Evolution of respiratory adaptations in hydrothermal vent scale-worms (Polynoidae)

Joana Cristina Projecto-garcia

Avertissement

THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE

Spécialité

Evolution moléculaire et adaptation

(Ecole doctorale IVIV)

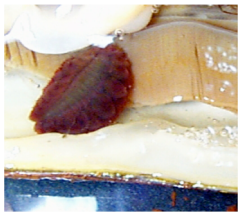
Présentée par

Mlle. Joana Cristina Projecto Garcia

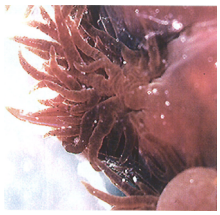
Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Evolution of respiratory adaptations in hydrothermal vent scale-worms (Polynoidae)







soutenue le 16 Novembre 2009

devant le jury composé de :

Dr. Stéphane HOURDEZ, CNRS, Station Biologique de Roscoff	Directeur de thèse
Prof. François H. LALLIER, UPMC, Station Biologique de Roscoff	Co-directeur de thèse
Prof. Charles R. FISHER, Pennsylvania State University	Rapporteur
Dr. Sylvia DEWILDE, Antwerpen University	Rapporteur
Prof. Dominique HIGUET, UPMC	Examinateurs

Once upon a time...

...I was in Regensburg, Germany doing my masters when I saw a PhD offer in Roscoff, France. Roscoff in Portuguese means something worth throwing away, something that does not have quality. Apparently it has something to do with a brand of watches that was made around Roscoff and that were not very good. But, who really knows if that is true... Anyway, I applied, and after a three days stay with interviews and presentations, I got it! But the story does not end here... The three days stay turned out in three years. During this time I met many people, I lived many different situations, both personally and professionally and I can say now that I am for sure a different person than when I arrived. Roscoff is a peculiar village with many charms, but where we can get very lonely. Most of the times the weather does not help (specially for a sun addict like me) and people become even more important in your day to day life. To those who were there for the better and the worse I would like to dedicate the following lines.

Tout d'abord I would like to thank Stéphane Hourdez for having proposed this interesting PhD subject to the ESTeam program. It was one of the fewest that captured my attention and the one that I applied as first option. But, most importantly Steph was with no doubt a huge support on my first months in Roscoff. His help in the lab, his little drawings with globins and phylogenetic trees (I kept some!), the ease of proposing solutions to upcoming problems of a new immigrant, are just some of the things that I will for sure remember. Thanks for the bike, the nice dinners/parties *chez toi* and other little things. Frequent sampling along the three years kept him sometimes away from the lab. Although Steph was most of the times reachable, I found that these periods were an opportunity to develop my independence inside the team and within my research project. He was always open to discuss new ideas, even if sometimes they were not feasible... let just say that the gills of hydrothermal scale worms are not easy to measure, but I still wanted to do it! Even in a hard time of his career (the loss of years of sampling due to the thaw of a -80ºC fridge!!!) he kept his strength and reassured me that we would find a solution and that my project was not in danger. His help in these last few months was precious! Writing a thesis is a marathon exercise that needs a lot of inner strength but is not possible without assistance, and Steph was there! *Merci infiniment!!!*

Because Steph was sampling at the time of the 3 days interview process, François Lallier was the one that presented me the project for the first time. I must say that it was quite weird to discuss about a future PhD subject with the cemetery just on the other side of the window. Fortunately, François kept the conversation at a very interesting level and I forgot the dead in no time. He was very supportive in those three days, showing me the entire lab, explaining the research subjects at the time and making me feel welcome, in case everything would go well, as it did. During the three years that followed he was always available if needed, and the only thing that I regret was not having discussed more frequently with him. As a teacher and his broad knowledge, his point of view is, with no doubt, a valuable one. *Merci beaucoup!*

Continuing at the professional level I would like to acknowledge Didier Jollivet. When I first saw Didier I recall that I had already seen his face on TV. But until today I am not sure if it was really he. Well, famous or not Didier gave an important input in my PhD work, especially in the final part. Converging my will to find more about the evolutionary history of the genes, with which I worked, and his knowledge on evolution produced extensive work using phylogenetic methods that I was not previously aware. His presence in my thesis committee was also very significant, helping to define clear working hypothesis. *Je te remercie!*

Jean Marie was one of the persons with whom I shared my first office. Since I did not speak French very well we did not have many conversations. Just casuals "Vamos comer" at lunchtime that he would say with a Spanish accent. When I started to speak the native language (no, not Breton!), it was a completely different story. Meanwhile, I changed office, I went to be with the other PhD students from the lab. Jean came regularly to this office to release the "saying of the day" and to check if we were working... and that reminded me so much of my father that I baptize him as *le papa du bureau* 116. As my work progressed I had more and more interactions with Jean about protein activities, structure, modeling and also discussions about generalities of life. He was there to help, every time. *Un grand merci!*

The lab work is not always straightforward, and many times things do not go as we planed or wished. For helping me to see clearly in some lab methodology I have to thank Isabelle Boutet and Arnaud Tanguy. Their explanations were very clear and their advices very useful. *Merci! Merci aussi* a Ann Andersen for the nice discussions that we had and the courses that she gave for the ESTeam program and also to Dominique Le Guen (Dom) for the nice ambiance in the lab and guaranteeing that everything would in place for our work.

Although she was not in my team, Mirjam Czjzek also made part of my thesis committee, always sharing her knowledge and ideas to improve my project. She is a wonderful teacher with an enormous capacity to motivate the ones around her. Du bist super! Other member of my thesis committee was Chuck Fisher, or should I say Charles Fisher?! Either way Chuck was amazing just in accepting to be a part of this project, as we do not work side by side. He came all the way from Penn State (USA) (twice!), to hear me talk for about an hour and patiently give his scientific input. With all the work that he had back home, he took some of his time to personally discuss with me how could I get the best of my PhD. I am much obliged! I really appreciated and it was a pleasure to accomplish the first task for the second year of this PhD: make cookies for the thesis committee.

To end the professional part I must mention some of the people that made the ESTeam project possible. Bernard Kloareg the director of the Station Biologique de Roscoff (SBR) with whom I had interesting conversations and I learned that Brittany is a special region and its language is to be maintained. Michèle Barbier, no longer among the SBR staff, but who was a very friendly contact between the students and the European Commission. All the SBR staff that help to rise and manage this project and all the

researchers that took some of their time to give us, ESTeam students, the annual courses. To all of them a great thank you! A particular *Merci* to Céline Manceau, you were great in handling all the paper works and being with you was and still is great fun!

Friends are precious, and to the old and the new ones I would like to thank them from all my heart. To start I cannot escape my ESTeam colleagues. Jan, Stefan, Simon, António, Dianne, Sarah, Christian and Aga. You made the first times of this "journey" a real pleasure. What would I do without the cars of Jan, Stefan, Simon, Christian and António. You were great just by driving me to the supermarkets or lending me you *voiture*. If I have a fridge, an apartment (and curtains on the windows), a car and a stove is because of you! Dianne and Sarah I love to hear you laughing (each one is a particular case...), you were great friends in times of need, thanks for being there. Aga you were very cool to hang out with and thanks for your nice parties, always with the side effect of vodka...

Ruth, Ula and Florence are friends that are no longer at the SBR. I thank you all for the great moments that we spend together. Ruth *gracias* for helping me to practice by Spanish and to know more about Peru. Ula you were simply a crazy girl that made me laugh even when I least expected. Thank you! Florence, what would I do without your translations between me and the lessor of my apartment? Your had infinite patience to deal with boring situations were both sides do not understand each other. *Un grand merci!*

My *bureau* mates, Constance, Matthieu and recently Sophie, you were fantastic. At 17h pm of almost every day we would go crazy in the office and say and do strange things... Mat took note of almost everything. The file is in high level of security in computer (Ouf!). Mahdi I cannot forget you and the effort that you have done, along with Mat, to render these last few months less painful. Merci, merci!!! Jéremy and Angelique... what can I say?! Good times when you where there. All the wonderful times at the Rockroum beach... and my singing pig still misses you Jérem! Nathalie, Virginie and Daphne I appreciated all the moments that we spent together. It was great fun and we had wonderful discussions. Sophie Sanchez and Yoan, I thank you for the lovely times that I spent at your place and the contribution for the *bureau* 17h pm craziness. The Portuguese connection, Miguel, Susana and again António. It was wonderful to have you here, to talk my mother language and to profit from you company. Obrigada! To all my other Portuguese friends, namely Nocas and Joana. My world without you it would not be the same. Obrigada, muito obrigada (e beijinhos)!

One of the last but definitely not the least is Anis. He was (and still is) a very important part of my life. With him I learned many different things, ways of seeing the world, other culture, the French, and mostly to accept differences. I will keep you in my heart forever... Shokran!

And a last word for my parents, that although far away in my dear Portugal, they have always supported me and have been there for me. Amo-vos!

Abstract

Hydrothermal vents are deep-sea ecosystems that are characterized by widely changing chemical and physical characteristics that are the result of the chaotic mixing of the hydrothermal fluid, rich in toxic compounds, such as sulfide and heavy metals, with low pH, high temperature and no oxygen; with the deep-sea water that contains no heavy metals, has a pH of ~7.8, low temperature and normal oxygen concentrations. Among others constraints, hypoxia and anoxia pose a serious challenge for the fauna colonizing hydrothermal ecosystems.

Scale-worms (Polynoidae) living at hydrothermal vents are abundant, diverse and widely distributed in the range of chemical and thermal conditions. To obtain the benefits of living there (abundant local primary production), they need to cope with the harsh conditions that characterize these ecosystems. Respiratory pigments play a central function in the adaptation to hypoxia (uptake, transport, and storage of the oxygen). Hydrothermal vent polynoids possess hemoglobins, a feature that clearly sets them aside from shallow water relatives. These hemoglobins are unique among annelids and their evolutionary history was the interest of this work.

There are two main types of hemoglobins in hydrothermal vent polynoids: single- and tetra-domain globins. Although they are extracellular, both globin types are more closely related to intracellular globins that to the typical annelid extracellular ones. This indicates a distinct origin for these hemoglobins. We first studied the evolutionary history of the multi-domain gene hemoglobin and found out that it originates from the tandem duplication of a myoglobin/like ancestral.

Both globin types possess residues in the heme pocket that have been shown to be responsible for the high oxygen affinity in the nematode *Ascaris* hemoglobin. These amino acids are also found in the globin from *Harmothoe*, a non-vent-endemic species. This indicates that the adaptive value of these hemoglobins most likely resides in their expression at high levels in the body and did not require specific adaptations in the heme pocket in the lineage to gave rise to all the vent species.

In addition, some amino acid sites were shown to be under positive selection in some lineages. Some of these amino acids are located in the heme pocket where they will likely affect the functional properties, and potentially provide protection against sulfide. The other amino acids under positive selection are located in areas where they will probably affect interactions between subunits for the single-domain globins, and between domains for the tetra-domain globins.

TABLE OF CONTENTS

1. Introduction

All the symbolism of red blood ultimately rests in the simple chemical bonding between oxygen and haemoglobin in our red blood cells.

Nick Lane *in* Oxygen – The molecule that made the world

Usually when someone thinks about blood or even the molecule that transports the oxygen in ones body, that someone would never imagine that it is one of the most widespread proteins in life. Oxygen Binding Proteins (OBPs) are present in bacteria, plants and animals (Suzuki and Imai 1998). Although the vertebrate OBPs were one of the first to be studied, due to the strong medical interest, they are not the respiratory pigments that show the highest structural and functional diversity (Weber and Vinogradov 2001). OBPs represent a valuable adaptation to environments that are characterized by low oxygen concentrations, such as hydrothermal vents (Hourdez and Lallier 2007). These peculiar ecosystems represent an oasis in the middle of the deep-sea and at the same time a colonization challenge for most (Tunnicliffe 1991).

In this thesis the small (big!) world of globins will be presented and discussed, from hydrothermal vent invertebrates in general and in particular from a family of hydrothermal scale worms, focusing on its phylogenetic history. To this end, this thesis opens with an introduction (chapter 1) where the hydrothermal characteristics are described, followed by an overview of how the animals adapt to this harsh environment and what the role of OBPs is, specifically globins and hemoglobins, in the mechanisms of adaptation. The introduction ends with the interest of choosing the Polynoidae as the biological model.

The two following chapters (2 and 3) report on the study of the evolution of the unique globins that these polychaetes possess, in an evolutionary perspective, but also regarding the relationship between structure and function. The hypotheses created along this study are discussed in a final chapter (4) were future research directions are also mentioned. In the end, I hope to send the message that the symbolism of red blood does not rest only in the simple chemical bonding between oxygen and an OBP, but a world of other factors is involved.

1.1 Hydrothermal vents: an example of a deep-sea oasis

There are around 70,000 km of mid-ocean ridges in the world's oceans, and, since the discovery of the first hydrothermal vent field in the Galápagos Rift in 1977, the number of identified vents has not ceased to increase, with around 80 hydrothermal vents sites to date (Fig. 1.1) (Corliss et al. 1979, Desbruyères et al. (eds) 2006, Fowler and Tunnicliffe 1997). The comparison between the usual deep-sea and chemoautotrophic ecosystems is inevitable because the former presents itself as an oligotrophic environment, rich in fauna diversity but poor in biomass (Hessler and Jumars 1974, Sokolova 1972), and hydrothermal vents, cold-seeps and whale falls (the last one to a smaller extent) are ecosystems that have the opposite characteristics (Corliss et al. 1979, Sibuet and Olu 1998, Smith et al. 1989). The latter three ecosystems burst with life in a much smaller surface than the deep-sea. The comparison can be done for several disciplines, such as geophysics, geochemistry, water composition, temperature and biota, and in all we can have very interesting results, as both types of ecosystems are so different. The cold-seep ecosystems will not be presented in the same extent of detail as the hydrothermal vents, and the whale falls will not be mentioned further.

1.1.1 Geological nature of the vents

Hydrothermal vents are expressions of intensive heat radiation from the earth's crust, being localized at the frontier of divergent tectonic plates and active convergent margins (that generate island arcs), in the ocean floor at depths approximately between 800 m and 3700 m (Fornari and Embley 1995, Fowler and Tunnicliffe 1997, Tunnicliffe 1991) (Fig. 1.1). The association of hydrothermal vents and the mentioned geological events happens due to the magma ascension, from the mantle, to form the new crust, and the older one that surrounds the spot of divergence gets more permeable, leading to the infiltration of deep-sea water (Lister 1972, Vine and Matthews 1963). This water will later rise up to the surface of the ocean crust in the form of a hydrothermal vent fluid (in a process that will be detailed in the following section). This process is chaotic, making hydrothermal vents a short-life and ever-changing ecosystem (Hessler and Kaharl 1995).

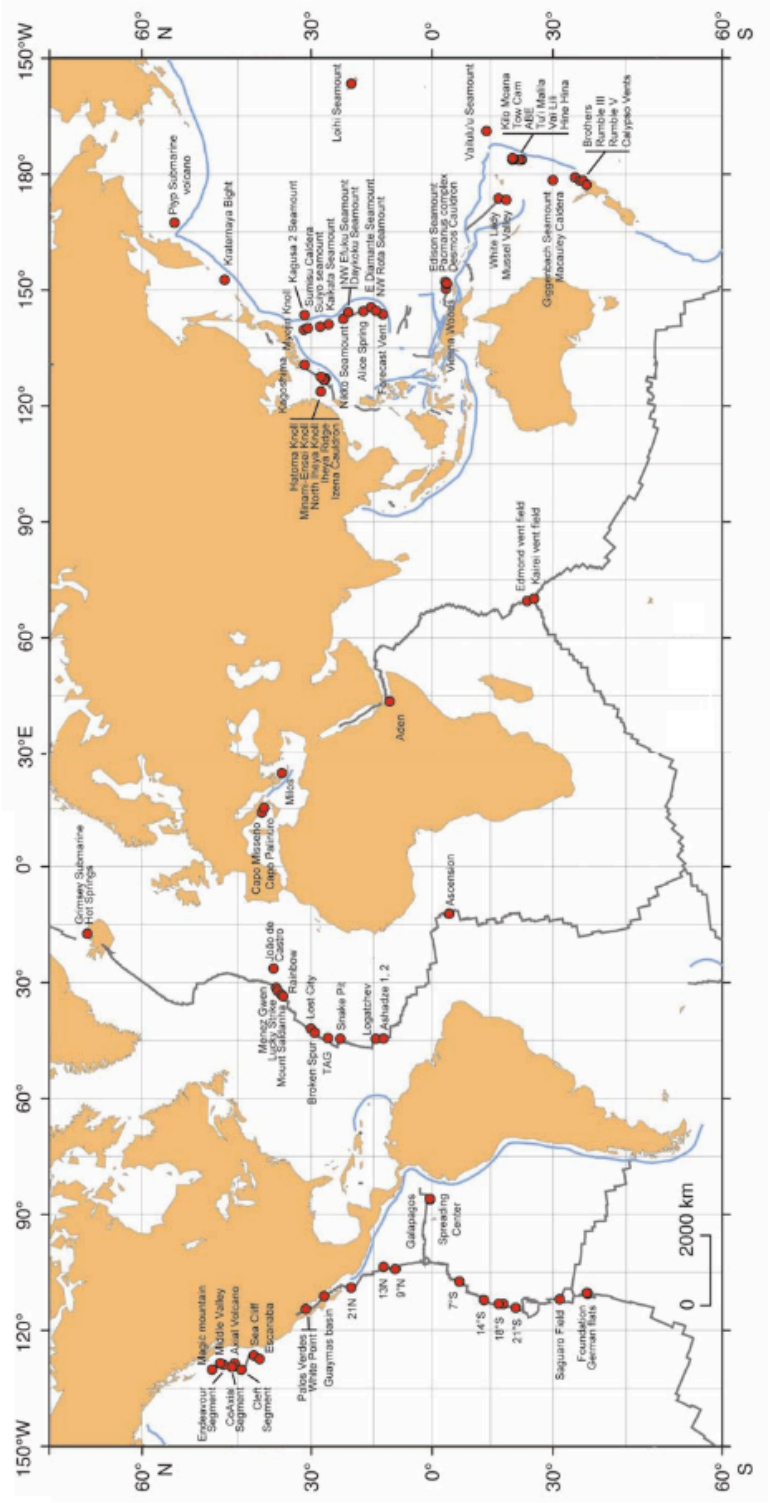


Figure 1.1. Localization of the main hydrothermal vent sites in the ocean floor. From Desbruyères et al. (eds) 2006.

The exact causes that determine vent sites are not completely understood, but they can be associated with all ridges spreading speeds and according to this fact the age of the vent fields is also variable (Fornari and Embley 1995). There are estimates of at least 125,000 years for the TAG field (Lalou et al. 1993) in the Mid-Atlantic Ridge (MAR) and only around 2 000 years (Goldstein et al. 1994) in the East Pacific Rise (EPR). In Fig. 1.2 we can observe several examples of hydrothermal vent sites in three of the most visited areas; MAR, EPR and West Pacific (Basin Lau).

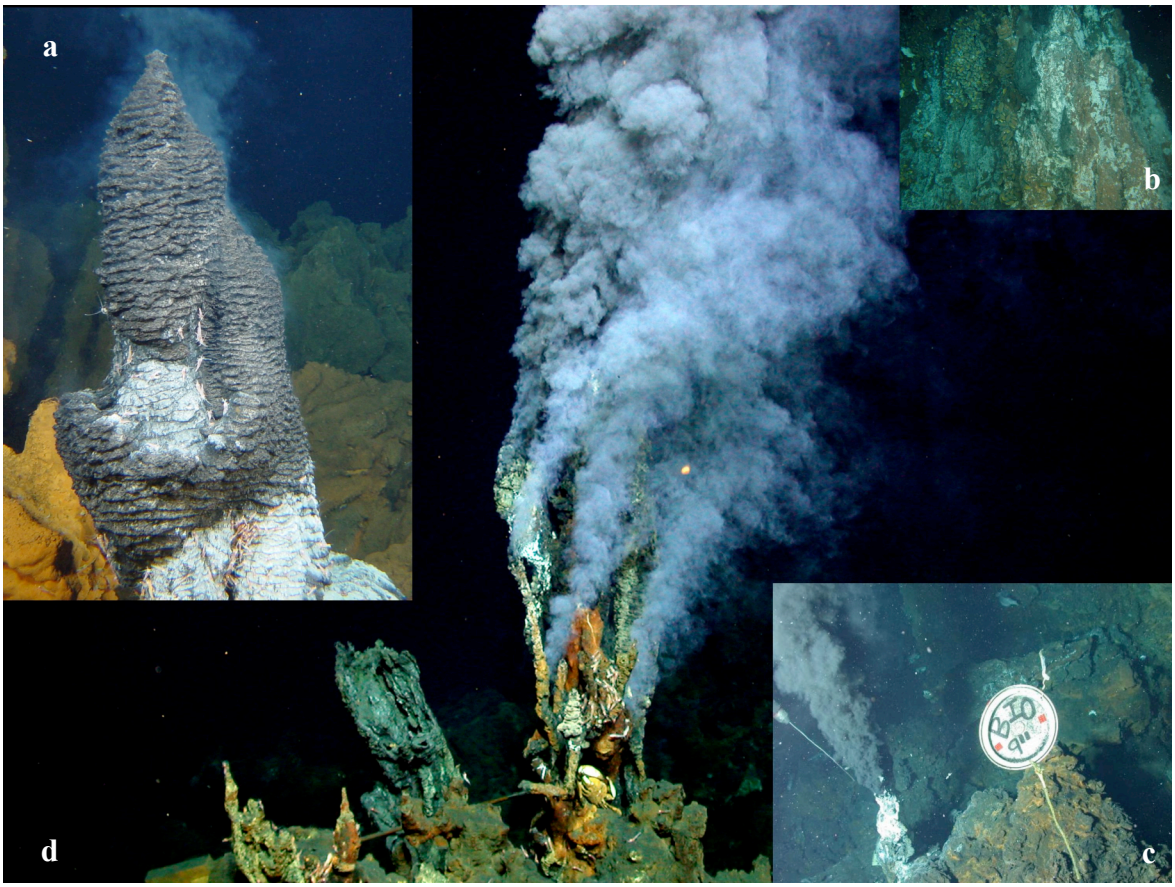


Figure 1.2. Examples of hydrothermal vents sites in the Mid Atlantic Ridge (MAR): (a) vent chimney, Rainbow; (b) bottom surface, Lucky Strike; East Pacific Rise (EPR): (c) Bio9": 9˚N 104˚W; and Pacific West: (d) Mariner site, Lau Basin 2009.

The geographical distribution of hydrothermal expressions can have different levels of organization (Fig. 1.3):

'vent' – single localized emission of hydrothermal fluid;

'vent site' –several emissions few meters apart, linked by subterranean ducts;

'vent field' – cluster of vent sites located a few hundreds of meters apart;

'venting areas' or 'sectors' – oceanic ridge segments.

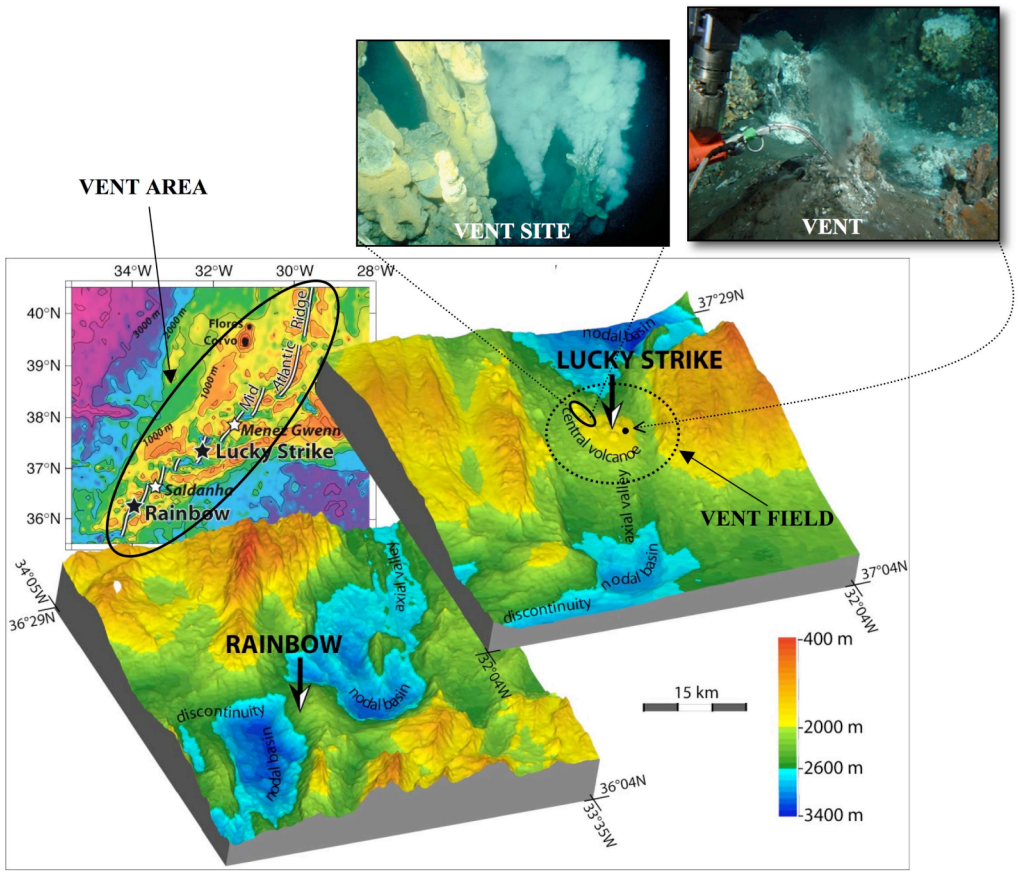


Figure 1.3. Representation of the topography and localization of two hydrothermal vent sites in the MAR: Rainbow and Lucky Strike. The different geographical distribution that characterizes the hydrothermal vent in general is represented. Adapted from http://www.ipgp.fr/rech/lgm/MOMAR_FR/Momarfr_Info.html.

Cold seeps are a manifestation of fluid or mud seepage from the sediment and are also a result of a geological process, but this time related with accreted sediment wedges (e.g. tectonically induced high-fluid pressures, petroleum or natural gas escape), or salt tectonics occurring in passive and active continental margins (Sibuet and Olu 1998, Tunnicliffe 1991).

1.1.2 Abiotic characteristics at hydrothermal vents

Although they are found at such deep locations in the ocean floor, their influence in the ocean chemical composition, deep and mid-depth circulation, thermohaline circulation, oceanic crust and atmospheric gases inputs, are far from being negligible (Lowell et al. 1995, Lupton et al. 1985, Stommel 1982, Von Damm 1995).

One of the most striking factors that characterize these ecosystems is the rapid abiotic variations that the whole ecosystem experiences (Tunnicliffe 1991). These chaotic changes, mainly in temperature and water chemistry, are due to the mixing of ascending water, from within the crust, rich in various dissolved inorganic compounds, acidic, and with temperatures around 300-400ºC, with the surrounding sea water that exhibits a different inorganic composition and a much lower temperature (around 2ºC) (Edmond et al. 1979, Corliss et al. 1979, Fowler and Tunnicliffe 1997, Tunnicliffe 1991).

The cycle can be divided into three phases, the first being the penetration of the cold and dense deep-sea water, through the permeable crust floor adjacent to sites of tectonic plates expansion (Edmond et al. 1979, Fornari and Embley 1995, Lister 1972, Tunnicliffe 1991). With the help of high temperature and pressure several inorganic compounds are dissolved in the water during the transit time through the crust (Fig. 1.4) (phase two) (see Alt 1995 for a review). The hydrothermal fluid is then rich in metals such as iron, zinc, copper, and manganese, as well as sulfur compounds, that afterwards precipitate in contact with the cold water forming smokers or chimneys (phase three) (Childress and Fisher 1992, Edmond et al. 1979, Fowler and Tunnicliffe 1997, Von Damm 1995) (Fig. 1.4).

The mixture of deep-sea water and precipitated minerals from the hydrothermal fluid can originate less dense elements than the deep-sea water, which form black or white plumes, according to the rock's crust initial composition (Lowell et al. 1995). If the hydrothermal fluid is rich in iron sulfide minerals, zinc sulfide (ZnS) and copper-iron sulfides ($CuFeS_2$) these compounds turn black when they precipitate, and pyrite (FeS_2), barite ($BaSO_4$) and silica (SiO_2) turn white (Fowler and Tunnicliffe 1997, Spiess et al. 1980, Von Damm 2001). These plumes can extend for a long distance in to the water column changing the chemistry of the mid-depth ocean (Corliss et al. 1979, Lowell et al. 1995, Von Damm 1995) and helping researchers in localizing new vent spots (Fowler and Tunnicliffe 1997, Tunnicliffe 1991). The process described above is known as the 'hydrothermal circulation' and it can be represented as in Fig.1.5, showing the connection

between the surrounding seawater, the crust and the products of this mixing (vent structure and plume).

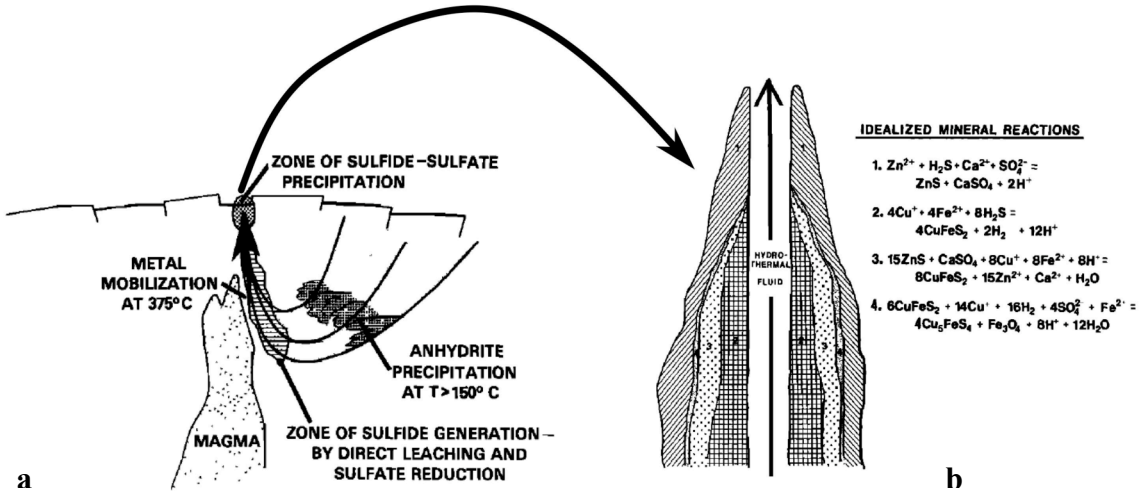


Figure. 1.4. Schematic representations of (a) a hydrothermal site and the processes affecting the sulfur system during deep circulation and of (b) chimney growth, reactions 1 and 2 occur by direct precipitation of sulfide and reaction 3 and 4 occur by replacement of sulfides with the incorporation of reduced sulfate. Adapted from Woodruff and Shanks (1988).

Von Damm (1990) suggests that the chemical composition of the vent fluids depends on several factors, such as the subsurface and seafloor temperature, pressure, rock composition among others. Also the characteristics of the interface region are very unstable and depend on the departure values of the hydrothermal fluids before being diluted in the surrounding deep-ocean water (Childress and Fisher 1992) (Fig. 1.6). Besides the temperature differences, the surrounding water also does not contain heavy metals, sulfide compounds, has alkaline pH values, and can have oxygen concentrations of about 0.2 mM (Hourdez and Lallier 2007). Table 1.1 represents a summarized comparison of deep-sea water and hydrothermal fluids compositions at different hydrothermal sites. The difference between both fluids is usually clear, with the metal compounds, sulfide and CO_2 found at high values in the hydrothermal vent fluid. The composition of elemental ratios can be rather constant between individual vents, or even between different vent areas, but they always show some degree of individual differences and they are mostly due to temperature differences of the phase separation of chemical compounds (Von Damm 1990, Von Damm 1995).

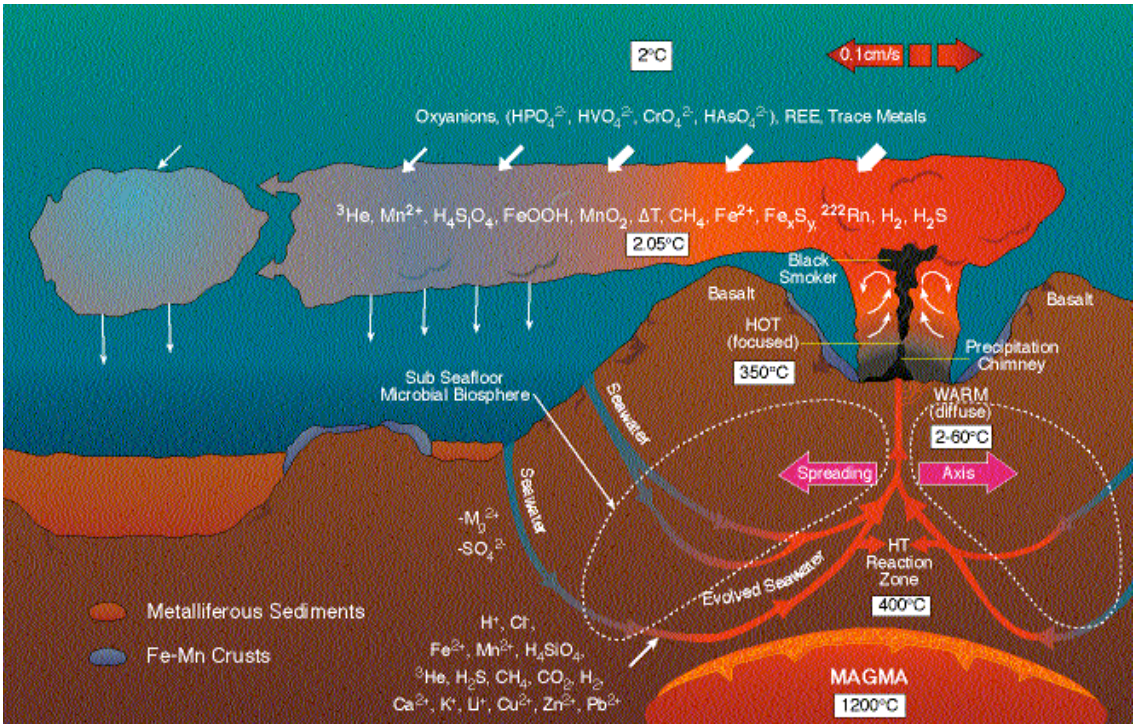


Figure 1.5. Representation of the 'hydrothermal circulation' in a hydrothermal vent system on a spreading axis. Represented structures are not at the same scale. All three phases, water infiltration, water transport through the crust and expulsion as hydrothermal fluid, are depicted. Also represented are the several chemical species that contribute for the water chemical transformation. Notice the deep-sea water and hydrothermal fluid temperatures. The plume that is generated can reach several meters above the bottom and eventually precipitates further away from the hydrothermal source. (Image from http://www.indiana.edu/~g105lab/images/gaia_chapter_13/vent_communities.htm, originally published by Massoth et al. 1988)

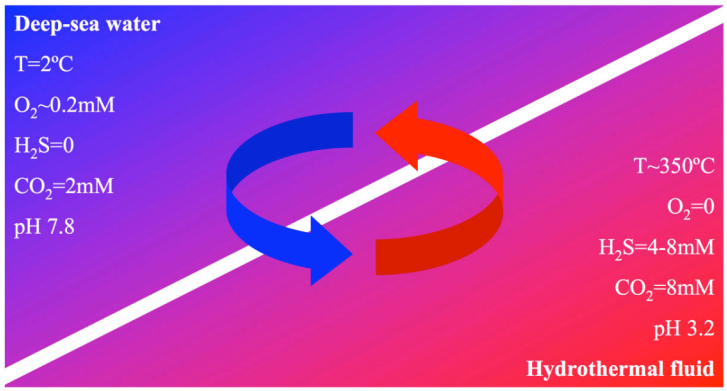


Figure 1.6. Representation of the interface region between deep-sea water (blue) and hydrothermal fluid (red). The typical characteristics for each fluid are shown. The chaotic element of this region is represented by the arrows. From Hourdez and Lallier 2007.

The hydrothermal vent ecosystems are thus habitats of a great potential for a faunal community to thrive, maybe one of the most ancient ones (Reysenbach and Shock 2002). This is possible by making available chemical compounds that when oxidized release a significant amount of energy, usable by autotrophs. But these energy sources are also toxic to metazoa, and adding this fact to the high temperatures and pressures and low pH and oxygen concentrations, makes the vents a *sui generis* environment with a lot to offer if one possesses the right adaptations to overcome the physiological obstacles (Childress and Fisher 1992).

Table 1.1. Deep-sea water and hydrothermal fluid chemical and physical composition of representative vent sites from the most sampled oceanic ridges (after Charlou et al. 2000, Le Bris et al. 2003, Hourdez and Lallier 2007 and Von Damm 1995). Oxygen values were not available in the present studies, but previous studies have shown that hydrothermal fluid is totally anoxic (Corliss et al. 1979, Edmond et al. 1979).

Physical-chemical parameters	13ºN (EPR)	Juan de Fuca Ridge (North EPR)	Lau Basin (West Pacific)	Lucky Strike (MAR)	Deep-sea water
Temperature (ºC)	317 – 380	328	334	172-324	2
pH	3.1 – 3.3 (25ºC, 1 bar)	3.5 (no information)	2 (no information)	3.65 (no information)	7.8 (25ºC, 1 bar)
O_2 (mM)	-	-	-	-	0.2
CO_2 (mM)	11.8 – 18.4	50	-	24.2	2.3
CH4	0.051	-	-	0.52	$4x10^{-7}$
H_2S (mM)	2.9 – 8.2	7.1	-	2.7	0
SO4 (mM)	0	0	0	-	27.9
SiO2 (mM)	22	22.7 – 23.3	14	14.4	0.16
Mn (mM)	1 – 2.93	1150	5.8 – 7.1	0.25	<0.001
Fe (mM)	1.45 – 10.8	1065	1.2 – 2.9	0.3	<0.001
Cu (mM)	-	9.9	15 – 35	0.01	0.007
Zn (mM)	-	111	1200 – 3100	0.03	0.01

1.1.3 Autotrophs and symbioses with macrofauna

Corliss et al. (1979) suggest that the energy source of hydrothermal vent ecosystems was based in the chemolitotrophic sulfur-oxidizing bacteria, and state that the primary productivity in the surface water as the origin of nutrients is, at most, marginal. The high amounts of sulfide, the high concentrations of sulfur-oxidizing bacteria in the water and the fact that the sampled hydrothermal vent (Galápagos Rift) had high depth and high abundance of fauna, lead them to these conclusions. In fact the hydrothermal vents (as well as cold seeps) represent an example of ecosystems that can virtually rely on metabolic mechanisms other than photosynthesis to survive and prosper (Reysenbach and Shock 2002).

Chemolithoautotrophs are organisms that fix inorganic carbon, using chemical energy issued from the oxidation of reduced compounds, such as reduced sulfur compounds (H_2S, HS^-), molecular hydrogen (H_2), methane (CH_4), reduced metals and ammonium (NH_4^+), using oxygen (O_2), nitrate (NO_3^-), oxidized sulfur compounds (SO_4^{2-}) or carbon dioxide (CO_2) as electron acceptors (Reysenbach and Shock 2002, Nakagawa and Takai 2008). From these reactions useful energy is available in the form of ATP and NADH molecules that will supply the Calvin-Benson cycle, or reverse TCA cycle, where the CO_2 will be used for biomass production (Kuenen 1999, Nakagawa and Takai 2008) (Fig. 1.7). The chemolithoautotrophs that can profit from the vents fluids (rich in reduced compounds) by being meso- or hyper-thermophilic, are organisms from the *Bacteria* and *Archaea* domains (Zierenberg et al. 2000). Along with other chemoheterotrophs, they form the source of nutrition for macrofauna, supporting the hydrothermal vents communities as primary producers. Interestingly, some of the strains can survive temperatures higher than 90ºC (hyper-thermophilic strains), as they colonize even the chimney walls. Their life strategies, along with other thermophiles, remain a puzzle for the community of biochemists and microbiologists, since high temperature is usually a hazardous factor for protoplasm maintenance (Reysenbach and Shock 2002, Zierenberg et al. 2000).

Their abundance and their life style (free or symbiotic with invertebrate hosts) will influence the metazoans distribution in the vent ecosystems. The bacterial communities that live freely, either in suspension in the mixing zone or over the sea-floor surface in the form of bacterial mats, are crucial for filter feeders and grazers, respectively. These bacteria can even colonize biological surfaces, such as the tube wall from tubeworms

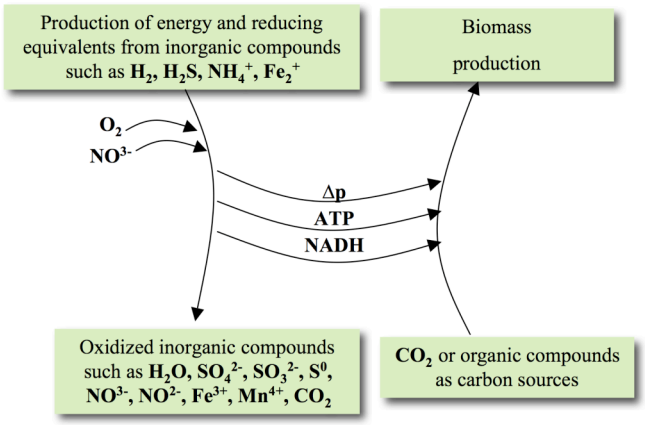


Figure 1.7. Representation of the major steps of chemolithoautotrophy, with the input and output compounds. The energy conversion is done from the oxidation of inorganic compounds and of autotrophic carbon dioxide fixation. (Δp: proton potential). Adapted from Kuenen 1999.

(vestimentiferans, e.g. *Riftia*) and shells from bivalves (mussels, vesicomyid clams) and gastropods (Childress and Fisher 1992). These latter graze on filamentous bacteria that grow on the surface of shells of other species or their own, or even on the tube wall of tubeworms (de Angelis et al. 1991, Childress and Fisher 1992). In fact many available surfaces, that still are bathed by diluted vent fluids and their reduced compounds, are colonized by chemolitoautotrophs, so other animal tissues such as the gills of some filter feeders are also used as a colonization surface (de Angelis et al. 1991, Fisher 1996).

The colonization of animal surfaces is taken as the first step for the origin of the endosymbioses that are known in hydrothermal and cold-seep environments. For example in a limpet species from the Juan de Fuca Ridge, the bacterial colonization of the surface of the gills is usually followed by the endocytosis of bacteria that are then digested in the cell (de Burgh and Singla 1984). Some alvinellid polychaetes (e.g. *Alvinella* spp) have the most highly developed ectosymbioses known for these ecosystems and may represent an initial stage regarding the level of symbioses found in gastropods (e.g. *Alviniconcha* spp., Provannidae), mussels (*Bathymodiolus* spp., Mytilidae), other bivalves like vesicomyid clams (e.g. *Calyptogena* spp., Vesycomyidae) or vestimentiferans annelids (*Riftia*, Siboglinidae) (Childress and Fisher 1992, Gaill et al. 1988). Although *Bathymodiolus* spp. and *Alviniconcha* spp. are symbiotic (bacteria are kept in special cells in the gill epithelium called bacteriocytes), they continue to have a functional digestive tract (Le Pennec 1988,

Le Pennec et al. 1990, Stein et al. 1988). In contrast, in *Calyptogena* spp., the animals have a mouth but a reduced digestive system, indicating that they largely depend on their symbionts for nutrition (Le Pennec and Fiala-Medioni 1988, Le Pennec et al. 1990). *Riftia* in turn does not have any digestive structure as adults and uses the branchial plume to capture gases and transports them into the trophosome, an organ in the trunk part of the body that contains the symbionts (Arp and Childress 1983, Cavanaugh et al. 1981, Jones 1988).

