTHGMASFPHWHRLFVKQMEDALAAHGAH   2608
tr|Q6KC55|Q6KC55_MEGCR          DDDGPNGFQAIAAYHGSPPMCHMHDGRDVACCTHGMASFPHWHRLFVKQMEDALAAHGAH   2608

tr|Q53IP9|Q53IP9_MEGCR          ADHGPNGYQAIAAFHGNPPMCPMPDGKNYSCCTHGMATFPHWHRLYTKQMEDALTAHGAR   2597
4BED:A-B|PDBID|CHAIN|SEQUENCE   ADHGPNGYQAIAAFHGNPPMCPMPDGKNYSCCTHGMATFPHWHRLYTKQMEDALTAHGAR   2587
                                . *   :: :.: ** ** *    **:. :** ****:** ****.:;*** **: .**:


slipper_Hc1_-_CDS_translation   VGIPYWDWTEAFTELPALVREEENNPFHHGRIPGTDTVTTRAPRPQLFRDPEHGDESFFY   2679
tr|Q1MVA1|Q1MVA1_MEGCR          IGIPYWDWTSAFSHLPALVTDHEHNPFHHGHIAHRNVDTSRSPRDMLFNDPEHGSESFFY   2668
tr|Q6KC55|Q6KC55_MEGCR          IGIPYWDWTSAFSHLPALVTDHEHNPFHHGHIAHRNVDTSRSPRDMLFNDPEHGSESFFY   2668
tr|Q53IP9|Q53IP9_MEGCR          VGLPYWDGTTAFTALPTFVTDEEDNPFHHGHIDYLGVDTTRSPRDKLFNDPERGSESFFY   2657
4BED:A-B|PDBID|CHAIN|SEQUENCE   VGLPYWDGTTAFTALPTFVTDEEDNPFHHGHIDYLGVDTTRSPRDKLFNDPERGSESFFY   2647
                                :*:**** * **: **::* :.*.******:*   .. *:*:** **.***:*.*****


slipper_Hc1_-_CDS_translation   RQVQLALEQRDYCDFEVQFEVIHNAIHSWIGGTSPYGMSTLEYSAYDPVFFIHHSNVDRQ   2739
tr|Q1MVA1|Q1MVA1_MEGCR          RQVLLALEQTDFCQFEVQFEITHNAIHSWTGGHTPYGMSSLEYTAYDPLFYLHHSNTDRI   2728
tr|Q6KC55|Q6KC55_MEGCR          RQVLLALEQTDFCQFEVQFEITHNAIHSWTGGHTPYGMSSLEYTAYDPLFYLHHSNTDRI   2728
tr|Q53IP9|Q53IP9_MEGCR          RQVLLALEQTDFCQFEVQFEITHNAIHSWTGGLTPYGMSTLEYTTYDPLFWLHHANTDRI   2717
4BED:A-B|PDBID|CHAIN|SEQUENCE   RQVLLALEQTDFCQFEVQFEITHNAIHSWTGGLTPYGMSTLEYTTYDPLFWLHHANTDRI   2707
                                *** ***** *:*:******: ******* ** :*****:***::***:*::**:*.**


slipper_Hc1_-_CDS_translation   FAIWQALQKYRGLDYNTANCNIQELRQVQEPFDRDDNPVVTTRHYSKAIDAFNYDQYGYQ   2799
tr|Q1MVA1|Q1MVA1_MEGCR          WAIWQALQKYRGFQYNAAHCDIQVLKQPLKPFSESRNPNPVTRANSRAVDSFDYERLNYQ   2788
tr|Q6KC55|Q6KC55_MEGCR          WAIWQALQKYRGFQYNAAHCDIQVLKQPLKPFSESRNPNPVTRANSRAVDSFDYERLNYQ   2788
tr|Q53IP9|Q53IP9_MEGCR          WAIWQALQEYRGLPYDHANCEIQAMKRPLRPFSDPINHNAFTHSNAKPTDVFEYSRFNFQ   2777
4BED:A-B|PDBID|CHAIN|SEQUENCE   WAIWQALQEYRGLPYDHANCEIQAMKRPLRPFSDPINHNAFTHSNAKPTDVFEYSRFNFQ   2767
                                :*******:***: *: *:*:** ::: .**.  *   *: :: * *:*.: .:*


