Prediction of the three-dimensional structure of *Cornu aspersus* hemocyanin and its characterisation

Mario García-Risco Aguado
Màster de Bioinformàtica i Bioestadística
Àrea de treball final

Elisabeth Ortega Carrasco
Carles Ventura Royo

01/2019

FITXA DEL TREBALL FINAL

Títol del treball:	*Prediction of the three-dimensional structure of Cornu aspersus hemocyanin and its characterisation*
Nom de l'autor:	*Mario García-Risco Aguado*
Nom del consultor/a:	*Elisabeth Ortega Carrasco*
Nom del PRA:	*Carles Ventura Royo*
Data de lliurament (mm/aaaa):	*01/2019*
Titulació o programa:	*Màster de Bioinformàtica i Bioestadística*
Àrea del Treball Final:	*TFM-Bioinformàtica i Bioestadística (Àrea 3)*
Idioma del treball:	*Anglès*
Paraules clau	*Hemocyanin, structure prediction, protein interaction, evolution.*

Resum del Treball (màxim 250 paraules): *Amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball*

El projecte ha intentat estudiar a grans trets la proteïna Hemocianina: a nivell evolutiu, a nivell d'estructura proteica i a nivell d'interacció amb altres proteïnes. Tot això ve motivat pel fet de conèixer en més profunditat aquesta proteïna: primerament, perquè ja s'ha estudiat en altres organismes propers, el que permet fer un treball comparatiu; segon, perquè es tracta d'una metal·loproteïna, les quals son molt complexes gràcies a que van associades a metalls i permeten fer funcions vitals (citocrom C, calmodulina, etc...); finalment, s'està estudiant el seu efecte anti-carcinogen i conèixer la seva estructura pot ajudar a resoldre incògnites de com es produeix aquest efecte. Els passos seguits en aquest treball són: Primer, s'han agafat les parts repetitives de la seqüència d'hemocianina (anomenades unitats funcionals) i s'han alineat amb AliView, s'han fet arbres filogenètics d'aquestes repeticions i de les tres isoformes d'hemocianina que es coneixen fins al moment. Després s'han agafat les unitats funcionals i s'ha predit la seva estructura terciària trobant el millor motlle fent prèviament BLAST. S'han comparat els models d'estructura resultants sobreposant-los amb Chimera. Finalment s'ha predit l'estructura d'una proteïna que es creu que intervé en la síntesi de l'hemocianina i s'ha fet un *docking* de les dues. Les conclusions són que les hemocianines provenen d'un mateix gen ancestral i que les unitats funcionals mantenen la seva estructura proteica. A més, hi ha una possible interacció entre unitat funcional i metal·lotioneïna.

Abstract (in English, 250 words or less):

This project has intended to shed light in to the knowledge of Hemocyanin: from the evolutionary level, protein structure level and its interaction with other proteins. This motivation is given by the peculiarities of this protein: (1) hemocyanin has been studied in close species before and it is possible to compare studies, (2) it is a metalloprotein, which have a lot of scientific interest due to its importance in living organisms due to the chemical properties given by the metal, (3) it has been reported that hemocyanin has anticancer effects and having more information about this protein may help in how this effect is produced.

This essay consists in taking the repetitive parts of the hemocyanin (these parts are called functional units) and they have been aligned with AliView, after that, several phylogenetic trees have been performed (one tree for each hemocyanin isoform). Then, the three-dimensional structure of the FUs have been predicted by SWISS-MODEL taking the best hit of the BLAST as a template. The models have been compared, superposing them with Chimera. Finally, the three-dimensional structure of a metallothionein (MT) have been predicted to perform a docking between a FU and the MT. As conclusions, one can take that all hemocyanin isoforms come from the same ancestor gene, that all the FUs share the most part of its protein structure and that there may be interaction between FUs and MT.

Índex

Llista de figures

1. Introducció

1.1 Context i justificació del Treball

En aquest treball s'intenta aprofundir en el coneixement bàsic de la biologia i la química de les hemocianines. Aquestes proteïnes transportadores d'oxigen que fan servir alguns invertebrats tenen un interès científic a en els àmbits de la immunologia i el càncer. En aquest projecte, es buscarà conèixer l'origen evolutiu de les diferents isoformes que existeixen de la proteïna i es predirà la seva estructura a partir d'altres estructures ja resoltes. Per últim s'intentarà resoldre la interacció d'aquesta proteïna amb una metal·lotioneïna (proteïna transportadora de coure) que es creu que té un paper clau en la renovació i síntesi de l'hemocianina.

1.2 Objectius del Treball

(1) Determinar l'origen evolutiu de les FU de l'hemocianina de *Cantareus aspersus*

(2) Predicció de l'estructura tridimensional de les FU i estudi comparatiu de les diferents isoformes

(3) Determinar si hi ha interacció entre les FU i les MTs de coure.

1.3 Enfocament i mètode seguit

Aquest treball consta de tres estratègies: fer alineament de seqüències i arbres filogenètics per trobar respostes al primer objectiu (determinar l'origen evolutiu de l'hemocianina), fer una predicció de l'estructura tridimensional de les unitats funcionals de l'hemocianina a través de programes de modelatge i fer un *docking* de la proteïna predita amb una metal·lotioneïna per constatar si hi ha interacció entre aquestes dues proteïnes.
Aquestes estratègies són les escollides principalment perquè són les tècniques bioinformàtiques que existeixen per estudiar evolució (alineaments), predir proteïnes (modelatge) i estudiar interacció de proteïnes (*docking*). Aquestes tècniques tenen variants i especificitats que s'aniran avaluant al llarg del treball per considerar finalment l'opció més òptima a l'hora d'assolir els objectius.

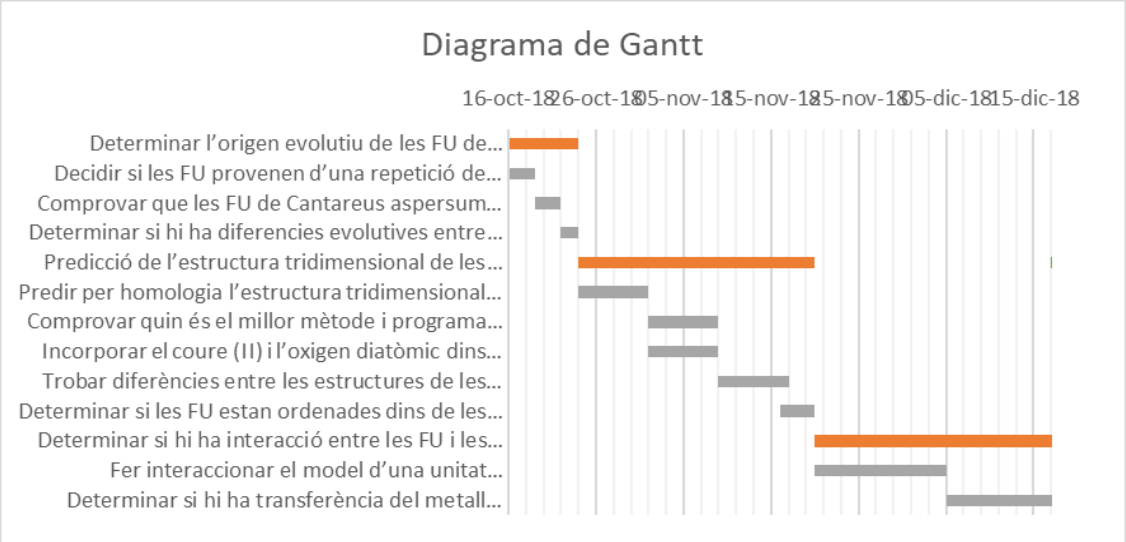
1.4 Planificació del Treball

Els recursos utilitzats fins a l'entrega de la PAC2 són recursos bioinformàtics a nivell de seqüència d'aminoàcids. Caldrà accedir a webs que facin l'alineament de seqüències i que facin la predicció del model tridimensional. Llavors, la fita per la PAC2 serà haver desenvolupat una sèrie d'arbres filogenètics que puguin respondre les preguntes del primer objectiu i haver predit l'estructura tridimensional de les unitats funcionals de l'objectiu dos. Pel que fa a la PAC3 és una feina de Chimera

majoritàriament. On es preparen les estructures per fer el *docking* i la dinàmica molecular. Amb la realització del *docking* es podria considerar assolida la PEC3:

(1) Determinar l'origen evolutiu de les FU de l'hemocianina de *Cantareus aspersus* (16 Octubre 2018 – 23 Octubre 2018)
1.1 Decidir si les FU provenen d'una repetició de seqüències (16 Octubre 2018 – 18 Octubre 2018)
1.2 Comprovar que les FU de Cantareus aspersus son evolutivament més properes a les d'altres mol·luscs terrestres que a mol·luscs aquàtics (19 Octubre 2018 – 21 Octubre 2018)
1.3 Determinar si hi ha diferencies evolutives entre les FU d'hemocianina de mol·lusc terrestre i mol·lusc aquàtic (22 Octubre 2018 – 23 Octubre 2018)
(2) Predicció de l'estructura tridimensional de les FU i estudi comparatiu de les diferents isoformes (24 Octubre 2018 – 19 Novembre 2018)
2.1 Predir per homologia l'estructura tridimensional de les unitats funcionals a través de la seqüència d'aminoàcids (24 Octubre 2018 – 31 Octubre 2018)
2.2 Comprovar quin és el millor mètode i programa per a aquest tipus de proteïnes (1 Novembre 2018 – 8 Novembre 2018)
2.3 Incorporar el coure (II) i l'oxigen diatòmic dins del model (1 Novembre 2018 – 8 Novembre 2018)
2.4 Trobar diferències entre les estructures de les diferents isoformes de les FU (9 Novembre 2018 – 16 Novembre 2018)
2.5 Determinar si les FU estan ordenades dins de les subunitats aleatòriament o hi ha una optimització de l'estructura tridimensional (16 Novembre 2018 – 19 Novembre 2018)
(3) Determinar si hi ha interacció entre les FU i les MTs de coure (20 Novembre 2018 – 17 Desembre 2018)
3.1 Fer interaccionar el model d'una unitat funcional amb una estructura de metal·lotioneïna (20 Novembre 2018 – 4 Desembre 2018)
3.2 Determinar si hi ha transferència del metall entre la metal·lotioneïna i la unitat funcional (5 Desembre 2018 – 17 Desembre 2018)

Finalment, es proporciona el diagrama de Gantt amb la planificació de les tasques a realitzar durant tot el projecte.



