

Evolution moléculaire de la fonction de liaison du sulfure d'hydrogène par les hémoglobines de Riftia pachyptila, annélide tubicole des sources hydrothermales profondes

Xavier Bailly

▶ **To cite this version:**

Xavier Bailly. Evolution moléculaire de la fonction de liaison du sulfure d'hydrogène par les hémoglobines de Riftia pachyptila, annélide tubicole des sources hydrothermales profondes. Biologie moléculaire. Paris 6, 2003. Français. NNT : . tel-01112979

HAL Id: tel-01112979

https://hal.sorbonne-universite.fr/tel-01112979

Submitted on 11 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Avertissement

Au vu de la législation sur les droits d'auteur, ce travail de thèse demeure la propriété de son auteur, et toute reproduction de cette oeuvre doit faire l'objet d'une autorisation de l'auteur. (cf Loi n°92-597; 1/07/1992. Journal Officiel, 2/07/1992)

THESE DE DOCTORAT DE L'ECOLE PRATIQUE DES HAUTES ETUDES
ET DE L'UNIVERSITE PIERRE ET MARIE CURIE, PARIS VI

Spécialité
Génétique Moléculaire et Evolution

Présentée par Xavier BAILLY

Evolution moléculaire de la fonction de liaison du sulfure d'hydrogène par les hémoglobines de *Riftia pachyptila*, annélide tubicole des sources hydrothermales profondes

Soutenue le 06 janvier 2003 devant le jury composé de:

Pr Emile Zuckerkandl	Institute of Molecular Medical Sciences, Palo Alto USA	Rapporteur
Dr Guy Hervé	CNRS, Université Pierre et Marie Curie, Paris VI	Rapporteur
Pr Pierre-Henri Gouyon	Universite d'Orsay, Paris XI	Examinateur
Pr Jean Deutsch	Université Pierre et Marie Curie, Paris VI	Examinateur
Pr Michel Veuille	Ecole Pratique des Hautes Etudes	Directeur de thèse
Pr André Toulmond	Université Pierre et Marie Curie, Paris VI	Directeur de thèse
Dr Didier Jollivet	CNRS, Station Biologique de Roscoff	Examinateur

A Vera, petit ange parmi les anges …

J'ai toujours formulé d'une façon ou d'une autre (demandez-le donc à mes parents), le souhait d'entrer en science …
Je sais aujourd'hui, en partie pourquoi.
Je souffre depuis longtemps d'un syndrome tenace et finalement sympathique que j'ai appelé très sérieusement (faute de ne pas l'avoir trouvé dans les manuels de psychologie) "le syndrome du précepteur". J'ai toujours été envieux des hommes et des femmes d'un autre temps qui savaient et enseignaient à la fois l'Astronomie, les Mathématiques, la Philosophie, les Sciences, l'Histoire, l'Art ou encore la Musique et la Poésie. Bien sûr je n'adresse pas de critiques à mes bons maîtres et maîtresses qui m'ont tout donné à l'école, mais j'y ai vu comme une injustice, et la part de l'enfant curieux qui persiste encore en moi jalouse ces élèves "préceptés", eux aussi d'un autre temps … tant de belles choses à connaître et tant de chose à découvrir, tant de questions et de *terra incognita* et toujours moins de réponses précises.
Je définirai donc le syndrome du précepteur comme un besoin effréné d'expliquer plus, de comprendre plus pour contempler le monde avec une pointe d'émerveillement naïf de l'enfance … des pouponnières d'étoiles aux bons poly-phénols des vins de Bordeaux en passant par la voix de Jessy Norman sur les 4 derniers leader de Strauss ou les droseras carnivores des monts d'Arée et les analyses du commissaire Laviolette … la liste est en fait bien plus longue …
Devenir et être un apprenti chercheur en biologie me permet d'embrasser tout ça et de jouir encore librement du pouvoir de l'imagination, d'explorer et de poser des hypothèses, de postuler et de rechercher … de trouver des réponses … d'essayer de bien le faire.
J'aime vivre et travailler de cette façon.
Je dois cela à de nombreuses personnes. Force m'est d'avouer que plus j'y pense, plus la liste est là aussi bien longue.
D'abord, solennel merci à Fédérico Gomez-Para, qui m'a permis il y a quelques années au cours d'un rendez-vous formel et professionnel de prendre mon envol vers ce jour où j'écris ces quelques mots … A toi, ma sempiternelle amitié.
Merci à Hervé Le Guyader qui m'a accueilli au sein du DEA Biodiversité au retour d'une année passée en Italie, et à qui je dois cette fabuleuse aventure Roscovite.
Merci à André Toulmond et à Michel Veuille d'avoir accepté d'être respectivement mes Directeurs de thèse de l'Université Pierre et Marie Curie et de l'Ecole Pratique des hautes études et pour leurs conseils.
A Didier Jollivet, d'interminables remerciements pour l'encadrement de ce travail, pour m'avoir supporté et porté jusqu'ici. Je devrais dire pour m'avoir guidé, pour m'avoir aussi permis de voir ces fameuses sources hydrothermales derrière les hublots d'un sous-marin par 2500 m de fond au large du Mexique. Merci de tout cœur.
A Jean Deutsch, à l'attention, à la confiance et à son amitié qui me sont si chères et dont j'ai pu bénéficier pendant ces années d'apprentissage, j'adresse ma profonde reconnaissance, et mon amitié fidèle. Je suis heureux qu'il m'ait fait l'honneur d'être le président du jury de ma thèse.
A Guy Hervé je veux dire combien son amitié, ses mots et ses encouragements ont été décisifs, combien son regard sur mon travail et son écoute furent prépondérants. Merci d'avoir été l'un des deux rapporteurs.
A Emile Zuckerkandl je réitère un sentiment ineffable, mélange d'une immense admiration et d'une affection spirituelle naturelle: pour la chaleur de ses missives, pour l'écoute attentive, pour la simplicité du grand homme auquel j'ai pu exposer un jour mes travaux, pour ses

encouragements et cette rencontre inoubliable. Pour ne jamais m'avoir laissé sans réponse et pour m'avoir offert son amitié. Comme dans un songe heureux …
A Pierre-Henri Gouyon, pour ses (dis) cours de génétique évolutive, la force, l'originalité et la qualité de son enseignement, pour avoir accepté d'expertiser ce travail de thèse, j'exprime ma respectueuse reconnaissance, le grand plaisir et l'honneur d'avoir pu être l'un de ses élèves.
Grands mercis à toutes les personnes des Equipes Ecophysiologie et EGPM et à la grande famille de la Station Biologique de Roscoff.
Aux amis et rencontres roscovites je suis redevable d'un accueil simple et chaleureux. Je pense particulièrement à mon ami Guillaume de Lannurien dit "Barazer", le peintre, avec qui et à force de naviguer au large de Roscoff sur son cotre en bois, j'ai pu apprendre plus et voir différemment. Merci à toi, Marco, qui tient la barre au mythique Winch Pub à Roscoff, pour ton amitié et tes cafés du matin. A toi Stefano Vanin et à ta famille pour toutes ces années d'amitié, de fraternité et de collaborations scientifiques autours des hémérythrines. A vous et vos familles respectives, Remy Bertre, Bertrand Compain et Olivier Meyer : "mes fondamentaux". Merci à mes chers amis coincheurs et surfeurs … Merci aux vagues du Dossen, de Moguériec et de la Palue.
Des indicibles mercis à Gaëlle Correc qui a su "prendre en marche" cette vie et cette thèse, et qui a si merveilleusement bien négocié tous ces courants et vents parfois contraires, pour tout ce bonheur que nous te devons.
Enfin tous ces moments ne seraient pas là sans le soutien de ma famille: mes parents et mon frère.
A toi, petit frère, pour tes encouragements passionnés. Pour ton appétit de savoir et ton opiniâtreté.
A vous mes parents, à l'amour que vous nous donnez, à votre précieuse et irremplaçable présence, à vos conseils toujours bons, à cette éducation que pour rien au monde je n'aurais voulue différente, au bonheur que vous nous permettez de vivre, à cette vie que je vous dois. Je vous dédie cette thèse et cette aventure qui me plaît tant …

Préambule

PREAMBULE

Les annélides occupent une aire géographique considérable à l'échelle de notre planète. Ces organismes retrouvés aussi bien dans les écosystèmes terrestres que marins possèdent souvent du sang. Nous nous sommes principalement intéressés pendant cette thèse à l'étude de l'évolution moléculaire de la fonction de fixation du sulfure d'hydrogène (H_2S) par les hémoglobines extracellulaires des annélides. Ce travail repose sur la description de la famille multigénique codant les globines de *Riftia pachyptila* qui vit à proximité de sources hydrothermales profondes riches en sulfure d'hydrogène. Les annélides exposées à de fortes concentrations en H_2S vivent dans des milieux que nous appellerons milieux réduits par opposition aux milieux sans sulfure d'hydrogène, dits oxydés. Nous retiendrons que la globine est une protéine et que son association avec un groupement prosthétique forme l'hémoglobine. Enfin, nous avons utilisé la notion d'adaptation moléculaire dans le sens d'une modification d'un ou de plusieurs acides aminés au sein d'une séquence codante, en réponse à une ou des nouvelles contraintes évolutives.

Le premier chapitre montre que l'hémoglobine est le principal support "historique" qui a servi à la définition des principaux concepts en évolution moléculaire. Il présente également les différentes fonctions de l'hémoglobine au sein des organismes vivants.

Le second chapitre présente *Riftia pachyptila* et son environnement, ainsi que les travaux déjà réalisés sur ses hémoglobines et leur capacité de liaison du sulfure d'hydrogène.

Le troisième chapitre décrit la famille multigénique de *Riftia pachyptila* en faisant le parallèle des résultats obtenus par la biologie moléculaire au cours de cette thèse et les précédents résultats obtenus par spectrométrie de masse. Il replace également la famille multigénique des globines de *Riftia pachyptila* au sein des autres familles multigéniques de globines d'annélides.

L'article présenté dans le quatrième chapitre concerne l'évolution moléculaire de la fonction de liaison du sulfure d'hydrogène par la famille multigénique de *Riftia pachyptila*. Le cinquième chapitre présente un deuxième article qui suggère que la fonction de fixation de H_2S a été perdue par les annélides des milieux sans sulfures.

Enfin, le sixième chapitre est composé d'une discussion et d'une conclusion générale suivies de perspectives.

Sommaire

SOMMAIRE

Chapitre 1. Généralités sur l'hémoglobine 1

I. Définition, structure et fonction de l'hémoglobine 1

II. L'hémoglobine au cœur de l'émergence des concepts en évolution moléculaire 6

1. L'hémoglobine: marqueur évolutif et homologie moléculaire 6

2. L'hémoglobine: évolution par duplication des gènes et diversification 8

3. L'hémoglobine: horloge moléculaire, évolution moléculaire neutre et sélection darwinienne positive 9

III. Evolution des hémoglobines: structure des gènes et l'histoire des hémoglobines 12

1. Relation entre la structure des gènes et les domaines protéiques 12

2. Position des introns: marqueur de l'évolution des gènes codant l'hémoglobine 13

3. Les hémoglobines sont membres d'une superfamille 14

4. Hémoglobines tronquées 16

5. Héxacoordination et pentacoordination des hémoglobines 17

IV. L'hémoglobine dans les règnes du vivant 17

1. Hémoglobines des levures et bactéries 18

2. Hémoglobines des plantes supérieures 20

3. Hémoglobines des algues 21

4. Hémoglobines de protozoaires ciliés 21

5. Hémoglobines de vertébrés, de mammifères et de l'homme 21

a. Fixation du monoxyde d'azote (NO) 22

b. Fonctions marginales 23

c. Les neuroglobines 24

d. Les cytoglobines 25

6. Les hémoglobines d'invertébrés (eucaryotes) 27

V. La fonction de fixation du sulfure d'hydrogène par les hémoglobines de mollusques et d'annélides 29

1. Fixation et transport de H_2S par les hémoglobines cytoplasmiques (intracellulaire) de mollusques 29

2. Fixation et transport du sulfure d'hydrogène par les hémoglobines extracellulaire d'annélides des milieux riches en H_2S 30

VI. Conclusion sur la diversité des hémoglobines et leurs fonctions multiples : scénario évolutif 32

Chapitre 2. Les adaptations aux milieux riches en sulfure d'hydrogène. Une illustration : le cas des hémoglobines de *Riftia pachyptila*. 33

I. la vie en présence de sulfure d'hydrogène : généralités 33

II. Stratégies de détoxication de H_2S 33

1. Les corps d'inclusion de sulfures 34

2. Oxydation biologique de H_2S 35

III. L'environnement hydrothermal 36

IV. Sources hydrothermales et suintements froids : réseaux trophiques 39

V. *Riftia pachyptila* : Vestimentifère, phylum des annélides 40
VI. Les hémoglobines de *Riftia pachyptila* : analyse comparée avec les hémoglobines d'annélides 43
1. Les hémoglobines extracellulaires de Riftia pachyptila 43
2. Cystéines : pont dissulfures intra et inter-chaînes et fixation du sulfure d'hydrogène 45
VI. Conclusion 48

Chapitre 3. Analyse de la famille multigénique codant les hémoglobines de *Riftia pachyptila*. 50

I. Introduction 50
II. Matériel et méthodes 51
III. Résultats et Discussion 51
1. Séquençage des ADNc de globines et organisation de la famille multigénique 51
2. Comparaison des séquences obtenues et des données de spectrométrie 53
3. Evolution de la famille multigénique des globines de *Riftia pachyptila* 55
4. Evolution de la famille multigénique des globines au sein des annélides 56
5. Chaîne de structure des hémoglobines vasculaires V1 (HBL-Hb) de *Riftia pachyptila* 60
6. Structure du gène A2 et B2 de globine de *Riftia pachyptila* : séquence d'adressage, position des introns et séquence 3'UTR 60
a. Séquence d'adressage 60
b. Introns et séquences 3' UTR 61
7. Conclusion 63

Chapitre 4. Premier Article : Evolution de la fonction de fixation des sulfures dans la famille multigénique de globines chez le ver tubicole *Riftia pachyptila* des sources hydrothermales profondes. 65

Chapitre 5. Deuxième Article : La perte de la fonction de fixation du sulfure d'hydrogène chez les annélides des milieux oxydés s'explique par une adaptation moléculaire, résultat de la sélection darwinnienne positive. 97

Chapitre 6. Discussion, conclusion générale et perspectives 126

I. Emergence des hémoglobines extracellulaires chez les annélides 126
II. La famille multigénique des hémoglobines de *Riftia pachyptila* 128
1. Rejet de l'hypothèse alternative : colonisation des milieux riches en H_2S par des annélides venant des milieux oxydés 129
2. Les sources hydrothermales et les suintements froids ont-ils été colonisés par des annélides littoraux ou sub-littoraux équipées de telles adaptations ? 129
3. Fonction de densité des hémoglobines 131
III. Détecter de l'adaptation moléculaire sans estimer les substitutions nucléotidiques 132