1.1.4. Community spatial variability and vent fauna biogeography

The relationship level between the symbionts and their hosts, the availability of reduced chemical compounds and O_2, the capacity of the hosts to withstand the toxicity and the high temperatures of the environment are some of the main factors for the spatial distribution of the macrofauna in the hydrothermal vents (Grassle 1985, Sarrazin et al. 1999, Tunnicliffe 1991). Usually the distribution around chimneys or diffusion zones can be arranged more or less concentrically and divided into three zones (Hessler and Kaharl 1995, Tunnicliffe 1991). On the EPR and in Juan de Fuca ridge the chimney walls (first zone, Fig. 1.8) are colonized by *Alvinella pompejana* and *Paralvinella sulfincola*, respectively (Chevaldonné et al. 1992, 2000, Rinke and Lee 2009). Other polychaetes such as the polynoid *Branchinotogluma segonzaci*, in the Basin Lau and the hesionid *Hesiolyra bergi*, on the EPR, can also be observed in this warmer zone, the latter competing with *Alvinella* for food (Desbruyères et al. 1998, (eds) 2006). In the following, less warm, zone (Fig. 1.8) most of the biomass corresponds to the macrofauna that rely exclusively or almost exclusively on their symbionts for nutrition, such as vestimentiferans annelids, vesycommyid clams (that avoid *Riftia* clusters), gastropods and mussels (Grassle 1985, Hessler and Kaharl 1995, Sarrazin et al. 1997). Bresiliid shrimp occupies this zone in a very high density at the MAR hydrothermal vents (Hessler and Kaharl 1995). Other smaller animals are usually found associated with the symbiotic ones, such as limpets, polychaetes worms, crabs and fish. The cohabitation can be such that *Branchypolynoe* scale-worms (Polynoidae) live inside the mussels *Bathymodiolus* sp. as a commensal (Grassle 1985). The third zone (Fig. 1.8), on the EPR, is dominated by serpulid

polychaetes, barnacles, anemones and scallop-like bivalve mollusks, that are suspension feeders and sessile. The outer zone is typically inhabited by non-vent animals (Hessler and Kaharl 1995, Tunnicliffe 1991).

At a global scale the hydrothermal species distribution is not the same between the different vent sites, not even in the same ocean ridge. To date we distinguish six biogeographic provinces, and the EPR (north and south) being an important center of dispersal for the hydrothermal fauna (Bachraty et al 2009, Van Dover 2002). Between the two most studied provinces, the EPR and the MAR, some differences may be pointed out, such as the alvinellid polychaetes notoriously absent on the MAR, but typically present on the EPR. Instead, the former region has a great abundance of the shrimp *Rimicaris* and the mussels *Bathymodiolus* spp that are present in all known hydrothermal vents (except the Juan de Fuca Ridge, Hourdez pers. comm.), and also at cold seeps (Ramírez-Llodra et al. 2003, Tunnicliffe 1991, Tyler and Young 2003). Regarding the smaller associated heterotrophic species an example of difference between regions for scale-worms is the presence of *Branchipolynoe symmytilida* on the EPR, *B. seepensis* on the MAR and *B. pettiboneae* in the Northwest and Soutwest Pacific, all occupying the same niche, the pallial cavity of the mussels *Bathymodiolus* (Chevaldonné et al. 1998). The reasons for these differences are beyond the scope of this work but Bachraty et al. (2009), Van Dover (2002), Tunnicliffe (1988) and Tyler and Young (2003) present and discuss some possible causes for the observed patterns.

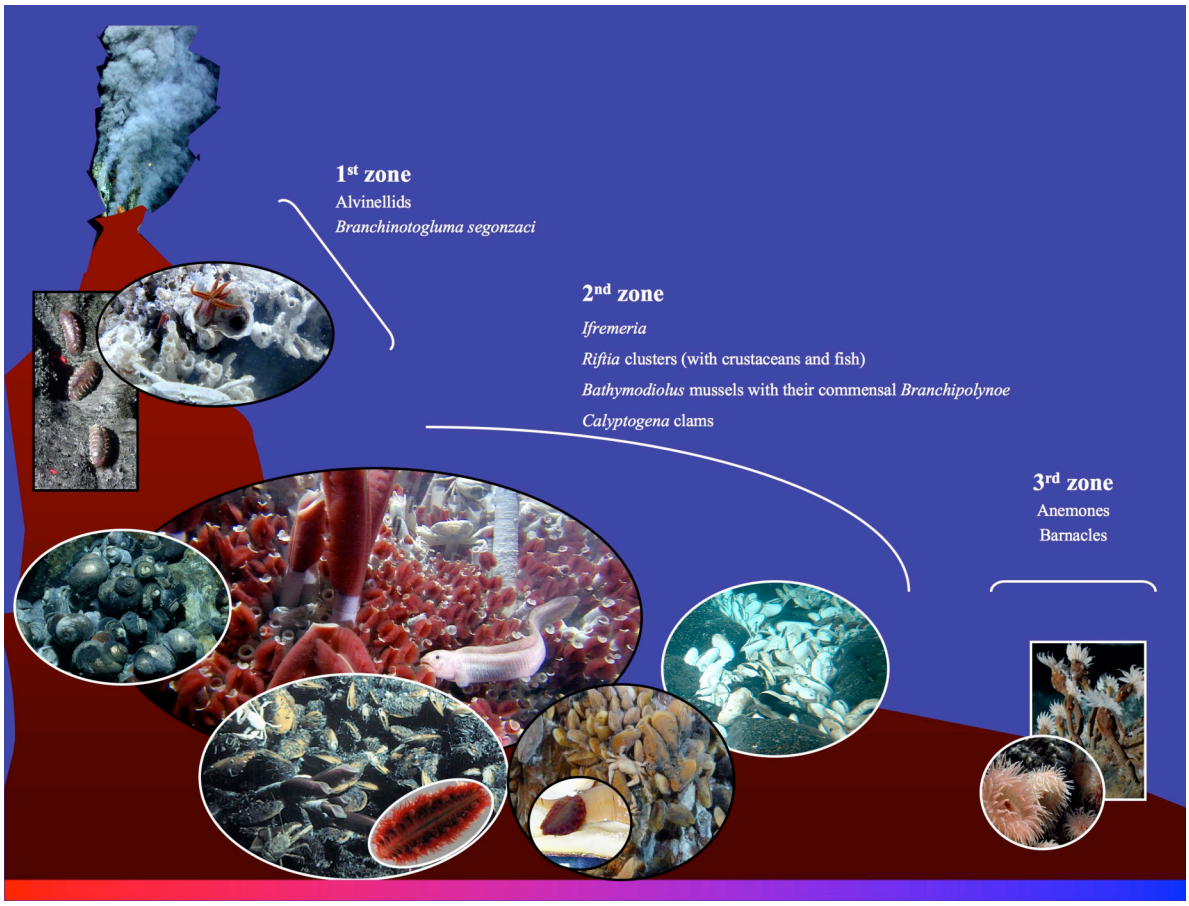


Figure 1.8. Schematic illustration representing the distribution of fauna around a vent chimney. Photograph credits counterclockwise: *Alvinella*: Stephen Low Productions, from http://www.noc.soton.ac.uk/chess/event/galapagos_event.php; *Branchinotogluma segonzaci* polychaete: ABE site, Lau Basin 2009; *Ifremeria* gastropods: ABE site, Lau Basin 2009; *Riftia* with accompanying fauna: http://fondation.total.com/environnement/recherche-sur-la-biodiversite/grands-fonds/inventaire-mondial-de-la-biodiversite-marine-800038.html&projectMedia=610#mediaBloc; *Bathymodiolus thermophilus*: http://www.whoi.edu/oceanus/viewImage.do?id=5129&aid=2420, and its commensal polychaete *Branchipolynoe symmytilida*: T. Shank, from http://www.ridge2000.org/SEAS/for_students/data/field_notes_05.html; *Bathymodiolus azoricus*: Ifremer Momareto 2006, with *B. seepensis*: S. Hourdez, from Hourdez and Weber 2005; *Calyptogena* clams: EPR 21ºN, from http://divediscover.sr.unh.edu/010902.html; Anemones: Tow Cam site, Lau Basin 2009; Barnacles: bogleech.com/barnacles.html. Bar on the bottom represents the temperature range, from the warm (red) to cold (blue).

1.2 Physiological adaptations: respiration and respiratory pigments

According to Jaenicke (1981) there are three ways through which an organism can cope with the extreme conditions of the ecosystem; (1) to exclude the disabling factor, (2) to detoxify its cells from the disabling factor and (3) to adapt to the disabling factor up to a level of co-habitation. Apparently the last mechanism seems to be the most common, with the organism becoming completely dependent on the factor that turns out to no longer be disabling. This affirmation is completely true for the vent organisms that show a high level of adaptation to factors such as pressure, temperature, pH, hypoxia and toxic compounds (Childress and Fisher 1992, Childress 1995, Hourdez and Lallier 2007). Before addressing specifically the adaptations regarding the respiratory process and the respiratory pigments involved, I will mention, in a non-exhaustive manner, some of the general effects and adaptations to cope with major disabling factors.

1.2.1 Pressure, temperature and pH

As deep-sea communities, the animals living at hydrothermal vents have to cope with the high pressures and high temperatures typical from these ecosystems. These two parameters are very important to biological processes and they usually interplay (Childress and Fisher 1992, Gekko and Hasegawa 1989, Gross and Jaenicke 1994). Pressure does not perturb biological processes when no change in volume is associated (Somero 1992). But usually volume changes do happen and protein structure and function, especially membrane-bounded enzymes, can be greatly influenced by pressure (Childress and Fisher 1992, Jaenicke 1981, Somero 1992 and references therein). A rise in temperature also influences proteins because of the weakening of the hydrogen bonds that have low bond energy and are responsible for enzyme-ligand complexes (function) and higher orders of protein structure (structure) (Jaenicke 1981, Kauzmann 1959 in Jaenicke 1981, Low and Somero 1975). Although pressure and temperature are both crucial for life at hydrothermal vents, it seems that the thermal variations are more

relevant for the need of adaptations (Low and Somero 1975).

An example of pressure adaptation is the rigid structure adopted by several oligomeric proteins (or part of the proteins) (Jaenicke 1981, Kundrot and Richards 1987), since high pressure, depending on the temperature, can sometimes dissociate them into monomers (Jaenicke and Koberstein 1971, Jaenicke 1981, Somero 1992). Temperature adaptations regarding mitochondrial membrane enzymes in vent and littoral congeneric species revealed that the species that inhabit cooler habitats (in vents or littoral ecosystems) do not show the thermal adaptations that occur in species that live in warmer areas. So, in vent endemic species like *Riftia* or *Alvinella*, that live closer to vent emissions, thermal adaptations are different from those observed in vesicomyid clams and mussels (Dahlhoff et al. 1991a).

pH is also an important factor establishing the communities' boundaries, mostly due to its effect on cell surface charges (Jaenicke 1981). Nevertheless, in microbes, it is usually only studied concomitantly with pressure and mostly temperature experiments because the majority of acidophiles are also thermoacidophiles (e.g. Jaenicke 1983 and references therein). For the respiratory metabolism pH effect is quite crucial, since with an increase of P_{CO2} the pH will fall, creating inhibitory interactions between the globin's heme group and oxygen (Weber and Vinogradov 2001).

Although factors cited above are important in the success of organisms colonizing environments such as hydrothermal vents, I would like to focus on other important factors, in particular for the respiratory process in organisms that inhabit those ecosystems.

1.2.2 Sulfide and other toxic elements

As discussed earlier, the key for the hydrothermal vents endemic communities to thrive is based on the local primary production by bacteria, free-living or symbiotic with invertebrates. This fact forces the organisms to live and forage in the mixing zone between the deep-sea water and the toxic hydrothermal fluid, in order to obtain the O_2 and also the reduced chemical compounds (Childress and Fisher 1992).

Among the reduced compounds coming out of the hydrothermal fluid, sulfide (H_2S,

HS^-) and CH_4 are the most used for the chemoautothrophy process at hydrothermal vents and cold seeps (Fisher 1996). However sulfide is known to be highly toxic for aerobic organisms, as it is a potent inhibitor of mitochondrial cytochrome *c* oxidase and also impairs oxygen transport by irreversibly binding to hemoglobins (Carrico 1978, Keilin 1929 in Childress and Fisher 1992, Grieshaber and Volkel 1998). Still, most of the vent animals studied by Hand and Somero (1983), namely *Riftia pachyptila*, *Calyptonena magnifica*, the crab *Bythograea thermydron* and *Alvinella pompejana*, showed cytochrome *c* oxidases activities similar to littoral species. This indicates that hydrothermal vent species do rely on aerobic respiration like their littoral relatives. In fact, in a recent study, Rinke and Lee (2009) show that in *Paralvinella sulfincola*, one of the species that live on the chimney walls (in presence of high sulfide concentration) uses both aerobic and anaerobic metabolism, with the former dominating in the branchial tissues.

The way organisms cope with sulfide depends on the species adaptations and even if they share mechanisms for defense against sulfide toxicity, the level of success is variable (Powell and Somero 1986). There are several mechanisms for sulfide tolerance and protection. It can start immediately from the body wall (partial impermeability and/or presence of sulfide-oxidizing bacteria), involve insensitive cytochrome *c* oxidase, reversible sulfide binding to respiratory pigments, sulfide oxidation in the mitochondria (e.g. Sulfide Quinone Reductase) coupled with the production of ATP and thiosulfate, anaerobiosis and symbioses with sulfide oxidizing bacteria (Childress and Fisher 1992, Grieshaber and Völkel 1998 and references there in, Hildebrandt and Grieshaber 2008, Theissen and Martin 2008a). One of the most important and primary way of defense of the hydrothermal vent species, with or without symbiotic relationships is the oxidation of sulfide (e.g. Vetter et al. 1987). In its trophosome (symbiont-containing tissue), *Riftia* has higher levels of sulfide oxidation than for example *Bathymodiolus* mussels (gills) (Powell and Somero 1986). Alvinellids are the ones that not only withstand the highest temperatures but also tolerate high concentrations of sulfide (20-100 $\mu mol.l^{-1}$) (Rinke and Lee 2009, Sarrazin et al.1999) although most of this sulfide is FeS (Luther et al. 2001). The mechanism to cope with this toxic compound is not clear yet, but several authors suggest the chemoautotrophy of the epibionts in *Alvinella* or a hemoglobin with a similar function to that of *Riftia* as processes of detoxification (Gaill et al. 1987, Martineu et al. 1997).

Sulfide oxidation and the reduction of oxygen by the respiratory chain produce oxygen radicals, such as O_2^-, HO_2^-, H_2O_2 and OH, known for rapidly oxidizing biological molecules (Fridovich 1978, Tapley et al. 1999). Although it seems surprising, since O_2 is essential for aerobic respiration, by these means it is also a toxic compound. To deal with the products of aerobiosis, organisms possess enzymes that catalyze the reduction of the oxygen radicals, such as superoxide dismutases and catalases (Fridovich 1978). This kind of responses to oxygen toxicity were found in *Riftia*, and *Calyptogena* by Blum and Fridovich (1984 in Childress and Fisher 1992), and in *Paralvinella grasslei* (Marie et al., 2006).

Heavy metals are also considered toxic elements for the fauna living at the vents, being accumulated in high concentrations in their tissues (Shmelev et al. 2009). An interesting adaptation, because of their proximity to the vent chimneys, is the resistance of alvinellids to heavy metals. A variety of heterotrophic metal-resistant bacteria lives in the mucus that the worm produces (Jeanthon and Prieur 1990) and Gaill et al. (1987) suggested that these bacteria probably help the annelid to detoxify its environment. How this process takes place is still unclear (Martineu et al. 1997).

1.2.3 Adaptations to hypoxia

Hypoxia is a basic challenge that hydrothermal-vent (and cold-seep) species face and need to overcome to thrive in these food-rich environments (Hourdez and Lallier 2007). In general vent organisms do not show reduced metabolic rates compared to their shallow-water relatives (Childress and Mickel 1985). This requires the development of adaptations at different levels to meet the metabolic needs. These adaptations will aim at improving the oxygen uptake and its transport, before having to rely on anaerobiosis (Childress and Fisher 1992, Hourdez and Lallier 2007). In this section I would like to review some of these levels of adaptation in hydrothermal invertebrates. Anaerobic metabolism will not be developed in this introduction, but hydrothermal vent fauna most likely rely on this process when exposed to high sulfide concentrations as high sulfide is usually associated to low oxygen and it also inhibits the respiratory chain in

the mitochondria (Childress and Fisher 1992, Grieshaber and Volkel 1998, Hand and Somero 1983).

The uptake and transport of oxygen depends on several steps from the water surrounding the organism to its cells, where the oxygen is consumed by the mitochondria. Each step relies on the difference of partial pressure of oxygen between two sides of a diffusion barrier, and is driven by convective or diffuse processes (Hourdez and Lallier 2007) (Fig. 1.9).

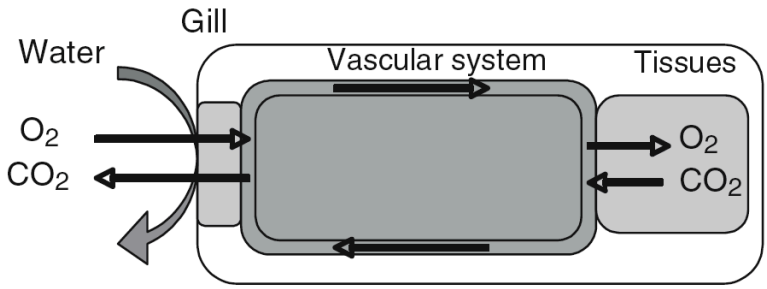


Figure. 1.9. Representation of gas flow in a metazoan with a circulatory system. Different conceptual barriers are represented (gill and tissues walls). From Hourdez and Lallier 2007

1.2.3.1 Oxygen Extraction

The gills form the first barrier to overcome in the oxygen pathway. Their ventilation, surface areas, and diffusion distances are major factors that will exhibit modifications in hydrothermal vent invertebrates compared to shallow water ones. Crustacea and mollusks possess a cavity in which the gas exchange organ is located and where they can control the water flux. They can therefore potentially affect the ventilation of their respiratory surface. In hydrothermal annelids however, the gills are not protected by any special structure, so the ventilation of the branchiae cannot be regulated by the individuals at a macroscale but the cilia that cover their gills can renew the diffusion layer at the surface (Hourdez and Lallier 2007). Enlarged branchiae (surface area) can help enhance oxygen uptake. Such enlarged gas exchange surfaces have been found in *Alvinella pompejana* and *Paralvinella grasslei* that show the largest specific gill surface areas known in polychaetes (Table 1.2, Jouin and Gaill 1990). This is especially meaningful for the scale-worms *Branchipolynoe* whose shallow-water relatives are usually devoid of gills and gas exchange occurs only through the body wall. Enlarged gill surface areas however can also be detrimental as the uptake is not selective and the

toxic compounds from the hydrothermal fluid can also more readily enter the body. In symbiotic vent and cold-seep species such as *Bathymodiolus* spp. and the provannid gastropods *Ifremeria nautilei* and *Alviniconcha hessleri*, the gills also have a high surface area, but it seems that this characteristic is mostly due to the housing of symbionts (Childress and Fisher 1992). These mollusks do not seem to be able to regulate their oxygen uptake, as shown in the cold-seep mussels (Childress and Fisher 1992).

Another possible way to compensate the low oxygen concentration is to reduce the diffusion distance (thinner branchial epithelium, Table 1.2). Good examples of this advantageous characteristic are again the alvinellids *A. pompejana* and *P. grasslei* (Jouin and Gaill 1990) and the siboglinid *Riftia pachyptila* (Andersen et al. 2002). In Table 1.2 we can observe more clearly the oxygen uptake functional characteristics of several hydrothermal and shallow water species. Studies on diffusion distances for other vent taxa other than polychaetes seem to be lacking (Hourdez and Lallier 2007).

Table 1.2. Gill characteristics for some hydrothermal vent species and shallow-water relatives. HV: hydrothermal vents, Lit.: littoral. After Hourdez and Jouin-Toulmond 1998, Hourdez and Lallier 2007.

Species	Environment	Type of gill	Specific gill surface area (cm^2/g)	Diffusion distance (μm)
Alvinella pompejana	HV	Vascular (lamellae)	12	1-3
Paralvinella grasslei	HV	Vascular (filaments)	47	4
Riftia pachyptila	HV	Vascular (lamellae)	12	1-3
Branchipolynoe symmytilida	HV	Coelomic branched	14.2	10
Branchipolynoe seepensis	HV	Coelomic branched	10.3	9
Branchipolynoe pettiboneae	HV	Coelomic branched	7.7	10
Glycera convoluta	Lit.	Coelomic, unbranched	1.5-2	?
Terebellides stroemi	Lit.	Vascular (lamellae)	6	5-8
Arenicola marina	Lit.	Vascular, branched	4	8-14

1.2.3.2 Oxygen transport

The type of oxygen transport system from the branchiae to the tissues differs among annelids. For example in alvinellids and *Riftia pachyptila* the gills are perfused with vessels from the vascular system, while in *Branchipolynoe* (Polynoidae) the coelomic fluid bathes all the gill (Andersen et al. 2002, Hourdez and Jouin-Toulmond 1998, Jouin and Gaill 1990). Regardless of the transport system, the conductance of oxygen will depend on the product of the capacitance of the blood (quantity of carried O_2 by unit of blood) and the blood flow (volume per minute) (Hourdez and Lallier 2007). This latter factor is modulated by the heart rate and stroke volume and can be affected by the body movements. The blood capacitance will depend on the presence, concentration, and functional properties of oxygen binding proteins (OBPs, Hourdez and Lallier 2007).

The influence of OBPs relies on several functional properties such as oxygen affinity, cooperativity and the strength of a Bohr effect (Hourdez and Lallier 2007). The cooperativity involves interactions between the heme groups, and the Bohr effect between heme groups and proton binding sites. The latter corresponds to a decrease of affinity when the pH decreases. Two main types of respiratory pigments can occur: storage OBPs and transport OBPs. Usually storage OBPs are characterized by a high oxygen affinity, lower cooperativity and no influence of pH. Transport OBPs have the inverse characteristics: low oxygen affinity, several degrees of cooperativity and exhibit a Bohr effect (sensitivity to pH changes). In Figure 1.10 are illustrated two hypothetical oxygen saturation curves, representing the functional properties described above. Both types of OBPs can occur in hydrothermal vent annelids and even co-occur in an organism, such as *R. pachyptila* where the different hemoglobins probably have complementary roles when the oxygenation conditions change (Hourdez and Weber 2005).

OBPs from hydrothermal-vent and cold-seep invertebrates usually exhibit a higher affinity for oxygen than OBPs from shallow-water relatives (Fig. 1.11). This high affinity allows the OBPs to extract oxygen from the environment, even when it is found at very low partial pressure. This high affinity is compensated by a pronounced Bohr effect that allows the release of oxygen near metabolically active tissues (local decrease of pH) (Hourdez and Lallier 2007). In *Branchipolynoe*, this release if further promoted by a specific CO_2 effect on one of the two main coelomic Hbs which affinity decreases when CO_2 concentration increases (Hourdez et al. 1999b).

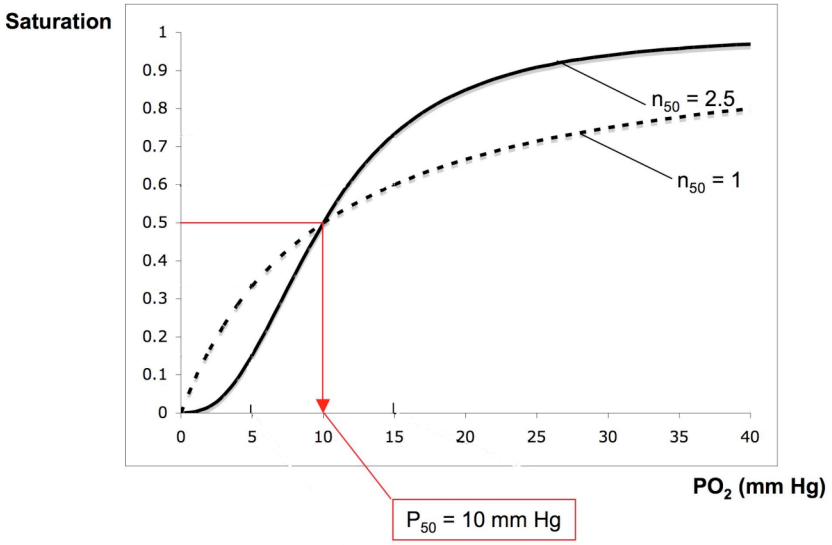


Figure 1.10. Saturation curves for hemoglobin (solid line) and myoglobin (dashed line) both exhibiting a P_{50} of 10 mm Hg (i.e. the same oxygen affinity), but with different levels of cooperativity (n_{50}). From Hourdez and Lallier 2007.

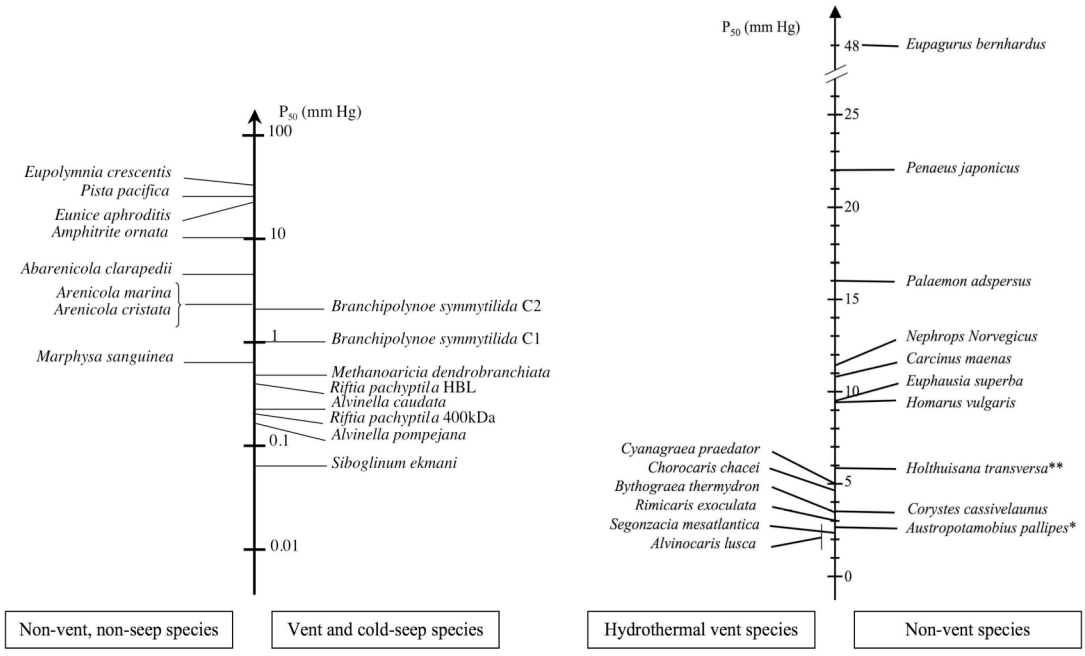


Figure 1.11. Representation of P_{50} values for annelids and decapod crustacea from hydrothermal and non-hydrothermal species. A higher P_{50} indicates a lower affinity. From Hourdez and Lallier 2007.

1.3 Invertebrate globins and hemoglobins

Respiratory pigments are very widespread in all the domains of life. They are known to be present in prokaryotes, unicellular eukaryotes and multicellular eukaryotes (plants, and animals) (Suzuki and Imai 1998). Although the diversity of the OBPs and their evolutionary history is fascinating, I will focus on the globins only (tissue globins and hemoglobins; Fig. 1.12) in one group of invertebrates, the annelids, as they are of the main interest for this thesis. The following section (1.4) will be devoted to the introduction of the Polynoidae and the occurrence of respiratory pigments in that family. For the other invertebrate groups the work of Vinogradov et al. (1993), Weber 1980, Weber and Vinogradov (2001) offer a complete review on the subject. I will use the designation of *globin* for oxygen-binding hemoproteins in general, *myoglobin* (Mb) when referring to tissue globins expressed in muscles, *neuroglobin* when expressed in the nerve cells, and *hemoglobin* (Hb) for circulating proteins (even if they are intracellular).

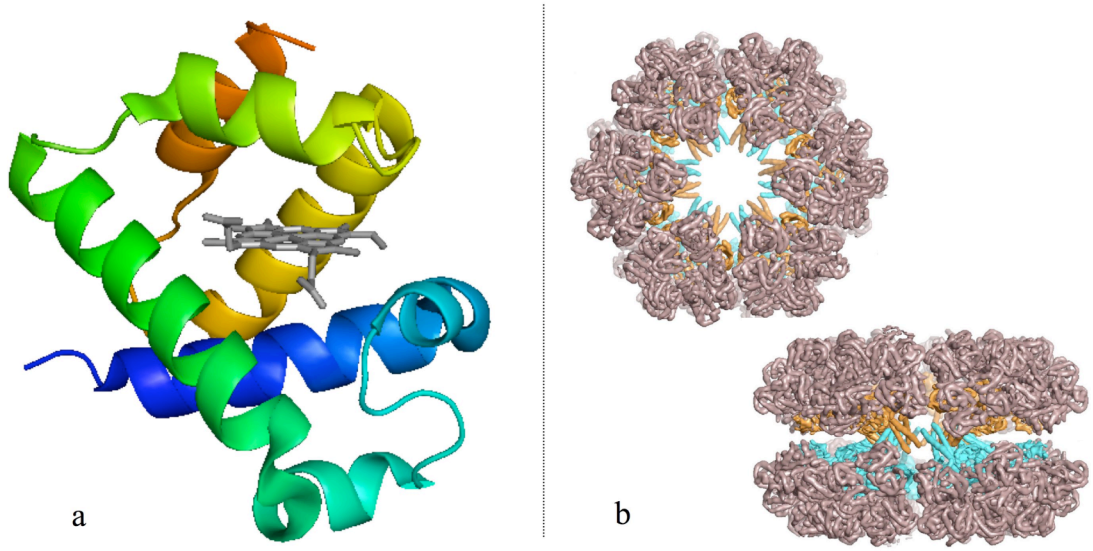


Figure 1.12. Example of (a) tissue monomeric globin (neuroglobin) from *Cerebratulus lacteus* (PDB database), and (b) extracellular multimeric hemoglobin from the lugworm *Arenicola marina* (from Royer Jr et al. 2007).

1.3.1 Structure and function

Globins from invertebrates exhibit a considerable heterogeneity in protein sequence and in quaternary structure when compared to vertebrate globins. However, despite this heterogeneity, the typical protein folding (the globin fold, that allows heme binding) shows an amazing structural conservation. This helps to group globins that *a priori*, based solely on sequence alignment data, would only share as much as 13% of identity (Rodionov and Johnson 1994). The quaternary structure diversity of invertebrate globins can be subdivided into five groups: 1) single-domain and monomeric, 2) single-domain and polymeric, 3) two-domain and polymeric, 4) multi-domain and polymeric and 5) single-domain and polymeric with a high molecular weights (Mw) (Vinogradov et al. 1993, Weber and Vinogradov 2001) (Fig. 1.13). The term globin domain (protein) will be used to refer to a functional protein unit: a polypeptide chain, of at least ~13 kDa that possesses one heme group and can reversibly bind O_2. The first, second and third groups can be both intracellular and extracellular. The fourth and fifth groups are only extracellular, and the high Mw is considered to be advantageous to avoid excretion and limit osmotic pressure (Weber and Vinogradov 2001). Accompanying the structural diversity there is also a functional diversity in invertebrate globins that does not have a parallel in vertebrate respiratory pigments (see Weber 1980, Weber and Vinogradov 2001 for a review).

Tissue globins can be found in several taxonomic groups such as nematodes, annelids, mollusks and arthropods, among others (Fox 1957, Schindelmeiser et al. 1979, Weber 1980, Weber and Vinogradov 2001, Wittenberg et al. 1965). In annelids Mbs are known to exist in the body wall of *Arenicola marina* (lugworm) that contains two isoforms of tissue globins (single-domain/monomeric) with similar oxygen affinity (Kreutzer and Jue 1997). In the sea-mouse *Aphrodita aculeata* there is a neuroglobin, a tissue ~17kDa globin specifically expressed in the nervous system (single-domain and monomeric), and with a lower oxygen affinity compared to the human Mb (Dewilde et al. 1996, Weber and Vinogradov 2001, Wittenberg et al. 1965).

Intracellular Hbs (from red blood cells), are found in the phyla Annelida, Mollusca, and in others like Echinodermata (Mangum 1992), and they usually have a relatively homogenous Mw of ~17kDa (Weber 1980, Weber and Vinogradov 2001).

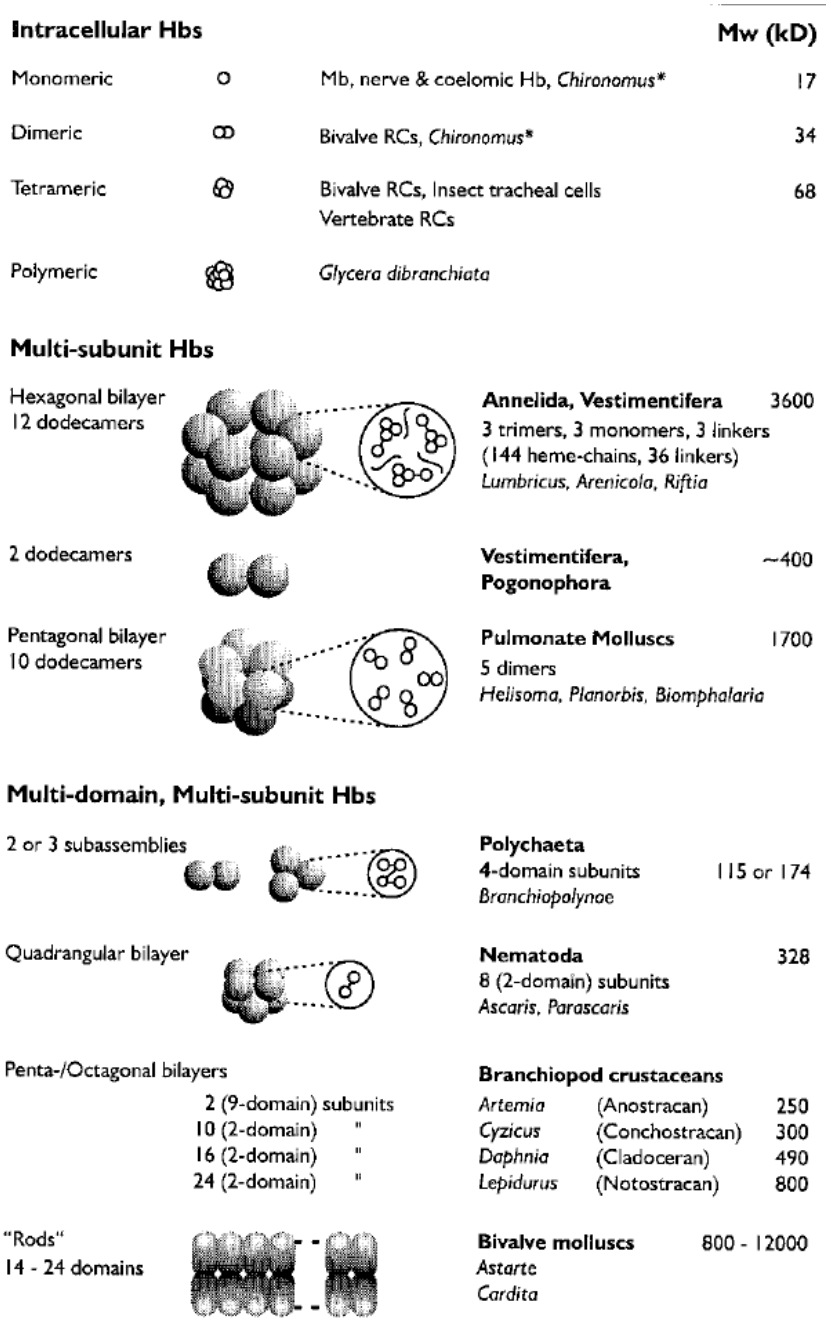


Figure 1.13. Schematic representation of the diversity of quaternary structures of invertebrate hemoglobins. From Weber and Vinogradov 2001.

Weber (1980) suggested that the monomeric state was the most common in annelid intracellular Hbs, with the exception of *Glycera* spp. (blood worm) (single-domain/monomeric and polymeric) that shows a pronounced heterogeneity in the Hb structure. Current knowledge indicates that several agglomeration states can be

observed in other annelid groups with intracellular Hbs (Vinogradov et al. 1993, Weber and Vinogradov 2001).

Extracellular Hbs (single-domain/multimeric and multi-domain/multimeric) are also widespread in several invertebrate groups, again in annelids and mollusks, but also in nematods and arthropods (Insecta) (Toulmond 1992). Annelids represent a special group as far as the structure of their extracellular Hb is concerned. Some annelid species such as *A. marina* and *Alvinella pompejana* (polychaeta), *Lumbricus terrestris* (oligochaeta), and *Hirudo medicinalis* (clitellata) possess a hexagonal bi-layer (HBL) globin structure (single-domain/multimeric) that comprises not only globin subunits but also linker chains (maintaining the HBL structure), the whole edifice reaching 3600 kDa (Terwilliger 1992, Weber 1980, Zal et al. 1997a) (Fig. 1.14).

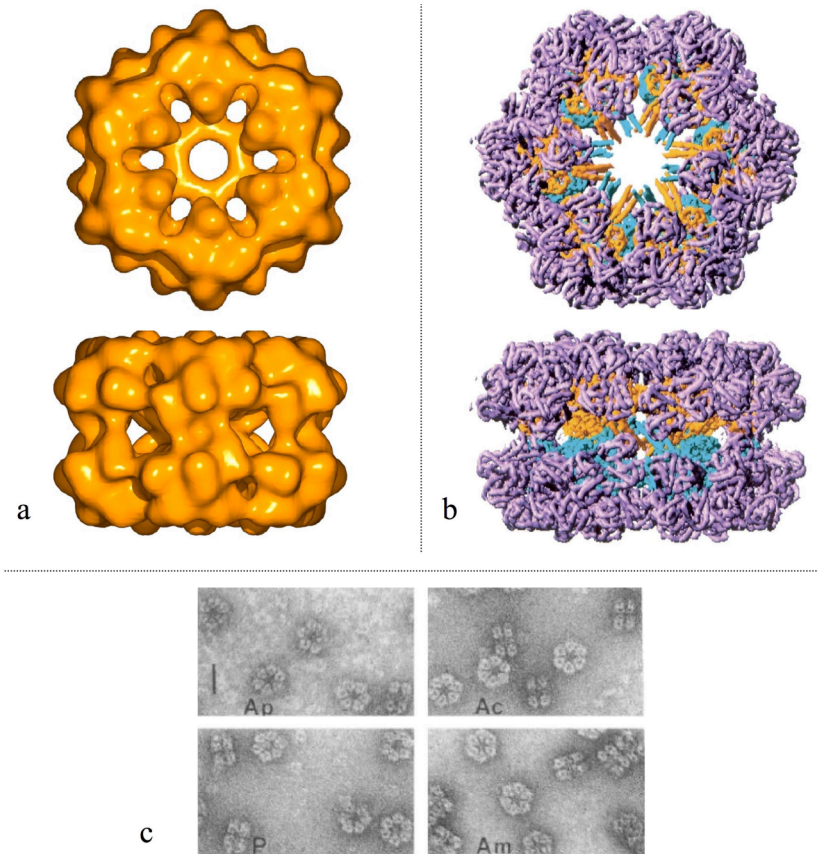


Figure 1.14. (a) Surface representation of the extracellular hemoglobin from *Alvinella pompejana*. From Jouan et al. 2003. (b) *Lumbricus terrestris* extracellular hemoglobin dodecamer structure. Hemoglobin chains are depicted in magenta, linker chains in yellow and blue. From Royer Jr et al. 2000. (c) Electron microscopy photograph from the HBL-Hb of 3 alvinellid species (Ap – *A. pompejana*; Ac – *A. caudata* and P – *Paralvinella grasslei*) and a littoral annelid (Am – *Arenicola marina*). From Toulmond et al. 1990. Scale-bar 25 nm. The scale differs between a, b, and c.

Smaller-sized extracellular Hbs can also be found in the polychaete family Siboglinidae (~400 kDa), sometimes co-existing with the previous type in the same organism, probably to enhance the O_2 transfer system (Terwilliger et al. 1980 in Arp et al. 1987, Toulmond 1992). A new extracellular globin with a unique structure for Annelida was discovered rather recently, exhibiting not only a multimeric assemblage but also a multi-domain structure, and having a fairly high oxygen affinity (Hourdez et al. 1999a, 1999b).

Hydrothermal vents species

Hydrothermal vent species with Hbs (vascular or coelomic), usually express them at high levels, resulting in high concentrations of these respiratory pigments, either in the coelome or the vascular system (Hourdez and Lallier 2007).

At hydrothermal vents, vestimentiferan tubeworms (siboglinid polychaetes), e.g. *Riftia pachyptila*, possess and use the already mentioned HBL Hb of ~ 3600 kDa in its vascular system. This HBL Hb not only reversibly binds O_2 but also sulfide, in order to "feed" their symbionts and to avoid exposing their tissues to sulfide, a toxic compound (Arp and Childress 1983, Arp et al. 1987, Childress et al. 1984). Sulfide does not bind to the heme group itself, and therefore does not form sulfhemoglobin, a form of Hb that is not capable of reversibly binding oxygen (Childress et al. 1984, Childress and Fisher 1992 and references there in). This vascular Hb is accompanied by another Hb of 400 kDa, as well as a somewhat different 400 kDa Hb in the coelome (Arp et al. 1990, Zal et al. 1996). These extracellular Hbs all have a high but slightly different oxygen affinity (Arp et al. 1990) that varies with pH. This supports the view that there is cooperation and a division of labor between different types of globins in the same organism under different situations (Arp et al. 1990, Hourdez and Weber 2005).

In alvinellids (Zal 1997b), although the globins do reversibly bind O_2, the capacity to bind sulfide does not seem to exist. Besides their vascular HBL-Hb they also possess a second type of Hb (single-domain/monomeric) in the coelome inside erythrocytes (Childress and Fisher 1992, Hourdez et al. 2000a, Jouin-Toulmond et al. 1996). A study on *Alvinella pompejana* showed that the oxygen binding properties of these two globin types are surprisingly similar (both with very high oxygen affinity) despite the different structural arrangement (Hourdez et al. 2000a). This high oxygen affinity is once more related to the low-oxygen environment. Oxygen affinity however decreases with

increasing temperature and *Alvinella* species are known to inhabit one of the hottest spots in the hydrothermal vent ecosystems, the upper part of the chimney walls (Toulmond et al. 1990). To explain this apparent contradiction, Childress and Fisher (1992) suggested that since the animal does not experience the same temperature in the whole body the higher temperature inside the tube would favor the release of O_2 and the lower temperature on the gill surface (outside the tube) would promote O_2 uptake. As discussed above, the sulfide detoxification mechanism in alvinellids is not yet well understood and the possibility that their HBL Hb could bind sulfide as *Riftia* HBL Hb is possible. This however would only temporarily immobilize sulfide that will then need to be detoxified.