1.5 Breu sumari de productes obtinguts

Bàsicament s'obtenen tres resultats d'aquesta memòria: arbres genealògics on es pot interpretar l'origen evolutiu de les seqüències amonioacídiques, les estructures resoltes de la proteïna (les seves subunitats) i la interacció d'aquestes amb la metal·lotioneïna (*docking*).

1.6 Breu descripció dels altres capítols de la memòria

La memòria presenta, a part dels capítols obligatoris en la normativa del treball, un capítol d'introducció on es descriuen els conceptes dels que es parlarà durant la memòria, un capítol de material i mètodes que exposa les tècniques i plataformes utilitzades per realitzar el treball i un capítol de resultats i discussió que representa el cos de la memòria i que descriu tots els resultats obtinguts a partir de la seqüència d'aminoàcids de la proteïna que s'ha estudiat. Després d'exposar els resultats i discutir que poden significar hi ha un apartat de conclusions on es resumeixen els punts més importants del projecte.

2. General Concepts

2.1. Biological concepts

2.1.1. Protein structure levels

Proteins are polymers made by the repetition of amino acids, which the consecutive union of them forms the primary structure of the protein. Secondary structure is known as the particularly stable disposition in the space of some amino acids giving structural patterns[1]. Tertiary structure refers to the three-dimensional folding of the amino acid sequence. Finally, quaternary structure is the arrangement of all the particular subunits that may conform a protein[1].
Then, this project starts with the primary sequence of amino acids of some proteins mentioned below and it is intended to elucidate how these sequences are evolutionary related between them. Besides, it is going to solve its tertiary structure (i.e.: how they are displayed in the space). Thus, these protein structure levels are going to be present all along the project.

2.1.2. *Cantareus aspersus*

Before explaining the proteins of study, here comes a section to put in context of the species is going to be studied. *Cantareus aspersus* is an animal from the Mollusca phylum. It belongs to the group of the pulmonated terrestrial gastropod [2,3]. Which means that uses lungs to breath and, therefore, aerobic respiration. Oxygen is captured by the hemocyanin solved in its haemolymph plasma, that is a liquid that travels all along the whole animal [3]. It is an interesting model to study because it has not been studied before, but its closest species has. Besides, the proteins found in this project are well conserved in molluscs[4], which allows two things: (1) align the sequences with close sequences to perform a robust phylogenetic tree and (2) to perform a high-rated structure prediction due to the sequences high similarity.
Then, the proteins studied in this project are the following.

2.1.2. Hemocyanin

Hemocyanin is an oxygen-transporter protein present in some invertebrates that is characterised, unlike haemoglobin, to present copper in its active site. Hemocyanin is freely solved in *Cantareus* haemolymph plasma[5]. Its binding site consists in six histidine residues that complex two copper ions[3]. These Cu(I) are oxidised to Cu(II) when binding dioxygen molecule, that is reversibly bound to them[3]. It is thought that metallothioneins (see below) are connected with the synthesis of hemocyanin because is this protein that provides with copper to the hemocyanin[6].

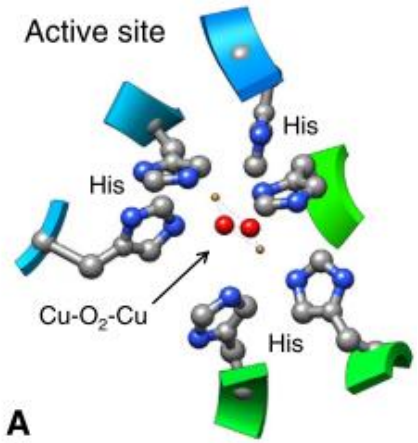


Figure 1. Figure taken from Markl[3]. Active site of hemocyanin, histidine is coordinating two copper ions.

With regards to the protein structure, hemocyanin is a large complex formed by subunits. There are three kinds of subunit: alphaD, alphaN and beta[3,7]. Each subunit is formed by 8 functional units (known as FU, they are coded from *a* to *h*). That means, all FUs are coded in the same gene to form a single amino acid sequence that forms one subunit. Subunits gather in groups of ten forming decamers and these gather in pairs forming dimers[3,7].

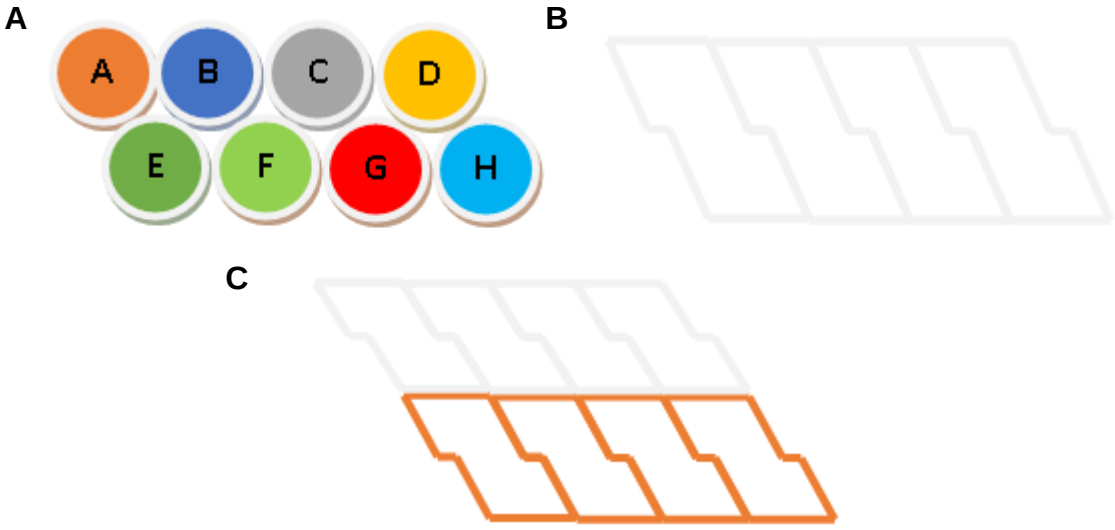


Figure 2. (A) Schematic disposition of the FUs in a subunit. (B) Schematic disposition of the subunit forming decamers. (C) Scheme of a dimer of decamers.

2.1.3. Metallothioneins

Metallothioneins (MT) are ubiquitous proteins found in all eukaryotes and reported in some prokaryotes[8]. They are characterised for being short amino acid sequences rich in cysteines, which confer them the ability of coordinating d^{10} transition metal ions[9]. They are natively bound to Zn(II) and Cu(I)[8], although they have been found to have detoxifying effects due to their ability to bind to heavy metals like Cd(II)[10]. They lack of secondary and three-dimensional structure by themselves, its structure is dictated by metal ions bound to them[8].
MTs have been reported to have an important role in metal homeostasis of the biological systems[11,12]. Since they attach, among other metal ions, copper and hemocyanin contains copper, they are thought to form part of the hemocyanin synthesis[13,14].

2.2. Bioinformatics concepts

Bioinformatics is a quite recent field that uses informatics tools to solve biological problems. By means of computational techniques, mathematics and statistics, biological data can be processed and understood to respond biological questions. The following sections explain the bioinformatics tools used in this project.

2.2.1. Sequence alignments and phylogenetics trees

The first part of this project consists in studying the origin of the different amino acid sequences given to us. This can be performed by doing alignments and phylogenetics trees. An alignment consists in comparing two or more sequences and highlighting their differences and similarities[15].
Once this is known, it is possible to determine a similarity percentage. This term (similarity) is commonly used in genetics analysis and sometimes misused together with the term homology. Similarity is a quantitative measure that can be used as an indication of biological relationship. On the other hand, homology is the fact that two (or more) sequences come from the same ancestral gene[15].
Thus, one can say that similarity gives a well understanding about the homology of the sequences, although it is not always right that a good similarity corresponds to a homology.
Then, with the results of the alignments it is possible to perform a phylogenetic tree to conduct tree-specific analyses [16].
Alignments were performed using AliView. This is a speed tool that performs alignments of large sequences causing a very low memory impact. Since the same program do not perform phylogenetics trees, once the alignment is done, another platform is used to perform the phylogenetics trees.

2.2.2. Structure modelling by homology

Homology modelling is a method that predicts the three-dimensional structure of a certain protein using the solved structure of another similar protein as a template. It is based on the fact that if two proteins share part of the sequence, they likely share part of their three-dimensional structure [15].
To find this similar sequence, BLAST is done. BLAST is a tool provided by NIH that performs a quick alignment with all the sequences present in the NIH database. Those with more similarity are displayed in results. The best hit with

known three-dimensional structure is used as template. The prediction is performed in SWISS-MODEL website, that is a platform dedicated to modelling proteins by homology.