slipper_Hc1_-_CDS_translation   YDNLNFHGMTISQLDEMLEKKKQEDHVFANFMLHGIQTSADVVFDLCDAKGKCNFAGTFA   2859
tr|Q1MVA1|Q1MVA1_MEGCR          YDTLTFHGHSISELDAMLQERKKEERTFAAFLLHGFGASADVSFDVCTPDGHCAFAGTFA   2848
tr|Q6KC55|Q6KC55_MEGCR          YDTLTFHGHSISELDAMLQERKKEERTFAAFLLHGFGASADVSFDVCTPDGHCAFAGTFA   2848
tr|Q53IP9|Q53IP9_MEGCR          YDNLRFHGMTIKKLEHELEKQKEEDRTFAAFLLHGIKKSADVSFDVCNHDGECHFAGTFA   2837
4BED:A-B|PDBID|CHAIN|SEQUENCE   YDNLRFHGMTIKKLEHELEKQKEEDRTFAAFLLHGIKKSADVSFDVCNHDGECHFAGTFA   2827
                                **.* *** :*.:*: *:::*:*::.** *:***: **** **:*  .*.* ******


slipper_Hc1_-_CDS_translation   ILGGPLEMPWSFDRLFKYDVTSVFKQMRLRPDSEYSFRVSLTAVNGTQLDSRLIEAPSVS   2919
tr|Q1MVA1|Q1MVA1_MEGCR          VLGGELEMPWSFERLFRYDITKVLKQMNLHYDSEFHFELKIVGTDGTELPSDRIKSPTIE   2908
tr|Q6KC55|Q6KC55_MEGCR          VLGGELEMPWSFERLFRYDITKVLKQMNLHYDSEFHFELKIVGTDGTELPSDRIKSPTIE   2908
tr|Q53IP9|Q53IP9_MEGCR          ILGGEHEMPWSFDRLFRYDITQVLKQMHLEYDSDFTFHMRIIDTSGKQLPSDLIKMPTVE   2897
4BED:A-B|PDBID|CHAIN|SEQUENCE   ILGGEHEMPWSFDRLFRYDITQVLKQMHLEYDSDFTFHMRIIDTSGKQLPSDLIKMPTVE   2887
                                :*** ******:***:**:*.*.*:***.*. **:: *.: :  ..*.:* *  *: *::


slipper_Hc1_-_CDS_translation   FVPGNKPGSKTSAAHEDPVPLDTGDVTRYEVSALSLAQVTNLRDALYKLQNDHGPNGFEA   2979
tr|Q1MVA1|Q1MVA1_MEGCR          HHGGDHHGGDTSGHDHSER---HDGFFRKEVGSLSLDEANDLKNALYKLQNDQGPNGYES   2965
tr|Q6KC55|Q6KC55_MEGCR          HHGGDHHGGDTSGHDHSER---HDGFFRKEVGSLSLDEANDLKNALYKLQNDQGPNGYES   2965
tr|Q53IP9|Q53IP9_MEGCR          HSPGGKHHEK---HHEDHH---EDILVRKNIHSLSHHEAEELRDALYKLQNDESHGGYEH   2951
4BED:A-B|PDBID|CHAIN|SEQUENCE   HSPGGKHHEK---HHEDHH---EDILVRKNIHSLSHHEAEELRDALYKLQNDESHGGYEH   2941
                                .  *.:  -     ...      . . * :: :** :. :*::********..  .*:*
```

```
slipper_Hc1_-_CDS_translation    IASFHGAPGLCPENATDHYACCRHGMPAFPHWHRLLTVQFERALKDKGAVVGVPYWDWTR    3039
tr|Q1MVA1|Q1MVA1_MEGCR           IAGYHGYPFLCPEHGEDQYACCVHGMPVFPHWHRLHTIQFERALKEHGSHLGIPYWDWTK    3025
tr|Q6KC55|Q6KC55_MEGCR           IAGYHGYPFLCPEHGEDQYACCVHGMPVFPHWHRLHTIQFERALKEHGSHLGIPYWDWTK    3025
tr|Q53IP9|Q53IP9_MEGCR           IAGFHGYPNLCPEKGDEKYPCCVHGMSIFPHWHRLHTIQFERALKKHGSHLGIPYWDWTQ    3011
4BED:A-B|PDBID|CHAIN|SEQUENCE    IAGFHGYPNLCPEKGDEKYPCCVHGMSIFPHWHRLHTIQFERALKKHGSHLGIPYWDWTQ    3001
                                 **.:** * ****:.  ::* ** ***  ****** *:***** .:*.  :*:******