Besides annelids other invertebrate groups such as Mollusca and Arthropoda possess extracellular Hbs (Hourdez and Weber 2005, Weber and Vinogradov 2001). At hydrothermal vents, the vesicomyid clams *Calyptogena* spp. are the only bivalve that was shown to have circulating respiratory pigments (Childress and Fisher 1992, Hourdez and Weber 2005). Their Hbs are erythrocytic, assembled into tetramers with relatively high oxygen affinity and a low Mw (Terwilliger and Terwilliger 1983, Zal et al. 2000). These bivalves possess symbionts in their gills and they use their foot to obtain sulfide from fissures in the basalt and transport it to their symbionts with a sulfide-binding compound that is not a hemoglobin (Childress et al. 1992, Zal et al. 2000). Among the arthropods living around hydrothermal vents, only the subclass Copepoda comprises species with Hb (single-domain/monomeric). These Hbs have very high oxygen affinities and low cooperativity, properties similar to thoses of myoglobins (Hourdez et al. 2000b, Sell 2000).

1.3.2 Adaptation and evolution

The primitive annelid globin was most likely monomeric and the original split between the intracellular and the extracellular annelid globins occurred about 570 millions years ago (Goodman et al. 1988). A phylogeny of the annelid globin sequences indicates that the intracellular and extracellular globins have a distinct evolutionary history (Fig. 1.15). The latter type evolved from a duplicated intracellular ancestral

globin, and several duplication events followed to produce the present-day globin diversity found in the complex extracellular HBL Hb (Bailly et al. 2007).

In this phylum, most of the globin gene sequences, whether they code for an intracellular or an extracellular globin, exhibit the typical 3 exon/2 introns vertebrate structure (Hardison 1996, Bailly et al. 2007). There is, however, the noticeable exception of the nerve globin from the aphroditid scale-worm *Aphrodita aculeata* which is missing the first intron (Dewilde et al. 1996). Nevertheless in the other annelid species the intron positions are highly conserved, at positions B12.2 and G7.0 (Bailly et al. 2007). This indicates that the ancestral gene for intra- and extra-cellular annelid globins had the same exon/intron structure and the same intron positions.

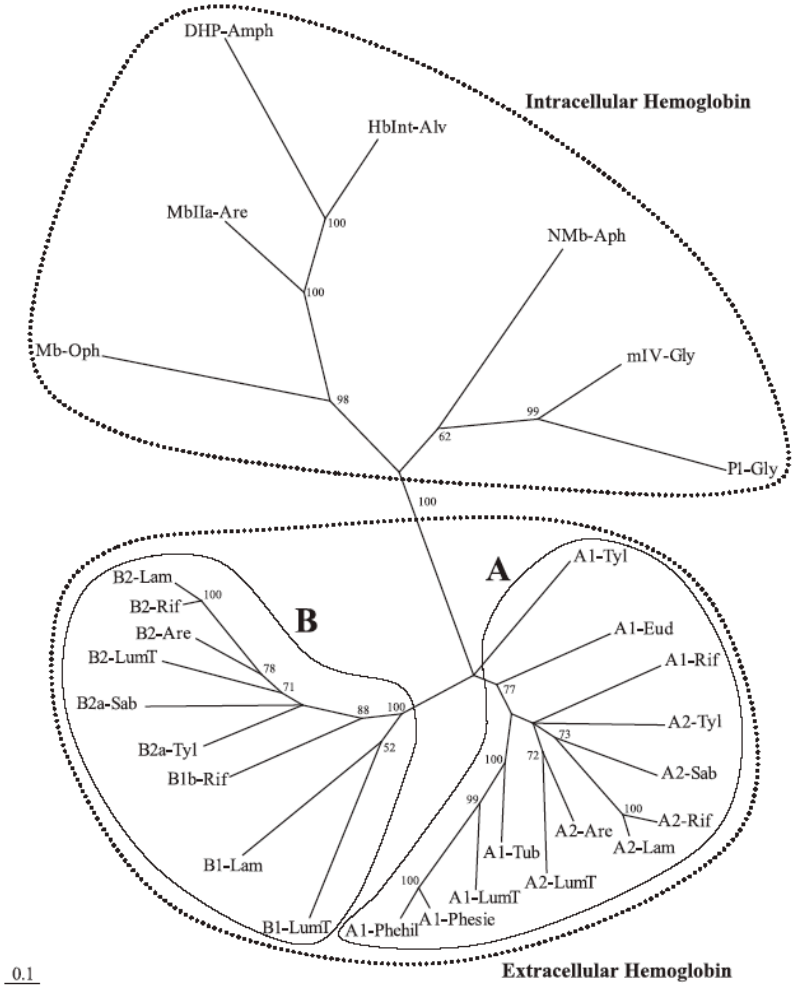


Figure 1.15. Bayesian phylogenetic tree based on annelid intra- and extra-cellular globins. Posterior probability values are shown near the nodes. From Bailly et al. 2007.

The polymerization of the globin subunits can be advantageous as it leads to the possibility of having cooperativity between subunits and a Bohr effect (Weber and Vinogradov 2001). The apparition of the HBL-Hb structure in annelids probably had the same goal, and in addition they present higher O_2 transport capacitance for the same colloid osmotic pressure than subassemblies or smaller subunits, and can avoid excretion (Weber and Vinogradov 2001). Other invertebrate groups, such as nematodes (e.g. *Ascaris*), mollusks (e.g. *Biomphalaria glarbata*), arthropods (e.g. *Daphnia*, *Artemia*) have multi-domain extracellular globins, in addition to being multimeric (Lieb et al. 2006, Manning et al. 1990, Sherman et al. 1992, Tokishita et al. 1997). In annelids multi-domain subunits were not known until the discovery of multi-domain extracellular Hbs in hydrothermal vent scale-worms from the family Polynoidae (Hourdez et al. 1999a).

Hydrothermal vents species

Respiratory pigments are undoubtedly a key to cope with the hypoxia/anoxia found in hydrothermal vents habitats (Hourdez and Lallier 2007). The majority of studies on respiratory pigments from hydrothermal vent organisms are however mainly on their physiology and not on their evolutionary history.

The first step was taken when Zal et al. (1997c) determined the primary structure of a globin chain from *Riftia pachyptila* Hb. In this work the globin phylogeny for that species showed that its respiratory pigments are close to other siboglinid Hbs. It also revealed the existence of free-cysteine residues (not involved in subunit interactions) in the protein sequence. These residues were earlier suggested as a potential site of sulfide binding (Suzuki et al. 1990). Later, Zal et al. (1998) indicated that this cysteine was one of the sulfide binding sites on this Hb, known to be the responsible for the sulfide transport in *Riftia*. This view was later challenged by Flores et al. (2005), and Flores and Hourdez (2006) who showed that the free-cysteine residues were not directly involved in sulfide binding, but instead the responsible were zinc ions bound to other specific amino acids, at least in the 400 kDa *Riftia* Hb. Interestingly however, these cysteines are also shared with other species inhabiting sulfidic habitats, including a littoral species (*Arenicola marina*) (Bailly et al. 2003). The evolutionary pathway for the presence of the free-cysteines residues was later investigated by Bailly et al. (2002, 2003 and 2007) showing that these amino acids are well conserved in annelids living in

sulfidic habitats and that they were the target of positive selection. This most likely indicates a function for these specific cysteine residues in annelids from sulfidic habitats but the function remains unclear.

1.4 The Polynoidae, an interesting working model

1.4.1 Hydrothermal vent Polynoidae diversity and evolutionary history

The family Polynoidae belongs to the Phyllodocida, a suborder of the phylum Annelida (Rouse and Fauchald 1997) (Fig. 1.16). This family of scale-worms is widely distributed in ecosystems from the littoral to the deep-sea having free-living or commensal lifestyles (Rouse and Fauchald 1997, Tunnicliffe 1991). In general scale-worms (e.g. Aphroditidae, Polynoidae) are known for only having tissue globins in the nervous system (neuroglobin) (Weber 1978, Dewilde et al. 1996), although they possess a reduced closed circulatory system, but with no heart and no respiratory pigment (Fauchald and Rouse 1997, Jaquet 1886 in Hourdez and Jouin-Toulmond 1998). They can possess epidermal extensions under the elytrae but these structures do not form true gills, with the possible exception of the symbiotic species *Arctonoe vittata* (Martin and Britayev 1998), although the branchial function of the appendage found in *Arctonoe* has not been studied.

In contrast, the symbiotic hydrothermal vent scale-worms, from the genus *Branchipolynoe*, show developed branchiae over the parapodes, that are filled with red coelomic fluid (Hourdez and Jouin-Toulmond et al. 1998) (Fig. 1.17c). This genus is symbiotic with several hydrothermal mussels species, depending on the geographical area where they occur: *B. symmytilida* on the East Pacific Rise having as host *Bathymodiolus thermophilus*, *B. seepensis* on the North Atlantic Ridge with *B. puteoserpentis* and *B. azoricus* as hosts (and with other mussel species at cold seeps, where it is also found, although the seep species is genetically distinct) and finally *B. pettiboneae* in the West Pacific living in *B. brevior* and *B. elongatus* (Chevaldonné et al. 1998, Hourdez 2000). The phylogeny of both symbiont and host were compared by Jollivet et al. (1998), using ITS2 and ribosomal DNA. These authors showed that *Branchipolynoe* evolved slightly faster than their hosts, but in contrast to what seems

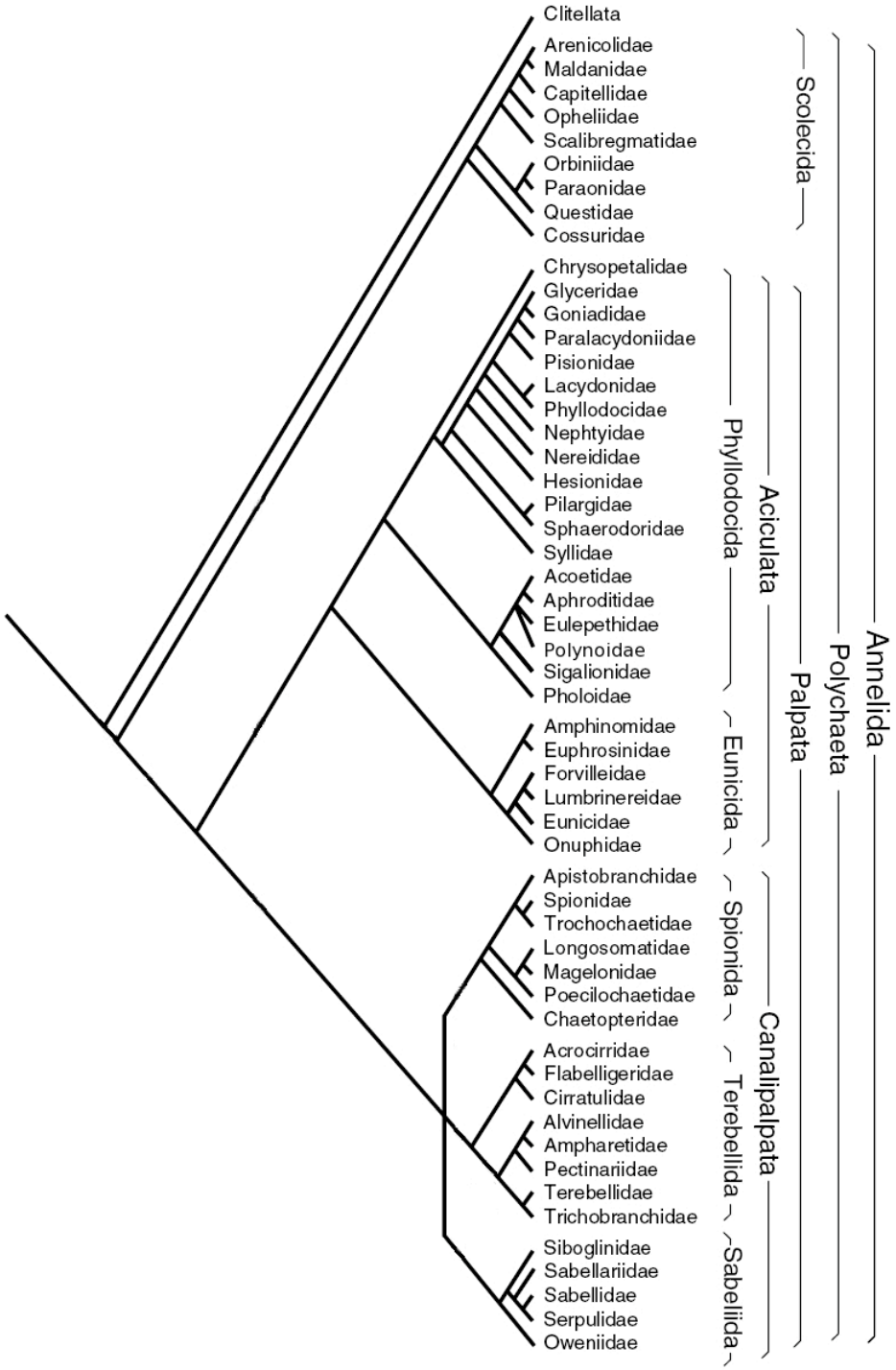


Figure 1.16. Annelid phylogeny from Rouse and Fauchald (1997) and modified by Bartolomaeus et al. (2005).

normal in symbiotic relationships, there is more specific diversity in the hosts than in the symbionts group. An interesting study on the kinship of *B. seepensis* females and juveniles found in the same host mussel revealed that they are most often not related, and that the juveniles are likely released at a relatively early stage of development (Plouviez et al. 2008). The relationship of *Branchipolynoe* with its host remains debated. Some studies indicate that they likely are opportunistic commensals (Chevaldonné et al. 1998, Desbruyères et al. 1985, Pettibone 1984, 1986). A study on carbon isotopes ($\delta^{13}C$) showed that there is a close nutritional relationship between *Branchipolynoe* and its host mussels, however is not clear if both eat the same source of food or if the worms feed on the mussels. The authors suggest that *B. symmytilida* feeds on bacteria, mucus, and pseudo-feces from the mussel (Fisher et al. 1988).

About 45 other Polynoidae species that are not symbiotic have been described from hydrothermal vents, and all together represent about 10% of all the endemic vent species (Tunnicliffe 1991). Most of these species belong to endemic subfamilies (Lepidonotopodinae, Branchpolynoinae, Branchiplicatinae, Branchinotogluminae), and some species belong to the Iphioninae (genera *Iphionella* and *Thermiphione*), a small subfamily that also comprises species collected mostly on coral reefs, and the "usual" (i.e. non-hydrothermal vent) deep-sea. They occupy all the microhabitats available to metazoa, from the coldest (and most oxygenated), to the warmest (and usually the most hypoxic) (Hourdez 2000) (Fig. 1.17). Among them the genus *Branchinotogluma* is particularly interesting as one species lives on the chimney wall in the Lau Basin (Pacific), close to *Paralvinella fijiensis*, likely experiencing high temperatures and sulfide concentrations, and low oxygen concentrations (Hourdez pers. comm) (Fig. 1.17). In the same area other *Branchinotogluma* species occur but not as close to the source of fluid. *B. trifurcus* lives among the gastropods *Ifremeria* and the mussels *Bathymodiolus* and *B. jasoni* at the periphery of the vents ecosystems. Other polynoids that live among the mussels are *Thermopolynoe branchiata*, *Levesteniella raisae* and *Lepidonotopodium minutum*. *Thermiophionae fijiensis* occurs on the periphery along with *B. jasoni* (Hourdez pers. comm.). Still in the Pacific, but on the EPR, *Branchiplicatus cupreus* occurs in *Riftia pachyptila* aggregations but is more common in the mussel communities (Govenar et al. 2005, Hourdez pers. comm.). *Riftia* thickets also harbor *Lepidonotopodium williamsae*, *Branchinotogluma grasslei*, *B. sandersi*, *B. hessleri*, and *Opisthotrochopodus alvinus*. The chimney walls on the EPR are inhabited

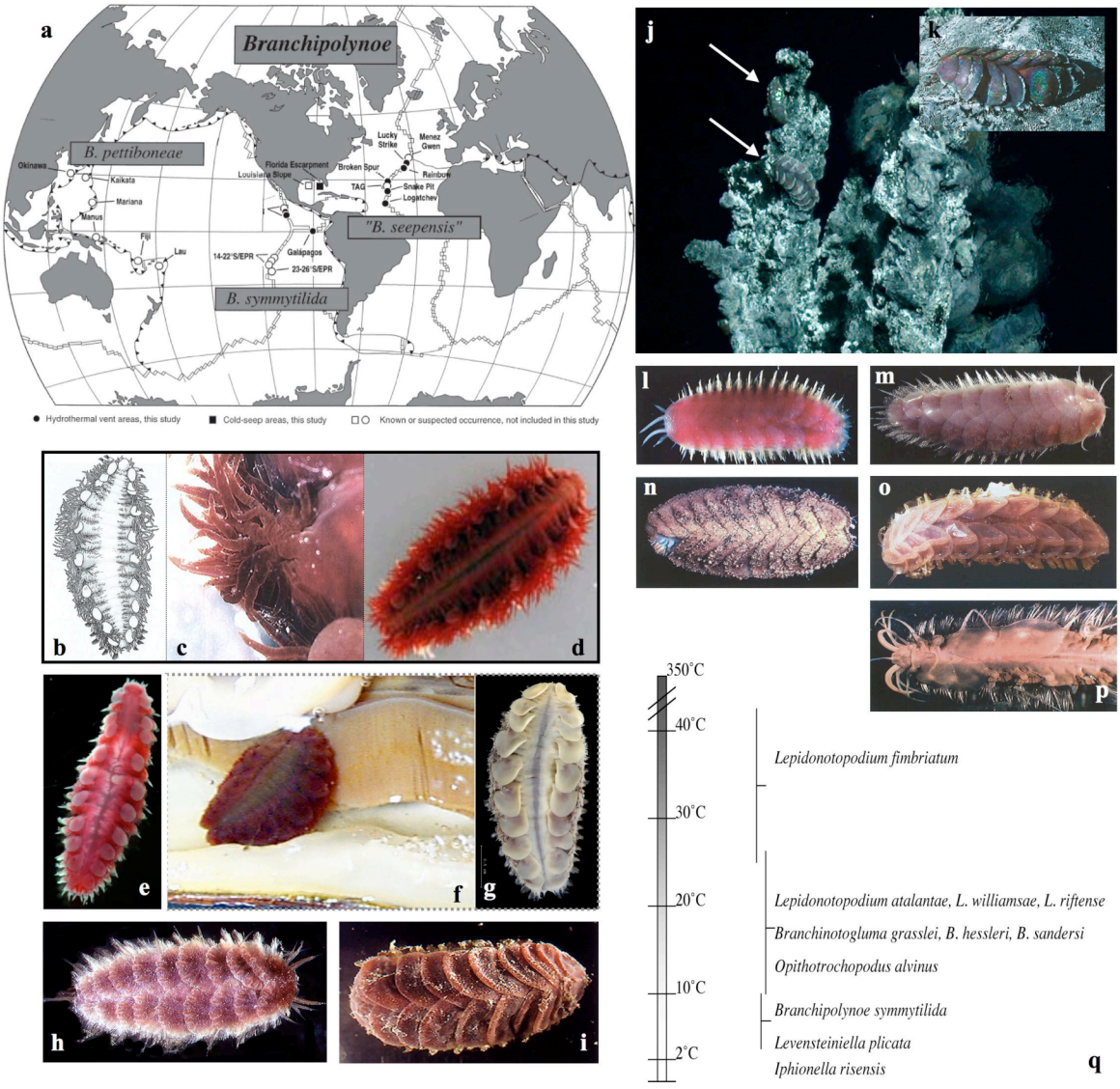


Figure 1.17. Several Polynoidae species. (a) distribution of the genus *Branchipolynoe*; (b) *Branchipolynoe symmytilida* illustration; (c) detail from *B. symmytilida* gill in vivo; (d) *B. symmytillida*; (e) *B. pettiboneae*; (f) *B. seepensis* inside hydrothermal mussel; (g) *B. seepensis*; (h) *Lepidonotopodium atalantae*; (i) *L. williamsae*; (j) *Branchinotogluma segonzaci* on a chimney wall; (k) *B. segonzaci*; (l) *B. trifurcus*; (m) *B. sandersi*; (n) *Iphionella risensis*; (o) *Levensteiniella plicata*; (p) *Branchiplicatus cupreus*; (q) thermal distribution of some of scale-worms in the EPR (9º50'). Credits: (a) from Chevaldonné et al. 1998; (b) V. Martin, (c) S. Hourdez, (g) A. Fifis, (h) and (i) Ifremer, (k) courtesy of C. R. Fisher, (l) courtesy of F. Pleijel, (m) P. Briand, (n) and (o) Ifremer, (p) P. Briand, all from Desbruyères et al. 2006; (f) S. Hourdez, from Hourdez and Weber 2005; (d) T. Shank, from http://www.ridge2000.org/SEAS/for_students/data/field_notes_05.html; (e) G. Rouse, from http://www.noc.soton.ac.uk/chess/easter05/29marzo_pt.html; (j) Lau Basin 2009 cruise.

by *Alvinella*, and the caterpillar worm *Hesiolyra bergi*, the scale-worm *Lepidonotopodium fimbriatum* are often found crawling among *Alvinella* tubes. Two species of the genus *Levensteiniella*; *L. intermedia* and *L. plicata* also occur among the mussels in the EPR. The Mid Atlantic Ridge is poorer in the occurrence of polynoid species, besides the symbiont *B. seepensis*, we can also find *L. iris*, *L. jouinae* and two species of *Branchinotogluma* in the mussel beds (Hourdez and Desbruyères 2003, Hourdez pers. comm.).

Hydrothermal vent scale-worms belonging to endemic subfamilies form a monophyletic group that probably originated about 65 Ma (Hourdez et al. in prep.) (Fig. 1.18). This suggests that there most likely was a single colonization event that gave rise to all the species of these endemic subfamilies. In addition, the short internal branches indicate that after the initial colonization, there was a rapid radiation and speciation, and the species most likely colonized the various microhabitats. The genera *Iphionella* and *Thermiphione* (Iphioninae) cluster with the shallow-water genus *Iphione*, in a group clearly distinct from all other hydrothermal vent species. They however correspond to species that are collected on the outskirts of the vent communities (or among the assemblages that live in the coldest and least constraining part of the vent). They may actually correspond to typical deep-sea species (i.e. not vent endemic) that are simply captured there because that is were the sampling effort is concentrated.

Because we have numerous species that occupy such contrasted environments and their phylogeny indicates that the vents scale-worm species are so closely related, we have an ideal setting to study adaptations. The observation we make at the morphological, biochemical, or molecular level can more easily be compared and interpreted in a phylogenetic context.

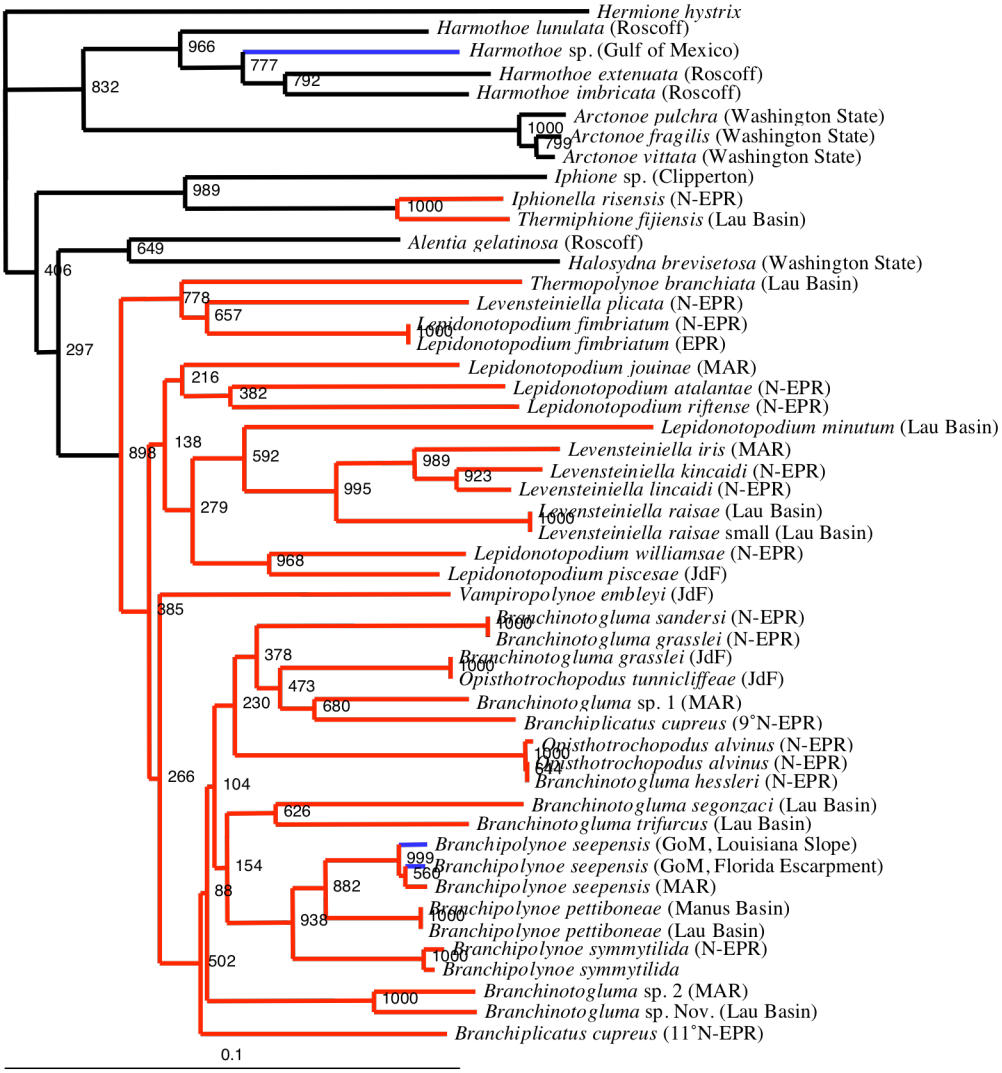


Figure 1.18. Phylogeny of the Polynoidae family based on 515 bp alignment of the mitochondrial 16S gene. Bootstrap values are indicated near the nodes (1000 replicates). Red branches: species collected at hydrothermal vents, blue branches: species collected at cold-seeps, black branches: littoral species.

1.4.2 Polynoidae respiratory adaptations

1.4.2.1 Presence of gills

In contrast to the littoral species, 17 out of 45 hydrothermal polynoids species exhibit ramified branchiae. These belong to the genera *Branchipolynoe* (Branchipolynoinae), *Branchinotogluma*, *Opisthotrocopodus* and *Peilaneopolynoe* (all three Branchinotogluminae), in addition to *Branchiplicatus cupreus* (Branchiplicatinae) and *Thermopolynoe branchiata* (Lepidonotopodinae) (Hourdez and Jouin-Toulmond 1998). Most of these species form a monophyletic group, indicating a single origin for the gills (Hourdez et al. in prep.). *Thermopolynoe branchiata* however is an exception as it clearly belongs to the Lepidonotopodinae and the presence of gills in this species is most likely the result of convergent evolution. This is supported by the fact that the gill structure appears very different in this species compared to the other branchiate group (Hourdez et al. in prep.).

The most developed branchiae are probably in the symbiotic genus *Branchipolynoe* (Fig. 1.17c). All three species in this genus exhibit interesting respiratory adaptations such as a large gill surface area and a smaller diffusion distance between the coelome and the environment than in other parts of the body (Fig. 1.19). These adaptations are similar to those observed in other species inhabiting sulfidic habitats, such as *Alvinella pompejana* and *Paralvinella grasslei* (Hourdez and Jouin-Toulmond 1998). In *Branchipolynoe* however, the gills correspond to mere expansions of the body wall and are not vascularized but rather perfused by the coelomic fluid. This state is radically different from most other known annelid gills, with the exception of *Glycera* that also has coelomic gills. The coelomic fluid circulates thanks to the action of cilial and myoepithelial contractions (Hourdez and Jouin-Toulmond 1998).

1.4.2.2 Presence of respiratory pigments

All species from hydrothermal vent endemic subfamilies have a pink- to red-colored coelomic fluid, a color due to the presence of extracellular Hbs (Hourdez pers. comm.). To date, only the extracellular coelomic Hbs from *Branchipolynoe* have been characterized (Hourdez et al. 1999a, 1999b). There are two types of quaternary structure: trimers (HbC1) and dimers (HbC2) with ~153 kDa and ~123 kDa,

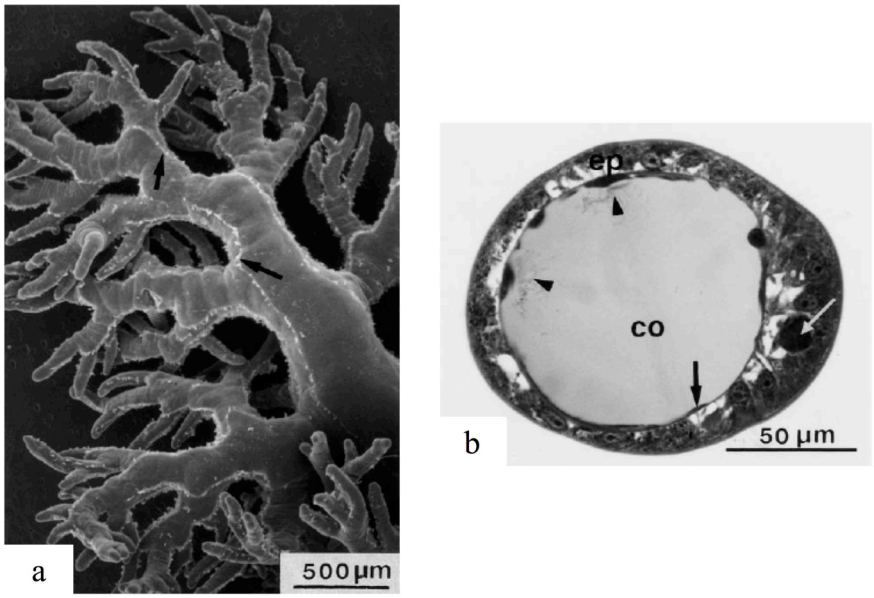


Figure 1.19. Gill details from *B. seepensis*. (a) Scanning electron micrograph of a part of the gill. Arrows indicate ciliated lines. (b) Histological section of a ramification from the gill; ep – epidermis, co – coelome, black arrow – coelomic epithelium, white arrow – lysosome, arrow heads - cillia. After Hourdez and Jouin-Toulmond 1998.

respectively, a Mw that is very unusual for extracellular annelid Hbs (Fig. 1.20). Besides the polymeric structure, it was also shown that *Branchypolynoe* extracellular Hbs where multi-domain, a unique characteristics for annelids (Hourdez et al. 1999a). Each subunit corresponds to a tetra-domain, a succession of 4 globin domains, each with one heme group capable of reversibly binding a dioxygen. In the same study a smaller extracellular Hb (HbC3), probably single-domain was also observed.

The oxygen binding properties of these globins showed a low cooperativity, a high affinity for oxygen (both characteristics reminiscent of myoglobins), and a strong Bohr effect. The high affinity for oxygen and the pronounced Bohr effect are characteristics that are typical of OBPs from species adapted to hypoxic environments. Usually CO_2 has a very limited or no effect on extracellular Hbs. Interestingly, HbC1 and HbC2 exhibit different responses towards P_{CO2} variation, the first is insensitive to P_{CO2} variations and the second is affected, decreasing its oxygen affinity (Hourdez et al. 1999b, Hourdez and Weber 2005). In addition, these coelomic respiratory pigments allow the blood to have a large buffer capacity, mostly because they probably bind CO_2 (Hourdez et al. 1999b).

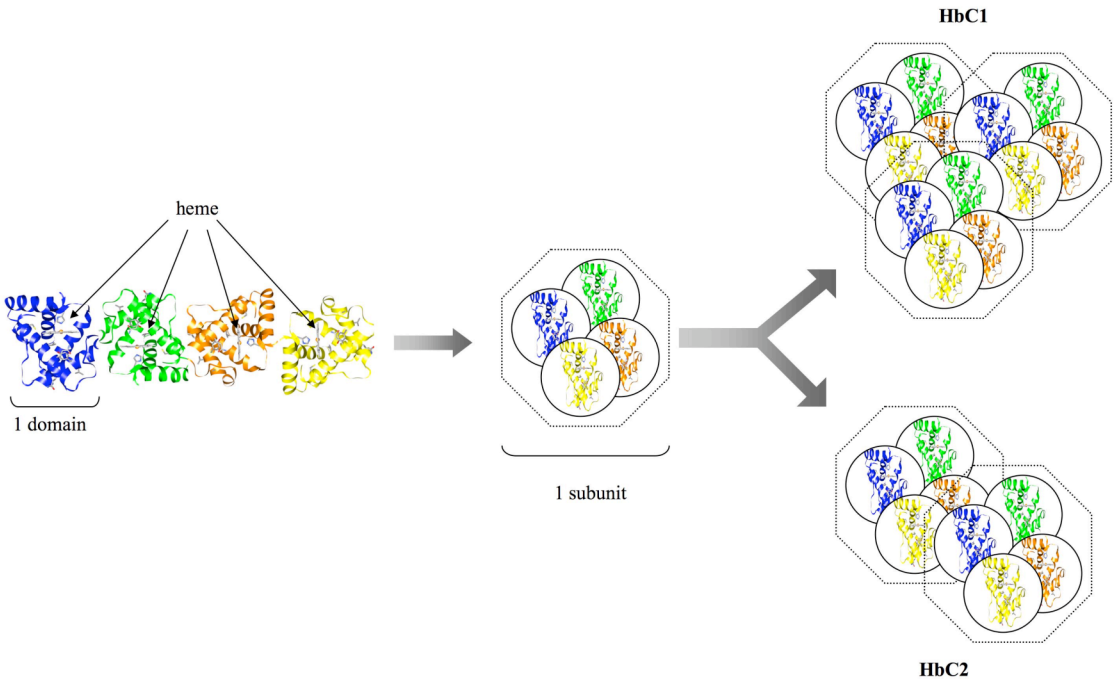


Figure 1.20. Model of the two quaternary forms of the *Branchipolynoe symmytilida* extracellular tetra-domain hemoglobins; HbC1: trimer (153 kDa) and HbC2: dimer (124 kDa). Each color represents a different protein domain, each one with a heme group.

Taking all these properties into account, and adding the fact that they do not circulate in one specific direction in the gill, they most likely have an oxygen storage function instead of transport (Hourdez and Jouin-Toulmond 1998, Hourdez et al. 1999b). They probably allow the worms to buffer the variations of environmental oxygen partial pressure. For example, based on Hb concentration and metabolic rate, it was estimated that the Hb-bound oxygen would be sufficient to meet the aerobic metabolic needs of a 1-gram worm for about 90 minutes (Hourdez and Weber 2005).

Interestingly, the reduced vascular system in *Branchipolynoe* also contains a Hb that seems to have a typical annelid extracellular HBL-Hb structure (Hourdez and Jouin-Toulmond 1998, Hourdez et al. 1999b). It is thought that having different Hbs in different compartments may allow gas transfer and internal buffering of different environmental O_2 tensions (Hourdez and Weber 2005). However, one can speculate that the 3600 kDa vascular globin will not have a major role on the general oxygen

storage/transport because the coelomic extracellular Hbs (HbC1 and HbC2) are expressed at much higher concentration, are distributed in the whole body, and the gills are not vascular. The potential role of this vascular Hb, and its overall significance for the worms remain to be determined.

1.5 Thesis objectives

The hydrothermal environment is without a doubt a very constraining ecosystem where only the organisms with the appropriate adaptations (respiratory, toxicity, thermal, etc.) can live and thrive. Numerous studies have been devoted to the adaptations of hydrothermal vent emblematic species, in particular in the vestimentiferan tubeworms, and the 'Pompei' worm *Alvinella pompejana*. Not all annelid hydrothermal species have received the same attention, even though they might represent equally interesting cases of adaptation and evolution, than the most studied species. The scale-worms from the family Polynoidae are certainly one of those cases, specifically in regard to respiratory adaptations. Having the evolution of the respiratory adaptations in hydrothermal scale-worms (Polynoidae) as the subject of this thesis, the study of the unique annelid multi-domain extracellular globins was inevitable. According to this, two major goals were established.

The first dealt with the study of the evolutionary history that lead to the appearance of this kind of extracellular Hb in polynoids. As it was a previously unknown structure in this phylum several questions arose:

1) would the origin and evolutionary history of these globins be similar to other annelids extracellular globins;

2) or to other groups multi-domain globins;

3) what would be their evolutionary and physiological advantage of having several domains in correlation with the polynoids' habitat; and

4) what structural differences/advantages would these globins present.

The second chapter of this thesis addresses these questions, specifically on the extracellular tetra-domain Hb gene of two *Branchipolynoe* species: *B. symmytilida* and *B. seepensis*. This chapter was divided into two parts, the first deals with the evolutionary history of the formation of this gene by duplication, and the second aims at elucidating what drove the evolution of this structure, at the amino acid level, and potentially the advantages that this gene represents in front of a hypoxic environment.

The second objective was to understand the evolution of the extracellular single-domain Hb gene, found in all hydrothermal endemic polynoid species. These species occupy a range of different habitats at hydrothermal vents, with different constraints, in particular in regard to aerobic respiration. In this context, several hypotheses/questions were raised:

1) is there structural differences of the single-domain extracellular Hb in hydrothermal polynoid species occupying different habitats, and potentially affecting the function of these globins;

2) if so, which residues (nature and position on the 3D structure) would be responsible for the variability; and

3) what would be their role in the adaptation to hypoxia in these annelids.

Chapter three comprises different analyses that were performed to answer these different hypotheses and questions.

This thesis attempts to understand how the strong constraints found at hydrothermal vents can can lead to evolutionary innovations, through the tinkering of existing genes to produce a protein expressed in a different context (coelomic globin originating from an intracellular myoglobin-like globin), and how minute changes in the sequence can have tremendous effects at the protein level that will be reflected in a better survival under challenging conditions. Hydrothermal vents very clearly offer a drive for species to colonize it (local primary production in large amounts) and yet they also require specific adaptations. This setting is very likely to facilitate and encourage innovation, a result of positive Darwinian selection.

2. Origin and evolution of the tetra-domain hemoglobin from the hydrothermal vent scale-worm *Branchipolynoe*

2.1 Introduction

Proteins with a multi-domain structure are not rare in Nature. They actually account for half of all the sequenced proteomes, and are more abundant in eukaryotes than in prokaryotes (Han et al. 2007). Multidomain globins can be intracellular, such as the two-domain hemoglobin (Hb) in *Barbatia* (Bivalvia, Mollusca) (Grinich and Terwilliger 1980, Suzuki and Arita 1995), or extracellular such as in crustaceans, with two (Conchostraca, Cladocera and Notostraca) to nine domains (Anostraca); nematodes (e.g. *Ascaris suum*) with a didomain structure; molluscs with 10-12 and 14-24 domains in gastropods (Planorbidae) and bivalves (Astartidae and Carditidae), respectively (Weber and Vinogradov 2001).

The proposed evolution processes for multi-domain globins involve alternative splicing, fusion/fission and shuffling events, depending on the organisms (Kriventseva et al. 2003, Pasek et al. 2006, Pathy 1999). For example, to form the intracellular two-domain hemoglobin from *Barbatia*, it was proposed an unequal crossing-over without a creation of a linker region for *B. reeveana* (Naito et al. 1991), and a duplication of an ancestral δ globin gene, with the loss of the stop codon, for *B. lima* (Suzuki and Arita 1995, Suzuki et al. 1996). As far as extracellular globins are concerned, in crustaceans, duplications of an ancestral *Artemia* monodomain gene are thought to be the base process for its evolution into a 9-domain hemoglobin (with a linker sequence between each one domain) (Manning et al. 1990, Trotman et al. 1994). Jellie et al. (1996) suggest that the duplication first formed triplets, which were in turn duplicated to form the 9 domains. Later, Matthews et al. (1998) suggested, and gave evidence for the whole duplication of the multidomain hemoglobin gene *Artemia* around 60 MY ago, giving birth to two copies that are since then transcribed into two polymers (T and C). For *Daphnia magna* Tokishita et al. (1997)

suggested very early gene duplication or possibly the fusion between two different globin genes, since they present a very low homology (24%) between them. Later, in a study on *D. pulex*, the mechanism attributed to the hemoglobin evolution was the same as for *B. reeveana*, unequal crossing-over, along with the lack of linker sequence between domains (Dewilde et al. 1999). The same authors confirm that only a preA sequence (a sequence before the A helix) is used to link both domains. In nematodes, Dixon et al. (1992) suggest an unequal crossing-over, duplication of the whole gene or even a mispaired gene conversion for the existence of two-domain globin in *Pseudoterranova decipiens*. For *Ascaris*, only gene duplication is discussed, as the intron positions are conserved between domains (Goldberg et al. 1995), contrary to *P. decipiens* (Dixon et al. 1992). In nematodes there is no linker structures between the domains (De Baere et al. 1992, Dixon et al. 1992), in contrasst to *Artemia*. Studies on the extracellular multi-domain Hb from the gastropod *Biomphalaria glabrata* suggest multiple gene duplications and gene fusion events from an ancient pulmonate myoglobin to produce the 10-13 Hb domains (Arndt et al. 1998, Lieb et al. 2006). In annelids only the Polynoidae family (two hydrothermal species), possess extracellular multi-domain globins. The quaternary structure of these respiratory pigments is also peculiar, corresponding to dimers and trimers (Hourdez et al. 1999a) instead of the HBL-Hb structure that is typical for annelids. The presence of Hb in these species most likely has an adaptive value in the hydrothermal vents environment, characterized by chronic hypoxia (Hourdez and Lallier 2007).

As no study has yet addressed the gene structure and the evolution history of these unique annelid globins, we decided to study and describe the tetradomain hemoglobin gene structure from *Branchipolynoe* (Polynoidae). This work was divided in two parts, the first with the goal to describe the gene structure and understand the duplication history of the different domains. This work constitutes a scientific article submitted to Molecular Biology and Evolution, currently *in press*. In the second part, we looked for signs of positive selection in specific *Branchipolynoe* globin domain lineages in order to understand the adaptations that were necessary in the evolution of this protein.

2.2. Origin and evolution of the tetra-domain hemoglobin from the hydrothermal vent scale-worm *Branchipolynoe*

This part of chapter 2 will be presented in the format that was accepted in the journal of Molecular Biology and Evolution, with the exception of the figures position. They will appear close to the section of text where they are mentioned. We were asked to move some of the methods and images to supplementary material. These will appear at the end of the article. Some of the bibliography included in this work will not be a part of the references at the end of this thesis, and only appear in the article.