2.2.3. Docking

Docking is a technique used in the last part of this project. It consists in to find interactions between two molecules. The tools analyse these interactions, by searching Van der Waals forces, hydrogen-bridges… Normally, this technique is used to determine if a certain ligand fits in the active site of a protein (normally an enzyme) and the atoms of the side chains of the amino acids interact with the atoms of the ligand. In this case, the interaction is studied for two proteins (hemocyanin and metallothionein). For this reason, ClusPro platform is used to perform this task. The algorithm used in tool is specialised in protein-protein interactions.

3. Materials and methods

This essay works some aspects of the bioinformatics: seeking and matching of sequences, alignment of sequences, modelling of proteins and protein interaction. Each aspect was achieved by using specific tools.

First, alignments of all functional units (FU) were performed to determine whether they have a unique origin, or they come from multiple mutations. To do so, AliView was used. 24 amino acid sequences of *Cantareus aspersus* hemocyanin FU where upload in to the program (8 FU for each 3 of the hemocyanin subunits). After performing the alignment, it was saved in phylip format. This phylip format file was upload in http://www.atgc-montpellier.fr/phyml/ . This portal performs a phylogenetic tree out of an alignment that wil be used to determine the first objective.

Secondly, BLAST tool of NCBI website was used in seeking and matching protein sequences. This tool was used extensively when looking for hemocyanin sequences of other species close to the sequences of reference. When choosing sequences for compiling phylogenetic trees, the same criteria was always followed: BLAST the amino acid sequence of interest, select “Multiple alignment”, select the sequence of reference and the matches with best scoring taking care of not repeating any species and selecting the same isoform when possible (i.e.: when working with alphaN subunit, to select alphaN subunit of all possible species). After selecting the sequences, an online realignment is performed, and a phylogenetic tree is carried out using the very website of NCBI where the alignment is performed.

Regarding protein prediction, two programs were used: SWISS-MODEL website and MODELLER platform. Besides, for both, some of their modalities were tried in order to determine which is the best option to achieve the best model. In the case of SWISS-MODELER, “Sequence”, “Target Template Alignment” and “User template” modalities were used. For “Sequence” a FASTA amino acid sequence was uploaded and the program was run. “Target-Template Alignment” option needs an alignment between the target sequence and the template sequence. This template sequence was BLAST previously in the NCBI

website, specifying “PDB” as the database where to aim the BLAST. The best scored with good resolution protein was selected to perform the alignment. Finally, “User template” was used following the same steps than before but, instead of performing an alignment, PDB file of the template was used. In the case of MODELLER, tutorials uploaded in its website were used to perform different modalities.

After getting the models of all 24 FU, they were visualised, superposed and structurally aligned using UCSF Chimera. As before, tutorials on its website were used to perform all these activities. Pictures of the results were taken after.

For the last objective, two techniques were used: *docking* and *molecular dynamics simulation.* The *docking* was performed by using the ClusPro platform in the following website https://cluspro.org/login.php. After preparing both receptor (hemocyanin FU) and ligand (MT) for *docking* with Chimera, they were uploaded to ClusPro website. The website provides a pack of *docking* models in PDB format.

4. Results and discussion

4.1 Determination of the evolutive origin of *Cantareus aspersus* hemocyanin FU

4.1.1 Study to determine whether each FU has the same genetic origin

Upon performing the alignment between all the FUs corresponding to 8 FUs (a,b,c,d,e,f,g, and h) of 3 subunits (alphaD, alphaN, and beta). The resulting phylogenetic tree is the following:

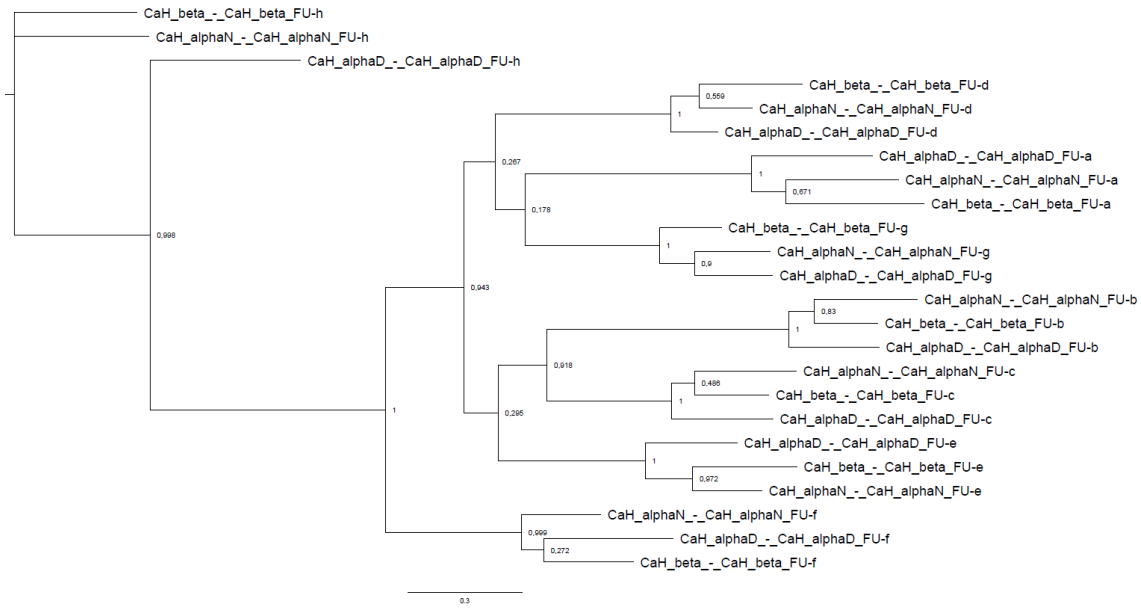


Figure 3. Phylogenetic tree of 24 functional units present in *Cantareus aspersus* hemocyanin

As shown in the tree, all sequences are grouped by functional units (FU-a of subunit alphaN is grouped with FU-a of alphaD and FU-a of beta). This exposes that FUs are closer to the same FUs of the other subunits than to the FUs of the same subunit. Thus, this fact outlines the idea that the three subunits come from the same gene. The nodes that group all three FUs in the same chunk, have a good scoring, giving consistency to the tree and supporting the theory that FUs come from the same gene. It is interesting to remark that FUs-h are grouped out of the tree, as if they were an outsider group. This may indicate that FUs-h are the most different FU among them. This goes in concordance with the fact that some invertebrate species present this specific isoform and some others do not [3]. Thus, this may mean that FU-h may have appeared in this gene later than the rest of FU.

Another test was performed to study if each subunit is formed by a repetition of amino acid sequences coming from a common ancestor. Thus, a similarity matrix has been performed taking the FU amino acid sequences to elucidate whether they are similar enough to have a possible common origin. In Figure 4 is presented the similarity matrix:

1: CaH_alphaD_-_CaH_alphaD_FU-h	100.00	38.85	34.43	39.45	40.50	39.35	38.56	40.70
2: CaH_alphaD_-_CaH_alphaD_FU-f	38.85	100.00	38.79	43.50	40.00	41.73	41.91	38.08
3: CaH_alphaD_-_CaH_alphaD_FU-a	34.43	38.79	100.00	41.81	39.50	41.90	40.75	40.35
4: CaH_alphaD_-_CaH_alphaD_FU-d	39.45	43.50	41.81	100.00	38.46	46.31	44.20	41.65
5: CaH_alphaD_-_CaH_alphaD_FU-b	40.50	40.00	39.50	38.46	100.00	41.46	43.48	45.61
6: CaH_alphaD_-_CaH_alphaD_FU-g	39.35	41.73	41.90	46.31	41.46	100.00	47.13	45.45
7: CaH_alphaD_-_CaH_alphaD_FU-e	38.56	41.91	40.75	44.20	43.48	47.13	100.00	46.25
8: CaH_alphaD_-_CaH_alphaD_FU-c	40.70	38.08	40.35	41.65	45.61	45.45	46.25	100.00

Figure 4. Percent identity matrix performed by Clustal2.1 using as samples the FU sequences of subunit alphaD

As can be seen, all the sequences, compared in pairs, have identities over 38%. This is they share almost half of their positions in the sequence [15]. Which means that they adopt the same three-dimensional structure [17]. Remarking that the sequences exposed above in the identity matrix come from the same protein, which has been split in those eight fragments. It is hard to conceive that eight equal structures could be found within the same amino acid sequence (i.e.: inside the same gene) without thinking in gene duplication.

Taking all this information in to consideration, here is postulated that hemocyanin subunits have the same origin, thus, they come from the same gene that evolved in different subunits. Moreover, it might be that each subunit is formed by eight repetitions of the same ancestral sequence (i.e.: formation of the so-called FUs), that repeated itself before duplicating in the three subunits.

4.1.2 Study whether FUs amino acid sequences of *Cantareus aspersus* are closer to terrestrial molluscs rather than to marine molluscs

To reveal if the FUs of *Cantareus aspersus* are closer to terrestrial molluscs rather than to marine molluscs, three alignments were performed. The complete subunit sequences of *Cantareus aspersus* were taken and aligned with other sequences of other hemocyanin, using, when possible, similar subunits (i.e.: using alphaN subunit of other species when working with alphaN of *Cantareus aspersus*). After aligning them, phylogenetical trees were performed. The following trees were resulted.

