

## Article

# Bioinformatic Characterization and Molecular Evolution of the *Lucina pectinata* Hemoglobins

Ingrid M. Montes-Rodríguez <sup>1</sup>, Carmen L. Cadilla <sup>2,\*</sup>, Juan López-Garriga <sup>3</sup> and Ricardo González-Méndez <sup>4,\*</sup>

<sup>1</sup> Cancer Biology Division, PROMIC, Comprehensive Cancer Center of the University of Puerto Rico, San Juan, PR 00936-3027, USA

<sup>2</sup> Department of Biochemistry, School of Medicine, University of Puerto Rico-Medical Sciences Campus, San Juan, PR 00936-5067, USA

<sup>3</sup> Department of Chemistry, Faculty of Arts and Sciences, University of Puerto Rico—Mayagüez Campus, Mayagüez, PR 00681-9000, USA

<sup>4</sup> Department of Radiological Sciences, School of Medicine, University of Puerto Rico-Medical Sciences Campus, San Juan, PR 00936-5067, USA

\* Correspondence: carmen.cadilla@upr.edu (C.L.C.); ricardo.gonzalez7@upr.edu (R.G.-M.)

**Abstract:** (1) Introduction: *Lucina pectinata* is a clam found in sulfide-rich mud environments that has three hemoglobins believed to be responsible for the transport of hydrogen sulfide (HbI<sub>LP</sub>) and oxygen (HbII<sub>LP</sub> and HbIII<sub>LP</sub>) to chemoautotrophic endosymbionts. The physiological roles and evolution of these globins in sulfide-rich environments are not well understood. (2) Methods: We performed bioinformatic and phylogenetic analyses with 32 homologous mollusk globin sequences. Phylogenetics suggests a first gene duplication resulting in sulfide binding and oxygen binding genes. A more recent gene duplication gave rise to the two oxygen-binding hemoglobins. Multidimensional scaling analysis of the sequence space shows evolutionary drift of HbII<sub>LP</sub> and HbIII<sub>LP</sub>, while HbI<sub>LP</sub> was closer to the *Calyptogena* hemoglobins. Further corroboration is seen by conservation in the coding region of hemoglobins from *L. pectinata* compared to those from *Calyptogena*. (3) Conclusions: Presence of glutamine in position E7 in organisms living in sulfide-rich environments can be considered an adaptation to prevent loss of protein function. In HbI<sub>LP</sub> a substitution of phenylalanine in position B10 is accountable for its unique reactivity towards H<sub>2</sub>S. It appears that HbI<sub>LP</sub> has been changing over time, apparently not subject to functional constraints of binding oxygen, and acquired a unique function for a specialized environment.

**Keywords:** *Lucina pectinata*; *L. pectinate*; hemoglobins; sulphur-binding hemoglobins; bioinformatics; phylogenetics of hemoglobins in mollusks



**Citation:** Montes-Rodríguez, I.M.; Cadilla, C.L.; López-Garriga, J.; González-Méndez, R. Bioinformatic Characterization and Molecular Evolution of the *Lucina pectinata* Hemoglobins. *Genes* **2022**, *13*, 2041. <https://doi.org/10.3390/genes13112041>

Academic Editor: Haiyan Wang

Received: 12 September 2022

Accepted: 26 October 2022

Published: 5 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Invertebrate hemoglobins (Hbs) display higher variability in terms of amino acid chain constitution, quaternary structures, and functional properties when compared to their vertebrate counterpart [1]. Some are monomeric, while others can form large polymers. In invertebrates, Hbs can be expressed differentially in different anatomical sites, and show differences in their ligand kinetics [1]. Since invertebrates can be found in a wide range of habitats, the variation encountered in their Hbs is indicative of how these molecules must adapt to fulfill their function regardless of the environmental challenges in the surrounding areas where they live [2].

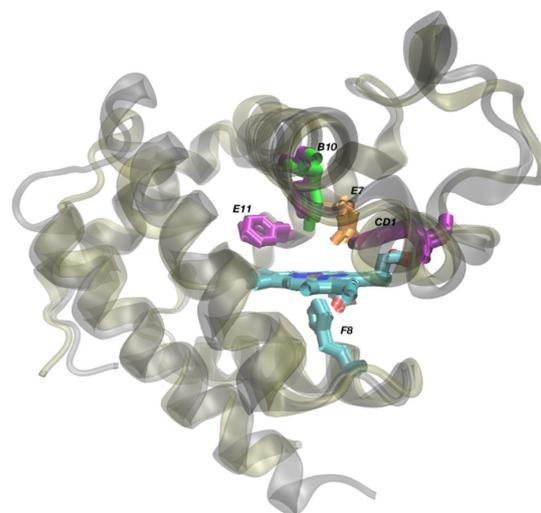
The globin fold is a well characterized protein structure, first described for sperm whale myoglobin (Mb) and horse oxyhemoglobin, composed of up to eight  $\alpha$  helices, labeled A to H, and residues within each helix segment are numbered from the amino end [3–5]. Perutz [6] established structural similarities between various globins and sperm whale myoglobin that allows the use of the same notation for equivalent structural residues in all hemoglobins, regardless of additions or deletions in their primary structure. Bashford,

Chothia, & Lesk [7] postulated that at least 32 internal conserved regions, possessing mostly hydrophobic residues, define the overall globin fold structure. Lesk & Chothia [8] demonstrated that the differences in amino acid sequences between these molecules lead to the formation of a well conserved globin structure by a series of low-energy mutations producing limited structural changes and by the functional advantage of any modulation of the ligand affinity caused by these changes.

In sulfide-rich habitats, Hbs are essential for the sustainability of the chemoautotrophic based symbiosis found in many invertebrates [9,10]. The hosts supply the symbionts with inorganic substrates, such as sulfide, while the symbionts are believed to provide the host with autotrophically fixed carbon [11]. A remarkable aspect of the Hbs of these organisms is their ability to retain their function in presence of hydrogen sulfide ( $H_2S$ ) without the covalent modification of the heme pyrrole rings that leads to formation of sulfhemoglobin, as has been seen in most mammalian myoglobins (Mbs) and Hbs [12]. However, the biochemical characteristics of the Hbs from these species differ substantially.

Chemosymbiosis with sulfide-oxidizing bacteria is found in members of the *Lucinidae* family, which have an extensive variety of species and are found across the globe [13]. The clam *Lucina pectinata* is found in the southwest coast of Puerto Rico and throughout the Caribbean Sea. As it lives in sulphide-rich muds, it has developed a symbiotic relation with intracellular bacteria for which the clam provides both hydrogen sulfide ( $H_2S$ ) and oxygen ( $O_2$ ). Three different Hbs have been identified in *L. pectinata*: hemoglobin I ( $HbI_{LP}$ ), a sulfide-reactive hemoglobin, and hemoglobin II ( $HbII_{LP}$ ) and hemoglobin III ( $HbIII_{LP}$ ), which are oxygen-reactive hemoglobins [14].

The sulfide binding  $HbI_{LP}$  is a monomeric globular protein, composed of 143 amino acid residues [15].  $HbI_{LP}$  has a conventional globin fold lacking the D helix. This is similar to plant and truncated hemoglobins [16]. “ $HbI_{LP}$  is characterized by an unusual distribution of phenylalanines at the heme distal positions B10, CD1 and E11” [17].  $HbII_{LP}$  and  $HbIII_{LP}$  (150 and 153 amino acid respectively) have a tyrosine residue at position B10 [18] (Figure 1 below). Structural alignment of  $HbI_{LP}$  protein structure (PDB 1FLP) and Sperm Whale Myoglobin (SWM) (PDB 1A6M) showing their amino acid distribution in the heme pocket. For SWM there is a leucine at position B10, a histidine at position E7 and a valine in position E11. (see Supplementary Figure S1).



**Figure 1.** Structural alignment of  $HbI_{LP}$  protein structure (PDB 1FLP) and  $HbII_{LP}$  (PDB 2OLP) showing their amino acid distribution in the heme pocket. Alignment was done using the STAMP algorithm as implemented in MultiSeq in VMD. The only difference between the two heme pockets is the amino acid in position B10, a phenylalanine for  $HbI_{LP}$  and a tyrosine for  $HbII_{LP}$ . Both hemoglobins have a phenylalanine at positions CD1 and E11, a glutamine at position E7 and a histidine at position F8.  $HbII_{LP}$  and  $HbIII_{LP}$  have the same heme pocket amino acid distribution. The RMSD for the backbones (residues 2–147) is 6.54 angstrom.

Amino acid sequences of HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> have a 65% of identity and they both have lower sequence identity with HbI<sub>Lp</sub> (32% of identity with HbII<sub>Lp</sub> and 33% of identity with HbIII<sub>Lp</sub>). In these three hemoglobins the heme distal E7 position has a glutamine residue instead of the typical distal histidine (Figure 1 shows the comparison of the heme pockets of HbI<sub>Lp</sub> and the sperm whale myoglobin). These three Hbs have been extensively studied in order to understand their chemistry, the factors that affect their specificity, and they have been used as models to understand distal ligand binding control [19–23]. How the environment influences the evolution of function in these hemoglobins has yet to be elucidated. We compared the *L. pectinata* Hbs to other homologous globin sequences from mollusks to identify conserved features and evolutionary relationships. We also compared the coding sequences of *L. pectinata* Hbs between them and with those available for *Calyptogena* spp. in order to obtain a better understanding of globin gene structure and sequence conservation in mollusks.

## 2. Materials and Methods

### 2.1. Bioinformatic Analyses

The sequences of *L. pectinata*'s Hbs (mRNA: HbI<sub>Lp</sub>-AF187049; HbII<sub>Lp</sub>-AY243364; HbIII<sub>Lp</sub>-EU040120; protein: HbI<sub>Lp</sub>-AAG01380; HbII<sub>Lp</sub>-AAO89499; HbIII<sub>Lp</sub>-ABS87592) were retrieved from the National Center for Biotechnology Information (NCBI) GenBank and GenPept databases [24]. Retrieval of sequences homologous to the three protein sequences of the *L. pectinata* hemoglobins was performed with the Basic Local Alignment Search Tool (BLAST) [25–27] using blastp with default parameters and selecting to search only in the mollusks (taxonomic id 6447). Thirty-two unique protein sequences were retrieved from the three BLAST homology searches with an expectation value threshold smaller than 0.0001 for significance and/or sequence similarity of approximately 25% or greater. The accession numbers of these sequences together with the BLAST search result summaries are shown in Supplementary Table S1.