Origin and evolution of the unique tetra-domain hemoglobin from the hydrothermal vent scale-worm *Branchipolynoe*

Projecto-Garcia[1,2], J., N. Zorn[3], D. Jollivet[1,2], S. W. Schaeffer[4], F. H. Lallier[2,5] and S. Hourdez[1,2]

[1] Équipe GAME, CNRS UMR 7144, Station Biologique de Roscoff, BP 74, Places Georges Teissier, 29682 Roscoff cedex, France

[2] Université Pierre et Marie Curie, Laboratoire Adaptation et Diversité en Milieu Marin, Place Georges Teissier, 29682 Roscoff cedex, France

[3]LDSM2 Institut de Chimie de Strasbourg, UMR 7177 CNRS-UDS, 1 rue Blaise Pascal - BP 296 R8, 67008 Strasbourg Cedex, France

[4] Department of Biology and Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802, USA

[5] Équipe EIMME, CNRS UMR 7144, Station Biologique de Roscoff, BP 74, Places Georges Teissier 29682 Roscoff cedex, France

Corresponding author:

Stéphane Hourdez (hourdez@sb-roscoff.fr)

Équipe GAME, CNRS UMR 7144, Station Biologique de Roscoff, BP 74, Places Georges Teissier, 29682 Roscoff cedex, France

Tel: +33 (0)2 98 29 23 40

Fax: +33 (0)2 98 29 23 24

Keywords: extracellular globins, tetra-domain, tandem duplication, evolution, annelids, *Branchipolynoe*

Running head: Scale-worm tetra-domain globin evolution

Abstract

Hemoglobin is the most common respiratory pigment in annelids. They can be intra- or extra-cellular, and this latter type can form large multimeric complexes. The hydrothermal vent scale-worms *Branchipolynoe symmytilida* and *B. seepensis* express an extracellular tetra-domain hemoglobin that is unique in annelids. We sequenced the gene for the single-domain and tetra-domain globins in these two species. The single-domain gene codes for a mature protein of 137 amino acids, and the tetra-domain gene codes for a mature protein of 552 amino acids. The single-domain gene has a typical 3 exons/2 introns structure, with introns located at their typical positions (B12.2 and G7.0). This structure is repeated four times in the tetra-domain gene, with no bridge introns or linker sequences between domains. The phylogenetic position of *Branchipolynoe* globins among known annelid globins revealed that, although extracellular, they cluster within the annelid intracellular globins clade, suggesting that the extracellular state of these hemoglobins is the result of convergent evolution. The tetra-domain structure likely resulted from two tandem duplications, domain 1 giving rise to domain 2 and after this the two-domain gene duplicated to produce domains 3 and 4. The high O_2 affinity of *Branchipolynoe* extracellular globins may be explained by the two key residues (B10Y and E7Q) in the heme pocket in each of the domains of the single and tetra-domain globins, which have been shown to be essential in the oxygen-avid hemoglobin from the nematode *Ascaris suum*. This peculiar globin evolutionary path seems to be very different from other annelid extracellular globins and is most likely the product of evolutionary tinkering associated with the strong selective pressure to adapt to chronic hypoxia that characterizes hydrothermal vents.

Introduction

Globins produced by invertebrates exhibit considerable heterogeneity in protein sequence and in quaternary structure when compared to vertebrates. Despite this heterogeneity, these proteins fold into the highly conserved globin fold that allows heme binding. The quaternary structure diversity encompasses simple monomers, polymeric subunits made of single-domains, and polymeric multi-domain subunits (see Weber and Vinogradov 2001 for a review). Multi-domain globins can be intracellular, such as in the bivalve *Barbatia* (Grinich and Terwilliger 1980, Suzuki and Arita 1995), or extracellular such as in crustaceans, nematodes, and molluscs (see Weber and Vinogradov 2001 for a review). This type of structure was not known in annelids until its discovery in two closely related hydrothermal vent species of annelids, *Branchipolynoe symmytilida* and *B. seepensis*, which possess tetra-domain globins (Hourdez et al. 1999a).

The presence of hemoglobin (Hb) in *Branchipolynoe* spp. may have an adaptive value in their chronically hypoxic environment (Hourdez and Lallier 2007). *Branchipolynoe* belongs to the Polynoidae, a family of scale-worms that is widely distributed in marine ecosystems from the littoral to the deep-sea. They are very diverse at hydrothermal vents (Tunnicliffe 1991) where they occupy all the microhabitats available to metazoa, from the coldest (and most oxygenated), to the warmest (and usually the most hypoxic) waters. Scale-worms in general (comprising several families, and including the Polynoidae) were known for having only tissue globins (neuroglobin and myoglobin) (Weber 1978; Dewilde et al. 1996). In contrast to their littoral relatives, most hydrothermal vent polynoid species have extracellular Hbs giving them a red-pigmented coelomic fluid (Hourdez et al. 1999a; S.H. pers. obs.). These Hbs can facilitate the diffusion of oxygen from the hypoxic environment, especially given their high affinity for oxygen (Hourdez et al. 1999b), and represent a significant oxygen storage for periods of complete anoxia (Hourdez and Weber 2005).

A phylogeny of the annelid globin sequences showed that the extracellular globins all evolved from a duplicated intracellular ancestral globin, and several duplication events followed to produce the present-day globin diversity found in the complex extracellular hexagonal bilayer Hb (HBL-Hb; Bailly et al. 2007). The original split between the intracellular and the extracellular globins occurred about 570 millions years ago in annelids (Goodman et al. 1988).

The hydrothermal vent environment was colonized by scale-worms about 65 million years ago (Hourdez et al. in prep.). All the vent-endemic subfamilies form a monophyletic group that rapidly radiated after this initial colonization event. All of these species express Hbs (single- or tetra-domain), suggesting that this trait was already present in the common vent ancestor (Hourdez et al. in prep.). We sequenced both the single- and tetra-domain (hereafter referred to SD, and TD, respectively) globin genes in *B. symmytilida* and *B. seepensis* to examine their exon/intron structure and to understand the origin and evolution of the tetra-domain globin.

Material and methods

Animal collection

Branchipolynoe symmytilida specimens were collected from the 9˚50' N locality on the East Pacific Rise in 2001 (9º46' N, 104˚21' W, 2515 m depth), and *B. seepensis* were collected from the Lucky Strike site in 2001 (Mid-Atlantic Ridge, 37º18' N, 32º16' W, 1700 m depth). The worms were removed from the pallial cavity of their host mussels. The specimens were identified and immediately frozen in liquid nitrogen on board, transported back to the laboratory and stored at -80ºC until analyzed.

Nucleic acids extraction, and cDNA synthesis

DNA, RNA extractions, and cDNA synthesis were performed with standard procedures that are described in Supplementary materials.

Globin cDNA and gene sequencing for *Branchipolynoe symmytilida*

Protein microsequencing and PCR primer design

The two coelomic hemoglobins (HbC1 and HbC2) from *Branchipolynoe symmytilida* were purified as described earlier (Hourdez et al. 1999a). We used Edman degradation for microsequencing and the released amino acids were identified by HPLC (High Performance Liquid Chromatography). The N-terminus in HbC1 was blocked but HbC2 yielded the microsequence N-terminus VSAAQKAAIK. Based on this microsequence, degenerate primers were designed to amplify HbC2 by PCR (Table S1, Supplementary material).

Initial globin amplification and sequencing

Primers BSY_E1D1_C2F and an anchored oligo(dT) (Table S1) were used on the synthesized cDNA, to amplify part of the tetra-domain globin gene. Detailed PCR conditions and cloning procedures are given in Supplementary material. This initial PCR reaction produced two insert size classes: 460-600 bp inserts that include one domain (400 bp) with different sized 3'UTRs, and 900 bp inserts corresponding to two domains (800 bp) with different sized 3'UTRs (See Results, Fig. S1). Two sequences representing the two size classes of inserts were chosen because they were the most abundant among the clones, designated BSY3 and BSY7. These sequenced clones were used to design specific primers to amplify the coding sequence and introns of the corresponding genes (single-domain and tetra-domain) (Table S1).

Bridging oligonucleotide rapid amplification of cDNA ends (BO-RACEing) for the single-domain globin

The 5'UTR for sequence BSY3 was amplified by BO-RACEing (Shi et al. 2002). Specific primers and conditions are given in Supplementary material. This approach did not yield any amplification for sequence BSY7 and chromosome walking was used instead to obtain the missing portion of the sequence.

Chromosome walking on genomic DNA

Sequencing of globin BSY7 from *B symmytilida* was completed by directional genome walking using PCR (Mishra et al., 2002). The detailed procedure is given in the Supplementary material. Directional genome walking was also used to obtain the promoter regions of both the TD and SD globins, in *B. symmytilida* and in *B. seepensis*.

Northern blot

A standard northern blot protocol (Sambrook et al. 1989) was used for *Branchipolynoe symmytilida*. The detailed procedure is given in the Supplementary material.

Globin amplification and sequencing in *Branchipolynoe seepensis*

The amplification of the SD and TD globin genes (cDNA and gDNA) from *B. seepensis* took advantage of the primers designed for the same genes in *B. symmytilida* and of an EST library from which 2500 clones were sequenced (Hourdez and Tanguy, unpub. data).

Amplification and sequencing followed a standard procedure detailed in Supplementary materials. Chromosome directional walking (see above) was used to sequence the 5'UTR and the promoter region of the globin genes.

Phylogenetic Analyses

The sequences were assembled and nucleotide positions with conflicts were resolved based on their chromatograms with CodonCode Aligner© 2.0.6 (http://www.codoncode.com/aligner/index.htm). Multiple nucleotide and amino acid sequence alignments were performed by using ClustalX 2.0.10 (Larkin et al. 2007) and, when necessary, manually optimized by using the sequence aligner editor Se-Al 2.0a11 Carbon (http://tree.bio.ed.ac.uk/software/seal/). During optimization, we minimized the number of indels and nucleotide alignment was constrained by the amino acid sequences alignment.

Before all analyses, we verified that there was no saturation of the signal by plotting K2P pairwise differences using all nucleotide sites against K2P pairwise differences estimated only on the first two codon positions (data not shown).

jModelTest was used to choose a model of nucleotide substitution for use in phylogenetic analysis that would best fit our data (Guindon and Gascuel 2003, Posada 2008). Neighbor-joining (Saitou and Nei 1987) trees were constructed using MEGA4 (Tamura et al. 2007), maximum-likelihood trees were computed using Phylip (http://evolution.genetics.washington.edu/phylip.html) and Phyml Online (http://atgc.lirmm.fr/phyml/, Guindon and Gascuel 2003, Guindon et al. 2005). Finally, a Bayesian analysis was performed with Mr. Bayes (Ronquist and Huelsenbeck 2003, Huelsenbeck and Ronquist 2001). PAML 3.14 (http://abacus.gene.ucl.ac.uk/software/paml.html, Yang 1997) was later used to test different tree topologies issued from the different phylogenetic analyses, by maximum likelihood relative ratio tests, and the codon substitution model derived from Goldman & Yang (1994). The tree topologies were rooted only when a molecular clock hypothesis was being tested, and under every test condition the single-domains were forced to be monophyletic. To test for molecular clock and possible differences in selective pressures between paralogous domains of the tetra-domain globin, we calculated pairwise distances between species for each domain, using the K2P model for nucleotides and the Dayhoff matrix for amino acids (MEGA4).

Results

cDNA cloning and sequencing

Amplification of *B. symmytilida* cDNA with degenerate primers and oligo(dT) yielded six distinct sequences that can be separated into two types (Fig. S1). The first type has a 400 bp coding region, corresponding to a globin domain, followed by a 3'UTR sequence that differed in length (sequences BSY1 and BSY3). The second type has a 800 bp coding region, corresponding to two globin domains (sequences BSY2, BSY4, BSY5, and BSY7), followed by a 3'UTR sequence also presenting differences in length. These two globin domains will be hereafter referred to as D3 and D4 (from 5' to 3'), in reference to the *Branchipolynoe* hemoglobin model composed of 4 domains (Hourdez et al. 1999a) (the same nomenclature will be used for the upstream domains: D1 and D2). The 3'UTR for sequences BSY2, BSY4, and BSY7 are nearly identical and a few differences appeared in their coding regions. The three remaining 3'UTR sequences (corresponding to sequences BSY1, BSY3, and BSY5) are clearly different. A Northern blot on *B. symmytilida* cDNA samples probed with a portion of sequence BSY3 and BSY7 (the most common sequences among the clones) revealed that these correspond to different transcript sizes. The BSY3 probe revealed a single band of ~610 bp in length, the size expected for a full-length mRNA coding for a single-domain globin (Fig. S2A). The BSY7 probe revealed a single band of ~1980 bp in length, consistent with the expected size for full-length mRNA coding for a tetra-domain globin (Fig. S2B).

The complete BSY3 cDNA sequence encodes a SD globin (coding sequence of 417 bp, including the stop codon), with a 67-base 5'UTR (Fig. S3), and a 58-base 3'UTR. After removal of the initial methionine, the deduced amino acid sequence (137 codons) would produce a protein with a molecular weight (MW) of 13826.8 Da.

The complete BSY7 cDNA sequence has a 79-base 5'UTR, a 1674 bp coding sequence (including the stop codon), and a 102-base 3'UTR (polyA tail not included). The mature protein encoded corresponds to a 4-domain globin (552 codons) (Fig. S4), with a MW of 57691.9 Da, a mass that closely matches the value found for the native HbC2 subunit (Hourdez et al, 1999a).

The cDNA of *B. seepensis* for both the SD and TD globins have the same length and number of codons as the homologous sequences from *B. symmytilida* (Figs. S3 and S4).

The corresponding proteins have a MW of 13798.7 and 57742.9 Da for the SD and TD globins, respectively.

Protein primary structure

The protein sequences for both *Branchipolynoe* species were aligned with other intracellular and extracellular annelid globins, two nematode (*Ascaris suum*) globin sequences, and the myoglobin from the sperm-whale *Physeter catodon* (Fig. 1). Only two positions were invariant: the proximal histidine (F8H, i.e. helix F, position 8 with the *Physeter* myoglobin sequence as a reference), and phenylalanine on the corner between helices C and D (CD1F). A tryptophan in helix A (A12W) was conserved in all but the nematode sequences. For all *B. symmytilida* and *B. seepensis* sequences, the distal histidine is replaced by a glutamine (E7Q), a feature also observed in *Ascaris*. Similarly, the B-helix tyrosine (B10Y), essential in modulating the oxygen affinity in *Ascaris* (De Baere et al. 1994), is also found in all *Branchipolynoe* spp. sequences. A tryptophan (H8W) is present in all sequences except in *P. catodon*. The two conserved cysteines in extracellular globins that form the characteristic intrachain disulfide bridge (A2C and H11C) are not found in any of the *Branchipolynoe* sequences. The TD sequences also have a cysteine (E18C) 11 residues after E7Q, a feature that is found in some globins of the tubeworms *Riftia pachyptila* and *Lamellibrachia* sp. (from hydrothermal vents and cold seeps, respectively). For all *Branchipolynoe* sequences, there is no pre A helix, signal peptide, or linker sequences.

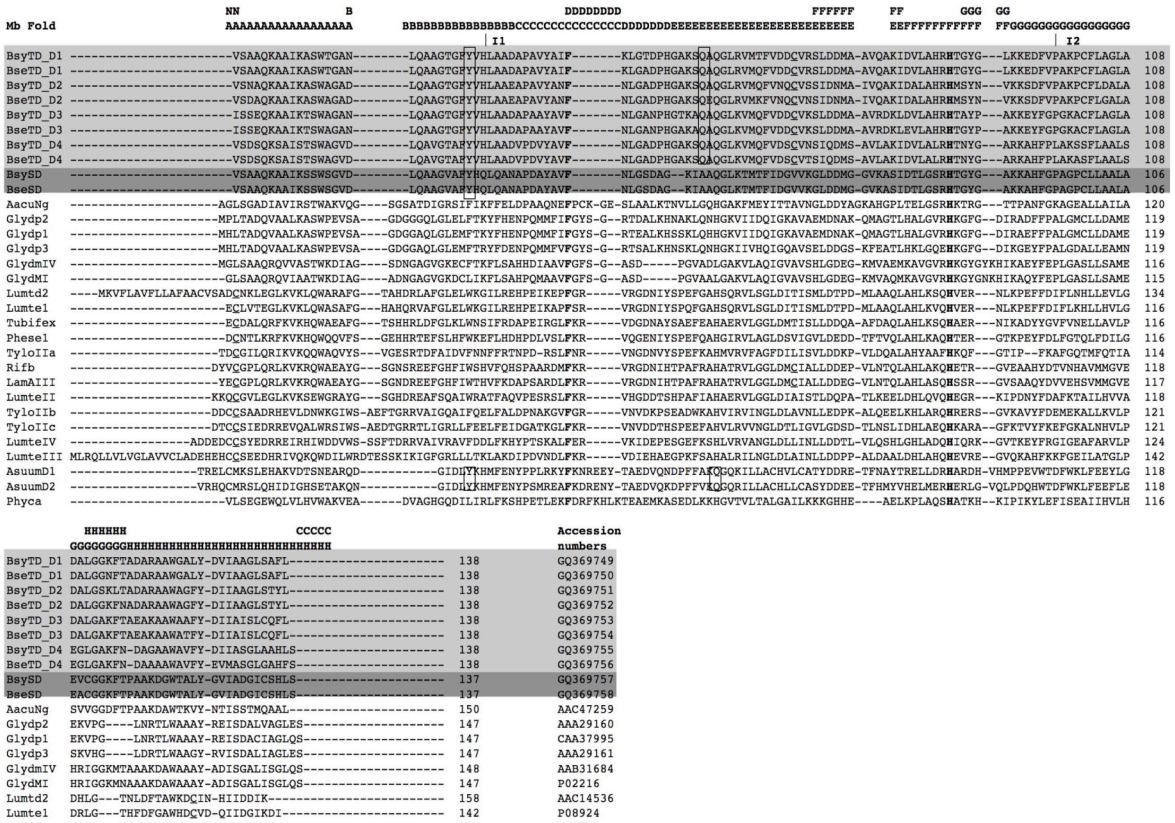


Figure 1. Globin sequences from annelids, nematodes and a vertebrate. *Branchipolynoe* globin sequences are shaded, TD globins shaded in light grey and SD globins in dark grey. Conserved residues shown in bold (CD1F and F8H), heme pocket residues that may explain high O_2 affinity in *Branchipolynoe* are boxed. Cysteines forming an intrachain disulfide bridge in typical extracellular annelid globins (A2C and H11C) underlined. Cysteine E18 underlined in the TD globins from *Branchipolynoe* spp., and for *Riftia* and *Lamellibrachia*. Intron (I1 and I2) conserved positions shown above the sequences. Bsy: *B. symmytilida*; Bse: *B. seepensis*; TD: tetra-domain; SD: single-domain; D1-D4: domains 1 through 4; AacuNg: *Aphrodite aculeata* neuroglobin; Gly: *Glycera* sp.; Lumt: *Lumbricus terrestris*; Tubifex: *Tubifex tubifex*; Phese: *Pheretima seiboldi*; Tylo: *Tylorhynchus heterochaetus*; Rifb: *Riftia pachyptila*; Lam: *Lamellibrachia* sp.; Asuum: *Ascaris suum*; Phyca: *Physeter catodon*.

Origin of the globin gene and relationships with other annelid globins

For both *Branchipolynoe* species, polymorphism was present in the different amplified sequences, therefore, a consensus nucleotide sequence was produced based on the majority of clones obtained for each species. These consensus sequences were translated and used in all phylogenetic analyses.

The unrooted phylogenetic tree clearly separates the globins that are typical extracellular HBL-Hbs from all intracellular globins (Fig. 2). The SD globin and the four domains of the TD globin from both *Branchipolynoe* species clearly group with the intracellular annelid globins, indicating that these extracellular globins have a distinct origin from all other annelid extracellular globins.

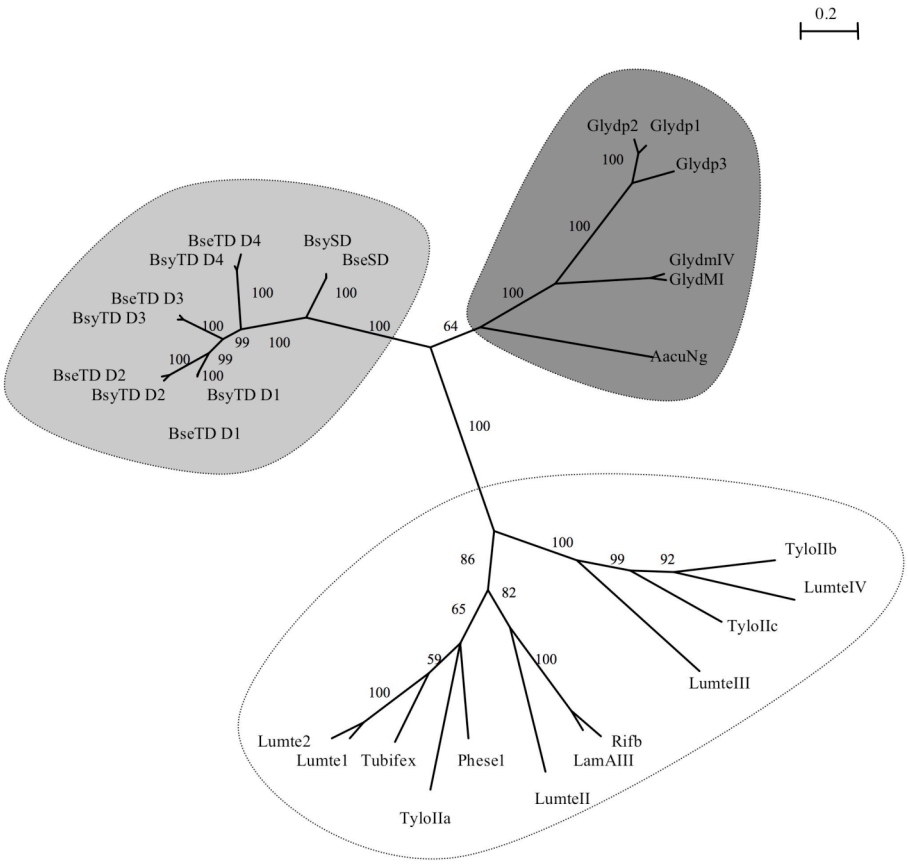


Figure 2. Bayesian phylogenetic tree based on annelid globins corresponding to the alignment in Fig. 1. White bubble: extracellular globins; light-grey bubble: *Branchipolynoe* globins (single-domain: SD and tetra-domain: TD); dark-grey bubble: annelid intracellular globins. Posterior probability values are indicated above the branches. See Fig. 1 for abbreviations and accession numbers.

Evolution of the tetra-domain structure

A phylogeny of the various *Branchipolynoe* globin sequences was determined using the SD as an outgroup because of its more ancestral status compared to the tetra-domain sequences (Fig. 2). Several nucleotide substitution models were used, and when using Mr. Bayes the codon model was also chosen. These analyses yielded different tree topologies depending on the phylogenetic method used and the selected nucleotide substitution model (Fig. S5).

The phylogenetic tree based on amino-acid sequences suggests that D4 is the most ancestral of the domains followed by D3, D2, and D1 in the TD gene. Although this phylogeny is well supported, it may not reflect the history of domain duplications as these can be going through periods of relaxed selective pressure and/or positive selection. Analysis based on nucleotides provided a different result, suggesting that D1 (or D3) is the most ancestral domain. The best resolved trees correspond to 4 equiprobable topologies (Fig. S5). In each tree topology, the globin domains of both species form well-supported pairs, indicating that duplications giving birth to the TD occurred prior to the radiation between the two *Branchipolynoe* species. The deeper branches of the nucleotide-based topologies, however, were not well supported (bootstrap values usually no higher than 65%, and at best 73%), suggesting that duplications occurred very close to each other over time, or that variable selective pressure over the various branches reduced the phylogenetic signal.

The different topologies were tested against each other using maximum likelihood relative ratio tests in CodeML (Yang 1997). Pairwise comparison of these 4 main topologies revealed topology (c) was the most probable, regardless of the d_N/d_S ratio model selected. The pairwise differences between the different TD domains and the SD indicated that D1 possesses the fewest differences with SD, followed by D3 (Table 1). In addition, D1 also displayed several motifs of adjacent amino acids that were nearly identical with the SD globin, reinforcing the hypothesis that D1 is more closely related to the ancestral state. D1 and D3 were also the domains with the smallest divergence between species for either nucleotide or amino acid sequences, suggesting that these domains are more constrained by purifying selection. To verify this hypothesis, we used BaseML (PAML) to test for a molecular clock using the theoretical topology D1D3 vs D2D4. The results indicate that the molecular clock is rejected in the evolution of the TD ($L_{clock=0}$ = -2171.78932, $L_{clock=1}$ = -2190.984727; RRT = 2*($L_{clock=0}$ - $L_{clock=1}$) = 38.390814 >> 10.83 (p

= 0.001)). This supports the idea that D1 and D3 may be under stronger selective constraints and evolving slower than the other domains.

Table 1. Pairwise differences between orthologous domains and the SD. Nucleotide differences were calculated by the Kimura-2 parameter model and amino acid differences based on the Dayhoff index matrix. BSY: *Branchipolynoe symmytilida*, BSE: *B. seepensis*, D: domain, SD: single-domain. Shaded values represent the lowest values between the SD and all the other domains.

BSY/BSE	Nucleotides (K2P)	Amino Acids (Dayhoff matrix)
D1/D1	0.030	0.022
D2/D2	0.035	0.058
D3/D3	0.025	0.022
D4/D4	0.065	0.058
SD/SD	0.017	0.007
SD/D1	0.458	0.592
SD/D2	0.506	0.725
SD/D3	0.485	0.653
SD/D4	0.564	0.689
D1/SD	0.462	0.549
D2/SD	0.511	0.713
D3/SD	0.480	0.625
D4/SD	0.554	0.657

Promoters and 5'UTRs

A portion of the promoter region was sequenced for the SD globin from *B. symmytilida*, and for the TD globin for both species. For both genes, the TATA box is located 32 bp upstream of the transcription start in *B. symmytilida*. For the SD sequence, only one base was different for the portion of 5'UTR for which there are data for both species (48 nucleotides, i.e. 97.9% identity) (Fig. S3). The promoter for the TD globin, as well as the 5'UTR, from both species, also exhibit a high identity level (97.5% for the 5'UTR) (Fig. S4). A search for transcription regulatory signal binding sites in *B. symmytilida* revealed a site for CP2, a factor that is known to enhance the transcription of globin genes in erythroid cells (Chae and Kim 2003). This element was however not found in *B. seepensis*. Other universal transcription factors, such as Oct-1, were identified in the promoter of both species (data not shown). Although these are extracellular Hbs no signal sequence for secretion was found. In *B. seepensis*, we sequenced a 700 bp fragment of the promoter. This region contains motifs for the binding of two transcription factors relevant for this gene: one site for hypoxia inducible factor 1 (HIF-1); and two sites for GATA-1, which

plays an important role in erythroid development (De Maria et al. 1999) and has been reported to enhance erythropoeisis in response to tissue hypoxia (Krantz 1991, Zon et al. 1991). The sequenced promoter region for *B. symmytilida* does not cover the region where these transcription factors' binding sites are located in *B. seepensis*.

Introns and gene structure

In both the SD and TD globin genomic sequences from both species, the typical 3 exons/2 introns globin gene structure was present, and this motif is repeated four times in the TD gene (Fig. 3). For all the genes, the introns are located at positions B12.2 and G7.0, corresponding to the typical globin introns positions. In the SD globin gene, the second intron exhibits a very high level of identity between the two species (94.4%). Although this is also true for the first 160 bp in intron 1, the identity then drops sharply because *B. seepensis* possesses a large (440 bp) insertion (Table S2). In the TD globin genomic sequence, each intron was usually 300 bp long, with one exception at 700 bp again for the first intron of D1 (Table S2). Interestingly, neither introns nor linker segments separate the four globin domains (Fig. S4). Pairwise intron sequence comparisons between the two species revealed a very high level of conservation (on average 90% of identity) of all but the second intron of D4. The major differences in the orthologous introns correspond to indels of varying size, from 3 to 41 bp. Comparisons of intron sequences among domains failed to reveal significant sequence similarity. In intron 2 from D1 and D2 from *B. seepensis*, there is a regulatory signal for GATA-1 and HIF-1, respectively.

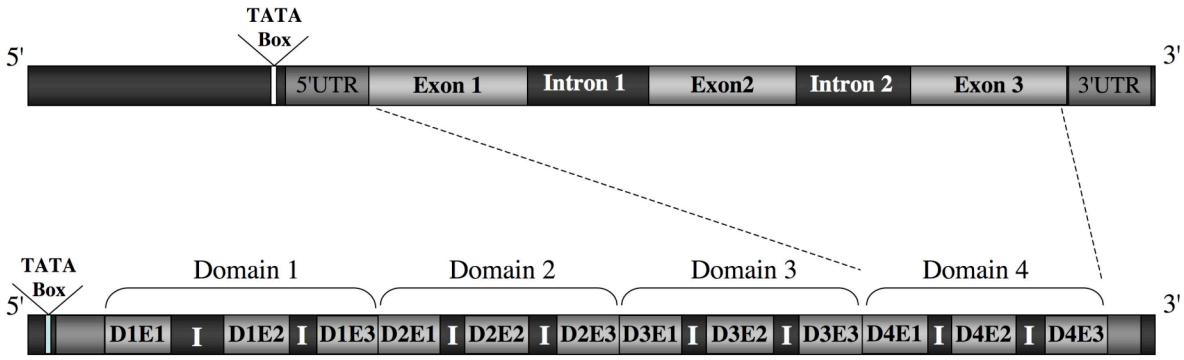


Figure 3. Gene structure from the single- (top) and tetra-domain (bottom) globins from *Branchipolynoe* spp. Promoter region: dark grey; UTR: medium grey; exons: light grey; introns: dark grey with "I" or "Intron". Gene sizes and components are not to scale.

Discussion

The coding sequence for the unique TD and SD globins genes, as well as the position and size of the introns have been determined to understand the origin and evolution of these respiratory pigments. The TD globin with a multi-domain gene structure represents the emergence of a new type of hemoglobin in annelids (Terwilliger 1992).

Clarification of the protein structure

Hourdez et al. (1999a) showed that the *Branchipolynoe* extracellular TD globins form dimers (HbC2) or trimers (HbC1) and our results suggest that we most likely amplified the former because the estimated MW most closely matches the one for HbC2 subunits. Large polypeptidic assemblages are not unusual in invertebrate respiratory pigments (Terwilliger, 1998). In all cases, the formation of large polypeptide complexes has been attributed to the necessity to avoid their excretion (Weber and Vinogradov, 2001). In *Daphnia* and *Artemia*, the didomain and nine-domain Hbs, respectively, can form homo- or hetero-dimers (Dewilde et al. 1999, Mansfield et al. 1980). In the nematode *A. suum* the didomain Hbs assemble into octamers of 328 kDa, through a C-terminus extension that forms a β-strand and each strand binds to each other with H-bonds in a zipper-like structure (Darawshe et al. 1987, De Baere et al. 1992). In *Branchipolynoe*, there is no such extension and there are

no interchain disulfide bridges to assemble into trimers or dimers, and only inter-subunit non-covalent bonds are involved.

Based on subtilisin partial digestion patterns and heme content determination, Hourdez et al. (1999a) suggested that each Hb subunit was composed of 4 globin domains (with one likely truncated) each with one heme group. The sequences we obtained clearly show that there are indeed 4 domains but each of these domains is full-length. This may indicate that the interdomain regions are not well exposed in the native Hb, and susceptible to peptidic cleavage by a wide-spectrum peptidase.

Interestingly, the globins from *Branchipolynoe* do not possess the two cysteine residues necessary to form the conserved disulfide bridge conserved in all other extracellular annelid globins (Bailly et al., 2007). This indicates that this character is not necessary for the globins to function outside of cells, although its conservation in the lineage of the HBL-Hb globins suggests that purifying selection is constraining these amino acid positions.

Structure/function relationship

The amino acid alignment revealed the presence of a glutamine instead of a distal histidine (E7H). Although this amino acid is usually conserved in vertebrates, some invertebrate species use other amino acids. In the nematode *A. suum* Hb, known to have one of the highest oxygen affinities among invertebrates (see Weber and Vinogradov 2001 for a review), the position is also occupied by a glutamine (E7Q). This high affinity can be explained in part by the interaction of the E7Q with a tyrosine in position B10Y (De Baere et al. 1994, Yang et al. 1995). Interestingly, *Branchipolynoe* also has a tyrosine in position B10. This could explain the high affinity for oxygen that was found in *B. symmytilida* Hbs (Hourdez et al. 1999b).

In the TD globins, but not in the SD ones, there is a cysteine located 11 residues after the E7Q (E18C). This cysteine is known to be under strong selection in some annelids living in reduced habitats (Bailly et al. 2003), suggesting a key function for E18C in these habitats. It was hypothesized to be a key residue for reversible sulfide binding, a characteristic that is essential for vestimentiferan tubeworms that live symbiotically with sulfide-oxidizing bacteria (Zal et al. 1998). It was later shown that the E18 cysteine was not likely the binding site for transporting sulfide in *Riftia pachyptila* and that, at least for the 400-kDa Hb, zinc atoms were involved (Flores et al. 2005, Flores and Hourdez 2006).

The SD globin from *Branchipolynoe* is devoid of this specific residue, indicating that it most likely represents an adaptive convergence to life in a sulfide-rich habitat in the TDs. Its presence in the heme pocket may have a protective role for the heme group that usually reacts with sulfide to irreversibly form sulfhemoglobin, an altered molecule incapable of binding oxygen.

Secretion into the coelomic fluid

Although *Branchipolynoe* SD and TD Hbs are extracellular (secreted in the coelomic cavity), their genes do not possess a signal peptide for secretion whereas all other annelid extracellular globins have such a signal (Riggs 1991; Bailly et al. 2007). The absence of a signal peptide in all *Branchipolynoe* globins could be due to the fact that we did not obtain the whole sequence and that the upstream sequence (promoter) actually corresponds to an intron. This however is unlikely as we did locate the TATA box and the sequence corresponding to the 5'UTR, and a 700 bp stretch upstream did not reveal any other open reading frame in *B. seepensis* TD globin gene. This may indicate that the secretion of *Branchipolynoe* Hbs is holocrine (by rupture of the cell membrane, and the release of all the components of the cytoplasm). This is supported by the fact that, in at least one other vent species (*Lepidonotopodium piscesae*), the SD globin expressed in the muscles (myoglobin) has exactly the same molecular weight as the one found in the coelomic cavity (Hourdez, unpub. data).

Origin and evolution of the tetra-domain globin

Phylogenetic relationships among annelid globins indicate that extracellular globins have evolved independently from the intracellular (circulating or non-circulating) ones, and diverged about 570 millions years ago (Goodman et al. 1988), an origin distinct from extracellular globins found in other phyla (Bailly et al. 2007). *Branchipolynoe* extracellular globin sequences have a distinct history, not only from other phyla's extracellular globins (including multidomain globins), but also from the typical extracellular annelid globins. It appears that they arose from an intracellular annelid globin more recently than the original split between intra and extracellular globins in annelids. Despite this different evolutionary history, the SD and TD globins conserved the basic globin gene structure of 3 exons and 2 introns, including intron positions. This arrangement differs from that of *Aphrodite aculeata* myoglobin (Dewilde et al. 1996) in which the first intron is missing. This species

belongs to the scale-worm family Aphroditidae, a close related family to the Polynoidae. This indicates that the ancestor to all scale-worms most likely did possess this first (B12.2) intron, and that it was later lost in the lineage leading to *Aphrodite aculeata*. Some of the introns showed a high level of conservation between the two species, but in both the SD and the TD globin genes there is a strongly divergent intron (intron 1 in the SD, intron 2 in D4 for the TD). This may indicate strong selective constraint on the intron sequence for a possible role related to a regulatory function.

The phylogenetic analyses yielded trees with only limited support for the internal branches. This could be indicative of a rapid succession of the duplication events to lead to this structure, and/or to different selective pressure affecting the various domains. We found that D1 and D3 are probably under stronger purifying selection than D2 and D4. This would then yield to different branch lengths, and the absence of a molecular clock. Although the phylogenetic trees do not allow us to clearly decide on a likely duplication scenario, other observations (pairwise distances, conservation of some amino acid motifs), suggest that the TD structure is likely the result of two tandem duplications, the first one giving rise to D1D2, and the second one affecting these two domains together to generate the whole TD. Domain D1 is likely the most ancestral domain, as (1) in both species, intron 1 from D1 has the same size as in the SD globin, (2) pairwise differences between the SD and D1 are smaller than all the other domains, and (3) there are some conserved amino acids motifs. This duplication scenario shows that duplication moved in a 5' to 3' direction where downstream duplicates were likely joined by intron losses. The alternative scenario where domain 4 is ancestral, and duplications proceeded in a 3' to 5' direction seems less likely because with each subsequent domain duplication the 5' regulatory region may have had to be reconstructed.

Duplication events

The mechanism responsible for these duplication events remains unclear for the moment. Interestingly, the TD globin gene from these scale-worms does not have any bridge introns. This contrasts with the didomain globins from the bivalve *Barbatia reeveana* (Naito et al. 1991), the water-flea *Daphnia pulex* (Dewilde et al. 1999), and some nematodes (*Ascaris suum* and *Pseudoterranova decipiens*; Dixon et al. 1992, Sherman et al. 1992). Similarly, in mollusk hemocyanins, the 7 or 8 domains are separated by bridge introns that are thought to be the remnants of the duplication of the gene (Lieb and Todt

2008). Amino acid linker sequences are also missing between the domains of the TD globins from *Branchipolynoe*. This kind of structure is known in other invertebrate multidomain proteins such as the nine-domain Hb from *Artemia* (Maning et al. 1990, Trotman et al. 1994), and the didomain Hb of the bivalve *B. reeveana* (Naito et al. 1991). It is however not found in the nematodes two-domain Hb (De Baere et al. 1992, Sherman et al. 1992). These linker sequences are thought to be the remnants of an inter-domain intron that lost its splicing signals and was eventually integrated into the coding sequence. All *Branchipolynoe* TD domains are full-length, indicating that there likely was inter-domain introns that were secondarily integrally lost, as suggested by Naito et al. (1991) for *B. reeveana*, Dewilde et al. (1999) for *D. pulex* Hb, and Dixon et al. (1992) for *P. decepiens*. The alternative, less likely, hypothesis would require two end-to-end tandem duplications that preserved (1) the whole length of the domains, and (2) the reading frame. In another water-flea, *Moina macrocopa*, the two-domain globin also lacks the bridge intron (Kato et al. 2001), and the authors suggest unequal crossing-over as the mechanism for the duplication of the gene, with a subsequent loss of the bridge intron when the Cladocera (water-fleas) families diverged. Finally, the TD structure could also correspond to the result of unequal crossing-overs between two copies of an initial didomain structure but this would yield chimeric domains and our search for such possible chimeras proved unsuccessful (data not shown).

Concluding remarks

Our work shows that *Branchipolynoe* extracellular Hbs have a different origin and evolutionary history from other annelid extracellular Hbs. It appears that an ancestral intracellular myoglobin was duplicated and evolved to form an extracellular globin by evolutive tinkering. The evolution of these Hbs, constitutively expressed at high levels, is most likely the result of strong selective pressure due to the chronic hypoxia that characterizes hydrothermal vents (Hourdez and Lallier, 2007). The multidomain structure allows higher concentration of oxygen binding sites (and thereby transport/storage capacity), without increasing the colloid osmotic pressure of the coelomic fluid.

Acknowledgments

The authors would like to thank the crews of the ships and submersibles, as well as the chief scientists, of the cruises ATOS 2001 (project funded by Ifremer and INSU), and EPR

2001 (project funded by a NSF grant to C.R. Fisher (NSF OCE-0002729)). We would also like to thank Isabelle Boutet-Tanguy and Arnaud Tanguy for technical advice in lab. This work is part of the project HYPOXEVO (Région Bretagne), and was supported by the ESTeam research Marie Curie grant under the 6th framework program from the European Commission.

Literature cited

Bailly X, Leroy R, Carney S, Collin O, Zal F, Toulmond A, Jollivet D. 2003. The loss of the hemoglobin H2S-binding function in annelids from sulfide-free habitats reveals molecular adaptation driven by Darwinian positive selection. P Natl Acad Sci USA. 100:5885-5890.

Bailly X, Chabasse C, Hourdez S, Dewilde S, Moens L, and Zal F. 2007. Globin gene family evolution and functional diversification in annelids. FEBS J. 274: 2641-2652.

Chae JH, Kim CG. 2003. CP2 Binding to the Promoter is Essential for the Enhanced Transcription of Globin Genes in Erythroid Cells. Mol Cells. 15: 40-47.

Church GM, Gilbert W. 1984. Genomic sequencing. P Natl Acad Sci USA 81:1991-1995.

Darawshe S, Tsafadyah Y, Daniel E. 1987. Quaternary structure of erythrocruorin from the nematode *Ascaris suum*. Biochem J. 242: 689-694.

De Baere I, Liu L, Moens L, Van Beeumen J, Gielens C, Richelle J, Trotman C, Finch J, Gerstein M, Perutz M. 1992. Polar zipper sequence in the high-affinity hemoglobin of *Ascaris suum*: Amino acid sequence and structural interpretation. P Natl Acad Sci USA. 89:4638-4642.

De Baere I, Perutz MF, Kiger L, Marden MC, Poyart C. 1994. Formation of two hydrogen bonds from the globin to the heme-linked oxygen molecule in *Ascaris* hemoglobin. P Natl Acad Sci USA. 91:1594-1597.

De Maria R, Zeuner A, Eramo A, Domenichelli C, Bonci D, Grignani F, Srinivasula SM, Alnemri ES, Testa U, Peschle C. 1999. Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1. Nature. 401:489-493.

Dewilde S, Blaxter M, Van Hauwaert M-L, Vanfleteren J, Esmans EL, Marden M, Griffon N, Moens L. 1996. Globin and Globin Structure of the Nerve Myoglobin of *Aphrodite aculeata*. J Biol Chem. 271:19865-19870.

Dewilde S, Van Hauwaert M-L, Peeters K, Vanfleteren J, Moens L. 1999. *Daphnia pulex* didomain hemoglobin: structure and evolution of polymeric hemoglobins and their coding genes. Mol Biol Evol. 16:1208-1218.

Dixon B, Walker B, Kimmins W, Pohajdak B. 1992. A nematode gene contains an intron previously thought to be unique to plants. J Mol Evol. 35:131-136.

Flores JF, Fisher CR, Carney SL, Green BN, Freytag JK, Schaeffer SW, Royer Jr WE. 2005. Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. P Natl Acad Sci USA. 102:2713-2718.

Flores JF, Hourdez S. 2006. The zinc-mediated sulfide-binding mechanism of hydrothermal vent tubeworm 400-kDa hemoglobin. Cah Biol Mar. 47: 371-377.

Goldman N, Yang Z.1994 A codon based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol. 11:725–736

Goodman M, Pedwaydon J, Czelusniak J, Suzuki T, Gotoh T, Moens L, Shishikura F, Walz D, Vinogradov SN. 1988. An evolutionary tree for invertebrate globin sequences. J Mol Evol. 27:236-249.

Grinich N P, Terwilliger RC. 1980. The quaternary structure of an unusual high-molecular-weight intracellular hemoglobin from the bivalve mollusc Barbatia reeveana. Biochem J. 189: 1-8.

Guindon S, Gascuel O. 2003. A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst Biol 52: 696-704.

Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML Online-a web server for fast maximum likelihood-based phylogenetic inference. Nucl. Acids Res. 33:W557-W559.

Hourdez S, Lallier FH, Green BN, Toulmond A. 1999a. Hemoglobins from deep-sea hydrothermal vent scale-worms of the genus *Branchipolynoe*: A new type of quaternary structure. Proteins. 34:427-434.

Hourdez S, Lallier FH, Martin-Jézéquel V, Weber RE, Toulmond A. 1999b. Characterization and functional properties of the extracellular coelomic hemoglobins from the deep-sea, hydrothermal vent scale-worm *Branchipolynoe symmytilida*. Proteins. 34:435-442.

Hourdez S, Lallier FH. 2007. Adaptations to hypoxia in hydrothermal vent and cold-seep invertebrates. Rev Env Sci Biotech. 6: 143-159.

Hourdez S, Weber RE. 2005. Molecular and functional adaptations in deep-sea hemoglobins. J Inorg Biochem. 99(1): 130-141

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17: 754-755.

Kato K, Tokishita S-i, Mandokoro Y, Kimura S, Ohta T, Kobayashi M, Yamagata H. 2001. Two-domain hemoglobin gene of the water flea *Moina macrocopa*: duplication in the ancestral Cladocera, diversification, and loss of a bridge intron. Gene. 273:41-

50.

Krantz SB. 1991. Erythropoietin. Blood. 77(3): 419-434.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947-2948.

Lieb B, Todt C. 2008. Hemocyanin in mollusks - A molecular survey and new data on hemocyanin genes in *Solenogastres* and *Caudofoveata*. Mol Phylogenet Evol. 49:382-385.

Manning A, Trotman C, Tate W. 1990. Evolution of a polymeric globin in the brine shrimp Artemia. Nature. 348: 653-656.