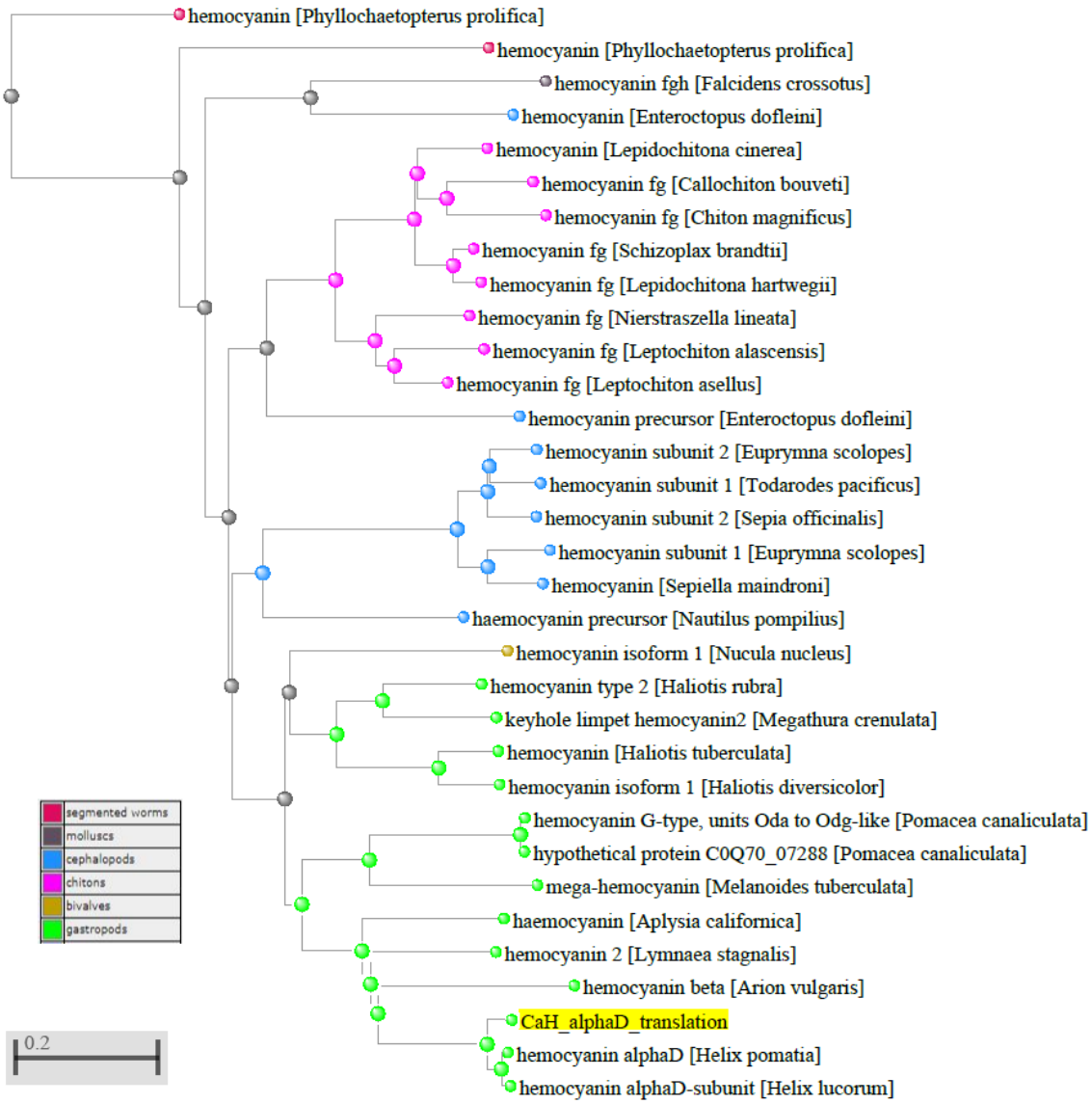


Figure 5. Phylogenetic tree of alphaD subunit of *Cantareus aspersus* hemocyanin (highlighted in yellow) and other hemocyanin sequences obtained by BLAST

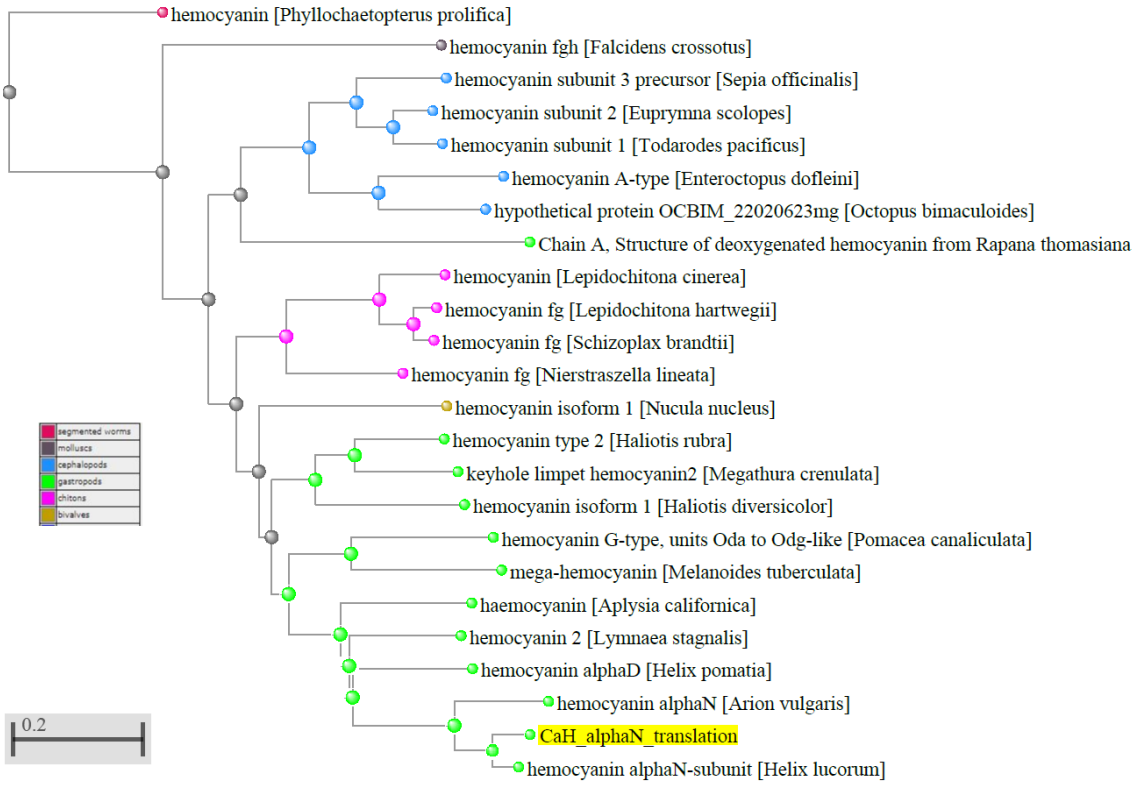


Figure 6. Phylogenetic tree of alphaN subunit of Cantareus aspersus hemocyanin (highlighted in yellow) and other hemocyanin sequences obtained by BLAST

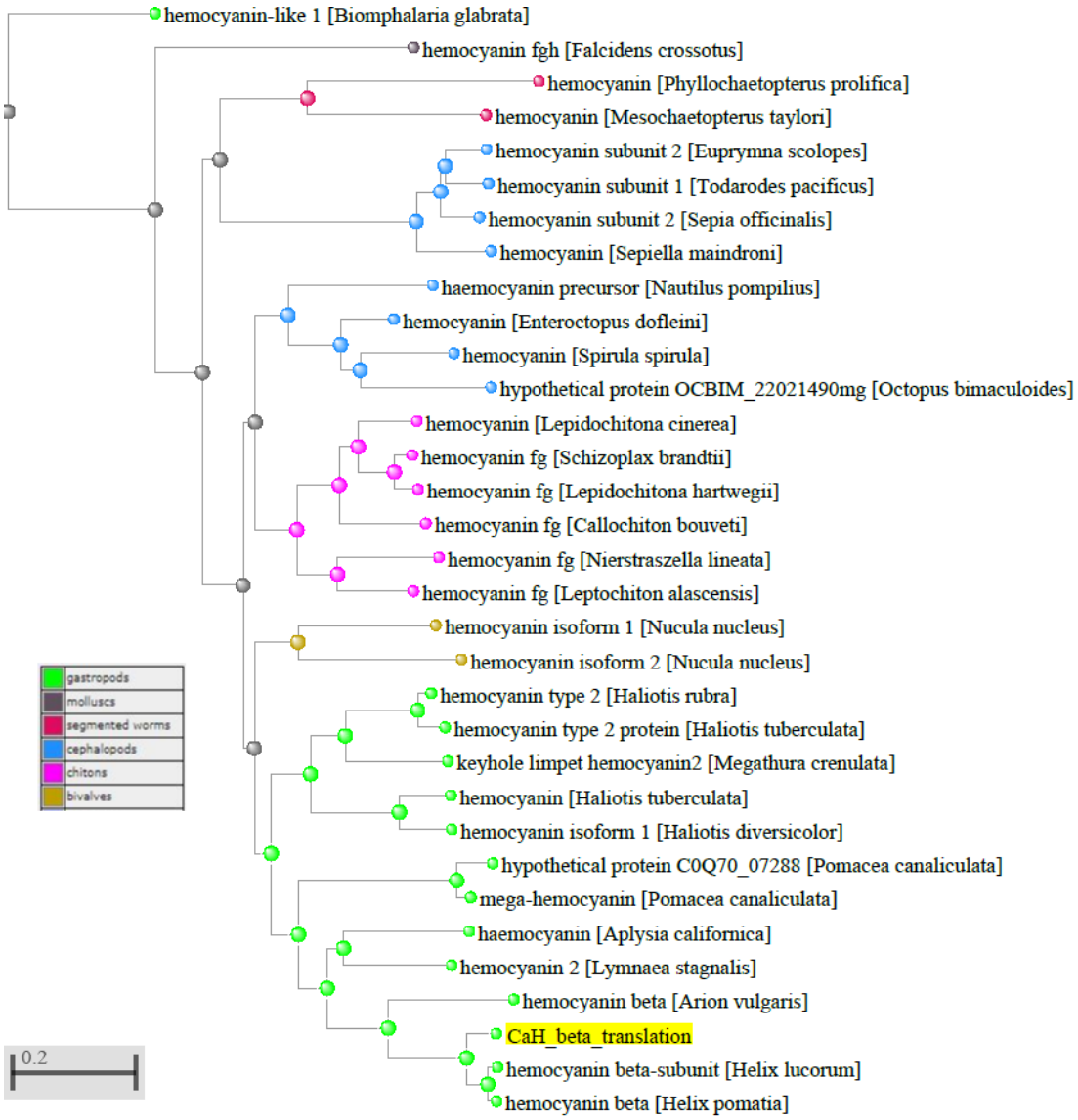


Figure 7. Phylogenetic tree of beta subunit of Cantareus aspersus hemocyanin (highlighted in yellow) and other hemocyanin sequences obtained by BLAST