Multiple sequence alignment (MSA) was performed with globin sequences homologous to HbI<sub>Lp</sub>, HbII<sub>Lp</sub> and HbIII<sub>Lp</sub>. The MSA includes globins from the deep-sea clam species of *Calyptogena*, marine gastropods from *Aplysia* spp., ark clams, also known as blood clams (*Anadara* spp., *Barbatia* spp., *Scapharca* spp. and *Tegillarca* spp.), the Atlantic surf clam *Spisula solidissima* and the Antarctic bivalve *Yoldia eightsi*. The *Physeter catodon* (Sperm Whale) myoglobin amino acid sequence (Accession number NP\_001277651) and the helical segments of its globin three-dimensional structure (Accession number 1EBC\_A) were included in the MSA as reference, to aid identification of amino acid residues based on the myoglobin fold [7,8] and the standard globin fold nomenclature [6].

Multiple sequence alignment (MSA) of these 32 homologous sequences was performed using the PSI-COFFEE alignment program. This algorithm is used to align distantly related proteins using the consistency function and profile information with homology extension [28]. Given the low sequence similarity, around the so-called 'twilight zone' for detection of homology, this is the method of choice for aligning these sequences. These results were visualized using GeneDoc [29].

The structural alignment of the protein structures of HbI<sub>Lp</sub> (PDB:1FLP) and HbII<sub>Lp</sub> (PDB:2OLP), and HbI<sub>Lp</sub> (PDB:1FLP) and Sperm Whale Myoglobin (SWM) (PDB:1A6M), showing the spatial orientation and distribution of amino acids in the heme pocket was performed using the MultiSeq bioinformatics analysis environment [30] using the Structural Alignment of Multiple Proteins algorithm [31] and the RMSD function in the VMD software [32]. These comparisons help illustrate the overall structural conservation of the hemoglobin fold and the heme O<sub>2</sub>/H<sub>2</sub>S binding pockets.

### 2.2. Phylogenetic Analysis

Phylogenetic analysis of these globin sequences was performed as follows. The MSA of the 32 amino acid sequences was trimmed using trimAl with the automated heuristic setting 'automated 1' [33]. The trimmed MSA was then converted to the PHYLYP format.

Using the PHYLIP suite of programs [34], an unrooted Maximum Likelihood phylogenetic tree with 1000 bootstrap replicates (SEQBOOT) was constructed using the protein maximum likelihood routine (ProML) in its parallel implementation (MPI-ProML) [35] using 16 processors in the BioU computer cluster at the Pittsburgh Supercomputing Center. A consensus phylogenetic tree was generated with the CONSENSE routine using the extended majority rule and visualized with FigTree [36]. A rooted species phylogenetic tree was built using the Taxonomy resource at NCBI [37,38] and was verified using the Tree-of-Life resource [39]. The gene tree and the species tree were then reconciled using the program NOTUNG [40,41] and a root was placed on the globin gene tree using the rooting procedure in NOTUNG [40–43]. Visualization of the reconciled phylogenetic trees was done with NOTUNG routines that are based on the ATV routine in FORESTER [44].

### 2.3. Metric Multidimensional Scaling Analysis

The trimmed MSA was used as input for metric multidimensional scaling (MMDS) analysis and exploration of the sequence space. The MMDS analysis was performed with the R package *bios2mds* [45], using a JTT similarity matrix to construct the distance matrix required for the analysis. A clustering analysis was performed as part of the MMDS using the routines in the package and optimized based on a Silhouette analysis. These were performed as described in [45,46].

### 2.4. Coding Region Alignments

Multiple Sequence Alignments (MSA) of the coding sequences of the three *L. pectinata* hemoglobin sequences were constructed using M-COFFEE [47]. Motif analysis was done using the Maximum Entropy Motif Elicitation program (MEME) in the zero-or-one-per-sequence mode, requesting 6 motifs [48]. All the alignment results were visualized using the program GeneDoc [29].

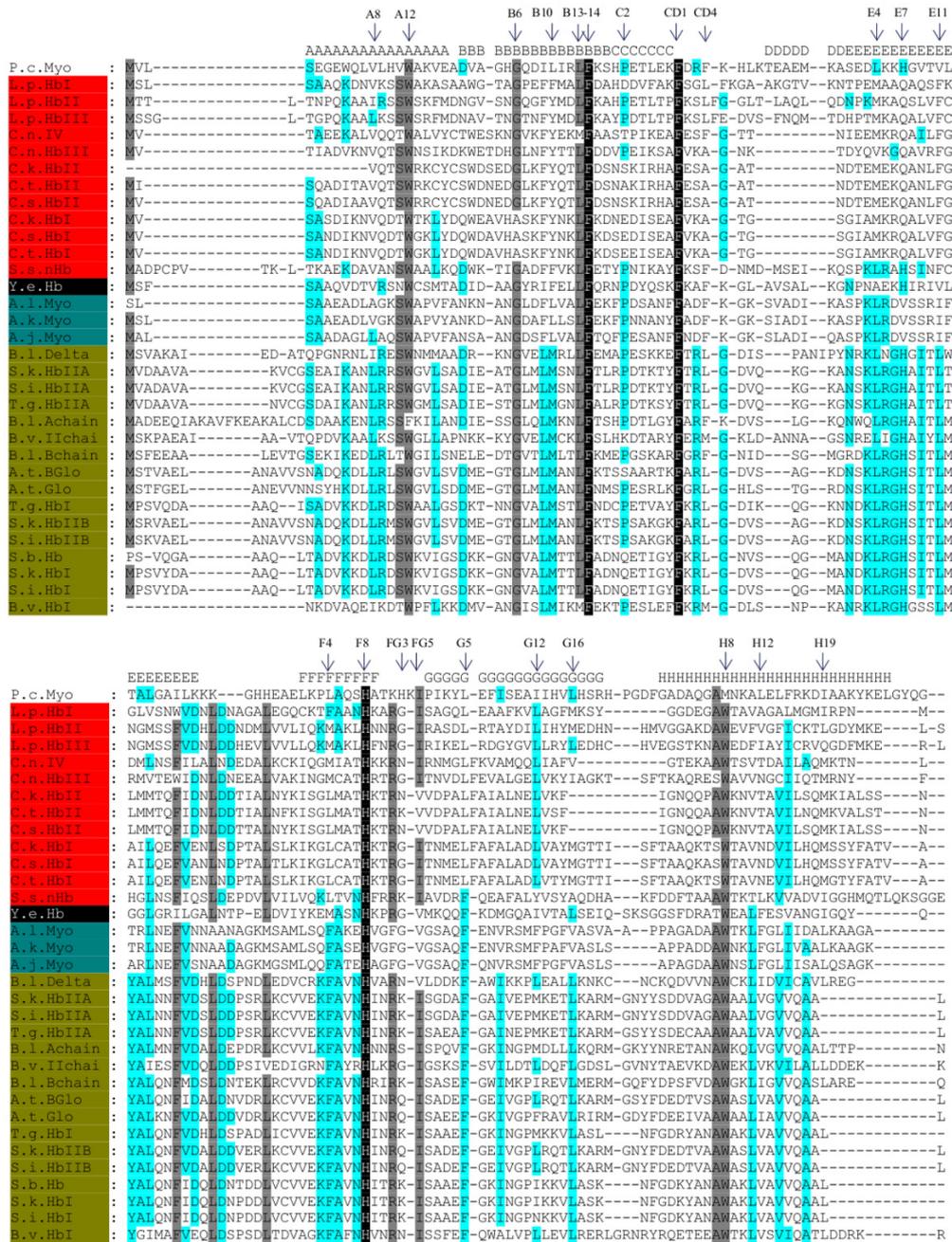
Retrieval of sequences homologous to the mRNA sequences of the *L. pectinata* Hbs was performed with BLAST [25–27] using blastn with default parameters and optimized for somewhat dissimilar sequences. The only sequences that showed homology with a significant expectation value were the *L. pectinata* hemoglobin sequences for HbII<sub>Lp</sub> and HbIII<sub>Lp</sub>, and that homology was with each other. HbI<sub>Lp</sub> did not retrieve any nucleotide sequence. The sequences of coding regions of *Calyptogenia* spp. Hbs (*C. nautiliei* HbIV-AB186050.1; *C. nautiliei* HbIII-AB186049.1; *C. soyoae* HbI-AB186047.1; *C. soyoae* HbII-AB186048.1; *C. tsubasa* HbI-AB186401.1; *C. tsubasa* HbII-AB186402.1; *C. kaikoi* HbI-AB186045.1; *C. kaikoi* HbII-AB186046.1) were also retrieved from the NCBI GenBank. MSA, MEME analysis and visualization was performed as indicated in the previous paragraph.

## 3. Results

### 3.1. Multiple Sequence Alignment

The MSA shows a high quality alignment confirmed by the conservation of the invariant proximal histidine at position F8 (His-F8), the highly conserved phenylalanine at position CD1 (Phe-CD1), and the phenylalanine at position B14 (Phe-B14) (Figure 2). A tryptophan at position H8 replaces the methionine found in sperm whale myoglobin. This is an amino acid replacement found in invertebrate Hbs [7,49]. The conservation of key positions in hemoglobins can be explained by the previous work of Ptitsyn and Ting [50] with the analysis of sequences from twelve globin subfamilies. In this analysis, Ptitsyn and co-workers showed that there are two common conserved clusters of amino acids between these globins. The first cluster includes residues in direct contact with the heme (CD1, F8, E11, FG5, F4, and G5), and heme neighboring residues (C2, B14, B10, E4, CD4, H19 and B13). The second cluster includes hydrophobic residues distant from the heme center that may play an important role in the first steps of the protein folding process (A8, A12, G12, G16, H8 and H12) [50]. Additionally, the work of Liang et al. [51] with sperm whale myoglobin showed the relevance of residues at positions F4, FG3, FG5, and G5, in retaining the porphyrin prosthetic group. They demonstrated that the residues in positions Leu-F4,

His-FG3, Ile-FG5, and Leu-G5 create a hydrophobic environment around the porphyrin preventing hydration of the heme. In our analysis, positions A12, B6, B13, B14, E4, FG5 and H8 show a high level of conservation (100% or >80% conservation level) presenting either the same residue or residues with the same physicochemical properties. The positions A8, B10, C2, CD4, E11, F4, G5, G12, G16, H12 and H19 have lower conservation (>50% conservation level), or no conservation among the sequences.