IV. Détecter de l'adaptation moléculaire en estimant les substitutions ncléotidiques 133
V. Un concept marginal : la perte de fonction peut refléter de l'adaptation moléculaire et non un simple effet de la dérive génétique 134
VI. Limites expérimentales 136
1. Pseudogènes et famille multigénique 136
2. Séquences nucléotidiques des globines disponibles: sous-échantillonnage ? 136
VII. Conclusion 138
VIII. Perspectives 141

Bibliographie 145

Annexes 160

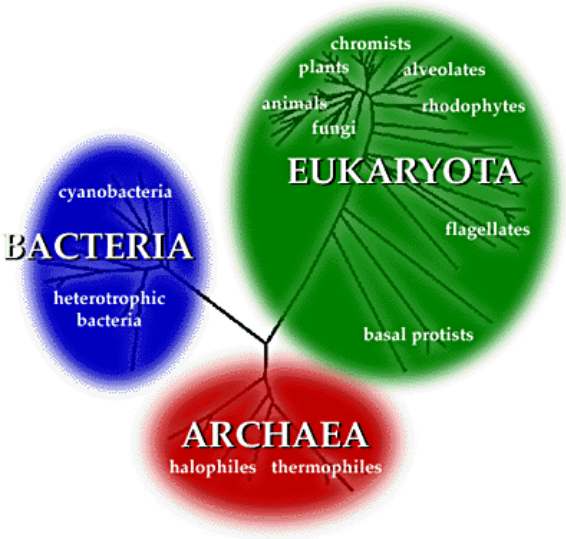
Annexe 1: 160
Légendes complètes: noms des espèces correspondant aux codes utilisés dans les phylogénies des figures 17, 18 et 29.
Annexe 2: 161
Matériel et Méthodes : A. Qu'est-ce qu'une amorce dégénérée ?
B. Extraction ADN génomique
C. Extraction des ARN Totaux
D. Purification des ARN messagers
E. Reverse transcription
F. Banque d' ADNc de *Riftia pachyptila*
Annexe 3: 164
Séquences nucléotidiques complètes ou partielles des clones contenant les globines
Annexe 4: 167
Alignements multiples
1. Alignement global des globines extracellulaires d'Annélides
2. Alignement de globines partielles d'Achètes
Annexe 5: 172
1. Alignement protéique des linkers de HBL-Hb d'annélides
2. Séquence nucléotidique et protéique partielles d'un linker de *Riftia pachyptila*
Annexe 6: 174
Séquences et positions relatives des deux introns au sein de la séquence protéique A2 et B2
Annexe 7: 175
1. Les 3 modèles d'estimations des substitutions synonymes et non-synonymes
2. Tableaux 2 et 3 du chapitre 4 (article 1)

Chapitre 1 :

Généralités sur l'hémoglobine

Chapitre 1. Généralités sur l'hémoglobine

Cette introduction générale met en évidence des caractères ancestraux et des innovations évolutives propres aux hémoglobines connues et analysées. Elle présente leur histoire et décrit leur diversité au sein des trois grands domaines du vivant Archea, Bacteria et Eucaryota (Woese et al. 1990 ; figure ci-contre). Nous verrons que certaines hémoglobines présentent des fonctions plutôt marginales eues égard à la fonction universelle de transporteur d'oxygène (O_2) qui leur a été d'abord attribuée.



Les 3 domaines du vivant d'après Woese (1990)

La finalité de ce travail est de pouvoir replacer à l'aide de concepts en évolution moléculaire, dans un contexte évolutif, adaptatif et fonctionnel, une fonction originale des hémoglobines extracellulaires d'annélides : la capacité de fixer et de transporter un composé toxique, le sulfure d'hydrogène (H_2S).

I. Définition, structure et fonction de l'hémoglobine

Le sang exerce une fonction respiratoire chez la plupart des animaux qui consiste principalement à transporter l'oxygène jusqu'aux cellules. L'hémoglobine, la protéine majoritaire du sang, est l'un des quatre transporteurs d'oxygène sélectionnés dans certaines lignées évolutives. Ces quatre structures protéiques sont des métalloprotéines qui se caractérisent par la présence d'un atome métallique directement impliqué dans la fixation de l'O_2 et qui confère la couleur spécifique à leur sang. C'est pour cette raison que ces structures ont été historiquement appelées pigments respiratoires. Trois classes de pigments respiratoires ont été caractérisés : les pigments héminiques, les hémérythrines et les hémocyanines. Parmi les pigments héminiques on trouve l'hémoglobine et la chlorocruorine. Ces pigments respiratoires peuvent être **intracellulaires circulants** (dans des cellules sanguines) ou **extracellulaires circulant**s (en solution dans le compartiment sanguin ou coelomique) ou encore **cytoplasmiques** (dans les cellules de certains tissus) (Tableau 1).

<table>
<tr><th colspan="2">Pigments héminiques</th></tr>
<tr><th>Hémoglobine</th><th>Chlorocruorine</th></tr>
<tr><td>Couleur : rouge ; Structure : protéine - porphyrine – fer
Intracellulaire :
Vertébrés, PM 68.000 Da
Cyclostomes, PM 19-23.000 Da,
Certains invertébrés. PM 23-36.000 Da
Extracellulaire :
Certains invertébrés. PM 1,5 à 3.000.000 Da
Cytoplasmique : Myoglobine. PM 17.000 Da</td><td>Couleur : verte. Structure : protéine - porphyrine - fer
Extracellulaire :
Sabellidés, Serpulidés, Chlorhæmidés, Ampharetidés
PM 2.750.000 Da</td></tr>
<tr><th colspan="2">Hémérythrine</th></tr>
<tr><td colspan="2">Couleur : rouge-violacé. Structure : protéine contenant du fer
Intracellulaire :
Sipunculiens, Priapulides, Brachiopodes, certains polychètes ; PM 108.000 Da</td></tr>
<tr><th colspan="2">Hémocyanine</th></tr>
<tr><td colspan="2">Couleur : bleue ; Structure : protéine contenant du cuivre
Extracellulaire :
Nombreux mollusques, crustacés et arthropodes ; PM 300.000 à 9.000.000 Da</td></tr>
</table>

Tableau 1. Noms et caractéristiques des quatre pigments respiratoires transporteurs d'oxygène.

Nous considérons les hémoglobines et les chlorocruorines comme des protéines homologues tel que Suzuki (1995) l'a précédemment suggéré. En effet, elles sont toutes les deux composées par l'association d'une partie protéique, la globine, et d'une partie prosthétique, l'hème. La globine est un polypeptide d'environ 150 acides aminés et possède une structure tridimensionnelle typique dite "globin-fold" (une succession de 7 à 8 hélicesα) (fig. 1). Chez ces deux pigments respiratoires l'origine évolutive de la partie protéique est manifestement la même (même structure du gène (Dewilde et al. 2001), même structure protéique (Suzuki et al. 1995), même structure quaternaire (Weber et Vinogradov 2001), même fonction). La seule différence est la substitution d'un groupement vynil par un groupement aldéhyde de l'hème qui confère la couleur verte pour les chlorocruorines (Weber et Vinogradov 2001). Conventionnellement, on attribue aux hélices α des globines une lettre, de A à H, de l'extrémité NH2-terminal vers l'extrémité COOH -terminal. Les segments inter-hélices portent les lettres des deux hélices qu'ils séparent (AB pour le segment entre l'hélice A et l'hélice B). NA et HC sont respectivement le segment qui précède la première et celui qui suit et la dernière hélice. L'hème est une protoporphyrine à laquelle est liée par coordinence un atome de fer ferreux (Fe^{2+}) à ses 4 azotes pyrroliques (fig.1). Le fer est aussi lié par une cinquième liaison de coordinence à une histidine de l'hélice F (histidine proximale).

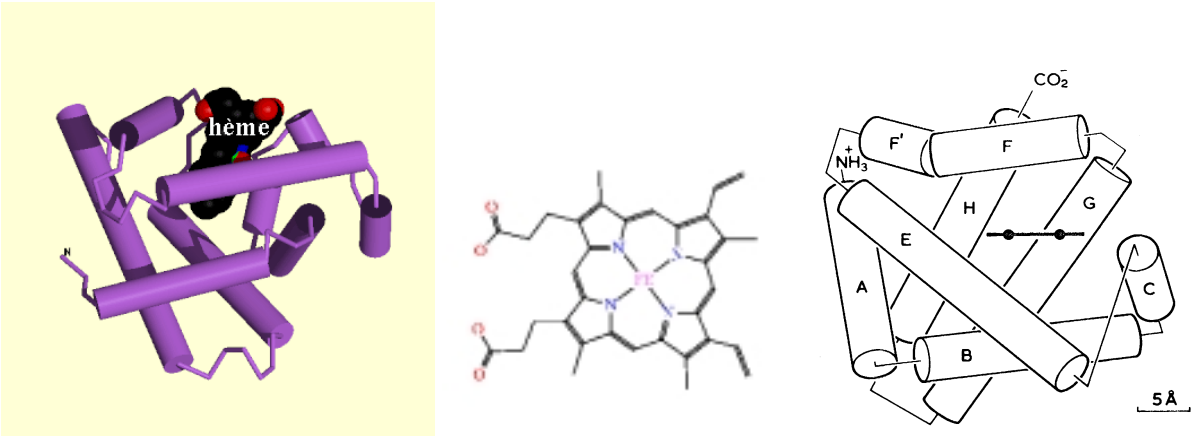


Figure 1. A gauche : structure de la myoglobine du cachalot (*Physeter catodon*) avec le groupement prosthétique (hème) en noir, et les hélices α représentées par des cylindres. Au centre : l'hème ou protoporphyrine IX. A droite : annotation des hélices α représentées par des cylindres. L'hème est plan entre les hélices E, F, G et H.

Chez l'homme, les mammifères, de nombreux vertébrés et invertébrés, la fonction (académique) de l'hémoglobine est de transporter l'oxygène en le fixant au niveau du fer de l'hème (l'atome de fer se rapproche de l'histidine proximale et l'oxygène s'interpose entre l'atome de fer et l'histidine distale). Chez les vertébrés, à l'exception de la lamproie (poisson cyclostome), les hémoglobines sont des tétramères de 68.000 Da résultant de l'association de 4 chaînes polypeptidiques combinée chacune à un hème (fig. 2). Chez l'homme adulte, on distingue deux chaînes α et deux chaînes β. Chez les cyclostomes (poissons vertébrés sans mâchoire) et certains invertébrés, il n'y aurait qu'une seule chaîne polypeptidique.

On trouve également une hémoglobine de faible poids moléculaire (17.000 Da) avec une seule chaîne protéique dans les muscles de différentes espèces, vertébrés comme invertébrés : la myoglobine, dont la fonction est de transporter et stocker l'oxygène au niveau cellulaire.

L'affinité de l'hémoglobine pour l'oxygène varie au sein d'un même organisme. L'oxygène est fixé avec une forte affinité dans l'organe qui se trouve à l'interface du milieu contenant l'oxygène (air ou eau) et relargué grâce à une diminution de l'affinité au niveaux des cellules, pour servir majoritairement d'accepteur final d'électron. Cette affinité est étroitement régulée par deux phénomènes majeurs connus: la coopérativité et l'effet Bohr.

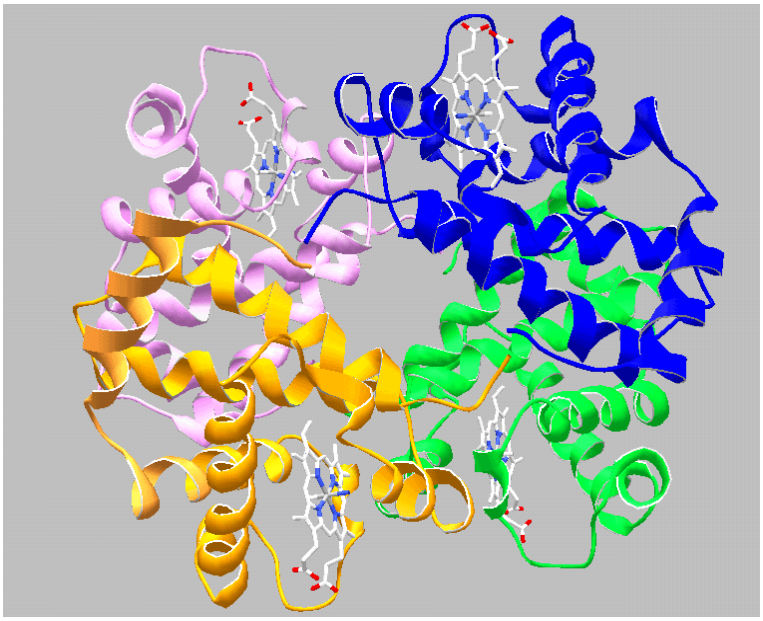


Figure 2. Structure quaternaire de l'hémoglobine tétramérique (2α et 2β). Chaque sous unité est identifiée par une couleur différente et possède un hème.

La majorité des organismes ont plusieurs chaînes de globines et chacune est codée par des gènes différents. On parle donc de **famille multigénique d'hémoglobines**. Retenons que l'émergence de ces familles s'explique par des duplications successives de gènes au cours de l'évolution des espèces. La figure 3 représente la succession des diverses sous unités d'hémoglobines au cours de l'évolution ontogénique de l'homme. Chez ce dernier, il existe deux commutations (switch) lors des passages de la vie embryonnaire à la vie fœtale, et de la vie fœtale à la vie adulte.

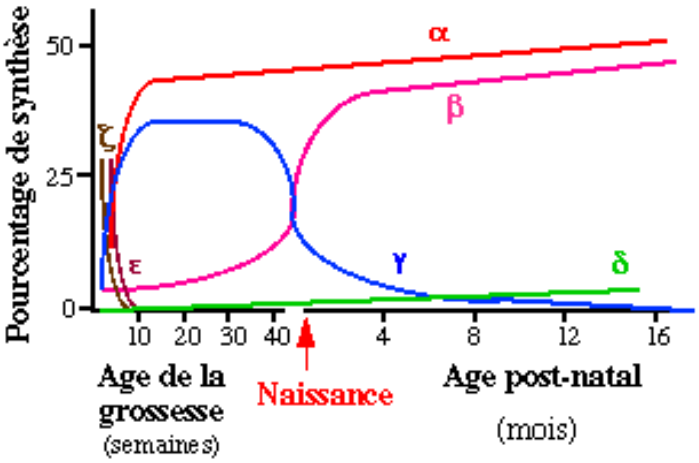


Figure 3.
Succession des diverses sous unités d'hémoglobines au cours de la vie chez l'homme

La famille multigénique des hémoglobines humaines se compose de neuf gènes de globine et d'un gène de myoglobine (fig. 4 et fig. 5).