slipper_Hc1_-_CDS_translation    PAKAMPSLFTDSYDNNPFQTYRMTFNDQYIQRDVSEELFNHPSEGDVESLFHQALETLEE    3099
tr|Q1MVA1|Q1MVA1_MEGCR           SMIALPAFFADSSNSNPFYKYHIMKAGHDTARSPSDLLFNQPQLHGYDYLYYLALSTLEE    3085
tr|Q6KC55|Q6KC55_MEGCR           SMIALPAFFADSSNSNPFYKYHIMKAGHDTARSPSDLLFNQPQLHGYDYLYYLALSTLEE    3085
tr|Q53IP9|Q53IP9_MEGCR           TISSLPTFFADSGNNNPFFKYHIRSINQDTVRDVNEAIFQQTKFGEFSSIFYLALQALEE    3071
4BED:A-B|PDBID|CHAIN|SEQUENCE    TISSLPTFFADSGNNNPFFKYHIRSINQDTVRDVNEAIFQQTKFGEFSSIFYLALQALEE    3061
                                 ::*::*:*:** :.*** .*::   .:    *. .: :*:: .   . ::: **:***

slipper_Hc1_-_CDS_translation    TNYCDFEVQYEMLHNAVHELIGGGNTYSMSTLEYSAFDPFFMVHHASIDRIWQIWQSLQK    3159
tr|Q1MVA1|Q1MVA1_MEGCR           DNYCDFEVHYEILHNAVHLWLGGTETYSMSSLAFSAYDPVFMILHSGLDRLWIIWQELQK    3145
tr|Q6KC55|Q6KC55_MEGCR           DNYCDFEVHYEILHNAVHLWLGGTETYSMSSLAFSAYDPVFMILHSGLDRLWIIWQELQK    3145
tr|Q53IP9|Q53IP9_MEGCR           DNYCDFEVHYEILHNAVHALWLGGAEKYSMSTLEYSAFDPYFMIHHASLDKIWIIWQELQK    3131
4BED:A-B|PDBID|CHAIN|SEQUENCE    DNYCDFEVQYEILHNEVHALIGGAEKYSMSTLEYSAFDPYFMIHHASLDKIWIIWQELQK    3121
                                 ******:**:*** **  :** :.****:* :**:** **: *:.:*:.* ***.***

slipper_Hc1_-_CDS_translation    LRHKPFNYASCASRSLYKPLEPFSYTSLNADPLTLNNAQFVHIFDTAKFHYHYDSLDLNG    3219
tr|Q1MVA1|Q1MVA1_MEGCR           LRKKPYNAAKCAGHMMDEPLHPFNYESANHDSFTRANAKPSTVFDSHKFNYHYDNPDVRG    3205
tr|Q6KC55|Q6KC55_MEGCR           LRKKPYNAAKCAGHMMDEPLHPFNYESANHDSFTRANAKPSTVFDSHKFNYHYDNPDVRG    3205
tr|Q53IP9|Q53IP9_MEGCR           RRVKPAHAGSCAGDIMHVPLHPFNYESVNNDDFTRENSLPNAVVDSHRFNYKYDNLNLHG    3191
4BED:A-B|PDBID|CHAIN|SEQUENCE    RRVKPAHAGSCAGDIMHVPLHPFNYESVNNDDFTRENSLPNAVVDSHRFNYKYDNLNLHG    3181
                                 * **  : ..**.  :  **.**.* * *  :*  **:  :.*: :*:*:**. ::.*

slipper_Hc1_-_CDS_translation    HGVHELHDMIEDMQATSRIFAGFVLSGISTSARVHVDVTRG--EDTVSVGNFYVLGGSSE    3277
tr|Q1MVA1|Q1MVA1_MEGCR           NSIQEISAIIHDLRNQPRVFAGFVLSGIYTSANVKIYLVRE-GHDDENVGSFVVLGGPKE    3264
tr|Q6KC55|Q6KC55_MEGCR           NSIQEISAIIHDLRNQPRVFAGFVLSGIYTSANVKIYLVRE-GHDDENVGSFVVLGGPKE    3264
tr|Q53IP9|Q53IP9_MEGCR           HNIEELEEVLRSLRLKSRVFAGFVLSGIRTTAVVKVYIKSGTDSDDEYAGSFVILGGAKE    3251
4BED:A-B|PDBID|CHAIN|SEQUENCE    HNIEELEEVLRSLRLKSRVFAGFVLSGIRTTAVVKVYIKSGTDSDDEYAGSFVILGGAKE    3241
                                 :.:.*: ::..::  *:*********** *:* *:: :    *   .*.* :*** .*