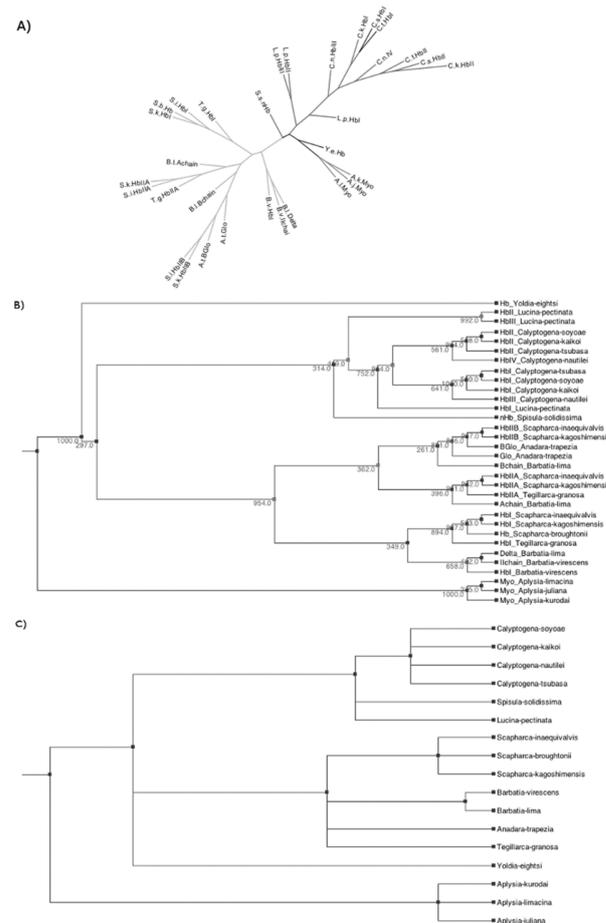


**Figure 2.** GeneDoc visualization of Multiple Sequence Alignment of 32 mollusk protein sequences using PSI-COFFEE. The conservation is shown in three levels based on similarity groups: Black for 100% conservation, Gray for 75% of conservation and light blue for 50% conservation. The abbreviations are the following: L.p (*L. pectinata*), C.n. (*C. nautili*), C.k. (*C. kaikoi*), C.t. (*C. tsubasa*), C.s. (*C. soyoae*), B.l. (*Barbatia lima*), S.s. (*Spisula solidissima*), Y.e. (*Y. eightsi*), A.l. (*Aplysia limacine*), A.k. (*Aplysia kurodai*), A.j. (*Aplysia juliana*), S.k. (*Scapharca kagoshimensis*), S.i. (*Scapharca inaequivalvis*), T.g. (*Tegillarca granosa*), B.v. (*Barbatia virens*), A.t. (*Anadara trapezida*), S.b. (*Scapharca broughtonii*) and P.c. (*P. catodon*). On top of the alignment

the secondary structure of *P. catodon* myoglobin is indicated, according to its crystallographic structure (accession number 1EBC\_A). The organism identifiers are shaded based on the clade division of phylogenetic analysis: red indicates the Bivalves belonging to the Heterodonta subclass; in ‘olive’, we indicate the Bivalves belonging to the Pteriomophia subclass; blue was used for the Gastropods belonging to the Opisthobranchia subclass; and in black the Bivalve, *Y. eightsi*, a member of the Protobranchia subclass.

### 3.2. Phylogenetic Analyses

Figure 3 shows the results of our phylogenetic analyses. A maximum likelihood (ML) gene tree constructed with 32 mollusk globin sequences is shown in Figure 3A. The unrooted ML gene tree shows three major clades: One clade corresponds to the Heterodonta subclass grouping the globins of *L. pectinata* with the *Calyptogena* spp. and *S. solidissima*. Another clade corresponds to those clams belonging to the Pteriomophia subclass grouping the globins of *Barbatia* spp., *Tegillarca* spp., *Anadara* spp., and *Scapharca* spp. A third clade clusters the members of the subclass Opisthobranchia, the myoglobins of the gastropods *Aplysia* spp.; *Y. eightsi*, a clam that belongs to the subclass Protobranchia is shown in black and does not cluster with any group. The rooted-species-reconciled gene tree is shown in Figure 3B. The species tree (Figure 3C), even though it is polytomic, has a well-defined root between the Gastropoda (*Aplysia* spp.) and the Bivalvia (all other groups in our analysis).



**Figure 3.** (A) Maximum-likelihood tree of based on 32 mollusk globin sequences with 1000 bootstraps (bootstrap values are shown in the reconciled tree). (B) Gene Tree. A rooted-species-reconciled phylogenetic gene tree was constructed using PHYLIP, followed by reconciling the gene tree with the species tree and rooting using NOTUNG, as described in Materials and Methods. Numbers represent the bootstrap values from the Maximum Likelihood phylogenetic analysis. (C) Species tree. A rooted species phylogenetic tree was built using the Taxonomy resource at NCBI and verified using the Tree-of-Life resource at <http://tolweb.org> (accessed on 12 November 2018).



We will discuss the similarities of those organisms that live in environments similar to that of *L. pectinata*, where the levels of hydrogen sulfide are high. The members of the *Calyptogena* spp., as well as *L. pectinata*, rely on the symbiosis with chemoautotrophic sulfur-oxidizing bacteria for their nutrition [52–56] and live in reduced environments where low to moderate concentrations of hydrogen sulfide are found, i.e., cold seeps and hydrothermal vents [57]. There are a few positions with conserved residues in this group that are different from the other groups: the distal heme positions Gln-E7, Phe-E11, position Phe-B9, and the B10 position, occupied mainly by a tyrosine, with the exception of HbI<sub>LP</sub> from *L. pectinata* that has phenylalanine. The topological positions B10, E7 and E11 are shown in the structural alignments of HbI<sub>LP</sub> vs. HbII<sub>LP</sub> (Figure 1) and HbI<sub>LP</sub> vs. SWM (Supplementary Figure S1).

Table 1 shows amino acids in positions B9, B10, E7 and E11 for this subgroup. The nerve Hb of *S. solidissima*, which is grouped in the Heterodonta clade with the Hbs from *L. pectinata* and *Calyptogena* spp., has His-E7 and inhabits sandy continental shelf habitats. Hence, this organism was not included when we looked for similarities between organisms living in sulfide-rich environments.

**Table 1.** Amino acids in positions conserved in organism living in sulfide-rich environments.

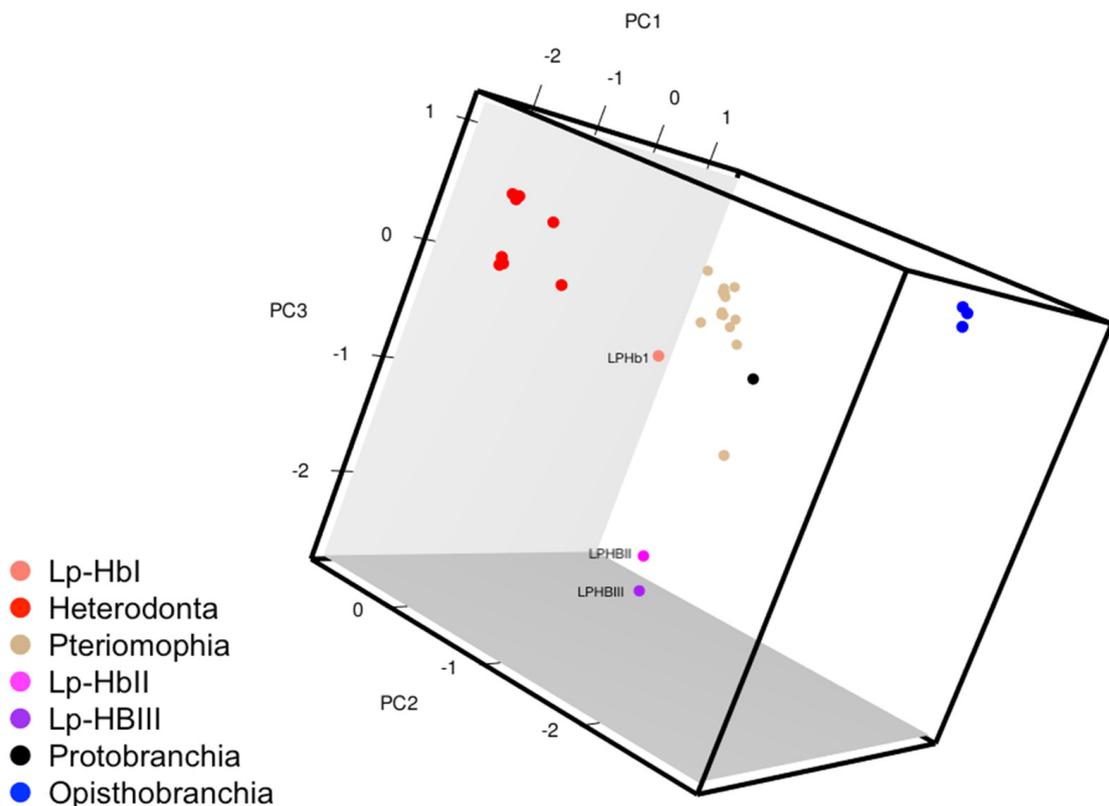
Organism	Globin	Abbreviation	Amino Acids in Positions:			
			B9	B10	E7	E11
<i>L. pectinata</i>	Hemoglobin I	HbI <sub>LP</sub>	Phe	Phe	Gln	Phe
<i>L. pectinata</i>	Hemoglobin II	HbII <sub>LP</sub>	Phe	Tyr	Gln	Phe
<i>L. pectinata</i>	Hemoglobin III	HbIII <sub>LP</sub>	Phe	Tyr	Gln	Phe
<i>C. nautili</i>	Hemoglobin IV	HbIV <sub>Cn</sub>	Phe	Tyr	Gln	Phe
<i>C. nautili</i>	Hemoglobin III	HbIII <sub>Cn</sub>	Phe	Tyr	Gln	Phe
<i>C. kaikoi</i>	Hemoglobin II	HbII <sub>Ck</sub>	Phe	Tyr	Gln	Phe
<i>C. tsubasa</i>	Hemoglobin II	HbII <sub>Ct</sub>	Phe	Tyr	Gln	Phe
<i>C. soyoae</i>	Hemoglobin II	HbII <sub>Cs</sub>	Phe	Tyr	Gln	Phe
<i>C. kaikoi</i>	Hemoglobin I	HbI <sub>Ck</sub>	Phe	Tyr	Gln	Phe
<i>C. soyoae</i>	Hemoglobin I	HbI <sub>Cs</sub>	Phe	Tyr	Gln	Phe
<i>C. tsubasa</i>	Hemoglobin I	HbI <sub>Ct</sub>	Phe	Tyr	Gln	Phe

### 3.3. Metric Multi-Dimensional Scaling Analysis

The Metric Multi-Dimensional Scaling analysis (MMDs) was performed with the trimmed alignment of the 32 mollusk globin sequences creating a multi-dimensional sequence space, with a 3D projection of major principal components shown in Figure 5.