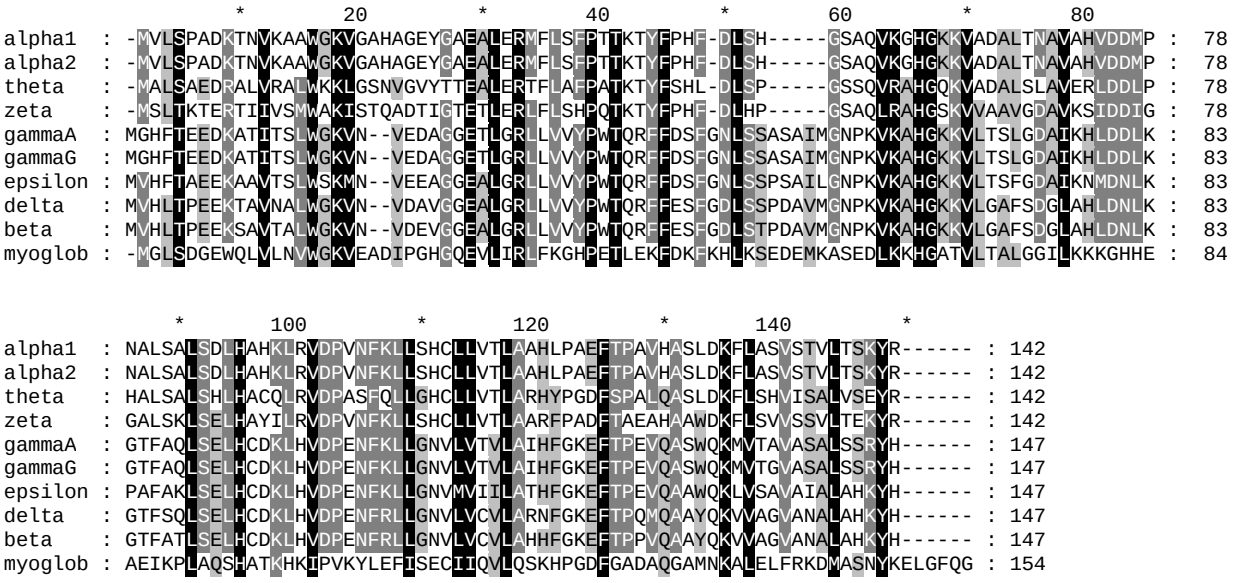


Figure 4. Alignement multiple des globines constitutives des hémoglobines et de la myoglobine humaine

Les différentes hémoglobines (structure quaternaire) humaines sont constituées de plusieurs chaînes. Sur le chromosome 11, on trouve les gènes des chaînes β, γA, γG, δ et ε. Sur le chromosome 16, on trouve les gènes des chaînes α1, α2, ζ et θ. Celui de la myoglobine se trouve sur le chromosome 22. Les hémoglobines adultes sont de type HbA (plus de 95%) formée de 2 chaînes α et de 2 chaînes β et type HbA2 formée de 2 chaînes α et de 2 chaînes δ. L'hémoglobine fœtale HbF est formée de 2 chaînes α et de 2 chaînes γ. Les hémoglobines embryonnaires Gower 1, Gower 2, et Portland sont respectivement formées de 2 chaînes ζ et de 2 chaînes ε, de 2 chaînes α et de 2 chaînes ε, de 2 chaînes θ et de 2 chaînes γ.

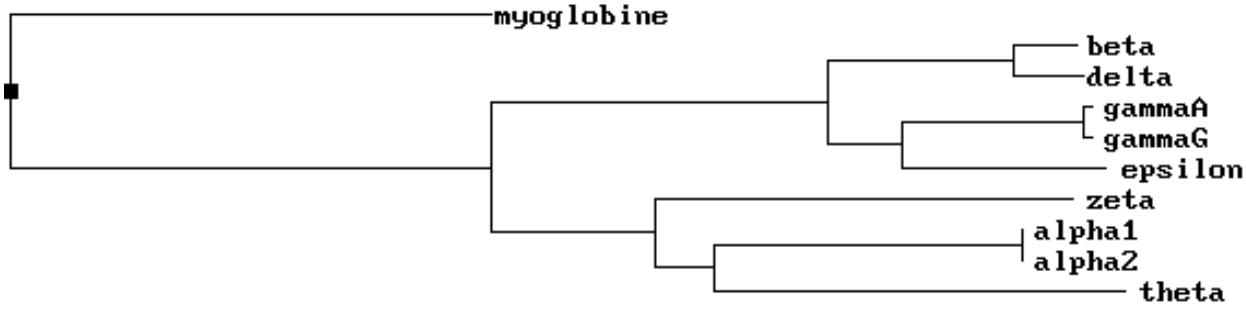


Figure 5. Arbre phylogénétique de la famille multigénique des globines humaines obtenu par Neighbor Joining (tous les nœuds ont une valeur de bootstrap supérieure à 95%)

II. L'hémoglobine : une molécule au cœur de l'émergence des concepts en évolution moléculaire

L'ère de l'évolution moléculaire en tant que discipline scientifique, a vraisemblablement commencé à partir de travaux portant sur l'hémoglobine, protéine ubiquiste dans le monde vivant. Sa diversité moléculaire au sein des organismes s'est rapidement révélée être un miroir de l'évolution et donc un modèle d'évolution moléculaire.

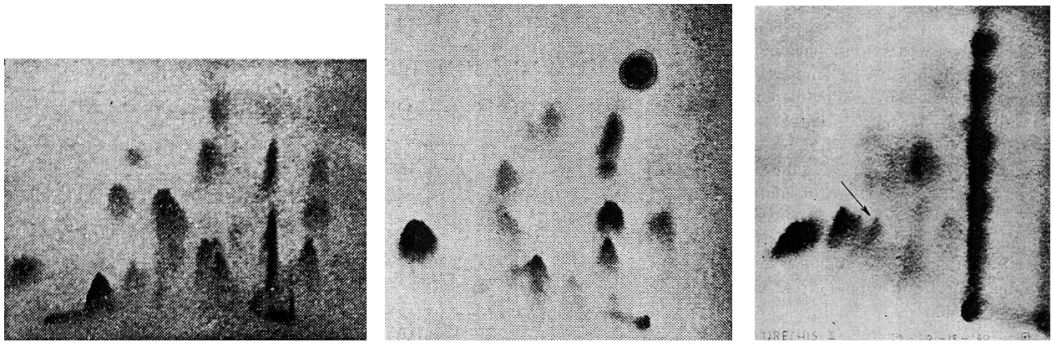
1. L'hémoglobine : marqueur évolutif et homologie moléculaire

Selon Jukes (2000) les premiers discours implicitement teintés d'évolution moléculaire apparaissent au début du XX ième siècle avec l'étude des hémoglobines de vertébrés. Reichert et Brown (1907, 1909) ont observé une divergence des propriétés cristallographiques des hémoglobines, corrélée à la séparation taxonomique des animaux étudiés. Les différences ainsi mises en évidence étaient, à juste titre, associées aux propriétés des globines car les cristaux d'hèmes des différentes espèces considérées étaient identiques.

Plus tard, l'étude de l'origine et des causes moléculaires de l'anémie falciforme (Ingram 1956), précédemment identifiée comme maladie génétique (Pauling et al. 1949), a apporté la première démonstration que la mutation d'un acide aminé au niveau du gène codant l'hémoglobine altérait les propriétés intrinsèques de la protéine (remplacement d'un acide glutamique par une valine) (Ingram 1957).

A cette époque, les techniques de "pré-séquençage" de protéines qui donnaient un patron approximatif dit "empreinte" de la protéine ont permis de révéler une variabilité des hémoglobines au niveau moléculaire et de pointer des différences subtiles existant entre des molécules pourtant homologues (fig. 6).

Le concept d'homologie moléculaire émerge peu à peu, et de façon sous-jacente, l'idée de parenté phylogénétique. Il est lui aussi basé, soit sur l'observation de variations de structures d'hémoglobines ("empreintes" de digestion tryptique) en relation avec l'éloignement phylogénétique des organismes (Zuckerdandl et al. 1960), soit sur la comparaison des structures primaires et tertiaires d'hémoglobines et de myoglobines (Watson et Kendrew 1961).



Humain **Requin** **Urechis caupo**

Figure 6. Chromatographies de produits de digestion tryptique d'hémoglobines chez différents organismes qui montrent que malgré des propriétés similaires, des différences au niveau moléculaire existent. D'après Zuckerdandl et al. (1960)

Les études tridimensionnelles portant sur des cristaux de myoglobine (Kendrew et al. 1960) et d'hémoglobine (Perutz et al. 1960 ; fig.7) ont également fourni des arguments en faveur d'une forte similitude structurale sans pour autant conclure à l'identité de ces structures. Ces travaux soulignent que des mutations affectant le repliement de la protéine auraient été et seraient fortement contre sélectionnées au cours de leur évolution (Ingram 1961).

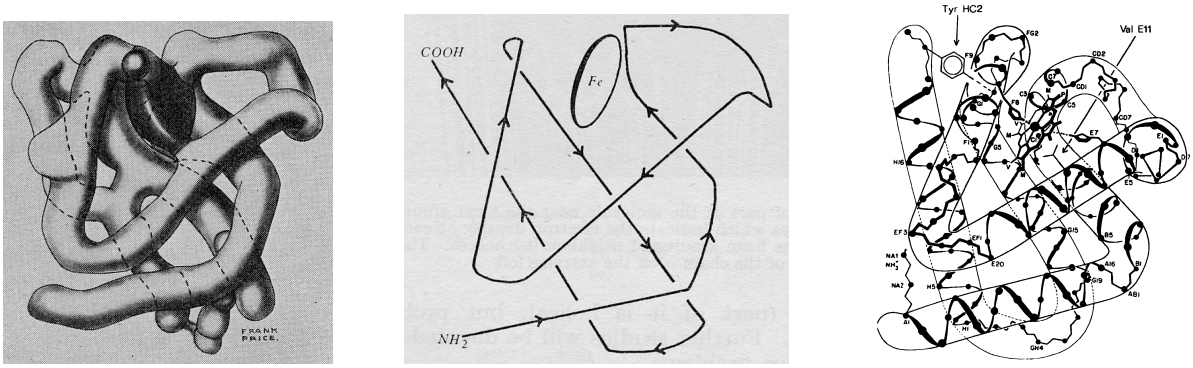


Figure 7. A droite : repliement de la structure tertiaire à 6-Å de la myoglobine (Kendrew et al. 1960) Au centre : orientation N-terminal – COOH-terminal de la chaîne polypeptidique de la myoglobine déduite de la résolution à 2-Å (Kendrew et al. 1960) A gauche : structure tertiare d'une chaîne de globine (Perutz 1983)

Les travaux menés par Riggs (1959) sur l'augmentation de l'effet Bohr en fonction de l'abondance de résidus cystéines dans les hémoglobines qu'il a étudiées, reposaient sur leur capacité de fixation du mercure (qui permet une évaluation du nombre de groupement thiol des cystéines) (fig. 8). Les résultats l'ont conduit à avancer, en terme d'adaptation moléculaire, "qu'il est physiologiquement avantageux pour un animal d'avoir un fort effet Bohr et que les

mutants avec un grand nombre de cystéines (corrélées à cet effet) correctement placées dans la molécule devraient être sélectionnés".

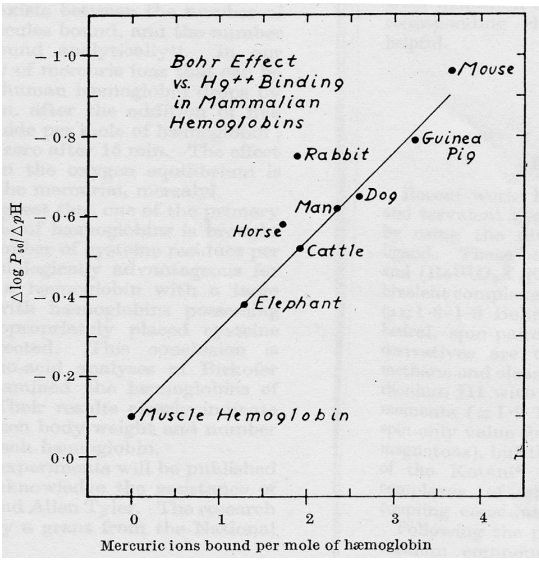


Figure 8. Relation entre effet Bohr et quantité de mercure lié par l'hémoglobine. D'après Riggs (1959)

A partir des années 1960 et 1961 de plus en plus de séquences primaires, partielles ou complètes d'hémoglobines commencent à être disponibles (voir tableau 2, p. 215 dans Zuckerdandl et Pauling (1962), Zuckerdandl (1965)) et les premiers scénarii évolutifs apparaissent. L'accès croissant à l'information moléculaire permet ainsi de replacer bon nombre de résultats dans un contexte évolutif et adaptatif.

2. L'hémoglobine : évolution par duplication des gènes et diversification

La première suggestion (prophétique) d'évolution par duplication d'un gène ancestral suivie par une évolution indépendante des gènes dupliqués a été proposée par Itano (1957), pour expliquer l'existence concomitante de la myoglobine et de l'hémoglobine. Il écrivait même qu'une deuxième duplication serait à l'origine de deux hémoglobines α et β.

Pour décrire l'évolution des gènes codant les hémoglobines humaines, Ingram (1961) avance 3 postulats : (i) la sélection ou l'élimination de ces gènes en fonction de l'intensité de la mutation au cours de l'évolution, (ii) la nécessité de la duplication et de l'évolution indépendante des duplicats, (iii) une différence des taux évolutifs entre les gènes dont les polypeptides correspondant interagissent (α et β par exemple).

Les bases de l'évolution moléculaire sont ainsi jetées, car dans cet article synthétisant l'avancée des connaissances moléculaires sur les hémoglobines, des concepts fondamentaux commencent à être discutés, à savoir, l'ancestralité des gènes, l'homologie, l'homoplasie, les pressions de sélection et les taux évolutifs ainsi que les mécanismes possibles d'innovation au niveau du génome : la duplication et la translocation des gènes. A ce titre, la famille multigénique des hémoglobines humaines (fig. 9) est considérée comme un modèle d'évolution moléculaire car suspectée d'être façonnée par des mécanismes moléculaires généraux. Ingram (1961) conclut, sans citer Itano qui était déjà arrivé à ces conclusions, qu'une unique protéine "myoglobine-like" est à l'origine des hémoglobines modernes.
L'historien des sciences de l'évolution moléculaire doit aussi rappeler que ces mêmes concepts fondamentaux publiés par Ingram en 1961, furent non seulement énoncés lors d'un séminaire à l'université de Dalhousie à Halifax en septembre 1960 par Zuckerkandl, mais étaient également prêts à être publiés. Dans un article rétrospectif, Zuckerkandl (1972) écrit : "La section relative à l'évolution des chaînes d'hémoglobines et la duplication des gènes en général allait être publiée en 1961, quand la parution du travail d'Ingram sur le même sujet faisait apparaître cette publication comme un duplication inutile. Contrairement à la duplication des gènes, celle des papiers ne contribue pas à l'évolution".

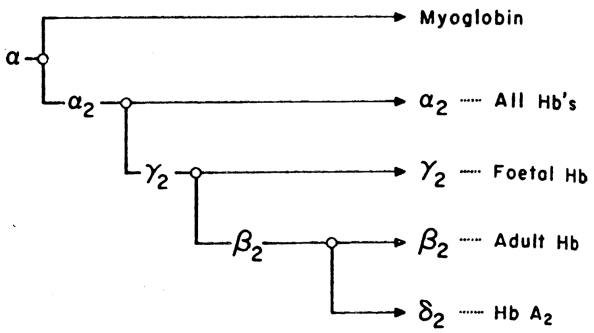


Figure 9.Evolution des chaînes d'hémoglobines selon Ingram (1961).