slipper_Hc1_-_CDS_translation    MPWAYERIYKLDMTEAASKLGLSSESTFHFKLTVTKYDGTALN-VTFPDPVIVKRAANSQ    3336
tr|Q1MVA1|Q1MVA1_MEGCR           MPWAYERIFKYDITEVANRLNMHHDDTFNFRLEVQSYTG-EMVTHHLPEPLIIYRPAKQE    3323
tr|Q6KC55|Q6KC55_MEGCR           MPWAYERIFKYDITEVANRLNMHHDDTFNFRLEVQSYTG-EMVTHHLPEPLIIYRPAKQE    3323
tr|Q53IP9|Q53IP9_MEGCR           MPWAYERLYRFDITETVHNLNLTD-DHVKFRFDLKKYDHTELDASVLPAPIIVRRPNNAV    3310
4BED:A-B|PDBID|CHAIN|SEQUENCE    MPWAYERLYRFDITETVHNLNLTD-DHVKFRFDLKKYDHTELDASVLPAPIIVRRPNNAV    3300
                                 *******::: *:.**.. .*.:   .:*::.: :    :  :* *:*: * :

slipper_Hc1_-_CDS_translation    HDVLVLPLSVANQLPPKIVVRQGTQVVFHASESGVS-SLREVGSYTNSVHCAIPPGQANL    3395
tr|Q1MVA1|Q1MVA1_MEGCR           YDVLVIPLGSGHKLPPKVIVKRGTRIMFHFVDDTVNRPVVDLGSHTALYNCVVPPFTYNG    3383
tr|Q6KC55|Q6KC55_MEGCR           YDVLVIPLGSGHKLPPKVIVKRGTRIMFHFVDDTVNRPVVDLGSHTALYNCVVPPFTYNG    3383
tr|Q53IP9|Q53IP9_MEGCR           FDIIEIPIGKDVNLPPKVVVKRGTKIMFMSVDEAVTTPMLNLGSYTAMFKCKVPPFSFHA    3370
4BED:A-B|PDBID|CHAIN|SEQUENCE    FDIIEIPIGKDVNLPPKVVVKRGTKIMFMSVDEAVTTPMLNLGSYTAMFKCKVPPFSFHA    3360
                                 .*:: :*:.   :****::*::**::*:  :: *.  :  :.::**:*  :* :** :

slipper_Hc1_-_CDS_translation    YDLDVAYSLEAGDYYFTSSDKTKCQQGS-RIQITVDDE        3432
tr|Q1MVA1|Q1MVA1_MEGCR           YELDHAYSLRDGHYYIAGPTKDLCTSGNVRIHIHIEDE        3421
tr|Q6KC55|Q6KC55_MEGCR           YELDHAYSLRDGHYYIAGPTKDLCTSGNVRIHIHIEDE        3421
tr|Q53IP9|Q53IP9_MEGCR           FELGKMYSVESGDYFMTASTTELCNDNNLRIHVHVDDE        3408
4BED:A-B|PDBID|CHAIN|SEQUENCE    FELGKMYSVESGDYFMTASTTELCNDNNLRIHVHVDDE        3398
                                 ::*.  **:. *.*.*::.   .  * ... **:: ::*
```

APPENDIX 11. *Graph shows Fourier shell correlation against Spatial Frequency revealing the resolution of 3D Reconstructed Image produced from Electron Microscopy, graph was produced by EMAN 2.2. Graphical data displays a lower resolution results in reduced similarity between even and uneven particle data sets which would then further correspond to two completely different 3D reconstruction images. At its lowest unmasked resolution of 41.2 Å resolution both even and uneven data sets are seemingly similar, and this has resulted in the reconstruction of the protein at 41.2 Å after this point the reconstructed images of hemocyanin produced by both uneven and even*

data sets are unsimilar and hence a lower more accurate reconstruction was not achieved.

APPENDIX 12. *Below are the stacked projections and class averages of SLH reconstruction data. There is a clear lack of data within the class averages which has directly resulted in the poor reconstruction of SLH and lack of collar structure.*

Stack Projections

Class Averages