It is not a surprise that in all three cases each subunit is more similar to the respective terrestrial mollusc subunit than to the marine mollusc subunits. *Cantareus aspersus* is a terrestrial mollusc, which means it is evolutionarily closer to other terrestrial molluscs than to marine molluscs. Here it is proved that hemocyanin keeps this evolution temporality, even when separating classes.

4.1.3 Determine whether there are differences between terrestrial molluscs hemocyanin and marine molluscs hemocyanin

This section can be solved with the results presented in the last section. As can be seen in all three phylogenetic trees, terrestrial molluscs are grouped together while marine molluscs are in a separated group. However, all gastropods form a bigger subgroup, as the other classes do. Thus, the phylogenetic tree is consistent when separating in classes, which makes it trustworthy.

4.2 Prediction of three-dimensional structure of FU and the comparative study of the different isoforms

4.2.1 Verification of the best method for predicting FUs

To achieve this part of the project, several strategies were used to elucidate which one was the best option. The following table shows all the methodologies used and the PDB code used in each one.

Table 1. Strategies used for predicting three-dimensional structure of FUs. In those methods were applicable, the 4 symbol PDB code was used.

SWISS MODEL			MODELLER
auto-finding	**alignment**	**user template**	**basic modelling**
error	error	4bed	4bed (bad model)
error	error	4bed	Not performed
error	error	4bed	
error	error	4bed	
error	error	1lnl and 4bed	
error	error	4bed_b	
error	error	4bed_b	
error	error	4bed_b	
error	error	4bed_a	
error	error	4bed_a	
error	error	4bed_a	
error	error	4bed_a	
error	error	1lnl and 4bed	
error	error	4bed_b	
error	error	4bed_b	
error	error	4bed_b and 1j8s	
error	error	4bed_b and 1j8s	
error	error	4bed	
error	error	4bed	
error	error	4bed	
error	error	1lnl and 4bed	
error	error	4bed_b	
error	error	4bed_b	
error	error	4bed_b and 1j8s	

Among all the strategies used, “user template” modality of SWISS-MODEL was the one used. It has been verified that the FU sequences were too long to be predicted by both “auto-finding” and “alignment” modalities of SWISS-MODEL. Moreover, Modeller was also used. However, the model obtained was rated very poorly. For all these reasons, “user template” modality was the preferred to develop this section of the essay.

4.2.2 Homology prediction of FU three-dimensional structure

From now on, all the models presented in this project are predicted using “user template” method of SWISS-MODEL. The 24 FU sequences of the three subunits were BLAST and the most fitting sequence (better scoring and identity percentage) was used as a template for each FU sequence. However, some of the PDB files of the best fitting sequences were not structured as SWISS-MODEL requires to do the prediction. In those cases, were this happened, the second most fitting sequence was used.

The results obtained are the following:

Table 2. Summary of the predictions made for all the FUs. Each row is one predicted model of each functional unit in which is indicated the template used to predict it and the SWISS-MODEL score obtained of that model.

	BLAST			SWISS-MODEL
FU NAME	*Identity (%)*	*Cover (%)*	*Hit (PDB code)*	*Qmean*
FU_BETA_A	47	100	4bed	-3.04
FU_BETA_B	51	100	4bed	-2.93
FU_BETA_C	62	99	4bed	-2.76
FU_BETA_D	57	99	4bed	-1.74
FU_BETA_E	56	96	1lnl and 4bed	-3.82 and -3.13
FU_BETA_F	54	99	4bed_b	-2.81
FU_BETA_G	56	99	4bed_b	-2.01
FU_BETA_H	47	99	4bed_b	-3.33
FU_ALPHAN_A	47	100	4bed_a	-2.86
FU_ALPHAN_B	49	98	4bed_a	-3.58
FU_ALPHAN_C	58	97	4bed_a	-2.66
FU_ALPHAN_D	59	98	4bed_a	-2.58
FU_ALPHAN_E	57	96	1lnl and 4bed	-3.63 and -2.46
FU_ALPHAN_F	54	100	4bed_b	-2.76
FU_ALPHAN_G	60	99	4bed_b	-2.47
FU_ALPHAN_H	45	98	4bed_b and 1j8s	-7.55 and -2.33
FU_ALPHAD_A	50	100	1j8s	-2.38
FU_ALPHAD_B	51	98	4bed	-2.99
FU_ALPHAD_C	62	99	4bed	-2.8
FU_ALPHAD_D	58	99	4bed	-2.15
FU_ALPHAD_E	57	96	1lnl and 4bed	-3.95 and -2.37
FU_ALPHAD_F	56	100	4bed_b	-2.86
FU_ALPHAD_G	60	98	4bed_b	-1.67
FU_ALPHAD_H	49	97	1j8s	-2.18

In some cases, more than one hit was used to predict the model. However, only one model was used in further sections. In those cases, the model with better Qmean was used.

Here is presented a picture of one FU as representation of all the models predicted.

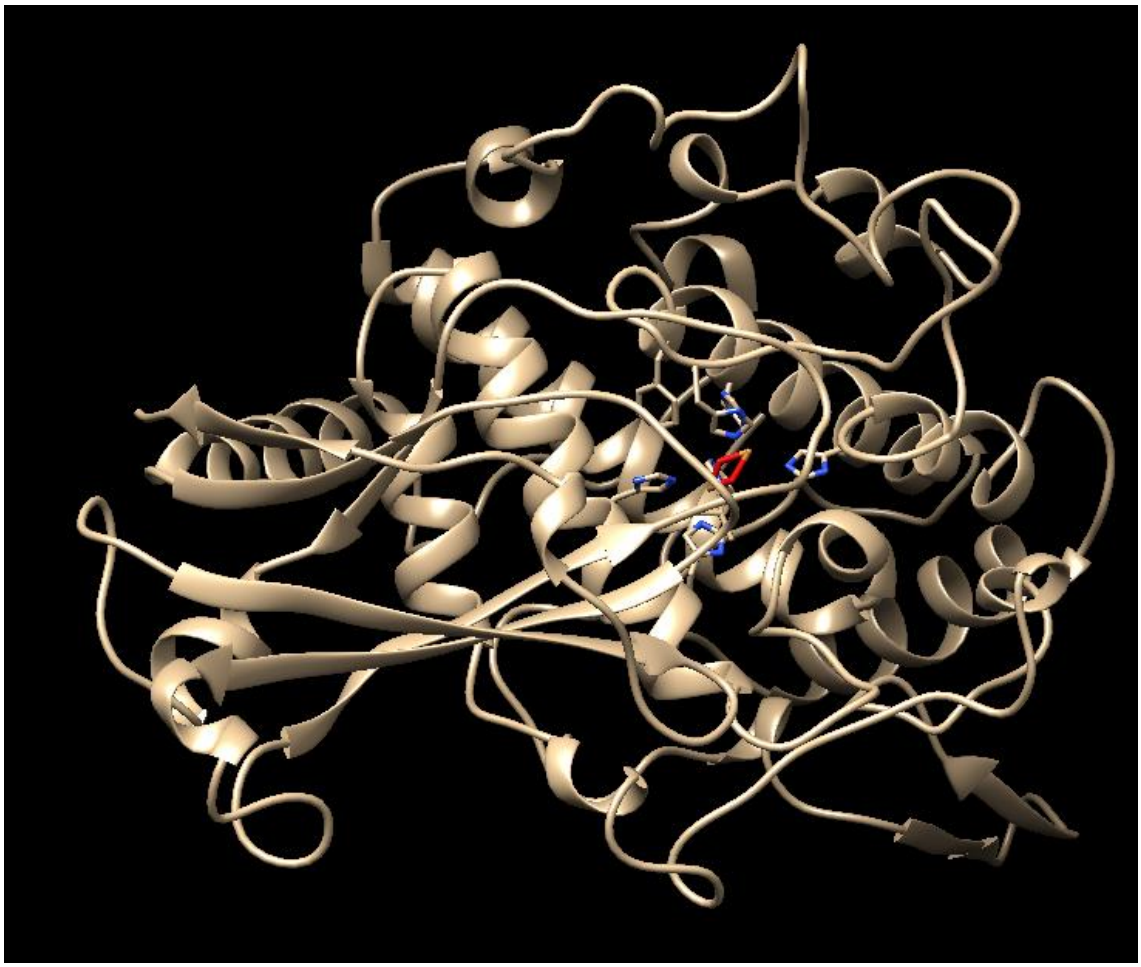


Figure 8. Predicted structure of alphaD_a FU. Exposed the active site of the protein: Copper ligand and superoxide radicals coordinated with histidines.