Mansfield BC, Krissansen GW, Smith MG, Tate WP. 1980. Isolation of intact larval haemoglobin from the brine shrimp *Artemia salina*. Prevention of degradation in vitro by proteases induced during larval development. Biochem Biophys Acta. 623:163-170.

Mishra RN, Singla-Pareek SL, Nair S, Sopory SK, Reddy MK. 2002. Directional Genome Walking Using PCR. BioTechniques. 33:830- 834.

Naito Y, Riggs C, Vandergon T, Riggs A. 1991. Origin of a "bridge" intron in the gene for two domain globin. P Natl Acad Sci USA. 88: 6672-6676.

Posada D. 2008. jModelTest: Phylogenetic Averaging. Mol Biol Evol. 25:1253-1256.

Riggs AF. 1991. Aspects of the Origin ad Evolution of Non-Vertebrate Hemoglobins. Am Zool. 31: 535-545.

Rodionov MA, Johnson MS. 1994. Residue-residue contact substitution probabilites derived from aligned three-dimensional structures and the identification of common folds. Protein Sci. 3: 2366-2377.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572-1574.

Saitou N, Nei M. 1987. The neighbor-joining method: A nem method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406-425.

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Nolan C. ed. Cold Spring Harbour :Cold Springs Harbor Press.

Sherman DR, Kloek AP, Krishnan BR, Guinn B. 1992. *Ascaris* hemoglobin gene: Plant-like structure reflects the ancestral globin gene. P Natl Acad Sci USA. 89:11696-11700.

Shi X, Karkut T, Chahmanhkah M, Alting-Mees M, Hemmingsen SM, Hegedus D. 2002.

5'-RACEing Across a Bridging Oligonucleotide. Biotechniques. 32:480-482.

Suzuki T, Arita T.1995. Two-Domain Hemoglobin from the Blood Clam, *Barbatia lima*. The cDNA-Derived Amino Acid Sequence. J Protein Chem. 14(7): 499-502.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24: 1596-1599.

Terwilliger NB. 1992. Molecular Structure of the Extracellular Heme Proteins Vol 13. In: Mangum CP editor. Advances in Comparative and Environmental Physiology. Blood and Tissue Oxygen Carriers, Berlin Heidelberg: Spriger-Verlag. 459p.

Terwilliger NB. 1998. Functional adaptations of oxygen-transport proteins. J Exp Biol. 201:1085-1098.

Trotman C, Manning A, Bray J, Jellie A., Moens L., Tate W. 1994. Interdomain Linkage in the Polymeric Hemoglobin Molecule of *Artemia*. J Mol Evol. 38:628-636.

Tunnicliffe V. 1991. The Biology of Hydrothermal Vents: Ecology and Evolution. Oceanogr Mar Biol. 29:319-407.

Weber RE, Vinogradov SN. 2001. Nonvertebrate hemoglobins: functions and molecular adaptations. Physiol Rev. 81:569-628.

Weber RE. 1978. Respiratory Pigments. Physiology of Annelids. P. J. Mill. London, Academic Press Inc.

Yang J, Kloek A., Goldberg D, Mathews F. 1995. The structure of *Ascaris* hemoglobin domain I at 2.2 Å resolution: molecular features of oxygen avidity. P Natl Acad Sci USA. 92:4224-4228.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput App BioSci. 13: 555-556.

Zal F, Leize E, Lallier FH, Toulmond A, Van Dorsselaer A, Childress J. 1998. S-Sulfohemoglobin and disulfide exchange: The mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins. P Natl Acad Sci USA. 95:8997-9002.

Zon LI, Youssoufian H, Mather C, Lodish HF, Orkin S. 1991. Activation of the erythropoietin receptor promoter by transcription factor GATA-1. P Natl Acad Sci USA. 88:10638-10641.

Supplementary material

Materials and methods

Nucleic acids extraction, and cDNA synthesis

For both species, total DNA was isolated from whole individuals, with a standard Phenol/Chloroform extraction following proteinase K digestion (Sambrook et al. 1989). Total RNA was extracted from the tissues using the TRI Reagent® protocol (Applied Biosystems) and mRNA was purified by affinity chromatography on oligo(dT) columns (mRNA purification kit, Pharmacia). cDNA were then synthesized by reverse transcription with MMLV- Reverse Transcriptase and oligo$(dT)_{18}$ or anchored oligo(dT) (Table S1).

Initial *B. symmytilida* globin amplification and sequencing

The PCR program for the initial amplification was as follows: one initial step at 95ºC (15'), to activate the Hot Start Taq polymerase (Qiagen), followed by 30 cycles (95ºC (1')/48ºC (45")/72ºC (2')), with a final incubation at 72˚C (10'), to ensure the terminal addition of an adenine. The PCR products were visualized on agarose gels with ethidium bromide under UV light. The PCR products were then cloned with the TOPO TA Cloning® kit (Invitrogen). Positive clones were sequenced with primers that flanked the cloned insert within the plasmid.

Bridging oligonucleotide rapid amplification of cDNA ends (BO-RACEing) for the single-domain globin

cDNA was produced using primer BSY3_RTP (Table S1), a phosphorylated reverse primer located in the 3'UTR portion of sequence BSY3. The reverse transcription mix contained a large excess of dGTP to ensure that the reverse transcriptase terminal transferase activity added a short string of the complementary base (cytidines). Primer BSY3_BO (Table S1) was then annealed, and the product was circularized by ligation. Finally, specific primers (BSY3_R3/BSY3_F3, nested with BSY3_R2/BSY3_F2) were used to amplify the unknown sequence, including the 5'UTR.

Northern-blot on *Branchipolynoe symmytilida* RNA

Branchipolynoe symmytilida RNA samples (20 µg of total RNA and, 1 µg and 3 µg of

mRNA) were denatured at 65ºC for 5 minutes in 1.6 volumes of denaturing charge buffer and loaded in a 1% agarose denaturing gel containing 17% formaldehyde and buffered with MEN (MOPS 20 mM pH 7.0, EDTA 1 mM, ammonium acetate 5 mM, final concentrations). The samples were electrophoresed at 60V for 4 hours, along with a Millenium molecular weight Marker (Ambion). The electrophoresis buffer was a MEN 1X, re-circulated between the anodic and cathodic buffer wells. RNAs were then transferred to a Hybond N membrane with a Vacuum Blotter in 20X SSC buffer (Standard Sodium 194 Citrate, 3 M sodium chloride, 300 mM trisodium citrate, pH 7.0) and bonded to the membrane by UV radiation. Specific probes (about 350 bp long) were produced from the cloned PCR products for sequences BSY3 and BSY7 from *B. symmytilida*. The probes were labelled with ^{32}P-labeled dCTP with the Megaprime DNA labelling system (Amersham), following the manufacturer's instructions. Hybridization was performed according to the Church and Gilbert method (1984). Briefly, the membrane was prehybridized in Church buffer (BSA 1%, EDTA 1 mM, NaHPO4 pH 7.2 0.5 M, and SDS 7%), the probe was then added at a final concentration of 10^6 disintegrations per minute and per milliliter. the pre-hybridization and hybridization temperatures were 42ºC. Ten-minute washings were also performed at 42˚C, first with 2X SSC, 0.5 % SDS, and then with 1X SSC, 0.1% SDS. The membranes were then sealed in bags and the resulting bands visualized on a STORM Phosphor-Fluor Imager after exposure of a radiation-sensitive screen for 4 hours.

Directional chromosome walking

Specific primers (Table S1) were designed in introns of the previously amplified genomic sequence of globin BSY7, and used in conjunction with a primer containing a random section in 3'. A nested PCR was then performed with a second specific primer in the intron and the non-random portion of the other primer (Mishra et al., 2002). The products were cloned in a TOPO TA Cloning® kit (Invitrogen). A series of clones of various sizes was sequenced and the overlap with the known sequence allowed us to determine which contained the sequence of interest. Each walking step covered 500-1000 bp, and a total of ~2600 bp were sequenced to cover the coding sequence, and the 5'UTR.

B. seepensis globin sequencing

The PCR reaction was performed for 30 cycles (94ºC (1')/50ºC (1')/72ºC (1')) with Uptitherm *Taq* polymerase (Interchim). Based on the primers that were used, another PCR program was used to improve the yield and the specificity: 10 cycles (94ºC (30")/52ºC (3')/72ºC (2')) + 30 cycles (94ºC (30")/52ºC (1')/72ºC (1')), and the $MgCl_2$ concentration

was increased to 2 mM. The PCR products were visualized on a 1.5% agarose gel containing ethidium bromide under UV light, and cloned with a TOPO TA Cloning kit (Invitrogen). Positive clones were sequenced, and the sequences used to produce specific primers for *B. seepensis* (Table S1).

Tables

Table S1. Sequences of the primers used to amplify the different parts of the single- (SD) and tetra-domain (TD) hemoglobin gene from *Branchipolynoe symmytilida* (Bsy) and *B. seepensis* (Bse).

Primers	5' Sequence 3'	Localization/Function
BSY3_RTP	P-GTT ACG GCT TGA TTT TAT TAC	Oligo for BO-RACEing on BsySD
BSY3_BO	GTA ATA AAA TCA AGC CGT AAC GGG	Bridging oligo for BO-RACEing on BsySD
BSY3_F2	CTC TAC GGA GTC ATC GCT G	Oligo in promoter sequence in BsySD (forward)
BSY3_F3	CGT CAC ACT GGG TAT GGT GC	Oligo in promoter sequence in BsySD (forward)
BSY3_R2	CCT GAA GGT CCA CTC CCG	Oligo in promoter sequence in BsySD (reverse)
BSY3_R3	CAT CTG GGG CAT TAG CCT GC	Oligo in promoter sequence in BsySD (reverse)
5UTR_TDF1	CCG CAC CAC CAG CAT CAT CCG C	Oligo in 5'UTR from BsyTD (forward)
5UTR_TDF2	GGG AAG AAC TAA TCA TCA CCA	Oligo in 5'UTR from BsyTD (forward)
5UTR_TDR1	TGG TGA TGA TTA GTT CTT CCC	Oligo in 5'UTR from BsyTD (reverse)
5UTR_TDR2	GAT GTG CTC ATG GGC TGT TGG	Oligo in 5'UTR from BsyTD (reverse)
BSY_E1D1_C2F	GCN GCN CAR AAR GCN GCN ATH	E1D1 from BsyTD HbC2 (forward)
BSY_E1D1F	GAT CTG CAA GCA GCT GGA ACA	Oligo in E1D1(in some allelic forms) (forward)
WBSY_I1D1F1	GCC AGA TGC ATG GAT ATG GCT	Walker oligo in I1D1 BsyTD (forward)
WBSY_I2D1R1	AAT GGA CTA ACG ATG GAA TAA GC	Walker oligo in I2D1 BsyTD (reverse)
WBSY_I1D1R4	ATG CAC ATA TGT AAT CAC AA	Walker oligo in I1D1 BsyTD (reverse)
WBSY_I1D1R5	TCA CAA TTG CTC CTG CCC TG	Walker oligo in I1D1 BsyTD (reverse)
WBSY_I1D1R6	ACA TTA ATC TGG AAG CCA TC	Walker oligo in I1D1 BsyTD (reverse)
WBSY_I1D2R1	TTA AGT GAG TCA CAT ATT ACA CC	Walker oligo in I1D2 BsyTD (reverse)
WBSY_I1D2R2N	ACT AGA CCT CTA CAG RCA GT	Walker oligo in I1D2 BsyTD (reverse)
WBSY_E1I1D2R3	GTC TTA CTG GAC ATA GAA TC	Walker oligo in E1&I1D2 BsyTD (reverse)
BSY_E3D2E1D3F	GAA AAY CAT GGT GTG TTT CAA G	Oligo in I2D2 from BsyTD (forward)

Primers	5' Sequence 3'	Localization/Function
BSY_E3D2E1D3R	ATA GCA AGA TGT CCA TGG CCC	Oligo in I1D3 from BsyTD (reverse)
BSY_E1D3BF	ATC AAG ACC TCA TGG GCA GG	Oligo in the beginning of E1D3 from BsyTD (forward)
BSY_E1D3BF	CCT CAT GGG CAG GAG CCA ACT TGC	Oligo in the beginning of E1D3 from BsyTD (forward)
BSY_E1D3BR	GTT GGC TCC TGC CCA TGA GGT C	Oligo in the beginning of E1D3 from BsyTD (reverse)
BSY_E3D3EF	GGT CTG GCA GAT GYT CTT GG	Oligo at the end of E3D3 from BsyTD (forward)
BSY_E3D3ER	CAG CTY TTG CTT CAG CGG	Oligo in E3D3 from BsyTD (reverse)
BSY_E1D4BF	CCT GGR CTG GAG TTG GAC CTC	Oligo in E1D4 from BsyTD (forward)
BSY_E1D4BR	CCA GGA CGT GGA GAT GGC GG	Oligo in E1D4 from BsyTD (reverse)
BSY_E3D4EF	CTA TGA TAT CAT CGC TTC TGG	Oligo in E3D4 from BsyTD (forward)
BSY_E3D4ER	AGA CAG CCC AKG CAG CNG CAG	Oligo in E3D4 from BsyTD (reverse)
BSE_E1D1_F1	TGG TTT GAT TGT GAT GGC	Oligo in E1D1 BseTD (forward)
BSE_E1D1_F2	GCT GGA ACA GGA TTC TAT GTC C	Oligo in E1D1 BseTD (forward)
BSE_E2D1_R2	GCC CAA CTT GAA GAY GGC ATA	Oligo in E2D1 BseTD (reverse)
BSE_E2D1_R1	GCA TCA CTC TCA GTC CCT G	Oligo in E2D1 BseTD (reverse)
BSE_I1D2_Rn	CCA CAG ACT ACT ACT AGA CCT CTA C	Oligo in I1D2 BseTD (reverse)
BSE_E2D2_R	CTC TGA GTC CCT GCT CTT GTG	Oligo in E2D2 BseTD (reverse)
BSE_E1D4_F1	GTC TCT GAC AGC CAG AAA TYC	Oligo in E1D4 BseTD (forward)
BSE_E1D4_F2	CTC CAG GCT GTT GGA ACT	Oligo in E1D4 BseTD (forward)
BSE_E2D4_R2	ACA CAT CAG GGA CAT CGG	Oligo in E2D4 BseTD (reverse)
BSE_E2D4_F1	CAG GGT CTC AAA GTG ATG C	Oligo in E2D4 BseTD (forward)
BSE_E2D4_R1	GCA TCA CTT TGA GAC CCT G	Oligo in E2D1 BseTD (reverse)
BSE_E2D4_F2	GCC AGG AAG GCA CAC TTC	Oligo in E2D4 BseTD (forward)
BSE_E3D4_R1	GCC AGG AAG GAA GCC TTT	Oligo in E3D4 BseTD (reverse)
Oligo $(dT)_{18}$	TTT TTT TTT TTT TTT TTT T	Upstream amplification from cDNA
Anchored OligodT	CTC CTC TCC TCT CCT C(T)17	Upstream amplification from cDNA

Table S2. Initial and terminal sequences (6 nucleotides), and total length for all introns from *Branchipolynoe* single- and tetra-domain globin genes (SD and TD respectively).

Sequence	**Beginning**	**End**	**Length (bp)**
Bsy-SD-Intron1	GTAAGT	CTATAG	305
Bse-SD-Intron1	GTAAGT	CTCTAG	745
Bsy-TD-Domain1 Intron1	GTAAGT	ACCTAG	707
Bse-TD-Domain1 Intron1	GTAAGT	ACCTAG	715
Bsy-TD-Domain2 Intron1	GTAAGA	AACCAG	215
Bse-TD-Domain2 Intron1	GTAAGA	AACCAG	257
Bsy-TD-Domain3 Intron1	GTAAGA	CACTAG	281
Bse-TD-Domain3 Intron1	GTAAGA	CACTAG	271
Bsy-TD-Domain4 Intron1	GTAAGA	CACTAG	384
Bse-TD-Domain4 Intron1	GTAAGA	CACTAG	384
Consensus intron 1	GTAAGW	MHMYAG	
Bsy-SD-Intron2	GTAAGT	TTCCAG	280
Bse-SD-Intron2	GTAAGT	TTCCAG	282
Bsy-TD-Domain1 Intron2	GTAAGT	TTTCAG	323
Bse-TD-Domain1 Intron2	GTAAGT	TTACAG	323
Bsy-TD-Domain2 Intron2	GTGAGT	TTGCAG	278
Bse-TD-Domain2 Intron2	GTGAGT	TTGCAG	278
Bsy-TD-Domain3 Intron2	GTGAGT	TTGCAG	378
Bse-TD-Domain3 Intron2	GTGAGT	TTGCAG	380
Bsy-TD-Domain4 Intron2	GTGAGT	TTGCAG	309
Bse-TD-Domain4 Intron2	GTGAGT	GTCTAG	358
Consensus intron 2	GTRAGT	KTNYAG	

Figures S1, S2, S3, S4, and S5

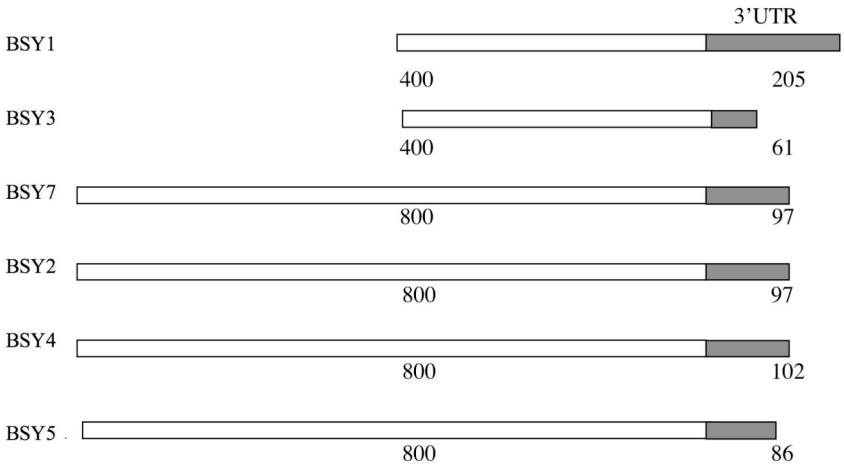


Figure S1. Schematic structure of the six initial cDNA sequences obtained for *B. symmytilida*. The coding sequence is shown in white and the 3'UTR in grey. The length of each part of the sequences is shown. BSY: *Branchipolynoe symmytilida*

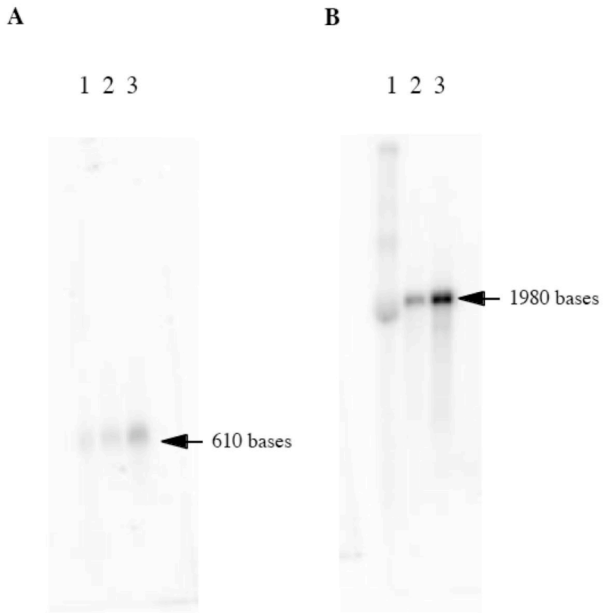


Figure S2. Northern-Blot results using probes from the sequences BSY3 (A) and BSY7 (B). The size of the obtained bands is indicated (the molecular weight marker is not visible). Lane 1: 20 µg of total RNA; Lane 2: 1 µg of poly A + mRNA; Lane 3: 3 µg of poly A + mRNA.

```
BsySD  AACATTGCAACATTTTCAGTGCTATAAAAGGCCAATGGGGTTCCACAGAGAAGTCACATCCcacactcaacatcatagcaa 80
BseSD  ---------------------------------------------------------------------------------

       ****************************************** ********************************
BsySD  cacattcagatactcggagggaagaactcttcagcatcaacagtaacaATGGTTTCTGCTGCACAGAAAGCTGCCATCAA 160
BseSD  cacattcagatactcggagggaagaactcttcagcatcaacagtaaccATGGTTTCTGCTGCACAGAAAGCTGCCATCAA 80

      ************************************************** ************************
BsySD GAGCTCCTGGTCGGGAGTGGACCTTCAGGCCGCTGGGGTCGCATTCTACCATCA|ACTGCAGGCTAATGCCCCAGATGCCT 240
BseSD GAGCTCCTGGTCGGGAGTGGACCTTCAGGCCGCTGGGGTCGCATTCTACCATCA|GCTGCAGGCTAATGCCCCAGATGCCT 160

       ****************************************************************************
BsySD  ATGCAGTCTTCAACTTAGGAAGTGATGCTGGAAAGATAGCAGCTCAGGGCCTCAAGACTATGACCTTCATTGATGGAGTA 320
BseSD  ATGCAGTCTTCAACTTAGGAAGTGATGCTGGAAAGATAGCAGCTCAGGGCCTCAAGACTATGACCTTCATTGATGGAGTG 240

       ******** ********************************************************* ******** *****
BsySD  GTCAAGGGCCTTGATGATATGGGTGGTGTCAAGGCAAGCATTGATACCCTGGGCAGCCGTCACACTGGGTATGGTGCCAA 400
BseSD  GTCAAGGGTCTTGATGATATGGGTGGTGTCAAGGCAAGCATTGATACCCTGGGCAGCCGTCACACGGGGTATGGGGCCAA 320

      *************** ************************************ ************************
BsySD GAAGGCCCACTTTGGG|CCTGCCGGACCATGCCTGCTCGCAGCACTGGCTGAGGTCTGTGGTGGAAAGTTCACCCCAGCTG 480
BseSD GAAGGCCCACTTTGGG|CCTGCCGGACCATGCCTGCTCGCAGCACTGGCTGAGGCCTGTGGTGGAAAGTTCACCCCAGCTG 400

      ****************** ********************************************
BsySD CCAAGGATGGCTGGACCGCCCTCTACGGAGTCATCGCTGATGGCATCTGCTCCCATCTGAGCtag 545
BseSD CCAAGGATGGCTGGACCGCACTCTACGGAGTCATCGCTGATGGCATCTGCTCCCATCTGAGCtag 465
```

Figure S3. cDNA sequence of the single-domain globin of *B. symmytilida* (BsySD) and *B. seepensis* (BseSD). The 5'UTR sequence and STOP codon are in lower-case. TATA box highlighted. Intron positions indicated by vertical lines. 3'UTR not shown.

Figure S4. cDNA sequence of the tetra-domain globins of *B. symmytilida* (BsyTD) and *B. seepensis* (BseTD). The 5'UTR and STOP codon are in lower-case. The TATA promoter binding site is highlighted. Domains separated by vertical lines. The 3'UTR not shown.

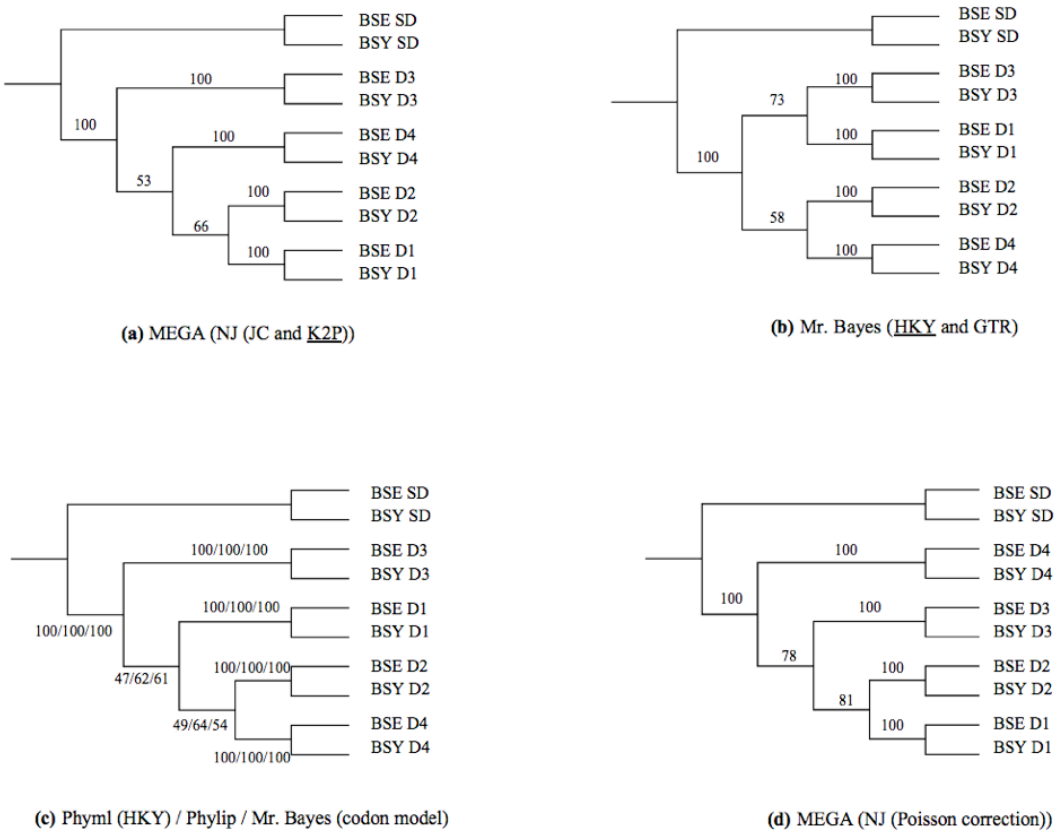


Figure S5. Different topologies obtained from the various phylogeny softwares and used in the PAML analysis. All trees are unrooted. Used softwares are indicated (the method and nucleotide substitution model are shown between parentheses). The bootstrap or posterior probability (Mr. Bayes) values are shown above/under the branches and the nucleotide substitution model to which they correspond is underlined. (a) to (c): topologies based on nucleotide sequences, (d): topology based on amino acid sequences. Bsy: *B. symmytilida*; Bse: *B. seepensis*; TD: tetra-domain; SD: single-domain; D1-D4: domains 1 through 4.

2.3 Identification of positively selected amino acid residues during the evolution of the tetra-domain globin

Based on the previous work on the evolution of the tetra-domain Hb from *Branchipolynoe* it was not clear which mechanism lead to the duplication of a single-domain gene. However, the domain duplication scenario can lead to two possible topologies. The main topology (A) favors D1D2 as the result of a first duplication, followed by a second one to generate the D3D4 (see Fig. 2.1 in 2.3.2). The other topology (B) depicts a scenario with three duplications starting with D4 as the most ancestral domain, followed by D3, D2 and D1. In addition, D1 and D3 do not seem to evolve the same way as D2 and D4, showing more resemblance with the SD, so they are probably under a stronger purifying selection.

To shed some light on these problems and to know which important amino acid residues could be crucial for the function and evolutionary history of these Hbs, in particular in the heme pocket, an analysis of positive selection was performed in all orthologous domains from *B. symmytilida* and *B. seepensis*. The methodology for the search of positive selection used on this work is not very recent (~10 years), but there are not as many studies as the ones using other phylogenetic methodologies (e.g. Neighbor-Joining, classical Maximum Likelihood) and the output is not as simple and intuitive as a tree. For these reasons I would like to describe the general assumptions and the methodology that was implemented. Therefore, a small section where I explain the tests that were done using the software PAML (Codeml) follows. This section will also be useful for the third chapter of this thesis.

2.3.1 Search for Darwinian positive selection – methods and assumptions

Nonsynonymous mutations imply a change of amino acid residue on the protein sequence, and synonymous mutations imply no change at all in the protein sequence. Natural selection works on the proteins and therefore affects differently the rates of nonsynosymous or synonymous mutations, favoring one of the two, according to the

environmental conditions at a certain period of time. The nonsynonymous/synonymous rate ratio ($\omega = d_N/d_S$) therefore becomes a sensitive measure of the selective pressures and DNA sequence evolution (Yang and Nielsen 2000, 2002). When the amino acid change offers a selective advantage it is fixed at a higher rate than the synonymous change. This implies that $\omega > 1$ is then a sign of positive selection, in opposition $\omega < 1$ indicates a purifying selection, and $\omega = 1$ indicates neutral selection (Yang 1998, Yang and Bielawski 2000).

Although earlier methods use the d_N and d_S rates, calculating them and averaging over the total site (codon) number, between two sequences (pairwise approach) (e.g. Nei and Gojobori 1986), this method has its flaws. This type of approach can hide residues that are the targets of positive selection, by the simple fact that their signature will be diluted by averaging over the total number of amino acid residues, which usually are highly invariable and under purifying selection ($\omega < 1$) in a protein sequence (Yang and Nielsen 2002). Other methods were developed in order to overcome this disadvantage (e.g. Goldman and Yang 1994), already using maximum likelihood to estimate d_N and d_S, but just as for the previous methods this one also did not take in to account that the different parts of a protein can be under different selective pressures (Nielsen and Yang 1998). The actual positive selection search method, using an estimation of by maximum likelihood calculations, mostly developed by Ziheng Yang and Rasmus Nielsen, and available in the software PAML, relies on different approaches to overcome the previous drawbacks. I will explain the general procedure and the tests used for the analyses that were performed in the scope of this thesis.

A basic codon substitution model developed by Yang, Nielsen and Hasegawa (1998) is the starting model for a set of possible analyses. This model is a simplified version of previous codon models (e.g. Goldman and Yang 1994, Grantham 1974) in the sense that it does not use the physical-chemical amino acid distances (that lead to the higher or lower probability of a residue being replaced by another), avoiding the caveats of the previous models (Yang 2004). Nevertheless this model accounts for the genetic code structure, transition/transversion rate bias and different base frequencies at codon positions (Yang 1998). Several important parameters are estimated (by maximum likelihood) in this model:

κ - transition/transversion rate ratio

ω - nonsynonymous/synonymous rate ratio

π_j – equilibrium frequency of codon *j*

Under this model $\omega = d_N/d_S$, d_N being the number of nonsynonymous substitutions per site and d_S the number of synonymous substitutions per site. From the previous model other more elaborate models can be developed, in order to allow different levels of heterogeneity of the d_N/d_S ratio among lineages and sites (Yang 1998, Yang and Nielsen 2002)

In this thesis two main analyses were performed using the tools in PAML, first the search for positive selection **between lineages** (branches) and second the search for positive selection **among sites in a given lineage**. The first step before any further analyses is to know if there is ω heterogeneity in the phylogenies that we want to analyze. For that we employ several *branch models* (software model designation between parentheses) (Table 2.1):

'one-ratio model' (model = 0) – is the simplest model and assumes the same ratio for all branches in the phylogeny;

'free-ratio model' (model = 1) – assumes a different d_N/d_S ratio for each branch.

From the comparison of the likelihood value of each of these two extreme models one can test the prediction that the d_N/d_S ratio is identical among lineages. This test is called **Likelihood Ratio Test** (LRT = $2\Delta\ell = 2*(\ell_1-\ell_0)$) and it is expected to have a χ^2 distribution. The result can then be compared to a probability χ^2 probability table, with the number of degrees of freedom equal to the difference in the number of parameters between the models (Yang 1998). The power and accuracy of the LRT were evaluated by Anisimova et al. (2001) with good results. When the value of the LRT is significant we can reject that the two models equally fit the data, and chose the one that has the highest likelihood value.

For the second step, assuming that indeed the 'free-ratio model' was indeed the best-fit model, we can compare the d_N/d_S ratio of specific or foreground branches (with ω_1) against the background branches (with ω_0) to know which ones were responsible for the ω heterogeneity among lineages. For this we use a 'two-ratio' or a 'three-ratio' model (model = 2) where we specify up to two branches for analysis against the background lineages, comparing up to three ω; ω_0 (background), ω_1 and ω_2 (representing two foreground branches) (Table 2.1). LRTs can again be used to decide which branch has a d_N/d_S ratio that significantly differs. This however does not give the total guarantee that a lineage is

the target of positive selection. Since the ω can be estimated or fixed at a given value, this is a useful tool to point out which branch really has a ω > 1. We can achieve that by forcing the d_N/d_S ratio of the foreground branch (one lineage at a time) to be equal to 1 ($\omega_1 = 1$). If ω was higher than 1 already when it was estimated, and if we force it to be equal to 1 with a significant result for the new LRT, we exclude with all certainty the null hypothesis, ω < 1 (Yang 1998). Applying the LRT between these tests and the previous one we can decide with assurance if the lineage is the target of positive selection. The complexity is greater when we compare 3 lineages (two foreground and one background).

The third step was the search for positively selected sites on the previously identified lineage, using a *branch-site model* (Yang and Nielsen 2002) (Table 2.1). This model combines the assumptions of the previous *branch model*, plus the ones from *site models*. In this model the d_N/d_S ratios cannot only vary among lineages but also among amino acid sites. Another advantage that came from the *site-models* is the development of a Bayesian approach to identify positively selected residues, based on maximum likelihood estimates of parameters and their sampling errors – Bayes empirical Bayes (BEB, Nielsen and Yang 1998, Yang 2004). The applied model, model A (Table 2.1), assumes that the d_N/d_S ratio of the chosen lineage can be different from the background (model = 2) and applies a positive selection *site model* (Nsites = 2) to the amino acid sites of that branch. Comparing, through LRT, to a simple *site model* assuming a nearly neutral theory (model = 1, Nsites = 0), we can discriminate which model best fits our data, and if there is variability of the d_N/d_S ratio among sites on the foreground branch. To be sure of the existence of positive selection again, in model A, we forced the ω of the foreground branch to be equal to 1 (MA $\omega_1 = 1$) and compared it through LRT to the previous model A with the estimated ω (Yang and Nielsen 2002, Yang 2004).

A search for positive selection among sites (*site model*, Nielsen and Yang 1998, Yang et al. 2000) was also performed, but since it does not allow the d_N/d_S ratio to vary among lineages, the detection of positively selected sites would depend on wether the average of d_N, over all lineages, is higher than the average of d_S (Yang and Nielsen 2002). So, I decided to give preference to the *branch-site model*. A summary of what was previsouly described is given in the following table.

Table 2.1. Summary of the available models that can be implemented in the positive selection likelihood analyses. In green are selected the two base models that constitute the *branch-site model* MA.

Branch model	*Site model*	*Branch-site model*
'one-ratio' model (model = 0)	M1 - nearly neutral (Nsites = 1)	
'free-ratio' model (model = 1)	M2 – positive selection (Nsites = 2)	MA (model = 2 & Nsites = 2)
'two-ratio' model (model = 2)	M7 – beta distribution (Nsites = 7)	
'three-ratio' model (model = 2)	M8 – beta distribution and ω (Nsites = 8)	

2.3.2 Results and discussion

The search for positive selection was performed using two separate topologies; topology A (D1D3_D2D4) and topology B (D4D3D2D1) (Fig. 2.1). The LRT between the *one-ratio model* and the *free-ratio model* gave significant results for both topologies (topology A: LRT = 22.9, df = 9, p = 0.01 and topology B: LRT = 22.19, df = 9, p = 0.01). From these results in both topologies the *free-ratio model* better fits the data (its likelihood value was always higher that the one attributed to *one-ratio model*). This means that in topology A and topology B the d_N/d_S ratios (ω) are heterogeneous among lineages, an heterogeneity that could be caused by positive selection (Yang 1998). We proceeded with the *branch-site model* (Model A: model = 2; nsites = 2) to look individually at each amino acid position along different branches of interest (foreground branches) and see whether we could detect positively selected residues. For these analyses we chose the branches that showed a ω > 1, and an estimated number of nonsynonymous substitutions for the entire lineage ($N*d_N$) greater than 2, in particular when the number of synonymous substitutions ($S*d_S$) was zero (Fig. 2.1). In topology A this test was therefore applied to branches *a*_A, *b*_A and *c*_A and in topology B to branches *a*_B, *b*_B, *c*_B and *d*_B. The results of the LRTs and amino acid residues determined to be positively selected are summarized in table 2.2.

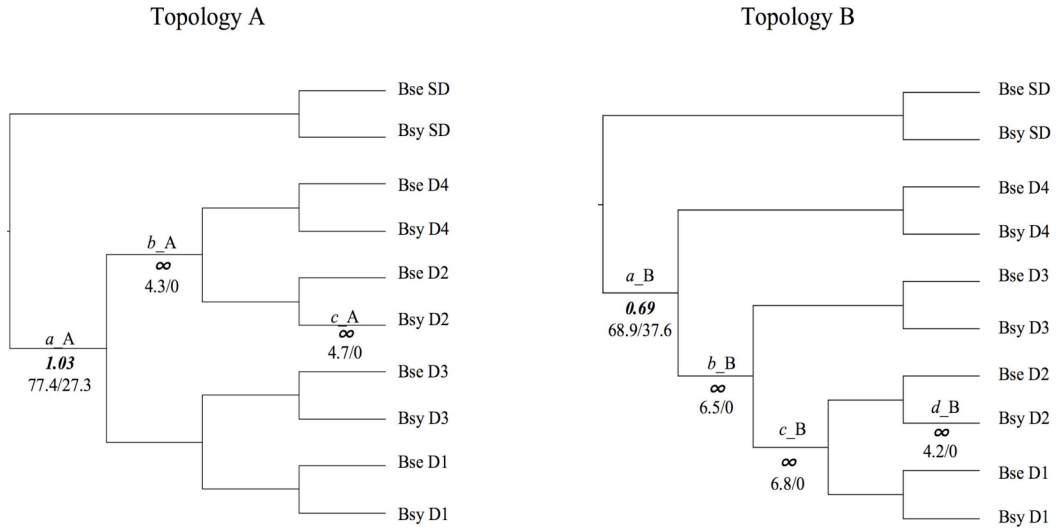


Figure 2.1. The two different topologies of the single- and tetra-domain extracellular hemoglobins from the genus *Branchipolynoe*, used for the search of positive selection. Each chosen branch is identified and the respective ω value is in bold and italics, the ratios indicate the maximum likelihood estimates for the numbers of nonsynonymous over the synonymous substitutions for the entire gene along the lineage.

For both topologies (A and B), the ω of the branch that separates the single-domain from the tetra-domain lineage indicates the presence of positively selected residues (Table 2.2). This suggests that these amino acid changes were necessary for formation of the multidomain globin, for structural and/or functional reasons. Duplication events are common in eukaryote lineages, producing gene families, with paralogs carrying on the function of the parent gene or giving rise to novel functions (Britten and Kohne 1968, Duboule and Wilkins 1998). Assuming an intracellular globin gene as the parental gene for the tetra-domain lineage (as discussed in the previous section), the paralogs seem to have functions similar to their ancestor. Hourdez et al. (1999b) showed that *Branchipolynoe* tetra-domain globins possess high oxygen affinity and a low cooperativity, both characteristics reminiscent of myoglobins function. *Branchipolynoe* extracellular hemoglobins can form dimers or trimers (Hourdez et al. 1999a). Dimerization is also known in body wall myoglobins from *Travisia foetida* (polychaete) (Terwilliger et al. 1980), and in the radular myoglobin of some gastropod mollusks (Brittain 1990).

Table 2.2. Results for the search for positive selection in the different lineages of the two topologies. For each branch the likelihood (ℓ) results of each model and the Likelihood Ratio Test (LRT) is shown. When model of selection (MA) was the fittest, its value is in green and the positively selected sites (PSS) are given, according to the BEB analysis (see section 2.3.1). p: number of estimated ω; df: degrees of freedom.

	Topology A												**Topology B**								
	***a*_A**			***b*_A**			***c*_A**			***a*_B**			***b*_B**			***c*_B**			***d*_B**	
Model/Test	**p**	**ℓ**		**p**	**ℓ**		**p**	**ℓ**		**p**	**ℓ**		**p**	**ℓ**		**p**	**ℓ**		**p**	**ℓ**	
M1	1	-2017.46		1	-2017.46		1	-2017.46		1	-2019.35		1	-2019.35		1	-2019.35		1	-2019.35	
MA	3	-2010.70		3	-2023.21		3	-2021.28		3	-2009.71		3	-2020.58		3	-2023.42		3	-2023.07	
	df	**LRT**	**PSS** (BEB)	**df**	**LRT**	**PSS** (BEB)	**df**	**LRT**	**PSS** (BEB)	**df**	**LRT**	**PSS** (BEB)	**df**	**LRT**	**PSS** (BEB)	**df**	**LRT**	**PSS** (BEB)	**df**	**LRT**	**PSS** (BEB)
M1 *vs* MA	2	**13.51** (p<0.01)	27V* 30A* 49Q* 62C*	2	11.5 (p<0.01)	-	2	7.64 (p<0.1)	-	2	**19.29** (p<0.001)	27V* 30A* 49Q* 55V* 62C**	2	2.46 (p<0.1)	-	2	8.14 (p<0.1)	-	2	7.43 (p<0.1)	-

BEB significance level: * - $p < 0.05$; ** - $p < 0.01$.

The positively selected residues, in the lineage leading to the tetra-domain hemoglobins, were represented in a 3D homology model[1] of domain 1 from *B. symmitilida* (Fig. 2.2). Their location (except for valine – 58V and cysteine – 65C), in the B helix and close to the DE corner lead us to believe that they could participate in the formation of polymers. However the majority of these amino acids are not located in the typical E and F helix that are known to possess special roles in the polymerization in other species (Royer et al. 2005). The glutamine in position 52 (52Q) is indeed in the helix E, and could be a part of a classical way to produce polymers. On the contrary the residues 30A and 27V are located in the B helix, a possible interface between two domains. 58V and 65C are located in the heme pocket, this latter one in the far end. The role of the valine is not quite clear, since it cannot form a hydrogen bond with the oxygen bound to the heme group. At the same time it was only found to be under positive selection in one of the two topologies (Table 2.2), so its identification as being under positive selection may be an artifact of analysis generated by a forced topology. The cysteine is particularly interesting as it is under positive selection in some extracellular globin from other annelid that live in sulfidic environments (Bailly et al. 2003). This amino acid in that same location was thought to be involved in the reversible binding of sulfide in *Riftia* (Zal et al. 1998) but this mechanism of sulfide binding has been challenged since (Flores et al. 2005, Flores and Hourdez 2006). Its occurrence in a lineage of extracellular globins different from the typical HBL-Hb globins suggests that it is the result of convergent evolution. Its function remains unclear but it may have role in protecting the heme to prevent the reaction of sulfide with the iron atom (that would form sulfhemoglobin, incapable of binding oxygen).

In the other branches, in both topologies, there were some cases of heterogeneity of the ω when comparing M1 *vs* Ma, with always M1 being chosen as the best fitting model. There was then no sense in doing the test MA *vs* $MA_{\omega=1}$ (Table 2.2). This could indicate that the heterogeneity of ω could be due to a nearly neutral evolution. However all the other external branches that were not chosen for the analyses of positive selection, had a ω

[1]The 3D model was obtained by homology modeling using the tools available on the website from SWISS-MODEL. In summary, this program constructs an optimized structural model based on the 3D structure of a template amino acid sequence that has the best psi-blast with our sequence of interest. *Chimera* was used to visualize and optimize the resulting structure. For more details see chapter 3.

< 1 (data not shown), which usually indicates purifying selection. This can be explained by the fact that after the positive selection for the first duplication event, the dominant pressures were to maintain this previously improved structure through purifying selection, just like in the globin family in *Caenorhabditis* and in the mammalian α and β globins (Goodman et al. 1975, Hoogewijs et al. 2008).