This type of analysis is a powerful method used to visualize distance between sequences and complements phylogeny providing evolutionary information [46]. This sequence space is defined by the principal components (PCs). In Figure 5 we show the 3-D projection of the first three PCs. It is highly informative about the evolution and drift of these proteins. In this analysis we observed, as expected, that these globins are also clustered according to their subfamilies. However, the 3D sequence space clearly shows how HbII<sub>LP</sub> and HbIII<sub>LP</sub> are close to each other and separated from the other members of the Heterodonta subfamily. On the other hand, HbI<sub>LP</sub> is closer to other members of the Heterodonta subfamily despite its unique hydrogen sulfide reactivity. This data shows that HbII<sub>LP</sub> and HbIII<sub>LP</sub> have drifted evolutionarily in the sequence space. The 2D projections for pairs of these three PC are also shown in Supplementary Figure S2 as B, C and D. In this analysis we observed, as expected, that these globins are also clustered according to their subfamilies. The 2D sequence space of PC1 and PC2, Figure S2B, resembles the unrooted phylogenetic tree previously shown. However, the 3D sequence space clearly shows how

HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> are close to each other and separated from the other members of the Heterodonta subfamily. This can also be seen in the other two 2D sequence space shown in Figure S2C,D. These results revealed that even though HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> are O<sub>2</sub> transporting Hbs like the ones found in the *Calypptogena* spp., they are evolving independently as an outgroup while the HbI<sub>Lp</sub> sequence conserves a higher degree of evolutionary similarity to the *Calypptogena* spp. Hemoglobins.



**3-D Plot of Metric Multi Dimensional Scaling of Mollusk Hemoglobins**

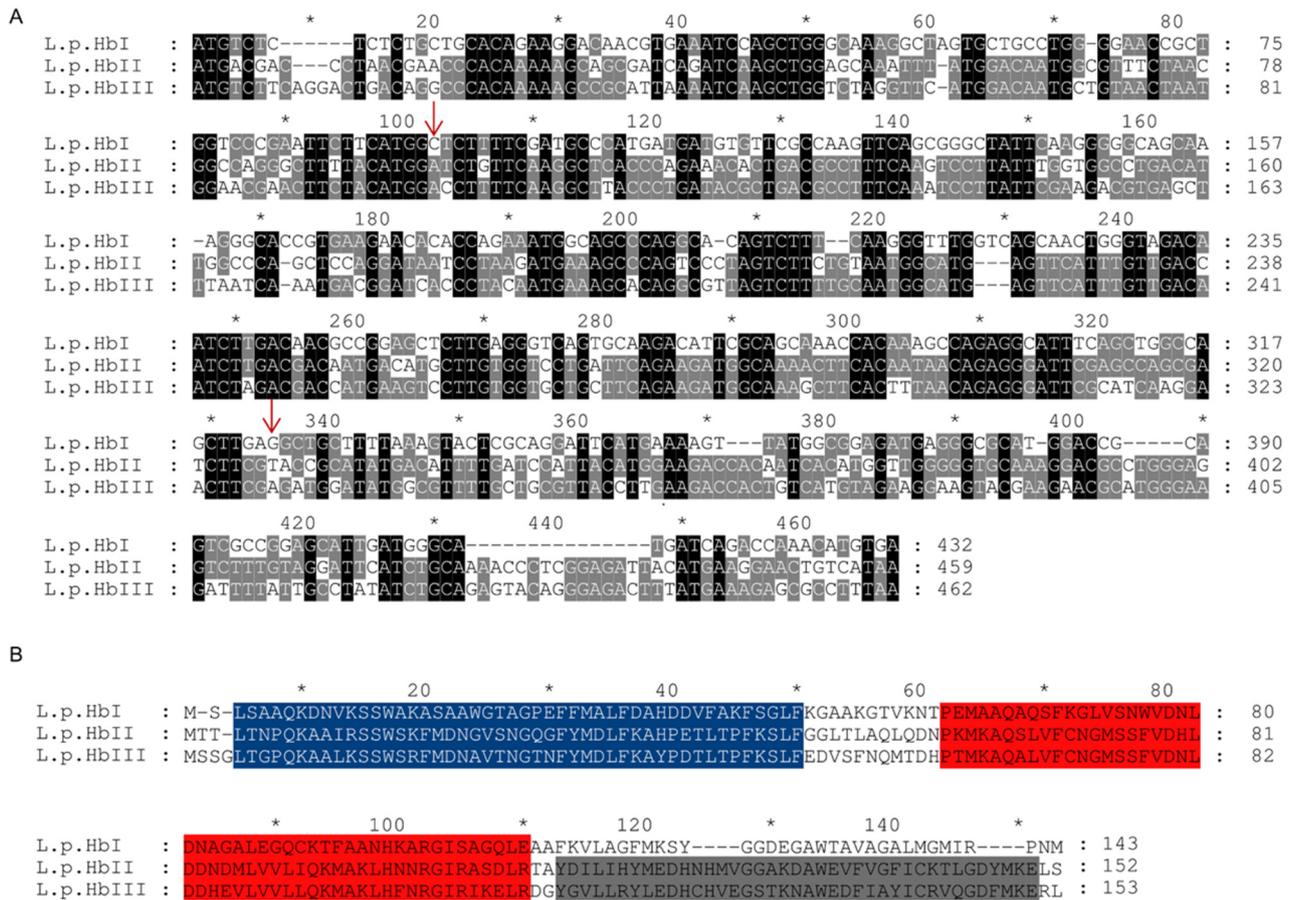
**Figure 5.** 3D representation of the mollusk globins sequence space. A trimmed multiple sequence alignment of 32 homologous mollusk globins was analyzed by MMDS, with distances based on the JTT distance matrix. The 3D space is defined by the first three components of the MDS analysis. HbI<sub>Lp</sub> is colored ‘salmon’, HbII<sub>Lp</sub> is colored in magenta and HbIII<sub>Lp</sub> is colored purple. The rest of the sequences are color coded according to mollusk sub-families they belong to: colored in red are Heterodonta subclass, in tan are the Pteriomophia subclass, in blue are the Opisthobranchia subclass, and in black the Bivalve, *Y. eightsi*, member of the Protobranchia subclass.

### 3.4. Coding Region Bioinformatic Analyses

The level of conservation between the three coding region sequences of HbI<sub>Lp</sub>, HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> of *L. pectinata* is presented in the multiple sequence alignment in Figure 6A.

The three Hbs from *L. pectinata* have intron insertions at the conserved positions B12.2 and G7.0 [58], indicated by the arrows in Figure 6A that mark these intron-insertion positions, respectively (B12.2 is the intron inserted after the second base of the codon of amino acid 12 on the B-helix and G7.0 is the intron inserted between the codons of amino acids 6 and 7 on the G-helix). For the three coding regions we observed high conservation surrounding the regions of intron insertion at the B12.2 position (after nucleotides 95, 98 and 101 in the coding region of HbI<sub>Lp</sub>, HbII<sub>Lp</sub> and HbIII<sub>Lp</sub>, respectively), while there is more variability in the region of the third intron insertion at position G7.0 (after nucleotides 324, 327 and 330 in the coding regions of HbI<sub>Lp</sub>, HbII<sub>Lp</sub> and HbIII<sub>Lp</sub>, respectively). Overall, the region before the second and third intron insertion is more conserved between the three

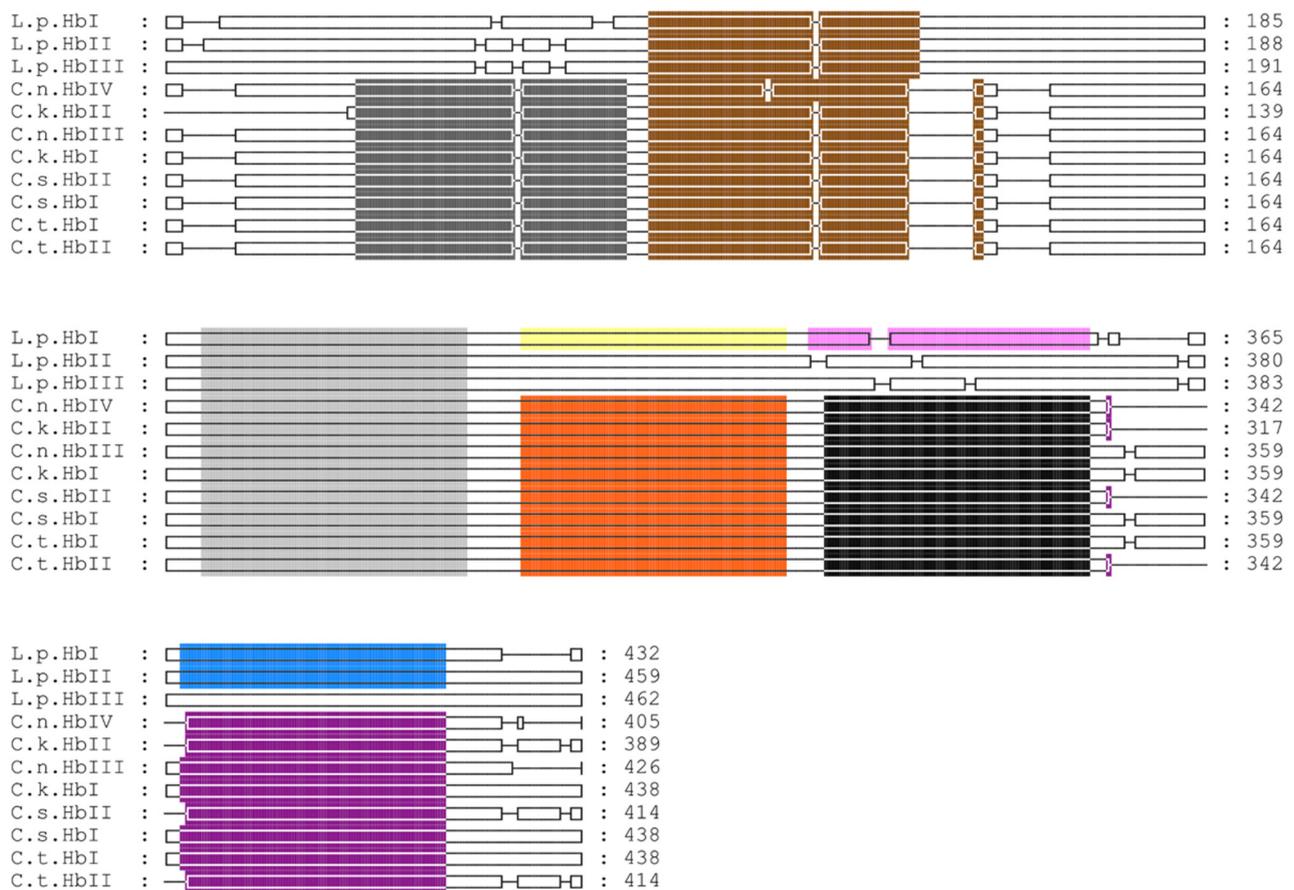
Hb, while the region after the intron insertion at G7.0 is more conserved between HbII<sub>Lp</sub> and HbIII<sub>Lp</sub>, but not for HbI<sub>Lp</sub>. This variability is corroborated at the protein level where motif analysis using MEME showed three major motifs (see Figure 6B), two of them are in the three Hbs (blue and red motifs), while the last motif (colored in gray) is not detected in HbI<sub>Lp</sub>.