3. L'hémoglobine : horloge moléculaire, évolution moléculaire neutre et sélection darwinienne positive

A la même époque, l'étude de deux autres protéines, les cytochromes et le trypsinogène conduisent à des concepts identiques (Margoliash 1963).
Dans "Divergence and convergence in proteins", Zuckerdandl et Pauling (1965) introduisent l'idée d'une horloge moléculaire basée sur l'évolution des polypeptides de globines, c'est-à-

dire l'existence d'une apparente proportionnalité entre les remplacements d'acides aminés dans une protéine et le temps de divergence des espèces considérées. Pour étayer cette théorie, les auteurs suggèrent que " les substitutions d'acides aminés seraient limitées pratiquement à des changements fonctionnellement presque neutre (nearly neutral changes) ".

Ces considérations nouvelles qui introduisent la notion d'un effet neutre ou sélectif des mutations, mitigent avant l'heure, comme une mise en garde, les débats qui opposeront les neutralistes et les sélectionnistes à partir de 1970.

La description de la double hélice d'ADN (Watson et Crick 1953) et les travaux de Nirenberg et al. (1965) sur la caractérisation du code génétique vont permettre de progresser dans la compréhension fine des mutations, d'abord observées au niveau des polypeptides, et de décrire plus précisément les fréquences de mutations et les pas évolutifs probables qui séparent des protéines homologues, notamment à partir de l'étude des hémoglobines anormales (Beale et Lehmann 1965 ; Vogel et Röhrborn 1966). Les premières matrices de substitutions basées sur le code génétique apparaissent, et visent à la fois à perfectionner la détection des homologies moléculaires et à mieux comprendre l'impact des mutations nucléiques au sein des protéines (Fitch 1966, Fitch 1967).

Toutes ces nouvelles données vont permettre les premières reconstructions phylogénétiques assistées par ordinateur (Fitch et Margoliash 1967). Dans un souci d'obtenir des phylogénies cohérentes Fitch (1970) clarifie le concept d'homologie en introduisant les termes d'orthologie et de paralogie (en prenant l'exemple des hémoglobines α et β). Il énonce la règle fondamentale selon laquelle c'est l'histoire des gènes apparus par spéciation, les gènes orthologues, qui reflète l'histoire des espèces et non celle des gènes paralogues, qui traduit l'émergence d'une famille multigénique.

D'autre part, en se basant sur des calculs de taux évolutifs obtenus à partir de séquences d'hémoglobines et de cytochrome c, Kimura (1968) considère que les mutations produites par remplacement d'un nucléotide sont beaucoup trop fréquentes pour ne pas altérer le génome des espèces de mammifères : selon lui, les mutations sont pratiquement sélectivement neutres et la force évolutive qui prévaut sur la structuration génétique des populations biologiques est la dérive génétique.

A partir d'arguments d'ordre fonctionnel basés sur l'observation des hémoglobines et d'autres protéines, King et Jukes (1969) adhèrent à la théorie neutraliste dans un article au titre provocateur " Non-Darwinian evolution ". Jukes (2000) aurait préféré *a posteriori,* le titre de "Non-adaptative evolution " pour renforcer l'idée que les mutations qui s'accumulent au sein des protéines au cours de l'évolution ne sont pas toutes éliminées par la sélection naturelle. Il

précise aussi que "la plupart des protéines contiennent des régions où les substitutions d'acides aminés ne produisent pas de changements sensibles dans la fonction de la protéine" : ceci suggère une pression évolutive plus subtile oscillant entre sélection et évolution neutre.
Toujours à propos des hémoglobines prises comme modèle d'évolution moléculaire, un débat assez virulent va opposer Goodman (Goodman et al. 1971 ; Goodman et al. 1975) et Kimura (1981) sur : (i) l'accélération du taux de substitution des acides aminés au sein des globines de vertébrés et (ii) sur l'interprétation de cette accélération traduite par Goodman comme de la sélection darwinienne positive en tant que processus adaptatif lié à l'acquisition de nouvelles fonctions après duplication. A cet égard, Kimura prône un relâchement des contraintes sélectives qui permet de révéler un polymorphisme par l'existence de mutants devenus sélectivement neutres et donc mimant un effet d'accumulation de mutations.

Le débat sélectionniste-neutraliste, qui est en fait une controverse conceptuelle, a permis de générer de nombreux travaux sur les familles multigéniques, dont celles des hémoglobines, et a ouvert le débat sur l'innovation et l'adaptation moléculaire liées aux étapes suivant immédiatement un évènement de duplication. Ce débat a été progressivement atténué par les travaux, majoritairement théoriques d'Ohta qui a élaboré la théorie "presque neutre" de l'évolution dans laquelle la dérive génétique et la sélection agissent, mais à des moments différents. Ohta (1988) maintient que la sélection Darwinienne positive est nécessaire pour obtenir de nouveaux gènes avec des fonctions différentes après duplication, alors que les gènes dont la fonction a été fixée depuis longtemps évoluent majoritairement par la dérive génétique. Cette théorie plutôt unificatrice est confortée par les données moléculaires de plus en plus nombreuses dont certaines sont issues de travaux sur les hémoglobines (Fitch et al. 1991). Elle a pratiquement réconcilié les détracteurs de l'une ou l'autre des théories (Ohta 1993 ; Ohta 1994). Aujourd'hui, la sélection darwinienne positive qui opère après un évènement de duplication, bien qu'elle soit techniquement et conceptuellement difficile a démontrer, est de plus en plus invoquée et caractérisée (Chapitre 5 et 6). Elle est considérée comme un moteur évolutif puissant de l'évolution des génomes et de l'acquisition de nouvelles fonctions ou sub-fonctionnalités par l'accumulation de substitutions non-synonymes (qui changent la nature des acides aminés) par rapport aux substitutions synonymes. La détection d'évènements d'adaptation moléculaire est réalisée par des méthodes statistiques qui reposent sur des algorithmes de maximum de vraisemblance et des probabilités bayesiennes (Yang et Bielawski 2000). Ces différents modèles sont décrits dans le chapitre 4.

III. Evolution des hémoglobines : structure des gènes et l'histoire des hémoglobines

Evolutionary History Written in Globin Genes

Science,
vol.214,
Octobre1981

Genes that duplicate, jump, and diverge: these are the key elements in the evolution of globin gene clusters

(Lewin 1981)

Parce qu'elle est abondante et facilement purifiable, que l'on a également identifié des pathologies et des mutations correspondantes, l'hémoglobine a été une protéine de choix et un modèle pour la compréhension des mécanismes moléculaires. Elle est au cœur des premiers travaux sur les relations existant entre structure des gènes (exons et introns), domaines fonctionnels et fonction de la protéine.

1. Relation entre la structure des gènes et les domaines protéiques

La découverte de l'existence de séquences non codantes (introns) insérées entre les séquences codantes (exons) d'un gène, en d'autres termes, la mise en évidence de la structure en mosaïque des gènes d'eucaryotes, a soulevé des questions en termes de relation entre la structure du gène et la fonction de la protéine. L'hémoglobine humaine et celle de certains vertébrés dont les deux introns avaient été localisés et dénombrés (voir références dans Craik et al. (1980)) a servi de modèle pour tester l'hypothèse de Gilbert (1978) selon laquelle les exons correspondent à des domaines fonctionnels de la protéine. Implicitement, il suggérait que la fonctionnalité d'une protéine s'explique par une synergie de propriétés intrinsèques portées par différents domaines de la protéine.

Fondée sur les travaux qui avaient montré la présence de l'histidine proximale dans l'exon central, son hypothèse prédisait que l'exon central de la globine devait fixer spécifiquement l'hème. En isolant le polypeptide correspondant à l'exon central, Craik et al (1980) ont démontré et confirmé sa capacité à lier le groupement prosthétique.

Cependant, la poursuite des expériences de Craik et al (1981) sur la capacité du seul exon central à lier réversiblement l'oxygène a montré que le seul second exon était trop court et qu'un repliement incomplet du polypeptide ne permettait pas de restaurer la fonction de liaison de l'oxygène.

2. Position des introns : marqueur de l'évolution des gènes codant l'hémoglobine

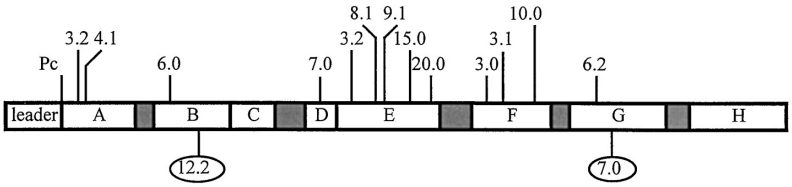
En cherchant une éventuelle correspondance entre les exons et des unités structurales de l'hémoglobine, Go (1981) a prédit que l'exon central devait être le résultat de la fusion de deux exons, avec la perte d'un intron. Cette remarquable prédiction fut confirmée par la description d'un intron supplémentaire aux positions attendues dans le gène codant la leghémoglobine des plantes légumineuses comme le soja (Jensen et al. 1981) puis dans le gène codant l'hémoglobine de plantes non-légumineuses (Bogusz et al. 1988). Plus tard, la découverte de ces deux gènes dans le même génome de ces plantes (Andersson et al. 1996 ; Trevaskis et al. 1997) a renforcé l'idée d'une hémoglobine ancestrale aux hémoglobines et aux leghémoglobines.

Ces travaux portant sur la localisation des introns et sur la correspondance entre exons et unités structurales fonctionnelles, ont suggéré l'existence d'un gène ancestral d'hémoglobine présentant le même motif tridimensionnel (le globin-fold) commun aux plantes et aux animaux (Hardison 1996).

Les gènes d'hémoglobines sont retrouvés chez de nombreuses espèces appartenant à des phyla eucaryotiques différents et contiennent majoritairement deux introns au niveau des hélices B et G. Dans la famille des hémoglobines d'arthropodes la position et le nombre d'introns peut varier (Dewilde et al. 1998a, Burmester et Hankeln 1999 ; fig. 10). En fait, de nombreux travaux ont montré que la diversité dans le nombre et la position des introns pouvait être attribuée à des insertions aléatoires de type transposition : la perte des introns, les positions différentes ou l'insertion de nouveau introns sont autant d'arguments qui démontrent la mobilité des introns ("intron sliding").

Ces cas de figure semblent rapprocher les deux théories opposées sur l'origine des introns : l' "intron early theory" selon laquelle les introns auraient été utilisés pour assembler les premiers gènes, et l' "intron late theory" selon laquelle les introns auraient été ajoutés au cours de l'évolution pour scinder les gènes continus originaux (Gilbert et al. 1997).

La présence d'un troisième intron (central) dans l'hélice E, (en plus des deux autres conservés dans les hélices B et G) dans le gène de la globine du nématode *Pseudoterranova sp*, en position similaire (mais pas identique) à celui trouvé chez les plantes, renforce un peu plus l'hypothèse d'un gène d'hémoglobine ancestral commun entre les plantes et les animaux, avant leur divergence il y a 1500 millions d'années (Hardison 1996).



	1							
Vertebrates	-		B12.2			E15.0		G7.0
Plants	-		B12.2			E15.0		G7.0
Chironomus thummi IIB	+					E9.1		
Chironomus melanotus	+					E15.0		
Gastrophilus	-				D7.0			G7.0
Artemia T3			B12.2				F3.1	
Artemia T4								G6.2
Artemia T6			B12.2				F3.0	
Anadara	+	Pc^2	B12.2					G7.0
Barbatia	+	Pc^2	B12.2					G7.0
Biomphalaria	-	A3.2	B12.2					G7.0
Ascaris	+	A4.1	B12.2			E8.1		G7.0
Pseudoterranova	+	A4.1	B12.2			E8.1		G7.0
Caenorhabditis	-					E3.2		
Nippostrongylus	-		B12.2			E3.2		G7.0
Chlamydomonas	-		B6.0			E20.0	F10.0	
Paramecium	-						F3.0	

Figure 10. Diversité des positions d'insertion des introns dans la famille des hémoglobines. **B12.2** : intron dans le codon 12 en phase 2 (après la seconde base) de l'hélice B. **G7.0** intron inséré entre le codon 6 et 7 de l'hélice G. **Colonne 1** : présence (+) ou absence (-) d'une séquence d'adressage dans le gène considéré. **Pc^2** : intron précédent la séquence codante. D'après Dewilde et al (1998a)

3. Les hémoglobines sont membres d'une superfamille

On rassemble dans une superfamille des protéines qui présentent certaines similitudes d'ordre fonctionnel et structural. Un exemple assez bien connu, la superfamille des hémocyanines d'arthropodes, composée des phénoloxydases d'arthropodes (tyrosinases), des pseudohemocyanines de crustacés (cryptocyanines) et des hexamérines d'insectes, montre que des familles de protéines ont pu évoluer à partir de familles pré-existantes. En effet, il a été avancé que les hémocyanines d'arthropodes auraient émergées à partir des phénoloxydases (Burmester 2001).

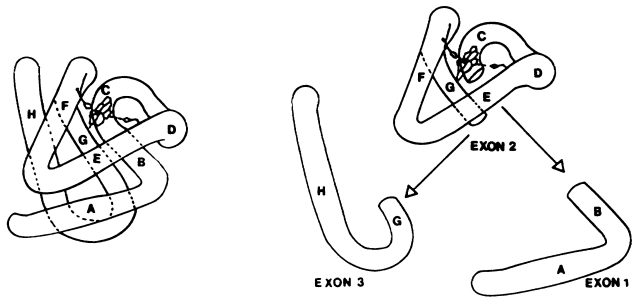
Les hémoglobines seraient, elles aussi, membres d'une superfamille.

Lors des travaux portant sur la fonctionnalité singulière des peptides codés par les exons, le deuxième exon de l'hémoglobine, privé de ses exons flanquants a montré un spectre d'absorption comparable à ceux des cytochromes ainsi qu'un cycle de réduction et ré-oxydation similaire avec un traitement alternatif dithionite / oxygène moléculaire (Craik et al. 1981).

La nature et la distribution spatiale des résidus de la globine en contact avec l'hème correspondent à celles trouvées chez plusieurs types de cytochromes b. Cette observation a permis à ces auteurs de suggérer que l'exon central pouvait être une unité évolutive ancienne commune, codant pour une structure plus universelle d'une hémoprotéine ancestrale, de type "hème binding" ou plus généralement "porphyrin binding". A partir de travaux basés sur la comparaison de séquences primaires de globines et de cytochromes de type b, Runnegar (1984) a proposé que les globines dériveraient de ces mêmes protéines. Ainsi les gènes codant les globines pourraient résulter de l'addition de deux exons flanquants (exon 1 et exon 3 modernes), à partir d'un exon codant pour un peptide fixant l'hème commun aux cytochromes et aux globines (Holland et Blake 1987, fig. 11). Cette adaptation aurait permis des conditions structurales conférant la capacité de lier réversiblement l'oxygène, ce qui n'est pas possible avec le produit de l'exon central seul (cf ci-dessus Craik et al. 1981). Des travaux portant sur les phycocyanines, protéines des phycobilisomes, complexes protéiques destinés à la capture des photons que l'on trouve chez les cyanobactéries et les algues rouges, montrent qu'elles auraient un ancêtre commun avec les hémoglobines. Sur la base d'études tridimensionnelles, ces deux familles de molécules montrent une même organisation spatiale dans leur repliement (Schirmer et al. 1985 ; Pastore et Lesk 1990).