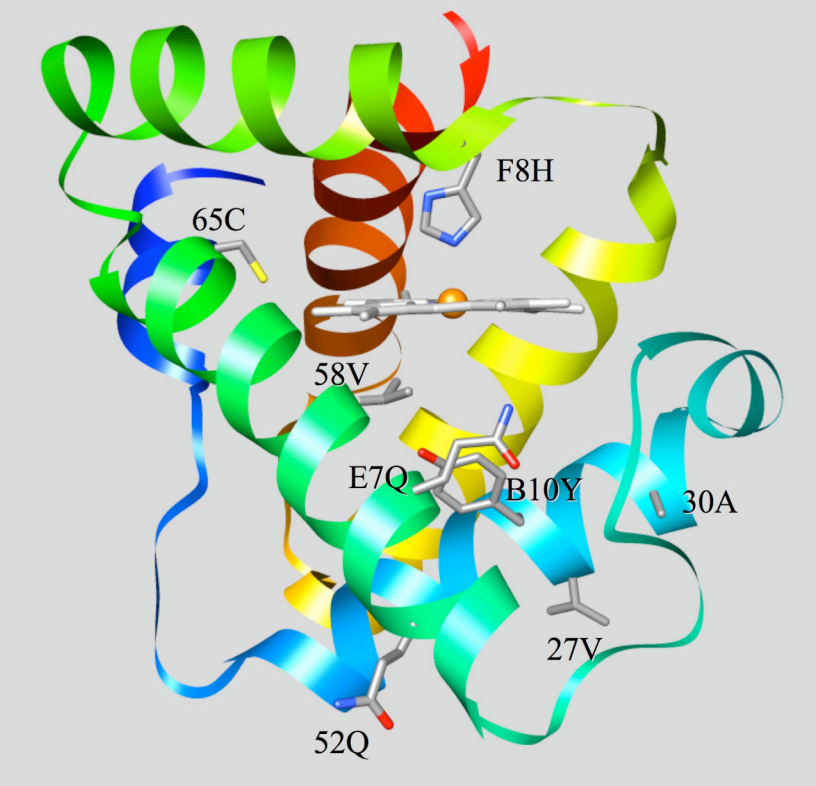


Figure 2.2. 3D homology model from *Branchipolynoe symmytilida* domain 1, based on a *Lumbricus terrestris* hemoglobin quaternary structure (see box on previous page). The represented amino acid residues are the ones under positive selection (27V, 30A, 52Q, 58V and 65C), the ones that are responsible for the high oxygen affinity in *Ascaris* Hb (B10Y and E7Q) and the conserved proximal histidine (F8H). The heme group is represented in the heme pocket with the Fe atom in orange.

2.5 General conclusions

Gene duplication is considered as a major force of evolution, because natural selection frequently does not affect copies of a gene, letting them evolve freely in other, different, directions, possibly giving rise to novel functions (Ohno 1970). *Branchipolynoe* multi-domain hemoglobins are unique in the annelid phyllum. First because they are extracellular multi-domain hemoglobins and second they show a different evolutionary history from other extracellular annelid hemoglobins. The typical gene structure has been conserved, and it is repeated 4 times in the gene for the tetradomain hemoglobin. Their ancestor must have been an intracellular myoglobin-like globin, from which they still possess similar functional characteristics (Hourdez et al. 1999b). The event that led to this duplication is still unknown, although we tried to unveil which molecular history could be at the origin of the tetra-domain structure that we observe today. Our results indicate that the most likely duplication scenario involves two successive duplication events, with the first one giving rise to domains 1 and 2, and this group duplicating as a whole to form domains 3 and 4.

The expression of this globin was a clear advantage for these scale-worms that inhabit hydrothermal vents and it was probably evolutionarily promoted. *Branchypolynoe* spp. hemoglobin possesses high oxygen affinity, low cooperativity, is found in the coelome in high concentrations and most likely has an oxygen storage role (Hourdez et al. 1999a, 1999b). The amino acid sequence showed residues that are similar to some key amino acid residues, responsible for the very high affinity in the most oxygen-avid hemoglobins known, *Ascaris* di-didomain extracellular hemoglobin (De Baere et al. 1994). In addition, other residues were identified in locations involved in subunit interactions, residues potentially involved in the stabilization of bound oxygen (with a possible effect on oxygen affinity), and finally a cysteine whose function could relate to protection against sulfide.

3. Single-domain hemoglobin adaptations in the hydrothermal polynoids

3.1 Introduction

Inhabiting ecosystems that are characterized by extreme conditions requires special adaptations, such as the ones seen in the hydrothermal vent fauna (Childress and Fisher 1992, Hourdez and Lallier 2007). In this environment one of the most important constraints is the extremely variable oxygen availability, making the uptake of this element not an easy task (Hourdez and Lallier 2007). Organisms that live in hypoxic/anoxic environments can cope with these difficulties at several levels; changing their behavior, expressing different morphology and/or changing their physiology (Childress and Seibel 1998, Hourdez and Lallier 2007). This last level of adaptation is commonly used and can comprise several enzyme/protein families all working for the same final end: to reliably produce energy for the metabolic needs (Powell and Somero 1986, Weber and Vinogradov 2001). Globins are one of the most studied proteins in respiratory adaptation studies, with a wealth of information on its function and structure (Hourdez and Weber 2005, Weber and Vinogradov 2001). Although carrying oxygen is not the only task of globin-like proteins (Imai 1999), it is one of the main roles for which they present some adaptations (see Weber and Vinogradov for a review).

Most hydrothermal vent organisms possess repiratory pigments, mainly hemoglobins, but also hemocyanins in crustacea and gastropods. Hydrothermal vent annelids possess extracellular hemoglobins with high oxygen affinity (Hourdez and Lallier 2007, Hourdez and Weber 2005). One of the most remarkable examples is the Hb from the siboglinid annelid *Riftia pachyptila*, with a HBL structure, high oxygen affinity and the unique characteristic of binding O_2 and sulfide reversibly (Arp and Childress 1983). Other hydrothermal annelids, such as *Alvinella* spp., also possess HBL-Hb, with very high oxygen affinity (Terwilliger and Terwilliger 1984, Toulomond et al 1990, Hourdez et al. 2000). This structure is typical for an annelid extracellular Hb, and is found non-hydrothermal vent species such as the earthworm *Lumbricus terrestris* or the lugworm *Arenicola marina* (Krebs et al. 1996, Zal et al. 1997a). Their functional properties however

differ, in particular in their O_2 affinity, Bohr effect and cooperativity (Toulmond et al. 1990, Krebs et al. 1996, Arp et al. 1990 and Toulmond 1970 in Weber and Vinogradov 2001). In the nematode *Ascaris* spp. the extracellular Hb is structurally very different (octameres of di-domain globins), and is one of the most oxygen avid respiratory pigments, most likely an adaptation to its hypoxic environment, the pig intestine (Darawshe et al. 1987, Gibson and Smith 1965, Weber and Vinogradov 2001).

These high oxygen affinities are the result of selective pressures that act to produce different Hb structural arrangements. Not only at the polymerization level, that might increase cooperativity effects (Weber and Vinogradov 2001), but also at the amino acid residue level, where the substitution of a single amino acid in the heme pocket can increase the oxygen affinity by a factor of 100, as in the *Ascaris* Hb (De Baere et al. 1994). In this example the conserved distal histidine (E7H) is replaced by a glutamine (Q) and in position B10 the usual residue (leucine) is replaced by a tyrosine (Y). Both these amino acids can establish hydrogen bonds with the dioxygen, which will decrease the off-rate (O_2 dissociation from the Fe from the heme group) (De Baere et al. 1994). The replacement of the typical vertebrate E7H by a glutamine is relatively common in invertebrate globins (Weber and Vinogradov 2001), but oxygen affinity is highly variable and could mostly depend on the heme pocket conformation (Peterson et al. 1997). *Ascaris* and *Lucina pectinata* (a bivalve from sulfidic environments) globins possess both the B10Y and E7Q, and show high oxygen affinities. The P_{50} in *Ascaris* Hb is however higher than in the bivalve, most likely because the nematode heme pocket forms a tight cage, making the release of oxygen more difficult (Peterson et al. 1997).

In the previous chapter, we showed that residues E7Q and B10Y also exist in the *Branchipolynoe* single- and tetra-domain Hb and their high oxygen affinity values (Hourdez et al. 1999b) might be due to the occurrence of those amino acids. The presence of these residues in other hydrothermal polynoidae species is a relevant issue, since not all scale-worms live in the same habitat inside the vent area, so they do not have the same respiratory constraints, which could lead to different selective pressures on their Hbs. This chapter deals exactly with this question, where the single-domain Hb from several hydrothermal polynoid species representing various environments (and thereby potential hypoxia level) were studied and searched for positive selection. The identified residues were located on a model of their 3D structures to understand the specific role of each residue.

3.2 Positive selection drives the evolution of the single-domain globin in the hydrothermal vent scale-worms *Branchipolynoe* spp. - Manuscript

This work will be presented in the format of a scientific article that will be submitted to the Proceedings of the National Academy of Sciences of the United States of America (PNAS). In order to facilitate the consultation, the main titles of this article are a continuation of the chapter headings. Again, as in the previous chapter, some of the bibliography of the following work will not be included in the final list of references of this thesis.

Positive selection drives the evolution of the single-domain globin in the hydrothermal vent scale-worms *Branchipolynoe* spp.

Projecto-Garcia[1,4], J., T. Govindji[2], D. Jollivet[1,4], J. Mary[3,4], S. W. Schaeffer[2], F. H. Lallier[3,4] and S. Hourdez[1,4]

[1]GAME research group, CNRS UMR 7144, Station Biologique de Roscoff, BP 74, Place Georges Teissier 29682 Roscoff cedex, France

[2]Department of Biology and Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA 16802, USA

[3]EIMME research group, CNRS UMR 7144, Station Biologique de Roscoff, BP 74, Place Georges Teissier 29682 Roscoff cedex, France

[4]Université Pierre et Marie Curie UMR 7144, Place Georges Teissier, 29682 Roscoff cedex, France

Corresponding author:

Stéphane Hourdez (hourdez@sb-roscoff.fr)

GAME research group, CNRS UMR 7144, Station Biologique de Roscoff, BP 74, Place Georges Teissier 29682 Roscoff cedex, France

T: +33 (0)2 98 29 23 40

Fax: +33 (0)2 98 29 23 24

Keywords: extracellular, globin, single-domain, positive selection, heme, oxygen affinity, Polynoidae, dimer interactions

Abstract

Deep-sea hydrothermal vents represent very challenging environments for metazoans. In particular, hypoxia makes the acquisition of oxygen difficult to meet the metabolic needs, and requires specific adaptations. Hydrothermal vent scale-worms (annelids of the family Polynoidae) express large amounts of extracellular hemoglobins in their coelomic fluid, in sharp contrast with their shallow-water relatives that are completely devoid of circulating hemoglobins and only possess a small amount of intracellular globins in their nervous system. We sequenced the gene for single-domain globins from seven species of polynoids representing various vent microhabitats (and associated constraints) and a non vent species to look for adaptations at the molecular level in this family of worms.

Although extracellular, all the single-domain globins form a monophyletic clade that clusters with intracellular globins from other annelids. This suggests that these hemoglobins have evolved from an intracellular form, after the original duplication that gave rise to the other extracellular globins in annelids. The search for positively selected amino acids revealed their presence only in the lineage leading to the two species of *Branchipolynoe* (that live inside the mantle cavity of mussels), indicating that the adaptive value of the globins in vent species did not require modifications of the sequence for all the free-living species. This suggests that the sheer expression of these globins in the coelomic fluid may be the main advantage for the species living in a hypoxic environment. In *Branchipolynoe*, the multiplicity of hemoglobins probably relieved the selective pressure on the single-domain hemoglobin, allowing the acquisition of novel functions by positive Darwinian selection. Two of the three amino acids under positive selection are located in positions where they can affect the formation of multimers, and the third one is located in the heme pocket where it can affect the binding on oxygen.

3.2.1 Introduction

Hydrothermal vent ecosystems are located along the oceanic ridges or active convergent margins on the ocean floor. These areas offer harsh and challenging conditions for metazoans, because of the presence of heavy metals and sulfide (both toxic compounds), hypoxic conditions, high temperatures, and low pH (Childress and Fisher 1992, Tunnicliffe 1991). Despite these conditions, hydrothermal vents are characterized by a high density of -mostly endemic- fauna, but also by a low species richness, because the constraints act as a filter for the colonization of species that are not adapted to these conditions. The adaptations of the hydrothermal species can be observed at several levels: trophic, morphological, enzymatic and ATP synthesis (Childress and Fisher 1992). In particular, hypoxia is possibly the most basic challenge that needs to be overcome by the organisms to thrive and reap the benefits of the local primary production.

Respiratory adaptations found in hydrothermal vent species can affect different organizational levels. They can affect the behavior (avoidance of some areas, variations in ventilation), the morphology (increased gills surface areas, reduced diffusion distances), the biochemistry (metabolism, presence of respiratory pigments), and the molecular level (for a review, see Hourdez and Lallier 2007). In particular, respiratory pigments usually exhibit high oxygen affinities when compared to littoral species that live in well-oxygenated environments (Hourdez and Weber 2005; Hourdez and Lallier 2007). In some annelids, circulating extracellular hemoglobins in high concentrations represent a significant form of oxygen storage. In addition, their high oxygen affinity fallows oxygen uptake from the environment even when its concentration is low. Finally, some have the capacity to reversibly bind both O_2 and sulfide, an ability that is essential for the functioning of the symbiosis in the vestimentiferan tubeworm *Riftia pachyptila* (Childress and Fisher 1992, Weber and Vinogradov 2001).

The Polynoidae scale-worms are very diverse in hydrothermal ecosystems, representing ~10% of all invertebrate species (Tunnicliffe 1991). Different species occupy all the available hydrothermal habitats where metazoa are found, ranging from the coldest areas (~2ºC) to the warmest areas on chimney walls (~40ºC with spikes of higher temperature). Before the discovery of hydrothermal vent species, this family was thought to possess only intracellular globins, in the muscles (myoglobin) and in the nerve cord

(neuroglobin) (Weber 1978; Dewilde et al. 1996). Interestingly hydrothermal polynoid species possess a red-colored coelomic fluid, due to the presence of extracellular hemoglobins (Hourdez et al. 1999a). In the genus *Branchypolynoe* two types of extracellular hemoglobins exist, formed of a single-domain or a tetra-domain globin subunits. This latter type is likely the result of evolutionary tinkering of a single-domain intracellular globin (Projecto-Garcia et al. in press). Although not all species possess tetra-domain hemoglobins, they all possess at least single-domain extracellular hemoglobins on which we focused our attention for the present study of adaptations.

The challenging conditions encountered at hydrothermal vents could lead to functional innovations (evidenced by positive selection) essential for the survival of the species. The detection of adaptation signatures and of the action of the selective pressure at the amino acid level can be performed by the use of the nonsynonymous/synonymous substitution rate ratio ($\omega = d_N/d_S$) (Yang 1998, Yang and Nielsen 2002). Using this tool we investigated the possible respiratory adaptations to hydrothermal conditions using the single-domain extracellular globin. The scale-worm species studied here represent a range of hydrothermal vent conditions (and associated levels of constraints), and different life styles (free-living or commensal). Their phylogenetic relationships are known (Hourdez et al. in prep.), giving a framework in which to study adaptations. The results of these analyses are discussed against other extracellular annelid globins sequences and structures, and also in connection with the microhabitat of the different polynoid species we used.

3.2.2 Materials and methods

Animal collection

The collected species, sampling area, and habitat are detailed in Table 1 (at the end of the article). All the specimens were identified on board the research vessel, and immediately frozen and stored at -80ºC until used. The species were chosen to represent various microhabitats at hydrothermal vents, from the coldest with the least hydrothermal influence, to the warmest on the chimney walls. *Branchinotogluma segonzaci* is a representative of the warmest habitat, on the chimney wall. *B. trifurcus* and *Branchiplicatus cupreus* are usually found in colder areas, farther away from from the source of the fluid. The deep-sea species *Harmothoe* sp. collected in the Lau Basin near hydrothermal vent sites was chosen as a deep-sea, non-hydrothermal reference. *Branchipolynoe seepensis* and *B. symmytilida* were chosen to represent the commensal life-style. Finally, *Lepidonotopodium williamsae* represents a free-living, non-branchiate endemic hydrothermal species.

Table 3.1. Sampling locations, habitat and depth for the species used in this study.

Species	Sampling area	Habitat and depth
Branchipolynoe symmytilida	9º50'N area (9º46'N, 104º21'W), East Pacific Rise	Mussel beds (commensal in the mussel pallial cavity), 2500 m
Branchipolynoe seepensis	Lucky Strike site (37º18'N, 32º16'N), Mid-Atlantic Ridge	Mussel beds (commensal in the mussel pallial cavity), 1700 m
Branchiplicatus cupreus	9º50'N area (9º46'N, 104º21'W), East Pacific Rise	Mussel beds, free-living, 2500 m
Lepidonotopodium williamsae	11˚N area (11˚25'N, 103˚47'N), East Pacific Rise	Mussel beds and tubeworm aggregations, free-living, 2500 m

Species	Sampling area	Habitat and depth
Branchinotogluma segonzaci	Lau Basin 1. ABE (20°46'S, 176°11'W), Lau Basin 2. Tow Cam (20°06'S, 176°34'W), Lau Basin	Chimney walls, free-living, 2150 and 2700 m
Branchinotogluma trifurcus	Lau Basin 1. Kilo Moana (20°03'S, 176°08'W), Lau Basin 2. Tu'i Malila (21°59'S, 176°34'W), Lau Basin	*Ifremeria nautilei* clumps, free-living, 2600 and 1900 m
Harmothoe sp.	Kilo Moana (20°03'S, 176°08'W), Lau Basin	Peripheral areas, 2600 m

Nucleic acids extractions, and cDNA synthesis

A standard phenol/chloroform protocol following proteinase K digestion (Sambrook et al. 1989) was used to extract total DNA from *Branchipolynoe symmytilida*, *B. seepensis*, *Branchiplicatus cupreus*, and *Lepidonotopodium williamsae*. For *Branchinotogluma segonzaci*, *B. trifurcus*, and *Harmothoe* sp., total DNA was isolated following a CTAB + PVPP extraction protocol (Doyle and Doyle, 1987). For all species total RNA was extracted from the tissues using TRI Reagent® (Sigma). For all species, cDNA was then synthesized by reverse transcription using MMLV-Reverse Transcriptase with an oligo$(dT)_{18}$ or an anchored oligo(dT) primer (Table S1 – Supplementary material).

cDNA and gene sequencing

Degenerate primers to amplify the single-domain (SD) globin were designed based on the sequences from the Polynoidae *Branchipolynoe symmytilida* and *B. seepensis*, as well as the Aphroditidae *Aphrodita aculeata*. The PCR conditions and the type of template (cDNA or total DNA) differed according to the species that was used for amplification (see Table S2 – Supplementary material). The PCR products were visualized on a 1.5% agarose gel containing ethidium bromide under UV light, and cloned with the TOPO TA Cloning kit (Invitrogen). The positive clones were sequenced, and the sequences were used to produce specific primers for all the species (Table S1) for further work.

Chromosome directional walking on total DNA (see Projecto-Garcia et al. in press, for details) was used to sequence the missing parts of the coding sequences, the 5'UTR, and the promoter region of the globin genes.

Protein sequence and phylogenetic analyses

The nucleotide sequences were assembled, checked and, edited based on their chromatograms with CodonCode Aligner 2.0.6 (http://www.codoncode.com/aligner/index.htm). All cDNA sequences were translated into amino acid sequences using the universal genetic code. Multiple nucleotide and amino-acid sequence alignments were performed with ClustalX 2.0.10 (Larkin et al. 2007) and, when necessary, manually optimized with the sequence aligner editor Se-Al 2.0a11 Carbon (http://tree.bio.ed.ac.uk/software/seal/, developed by Andrew Rambaut). The optimization was based on minimizing the number of indels, and verifying that the nucleotide sequences alignment followed the amino acid sequences alignment.

Tree construction

A Bayesian analysis was performed with the software Mr. Bayes (Ronquist and Huelsenbeck 2003, Huelsenbeck and Ronquist 2001) on an alignment of all the Polynoidae, and other extracellular and intracellular annelid globin sequences.
A neighbor-joining mitochondrial 16S tree with only the polynoid species used here was constructed using MEGA4 (Tamura et al. 2007, Hourdez et al. in prep).

Positive selection and associated tests (Codeml)

The search for potential positive selection, among branches and sites, was performed by maximum likelihood (ML) following the procedure described by Nielsen and Yang (1998), Yang (1998), Yang and Nielsen (2002) and the PAML program documentation (Codeml).

Using the 16S phylogeny for the species of Polynoidae as a framework (Fig. 5, Hourdez et al. in prep.), we first tested whether the d_N/d_S (ω) ratios were different among lineages with a likelihood ratio test (LRT = $2\Delta\ell$) between the *one-ratio model* (same ω for all branches) and the *free-ratio model* (each branch with a different ω). The advantage of the LRT is that its results can be compared to a X^2 distribution. Power and accuracy of the LRT were evaluated by Anisimova et al. (2001) with good results. A *two-ratio model*, where the ratio for a chosen branch (ω_1) is distinct from the ratio for the remaining (background)

branches (ω_0), was tested against the *one-ratio model* with a LRT. This allowed us to determine which branch, if any, had a distinct ω. Once the branches with significantly different ω were identified, we searched for differences of ω ratio among sites on those specific lineages. Yang and Nielsen (2002) implemented a test that lets the ω ratio vary both among sites and among lineages (branch-site test). We performed a LRT test comparing MA, a combination of a *two-ratio model* (model = 2) with a *positive selection site model* (M2 – Nssites = 2) (Yang and Nielsen 2002), against the nearly neutral site model (M1 - NSsites = 1). A second test, comparing M1 against a MA fixing $\omega_2 = 1$ ($MA_{\omega=1}$), allowed us to test whether the site variability is actually due to positive selection. This test also identifies sites that were under positive selection, with Bayes empirical Bayes (BEB) probabilities higher than 95%.

Three-dimensional model and localization of key amino-acid replacements

To construct a 3D homology model of some of the polynoid species globin sequences we used the tools available on the SWISS-MODEL website (http://swissmodel.expasy.org//SWISS-MODEL.html). Briefly, this modeling instrument allowed us to obtain a 3D model of an amino acid sequence of interest based on the available 3D structure of a template sequence that has the best psi-blast score with our sequence. Atomic energy calculations and minimization of the force fields were optimized.

The product of this rough modelization was visualized using UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081) (Pettersen et al. 2004). This same software was also used to graphically enhance the model, to highlight some important residues, and to insert the heme group in the heme pocket of our model. For this manipulation we used the heme coordinates from the template sequence. For the structural alignment analysis we used the software Pymol (DeLano 2008).

The Hydrophobic Cluster Analyses (HCA) were produced using the tools available on the software website (http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA) (Callebaut et al. 1997, Gaboriaud et al. 1987).

3.2.3 Results

Single-Domain DNA and cDNA amplification and sequencing

Except for *Harmothoe* sp. for which we do not have the gene structure, the single-domain genes all exhibit the typical vertebrate globin gene structure with 3 exons/2 introns. The introns are located in the conserved positions B12.2 and G7.0 in reference to the *Physeter catodon* globin fold.

Coding sequences

For *Branchinotogluma segonzaci*, *B. trifurcus*, *Branchiplicatus cupreus*, *Lepidonotopodium williamsae*, and *Harmothoe* sp. several cDNA sequences were obtained, indicating a possible allelism or the presence of different globin loci in these species. Except for *B. cupreus*, for which the sequence differences were such that we most likely have two different loci, a consensus sequence was produced for the others species (considering the most abundant nucleotides between the sequenced clones) for the following analyses.

The complete cDNA sequences from the single-domain globin from *B. symmytilida* and *B. seepensis* have a coding sequence of 417 nucleotides including the stop codon. For *B. segonzaci* we also have a full-length coding sequence (cds) of 417 nucleotides. For *B. trifurcus*, *B. cupreus* and *L. williamsae* we could only amplify 366 bp (122 codons, including the initial methionine) of the cds. These partial sequences correspond to the first two exons, and most of the third (and last) exon. Finally, for *Harmothoe* sp., we sequenced a 323 bp portion that corresponds to roughly half of exons 1 and 3, and the complete exon 2.

Over the common 323 bp, there is no indel and there is a fairly high identity (67.2%) in nucleotides (Fig. 3.1).

```
                                                                |I1
          * **  * ** ** ***** ** **    ** *****    **  ** ***** ** * ** ***** **  * ******
BrcuSD2   CGGGAGTAGATCTTCAGGCTGCGGGAAATGCATTCTATGCTCAGTTGAAGGCTCATGCTCCAGATGCGTATGCTGTCTTC    80
BrcuSD1   CGGGAGTAGATCTTCAGGCTGCTGGAAATGCCTTCTATGCTCAGTTGAAGGCTAATGCCCCGGATGCGTATGCCGTCTTC    80
BsySD     CGGGAGTGGACCTTCAGGCCGCTGGGGTCGCATTCTACCATCAACTGCAGGCTAATGCCCCAGATGCCTATGCAGTCTTC    80
BseSD     CGGGAGTGGACCTTCAGGCCGCTGGGGTCGCATTCTACCATCAGCTGCAGGCTAATGCCCCAGATGCCTATGCAGTCTTC    80
BtriSD    CGGGAGTGGACCTTCAGGCTGCTGGGAATGCATTCTATGCCCAGCTGAAGGCTAATGCCCCCGATGCTTATGCGGTCTTC    80
HarmoSD   CGGGAGTGGACCTTCAGGCTGCTGGGAATGCATTCTATGCCCAGCTGAAGGCTAATGCCCCCGATGCTTATGCGGTCTTC    80
BsegSD    CTGGAATCGACCTTCAGGCGGCCGGCAATGCATTCTACCATCAGCTGAAGGCTAATGCCCCGGATGCTTATGCGGTCTTC    80
LewiSD    CTGGCGTGGACCTGCAGGCTGCTGGGATAGCATTCTACAGCCAGCTGCAGGCTTATACGCCAGATGCATACCCAGTCTTC    80

          ***** *** * *** * ** ****  ** ** ** ***** **    ****  *** ******* *  ** **     *
BrcuSD2   AACTTAGGAGGGGATGCTGGGAAGACAGCGGCACAGGGCCTGAAGGTTATGACCTTCATTGATGGAGCGGTCAAGGACCT   160
BrcuSD1   AACTTAGGAGGGGATGCTGGGAAGACTGCGGCACAGGGCCTGAAGGTTATGACCTTCATTGATGGAGCAGTGAAAGGTCT   160
BsySD     AACTTAGGAAGTGATGCTGGAAAGATAGCAGCTCAGGGCCTCAAGACTATGACCTTCATTGATGGAGTAGTCAAGGGCCT   160
BseSD     AACTTAGGAAGTGATGCTGGAAAGATAGCAGCTCAGGGCCTCAAGACTATGACCTTCATTGATGGAGTGGTCAAGGGTCT   160
BtriSD    AACTTGGGAGGGGATGCTGGAAAGATAGCAGCTCAGGGCCTCAAAGTCATGACCTTCATTGATGGAGTTGTCAAGGGTCT   160
HarmoSD   AACTTGGGAGGGGATGCTGGAAAGATAGCAGCTCAGGGCCTCAAAGTCATGACCTTCGTTGATGGAGTTGTCAAGGGTCT   160
BsegSD    AACTTAGGAGGGGATGCCGGGAAGATAGCAGCTCAGGGCCTCAAGGTCATGACCTTCATTGATGGGGTTGTCAAGGGCCT   160
Lewi_SD   AACTTAGGAGGAGATCCTGGAAAGACTGCAGCACAAGGCCTGAAGGTCATGAATTTCATTGATGGAGCTGTCAAAAACAT   160

           ** ** ***  ** ******   ***  * **    *  *     ** ***** *  ** ** *** * ****  ****
BrcuSD2   TGACGATATGGGTGCTGTCAAGGCAAGTGTTGATGCACTTGGACAACGCCACACTGGGTATGGCGCCAAGAAGGAACACT   240
BrcuSD1   TGATGATATGAGTGCTGTCAAGTCAAGTATCGATGCACTCGGACAACGCCACACTGCATATGGTGCCCATAAGGGCCACT   240
BsySD     TGATGATATGGGTGGTGTCAAGGCAAGCATTGATACCCTGGGCAGCCGTCACACTGGGTATGGTGCCAAGAAGGCCCACT   240
BseSD     TGATGATATGGGTGGTGTCAAGGCAAGCATTGATACCCTGGGCAGCCGTCACACGGGGTATGGGGCCAAGAAGGCCCACT   240
BtriSD    GGATGACATGGGTGGTGTCAAGGCAAGCATTGATACCCCGGGCCAGCGTCACACCGGGTACGGTGCCAAGAAGGCCCACT   240
HarmoSD   GGATGACATGGGTGGTGTCAAGGCAAGCATTGATGCCCTGGGCCAGCGTCACACCGGGTACGGTGCCAAGAAGGCCCACT   240
BsegSD    GGATGACATGGGTGGTGTCAAAGCAAGCATTGACAACCTGGGCCAACGCCACACCGGATACGGCGCCAAGAAGGCCCACT   240
LewiSD    CGATGACATGATTGCTGTCAAGGGAAGCATTGATGCTCTGGCTCAGCGTCACACTGGATACGGTGCCAAGAAGGCCCACT   240

             |I2
            ***** **** ** ** ****** * ** **  * *******  ** ******** ************ *********
BrcuSD2   TTGGGCCTGCGGGACCCTGCCTGCTTGCAGCATTGGCTGAGGTCTGTGGTGGAAAATTCACCCCAGCTGCCAAGGATGCA   320
BrcuSD1   TTGGGTCTGCCGGGCCCTGCCTGCTTGCAGCACTGGCTGAGGTCTGCGGTGGAAAATTCACCCCAGCTGCCAAGGATGCC   320
BsySD     TTGGGCCTGCCGGACCATGCCTGCTCGCAGCACTGGCTGAGGTCTGTGGTGGAAAGTTCACCCCAGCTGCCAAGGATGGC   320
BseSD     TTGGGCCTGCCGGACCATGCCTGCTCGCAGCACTGGCTGAGGCCTGTGGTGGAAAGTTCACCCCAGCTGCCAAGGATGGC   320
BtriSD    TTGGGCCTGCTGGACCATGCCTGCTTGCAGCACTGGCTGAGGTCTGTGGTGGAAAGTTCACCCCAGCTGCCAAGGATGCA   320
HarmoSD   TTGGGCCTGCTGGACCATGCCTGCTTGCAGCACTGGCTGAGGTCTGTGGTGGAAAGTTCACCCCAGCTTCCAAGGATGCA   320
BsegSD    TTGGGCCTGCCGGACCATGCCTGCTTGCAGCGCTAGCTGAGGTCTGTGGTGGAAAGTTCACCCCAGCTGCCAAGGATGGC   320
LewiSDc   TTGGGCCTGCCGGCCCCTGCCTGTTGGCTGCACTTGCTGAGGTGTGCGGTGGAAAGTTCACCCCAGCTGCCAAGGATGCA   320

          ***
BrcuSD2   TGG   323
BrcuSD1   TGG   323
BsySD     TGG   323
BseSD     TGG   323
BtriSD    TGG   323
HarmoSD   TGG   323
BsegSD    TGG   323
LewiSD    TGG   323
```

Figure 3.1. Alignment corresponding to 323 bp of cDNA coding for the single-domain globin in common between all the species studied here. Intron positions are represented above the sequences.

Promoter regions, and UTRs

For *B. symmytilida*, *B. cupreus*, and *L. williamsae* our sequence covers the full 5'UTR (~68 bp), and for the former and latter, we also have about 440 bp of the promoter region. For *B. seepensis*, *B. trifurcus*, and *B. segonzaci*, we have 48 bp of the 5'UTR (Fig. 3.2).

For *B. symmytilida* and *L. williamsae* the promoter sequences were fairly conserved (62.3%). In both sequences the TATA box was located ~30 bp upstream of the beginning of the 5'UTR (Fig. 3.2). The identity between the amplified common parts of the 5'UTR (48 bp) for all polynoid species was ~80%.

In addition to the TATA box, only one more transcription signal (CAC binding protein motif) was found in *B. symmytilida* promoter sequence. On the contrary several were found in the promoter sequence from *L. williamsae*. Three GATA motifs, known for playing an important role in erythroid development (De Maria et al. 1999) and also to enhance erythropoeisis in response to tissue hypoxia (Krantz 1991, Zon et al. 1991) and a CAC binding protein motif, that, when mutated, produces anomalies in the transcription of the ß-globin gene (Kulozik et al. 1991). Some core sequences of heat shock elements (HSE) were found in both species sequences. However, these are only putative, as HSEs only perform their function in pairs (Cunniff et al. 1991), and we did not find 2 perfect core motifs in tandem.

```
BsySD    AGGTCTATTATCAAACTGTGCCACACATAAT--GCCTCTACGTGACAAGTCGCAACATTGCAACATTTTCAGTGCTATAAA    79
BsegSD   ---------------------------------------------------------------------------------
BtriSD   ---------------------------------------------------------------------------------
BrcuSD2  ---------------------------------------------------------------------------------
BrcuSD1  ---------------------------------------------------------------------------------
BseSD    ---------------------------------------------------------------------------------
LewiSD   --ATGCACTAGCAAATTAGGGCGTGCTCAAATGGAGACTAATGACCCATACTCACCCTCCCCACAATTTGGTTGCTATAAA    79

                                                          *  *** ***************** ****
BsySD    AGGCCAATGGGGTTCCACAGAGAAGTCACATCCACACTCAACATCATAGCAACACATTCAGATACTCGGAGGGAAGAACTC   159
BsegSD   ------------------------------------------------------CACATTCAGATACTCGGAGGGAAGAACTC    29
BtriSD   ------------------------------------------------------CACATTCAGATACTCGGAGGGAAGGACTC    29
BrcuSD2  -----------------------------------------------------CTTCACATTCAGATACTCGGAGGGAAGTACTC    32
BrcuSD1  -----------------------------------------------------CTTCACATTCAGATACTCGGAGGGAAGCACTC    32
BseSD    ------------------------------------------------------CACATTCAGATACTCGGAGGGAAGAACTC    29
LewiSD   AGACCATCGGGGTGGGATTGAGGGGTTACTCCAACACTCAGCA-CACCAGAAGATTTTC-GATACTCGGAGGGAAGAACTC   157

         ********** ***  *
BsySD    TTCAGCATCAACAGTAACA   177
BsegSD   TTCAGCATCAACAGTCACC    47
BtriSD   TTCAGCATCAACAGTCACC    47
BrcuSD2  TTCAGCATCAACAGCAAAC    40
BrcuSD1  TTCAGCATCAACAGCAAAC    40
BseSD    TTCAGCATCAACAGTAACC    47
LewiSD   TTCAGCATCATCAGCAAAC   175
```

Figure 3.2. Sequences of the single-domain globin from all polynoid species upstream of the first codon. Part of the promoter region for *B. symmytilida* and *L. williamsae* is shown, and part of the 5'UTR for the rest of the species. The TATA box is shaded in grey.

Introns

Introns were amplified and sequenced in all species but *Harmothoe* sp. and intron 2 in *B. trifurcus*. Intron sequence length differed considerably, especially for intron 1 with sequence lengths ranging from 306 bp in *B. symmytilida* to 743 bp in *B. seepensis*. Intron 2 sequence length was also variable but with smaller differences, from 180 bp in *L. williamsae* to 295 bp in *B. seepensis*. The alignment between all orthologous intron sequences revealed limited identity. Within each genus for which we have two species (i.e.

Branchipolynoe, and *Branchinotogluma*), however, the identity is quite high for at least a portion of the sequence.

Some transcription regulatory signals were found in some of intron 1 sequences of all species except *B. symmytilida* and *L. williamsae*. A Sp1 regulatory signal was found in *B. seepensis*, a CCAAT box also in *B. seepensis* and *B. segonzaci*, *B. trifurcus* and *B. cupreus* SD1. GATA motifs were only found in *B. segonzaci* and in both forms of *B. cupreus* SD globin intron 1. A putative HSE was also found in *B. cupreus* SD2. In intron 2 a few transcription regulatory signals were also found. *B. cupreus* SD2 does not have any signal. In *B. symmytilida*, *B. seepensis* and *L. williamsae* we found probable motifs for GATA-1. In *B. cupreus* SD1 and *B. segonzaci* we observed a motif for GATA. In this latter one we also found a CACBP motif.

Amino acid sequences and protein structure

The SD sequences obtained here were aligned with other annelid globins (intra- and extra-cellular), two extracellular hemoglobin sequences from the nematode *Ascaris suum* (pig intestinal parasite) and the myoglobin from *Physeter catodon* (sperm-whale), as a vertebrate reference (Fig. 3.3).

The alignment shows only two conserved residues, the phenylalanine CD1F (first position of the corner between helix C and helix D, using *Physeter* myoglobin fold as reference hereafter) and the proximal histidine F8H on which the heme is bound. The tryptophan in position A12 was conserved in all the species except for *Ascaris*. All invertebrate sequences also have a conserved tryptophan (H8W) that is not found in *Physeter* myoglobin. The two conserved cysteines in the typical extracellular globins from annelids (positions A2 and H11) are not present in the Polynoidae globins. Over the section for which we had a sequence overlap (107 residues), the Polynoidae sequences showed an identity of ~76%. Several important amino acids in the heme pocket show interesting features. Two important residues that are known to stabilize the binding of oxygen on the heme group in *Ascaris*, tyrosine B10 and glutamine E7, are also present in all the Polynoidae sequences. The pogonophoran annelid *O. mashikoi* also possesses a glutamine in E7. Among the polynoid sequences, out of the 30 probable heme contacts (sperm-whale myoglobin heme contacts as a reference) (Fig. 3.3), only 6 residues positions are affected by changes (80% identity). These changes however, do not significantly affect the hydrophobic clusters in this region of the protein.

Figure 3.3. Alignment of globin sequences from annelids, nematodes and a vertebrate. Polynoidae single-domain globin sequences are shaded. Conserved residues are shown in bold (CD1F and F8H), heme pocket residues that explain the high O_2 affinity in *Ascaris* are boxed in the Polynoidae. The cysteines forming an intrachain disulfide bridge in typical extracellular annelid globins (A2C and H11C) are underlined. Intron (I1 and I2) conserved positions shown above the sequences. Bsy: *B. symmytilida*; Bse: *B. seepensis*; Bseg: *B. segonzaci*; Btri: *B. trifurcus*; Brcu: *B. cupreus*; Lewi: *L. williamsae*; Harmo: *Harmothoe* sp.; SD: single-domain; D1-D2: single-domain globin type; AacuNg: *Aphrodite aculeata* neuroglobin; Gly: *Glycera* sp.; Lumt: *Lumbricus terrestris*; Tubifex: *Tubifex tubifex*; Phese: *Pheretima seiboldi*; Tylo: *Tylorhynchus heterochaetus*; Rifb: *Riftia pachyptila*; Lam: *Lamellibrachia* sp.; Asuum: *Ascaris suum*; Omas: *Oligobrachia mashikoi*; Phyca: *Physeter catodon*.

Hydrophobic Cluster Analysis (HCA) revealed that the hydrophobic clusters are quite similar between polynoid sequences (HCA homology ranging between ~80-99%). These clusters clearly differ from those for *Ascaris* and *Lumbricus terrestris* sequences (Fig. 3.S1, supplementary material).

No signal peptide was found in any of the species for which we obtained sequences upstream of the initial methionine (all species but *Harmothoe* sp.).

Single-domain relationship with other annelids

The Bayesian phylogenetic tree (Fig. 3.4) shows a clear dichotomy between the globins involved in giant extracellular hexagonal bilayer hemoglobins (HBL-Hbs) on one hand, and the myoglobins and circulating intracellular hemoglobins on the other hand. The polynoid SD globins clearly cluster together, as a sister-group to the four domains from the tetradomain globins previously sequenced in *Branchipolynoe symmytilida* and *B. seepensis*. Interestingly, although expressed as extracellular proteins in the coelomic fluid, these globins more closely cluster with the intracellular group. Among the SD sequences some cases of politomy were observed, with an internal node with a relatively weak posterior probability (0.63). Some internal branches however have a strong support.

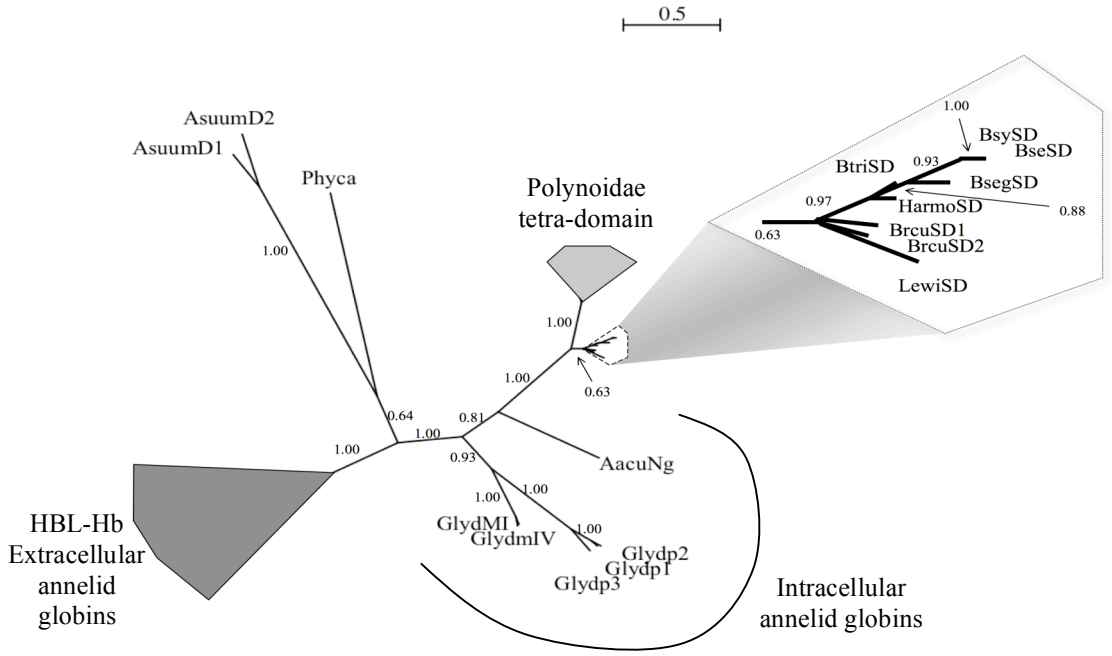


Figure 3.4. Bayesian phylogenetic tree based on annelid globins corresponding to the alignment in figure 3.3 (for details see legend of figure 3.3). The zoomed area represents the Polynoidae single-domain globins. Posterior probability values are indicated above or near the respective branch.

Variability of d_N/d_S ratios and search for positive selection

Among lineages

The search for positive selection was performed in the context of species phylogeny based on a partial sequence of the mitochondrial 16S (mt16S) gene (Figure 3.5; Hourdez et al. in prep.). We obtained two different single-domain globins for *Branchiplicatus cupreus*, and the following results were obtained with SD1. The same analyses were performed with SD2 and produced very similar results.

The LRT between the *one-ratio model* and the *free-ratio model* was significant, the latter one being more likely (LRT = 30.38, df = 5, $p > 0.001$). According to this result, the ω (d_N/d_S) ratios are significantly different among lineages (Yang 1998). For some branches, however, ω ratios greater than 1 were due to the absence of synonymous substitutions. When these values were associated to less than one non-synonymous substitution, we did not consider the branch for further analyses as this result can be merely due to chance. As a consequence, only three branches were chosen to test for potential differences in ω ratios: the branch leading to the two commensal species *Branchipolynoe symmytilida* and *B. seepensis* (hereafter referred to as branch *a*), the branch leading to *Branchinotogluma trifurcus* (hereafter referred to as branch *b*), and the branch linking *Harmothoe* sp. to all the other species (hereafter referred to as branch *c*) (see Fig. 3.5).