All the models predicted have similar features. They present alpha-helix secondary structures that surround the active site formed by 6 histidines coordinating 2 Cu (3 histidines each) and 2 O. In some cases, it seems that a cysteine may contribute in the coordination of Cu because of its proximity to the active site.

4.2.3 Incorporation of copper and oxygen ligand in to the model

One of the criteria used to choose a template for predicting was that the template contained the ligand in its model. If these molecules are included in the PDB file, SWISS-MODEL builds a model incorporating the same molecule. By choosing these templates, the model built is better scored and better structured. Besides, the difficulty of adding the ligand in the active site is avoided. This may explain why Modeller was not able to build a model well-structured.

Therefore, all the models predicted contain the proper ligand in their active site. As a reference, in Figure 8 is represented this structure.

4.2.4 Finding the differences between the structures of all FU isoforms.
One of the hypothesis is that all the FUs come from the same sequence. Even though the sequences are different, the structure should be maintained in all the FUs. For this reason, the FUs of each subunit were superposed to compare the structure and to detect differences between them.

This was performed using Chimera and the results are the following:

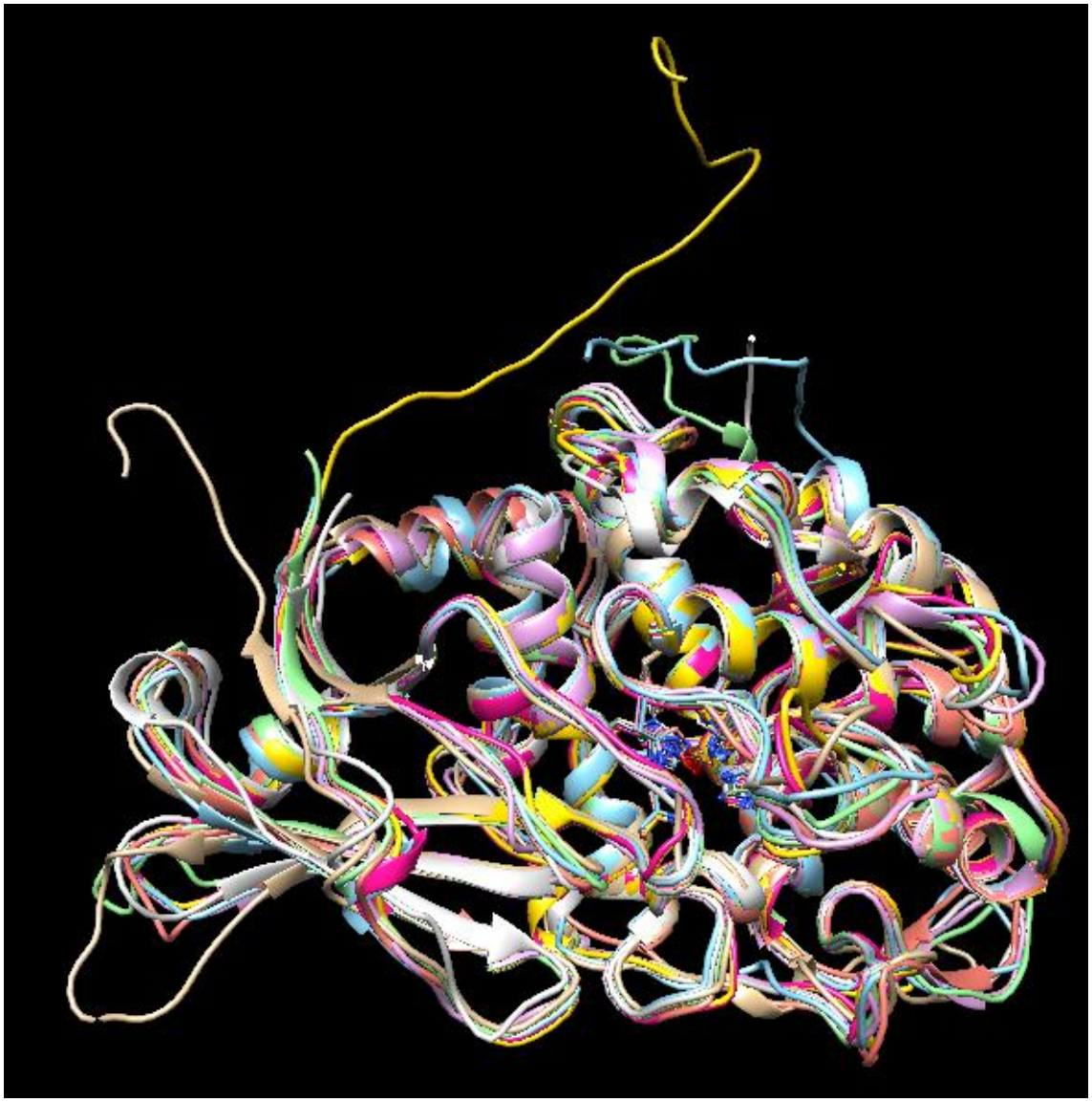


Figure 9. Superimposition of all beta subunit FUs

As can be seen in Figure 9, the structure is maintained along all the FU. Althought the same template is used to predict almost the totallity of the FUs, the predictions have good scorings (as seen in Table 2), which raises the probability that the models are consistent with the reality. Moreover, since all

the models have good overlapping, it gives strength to the postulate that the structures of the FUs in the same subunit are conserved.

4.3. Determining whether there is interaction between FU and Cu-MTs

4.3.1 Interaction between a FU model and a MT structure

To complete this part, it was needed the structure of the MT. There are very few protein structures of metallothioneins solved in PDB. This is because MT are very labile and it is very difficult to set the ideal conditions to perform either X-ray or NMR for these kind of proteins [8]. Moreover, due to its lack of secondary structure and because its tertiary structure is not ruled by the amino acid sequence but for its coordination to metal ions, MT resolution by bioinformatics tools is a complicated mission as well.

Considering all this and since no *Cantareus aspersus* MT structure is available in PDB, the preferred strategy was to perform a prediction of the protein by homology. An NMR-solved structure of *Littorina litorea* MT was taken as a template and the amino acid sequence of Cu-MT of *Cantareus aspersus* as a target. The methodology followed was the same as the used for predicting the FUs.

After obtaining the predicted model, it was checked that it made sense. As can be seen in Figure 10, the protein is disposed in two domains. As regarded in the bibliography, it is expected two domains of six coppers each domain [12]. Thus, it seems a good model so far. Moreover, cysteines are disposed aiming the centre of the domain, as if there should be some cluster in it.

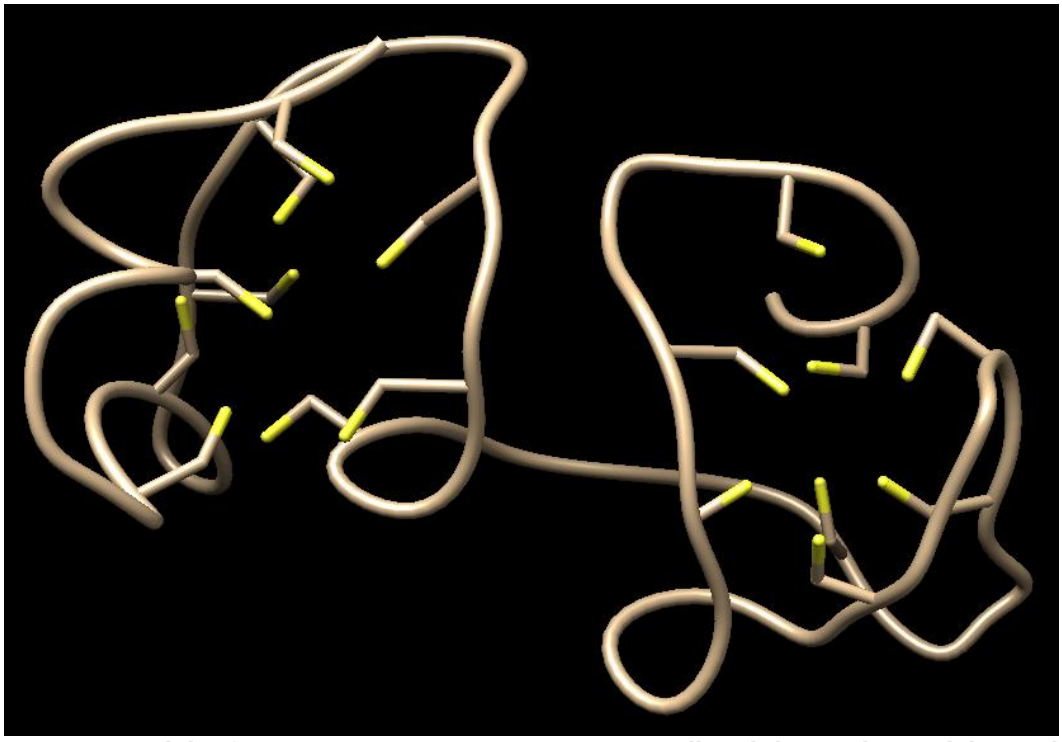


Figure 10. Model of *Cantareus aspersus* Cu-MT predicted by SwissModel. Tertiary structure where only cysteines are shown.

After that, copper ions were introduced manually in to the model (Figure 11).