Figure 11

Les sous-unités fonctionnelles (exons) de l'hémoglobine et leur position spatiale par rapport au groupement prosthétique. D'après Holland et Blake (1987)



La dehaloperoxidase est une enzyme isolée d'un polychète térébèllidé, *Amphitrite ornata.* Elle dégrade des composés halogénés répulsifs et toxiques émis par d'autres polychètes, dans le cadre d'une "guerre chimique" territoriale. Cette peroxydase à également une poche héminique et une histidine proximale, comme les globines et les cytochromes. De plus le cristal de cette enzyme a un repliement typique du "globin fold" (LaCount et al. 2000). Selon les auteurs, la déhaloperoxidase aurait évolué à partir d'un transporteur d'oxygène, de type hémoglobine.

L'ensemble de ces travaux suggère qu'une superfamille aurait émergé à partir d'une hémoprotéine ancestrale en se diversifiant et donnant des lignées moléculaires avec des caractéristiques fonctionnelles divergentes tout en gardant une unité fonctionnelle

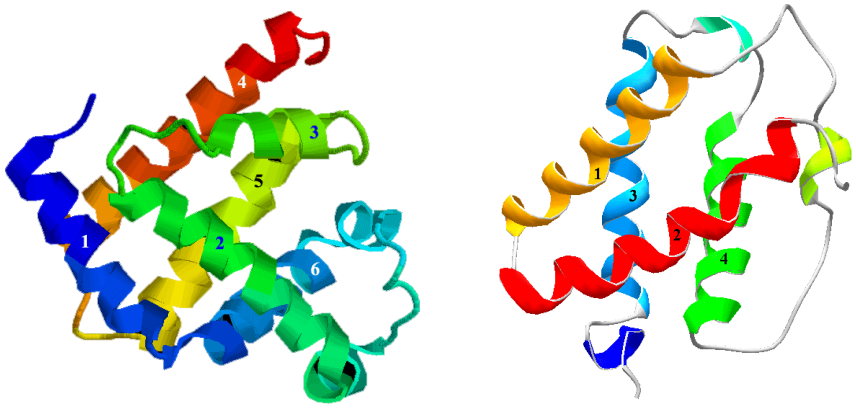
fondamentale de type "globin fold –like" capable de fixer une structure linéaire tétrapyrolique héminique ou dérivée de l'hème.

Retenons toutefois, qu'il est nécessaire de connaître un plus grand nombre de structures et/ou d'avoir recours à des analyses très fines de phylogénie moléculaire, pour s'affranchir complètement de la question de la possible convergence évolutive de ces protéines comme le montrent les travaux sur l'indoléamine "dioxygénase-like myoglobin" une enzyme à hème dégradant le tryptophane, chez le mollusque *Sulculus diversicolor* (Suzuki et al. 1996). Après la perte ou l'altération du gène codant la myoglobine de *Sulculus diversicolor*, l'activité enzymatique originale de l'indoléamine dioxygénase aurait été modifiée en faveur de la fixation de l'oxygène par des mutations dans ou proches de la poche de l'hème.

4. Hémoglobines tronquées

Il faut aujourd'hui, dans la classification des hémoglobines, prendre en compte une famille émergente d'hémoglobines, longtemps ignorées car présentes à de très faibles concentrations chez de nombreux organismes, les hémoglobines tronquées. Elles n'ont pas été trouvées chez les métazoaires et les archéobactéries et ne sont présentent que chez les eubactéries, les plantes, les protozoaires et les algues (Wittenberg et al. 2002). La présence simultanée d'hémoglobines tronquées chez *Mycobactérium avium* (Gram positive) et *Methyloccocus capsulatus* (protéobactérie), indique que différentes lignées d'hémoglobines tronquées ont divergé d'un ancêtre commun avant la séparation des lignées majeures de procaryotes. Ces hémoglobines un peu particulières, plus courtes de 20 à 40 acides aminés, dont on connaît maintenant la structure 3D, ont une organisation structurale dite "two-over-two" (deux hélices alpha faisant face à deux autres) alors que les hémoglobines classiques présentent un repliement de type "three-over-three" (Pesce et al. 2000, fig. 12).

Figure 12



A gauche : repliement "three-over-three" classique. Les hélices α 1, 2 et 3 font face aux hélices α 4,5 et 6.

A droite : repliement "two-over-two"des hémoglobines tronquées. Les hélices α 1 et 2 font face aux hélices α 3 et 4.

5. Hexacoordination et pentacoordination des hémoglobines

Avec les hémoglobines récemment caractérisées, une distinction supplémentaire d'ordre structural et fonctionnel apparaît : on parle dans de **pentacoordination** dans le cas des hémoglobines au sein desquelles l'atome de fer en absence du ligand est stabilisé au centre du noyau tetrapyrolique par 5 interactions de type électrostatique, et d'**hexacoordination (**fig 13) pour les hémoglobines qui utilisent un acide aminé supplémentaire pour stabiliser l'atome de fer.

Cette différence permet d'expliquer des fonctions atypiques des hémoglobines "hexacoordinées" qui sont capables d'affinités plus fortes et plus rapides pour leur ligand. En effet, le déplacement intramoléculaire de la chaîne latérale d'un acide aminé pour libérer le site de fixation d'un ligand a des conséquences sur les vitesses de fixation. Les mouvements moléculaires sous-jacents agiraient dans la régulation de fixation du ligand dans certaines conditions physiologiques. Des mécanismes récemment identifiés, plus subtils, devraient également permettre de décrire d'autres procédés moléculaires de fixation des ligands qui reposent sur l'état énergétique des atomes : les conformations "low spin hexacoordinate" et "high spin hexacoordinate" (Couture et al. 1999a).

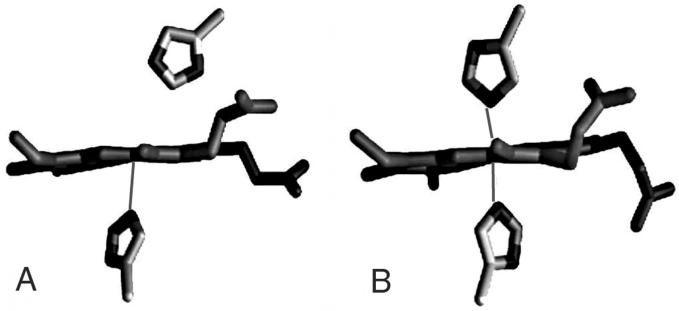


Figure 13. Etat de coordination de l'hème.
A, pentacoordination : seule l'histidine proximale stabilise le fer
B, hexacoordination : les deux histidines proximale et distale coordonnent le fer.
D'après Trent et Hargrove (2002)

IV. L'hémoglobine dans les règnes du vivant

Dans la famille des hémoglobines, au sein de cette lignée moléculaire qui a été très tôt décrite comme un pigment respiratoire liant réversiblement, stockant et transportant l'oxygène, la découverte d'hémoglobines aux fonctions originales, dans des organismes procaryotes et eucaryotes a montré que cette molécule était capable de fonctions moléculaires différentes et inattendues. Ces fonctions cryptiques ont été dévoilées grâce à la résolution croissante des appareils de mesures et à l'amélioration des techniques d'analyses. Ces découvertes détrônent

la fonction universelle unique très tôt attribuée aux hémoglobines, celle de transporter l'oxygène. De plus en plus, des hémoglobines putatives ("hémoglobin-like") ont été d'abord identifiées *in silico*, puis séquencées (Kaneko et al. 1996 ; Burmester et Hankeln 1999). Leur présence dans le génome, dans le transcriptome, leur expression et leur fonctionnalité attestent qu'elles ont été retenues par le crible de la sélection naturelle. Bien que leur fonction exacte n'ait pas toujours pu être identifiée, l'ensemble des travaux suivants montre combien une lignée moléculaire a pu être recrutée pour différentes fonctions dans des lignées évolutives aussi éloignées que des bactéries, des archéobactéries, des champignons, des plantes ou des animaux. Les exemples suivants, qui ne sont pas exhaustifs, illustrent la diversité fonctionnelle et parfois structurale des hémoglobines. Ils révèlent qu'une classification par type de fonction ne convient pas forcément. Nous décrirons ainsi la diversité des hémoglobines au sein des groupes suivants : plantes, bactéries et levures, algues, protozoaires, humaines et mammifères, puis invertébrés.

1. Hémoglobines des levures et bactéries

Chez les levures, dont *Saccharomyces cerevisiae*, et certaines eubactéries dont *Escherichia coli*, les hémoglobines décrites ont en plus d'un domaine NH_2-terminal contenant un hème et un "globin-fold", un domaine COOH-terminal à co-enzyme FAD ou $NADP^+$, ce dernier étant structurellement proche des réductases (fig. 14).

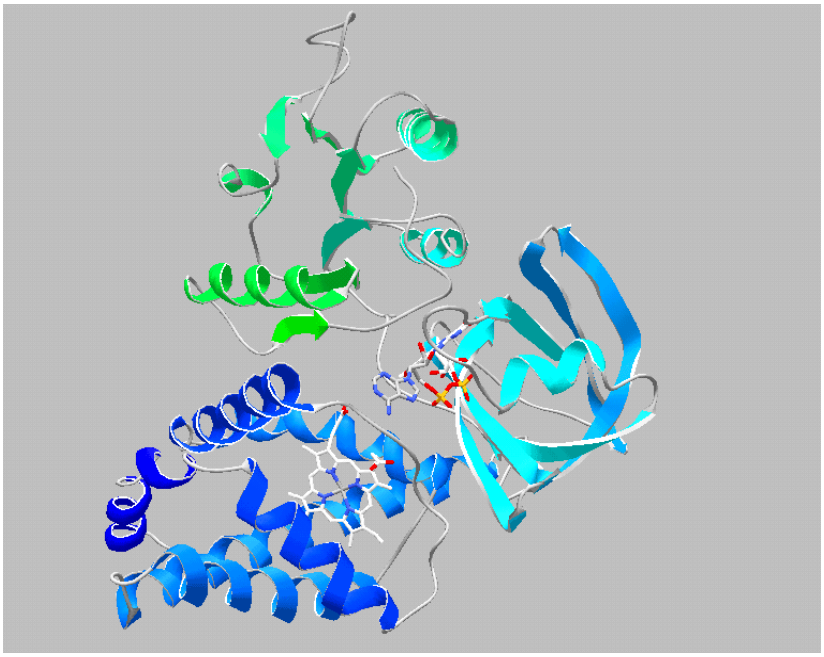


Figure 14.
Flavohémoglobine de *Alcaligenes eutrophus* : hémoglobine chimérique. Le "globin fold", avec au centre l'hème, est en bleu marine (hélices α), et le domaine à flavine est en bleu clair (feuillets β) et vert clair (hélices α et feuillets β) avec au centre le groupement FAD. (Structure 3D de la flavohémoglobine de *Alcaligenes eutrophus* n° accession 1CQX)

Ces hémoglobines sont appelées flavohémoglobines (Zhu et Riggs 1992). Elles ont été l'objet de nombreux travaux et semblent principalement avoir une fonction enzymatique de détoxication de la forme radicalaire du monoxyde d'azote (en formant des nitrates, NO_3) et participent ainsi à la gestion du stress oxydatif en tant qu'oxyde nitrique dioxygénase (Gardner et al. 1998 ; Poole et Hughes 2000). Récemment, il a été montré que la partie globine de cette molécule chimérique présente un site catalytique peroxidase-like suggérant que l'hème serait davantage impliqué dans la fixation d'espèces radicalaires de l'oxygène que dans le transport d'oxygène.

Un autre type d'hémoglobine a été identifié chez la bactérie *Vitreoscilla stercoraria*. Elle diffère des autres hémoglobines bactériennes, car elle ne possède pas de cofacteurs rédox (domaine à flavines comme dans l'exemple ci-dessus). Le domaine fixant l'hème partage jusqu'à 56 % d'identité avec le domaine de fixation de l'hème des flavohémoglobines. Le rôle de ces hémoglobines serait de faciliter la diffusion de l'oxygène jusqu'aux cytochromes oxydases (Tarricone et al. 1997), car leur surexpression dans *Escherichia coli* en conditions hypoxiques permet une croissance plus rapide et des cultures plus denses que celles obtenues avec les souches sauvages.

La bactérie filamenteuse actinomycète fixatrice d'azote atmosphérique et formant des nodules avec certaines plantes (l'aulne, *Alnus sp*) possède également une hémoglobine qui a pu être caractérisée (sans être séquencée). Ses caractéristiques spectrales, de constante d'association et de dissociation permettent de la rapprocher des hémoglobines tronquées (Tjepkema et al. 2002).

Chez l'archaébactérie *Halobacterium salinarum*, une protéine chimérique 'myoglobin-like' possède un domaine N-terminal de type globin-fold avec un hème et un domaine C-terminal de type MCP (methyl-accepting chemotaxis proteins). Les travaux portant sur cette hémoglobine (Hou et al. 2000) suggèrent que la fixation de l'oxygène sur l'hème entraîne un changement structural qui interfère avec le domaine MCP impliqué dans la reconnaissance de signaux intracellulaires, et permettrait ainsi à ces bactéries de s'orienter en fonction de l'abondance de l'oxygène, et peut-être d'autres gaz dans le milieu (fig 15). Ce chimiotactisme, et particulièrement cet aérotactisme, a également été retrouvé chez le bacille *Bacillus halodurans* qui possède une protéine similaire de type "globin-coupled oxygen sensor" (Hou et al. 2001a ; Hou et al. 2001b).

Fréquemment, les hémoglobines tronquées et les flavohémoglobines sont co-exprimées au sein de la même bactérie, suggérant des fonctions spécifiques pour chacune d'elles (Wittenberg et al. 2002).

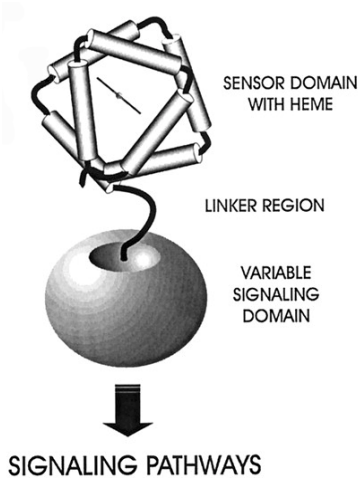


Figure 15.
Représentation schématique des hémoglobines bactériennes impliquées dans la reconnaissance du niveau d'oxygène dans le milieu. D'après Hou et al. (2001b)

Chez des bactéries pathogènes comme *Mycobacterium tuberculosis* l'hémoglobine tronquée servirait à protéger le bacille contre les émissions de formes radicalaires du monoxyde d'azote émis par les macrophages pendant les réactions immunitaires en transformant NO en nitrate (Couture et al. 1999b).