Different sets of statistical tests under PAML indicated that branches *a* and *c* (but not *b*) possess a ω ratio that is significantly different from the background value (ω_0). The $\hat{\kappa}$ values (transition/transversion rate ratio) were very similar between the different models and between the different sets, ranging from 2.07 to 2.11. Under the one ratio model ω_0 is smaller than one, indicating some purifying selection. In all the tests, branch *a* possesses a ω ratio greater than that of other branches (close to 1, or slightly greater than 1). Branches *b*, and *c* have an infinite ratio due to the absence of synonymous changes on these branches. As these ratios were calculated over the whole sequence, some amino acid positions can be under positive selection, and the average ratio may mask this positive selection. Only branch *a* was tested for this as it alone has a ratio close to 1 and a number of synonymous changes that was not null.

Among sites

We performed a branch-site model test (Yang and Nielsen 2002) on branch *a* that

revealed that the ω ratio significantly differed between amino acid positions (LRT = $2\Delta\ell$ = 29.2, df = 2, $p < 0.001$). This was confirmed when we tested for positive selection, by forcing $\omega_1 = 1$ in the MA (see Yang and Nielsen 2002) (LRT = $2\Delta\ell$ = 24, df = 2, $p < 0.001$). From this latter analysis, we identified the amino acid positions that are under positive selection (BEB; the posterior probabilities are given in parentheses and asterisks indicate significant positions): 9 V (0.927), 13 H (0.816), 16 Q (0.75), 30 S (0.978*), 42 T (0.994**), 68 S (0.998**) and 106 G (0.737).

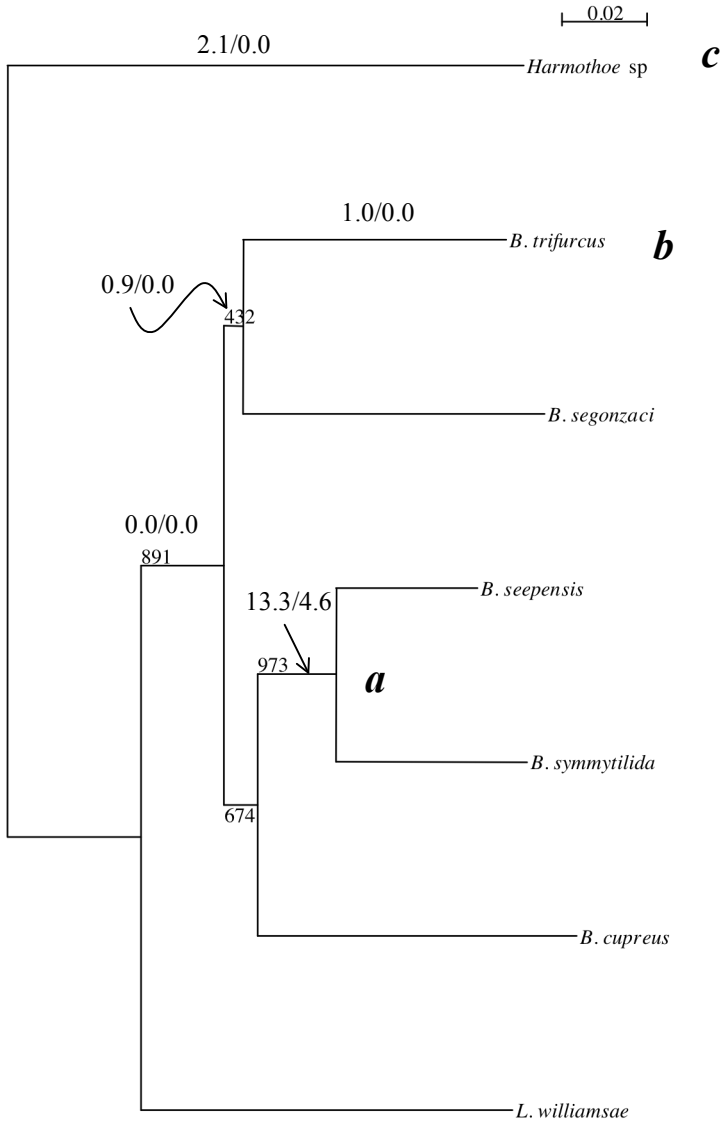


Figure 3.5. Phylogeny of a subset of seven polynoid species based on 515b bp portion of the mitochondrial 16S gene. Bootstrap values are indicated above/under the respective branches (1000 replicates). Ratios indicate the maximum likelihood estimates of the numbers of nonsynonymous over synonymous substitutions for the entire gene along the branch where they are located. ***a***, ***b*** and ***c*** represent the lineages chosen for the *branch-site model* test (see results).

Homology modeling – 3D approach

Homology models were created only for the species for which we had a complete sequence, *B. symmytilida* and *B. segonzaci* (Fig. 3.6). For both species the automatically chosen template sequence was the monomer chain of the hemoglobin from *Lumbricus terrestris* (available with a high resolution), that had ~20% of sequence identity with our sequences. The residues responsible for the high oxygen affinity in *A. suum* hemoglobin (B10Y and E7Q), and also found in the polynoid globin sequences, are highlighted on the structural models. Both residues point towards the distal side of the heme group. On the proximal side, F8H, crucial for the attachment of the heme appears with its expected orientation. The amino acids that were significantly identified as being under positive selection in the *Branchypolynoe* branch (positions E11, CD5 and F6, Fig. 3.5) are also shown on both models for comparison. E11T in *Branchipolynoe* spp. (E11V in *B. segonzaci*) is also located in the distal region of the heme pocket, and points in the same direction as E7Q and B10Y. The two remaining amino acids under positive selection, CD5S and F6S, in *Branchipolynoe* spp. (CD5G and F6Q in *B. segonzaci*) are located in helix regions that, in other annelid globins, are important for the formation of oligomers (formation of dimers by interaction of helices E and F; Royer et al. 2005). Both of these models were structurally compared to the 3D model of the *A. suum* hemoglobin (data not shown) and showed that the residues in the heme pocket have a very similar structural position.

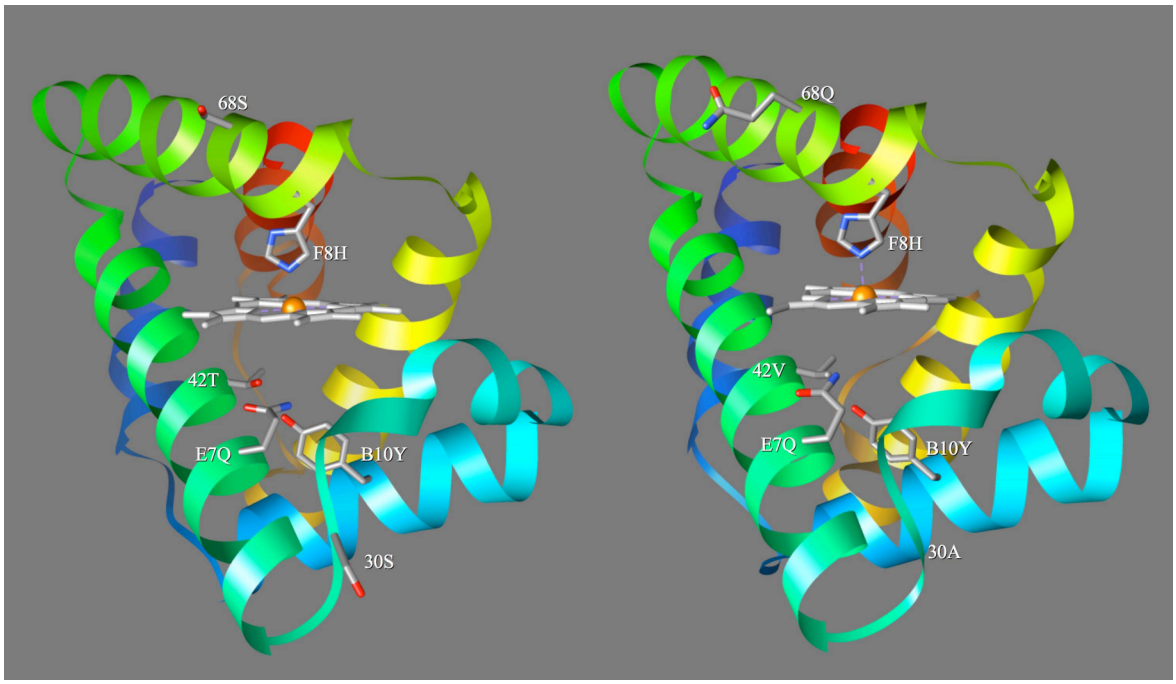


Figure 3.6. 3D structural model of *B. symmytilida* (left) and *B. segonzaci* (right) single-domain globin, based on *Lumbricus terrestris* 3D structure. Amino acid residues that are important in the heme pocket (B10Y, E7Q and F8H in both species) are shown. The three positions that are under positive selection in *B. symmytilida* (30S, 42T and 68S) are also shown. In *B. segonzaci*, residue 30 is an alanine, with only a hydrogen as a side chain, not visible here.

3.2.3 Discussion

Globins from invertebrates exhibit a great range of structural and functional diversity (Weber and Vinogradov 2001). This diversity is most likely the result of adaptations at the molecular level to the contrasted environments that these invertebrates inhabit. Hydrothermal vents are very challenging for aerobic organisms, especially in regard to hypoxia and the presence of sulfide (a potent inhibitor of aerobic metabolism) (Carrico 1978, Childress and Fisher 1992). The scale-worm species studied here represent a range of hydrothermal vent conditions, and different life styles (free-living or commensal). Such challenging conditions could lead to functional innovations (evidenced by positive selection) essential for the survival of the species.

Hemoglobin expression in vent species

Endemic hydrothermal vent polynoids typically possess extracellular hemoglobins in their coelomic fluid that confer it a red color. The sheer presence of hemoglobins in these species can be regarded as an adaptation to hypoxic conditions as these proteins represent a form of storage that will buffer variations of external oxygen concentrations. It was calculated that in *Branchipolynoe seepensis*, the amount of oxygen bound on the hemoglobins represents about 1h30 of aerobic metabolic needs (Hourdez and Lallier 2007). Although extracellular single-domain globins exist in all hydrothermal vent endemic species, tetra-domain globins were only detected in the genera *Branchypolynoe* and *Branchinotogluma* (Hourdez et al. in prep). The phylogenetic relationships indicate that all the studied polynoid extracellular globins (single- and tetra-domain) have a common evolutionary history, with a probable origin from an intracellular ancestor. These extracellular globins have a distinct evolutionary path from the other annelid extracellular globins that diverged from the intracellular ones about 570 millions years ago (Goodman et al. 1988).

Although extracellular, all the sequenced globins lack a signal peptide (except in *Harmothoe* sp. for which we have no sequence data in that region). This was already observed in the single- and tetra-domain globin from *Branchipolynoe seepensis* and *B. symmytilida* (Projecto-Garcia et al. in press) who suggested that the secretion could be holocrine (release of the components dissolved in the cytoplasm after the rupture of the

cell membrane). This hypothesis is corroborated by the fact that in another vent scale-worm *Lepidonotopodium piscesae*, the molecular weight of the intracellular (myoglobin) and extracellular (coelomic) globins are exactly the same (Hourdez, unpubl. data). It is also possible that we actually amplified the intracellular globin gene and not the extracellular single-domain globin gene. This however is unlikely as, in addition to the observations made on *L. piscesae*, none of the seven sequences we obtained possessed a signal peptide. This number includes 2 different globins in *B. cupreus* and the odds of only amplifying one type of globin are small.

Interestingly, the 5'UTR is very well conserved in all the vent species, and the promoter region also displays areas that are conserved between *B. symmytilida* and *L. williamsae*, the two most distantly related vent species. This may indicate some structural or regulatory function(s) for these regions.

Amino acid positions under positive selection

All the single-domain globin sequences of polynoids exhibit conserved amino acid residues (B10Y and E7Q) that are known to be responsible for the high oxygen affinity in the *Ascaris suum* globins (pig intestinal parasite). This high affinity is mostly due to the low oxygen dissociation rate that they provide (Davenport 1949 in Peterson et al. 1997, De Baere et al. 1994, Peterson et al. 1997). The replacement of the conserved distal histidine (E7H) by a glutamine (E7Q) and the B10L by a tyrosine (B10Y) seems common in many invertebrate globins (Weber and Vinogradov 2001). Not all invertebrate globins however possess the same oxygen affinity, depending mostly on the heme pocket conformation (Peterson et al. 1997).

The construction of a homology model of two polynoid globins, *B. symmytilida* and *B. segonzaci*, indicates that the B10Y and E7Q point towards the heme group, suggesting that these residues are likely responsible, like in *A. suum*, for the high oxygen affinity measured for *Branchypolynoe* hemoglobins (Hourdez et al. 1999b). However, in *A. suum* the affinity does not depend only on the presence of these residues, but also on the tightness of the heme pocket, making the oxygen dissociation rates much lower than in other invertebrate globins with the same residues in B10 and E7 (Petersen et al. 1997).

Some amino acid residues in globins from annelids living in high-sulfide habitats, such as hydrothermal vents, are under Darwinian positive selection (Bailly et al. 2003). In the *Branchipolynoe* lineage some important amino acids (CD5S, E11T and F6S) were found to

be under positive selection. Among these residues only the threonine (E11T) is located in the heme pocket. The relationship of this residue with E7Q and B10Y is not clear, but the hydroxyl group of E11T could form a hydrogen bond with either the dioxygen (thereby stabilizing the bound ligand) or a water molecule, keeping this molecule from interacting with the bound dioxygen, as in the globin from the bivalve *Lucina pectinata* that also lives in sulfidic habitats (Petersen et al. 1997). The equivalent residue in *B. segonzaci*, a valine, does not possess a hydroxyl group and cannot form these interactions.

The two remaining residues, located in CD5 and F6, could affect subunit interactions in *Branchipolynoe* single-domain globin. The dimer interactions in *Lumbricus terrestris* hemoglobin are established through residues in the E and F helices (Royer et al. 2000), an interaction in which F6S is likely to participate. In *L. terrestris*, dimers form tetramers, mainly by the interaction of the loop formed by the AB corner. CD5S is also located in a corner that could be involved interactions to form a multimer. The assembly of globin subunits can be quite beneficial in the way that through cooperativity (homotropic interactions between different heme groups that improve O_2 loading and unloading) and the Bohr effect (heterotropic interactions involving inhibition between hemes, for example) the uptake and release of O_2 can be optimized, depending on pH and O_2 partial pressure (Weber and Vinogradov 2001). The formation of multimeric assemblages may also be beneficial as these hemoglobins are extracellulalr and larger molecular weights avoid excretion (Weber and Vinogradov 2001). These results call for a study of the protein quaternary structure through crystallography, as well as functional properties.

Positive selection and molecular innovation

Only the *Branchipolynoe* lineage showed positively selected amino acid positions compared to the rest of the polynoid single-domain globins, located in the heme pocket and at the surface of the protein. All the known *Branchipolynoe* species live in the mantle cavity of mussels *Bathymodiolus* (Chevaldonné et al. 1998). The mussels rely on symbiotic chemoautotrophic or methanotrophic bacteria for at least part of their nutrition (Childress and Fisher 1992). These bacteria oxidize sulfide and/or methane, and the mussels need to flow water laden with these compounds to meet the bacteria's metabolic needs. This water therefore hypoxic as it contains a fair proportion of anoxic hydrothermal fluid. The commensal life-style likely more often exposes the worms to hypoxia when the mussels

close their valves completely for an undetermined period of time. With so many constraints for aerobic respiration, having hemoglobins with high oxygen affinity (Hourdez et al. 1999b) is most likely advantageous. The other species of scale-worms studied here are free-living and, although the hydrothermal vent environment is generally hypoxic (Hourdez and Lallier 2007), they can probably change location when exposed to conditions that are not adequate. Although *B. segonzaci* lives on the chimney walls, along with *Paralvinella fijiensis* in the Lau Basin where sulfide concentrations can be high and oxygen low, the chaotic mixing of the hydrothermal fluid and the deep-sea water transiently exposes the animals to the influence of oxygenated water.

We surprisingly did not find any evidence of positive selection that would reflect an adaptation to hydrothermal vent conditions in the branch that leads to the endemic hydrothermal polynoid species (all except *Harmothoe* sp.). This could be due to the fact that the main adaptation to the vent conditions resides in the expression of large amounts of globins in the coelomic cavity where it can be used as an oxygen storage, for periods of strong hypoxia and complete anoxia.

The two main hemoglobins (composed of tetradomain globins) found in the coelomic fluid of *Branchipolynoe* display different sensitivity to CO_2 (Hourdez et al. 1999b). This is reminiscent of 'class II' fish in which the different hemoglobins found in the eythrocytes have different functional properties and sensitivities to effectors that reflect a division of labor. In *Branchipolynoe*, this division of labor may be extended to the single-domain globins, also found in the coelomic fluid. After duplication of a gene, selective pressure can be lifted on some proteins that may then acquire novel functions. The detection of some amino acids under positive selection most likely corresponds to such an acquisition.

Acknowledgments

The authors would like to thank the crews of the ships and submersibles, and the chief scientists, of the cruises ATOS 2001 (project funded by Ifremer and INSU), Lau Basin (projects funded by two NSF grants to C.R. Fisher (NSF OCE 0240985 and NSF OCE 0732333)), and EPR 2001 (project funded by a NSF grant to C.R. Fisher (NSF OCE-0002729)). We would also like to thank I. Boutet-Tanguy and A. Tanguy for technical advice, M. Bruneaux, A. Bessadok, and M. Czjzek for protein modeling advice. This work is part of the project HYPOXEVO (Région Bretagne), and was supported by the ESTeam research Marie Curie grant under the 6th framework program from the EC.

References

Anisimova M, Bielawski JP, Yang Z. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585-1592.

Bailly X, Leroy R, Carney S, Collin O, Zal F, Toulmond A, Jollivet D. 2003. The loss of the hemoglobin H_2S-binding function in annelids from sulfide-free habitats reveals molecular adaptation driven by Darwinian positive selection. P Natl Acad Sci USA. 100:5885-5890.

DeLano, Warren L. The PyMOL Molecular Graphics System. 2009. DeLano Scientific LLC, Palo Alto, California, USA. http://www.pymol.org

Doyle JJ and Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15.

Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat, B, Mornon, J-P. 1997. Deciphering protein sequence information through Hydrophobic Cluster Analysis (HCA): current status and perspecrives. Cell Mol Life Sci 53:621-645.

Carrico RJ, Blumberg WE, Peisach J. 1978. The reversible binding of oxygen to sulfhemoglobin. J Biol Chem 253:7212-7215.

Chevaldonnée P, Jollivet D, Feldman RA, Desbruyères D, Lutz RA, Vrijenhoek RC. 1998. Commensal scale-worms of the genus Branchiolynoe (Polychaeta: Polynoidae) at deep-sea hydrothermal vents and cold seeps. Cah Biol Mar 39:347-350.

Childress JJ and Fisher CR. 1992. The biology of hydrothermal vent animals: physiology, biochemistry, and autotrophic symbioses. Oceanogr Mar Biol - An Annual Review 30:337-441.

Cunniff NFA, Wagner J, Morgan WD. 1991. Modular recognition of 5-base-pair- DNA sequence motifs by human heat shock transcription factor. Mol Cell Biol 11:3504-3514.

Davenport, HE. 1949. Ascaris Haemoglobin as an indicator of the oxygen produced by isolated chloroplasts. P R Soc London B 136:281-290.

De Baere I, Perutz MF, Kiger L, Marden MC, Poyart C. 1994. Formation of two hydrogen bonds from the globin to the heme-linked oxygen molecule in *Ascaris* hemoglobin. P Natl Acad Sci USA. 91:1594-1597.

DeLano, W.L. 2008. The PyMOL Molecular Graphics System. DeLano Scientific LLC, Palo Alto, CA, USA

De Maria R, Zeuner A, Eramo A, Domenichelli C, Bonci D, Grignani F, Srinivasula SM, Alnemri ES, Testa U, Peschle C. 1999. Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1. Nature 401:489-493.

Dewilde S, Blaxter M, Hauwaert, M-L, Vanfleteren J, Esmans EL, Marden M, Griffon N, Moens L. 1996. Globin and Globin Structure of the Nerve Myoglobin of *Aphrodite aculeata*. J Biol Chem 271:19865-19870.

Doyle JJ and Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15.

Gaboriaud C, Bissery V, Benchetrit T, Mornon JP. 1987. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett, 224: 149-155.

Goodman M, Pedwaydon J, Czelusniak J, Suzuki T, Gotoh T, Moens L, Shishikura F, Walz D, Vinogradov SN. 1988. An evolutionary tree for invertebrate globin sequences. J Mol Evol 27:236-249.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947-2948.

Hourdez S, Lallier FH, Green BN, Toulmond A. 1999a. Hemoglobins from deep-sea hydrothermal vent scale-worms of the genus *Branchipolynoe*: A new type of quaternary structure. Proteins. 34:427-434.

Hourdez S, Lallier FH, Martin-Jézéquel V, Weber RE, Toulmond A. 1999b. Characterization and functional properties of the extracellular coelomic hemoglobins from the deep-sea, hydrothermal vent scale-worm *Branchipolynoe symmytilida*. Proteins. 34:435-442.

Hourdez S and Lallier F. 2007. Adaptations to hypoxia in hydrothermal-vent and cold-seep invertebrates. Rev Environ Sci Biotechnol 6:143-159.

Hourdez S. and Weber RE. 2005. Molecular and functional adaptations in deep-sea hemoglobins. J Inorg Biochem 99:130-141.

Huelsenbeck J P and Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Krantz SB. 1991. Erythropoietin. Blood 77:419-434.

Kulozik AE, Bellan-Koch A, Bail S, Kohne E, Kleihauer E. 1991. Thalassemia intermedia: moderate reduction of beta globin gene transcriptional activity by a novel mutation of the proximal CACCC promoter element. Blood 77: 2054-2058.

Nielsen R and Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929-936.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem 25:1605-1612.

Peterson ES, Huang S, Wang J, Miller LM, Vidugiris G, Kloek AP, Goldberg DE, Chance MR, Wittenberg JB, Friedman JM. 1997. 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 36:13110-13121.

Projecto-Garcia J, Zorn N, Didier J, Shaeffer SW, Lallier FH and Hourdez S. (in press).Origin and evolution of the unique tetra-domain hemoglobin from the hydrothermal vent scale-worm Branchipolynoe. Mol Biol Evol (doi: 10.1093/molbev/msp218).

Ronquist F. and Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Royer Jr WE, Strand K, van Heel M, Hendrickson WA. 2000. Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. P Natl Acad Sci USA 97:7107-7111.

Royer WE, Zhu H, Gorr TA, Flores JF, Knapp JE. 2005. Allosteric hemoglobin assembly: Diversity and similarity. J Biol Chem 280:27477-27480.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599.

Tunnicliffe, V. (1991). "The Biology of Hydrothermal Vents: Ecology and Evolution." Oceanography and Marine Biology: An Annual Review 29: 319-407.

Sambrook J and Russell DW. 2001. Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Weber RE. 1978. Respiratory Pigments. Physiology of Annelids. Mill, P. J. London, Academic Press Inc.

Weber RE and Vinogradov SN. 2001. Nonvertebrate hemoglobins: functions and molecular adaptations. Physiol Rev 81:569-628.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568-573.

Yang Z and Nielsen R. 2002. Codon-substitution models for detecting molecular adaptations at individual sites along specific lineages. Mol Biol Evol 19:908-917.

Zon L I, Youssoufian H, Mather C, Lodish HF, Orkin SH. 1991. Activation of the erythropoietin receptor promoter by transcription factor GATA-1. P Natl Acad Sci USA 88:10638-10641

Suplementary material

Tables

Table 3.S1. Sequences of the primers used to amplify the different parts of the single-domain (SD) hemoglobin gene from the several speces from Polynoidae. Bsy: Branchipolynoe symmytilida, Brcu: *Branchiplicatus cupreus*, Bng: *Branchinotogluma*, Bseg: *Branchinotogluma segonzaci*, Lewi: *Lepidonotopodium williamsae*, CW: chromosome walking, E: exon, I: intron.

Primers	5' Sequence 3'	Localization/Function
93F	CGG GAG TGG ACC TTC AGG C	E1 from BsySD (forward), for *Branchipolynoe*
93R	GTT ACG GCT TGA TTT TAT TAC	3'UTR from BsySD (reverse), for *Branchipolynoe*
93F2	CTC TAC GGA GTC ATC GCT G	Oligo in promoter sequence in BsySD (forward)
93F3	CGT CAC ACT GGG TAT GGT GC	Oligo in promoter sequence in BsySD (forward)
93R2	CCT GAA GGT CCA CTC CCG	Oligo in promoter sequence in BsySD (reverse)
93R3	CAT CTG GGG CAT TAG CCT GC	Oligo in promoter sequence in BsySD (reverse)
5'UTRMb_F	CAC ATT CAG ATA CTC GGA GGG	Oligo in 5'UTR from BsySD (reverse), for Polynoidae
MbPolF	ATC AMG AGC TCM TGG NC	Oligo in E1 from BsySD (forward), for Polynoidae
MbPolR	CCA NGC ATC CTT GGC AGC TGG	Oligo in E3 from BsySD (reverse), for Polynoidae
BrcuSD1Int1R2	CTC TGC TCT TGG AAT ACT AGC	Oligo in I1 from BrcuSD, type 1 (reverse)
BrcuSD2Int1R2	CTA TCG TTC TAC ATT CTG CCC	Oligo in I1 from BrcuSD, type 2 (reverse)
BngSD_E1_F1	AGT GGA CCT TCA GGC YG	Oligo in E1 from BngSD (forward), for *Branchinotogluma* spp.
BngSD_E2_F2	GCT AAT GCC CCC GAT GCT TAT	Oligo in E2 from BngSD (forward), for *Branchinotogluma* spp.
BngSD_E2_R5	ATA AGC ATC GGG GGC ATT AGC	Oligo in E2 from BngSD (reverse), for *Branchinotogluma* spp.
BngSD_E2_F3	TGC CAA GAA GGT CCA YTT TGG	Oligo in E2 from BngSD (forward), for *Branchinotogluma* spp.
BngSD_E2_R4	CCA AAR TGG ACC TTC TTG GCA	Oligo in E2 from BngSD (reverse), for *Branchinotogluma* spp.
BngSD_E2I2_R3	ATA TTS ACT CAC CCC AAA RTG G	Oligo in I2 from E2 from BngSD (reverse), for *Branchinotogluma* spp.
BngSD_E3_F5	GAC CMT GCC TGC TYG CA	Oligo in E3 from BngSD (forward), for *Branchinotogluma* spp.
BngSD_E3_R2	CWG CRA GCA GGC AKG GTC	Oligo in E3 from BngSD

Primers	5' Sequence 3'	Localization/Function
		(reverse), for *Branchinotogluma* spp.
BngSD_E3_F6	GCW CTG GCY GAG GTC TGT	Oligo in E3 from BngSD (forward), for *Branchinotogluma* spp.
BngSD_E3_R1	AAC TTT CCA CCA CAG ACC TCR G	Oligo in E3 from BngSD (reverse), for *Branchinotogluma* spp.
BsegSD1_I1_F1	GTC TTC AAA TAT GCT CTG CTT TGT CCC	Oligo in I1 from BsegSD, type 1 (forward)
BsegSD1_I1_F2	TTA GTC TGT ATC CTA GCG TGT CTT CC	Oligo in I1 from BsegSD, type 1 (forward)
BsegSD2_E2_R1	CCT TGA CAA CTC CAT CAA TGA AGG TC	Oligo in I1 from BsegSD, type 2 (reverse)
BsegSD2_E2_F1	GAC CTT CAT TGA TGG AGT TGT CAA GG	Oligo in E2 from BsegSD, type 2 (forward)
LewiSDInt1R2	GAA ATA AAA ATT AGG GTG CTA ACG	Oligo in I1 from LewiSD(reverse)
Oligo $(dT)_{18}$	TTT TTT TTT TTT TTT TTT T	Upstream amplification from cDNA
Anchored OligodT	CTC CTC TCC TCT CCT C(T)17	Upstream amplification from cDNA
Walker 1 (W1)	CTA ATA CGA CTC ACT ATA GGG NNN NAT GC	CW
Walker common (WC)	CTA ATA CGA CTC ACT ATA GGG	CW

Table 3.S2. Different PCR amplification conditions for the different polynoid species.

Species	Target gene part	PCR conditions	Primer pair
Branchipolynoe symmytilida	Whole gene	30 cy: 94ºC(1')/55ºC(1')/72ºC(1'); 1.5 mM $MgCl_2$	93F/93R
Branchipolynoe seepensis	1/2 I1+E2+I2+E3	30 cy: 94ºC(1')/52ºC(1')/72ºC(1'); 1.5 mM $MgCl_2$	93F/93R
	5'UTR+whole gene+3'UTR	10 cy: 94ºC(30")/52ºC(3')/72ºC(2')+ 30 cy: 94ºC(30")/52ºC(1')/72ºC(1'); 2 mM $MgCl_2$	5'UTRMb/93R3
Branchiplicatus cupreus	Whole gene	30 cy: 94ºC(1')/55ºC(1')/72ºC(1'); 1.5 mM $MgCl_2$	MbPolF/MbPolR
	5'UTR+E1+1/2 I1	30 cy: 94ºC(1')/55ºC(1')/72ºC(1'); 1.5 mM $MgCl_2$	WC/BrcuMb1I1R2 & WC/BrcuMb2I1R2
Branchinotogluma segonzaci	1/2 E1+E2+1/2 E3	35 cy: 94ºC(1')/55ºC(1')/72ºC(1'); 1.5 mM $MgCl_2$ **nested** 10 cy: 94ºC(30")/52ºC(3')/72ºC(2')+ 30 cy: 94ºC(30")/52ºC(1')/72ºC(1'); 2 mM $MgCl_2$	MbPolF/MbPolR **nested** 93F/MbPolR
	1/2 E2+I2+1/2 E3	35 cy: 94ºC(1')/55ºC(1')/72ºC(1'); 1.5 mM $MgCl_2$ **nested** 10 cy: 94ºC(30'')/52ºC(3')/72ºC(2') + 30 cy: 94ºC(30")/52ºC(1')/72ºC(1'); 2 mM $MgCl_2$	MbPolF/MbPolR **nested** 93F3/MbPolR
	final part of E3+3'UTR	10 cy: 94ºC(30'')/52ºC(3')/72ºC(2') + 30 cy: 94ºC(30")/52ºC(1')/72ºC(1'); 2 mM $MgCl_2$	BngSD_E3_F6/AncredT
	E3+3'UTR	10 cy: 94ºC(1')/50ºC(1')/72ºC(1')+ 25 cy: 94ºC(1')/55ºC(30")/72ºC(1'); 2 mM $MgCl_2$	BngSD_E2F3/AncredT **nested** BngSD_E3F5/AncredT

Species	Target gene part	PCR conditions	Primer pair
	E3+3'UTR	10 cy: 94ºC(1')/50ºC(1')/72ºC(1')+ 25 cy: 94ºC(1')/55ºC(30")/72ºC(1'); 2 mM $MgCl_2$	BngSD_E2F3/Oligo dT **nested** BngSD_E3F5/Oligo dT
	5'UTR+E1+I1+beginning E2	10 cy: 94ºC(1')/50ºC(1')/72ºC(1')+ 25 cy: 94ºC(1')/55ºC(30")/72ºC(1'); 2 mM $MgCl_2$	W1/BsegSD2_E2R1 **nested** BngSD_E1F1/BngSD_E2R5
	1/2 I2+E2	10 cy: 94ºC(1')/50ºC(1')/72ºC(1')+ 25 cy: 94ºC(1')/55ºC(30")/72ºC(1'); 2 mM $MgCl_2$	BsegSD1_I1F/BngSD_E2I2R3 **nested** BsegSD1_I1F2/BngSD_E2R4
	5'UTR+E1+I1+beginning E2	10 cy: 94ºC(30")/52ºC(3')/72ºC(2')+ 30 cy: 94ºC(30")/52ºC(1')/72ºC(1'); 2 mM $MgCl_2$	5'UTRMb_F/BngSD_E2R5
	1/2 E2+I2+beginning E3	10 cy: 94ºC(1')/50ºC(1')/72ºC(1')+ 25 cy: 94ºC(1')/55ºC(30")/72ºC(1'); 2 mM $MgCl_2$	BngSD_E2F2/BngSD_E3R1 **nested** BsegSD2_E2F1/BngSD_E3R2
Branchinotogluma trifurcus	1/2 E1+E2+1/2 E3	10 cy: 94ºC(30")/52ºC(3')/72ºC(2')+ 30 cy: 94ºC(30")/52ºC(1')/72ºC(1'); 2 mM $MgCl_2$	93F/MbPolR
	1/2 I2+E3	10 cy: 94ºC(30")/52ºC(3')/72ºC(2')+ 30 cy: 94ºC(30")/52ºC(1')/72ºC(1'); 2 mM $MgCl_2$	93F3/MbPolR
	5'UTR+E1+I1+beginning E2	10 cy: 94ºC(30")/52ºC(3')/72ºC(2')+ 30 cy: 94ºC(30")/52ºC(1')/72ºC(1'); 2 mM $MgCl_2$	5'UTRMb_F/BngSD_E2R5
Lepidonotopodium williamsae		30 cy: 94ºC(1')/55ºC(1')/72ºC(1'); 1.5 mM $MgCl_2$	MbPolF/MbPolR

Species	Target gene part	PCR conditions	Primer pair
	Promoter regions+5'UTR+1/2 I1	30 cy:94ºC(1')/55ºC(1º)/72ºC(1') ; 1.5mM $MgCl_2$	WC/Lewi_I1_R2
Harmothoe sp.	1/2 E2+I2+ 1/2 E3	30 cy:94ºC(1')/50ºC(1º)/72ºC(1') ; 1,5mM $MgCl_2$	93F/MbPolR

Hydrophobic Cluster Analyses Plots

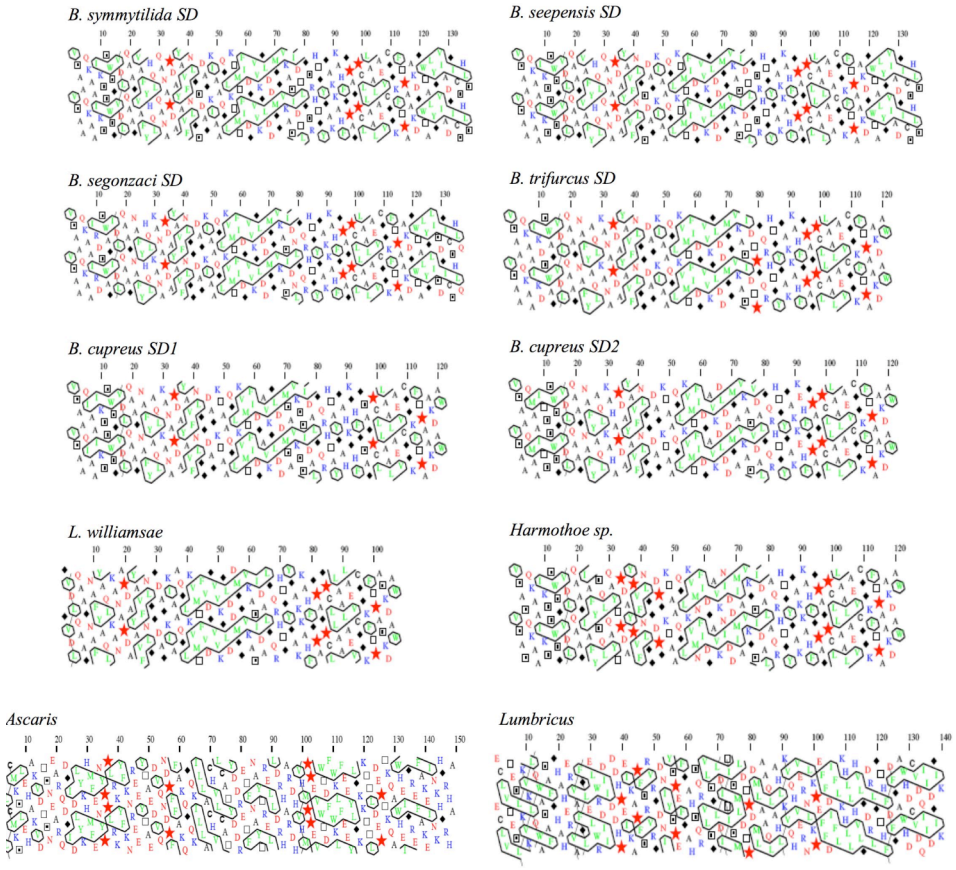


Figure 3.S1. Hydrophobic Cluster Analyses (HCA) results for the single-domain from all the polynoidae species used in this study, domain 1 from the di-domain globin from the nematode *Ascaris* and from the chain L from the polymeric *Lumbricus* globin.

3.3 General conclusions

Respiratory pigments have an important role in aerobic organisms living in hypoxic/anoxic environments such as hydrothermal vents, and globins are by far the most common ones, showing a great diversity among invertebrates (Hourdez and Weber 2005, Weber and Vinogradov 2001). All hydrothermal vent polychaetes studied to date possess hemoglobin(s) and these are expressed in great quantities (Hourdez and Lallier 2007). As hypothesized in the discussion of the previous section, this can be one of the major adaptations to hypoxia used by annelids in general and Polynoidae in particular. The effect of the high oxygen affinity, when compared to littoral congeneric species (Hourdez and Weber 2005), is then amplified when the globin concentration increases either in the vascular blood or in the coelomic fluid.

In the particular case of the *Branchipolynoe* species, we observed significant positive selection on one amino acid residue located in the heme pocket. This could mean that under almost anoxic conditions, when the *Bathymodiolus* mussels close their valves for an undetermined period of time, these polynoids take advantage of the more stable binding of the oxygen to the heme group, facilitated by the threonine, and extract oxygen from the pallial fluid, even at very low partial pressure. However, the mussel hosts are aerobic organisms and do not have any special mechanism to regulate the metabolic transport of gases (Kochevar et al. 1992), which means that they cannot remain "closed" for a long period of time, although their ability to rely on anaerobic metabolism has not been explored in depth (in particular how long). One possible scenario is that the positive selection observed in the single-domain Hb from *Branchipolynoe* is an adaptation to the hypoxia of the hydrothermal fluids filtered by the mussel but then you would expect to find the same type of selection for the free-living polynoid species studied here (e.g. *B. trifurcus* and *L. williamsae*). The alternative scenario is the division of labor for the different Hbs existing in the same organism, as observed in the 'class II' fish hemoglobins. Another example is the *Riftia* Hbs, where the different oxygen affinities could buffer the effects of variable oxygen tensions (Hourdez and Weber 2005).

Even though it was only found in the *Branchipolynoe* lineage, the result of positive selection in residues that can be involved in polymerization is quite exciting. In that lineage the tetra-domain Hb can form dimers or trimers (Hourdez et al. 1999a), but for the

other polynoids hydrothermal species there is no information on this subject. The location of these residues, 30S and 68S, in the DE corner and F helix, respectively, could strongly indicate that these residues either favor or prevent the formation of dimers in *Branchipolynoe* extracellular single-domain Hbs. This will have consequences on their possible cooperativity (Royer Jr et al. 2005). Such dimers are found in *Lumbricus terrestris* and the echinoderm *Caudina arenicola* Hbs, that are built with residues from the same E and F helix, what is commonly known as the 'EF dimer' structure (Mitchell et al. 1995, Royer Jr et al. 2005). In the echinoderm the dimmer interactions can be of several types, such as heme-F helix or E helix, E helix-E helix, E helix-F helix and B helix-F helix. In the nematode *Ascaris* sp. there is only E helix-E helix (Minning et al. 1995, Mitchell et al. 1995).

All these results start to shed some light on the characterization of the polynoid particular extracellular globins. The information obtained from this work corroborates the previous studied functional properties, regarding the high oxygen affinity and gives hints for the possible implications on the polymerization. The globin amino acid variability among the family most likely indicates different selective pressures according to the habitat (and life style) constraints.

4. Conclusion and perspectives

The major goal of this thesis was to study the evolution of respiratory adaptations in a group of annelids that is common at hydrothermal vents. We hoped that the species diversity, the different life-styles, and the various microhabitats occupied by these species would lead to various adaptations at a molecular level. Hydrothermal vent species of scale-worms, in contrast to their shallow-water relatives, express hemoglobin. This characteristic offers an obvious advantage to face the hypoxia that characterizes hydrothermal vents environment. In addition, the structure of some of these globins is unique in annelids. This family of proteins then appeared to be an ideal candidate to study the evolution of the respiratory adaptations in scale-worms.

We followed two complementary approaches to study these globins: the first one focused on the origin and evolutionary history of these globins, and the second focused on the structure-function relationships. Both approaches yielded results that allowed interesting discussions, and generated new hypotheses. Here I would like to go over the major conclusions of this work, and elaborate on their implications and on how future studies can help better understand the evolution of the globins.

4.1 Evolution of polynoids extracellular hemoglobins

The uniqueness of the polynoid extracellular single- and tetra-domain hemoglobins lead us to wonder about the origin of such globins and their corresponding genes. Although extracellular, the globins from hydrothermal vent species (single- and tetra-domain) exhibit a closer relationship to intracellular globins than to other extracellular annelid globins. Based on this closer relationship to intracellular globins, and the fact that these globins exhibit functional properties similar to myoglobins (high oxygen affinity and low cooperativity) (Hourdez et al. 1999b), a myoglobin-like globin was then considered as the most probable ancestor. All species from vent-endemic subfamilies of scale-worms possess at least one hemoglobin in their coelomic fluid. This strongly suggests that the common ancestor to all these species possessed at least one hemoglobin and the origin of this

ancestral form is at least 65 My old (estimated date of colonization of the vent environment by scale-worms) (Hourdez et al. in prep.). The tetradomain hemoglobin is only found in one lineage of vent species, likely indicating a younger origin for that molecule.

The duplication history of the tetra-domain gene, however, remains unresolved in regards of the order of gene duplication, but specially the mechanisms that it involved. Nucleotides and amino acid analyses showed different scenarios in the order of tandem duplications to form the same globin gene. A more likely scenario of tandem duplications was however established, where domain 1 is assumed to be ancestral giving rise to domain 2 and these two domains together duplicate once to give rise to domains 3 and 4. Our study did not allow us to exactly determine the duplication mechanism, most likely because the signatures of eventual crossing-over events were lost with the loss of the inter-domain introns. The phylogenetic uncertainty about the order of domain duplication may never be resolved because the successive duplication events seem to have occurred rapidly, leaving no detectable genetic signatures.

Other molecular statistical analyses may yield additional information, and possibly help us better understand the evolutionary history of the globins since their split from a myoglobin-like status. In particular, a parsimony analysis would be very interesting to compare to the maximum likelihood results. This analysis determines ancestral sequences at each node of a phylogeny. This would allow us to track which amino acid residues have changed in the course of the evolutionary history. This analysis applied to the two topologies reflecting the two possible scenarios of tandem duplication could help to decide which one is the most probable (i.e. the most parsimonious) evolutionary history for the tetra-domain globin gene.

4.2 Structure - function relationships

All the globins (single- and tetra-domain) we sequenced possessed some residues in the heme pocket that will most likely affect their functional properties (E7Q and B10Y). These key amino acids account for the high oxygen affinity in *Ascaris* hemoglobin (De Baere 1994). They are also found in the non-hydrothermal vent species *Harmothoe* sp. but are replaced by a histidine at the distal position (E7H), and a phenylalanine instead of the

tyrosine in B10 in *Aphrodita aculeata*. This suggests that these specific amino acids may not correspond to an adaptation to hydrothermal vents, and that the tissue globin also has a high affinity in non-vent species. The first and potentially most important adaptive value for the globins is probably their expression at very high levels in the coelomic cavity and in the muscles of the vent species.