**Figure 6.** (A) Multiple sequence alignment and conservation levels of the coding regions of hemoglobins from *L. pectinata*. Conservation levels between the three coding region sequences of HbI<sub>Lp</sub>, HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> of *L. pectinata* were visualized using GeneDoc. Nucleotides shaded in black represent 100% of conservation, in gray, 66% conservation, and nucleotides having less than 66% conservation were not shaded. Red arrows indicate position of intron insertions. (B) Multiple sequence alignment of *L. pectinata* hemoglobins. MSAs were carried out using M-COFFEE and visualized in GeneDoc. Motif analysis was performed using MEME and identified motifs are represented in different colors.

Motif analysis of the hemoglobin coding regions from *L. pectinata* and the genus *Calyptogena* showed that two out of six motifs, colored in brown and light-gray, are shared between all of the members of the Heterodonta clade (see Figure 7).

However, for the HbI<sub>Lp</sub> coding region we find a weak signal for the motifs shared with the *Calyptogena spp.* sequences, indicated in yellow, pink and blue, with *p*-values of  $7.99 \times 10^{-5}$ ,  $6.39 \times 10^{-5}$ , and  $3.01 \times 10^{-5}$  respectively. The HbII<sub>Lp</sub> coding region also shows a weak signal for a motif with a *p*-value of  $1.23 \times 10^{-5}$ , indicated in blue.



**Figure 7.** Multiple sequence alignment of coding regions of hemoglobins from *Lucina pectinata* and *Calyptogena*. Multiple sequence alignments were carried out using M-COFFEE, motif analysis was done using MEME and aligned sequences visualized in GeneDoc. The abbreviations are as follows: L.p. (*L. pectinata*), C.n. (*C. nautili*), C.k. (*C. kaikoi*), C.t. (*C. tsubasa*), C.s. (*C. soyoae*). Detected motifs are represented in different colors.

#### 4. Discussion

Bioinformatic analyses between *L. pectinata* hemoglobins and globin sequences from other mollusks living in sulfide-rich environments revealed sequence conservation in the B10, E7, and E11 residues that give these proteins a unique distal site structure allowing them to carry their normal function in such extreme environments. The mollusk globin sequences used for this analysis differ in terms of subunit structure and ligand specificities. For example, HbI<sub>Ck</sub> and HbII<sub>Ck</sub> from *C. kaikoi* can assume homodimeric or homotetrameric conformation in a concentration dependent manner. Since the amino acid sequence of HbI<sub>Ck</sub> and HbII<sub>Ck</sub> are identical to MbI<sub>Ck</sub> and MbII<sub>Ck</sub>, respectively, found in this clam's muscle tissue, it has been suggested that they may serve as oxygen storage molecules [59]. *C. nautili* hemoglobins HbIII<sub>Cn</sub> and HbIV<sub>Cn</sub> are monomeric hemoglobins while hemoglobins HbI<sub>Ct</sub> and HbII<sub>Ct</sub> from *C. tsubasa* form dimers [60]. In the case of *C. soyoae*, it has two homodimeric hemoglobins, HbI<sub>Cs</sub> and HbII<sub>Cs</sub>, and for HbII<sub>Cs</sub> the autoxidation rate under physiological temperature is three times slower than the one observed for human hemoglobin, which, in combination with its relatively high O<sub>2</sub> affinity, is indicative that it can transport O<sub>2</sub> under physiological conditions [53,54]. On the other hand, globins from blood clams tend to form dimers and tetramers that bind oxygen cooperatively [57,61,62]. Bao et al. [63] described a new role for the Hbs from the blood clam *T. granosa*, since they are involved in the immune defense after the clams were exposed to bacterial infection. Dewilde and collaborators [64] concluded that the nerve hemoglobin (nHb) from the surf

clam *S. solidissima* has a myoglobin-like function since it has a moderate O<sub>2</sub> affinity and phylogenetic analysis with mollusk Hb and vertebrate nHb placed this globin with mollusks [65]. Finally, the Antarctic clam *Y. eightsi* possesses a hemoglobin with O<sub>2</sub> affinity similar to that of mammalian Hb, at the clam's physiological temperature. This affinity decreases significantly at 25 °C, suggestive of this clam's adaptation to a low temperature environment [66].

We looked for similarities between the globins from members of the *Calyptogenia* spp. and *L. pectinata* since both organisms live in environments that have high concentration of H<sub>2</sub>S. For the set of mollusk globin sequences in this study, they show clade-specific conservation or structurally conservative substitutions. The positions with residues conserved in this group are the Gln-E7, Phe-E11, and the Tyr-B10 (with the exception of HbI from *L. pectinata* that has Phe-B10) and position Phe-B9. The globins from members of the Heterodonta subclass and *Y. eightsi* have a phenylalanine at E11, while the rest of the globin sequences possess an aliphatic residue (Val, Ile, or Leu). Several studies have demonstrated the importance of the residue occupying the E11 position since it is part of an overall kinetic barrier for the ligand in the heme pocket [67,68]. The residue at position E7 is considered to be essential for the stabilization of the bound ligand through hydrogen bonding, serves as a gate for ligand entry, and plays a crucial role in the prevention of the autoxidation of the heme [66,69–71]. The stabilization of ligands by hydrogen bonding with glutamine in the E7 position has been described for the three *L. pectinata* Hbs [17,21,72]. Previous work where the HbI<sub>Lp</sub> heme pocket was mutated to mimic the human myoglobin distal site demonstrated that His-E7, in the presence of H<sub>2</sub>S and H<sub>2</sub>O<sub>2</sub> or O<sub>2</sub>, is essential for the formation of sulfhemoglobin, where sulfide is covalently bound in the pyrrole of the porphyrin ring resulting in a non-functional hemoglobin [73]. Interestingly, polymeric-hemoglobin from symbiont-containing tubeworms living in deep-sea hydrothermal vents contain histidine at the E7 position [56] and are capable of binding simultaneously and reversibly O<sub>2</sub> (in their heme pockets) and H<sub>2</sub>S by the formation of zinc-sulfide complexes with Zn ions that are bound to specific residues (B7, B12, B16, CD10, G9, G13 and G16) across the globin chain [74,75]. This zinc-binding function has not been described in the *Lucinidae* spp. or *Calyptogenia* spp. hemoglobins.

We previously mentioned some studies that described the role of several positions across the globin structure and observed that some positions showed less conservation across the mollusk globin sequences in our analysis. Some of these positions correspond to residues located at the immediate distal heme pocket vicinity. Studies have shown that the variability in ligand affinity and autoxidation of these globins are mostly determined by the physicochemical properties of the residues occupying positions at the distal heme pocket. In their review, Springer et al. [66] discussed myoglobin ligand binding recognition in studies where site-directed mutagenesis of distal residues was combined with their structural and physicochemical characterization. Based on that review, it becomes clear that a combination of steric hindrance and polarity of the distal pocket governs hemoglobin ligand recognition, affinity and autoxidation. The positions B10, E11 and E7 of the distal heme environment play a predominant role in the control of ligand binding kinetics [76]. One of the most relevant similarities between these globins is the invariant presence of a glutamine residue in the E7 position. As mentioned before, the existence of a histidine residue in this particular position will promote the formation of sulfhemoglobin in the presence of H<sub>2</sub>S and O<sub>2</sub> or H<sub>2</sub>O<sub>2</sub>, resulting in a nonfunctional protein. This single amino acid substitution may be considered an evolutionary adaptation of these organisms in order to increase their fitness in these sulphide-rich environments.

Like HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> from *L. pectinata*, the hemoglobins from *Calyptogenia* spp. included in this study bind and transport O<sub>2</sub> in the presence of H<sub>2</sub>S and their identical distal site residues distribution may imply a similar binding mechanism. The main difference in the distal cavity between the sulfide reactive hemoglobin from *L. pectinata*, HbI<sub>Lp</sub>, and the oxygen reactive hemoglobins is a single substitution of a phenylalanine in the B10 position instead of a tyrosine, which has been demonstrated to be accountable of HbI<sub>Lp</sub>'s

unique reactivity towards H<sub>2</sub>S [77]. HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> have high O<sub>2</sub> affinities and they remain oxygenated in the presence of H<sub>2</sub>S [78,79] and their O<sub>2</sub> dissociation rates are three times slower than dissociation from HbI<sub>Lp</sub> and very slow compared when compared with other Hbs [14]. In presence of H<sub>2</sub>S, HbI<sub>Lp</sub> forms a ferric hemoglobin sulfide [14]. Studies with other hemoglobins that also have Gln-E7 and Tyr-B10 in their heme pockets, revealed that Tyr-B10 directly interacts with the bound O<sub>2</sub>, forming a hydrogen bond network between Gln-E7, B10 Tyr, and O<sub>2</sub>, that accounts for the low O<sub>2</sub> dissociation rate of these hemoglobins [80–82]. Several studies have shown how Phe-B9 affects the orientation of the aromatic residues in the positions B10 and E11 due to electrostatic aromatic interactions [17,83]. Therefore, comparing the hemoglobin sequences from *L. pectinata* to sperm whale myoglobin allows presenting that a single amino acid in the heme proteins: Gln-E7 leads to transport of hydrogen sulphide in *L. pectinata* hemoglobin I, while His-E7 leads to the modification of the heme group to generate sulfmyoglobin. Thus, a single amino acid substitution can lead to differences in these heme proteins' function. Studies with the sperm whale myoglobin triple mutant (LeuB10Phe, HisE7Gln, and ValE11Phe), mimicking the HbI<sub>Lp</sub> distal environment, showed a change in orientation, not only in Phe-B10 but also for Phe-E11, which assume a more perpendicular position to the heme plane, since the B9 position is a isoleucine residue, resulting in destabilization of the bound ligand [84].