Chez les cyanobactéries, dont l'activité photosynthétique serait à l'origine de la modification de l'atmosphère il y a entre 3,5 et 2 milliard d'années (Schopf et Walter 1982), des hémoprotéines de la famille des hémoglobines, les cyanoglobines sont synthétisées. Chez *Nostoc commune,* une cyanobactérie fixatrice d'azote atmosphérique, le rôle des cyanoglobines qui sont des hémoglobines tronquées, serait de chélater l'oxygène pour protéger l'activité enzymatique des nitrogénases (Potts et al. 1992). La localisation des cyanoglobines sur la face cytosolique de la membrane cellulaire suggère que ces hémoglobines présentent l'oxygène à un complexe de type cytochrome oxydase. D'autres cyanobactéries telles que *Synechocystis sp* possèdent également des cyanoglobines dont les fonctions n'ont pas encore été identifiées. La cyanoglobine hexacoordinée de *Synechocystis sp* confère un site de fixation du ligand (O_2) extrêmement réactif (Hvitved et al. 2001).

2. Hémoglobines des plantes supérieures

Les hémoglobines des nodules de plantes symbiotiques permettent à la fois de fournir les bactéroïdes en oxygène qui à partir duquel ils produisent de l'ATP pour la nitrogénase. Cette liaison de l'oxygène permet de maintenir une pression partielle faible en oxygène pour ne pas inhiber l'activité enzymatique de la nitrogénase (Appleby 1992). Les hémoglobines de plantes non symbiotiques sont hexacoordonnées (Arredondo-Peter et al. 1998) et ont probablement un rôle de diffusion de l'oxygène dans les tissus. Elles sont exprimées dans différents tissus de la

plante, pendant la germination également, et vraisemblablement dans des tissus en croissance, en intense besoin d'énergie (Sowa et al. 1999).
De récents travaux ont montré la présence d'hémoglobines tronquées chez les plantes (angiospermes et bryophytes). Elles présentent un domaine central similaire aux hémoglobines tronquées déjà identifiées et ont un état transitoire d'héxacoordination qui revient à un état de pentacoordination après déoxygénation. Ces caractéristiques laissent présager des fonctions spécifiques qui n'ont pas encore été décrites (Watts et al. 2001).

3. Hémoglobines des algues

Chez l'algue verte unicellulaire *Chlamydomonas eugametos* trois hémoglobines co-existent, deux "traditionnelles" et une tronquée. Les deux hémoglobines non tronquées sont exprimées par le génome nucléaire et adressées dans le chloroplaste. Pour l'une d'elles, différentes conditions de pH ont révélé différentes formes d'état de coordination (4, 5 et 6-coordinations) pH-dépendant (Couture et al. 1999a), mais pour l'instant les fonctions biologiques correspondantes n'ont pas encore été élucidées. L'hémoglobine tronquée est aussi exprimée à partir du génome nucléaire, en réponse à l'activation de la photosynthèse. Sa localisation après exportation, près des thylacoïdes dans le chloroplaste, suggère qu'elle est impliquée dans des transferts d'électrons (Couture et al. 1994).

4. Hémoglobines de protozoaires ciliés

Des protozoaires unicellulaires possèdent également des hémoglobines dont les fonctions biologiques ne sont pas encore connues. Les hémoglobines du thermophile *Tetrahymena pyriformis* (Iwaasa et al. 1989) et *Paramecium caudatum* (Iwaasa et al. 1990) ont été séquencées et présentent elles aussi un état d'hexacoordination.

5. Hémoglobines de vertébrés, de mammifères et de l'homme

La famille multigénique des hémoglobines humaines a été la première à être caractérisée, en termes de nombre de gènes, de localisation chromosomique, de lieu de synthèse, de succession et d'expression au cours du développement (Chapitre 1, I). Le rôle de transporteur d'oxygène a été rapidement élevé au rang de fonction universelle des hémoglobines car ce

pigment et ses principales caractéristiques ont été retrouvés chez les autres mammifères et tous les vertébrés.
Pourtant, la description d'hémoglobines ayant des fonctions différentes ne manque pas. Des hémoglobines humaines et de vertébrés possédant des fonctions nouvellement découvertes sont présentées ci-dessous.

a. Fixation du monoxyde d'azote (NO)

Les sous-unités β de l'hémoglobine humaine, de tous les mammifères et de la plupart des vertébrés possèdent une cystéine (cysβ93) avec un groupement thiol hautement réactif, dont la fonction est restée longtemps inconnue. Il a été montré chez les mammifères que cette cystéine fixe le monoxyde d'azote et forme de la S-nitrosohémoglobine (SNO-Hb) dans les globules rouges au niveau des poumons. Pendant le trajet artériel et veineux, le monoxyde d'azote est relargué et joue un rôle au niveau du contrôle de la pression artérielle (Jia et al. 1996). Les travaux qui portent sur le monoxyde d'azote et l'hémoglobine ont conduit à un nouveau paradigme sur la fonction de l'hémoglobine chez les mammifères, les oiseaux et poissons et de nombreux vertébrés. Comme le soulignent Durner et al (1999), "Ces nouvelles observations ont conduit à la révision de notre compréhension du cycle respiratoire, et à y inclure un troisième gaz, le monoxyde d'azote. Dans leur forme dé-oxygénée, les érythrocytes passent dans les capillaires et transportent NO fixé sur l'hème, ainsi que le CO_2. Dans les poumons, l'hémoglobine subit un changement allostérique de T vers R qui est couplé avec le transfert de NO de l'hème à la cystéine 93. La S-nitroso-oxyhémoglobine (avec NO lié au groupement thiol et O_2 sur l'hème) entre ensuite dans le circuit artériel. Exposé à de basse tension en O_2, la transition de l'état R vers l'état T relargue NO qui dilate les vaisseaux sanguins et facilite le relarguage de O_2 ". Ce mécanisme est illustré dans la figure 16.
Le cycle SNO-hémoglobine tel qu'il est décrit ci-dessous, a également été montré au niveau de la circulation fœtale.
A la question : " Est-ce que le nouveau paradigme du cycle SNO-hémoglobine devient obsolète avec des hémoglobines sans cystéine en position β93 ? La réponse est non ! En effet, de récents travaux sur le poisson perciforme *Leiostomus xanthurus* montrent que la cystéine fixatrice de NO n'a pas besoin d'être en position 93 (Bonaventura et Lance 2001).

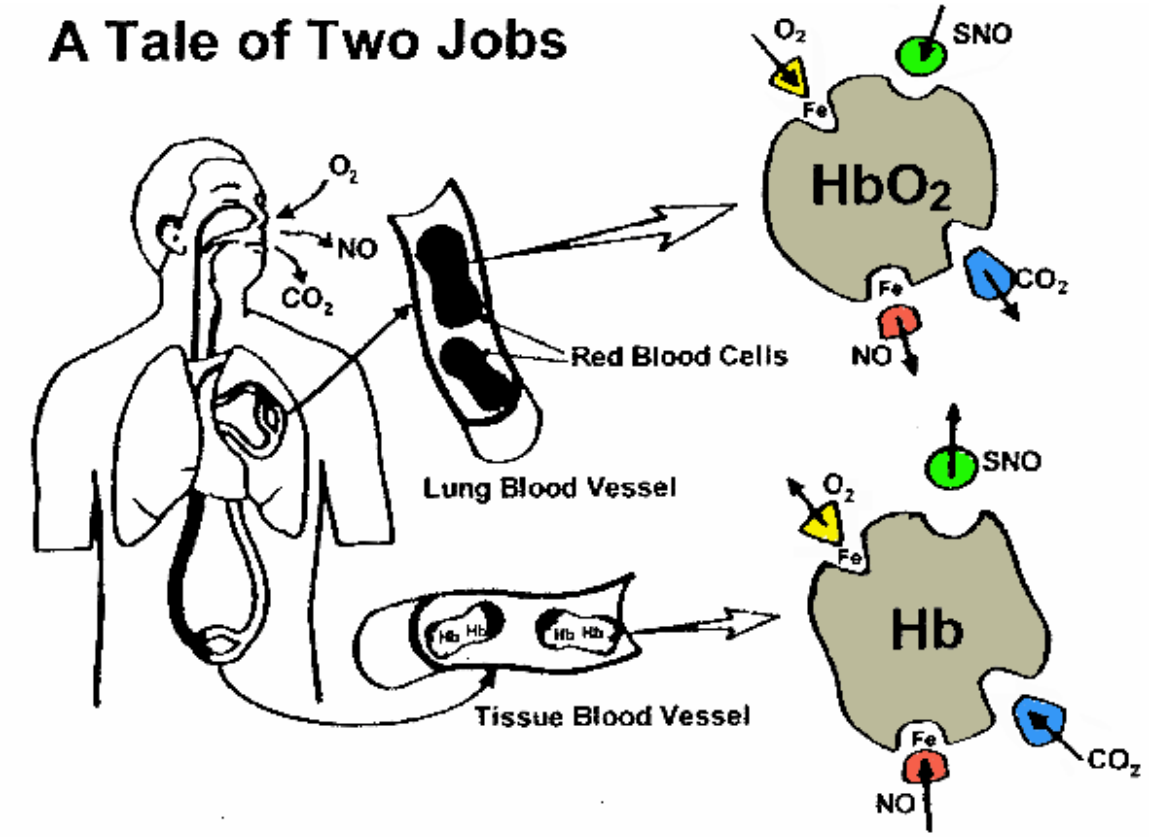


Figure 16. Le schéma (didactique et révisé) du transport de l'oxygène où interagissent O_2, CO_2 et NO. D'après Bonaventura et Lance (2001).

b. Autres fonctions

Des travaux sur des cellules hépatiques humaines par une approche protéomique, ont permis de séquencer une hémoprotéine qui a une activité enzymatique de type peroxydase et qui présente 40% d'homologie avec la myoglobine. De sérieux indices suggèrent que cette protéine pourrait aussi être un transporteur intracellulaire d'oxygène (Kawada et al. 2001).
L'adaptation à des apnées longues (jusqu'à une heure) de certains crocodiles est expliquée par un relarguage plus lent de l'oxygène. Cette capacité est expliquée par la présence au sein de la globine de quelques acides aminés clés qui confèrent un effet allostérique particulier. Des hémoglobines chimériques humaine-crocodile ont mimé *in vitro* cet effet allostérique, montrant que la substitution de quelques acides aminés permet de modifier l'affinité de l'hémoglobine pour l'oxygène. Ainsi, une modification mineure peut permettre l'émergence d'une nouvelle fonction qui confère un avantage sélectif et permet une adaptation à un environnement différent (Komiyama et al. 1995).

L'hémoglobine de *Salmo irideus* (la truite) possède une d'activité peroxydasique (Fedeli et al. 2001) sans que le mécanisme moléculaire soit identifié. Il reposerait davantage sur la configuration hétéropolymérique des hémoglobines et leur dissociation que sur une propriété portée intrinsèquement par les sous-unités. En effet, l'activité peroxydasique n'a pas été retrouvée à partir des chaînes de globines seules.

L'identification récente de nouvelles hémoglobines humaines, différentes des hémoglobines circulantes dans les globules rouges et de la myoglobine (cytoplasmique), est venue également révolutionner nos conceptions concernant les pigments respiratoires humains.
Deux nouveaux types de globines humaines ont été ainsi récemment décrits : les **neuroglobines** et les **cytoglobines** (ou histoglobines).

c. les neuroglobines

Les **neuroglobines** ont été identifiées *in silico* à partir des bases de données sur l'ADN humain et murin, puis séquencées et utilisées comme sonde pour de l'hybridation *in situ* pour chercher un éventuel patron d'expression. Leur expression dans le cerveau humain et les cellules neurales de différentes régions du cerveau de souris leur ont valu le nom de neuroglobines (Burmester et al. 2000). L'arbre phylogénétique incluant des myoglobines, des hémoglobines et ces neuroglobines, montre une association avec la myoglobine (probablement une neuroglobine) exprimée dans des tissus nerveux du polychète *Aphrodite aculeata* (Moens et Dewilde 2000 ; fig. 17). Des travaux précédents avaient également montré l'existence d'hémoglobines dans des tissus nerveux de divers autres mollusques, annélides et nématodes. La découverte de ces neuroglobines humaines et murines et leur présence chez des invertébrés suggèrent qu'elles ont probablement évolué avec le développement du système nerveux, et qu'il ne serait pas surprenant de les retrouver chez d'autres métazoaires. Ces neuroglobines ont été récemment retrouvées également chez des poissons téléostéens (Awenius et al. 2001) et le rat (Zhang et al. 2002).
Les neuroglobines auraient un effet neuroprotecteur contre l'hypoxie et ischémie (arrêt de l'irrigation) (Sun et al. 2001). Elles représentent le premier exemple d'hémoglobines "hexacoordinées" de vertébrés (Trent et Hargrove 2002), semblables aux hémoglobines "hexacoordinées" des bactéries. Ainsi, la présence d'hémoglobines hexacoordinées, chez les vertébrés, les plantes et les bactéries suggère une fonction plus générale de ces hémoglobines.