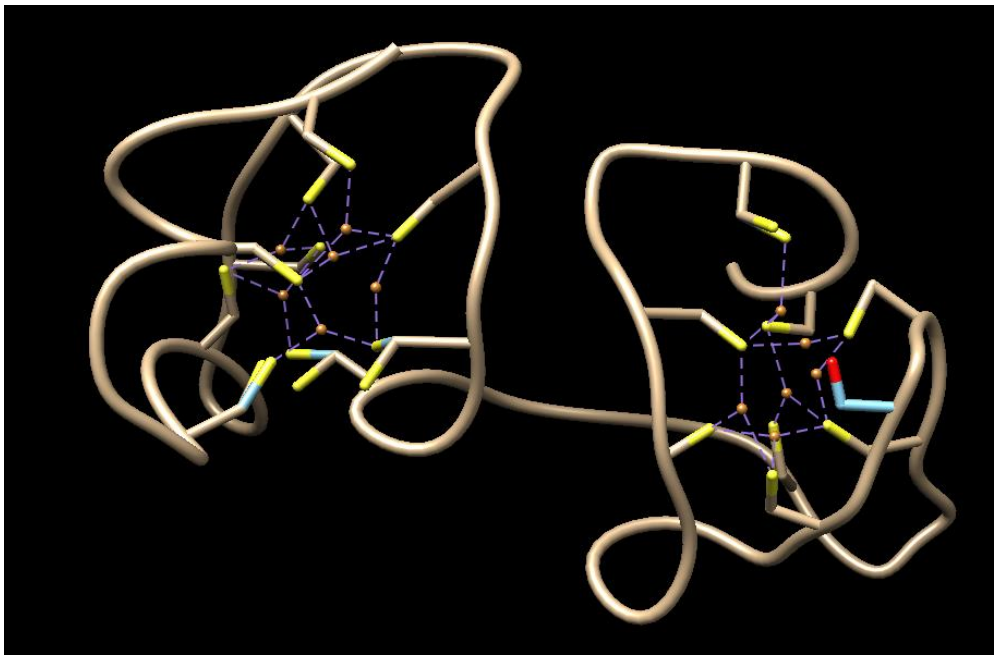


Figure 11. Cu-MT model with copper in coordination. Dashed lines represent coordinative bonds.

All bonds made between cysteines and copper are either linear or plain trigonal, as seen in the bibliography [9]. Taking all this information into account, it can be postulated that this model of MT is a model good enough to work on.
Once the *Cantareus aspersus* model is available, its interaction with a FU model is studied. The FU chosen is alphaD_g due to its good scoring. Both MT and FU were uploaded to ClusPro. The algorithm solved over 100 interactions between both peptides. In Table 3 are shown the 30 best predictions resulted from the tool.

Table 3. Results displayed in the ClusPro website. These are the 30 best models of docking predicted by the website.

Cluster	Members	Representative	Weighted Score	Cluster	Members	Representative	Weighted Score
0	120	Center	-583.2	**15**	22	Center	-586.8
		Lowest Energy	-670			Lowest Energy	-630.8
1	54	Center	-531.9	**16**	20	Center	-519.1
		Lowest Energy	-640.8			Lowest Energy	-575.4
2	52	Center	-528.6	**17**	18	Center	-569.6
		Lowest Energy	-650.2			Lowest Energy	-588
3	49	Center	-539.4	**18**	18	Center	-571.3
		Lowest Energy	-646.6			Lowest Energy	-571.3
4	46	Center	-527.9	**19**	17	Center	-602.3
		Lowest Energy	-591.9			Lowest Energy	-602.3
5	43	Center	-526.5	**20**	17	Center	-577
		Lowest Energy	-588.8			Lowest Energy	-577
6	42	Center	-540.4	**21**	16	Center	-534.3
		Lowest Energy	-584.1			Lowest Energy	-573.9
7	40	Center	-583.1	**22**	14	Center	-523.8
		Lowest Energy	-583.1			Lowest Energy	-570.3
8	32	Center	-617.4	**23**	13	Center	-539.4
		Lowest Energy	-634.1			Lowest Energy	-583.9
9	28	Center	-537.1	**24**	12	Center	-538.1
		Lowest Energy	-590			Lowest Energy	-581.8
10	28	Center	-626.6	**25**	12	Center	-520.5
		Lowest Energy	-626.6			Lowest Energy	-567.6
11	24	Center	-523.1	**26**	11	Center	-576.8
		Lowest Energy	-590.6			Lowest Energy	-576.8
12	24	Center	-610.4	**27**	11	Center	-576.2
		Lowest Energy	-610.4			Lowest Energy	-576.2
13	23	Center	-547.4	**28**	11	Center	-566.5
		Lowest Energy	-621.4			Lowest Energy	-566.5
14	23	Center	-522.1	**29**	11	Center	-558.7
		Lowest Energy	-615.8			Lowest Energy	-558.7

The criteria used to choose one of them was that the interaction had one of the lowest energies and that the interaction took place near the active site of the FU. The lowest energy is taken because it is the most favoured system energetically. On the other hand, the logic understanding is that the interaction should take place on the closest side to the active site because the less will travel the ion from the MT to the FU.
Considering this, Figure 12 is the model chosen as a one possible interaction between these two proteins.

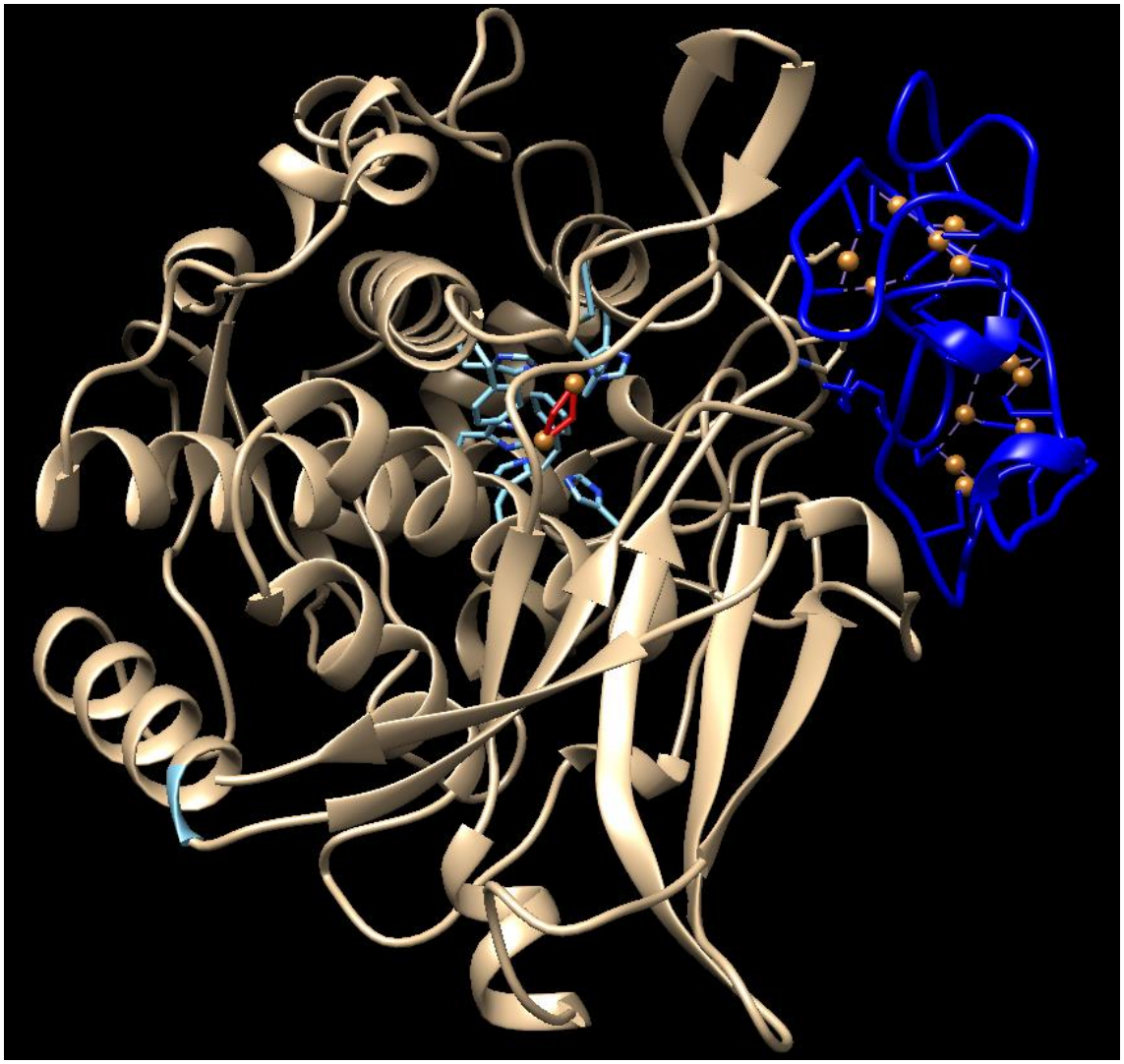


Figure 12. Interaction model after docking protein-protein. In beige is represented the FU, in blue the MT and the spheres represent copper ions.

In Figure 12 is reflected that the docking is done in one of the bits where the FU does not have secondary structure. It is noteworthy that the MT is labile and can adapt itself to the FU. Moreover, in this case the MT is re-structured so the bonds with the ions have changed passing from plain trigonal to lineal. This may have something to do with the process of ion exchange, which the fact of losing contact points eases the ion exchange.