The phylogenetic analysis shows a consistent and robust gene tree. Based on these analyses, we suggest that a first gene duplication at an early evolutionary stage resulted in the ancestor of the HbI<sub>Lp</sub> gene and an ancestor of the HbII<sub>Lp</sub>/HbIII<sub>Lp</sub> gene. A more recent gene duplication of the oxygen-binding gene occurred to give rise to genes for two oxygen-binding proteins, HbII<sub>Lp</sub>, and HbIII<sub>Lp</sub>. This hypothesis is supported with the fact that HbI<sub>Lp</sub> not only showed lower sequence similarity to both HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> but also has different ligand kinetics. Furthermore, this corroborates the similarities at the sequence level and kinetic mechanisms between HbII<sub>Lp</sub> and HbIII<sub>Lp</sub>. In addition, HbI<sub>Lp</sub> appears to have been changing over time, apparently not subject to the constraint of binding oxygen, acquiring a unique function for a specialized environment.

Metric Multi-Dimensional Scaling analysis of mollusk hemoglobins showed evolutionary drift of HbII<sub>Lp</sub> and HbIII<sub>Lp</sub>, while HbI<sub>Lp</sub> was closer to the *Calypptogena* spp. hemoglobins in sequence space. The MMDS shows that HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> have drifted evolutionarily in the sequence space. These results revealed that even though HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> are O<sub>2</sub> transporting Hbs like the ones found in the *Calypptogena* spp., they are evolving independently as a separate subgroup while the HbI<sub>Lp</sub> sequence maintains a higher degree of similarity to the *Calypptogena* spp. hemoglobins.

The coding region analysis of hemoglobins from *L. pectinata* with those from *Calypptogena* spp., suggest that HbI<sub>Lp</sub> from *L. pectinata* has features that are shared with the *Calypptogena* spp. hemoglobins. These similarities are supported by previous work by Suzuki and collaborators, who proposed that the *L. pectinata* and *Calypptogena* spp. hemoglobin chains evolved from a common ancestor [59]. The conservation of intron locations interrupting the coding region of their cognate genes is further evidence of the conservation of globin gene structure in invertebrate hemoglobins.

## 5. Conclusions

Bioinformatic analyses between *L. pectinata* hemoglobins and globin sequences from other mollusks living in sulfide-rich environments revealed what appears to be clade-specific sequence conservation in residues that give these proteins a unique distal heme pocket site structure that allows for normal function in such extreme environments. One of the most relevant similarities between these globins is the invariant presence of a glutamine residue in the E7 position. This single amino acid substitution may be considered an evolutionary adaptation of these organisms in order to increase their fitness in these sulphide-rich environments. Phylogenetics suggests that in *L. pectinata* a first gene duplication resulting in sulfide binding and oxygen binding genes. A more recent gene duplication gave rise

to the two oxygen-binding hemoglobins. Metric Multi-Dimensional Scaling analysis of mollusk hemoglobins showed evolutionary drift of HbII<sub>Lp</sub> and HbIII<sub>Lp</sub>, while HbI<sub>Lp</sub> was closer to the *Calyptogena* spp. hemoglobins in sequence space. This is further corroborated by the conservation seen in the coding region of hemoglobins from *L. pectinata* with those from *Calyptogena* spp.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes13112041/s1>: (a) Table S1: List of organism and globins used for MSA and abbreviations used for analysis, including BLAST results. Includes the abbreviations used, the description of the sequence, the percent similarity and E-value of the BLAST searches carried out using the *L. pectinata* hemoglobins, and the accession numbers of all sequences used for this work; (b) Figure S1: (A) Structural alignment of HbI<sub>Lp</sub> protein structure (PDB 1FLP) and HbII<sub>Lp</sub> (PDB 2OLP) showing their amino acid distribution in the heme pocket. Alignment was done using the STAMP algorithm as implemented in MultiSeq in VMD. The only difference between the two heme pockets is the amino acid in position B10, a phenylalanine for HbI<sub>Lp</sub> and a tyrosine for HbII<sub>Lp</sub>. Both hemoglobins have a phenylalanine at positions CD1 and E11, a glutamine at position E7 and a histidine at position F8. HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> have the same heme pocket amino acid distribution. (B) Structural alignment of HbI<sub>Lp</sub> protein structure (PDB 1FLP) and Sperm Whale Myoglobin (SWM) (PDB 1A6M) showing their amino acid distribution in the heme pocket. Amino acid residues of heme pocket of SWM are colored in blue. For SWM there is a leucine at position B10, a histidine at position E7 and a valine in position E11. Figure S2: (A) 3D representation of the mollusk globins sequence space. A typical multiple sequence alignment of 32 homologous mollusk globins was analyzed by MMDS, with distances based on the JTT distance matrix. The 3D space is defined by the first three components of the MDS analysis. HbI<sub>Lp</sub> is colored in gold, HbII<sub>Lp</sub> is colored in pink and HbIII<sub>Lp</sub> is colored purple. The rest of the sequences are color coded according to mollusk sub-families they belong to: colored in red are Heterodonta subclass, in green are the Pteriomophia subclass, in blue are the Opisthobranchia subclass, and in black the Bivalve, Y. eightsi, member of the Protobranchia subclass. (B) The 2D sequence space of mollusk globin formed by PC1 and PC2. (C) The 2D sequence space of mollusk globin formed by PC3 and PC2. (D) The 2D sequence space of mollusk globin formed by PC1 and PC3. We show the 3-D projection of the first three PC. It is highly informative about the evolution and drift of these proteins. The 2D projections for pairs of these three PC are also shown in Supplementary Figure S2 as B, C and D. In this analysis we observed, as expected, that these globins are also clustered according to their subfamilies. The 2D sequence space of PC1 and PC2, Supplementary Figure S2B, resembles the unrooted phylogenetic tree previously shown. However, the 3D sequence space clearly shows how HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> are close to each other and separated from the other members of the Heterodonta subfamily. This can also be seen in the other two 2D sequence space shown in Supplementary Figure S2C,D. This data shows that HbII<sub>Lp</sub> and HbIII<sub>Lp</sub> have drifted evolutionarily in the sequence space.

**Author Contributions:** I.M.M.-R. carried out parts of the bioinformatic and structural analysis and drafted and edited the manuscript. R.G.-M. contributed to the design of the study and contributed to the bioinformatic experimental design, analysis, and interpretation, performed the phylogenetic analyses, the structural alignment, the MMDS analysis, and helped to draft and edit the manuscript. C.L.C. helped with the interpretation of bioinformatic analysis and helped to draft and edit the manuscript. J.L.-G. contributed to the design of the study, helped with the structural interpretation, and helped to draft and edit the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** NIH grants U54MD007600 and P20GM103475, NSF grant MCB-0843608, UPRM RISE 2BEST NIH-R25GM088023, and UPRM-SLOAN grants provided partial support for this project. This research was also supported in part by the National Institutes of Health through resources provided by the National Resource for Biomedical Supercomputing (P41 RR06009), which is part of the Pittsburgh Supercomputing Center (PSC). This work, and in part RGM and IMMR, were supported by National Institutes of Health Minority Access to Research Careers (MARC) grant T36-GM095335 to the PSC. Specifically, this work used the computational resource BioU, which was made available by National Institutes of Health grant T36-GM008789 to the PSC.

**Institutional Review Board Statement:** Not Applicable. This work relies on a reanalysis of protein, nucleic acid sequences, and structures using publicly available data from international repositories.