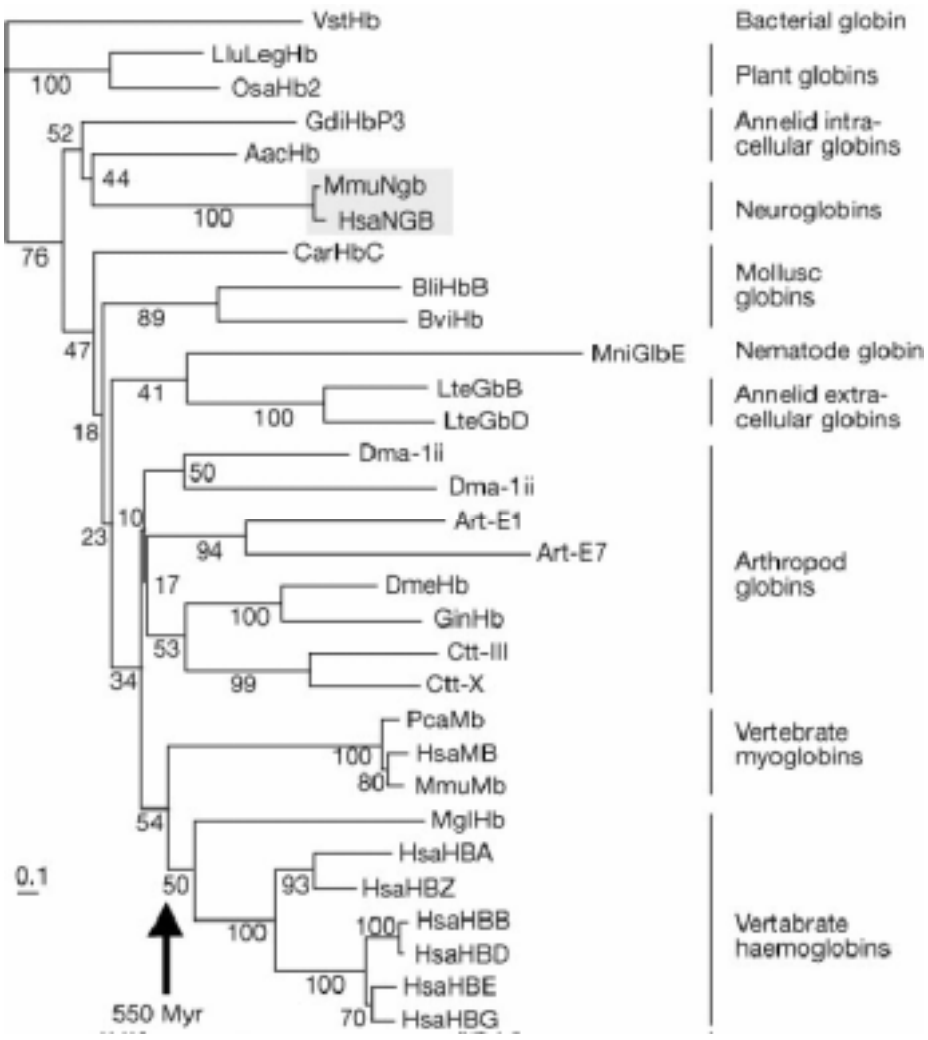


Figure 17. Phylogénie moléculaire de globines (Neighbour-joining). Légendes des taxons en annexe 1. D'après Burmester et al (2000)

d. Les cytoglobines

Les **cytoglobines** (Burmester et al. 2002) ou **histoglobines** (Trent et Hargrove 2002) sont des globines ubiquitaires exprimées dans tous les tissus de vertébrés (humains, murins et de poissons). L'étude phylogénétique montre qu'elles sont proches mais différentes des myoglobines et montre ainsi que l'évolution des globines est plus complexe qu'initialement envisagée (fig. 18).

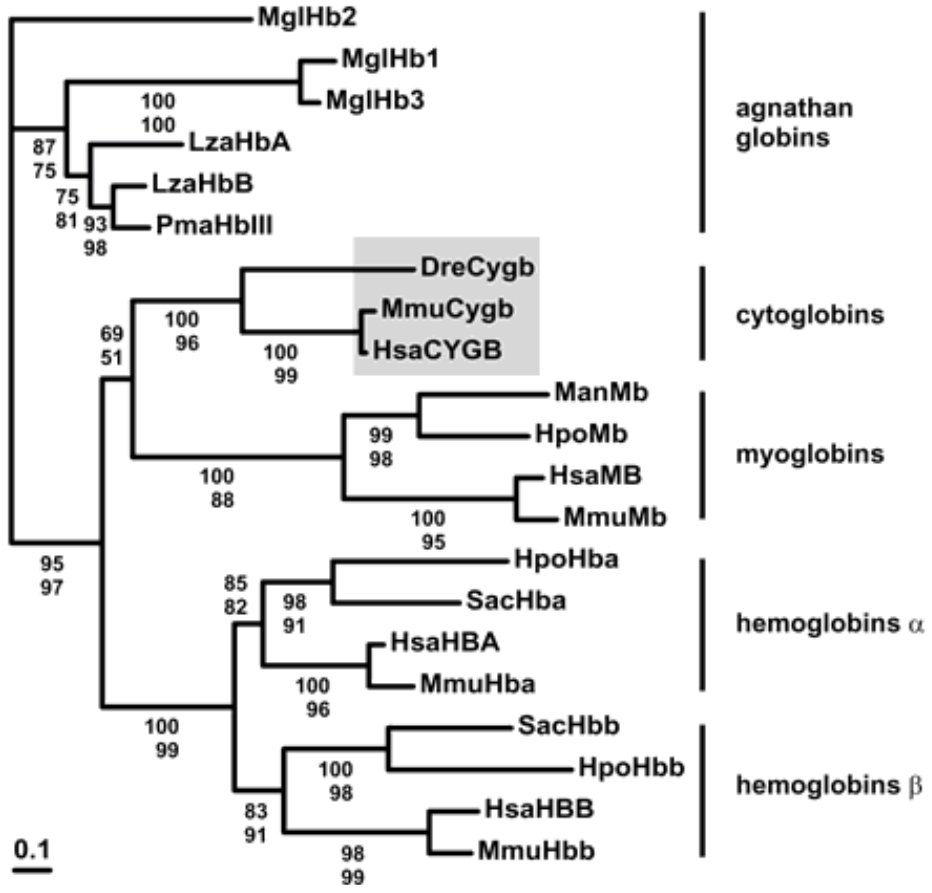


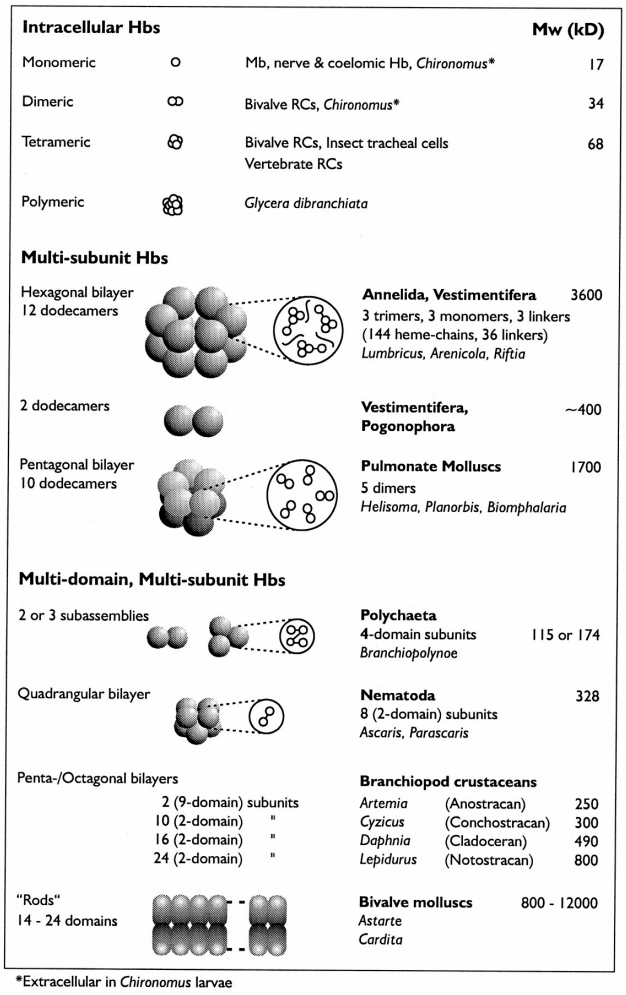
Figure 18. Phylogénie moléculaire de globines de vertébrés avec les valeurs de bootstraps (au-dessus) et de quartets (au-dessous). Légendes des taxons en annexe 1. D'après (Burmester et al. 2002)

Ces hémoglobines présentent un état d'hexacoordination, mais possèdent cependant des caractéristiques de fixation du ligand différentes des neuroglobines. Ce deuxième exemple d'hémoglobines "hexacoordinées" chez les vertébrés renforce l'idée d'une fonction répandue de ce type d'hémoglobine.

Avant la découverte des neuroglobines et des cytoglobines, une souris de laboratoire sans myoglobine a été générée par la technique du "gene knock-out". Cette mutante vit comme une souris sauvage qui aurait exprimé normalement la myoglobine (Garry et al. 1998). Ces travaux ont suscité des interrogations à propos du transport de l'oxygène au niveau cellulaire et suggéraient la présence de molécules alternatives, exprimées au niveau cellulaire à des concentrations très faibles, pouvant se substituer aux myoglobines.

6. Les hémoglobines d'invertébrés

Les hémoglobines d'invertébrés présentent une plus grande diversité de structures primaires et quaternaires que les hémoglobines de vertébrés. Cette variabilité reflète sans doute une spécialisation et une adaptation à un plus grand nombre de conditions et d'environnements. Dans l'encadré ci-contre, on peut voir les différentes structures quaternaires, du monomère au polymère de très haut poids moléculaire, en passant par les hémoglobines multi domaines, qu'elles soient intra ou extra cellulaires. Elles correspondent à autant d'adaptations à différents milieux (Weber et Vinogradov 2001).



Cette hétérogénéité de structure explique aussi en partie pourquoi ces hémoglobines ont été moins étudiées d'un point de vue fonctionnel et moléculaire que les hémoglobines de vertébrés. Le rôle de fixation et de transport de l'oxygène par les hémoglobines est largement reconnu et établi chez un grand nombre d'invertébrés tels que les nématodes, les plathelminthes, les annélides, les némertes, les mollusques (fig. 19).

Cependant d'autres fonctions et activités plus marginales ont été également répertoriées. On peut entre autre citer, les activités peroxidasiques chez des annélides polychètes, superoxide dismutase chez des annélides oligochètes ou des fonctions comparables à celle de la vitellogenine, en tant que protéine de stockage des œufs chez des diptères et même de flottabilité chez des hémiptères (Weber et Vinogradov 2001).

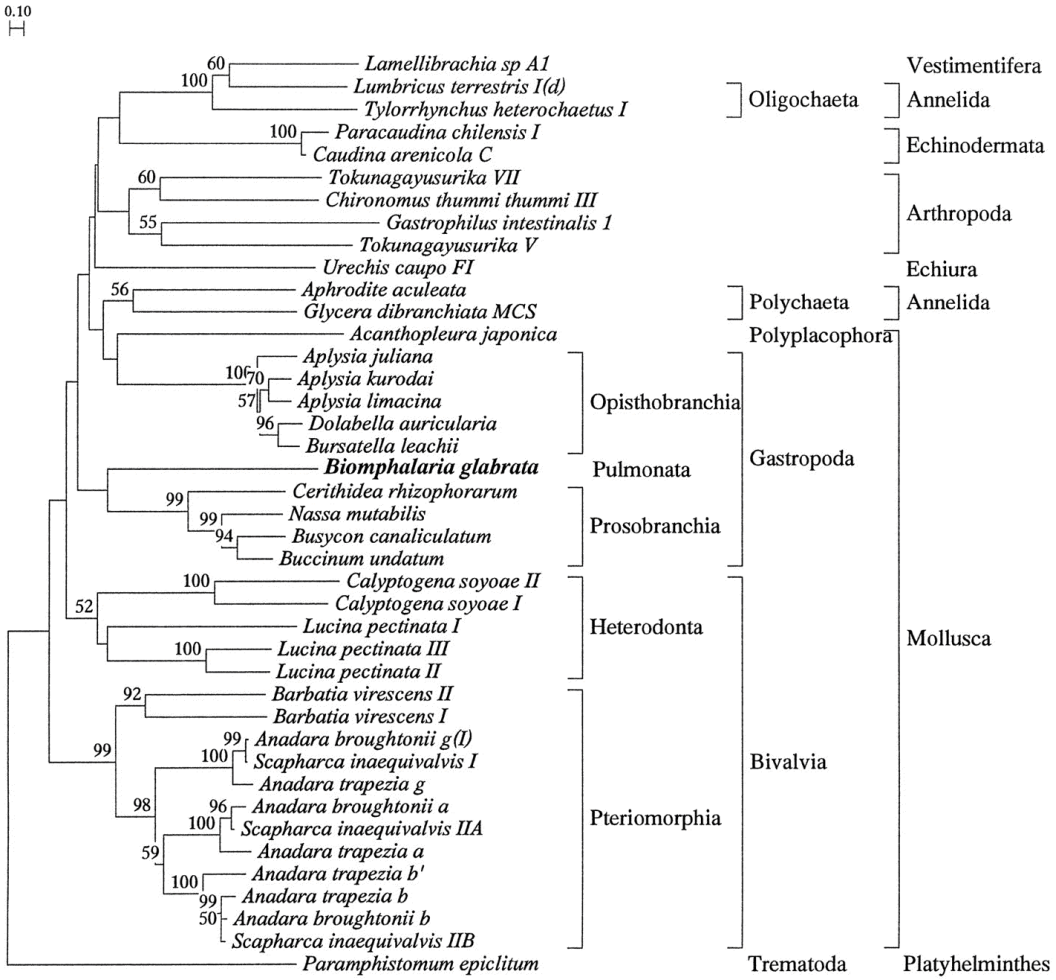


Figure 19. Phylogénie moléculaire de globines d'invertébrés. D'après (Dewilde et al. 1998b)

Deux exemples portant sur l'hémoglobine des nématodes montrent que cette protéine peut être à la fois une enzyme et une protéine fonctionnelle.

Ascaris lumbricoides, nématode parasite, possède une hémoglobine dont l'affinité pour l'oxygène est 25 000 fois plus forte que celle de l'hémoglobine humaine. Cette véritable chélation de l'oxygène permet le maintien de conditions hypoxiques vitales pour le nématode et qui illustre ici une fonction de détoxication de l'oxygène. D'abord soupçonnée d'intervenir dans la biosynthèse des stérols comme ferrihémoprotéine réductase NADPH-dépendante (Sherman et al. 1992), l'hémoglobine d'*Ascaris* est considérée aujourd'hui comme une déoxygénase utilisant le monoxyde d'azote (NO) pour maintenir un état hypoxique par oxidation de NO en nitrate, grâce à la présence simultanée de l'hème, d'un groupement thiol d'une cystéine de la globine et du NO (Minning et al. 1999).

Mermis nigrescens est un autre nématode parasite d'insectes orthoptères, dont la femelle gravide possède un ocelle pigmenté qui contient de l'hémoglobine. La phototaxie positive

observée chez ces femelles est attribuée à cette hémoglobine soluble, stable et qui absorbe fortement la lumière. Ces travaux suggèrent un recrutement des gènes codant une hémoglobine photoréceptrice lors de l'élaboration du système optique (Burr et al. 2000).

V. La fonction de fixation du sulfure d'hydrogène par les hémoglobines de mollusques et d'annélides

Le sulfure d'hydrogène est un composé toxique (Chapitre 2), présent dans les milieux dits "réduits" comme certaines zones de l'estran, les mangroves, les vasières. Le sulfure d'hydrogène a une origine biogénique car l'accumulation de matière organique, des conditions sévères d'hypoxie et la présence de bactéries sulfato-réductrices favorisent la production de fortes concentrations en sulfures. Dans d'autres milieux le sulfure d'hydrogène a une origine chimique et est intrinsèquement produit par le milieu, comme les sources hydrothermales et les suintements froids (chapitre 2) qui expulsent des fluides enrichis naturellement en ce composé. A l'inverse de l'hémoglobine des vertébrés dont l'hème réagit de manière covalente avec H_2S et forme la sulfohémoglobine non fonctionnelle, les invertébrés des milieux riches en sulfures (majoritairement des annélides et des mollusques) réagissent de différentes manières avec H_2S et ce sans altérer la fonction de transporteur d'oxygène.