The interactions between them are mainly electrostatic interactions. On the surface of both proteins there are present charged amino acids (lysine, histidine, etc.). Moreover, there is a H-bond between GLY 24 of the MT and the HIS 65 of the FU as can be seen in Figure 13. It is important to remark that this is a model chosen due to its low energy and logic positioning, but it is just a hypothesis of a possible docking.

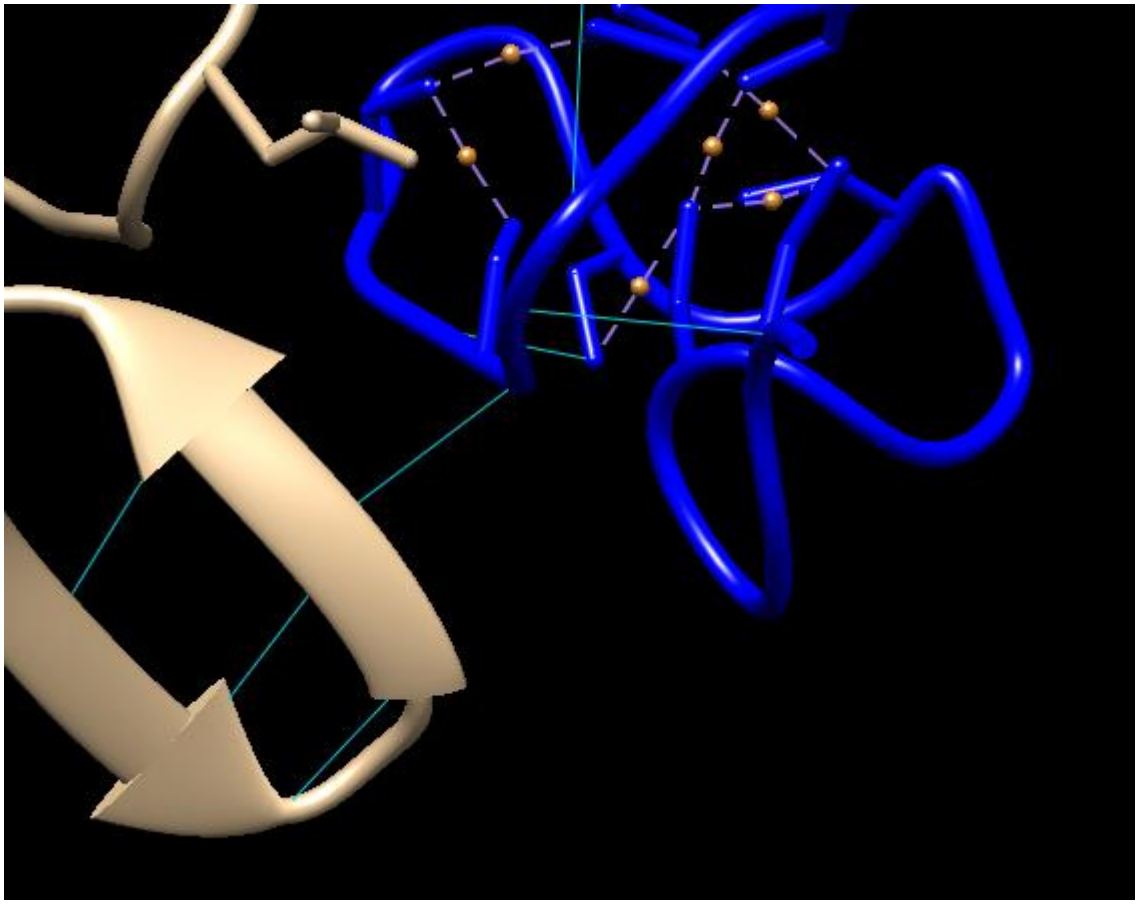


Figure 13. H-bond (solid blue line) between MT (blue) and the FU (beige).

In any case, here is demonstrated that it may exist a real interaction between these two proteins.

4.3.2 Elucidating whether there is metal exchange between FUs and Cu-MT

For this section, it was intended to demonstrate that a copper ion from Cu-MT is transferred to the FU active site by means of molecular dynamics. This process was tried using Chimera and using VMD. Both programs have their own particularities. For both programs one FU and a predicted model of Cu-MT was used. Only the Cu-MT contained Cu(I) and it was bound by coordinating bonds. After learning how these programs worked and trying this process, no result was obtained. It is thought that it is needed to configure the programs with more inputs, but it is needed to do a deeper study to elucidate this issue.

5. Conclusions

The conclusions are built on base of the initial objectives of this project.

(1) In this project it has been demonstrated that the three subunits of the hemocyanin come from the same ancient gene. Also, that hemocyanin in terrestrial molluscs is more similar to other terrestrial molluscs than to the marine ones, which make terrestrial hemocyanin different of marine hemocyanin.

(2) It has been demonstrated that the best option, for being more trustful and quick is to use MODEL-SWISS “user template” functionality and that the FUs were properly predicted. Besides, all the FU have a very similar structure, which can be postulated that the FU have a conserved structure.

(3) For the last objective it has been demonstrated that there can exist an interaction between the FU and the MT because the docking was completed but it could not be demonstrated that there is ion transfer between the two proteins.

Regarding the project planification, it has been performed according to the initial plan. There has been only one sub-objective that could not be achieved. The whole project was a complex and large project. It has been intended to explore genetics, proteomics and interactomics of hemocyanin. This has not permitted to study each of the parts deeply. Thus, instead of a focused work of a specific aspect of the hemocyanin, here are studied some general aspects of the protein. This has its weaknesses: lack of time to perform all the experiments. Thus, it can be concluded that the planification was good but the objectives too ambitious to be completed properly. In any case, here there are opened several lines of study and as a perspective of future it will be necessary to elucidate whether there is ion transfer between FUs and MTs.

6. Glossari

BLAST	Basic Local Alignment Search Tool
PDB	Protein Data Bank
FU	Functional Unit (as part of the hemocyanin)
MT	Metallothionein
NMR	Nuclear Magnetic Ressonance
NCBI	National Center for Biotechnology Information

7. Bibliografia

1. Nelson, D. L. & Cox, M. M. *Lehninger's Principles of Biochemestry*. *W. H. Freeman* (2004). doi:10.1017/CBO9781107415324.004
2. Raynova, Y., Doumanova, L. & Idakieva, K. N. Phenoloxidase activity of helix aspersa maxima (Garden Snail, Gastropod) Hemocyanin. *Protein J.* **32,** 609–618 (2013).
3. Markl, J. Evolution of molluscan hemocyanin structures. *Biochim. Biophys. Acta - Proteins Proteomics* **1834,** 1840–1852 (2013).
4. Yang, S. *et al.* Prediction and characterization of a novel hemocyanin-derived antimicrobial peptide from shrimp Litopenaeus vannamei. *Amino Acids* **50,** 995–1005 (2018).
5. Coates, C. J. & Decker, H. Immunological properties of oxygen-transport proteins: hemoglobin, hemocyanin and hemerythrin. *Cell. Mol. Life Sci.* **74,** 293–317 (2016).
6. Palacios, Ò. *et al.* Cognate and noncognate metal ion coordination in metal-specific metallothioneins: The Helix pomatia system as a model. *J. Biol. Inorg. Chem.* **19,** 923–935 (2014).
7. Velkova, L. *et al.* Structure of hemocyanin from garden snail Helix lucorum. *Comp. Biochem. Physiol. - B Biochem. Mol. Biol.* **157,** 16–25 (2010).
8. Capdevila, M., Bofill, R., Palacios, O. & Atrian, S. State-of-the-art of metallothioneins at the beginning of the 21st century. *Coord. Chem. Rev.* **256,** 46–62 (2012).
9. Calvo, J. & Meloni, G. Critical Review Copper Metallothioneins. 236–245 doi:10.1002/iub.1618
10. Chabicovsky, M., Klepal, W. & Dallinger, R. Mechanisms of cadmium toxicity in terrestrial pulmonates: Programmed cell death and metallothionein overload. *Environ. Toxicol. Chem.* **23,** 648–655 (2004).
11. Imam, H. T. & Blindauer, C. A. Differential reactivity of closely related zinc(II)-binding metallothioneins from the plant Arabidopsis thaliana. *J. Biol. Inorg. Chem.* **23,** 137–154 (2018).
12. Höckner, M. *et al.* Physiological relevance and contribution to metal balance of specific and non-specific Metallothionein isoforms in the garden snail, Cantareus aspersus. *BioMetals* **24,** 1079–1092 (2011).
13. Pérez-Rafael, S. *et al.* Cantareus aspersus metallothionein metal binding abilities: The unspecific CaCd/CuMT isoform provides hints about the metal preference determinants in metallothioneins. *Biochim. Biophys. Acta - Proteins Proteomics* **1844,** 1694–1707 (2014).
14. Dallinger, R. *et al.* Copper in *Helix pomatia* (Gastropoda) is regulated by one single cell type: differently responsive metal pools in rhogocytes. *Am. J. Physiol. Integr. Comp. Physiol.* **289,** R1185–R1195 (2005).
15. Ullah, I. *Probabilistic Models for Species Tree Inference and Orthology Analysis*. (2015).
16. Moses, V., Hatherley, R. & Tastan Bishop, Ö. Bioinformatic characterization of type-specific sequence and structural features in auxiliary activity family 9 proteins. *Biotechnol. Biofuels* **9,** 1–17 (2016).
17. Gilson, A. I., Marshall-Christensen, A., Choi, J. M. & Shakhnovich, E. I.

The Role of Evolutionary Selection in the Dynamics of Protein Structure Evolution. *Biophys. J.* **112,** 1350–1365 (2017).