**Informed Consent Statement:** Not Applicable. We did not use human data for this work.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** This work was done in partial fulfillment of the doctoral dissertation requirements of I.M.M.-R.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Weber, R.E.; Vinogradov, S.N. Nonvertebrate hemoglobins: Functions and molecular adaptations. *Physiol. Rev.* **2001**, *81*, 569–628. [[CrossRef](#)] [[PubMed](#)]
2. Goodman, M.; Pedwaydon, J.; Czelusniak, J.; Suzuki, T.; Gotoh, T.; Moens, L.; Shishikura, F.; Walz, D.; Vinogradov, S. An evolutionary tree for invertebrate globin sequences. *J. Mol. Evol.* **1988**, *27*, 236–249. [[CrossRef](#)] [[PubMed](#)]
3. Edmondson, A.B.; Hirs, C.H. The amino-acid sequence of sperm whale myoglobin. Chemical studies. *Nature* **1961**, *190*, 663–665. [[CrossRef](#)]
4. Perutz, M.F. Structure and function of haemoglobin: I. A tentative atomic model of horse oxyhaemoglobin. *J. Mol. Biol.* **1965**, *13*, 646–668. [[CrossRef](#)]
5. Perutz, M.F.; Muirhead, H.; Cox, J.M.; Goaman, L.C. Three-dimensional Fourier synthesis of horse oxyhaemoglobin at 2.8 Å resolution: The atomic model. *Nature* **1968**, *219*, 131–139. [[CrossRef](#)] [[PubMed](#)]
6. Perutz, M.F. Regulation of oxygen affinity of hemoglobin: Influence of structure of the globin on the heme iron. *Annu. Rev. Biochem.* **1979**, *48*, 327–386. [[CrossRef](#)]
7. Bashford, D.; Chothia, C.; Lesk, A.M. Determinants of a protein fold. Unique features of the globin amino acid sequences. *J. Mol. Biol.* **1987**, *196*, 199–216. [[CrossRef](#)]
8. Lesk, A.M.; Chothia, C. How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **1980**, *136*, 225–270. [[CrossRef](#)]
9. Childress, J.J. Uptake and transport of sulfide in marine invertebrates. In *Comparative Physiology: Life in Water and on Land*; Dejours, P., Bolis, L., Taylor, C.R., Weibel, E.R., Eds.; Springer: Berlin/Heidelberg, Germany, 1987; Volume 9, pp. 231–239.
10. Felbeck, H.; Childress, J.J.; Somero, G.N. Calvin-Benson Cycle and sulphide oxidation enzymes in animals from sulphide-rich habitats. *Nature* **1981**, *293*, 291–293. [[CrossRef](#)]
11. Kim, Y.W.; Yasuda, M.; Yamagishi, A.; Oshima, T.; Ohta, S. Characterization of the endosymbiont of a deep-sea bivalve, *Calyptogena soyocae*. *Appl. Environ. Microbiol.* **1995**, *61*, 823–827. [[CrossRef](#)]
12. Chatfield, M.J.; La Mar, G.N.; Kauten, R.J. Proton NMR characterization of isomeric sulfmyoglobins: Preparation, interconversion, reactivity patterns, and structural features. *Biochemistry* **1987**, *26*, 6939–6950. [[CrossRef](#)] [[PubMed](#)]
13. Taylor, J.D.; Glover, E.A. Functional anatomy, chemosymbiosis and evolution of the Lucinidae. *Geol. Soc. London, Spéc. Publ.* **2000**, *177*, 207–225. [[CrossRef](#)]
14. Kraus, D.W.; Wittenberg, J.B. Hemoglobins of the *Lucina pectinata*/bacteria symbiosis. I. Molecular properties, kinetics and equilibria of reactions with ligands. *J. Biol. Chem.* **1990**, *265*, 16043–16053. [[CrossRef](#)]
15. Rizzi, M.; Wittenberg, J.B.; Coda, A.; Fasano, M.; Ascenzi, P.; Bolognesi, M. Structure of the sulfide-reactive hemoglobin from the clam *Lucina pectinata*. Crystallographic analysis at 1.5 Å resolution. *J. Mol. Biol.* **1994**, *244*, 86–99. [[CrossRef](#)]
16. Gibson, Q.H.; Wittenberg, J.B.; Wittenberg, B.A.; Bogusz, D.; Appleby, C.A. The kinetics of ligand binding to plant hemoglobins. Structural implications. *J. Biol. Chem.* **1989**, *264*, 100–107. [[CrossRef](#)]
17. Rizzi, M.; Wittenberg, J.B.; Coda, A.; Ascenzi, P.; Bolognesi, M. Structural bases for sulfide recognition in *Lucina pectinata* hemoglobin I. *J. Mol. Biol.* **1996**, *258*, 1–5. [[CrossRef](#)]
18. Kraus, D.W.; Wittenberg, J.B.; Lu, J.F.; Peisach, J. Hemoglobins of the *Lucina pectinata*/bacteria symbiosis. II. An electron paramagnetic resonance and optical spectral study of the ferric proteins. *J. Biol. Chem.* **1990**, *265*, 16054–16059. [[CrossRef](#)]
19. Gavira, J.A.; Camara-Artigas, A.; De Jesús-Bonilla, W.; López-Garriga, J.; Lewis, A.; Pietri, R.; García-Ruiz, J.M. Structure and ligand selection of hemoglobin II from *Lucina pectinata*. *J. Biol. Chem.* **2008**, *283*, 9414–9423. [[CrossRef](#)]
20. Pietri, R.; León, R.G.; Kiger, L.; Marden, M.C.; Granell, L.B.; Cadilla, C.L.; López-Garriga, J. Hemoglobin I from *Lucina pectinata*: A model for distal heme-ligand control. *Biochim. Biophys. Acta* **2006**, *1764*, 758–765. [[CrossRef](#)]
21. Pietri, R.; Lewis, A.; León, R.G.; Casabona, G.; Kiger, L.; Yeh, S.R.; López-Garriga, J. Factors controlling the reactivity of hydrogen sulfide with hemeproteins. *Biochemistry* **2009**, *48*, 4881–4894. [[CrossRef](#)]
22. Ramos-Alvarez, C.; Yoo, B.K.; Pietri, R.; Lamarre, I.; Martin, J.L.; Lopez-Garriga, J.; Negreirer, M. Reactivity and dynamics of H<sub>2</sub>S, NO, and O<sub>2</sub> interacting with hemoglobins from *Lucina pectinata*. *Biochemistry* **2013**, *52*, 7007–7021. [[CrossRef](#)] [[PubMed](#)]
23. Ruiz-Martínez, C.R.; Nieves-Marrero, C.A.; Estremera-Andújar, R.A.; Gavira, J.A.; González-Ramírez, L.A.; López-Garriga, J.; García-Ruiz, J.M. Crystallization and diffraction patterns of the oxy and cyano forms of the *Lucina pectinata* haemoglobins complex. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2008**, *65 Pt 1*, 25–28. [[CrossRef](#)] [[PubMed](#)]

24. Benson, D.A.; Karsch-Mizrachi, I.; Clark, K.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2012**, *40*, D48–D53. [[CrossRef](#)] [[PubMed](#)]
25. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)]
26. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [[CrossRef](#)]
27. Johnson, M.; Zaretskaya, I.; Raytselis, Y.; Merezuk, Y.; McGinnis, S.; Madden, T.L. NCBI BLAST: A better web interface. *Nucleic Acids Res.* **2008**, *36*, W5–W9. [[CrossRef](#)]
28. Taly, J.F.; Magis, C.; Bussotti, G.; Chang, J.M.; Di Tommaso, P.; Erb, I.; Notredame, C. Using the T-Coffee package to build multiple sequence alignments of protein, RNA, DNA sequences and 3D structures. *Nat. Protoc.* **2011**, *6*, 1669–1682. [[CrossRef](#)]
29. Nicholas, K.B.; Nicholas, H.B.; Deerfield II, D.W. GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNW News* **1997**, *4*, 14.
30. Roberts, E.; Eargle, J.; Wright, D.; Luthey-Schulten, Z. MultiSeq: Unifying sequence and structure data for evolutionary analysis. *BMC Bioinform.* **2006**, *7*, 382. [[CrossRef](#)]
31. Russell, R.B.; Barton, G.J. Multiple protein sequence alignment from tertiary structure comparison: Assignment of global and residue confidence levels. *Proteins* **1992**, *14*, 309–323. [[CrossRef](#)]
32. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
33. Capella-Gutiérrez, S.; Silla-Martínez, J.M.; Gabaldón, T. TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)] [[PubMed](#)]
34. Felsenstein, J. PHYLIP-Phylogeny Inference Package, Version 3.2. *Cladistics* **1989**, *5*, 164–166.
35. Ropelewski, A.J.; Nicholas, H.B.; Gonzalez Mendez, R.R. MPI-PHYLIP: Parallelizing computationally intensive phylogenetic analysis routines for the analysis of large protein families. *PLoS ONE* **2010**, *5*, e13999. [[CrossRef](#)] [[PubMed](#)]
36. Rambaut, A. *FigTree, v.1.4.4*; Institute of Evolutionary Biology, University of Edinburgh: Edinburgh, UK, 2010; Available online: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 30 August 2022).
37. Schoch, C.L.; Ciufo, S.; Domrachev, M.; Hotton, C.L.; Kannan, S.; Khovanskaya, R.; Leipe, D.; Mcveigh, R.; O’Neill, K.; Rob-bertse, B.; et al. NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database* **2020**, *2020*, baaa062. [[CrossRef](#)] [[PubMed](#)]
38. Sayers, E.W.; Barrett, T.; Benson, D.A.; Bryant, S.H.; Canese, K.; Chetvernin, V.; Church, D.M.; DiCuccio, M.; Edgar, R.; Federhen, S.; et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2009**, *37*, D5–D15. [[CrossRef](#)]
39. Maddison, D.R.; Schulz, K.-S.; Maddison, W.P. The Tree of Life Web Project. *Zootaxa* **2007**, *1668*, 19–40. [[CrossRef](#)]
40. Durand, D.; Halldórsson, B.V.; Vernot, B. A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J. Comput. Biol.* **2006**, *13*, 320–335. [[CrossRef](#)]
41. Vernot, B.; Stolzer, M.; Goldman, A.; Durand, D. Reconciliation with non-binary species trees. *J. Comput. Biol.* **2008**, *15*, 981–1006. [[CrossRef](#)]
42. Chen, K.; Durand, D.; Farach-Colton, M. NOTUNG: A program for dating gene duplications and optimizing gene family trees. *J. Comput. Biol.* **2000**, *7*, 429–447. [[CrossRef](#)]
43. Felsenstein, J. *Inferring Phylogenies*, 2nd ed.; Sinauer Associates, Inc.: Sunderland, MA, USA, 2003.
44. Zmasek, C.M.; Eddy, S.R. ATV: Display and manipulation of annotated phylogenetic trees. *Bioinformatics* **2001**, *17*, 383–384. [[CrossRef](#)] [[PubMed](#)]
45. Pelé, J.; Abdi, H.; Moreau, M.; Thybert, D.; Chabbert, M. Multidimensional scaling reveals the main evolutionary pathways of class A G-protein-coupled receptors. *PLoS ONE* **2011**, *6*, e19094. [[CrossRef](#)] [[PubMed](#)]
46. Pelé, J.; Bécu, J.M.; Abdi, H.; Chabbert, M. Bios2mds: An R package for comparing orthologous protein families by metric multidimensional scaling. *BMC Bioinform.* **2012**, *13*, 133. [[CrossRef](#)] [[PubMed](#)]
47. Moretti, S.; Armougom, F.; Wallace, I.M.; Higgins, D.G.; Jongeneel, C.V.; Notredame, C. The M-Coffee web server: A meta-method for computing multiple sequence alignments by combining alternative alignment methods. *Nucleic Acids Res.* **2007**, *35*, W645–W648. [[CrossRef](#)] [[PubMed](#)]
48. Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1994**, *2*, 28–36.
49. Furuta, H.; Kajita, A. Dimeric hemoglobin of the bivalve mollusc *Anadara broughtonii*: Complete amino acid sequence of the globin chain. *Biochemistry* **1983**, *22*, 917–922. [[CrossRef](#)]
50. Ptitsyn, O.B.; Ting, K.L. Non-functional conserved residues in globins and their possible role as a folding nucleus. *J. Mol. Biol.* **1999**, *291*, 671–682. [[CrossRef](#)]
51. Liong, E.C.; Dou, Y.; Scott, E.E.; Olson, J.S.; Phillips, G.N. Waterproofing the heme pocket. Role of proximal amino acid side chains in preventing heme loss from myoglobin. *J. Biol. Chem.* **2001**, *276*, 9093–9100. [[CrossRef](#)]
52. Kraus, D.W. Heme Proteins in Sulfide-Oxidizing Bacteria/Mollusc Symbioses. *Am. Zool.* **1995**, *35*, 112–120. [[CrossRef](#)]
53. Suzuki, T.; Takagi, T.; Ohta, S. Amino acid sequence of the dimeric hemoglobin (Hb I) from the deep-sea cold-seep clam *Calyptogena soyoae* and the phylogenetic relationship with other molluscan globins. *Biochim. Biophys. Acta* **1989**, *999*, 254–259. [[CrossRef](#)]