We then performed some statistical molecular analyses to potentially detect amino acids that are under positive selection. This type of analyses reveals changes that are essential for genes to acquire novel functional or structural characteristics. This search did not reveal any amino acids under positive selective pressure in the branch leading to hydrothermal vent polynoid scale-worms, supporting the idea that the initial adaptive value of these globins resides in their high expression levels in the vent species. However, positively selected amino acids were detected in two lineages, one corresponds to the split of the first duplication in the tetra-domain globin, the second to the single-domain globin in *Branchipolynoe* spp.

4.2.1 Tetradomain globins

PAML analyses clearly showed that the duplication favored or required some key changes, with positively selected amino acid residues in the branch between the single- and the tetra-domain lineages. These residues are located in crucial parts of the protein, with probable structural, and functional effects. Three of these residues face outward in a region of the molecule where the B helix is located. This will most likely influence possible interactions in that region, either between domains within a subunit, or between subunits. The two other residues are located deeper in the molecule. One of them is located in the heme-distal part of the heme pocket and may affect oxygen binding. The remaining one, a cysteine, is located in the very back of the heme pocket. A cysteine in that position is also under strong selective pressure in extracellular globins from annelid species that live in sulfidic environments (Bailly et al. 2003). This suggests that this is the result of convergent evolution and points to an important role in species that are exposed to sulfide, a chemical that can irreversibly inhibit oxygen binding in hemoglobins. A cystein in that position may have a protective role to keep the hemoglobin functional. This could be tested by the

production of recombinant proteins, comparing this wild type to a mutated molecule where the cysteine is replaced by another amino acid.

4.2.2 Single-domain globins

As far as the single-domain globins are concerned, we only found positive selection in the branch that leads to the two *Branchipolynoe* species. In this genus, there are three types of hemoglobins in the coelomic fluid, two multimeric tetradomain hemoglobins and a single-domain globin. This probably allowed a division of labor, and lifted the selective pressure for one of them. The amino acid sequences showed some variability among species but the positively selected amino acid search pointed to three amino acids in *Branchipolynoe*. This time, one residue is located on the F-helix, in a position that is bound to affect the formation of dimers (typically involving both E and F helices in annelids). Similarly, a residue located in the DE corner could affect the formation of tetramers. Finally, a residue located in the heme pocket is likely to stabilize the binding of oxygen (and thereby increase the affinity by decreasing the *off* rate). The amplification and sequencing of the missing parts of the incomplete sequences would be most valuable, especially to verify is other amino acid positions can be under positive selective pressure. Afterwards a model could be constructed, similarly to the approach followed for *Branchipolynoe symmytillida* and *Branchinotogluma segonzaci*.

The homology modelization helped to locate the amino acids under positive selective pressure and generate hypotheses on their potential structural and functional effects. Nevertheless this is only a rough model, and although it can be quite accurate, it will never substitute the true protein structure. To achieve this, it would be good to obtain the real structure of these unique globins, through crystallization, as was done for some other species, cold-seep annelid *Oligobrachia mashikoi* (Nakagawa et al. 2005), and the vestimentiferan hydrothermal vent *Riftia pachyptila* 400-kDa hemoglobin (Flores et al. 2005).

4.3 Future research directions

The next obvious step would be the over-expression of the Polynoidae extracellular globins. This would represent a great achievement as the folding and quaternary assemblage of proteins is not always easy to accomplish under experimental conditions, and we may need large amounts of pure protein. This would greatly facilitate the study of these proteins because the constraints of sampling deep-sea species would no longer be a problem. In addition to producing the proteins in large amounts, the over-expression approach would allow us to specifically change some amino acids by directed mutagenesis.

This would bring definite information on the role of the identified residues in the heme pocket. The replacement of the glutamine in position E7 and tyrosine in B10, by the most common residues in the same positions, a histidine and a leucine, respectively, could reveal the interactions and final effect on oxygen affinity that these residues might have. Similar studies have already used this approach with very interesting results in *Ascaris* hemoglobin (De Baere 1994, Gibson et al. 1993). Still in the heme pocket, one residue in under positive selective pressure in *Branchipolynoe* single-domain globin. It could be replaced by a valine (as observed in *Branchinotogluma segonzaci*) to see the effect on the functional properties. The same type of approach can be used to determine the effect of each amino acid under positive selective, both on the functional and structural characteristics of the globins.

To complete this study in a larger respiratory context, adaptations can also be searched in enzymes involved in the switch between aerobic and anaerobic respiration. These polynoid scale-worms likely rely on anaerobiosis under temporary conditions, either when oxygen is too low in the environment (despite the adaptations to extract it), or when sulfide concentration is high and the aerobic metabolism is inhibited. Several enzymes can be interesting to study because they are at key positions in the metabolism. In particular, the mitochondrial malate dehydrogenase (MDH) is a crucial protein in the citric acid cycle (part of both aerobic and anaerobic metabolisms) and there is evidence from other invertebrate groups (Dahlhoff and Somero 1991b, 1993, Kirby 2000), that some intraspecific allelic variation, in the gene encoding this protein, can be correlated with environmental factors. This makes this protein a good molecular candidate to measure different responses to environmental constraints.

As an extreme environment, hydrothermal vents have proven in the course of this work that they can be a promoter of innovation and variability. The harshness of this environment and the succession of extinctions and recolonizations (and associated bottle-neck effect) likely favored innovation at the morphological and molecular level. Applied to one of the most successful adaptations of the aerobic way of life, the respiratory pigments, this thesis was able to show another evolutionary history inside the globin family and how small changes can have important effects in crucial body functions.

5. References

A

Alt JC. 1995. Subseafloor processes in mid-ocean ridge hydrothermal systems. *Seafloor Hydrothermal Systems*. Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. and Thompson, R. E. **Geophysical Monograph 91:** 85-114. Washington, American Geophysical Society.

Andersen AC, Jolivet S, et al. 2002. Biometry of the branchial plume in the hydrothermal vent tubeworm *Riftia pachyptila* (Vestimentifera; Annelida). *Canadian Journal of Zoology* **80**: 320-332.

Anisimova M, Bielawski JP, et al. 1991. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution* **18**: 1585-1592.

Arndt MHL, Nascimento DG, et al. 1998. The myoglobin and the hemoglobin of Biomphalaria glabrata, an evidence of gene duplications. *Memórias do Instituto Osvaldo Cruz* **93**: 171-172.

Arp AJ and Childress JJ. 1983. Sulfide binding by the blood of the hydrothermal vent tube worm *Riftia pachyptila*. *Science* **219**: 295-297.

Arp AJ, Childress JJ, et al. 1987. The sulphide-binding protein in the blood of the vestimentiferan tube-worm, *Riftia pachyptila*, is the extracellular haemoglobin. *Journal of Experimental Biology* **128**: 139-158.

Arp AJ, Doyle ML, et al. 1990. Oxygenation properties of the two co-occurring hemoglobins of the tube worm *Riftia pachyptila*. *Respiration Physiology* **80**:323-334.

B

Bachraty C, Legendre P, et al. 2009. Biogeographic relationships among deep-sea hydrothermal vent faunas at global scale. *Deep-Sea Research I* **56**: 1371-1378.

Bailly X, Chabasse C, et al. 2007. Globin gene family evolution and functional diversification in annelids. *FEBS Journal* **274**: 2641-2652.

Bailly X, Jollivet D, et al. 2002. Evolution of the sulfide-binding function within the globin multigenic family of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*. *Molecular Biology and Evolution* **19**: 1421-1433.

Bailly X, Leroy R, et al. 2003. The loss of the hemoglobin H_2S-binding function in annelids from sulfide-free habitats reveals molecular adaptation driven by Darwinian positive selection. *Proceedings of the National Academy of Sciences of the United States of America* **100**: 5885-5890.

Bartolomaeus T, Purschke G, et al. 2005. Polychaete phylogeny based on morphological data - a comparison of current attempts. *Hydrobiologia* **535/536**: 341-356.

Blum J and Fridovich I. 1984. Enzymatic defenses against oxygen toxicity in the hydrothermal vent animals *Riftia pachyptila* and *Calyptogena magnifica*. *Archives of Biochemistry and Biophysics* **228**: 617-620.

Brittain T. 1990. Co-operative functioning of the dimeric haemoglobin obtained from the radular muscle of the amphineurian mollusc *Amaurochiton glaucus*. *Comparative Biochemistry and Physiology Part B* **96**: 291-295.

Britten RJ and Kohne DE. 1968. Repeated sequences in DNA. *Science* **161**: 529-540.

C

Carrico RJ, Blumberg WE, et al. 1978. The reversible binding of oxygen to sulfhemoglobin. *Journal of Biological Chemistry* **253**: 7212-7215.

Cavanaugh CM, Gardiner SL, et al. 1981. Prokariotic cells in the hydrothermal vent tube worm Riftia pachyptila Jones: possible chemoautotrophic symbionts. *Science* **213**: 340-342.

Charlou JL, Donval JP, et al. 2000. Compared geochemical signatures and the evolution of Menez Gwen (37º50'N) and Lucky Strike (37º17'N) hydrothermal fluids, south of the Azores Triple Juntion on the Mid-Atlantic Ridge. *Chemical Geology* **171**: 49-75.

Chevaldonné P, Desbruyères D, et al. 1992. ... and some even hotter. *Nature* **359**: 593-594.

Chevaldonné P, Fisher, CR, et al. 2000. Thermotolerance and the 'Pompeii worms'. *Marine Ecology Progress Series* **208**: 293-295.

Chevaldonné P, Jollivet D, et al. 1998. Commensal scale-worms of the genus *Branchiolynoe* (Polychaeta: Polynoidae) at deep-sea hydrothermal vents and cold seeps. *Cahiers de Biologie Marine* **39**: 347-350.

Childress JJ. 1995. Are there physiological and biochemical adaptations of the metabolism in deep-sea animals? *Trends in Ecology and Evolution* **10**: 30-36.

Childress JJ, Arp AJ, et al. 1984. Metabolic and blood characteristics of the hydrothermal vent tube-worm *Riftia pachyptila*. *Marine Biology* **83**: 109-124.

Childress JJ, Fisher CR, et al. 1992. Sulfide and carbon dioxide uptake by the hydrothermal vent clam, *Calyptogena magnifica*, and its chemoautotrophic symbionts. *Physiological Zoology* **64**: 1444-1470.

Childress JJ and Fisher CR. 1992. The biology of hydrothermal vent animals: physiology, biochemistry, and autotrophic symbioses. *Oceanography and Marine Biology: An Annual Review* **30**: 337-441.

Childress JJ and Mickel TJ. 1985. Metabolic rates of animals from the hydrothermal vents and other deep-sea habitats. *Biological Society of Washington Bulletin* **6**: 249-260.

Childress JJ and Seibel BA. 1998. Life at stable low oxygen levels: adaptations of animals to oceanic oxygen minimum layers. *Journal of Experimental Biology* **201**: 1223-1232.

Corliss JB, Dymond J, et al. 1979. Submarine thermal springs on the galápagos rift. *Science* **203**: 1073-1083.

D

Dahlhoff E, O'Brien J, et al. 1991a. Temperature effects on mitochondria from hydrothermal vent invertebrates: evidence for adaptation to elevated and variable habitat temperatures. *Physiological Zoology* **64**: 1490-1508.

Dahlhoff E and Somero GN. 1991b. Pressure and temperature adaptation of cytosolic malate dehydrogenases of shallow and deep-living marine invertebrates: evidence for high body temperatures in hydrothermal vent animals. *Journal of Experimental Biology* **159**: 473-487.

Dahlhoff E and Somero GN. 1993. Kinetic and structural adaptations of cytoplasmic malate dehydrogenases of eastern pacific abalone (genus *Haliotis*) from different thermal habitats: biochemical correlates of biogeographical patterning. *Journal of Experimental Biology* **185**: 137-150.

Darawshe S, Tsafadyah Y, et al. 1987. Quaternary structure of erythrocruorin from the nematode *Ascaris suum*. *Biochemical Journal* **242**: 689-694-

de Angelis MA, Reysenbach A-L, et al. 1991. Surfaces of hydrothermal vent invertebrates: sites of elevates microbial CH_4 oxidation activity. *Limnology and Oceanography* **36**: 570-577.

De Baere I, Liu L, et al. 1992. Polar zipper sequence in the high-affinity hemoglobin of *Ascaris suum*: Amino acid sequence and structural interpretation. *Proceedings of the National Academy of Science USA* **89**: 4638-4642.

De Baere I, Perutz MF, et al. 1994. Formation of two hydrogen bonds from the globin to the heme-linked oxygen molecule in *Ascaris* hemoglobin. *Proceedings of the National Academy of Sciences* **91**: 1594-1597.

de Burgh ME and Singla CL. 1984. Bacterial colonization an endocytosis on the gill of a new limpet species from a hydrothermal vent. *Marine Biology* **84**: 1-6.

Desbruyères D, Chevaldonné P, et al. 1998. Biology and ecology of the "Pompeii worm" (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller of an extreme deep-sea environment: A synthesis of current knowledge and recent developments. *Deep-Sea Research II* **45**: 383-422.

Desbruyères D, Gaill F, et al. 1985. Polychaetous annelids from hydrothermal ven

ecosystems: An ecological overview. *Bulletin of the Biological Society of Washington* **6**: 103-116.

Desbruyères DI, Segonzac M, et al. Editors, 2006. *Handbook of deep-sea hydrothermal vent fauna.* Second completely revised edition. Linz, **Denise 18**: 544 pp.

Dewilde S, Blaxter M, et al. 1996. Globin and Globin Structure of the Nerve Myoglobin of *Aphrodite aculeata*. *The Journal of Biological Chemistry* **271**: 19865-19870.

Dewilde S, Van Hauwaert M-L, et al. 1999. *Daphnia pulex* didomain hemoglobin: structure and evolution of polymeric hemoglobins and their coding genes. *Molecular Biology and Evolution* **16**: 1208-1218.

Dixon B, Walker B, et al. 1992. A Nematode Gene Contains an Intron Previously Thought to be Unique to Plants. *Journal of Molecular Evolution* **35**: 131-136.

Duboule D and Wilkins AS. 1998. The evolution of 'bricolage'. *TIG* **14**: 54-59.

E

Edmond JM, Measures C, et al. 1979. On the formation of metal-rich deposits at ridge crests. *Earth and Planetary Science Letters* **46**: 19-30.

F

Fisher CR. 1996. Ecophysiology of primary production at deep-sea vents and seeps. *Deep-sea and extreme shallow-water habitats: affinities and adaptations*. Uiblein, F., Ott, J. and Stachowtisch, M. **Biosystematics and Ecology Series 11**: 313-336.

Fisher CR, Childress JJ, et al. 1988. Microhabitat variation in the hydrothermal vent mussel, *Bathymodiolus thermophilus*, at the Rose Garden vent on the Galapagos Rift. *Deep-Sea Research* **35**:1769-1988.

Flores JF, Fisher CR, et al. 2005. Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 2713-2718.

Flores JF and Hourdez S. 2006. The zinc-mediated sulfide-binding mechanism of hydrothermal vent tubeworm 400-kDa hemoglobin. *Cahiers de Biologie Marine* **47**: 371-377.

Fornari D and Embley RW. 1995. Tectonic and volcanic controls on hydrothermal processes at the mid-ocean ridge: An overview based on near-bottom and submersible studies. *Seafloor Hydrothermal Systems*. Humphris, S. E., Zierenberg, R. A., et al. **Geophysical Monograph 91**: 1-46. Washington, American Geophysical Society.

Fowler CMR and Tunnicliffe V. 1997. Hydrothermal vent communities and plate tectonics. *Endeavour* **21**: 164-168.

Fox HM. 1957. Haemoglobin in the Crustacea. *Nature* **179**: 148-149.

Fridovich I. 1978. The biology of oxygen radicals. *Science* **201**: 875-880.

Früh-Green GL, Kelley DS, et al. 2003. 30,000 years of hydrothermal activity at the Lost City vent field. *Science* **301**: 495-498.

G

Gaill F, Desbruyères D, et al. 1987. Bacterial communities associated with "Pompei Worms" from the East Pacific Rise hydrothermal vents: SEM, TEM Observations. *Microbial Ecology* **13**: 129-139.

Gaill F, Desbruyères D, et al. 1988. Relationships between the "Pompei worms" and their epibiotic bacteria. *Oceanologica Acta* **Nº SP 8**: 147-154.

Gekko K, Hasegawa Y. 1989. Effect of temperature on the compressibility of native globular proteins. *Journal of Physical Chemistry* **93**:426-429.

Gibson QH, Regan R, et al. 1993. Kinetics of ligand to Pseudoterranova decipiens and Ascaris suum hemoglobins and to Leu-29 → Tyr sperm-whale myoglobin mutant. *Journal of Biological Chemistry* **268**: 16993-16998.

Gibson QH and smith MH. 1965. Rates of reaction of *Ascaris* haemoglobins with ligands. *Proceedings of the Royal society of London B* **163**: 206-214.

Goldberg DE. 1995. The enigmatic oxygen-avid hemoglobin of *Ascaris*. *BioEssays* **17**: 177-182.

Goldman N and Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology an Evolution* **11**: 725-736.

Goldstein SJ, Perfit MR, et al. 1994. Off-axis volcanism at the East Pacific Rise detected by uranium-series dating of basalts. *Nature* **367**: 157-159.

Goodman M, Moore GW, et al. 1975. Darwinian evolution in the genealogy of haemoglobin. *Nature* **253**: 603-608.

Goodman M, Pedwaydon J, et al. 1988. An evolutionary tree for invertebrate globin sequences. *Journal of Molecular Evolution* **27**: 236-249.

Govenar B, Bris NL, et al. 2005. Epifaunal community structure associated with *Riftia pachyptila* aggregations in chemically different hydrothermal vent habitats. Marine *Ecology Progress Series* **305**: 67-77.

Grassle JF. 1985. Hydrothermal vent animals: distribution and biology. *Science* **229**: 713-717.

Grinich NP and Terwilliger RC. 1980. The quaternary structure of an unusual high-molecular-weight intracellular haemoglobin from the bivalve mollusc *Barbatia*

reeveana. *Biochemical Journal* **189**: 1-8.

Grieshaber MK, Völkel S. 1998. Animal adaptations for tolerance and exploitation of poisonous sulfide. *Annual Review of Physiology* **60**: 33-53.

Gross M, Jaenicke R. 1994. Proteins under pressure – The influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes. *European Journal of Biochemistry* **221**:617-630.

H

Han J-H, Batey S, et al. 2007. The folding and evolution of multidomain proteins. *Nature Reviews Molecular Cell Biology* **8**(4): 319-330.

Hand SC and Somero GN. 1983. Energy metabolism pathways of hydrothermal vent animals: adaptations to a foo-rich and sulfide-rich deep-sea environment." *Biological Bulletin* **165**: 167-181.

Hardison R. 1996. A brief history of hemoglobins: Plant, animal, protist and bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **93**: 5675-5679.

Hessler RR and Jumars PA. 1974. Abyssal community analysis from replicate box cores in the central North Pacific. *Deep-Sea Research* **21**: 185-209.

Hessler RR and Kaharl V. 1995. The deep-sea hydrothermal vent community: an overview. *Seafloor Hydrothermal Systems*. Humphris, S. E., Zierenberg, R. A., Mullineaux, L. S. and Thompson, R. E. **Geophysical Monograph 91:** 72-84. Washington, American Geophysical Society.

Hoogewijs D, De Henau S, et al. 2008. The *Caenorhabditis* globin gene family reveals extensive nematode-specific radiation and diversification. *BMC Evolutionary Biology* **8**: 279 (doi:10.1186/1471-2148-8-279)

Hourdez S. (2000). Adaptations respiratoires des annélides polychètes de milieux abyssaux hypoxiques. *PhD Thesis*: 197 pp., Équipe Ecophysiologie, SBR. Roscoff, Université Pierre et Marie Curie, Paris VI.

Hourdez S and Desbruyères D. 2003. A new species of scale-worm (Polychaeta: Polynoidae), *Levensteiniella iris* sp. nov., from the Rainbow and Lucky Strike vent fields (Mid-Atlantic Ridge). *Cahiers de Biologie Marine* **44**: 13-21.

Hourdez S and Jouin-Toulmond C. 1998. Functional anatomy of the respiratory system of *Branchipolynoe* species (Polychaeta, Polynoidae), commensal with *Bathymodiolus* species (Bivalvia, Mytilidae) from deep-sea hydrothermal vents. *Zoomorphology* **118**: 225-233.

Hourdez S, Lallier FH, et al. 1999a. Hemoglobins from deep-sea hydrothermal vent scaleworms of the genus *Branchipolynoe*: A new type of quaternary structure. *Proteins: Structure, Function, and Genetics* **34**: 427-434.

Hourdez S, Lallier FH, et al. 1999b. Characterization and functional properties of the extracellular coelomic hemoglobins from the deep-sea, hydrothermal vent scaleworm *Branchipolynoe symmytilida*. *Proteins: Structure, Function, and Genetics* **34**: 435-442.

Hourdez S, Lallier FH, et al. 2000a. Gas transfer system in *Alvinella pompejana* (Annelida Polychaeta, Terebellida): functional properties of intracellular and extracellular hemoglobins. *Physiological and Biochemical Zoology* **73**: 365-373.

Hourdez S, Lamontagne J, et al. 2000b. Hemoglobin from a deep-sea hydrothermal-vent copepod. *Biological Bulletin* **199**: 95-99.

Hourdez S and Lallier F. 2007. Adaptations to hypoxia in hydrothermal-vent and cold-seep invertebrates. *Reviews in Environmental Science and Biotechnology* **6**: 143-159.

Hourdez S and Weber RE. 2005. Molecular and functional adaptations in deep-sea hemoglobins. *Journal of Inorganic Biochemistry* **99**: 130-141.

I

Imai K. 1999. The haemoglobin enzyme. *Nature* **401**: 437.

J

Jaenicke R. 1981. Enzymes under extremes of physical conditions. *Annual Review of Biophysics and Bioengineering* **10**: 1-67.

Jaenicke R. 1983. Biochemical processes under high hydrostatic pressure. *Naturwissenschaften* **70**: 332-341.

Jaenicke R and Koberstein R. 1971. High pressure dissociation of lactic dehydrogenase. *FEBS Letters* **17**: 351-354.

Jaquet M. 1886. Recherches sur le système vasculaire des Annélides. *Mitt Zool Sta Neapel* **6**: 297–398.

Jeanthon C, Prieur D. 1990. Resistance to heavy metals of heterotrophic bacteria isolated from the deep-sea hysdrothermal vent polychaete, *Alvinella pompejana*. *Progress in Oceanography* **24**: 81-88.

Jellie AM, Tate WP, et al. 1996. Evolutionary history of introns in a multidomain globin gene. *Journal of Molecular Evolution* **42**: 641-647.

Jones ML. 1988. The Vestimentifera, their biology, systematic and evolutionary patterns. *Oceanologica Acta* **Nº SP 8**: 69-82.

Jollivet D, Comtet T, et al. 1998. Unexpected relationship between dispersal strategies and speciation within the association Bathymodiolus (Bivalvia) - *Branchipolynoe* (Polychaeta) inferres from the rDNA neutral ITS2 marker. *Cahiers de Biologie*

Marine **39**: 359-362.

Jouan L, Marco S, et al. 2003. Revisiting the structure of Alvinella pompejana hemoglobin at 20A resolution by cryoelectron microscopy. *Journal of Structural Biology* **143**: 33-44.

Jouin-Toulmond C, Augustin D, et al. 1996. The gas transfer system in alvinellids (Annelida Polychaeta, Terebellida). Anatomy and ultrastructure of the anterior circulatory system and characterization of a coelomic, intracellular, haemoglobin *Cahiers de Biologie Marine* **37**: 135-151

Jouin C and Gaill F. 1990. Gills of hydrothermal vent annelids: structure, ultrastructure and functional implications in two alvinellid species. *Progress in Oceanography* **24**: 59-69.

K

Kauzmann W. 1959. Some factors in the interpretation of protein denaturation. *Advances in Protein and Chemistry* **14**: 1-63.

Keilin D. 1929. Cytochrome and respiratory enzymes. *Proceedings of the Royal Society of London Series B* **104**: 206-252.

Kirby RR. 2000. Cloning and primary structure of putative cytosolic and mitochondrial malate dehydrogenase from the mollusc *Nucella lapillus* (L.). *Gene* **245**: 81-88.

Kochevar RE, Childress JJ, et al. 1992. The methane mussel: roles of symbiont and host in the metabolic utilization of methane. *Marine Biology* **112**: 389-401.

Krebs A, Kuchumov AR, et al. 1996. Molecular shape, dissociation, and oxygen binding of the dodecamer subunit of *Lumbricus terrestris* hemoglobin. *Journal of Biological Chemistry* **271**: 18695-18704.

Kreutzer U and Jue T. 1997. Metabolic response in *Arenicola marina* to limiting oxygen as reflected in the H-NMR oxymyoglobin signal. *European Journal Biochemistry* **243**: 233-239.

Kriventseva E, Koch I, et al. 2003. Increase of functional diversity by alternative splicing. *Trends in Genetics* **19**: 124-128.

Kuenen G. 1999. Oxidation of inorganic compounds by chemolithotrophs. *Biology of the Prokariots*: 234-260 Lengeler, J. W., Drews, G. and Schlegel, H. G. (eds). Stuttgart, Germany, Blackwell Science.

Kundrot CE and Richards FM. 1987. Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres. *Journal of Molecular Biology* **193**: 157-170.

L

Lalou C, Reyss J-L, et al. 1993. New age data for Mid-Atlantic ridge hydrothermal sites:

TAG and Snakepitchronology revisited. *Journal of Geophysical Research* **98**: 9705-9713.

Le Bris N, Sarradin P-M, et al. 2003. Contrasted sulphide chemistries in the environment of 13ºN EPR vent fauna. *Deep-Sea Research I* **50**: 737-747.

Le Pennec M. 1988. Alimentation et reproduction d'un Mytilidae des sources hydrothermales profondes du Pacifique oriental. *Oceanologica Acta* **NºSP 8**: 181-190.

Le Pennec M, Donval A, et al. 1990. Nutritional strategies of the hydrothermal ecosystem bivalves. *Progress in Oceanography* **24**: 71-80.

Le Pennec M and Fiala-Medioni A. 1988. The role of the digestive tract of *Calyptogena lauberi* and *Calyptogena phaseoliformis*, vesicomyid bivalves of the subduction zones of Japan. *Oceanologica Acta* **NºSP 8**: 193-199.

Lieb B, Dimitrova K, et al. 2006. Red blood with blue-blood ancestry: Intriguing structure of a snail hemoglobin. *Proceedings of the Natural Academy of Sciences of the United States of America* **103**: 12011-12016.

Lister CRB. 1972. On the thermal balance of a mid-ocean ridge. *Geophysics Journal International* **26**: 515-535.

Lowell RP, Rona PA, et al. 1995. Seafloor hydrothermal systems. *Journal of Geophysical Research* **100**: 327-352.

Low PS, Somero GN. 1975. Pressure effects on enzyme structure and functions in vitro and under simulated in vivo conditions. *Comparative Biochemistry and Physiology* **52B**: 67-74.

Lupton JE, Delaney JR, et al. 1985. Entrainment and vertical trasnport of deep-ocean water by buoyant hydrothermal plumes. *Nature* **316**: 621-623.

Luther III GW, Rozan TF, et al. 2001. Chemical speciation drives hydrothermal vent ecology. *Nature* **410**: 813-816.

M

Mangum CP. 1992. Respiratory function of the red blood cell hemoglobins of six animal phyla. *Advances in Comparative & Environmental Physiology* **13:** 118-149. Mangum CP, Springer-Verlag.

Manning A, Trotman C, et al. 1990. Evolution of a polymeric globin in the brine shrimp *Artemia*. *Nature* **348**: 653-656.

Marie B, Genard B, Rees J-F, Zal F. 2006. Effect of ambient oxygen concentration on activities of enzymatic antioxidant defences and aerobic metabolism in the hydrothermal vent worm, *Paralvinella grasslei*. *Marine Biology*: **150**, 273-284.

Martin D and Britayev TA. 1998. Symbiotic polychaetes: review of known species. *Oceanography and Marine Biology: an Annual Review* **36**: 217-340.

Martineu P, Juniper SK, et al. 1997. Sulfide binding in the body fluids of hydrothermal vent alvinellid polychaetes. *Physiological Zoology* **70**: 578-588.

Massoth GJ, Milburn HB, et al. 1988. The geochemistry of submarine venting fluids at Axial Volcano, Juan de Fuca Ridge: new sampling methods and a VENTS program rationale. *Global Venting, Midwater and Benthic Ecological Processes*. De Luca MP and Babb I. **National Undersea Research Program Res. Rep. 88-4**: 29-59. U.S. Dept. Commerce, Rockville, MD.

Matthews CM, Vanderberg CJ, et al. 1998. Variable substitution rates of the 18 domain sequences in Artemia hemoglobin. *Journal of Molecular Evolution* **46**: 729-733.

Minning DM, Kloek AP, et al. 1995. Subunit interactions in *Ascaris* hemoglobin octamer formation. *Journal of Biological Chemistry* **270**: 22248-22253.

Mitchell DT, Kitto GB, et al. 1995. Atructural analysis of monomeric hemichrome and dimeric cyanomet hemoglobins from *Caudina arenicola*. *Journal of Molecular Biology* **251**: 421-431.

N

Naito Y, Riggs CK, et al. 1991. Origin of a "bridge" intron in the gene for a two-domain globin. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 6672-6676.

Nakagawa T, Onoda S, et al. 2005. Purification, characterization and sequence analyses of the extracellular giant hemoglobin from *Oligobrachia mashikoi*. *Zoological Science* **22**: 283-291.

Nakagawa S and Takai K. 2008. Deep-sea vent chemoautotrophs: diversity, biochemistry and ecological significance. *FEMS Microbiology Ecology* **65**: 1-14.

Nei M and Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**: 418-426.

Nielsen R and Yang Z. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**: 929-936.

O

Ohno S. 1970. Evolution by gene duplication. Springer-Verlag, Berlin. 160 pp.

P

Pasek S, Risler J-L, et al. 2006. Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* **22**: 1418-1423.

Pathy L. 1999. Genome evolution and the evolution of the exon-shuffling - a review. *Gene* **238**: 103-114.

Peterson ES, Huang S, et al. 1997. A comparison of functional and structural consequences of the tyrosyne B10 and glutamine E7 motifs in two invertebrate hemoglobins (*Ascaris suum* and *Lucina pectinata*). *Biochemistry* **36**: 13110-13121.

Pettibone MH. 1984. A new scale-worm commensal with deep-sea mussels on the Galapagos hydrothermal vent (Polychaeta: Polynoidae). Proceedings of the *Biological Society of Washington* **97**: 226-239.

Pettibone MH. 1986. A new scale-worm commensal with deep-sea mussels in the seep-sites at the florida escarpment in the eastern Gulf of Mexico (Polychaeta: Polynoidae: Branchypolynoinae). Proceedings of the *Biological Society of Washington* **99**: 444-451.

Plouviez S, Daguin-Thiébaut C, et al. 2008. Juvenile and adult scale worms Branchipolynoe seepensis in Lucky Strike hydrothermal vent mussels are genetically unrelated. *Aquatic Biology* **3**: 79-77.

Powell MA and Somero GN. 1986. Adaptations to sulfide by hydrothermal vent animals: sites and mechanisms of detoxification and metabolism. *Biological Bulletin* **171**: 274-290.

R

Ramírez-Llodra E, Tyler PA, et al. 2003. Biogeography of deep-water chemosynthetic ecosystems (CHESS): Exploring the southern oceans. *Gayana* **67**: 168-176.

Reysenbach A-L and Shock E. 2002. Merging genomes with geochemistry in hydrothermal ecosystems. *Science* **296**: 1077-1082.

Rinke C and Lee RW. 2009. Pathways, activities and thermal stability of anaerobic and aerobic enzymes in thermophilic vent paralvinellid worms. *Marine Ecology Progress Series* **382**: 99-112.

Rodionov MA and Johnson MS. 1994. Residue-residue contact substitution probabilites derived from aligned three-dimensional structures and the identification of common folds. *Protein Science* **3**: 2366-2377.

Rouse GW and Fauchald K. 1997. Cladistics and polychaetes. *Zoologica Scripta* **26**: 139-204.

Royer Jr W, Omartian M, et al. 2007. Low resolution crystal structure of *Arenicola* erythocrurin: influence of coiled coils on the architecture of a megadalton respiratory protein. *Journal of Molecular Biology* **365**: 226-236.

Royer Jr WE, Strand K, et al. 2000. Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 7107-7111.

Royer Jr WE, Zhu H, et al. 2005. Allosteric hemoglobin assembly: diversity and similarity. *Journal of Biological Chemistry* **280**: 27477-27480.

S

Sarrazin J, Juniper SK, et al. 1999. Physical and chemical factors influencing species distributions on hydrothermal sulfide edifices of the Juan de Fuca Ridge, northeast Pacific. *Marine Ecology Progress Series* **190**: 89-112.

Sarrazin J, Robigou V, et al. 1997. Biological and geological dynamics over fours years on a high-temperature sulfide structure at the Juan de Fuca Ridge hydrothermal observatory. *Marine Ecology Progress Series* **153**: 5-24.

Schindelmeiser I, Kuhlmann D, et al. 1979. Localization and characterization of hemoproteins in the central nervous tissue of some gastropods. *Comparative Biochemistry and Physiology B* **64**: 149-154.

Grantham R. 1974. Amino acid difference formula to help explain protein evolution. *Science* **185**: 862-864.

Sell AF. 2000. Life in the extreme environment at a hydrothermal vent: haemoglobin in a deep-sea copepod. *Proceedings of the Royal Society of London, Series B: Biological Sciences* **267**: 2323-2326.

Sherman DR, Kloek AP, et al. 1992. *Ascaris* hemoglobin gene: Plant-like structure reflects the ancestral globin gene. *Proceedings of the National Academy of Sciences of the United States of America* **89**: 11696-11700.

Shmelev IP, Kuznetsov AA, et al. 2009. Heavy metals in the benthic animals from hydrothermal vents: results of neutron activation analysis. *Oceanology* **49**: 429-431.

Shock EL, McCollom T, et al. 1998. The emergence of metabolism from within hydrothermal systems. *Thermophiles: the keys to molecular evolution and the origin of life*: 59-76. Wiegel J and Adams MWW (eds). London, UK, Taylor & Francis.

Sibuet M and Olu K. 1998. Biogeography, biodiversity and fluid dependence of deep-sea cold-seep comunities at active and passive margins. *Deep-Sea Research II* **45**: 517-567.

Smith CR, Kukert H, et al. 1989. Vent fauna on whale remains. *Nature* **341**: 27-28

Sokolova MN. 1972. Trophic structure of deep-sea macrobenthos. *Marine Biology* **16**: 1-12.

Somero GN. 1992. Adaptations to high hydrostatic pressure. *Annual Review of Physiology* **54**: 557-577.

Spiess FN, Macdonald KC, et al. 1980. East Pacific Rise: Hot Springs and Geophysical Experiments. *Science* **207**(4438): 1421-1433.

Stein JL, Cary SC, et al. 1988. Chemoautothrophic symbiosis in a hydrothermal vent gastropod. *Biological Bulletin* **174**: 373-378.

Stommel H. 1982. Is the South Pacific helium-3 plume dynamically active? *Earth and Planetary Science Letters* **61**: 63-67.

Suzuki T and Arita T. 1995. Two-Domain Hemoglobin from the Blood Clam, *Barbatia lima*. The cDNA-Derived Amino Acid Sequence. *Journal of Protein Chemistry* **14**: 499-502.

Suzuki T and Imai K. 1998. Evolution of myoglobin. *Cell and Molecular Life Sciences* **54**: 979-1004.

Suzuki T, Kawasaki Y, et al. 1996. Two-domain haemoglobin of the blood clam *Barbatia lima* resulted from the recent gene duplication of the single-domain δ chain. *Biochemical Journal* **313**: 561-566.

Suzuki T, Takagi T, et al. 1990. Primary structure of a linker subunit of the tube worm 3000-kDa hemoglobin. *Journal of Biological Chemistry* **265**: 1551-1555.

T

Tapley DW, Buettner GR, et al. 1999. Free radicals and chemilumunescence as products of spontaneous oxidation of sulfide in seawater, and their biological implications. *Biological Bulletin* **196**: 52-56.

Terwilliger NB. 1992. Molecular structure of the extracellular heme proteins. *Advances in Comparative & Environmental Physiology* **13:** 193-229. Mangum CP, Springer-Verlag.

Terwilliger RC, Garlick RL, et al. 1980. Characterization of the hemoglobins and myoglobin of *Travisia foetida*. *Comparative Biochemistry and Physiology Part B* **66**: 261-266.

Terwilliger RC. and Terwilliger NB. 1983. Thermal vent clam (*Calyptogena magnifica*) hemoglobin. *Science* **219**: 981-983.

Terwilliger NB and Terwilliger RC. 1984. Hemoglobin from the "Pompei worm", *Alvinella pompejana*, ana annelid from a deep sea hot hydrothermal vent

environment. *Marine Biology Letters* **5**: 191-201.

Terwilliger RC, Terwilliger NB, et al. 1980. The structure of hemoglobin from an unusual deep-sea worm (Vestimentifera). *Comparative Biochemistry and Physiology* **65B**: 531-535.

Theissen U and Martin W. 2008. Sulfide:quinone oxidoreductase (SQR) from the lugworm *Arenicola marina* shows cyanide- and thioredoxin-dependent activity. *The FEBS Journal* **275**: 1131-1139.

Tokishita S, Shiga Y, et al. 1997. Cloning and analysis of a cDNA enconding a two-domain hemoglobin chain from the water flea *Daphnia magna*. *Gene* **189**: 73-78.

Toulmond A. 1970. La fixation du dioxide de carbone par le sang chez l'Arénicole [*Arenicola marina* (L.), Annélide, Polychète]. *Comptes Rendus de l'Académie des Sciences de Paris* **270**: 1487-1490.

Toulmond A. 1992. Properties and functions of extracellular heme pigments. *Advances in Comparative & Environmental Physiology* **13:** 231-256. Mangum, C. P., Springer-Verlag.

Toulmond A, Slitine FEI., et al. 1990. Extracellular homoglobins of hydrothermal vent annelids: structural and functional characteristics in three alvinellid species. *Biological Bulletin* **179**: 366-373.

Trotman C, Manning A, et al. 1994. Interdomain Linkage in the Polymeric Hemoglobin Molecule of *Artemia*. *Journal of Molecular Evolution* **38**: 628-636.

Tunnicliffe V. 1991. The Biology of Hydrothermal Vents: Ecology and Evolution. *Oceanography and Marine Biology: An Annual Review* **29**: 319-407.

Tunnicliffe V. 1988. Biogeography and evolution of hydrothermal-vent fauna in the eastern Pacific Ocean. *Proceedings of the Royal Society of London B* **233**: 347-366.

Tyler PA and Young CM. 2003. Dispersal at hydrothermal vents: a summary of recent progress. *Hydrobiologia* **503**: 9-19.

V

Van Dover CL, German CR, et al. 2002. Evolution and biogeography of deep-sea vent and seep invertebrates. *Science* **295**: 1253-1257.

Vetter RD, Wells ME, et al. 1987. Sulfide detoxification by the hydrothermal vent crab *Bythograea thermydron* and other decapod crustaceans. *Physiological Zoology* **60**: 121-137.

Vine FJ and Matthews DH. 1963. Magnetic anomalies over oceanic ridges. *Nature* **199**: 947-949.

Vinogradov SN, Walz DA, et al. 1993. Adventitious variability? The amino acid sequences

of nonvertebrate globins. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **106**: 1-26.

Von Damm KL. 1990. Seafloor hydrothermal activity: black smoker chemistry and chimneys. *Annual Review of Earth Planetary Sciences* **18**: 173-204.

Von Damm KL. 1995. Controls on the chemistry and temporal variability of seafloor hydrothermal fluids. *Seafloor Hydrothermal Systems*. Humphris SE, Zierenberg RA, Mullineaux LS and Thompson RE. Washington, American Geophysical Society. **Geophysical Monograph 91:** 222-247.

Von Damm KL. 2001. Lost City found. *Nature* **412**: 127-128.

W

Weber RE. 1980. Functions of invertebrate hemoglobins with special reference to adaptations to environmental hypoxia. *American Zoology* **20**: 75-101.

Weber RE. 1978. Respiratory Pigments. *Physiology of Annelids*. Mill, P. J. London, Academic Press Inc.

Weber RE and Vinogradov SN. 2001. Nonvertebrate hemoglobins: functions and molecular adaptations. *Physiological Reviews* **81**: 569-628.

Wittenberg BA, Briehl RW, et al. 1965. Haemoglobins of invertebrate tissues - nerve haemoglobins of *Aphrodite*, *Aplysia* and *Halosydna*. *Biochemical Journal* **96**: 363-371.

Woodruff LG and Shanks III WC. 1988. Sulfur isotope study of chimney minerals and vent fluids from 21ºN, East Pacific Rise: hydrothermal sulfur sources and disequilibrium sulfate reduction. *Journal of Geophysical Research* **93**: 4562-4572.

Y

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* **15**: 568-573.

Yang Z. 2004. PAML: Phylogenetic analysis by Maximum Likelihood, version 3.14. User's Guide. (http://abacus.gene.ucl.ac.uk/software/paml.html).

Yang Z and Bielawski JP. 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution* **15**: 496-503.

Yang Z and Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution* **17**: 32-43.

Yang Z, Nielsen R, et al. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431-449.

Yang Z and Nielsen R. 2002. Codon-substitution models for detecting molecular

adaptations at individual sites along specific lineages. *Molecular Biology and Evolution* **19**: 908-917.

Yang Z, Nielsen R, et al. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Molecular Biology and Evolution* **15**: 1600-1611.

Z

Zal F, Green BN, et al. 1997a. Quaternary structure of the extracellular haemoglobin of the lugworm *Arenicola marina* - a multi-angle-laser-light-scattering and electrospray-ionisation-mass-spectrometry analysis. *European Journal Biochemistry* **243**: 85-92.

Zal F, Green B, et al. 1997b. Investigation by electrospray ionization mass spectrometry of the extracellular hemoglobin from the polychaete annelid *Alvinella pompejana*: an unusual hexagonal bilayer hemoglobin. *Biochemistry* **36**: 11777-11786.

Zal F, Jollivet D, et al. 1995. The reproductive biology and population structure of the deep-sea hydrothermal vent worm *Paralvinella grasslei* (Polychaeta: Alvinellidae) at 13ºN on the East Pacific Rise. *Marine Biology* **122**: 637-648.

Zal F, Lallier F, et al. 1996a. The Muti-hemoglobin System of the Hydrothermal Vent Tube Worm *Riftia pachyptila*. I. Reexamination of the number and masses of its consituents. *Journal of Biological Chemistry* **271**: 8869-8874.

Zal F, Leize E, et al. 1998. S-Sulfohemoglobin and disulfide exchange: The mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins. *Proceedings of the National Academy of Sciences of the United States of America* **95**: 8997-9002.

Zal F, Leize E, et al. 2000. Haemoglobin structure and biochemical charateristics of the sulphide-binding component from the deep-sea clam *Calyptogena magnifica*. *Cahiers de Biologie Marine* **41**: 413-423.

Zal F, Suzuki T, et al. 1997c. Primary structure of the common polypeptide chain *b* from the multi-hemoglobin system of the hydrothermal vent tube worm *Riftia pachyptila*: An insight on the sulfide binding-site. *Proteins: Structure, Function, and Genetics* **29**: 562-574.

Zierenberg RA, Adams MWW, et al. 2000. Life in extreme environments: hydrothermal vents." *Proceedings of the National Academy of Sciences of the United States of America* **97**: 12961_12962.

Figures index

Tables index