54. Suzuki, T.; Takagi, T.; Ohta, S. Primary structure of a dimeric haemoglobin from the deep-sea cold-seep clam *Calyptogena soyoeae*. *Biochem. J.* **1989**, *260*, 177–182. [[CrossRef](#)] [[PubMed](#)]
55. Van Dover, C.L. *The Ecology of Deep-Sea Hydrothermal Vents*; Princeton University Press: Princeton, NJ, USA, 2000.
56. Vetter, R.D.; Wells, M.E.; Kurtsman, A.L.; Somero, G.N. Sulfide Detoxification by the Hydrothermal Vent Crab *Bythograea thermydron* and Other Decapod Crustaceans. *Physiol. Zool.* **1987**, *60*, 121–137. [[CrossRef](#)]
57. Suzuki, T.; Arita, T.; Kawasaki, Y. The cDNA-derived amino acid sequence of chain II of the heterodimeric hemoglobin from the blood clam *Barbatia virescens*. *Zool. Sci.* **1995**, *12*, 453–455. [[CrossRef](#)]
58. Montes-Rodríguez, I.M.; Rivera, L.E.; López-Garriga, J.; Cadilla, C.L. Characterization and Expression of the *Lucina pectinata* Oxygen and Sulfide Binding Hemoglobin Genes. *PLoS ONE* **2016**, *11*, e0147977. [[CrossRef](#)] [[PubMed](#)]
59. Suzuki, T.; Kawamichi, H.; Ohtsuki, R.; Iwai, M.; Fujikura, K. Isolation and cDNA-derived amino acid sequences of hemoglobin and myoglobin from the deep-sea clam *Calyptogena kaikoi*. *Biochim. Biophys. Acta* **2000**, *1478*, 152–158. [[CrossRef](#)]
60. Kawano, K.; Iwasaki, N.; Suzuki, T. Notable diversity in hemoglobin expression patterns among species of the deep-sea clam, *Calyptogena*. *Cell. Mol. Life Sci.* **2003**, *60*, 1952–1956. [[CrossRef](#)]
61. Chiancone, E.; Vecchini, P.; Verzili, D.; Ascoli, F.; Antonini, E. Dimeric and tetrameric hemoglobins from the mollusc *Scapharca inaequivalvis*. Structural and functional properties. *J. Mol. Biol.* **1981**, *152*, 577–592. [[CrossRef](#)]
62. Suzuki, T.; Kawasaki, Y.; Arita, T.; Nakamura, A. Two-domain haemoglobin of the blood clam *Barbatia lima* resulted from the recent gene duplication of the single-domain delta chain. *Biochem. J.* **1996**, *313 Pt 2*, 561–566. [[CrossRef](#)]
63. Bao, Y.B.; Wang, Q.; Guo, X.M.; Lin, Z.H. Structure and immune expression analysis of hemoglobin genes from the blood clam *Tegillarca granosa*. *Genet. Mol. Res.* **2013**, *12*, 3110–3123. [[CrossRef](#)]
64. Dewilde, S.; Ebner, B.; Vinck, E.; Gilany, K.; Hankeln, T.; Burmester, T.; Kreiling, J.; Reinisch, C.; Vanfleteren, J.R.; Kiger, L.; et al. The nerve hemoglobin of the bivalve mollusc *Spisula solidissima*: Molecular cloning, ligand binding studies, and phylogenetic analysis. *J. Biol. Chem.* **2006**, *281*, 5364–5372. [[CrossRef](#)]
65. Dewilde, S.; Angelini, E.; Kiger, L.; Marden, M.C.; Beltramini, M.; Salvato, B.; Moens, L. Structure and function of the globin and globin gene from the Antarctic mollusc *Yoldia eightsi*. *Biochem. J.* **2003**, *370 Pt 1*, 245–253. [[CrossRef](#)] [[PubMed](#)]
66. Springer, B.A.; Sligar, S.G.; Olson, J.S.; Phillips, G.N. Mechanisms of Ligand Recognition in Myoglobin. *Chem. Rev.* **1994**, *94*, 699–714. [[CrossRef](#)]
67. Egeberg, K.D.; Springer, B.A.; Sligar, S.G.; Carver, T.E.; Rohlfs, R.J.; Olson, J.S. The role of Val68(E11) in ligand binding to sperm whale myoglobin. Site-directed mutagenesis of a synthetic gene. *J. Biol. Chem.* **1990**, *265*, 11788–11795. [[CrossRef](#)]
68. Ramirez, E.; Cruz, A.; Rodriguez, D.; Uchima, L.; Pietri, R.; Santana, A.; López, G.E. Effects of active site mutations in haemoglobin I from *Lucina pectinata*: A molecular dynamic study. *Mol. Simul.* **2008**, *34*, 715–725. [[CrossRef](#)]
69. Springer, B.A.; Egeberg, K.D.; Sligar, S.G.; Rohlfs, R.J.; Mathews, A.J.; Olson, J.S. Discrimination between oxygen and carbon monoxide and inhibition of autooxidation by myoglobin. Site-directed mutagenesis of the distal histidine. *J. Biol. Chem.* **1989**, *264*, 3057–3060. [[CrossRef](#)]
70. Birukou, I.; Schweers, R.L.; Olson, J.S. Distal histidine stabilizes bound O<sub>2</sub> and acts as a gate for ligand entry in both subunits of adult human hemoglobin. *J. Biol. Chem.* **2010**, *285*, 8840–8854. [[CrossRef](#)]
71. Jensen, F.B. Comparative analysis of autooxidation of haemoglobin. *J. Exp. Biol.* **2001**, *204 Pt 11*, 2029–2033. [[CrossRef](#)]
72. Pietri, R.; Román-Morales, E.; López-Garriga, J. Hydrogen sulfide and heme proteins: Knowledge and mysteries. *Antioxid. Redox Signal.* **2011**, *15*, 393–404. [[CrossRef](#)]
73. Román-Morales, E.; Pietri, R.; Ramos-Santana, B.; Vinogradov, S.N.; Lewis-Ballester, A.; López-Garriga, J. Structural determinants for the formation of sulfhemeprotein complexes. *Biochem. Biophys. Res. Commun.* **2010**, *400*, 489–492. [[CrossRef](#)]
74. Arp, A.J.; Childress, J.J.; Vetter, R.D. The sulphide-binding protein in the blood of the vestimentiferan tube-worm, *Riftia pachyptila*, is the extracellular haemoglobin. *J. Exp. Biol.* **1987**, *128*, 139–158. [[CrossRef](#)]
75. Flores, J.F.; Hourdez, S.M. The zinc-mediated sulfide-binding mechanism of hydrothermal vent tubeworm 400-kDa hemoglobin. *Cah. Biol. Mar.* **2006**, *47*, 371–377.
76. Carlson, M.L.; Regan, R.; Elber, R.; Li, H.; Phillips, G.N.; Olson, J.S.; Gibson, Q.H. Nitric oxide recombination to double mutants of myoglobin: Role of ligand diffusion in a fluctuating heme pocket. *Biochemistry* **1994**, *33*, 10597–10606. [[CrossRef](#)]
77. Pietri, R.; Granell, L.; Cruz, A.; De Jesús, W.; Lewis, A.; Leon, R.; Garriga, J.L. Tyrosine B10 and heme-ligand interactions of *Lucina pectinata* hemoglobin II: Control of heme reactivity. *Biochim. Biophys. Acta* **2005**, *1747*, 195–203. [[CrossRef](#)] [[PubMed](#)]
78. Hockenhull-Johnson, J.D.; Stern, M.S.; Martin, P.; Dass, C.; Desiderio, D.M.; Wittenberg, J.B.; Walz, D.A. The amino acid sequence of hemoglobin II from the symbiont-harboring clam *Lucina pectinata*. *J. Protein Chem.* **1991**, *10*, 609–622. [[CrossRef](#)] [[PubMed](#)]
79. Hockenhull-Johnson, J.D.; Stern, M.S.; Wittenberg, J.B.; Vinogradov, S.N.; Kapp, O.H.; Walz, D.A. The amino acid sequence of hemoglobin III from the symbiont-harboring clam *Lucina pectinata*. *J. Protein Chem.* **1993**, *12*, 261–277. [[CrossRef](#)] [[PubMed](#)]
80. Peterson, E.S.; Huang, S.; Wang, J.; Miller, L.M.; Vidugiris, G.; Kloek, A.P.; Friedman, J.M. A comparison of functional and structural consequences of the tyrosine B10 and glutamine E7 motifs in two invertebrate hemoglobins (*Ascaris suum* and *Lucina pectinata*). *Biochemistry* **1997**, *36*, 13110–13121. [[CrossRef](#)] [[PubMed](#)]
81. Wittenberg, J.B.; Bolognesi, M.; Wittenberg, B.A.; Guertin, M. Truncated hemoglobins: A new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants. *J. Biol. Chem.* **2002**, *277*, 871–874. [[CrossRef](#)]

82. Igarashi, J.; Kobayashi, K.; Matsuoka, A. A hydrogen-bonding network formed by the B10-E7-E11 residues of a truncated hemoglobin from *Tetrahymena pyriformis* is critical for stability of bound oxygen and nitric oxide detoxification. *J. Biol. Inorg. Chem.* **2011**, *16*, 599–609. [[CrossRef](#)]
83. Tarricone, C.; Galizzi, A.; Coda, A.; Ascenzi, P.; Bolognesi, M. Unusual structure of the oxygen-binding site in the dimeric bacterial hemoglobin from *Vitreoscilla* sp. *Structure* **1997**, *5*, 497–507. [[CrossRef](#)]
84. Nguyen, B.D.; Zhao, X.; Vyas, K.; La Mar, G.N.; Lile, R.A.; Brucker, E.A.; Phillips, G.N.; Olson, J.; Wittenberg, J.B. Solution and crystal structures of a sperm whale myoglobin triple mutant that mimics the sulfide-binding hemoglobin from *Lucina pectinata*. *J. Biol. Chem.* **1998**, *273*, 9517–9526. [[CrossRef](#)]