1. Fixation et transport de H_2S par les hémoglobines cytoplasmiques (intracellulaires) de mollusques

Les trois hémoglobines des branchies qui abritent des bactéries sulfo-oxydantes symbiotiques du bivalve *Lucina pectinata* ont été les plus étudiées des mollusques des milieux riches en sulfures. Seule l'hémoglobine I fixe réversiblement H_2S (fig. 20) alors que les hémoglobines II et III restent oxygénées en présence de sulfure d'hydrogène. Le sulfure d'hydrogène est fixé de façon non covalente sur l'hème de l'hémoglobine I, dont le fer passe à l'état oxydé (ferrique) et diffuse ainsi vers les bactéries endosymbiotiques sulfo-oxydantes. La très faible vitesse de dissociation *in vitro* (environ 80 minutes) traduit la forte affinité de l'hémoglobine pour ce ligand et suggère qu'une simple dissociation spontanée n'est pas suffisante et nécessite d'autres acteurs moléculaires ou des conditions physiologiques particulières (Kraus et Wittenberg 1990).

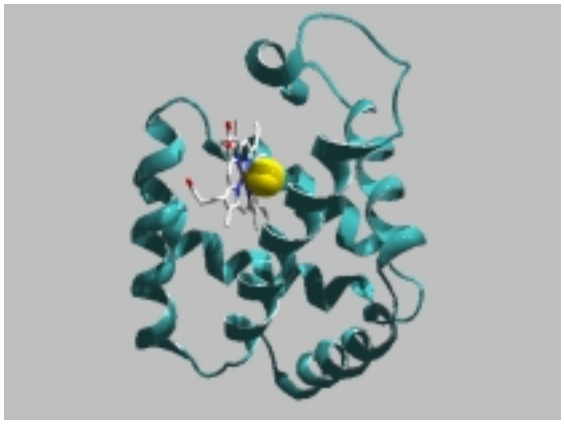
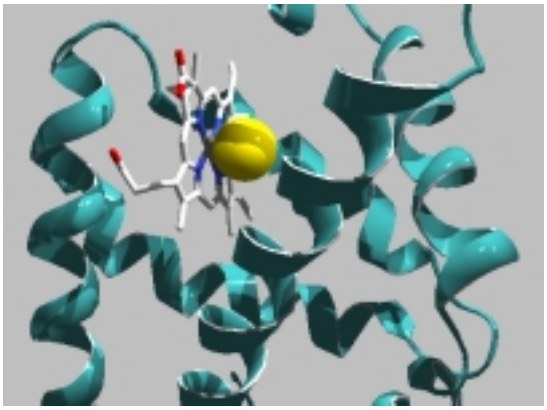
1.  2. 

Figure 20.1. Vue 3D de l'hémoglobine cytoplasmique I de *Lucina pectinata* et de H_2S (sphère jaune). 2. zoom sur la poche contenant l'hème. (n° accession dans la Protein Data Bank (PDB) 1MOH)

Chez *Solemya reidi*, la présence de H_2S diminue de 5 à 15 fois le taux de dissociation de l'oxygène de deux des trois hémoglobines. La fonction physiologique serait de maintenir un pool d'oxygène intracellulaire pendant les périodes d'intense exposition aux sulfures et/ou de faible disponibilité d'oxygène (Weber et Vinogradov 2001).

Pour le bivalve *Calyptogena magnifica* (famille des Vesicomidae) présent au niveau des sources hydrothermales profondes, l'hémoglobine ne fixe pas H_2S, mais un composé qui pourrait être une lipoprotéine contenant du zinc est probablement impliqué dans la fixation de H_2S (Zal et al. 2000).

2. Fixation et transport du sulfure d'hydrogène par les hémoglobines extracellulaires d'annélides des milieux riches en H_2S

On trouve trois types d'hémoglobines au sein des annélides : des hémoglobines non circulantes cytoplasmiques, des hémoglobines circulantes intracellulaires et extracellulaires dans le coelome et le réseau vasculaire. Les hémoglobines non circulantes cytoplasmiques sont les analogues des myoglobines des vertébrés et remplissent les mêmes fonctions de transfert de l'oxygène de l'hémoglobine vasculaire vers les mitochondries. Les hémoglobines intracellulaires circulantes sont restreintes aux annélides polychètes et ce dans les 6 familles suivantes Capitellidae, Ophelidae, Scalibregmatidae, Terebellidae, Alvinellidae et Glyceridae. Cependant les Terebellidae, Alvinellidae et Ophelidae possèdent aussi des hémoglobines extracellulaires. Les hémoglobines extracellulaires circulantes, d'abord intracellulaires car

synthétisées dans les cellules du tissu hématopoïétique puis secondairement exportées, sont largement répandues dans le phylum des annélides. Leur haut poids moléculaire lié à leur composition multimérique complexe peut atteindre 3 600 kDa (Weber et Vinogradov 2001). **On dit qu'elle sont en bicouche hexagonal** (**HBL-Hb : Hexagonal Bilayer Hemoglobin** ; fig 21).

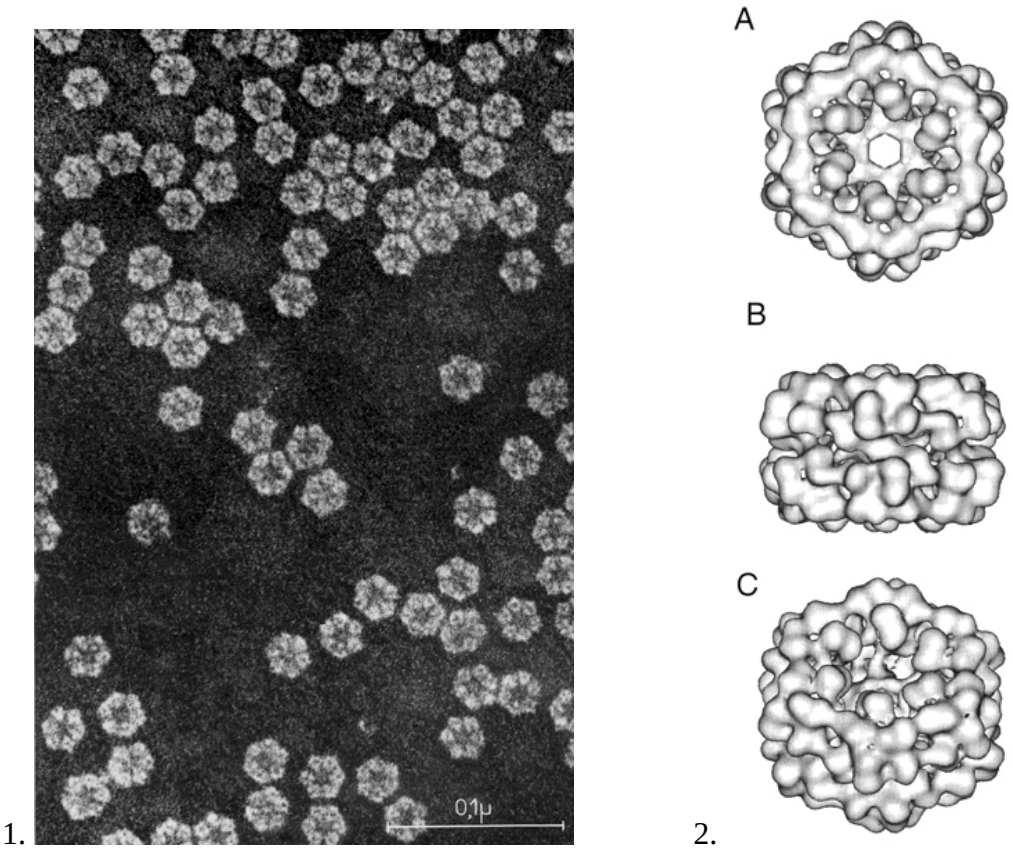


Figure 21. 1. Microscopie électronique des hémoglobines d'*Arenicola marina* (X 446 000) Tiré de (Breton-Gorius 1963). 2. Reconstruction tridimensionnelle à 22-Å de l'hémoglobine de *Lumbricus terrestris* (ver de terre) A : vue de dessus. B : vue latérale. C : vue à 45° de rotation autour de l'axe horizontal. D'après (Weber et Vinogradov 2001).

Chez les annélides, les premiers travaux portant sur les capacités de fixation réversible du sulfure d'hydrogène par les hémoglobines sont corrélés à la découverte et à l'exploration du milieu hydrothermal profond et aux espèces inféodés qui baignent dans un fluide hydrothermal riche en sulfures (Lonsdale 1977). La propriété de lier réversiblement H_2S semble être l'apanage des annélides symbiotiques et non symbiotiques des milieux riches en sulfures, car les annélides des milieux oxydés (sans sulfures) n'ont pas cette capacité et ne résistent pas au sulfure d'hydrogène.

La présence de bactéries sulfo-oxydantes endosymbiotiques est un moyen indirect de détoxication de l'organisme : chez ces annélides, c'est le moyen de transporter H_2S et donc de l'immobiliser pendant le transport qui représente une adaptation directe à la toxicité de H_2S. Pour les annélides sans bactéries symbiotiques vivant en présence de sulfures et dont l'hémoglobine peut fixer le sulfure d'hydrogène, c'est l'immobilisation temporaire ou la chélation qui représente un moyen de résistance.
La vie en présence des sulfures, la résistance à ce composé, les mécanismes de fixation et de transport par les hémoglobines des annélides sont développés dans le chapitre 2.

VI. Conclusion sur la diversité des hémoglobines et leurs fonctions multiples : scénario évolutif

A la lumière de ces exemples et de cette variété de fonctions certains auteurs ont ironiquement écrit " l'hémoglobine fixe également l'oxygène !" (Bonaventura et Lance 2001). Retenons cependant que la fonction principale la plus observée des hémoglobines à l'heure actuelle est celle de transporteur d'oxygène.
Ainsi, avant d'être une protéine liant réversiblement l'oxygène, l'hémoglobine ancestrale a peut-être joué le rôle de chélateur d'oxygène, afin de protéger les cellules de l'effet oxydant délétère (détoxication) (Moens et al. 1996). Cette hypothèse postule que le caractère "affinité pour l'oxygène" aurait été sélectionné dans un premier temps. Les concentrations en oxygène augmentant et le métabolisme aérobie émergeant, des espèces chimiques telles le monoxyde d'azote et le sulfure d'hydrogène devenues toxiques et ont été probablement prises en charge à leur tour. L'hypothèse retenue est qu'avant même d'acquérir la fonction de réversibilité de la liaison de l'oxygène, l'hémoglobine était aussi un piège à NO et H_2S et à d'autres molécules et radicaux (Bonaventura et Lance 2001).
Une autre hypothèse, plus centrée sur le monoxyde d'azote et basée sur les propriétés des flavohémoglobines, avance que l'hémoglobine était initialement une enzyme au fort pouvoir redox pour se débarrasser de NO et de son stress oxydatif : cette enzyme aurait perdu son domaine réductase mais incorporé des résidus réactifs (cystéines) pour continuer à gérer le monoxyde d'azote (Durner et al. 1999).
Le cours de l'évolution a conduit aujourd'hui à des hémoglobines ayant de multiples ligands qui sont autant d'indices pour découvrir l'histoire (ou les histoires) de ces pigments respiratoires les plus répandus dans le monde vivant.

Chapitre 2 :

Les adaptations aux milieux riches en sulfure d'hydrogène

Une illustration : les hémoglobines de *Riftia pachyptila*

Chapitre 2. Les adaptations aux milieux riches en sulfure d'hydrogène. Une illustration : le cas des hémoglobines de *Riftia pachyptila*

I. La vie en présence de sulfure d'hydrogène : généralités

Le sulfure d'hydrogène (H_2S) est un composé toxique pour le métabolisme aérobie. Il inhibe les métalloprotéines comme la cytochrome-c oxidase et les hémoglobines (Nicholls 1975). Sa fixation irréversible sur l'hème chez l'hémoglobine humaine forme la sulfohémoglobine. L'auto-oxydation de H_2S génère des produits non ou moins toxiques tels le soufre élémentaire (S), le sulfite (SO_3^{2-}), le thiosulfate ($S_2O_3^{2-}$) et le sulfate (SO_4^{2-}). Cette réaction, pH dépendant, est catalysée par des ions tels Ni^{2+}, Mn^{2+}, Cu^{2+}, et Fe^{2+}. Malgré cette réactivité, la demi-vie de H_2S dans de l'eau de mer saturée en oxygène est de l'ordre de quelques heures (Grieshaber and Volkel 1998). Le sulfure d'hydrogène d'origine biogénique (les vases de l'estran, les mangroves) est issu de l'activité de bactérie sulfato-réductrices anaérobies et hétérotrophes qui minéralisent le carbone organique (chapitre 1). Le degré d'oxydation de H_2S varie en fonction du produit d'oxydation final. Deux moles de O_2 sont consommées pour oxyder 1 mole de H_2S en SO_4^{2-} alors que 1,5 moles de O_2 sont consommées pour oxyder 2 moles de H_2S en 1 mole de $S_2O_3^{2-}$. La répartition de la méiofaune de l'estran en fonction de la profondeur du substrat est en partie dictée par ces rendements d'oxydation de H_2S et par la limitation en oxygène. Dans ces conditions souvent hypoxiques et parfois anoxique, il est important de dépenser un minimum de molécule d'O_2 pour la détoxication de H_2S (Grieshaber and Volkel 1998).
Les sources hydrothermales et les suintements froids des zones d'accrétions et de subduction océaniques sont des environnements où H_2S abonde également (chapitre 2. II).

II. Stratégies de détoxication de H_2S

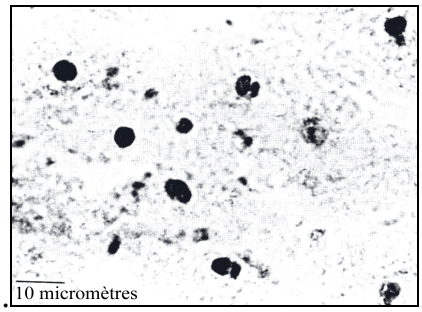
Dans les sédiments marins, les bactéries sulfato-réductrices libèrent H_2S qui se retrouve également dans la colonne d'eau. Ainsi les animaux vivant à ces endroits de l'estran rencontrent à la fois des conditions variables d'hypoxie (modérée à sévère) couplées à de fortes concentrations en H_2S. Celles-ci varient en fonction des saisons avec des points culminants en été. Des annélides polychètes comme *Arenicola marina* peuvent être exposées à des concentrations d'H_2S de 330 µM (Grieshaber and Volkel 1998).
Plusieurs mécanismes de résistance et de tolérance aux effets délétères de H_2S ont été proposés pour expliquer l'adaptation aux milieux riches en sulfures.

Chez les organismes dépourvus de bactéries symbiotiques sulfo-oxydantes, l'imperméabilité aux sulfures, la présence de cytochrome c oxydases insensibles à H_2S, la fixation réversible à des composés du sang (autres que l'hémoglobine) sont des stratégies de détoxication ou de transport de H_2S qui ont été décrites chez les annélides et les mollusques des milieux réduits (riches en sulfures).
Les espèces vivant dans ces milieux possèdent en plus de cette panoplie de résistance au sulfure d'hydrogène, des moyens biologiques d'oxydation de H_2S.

1. Les corps d'inclusion de sulfures

Dans ses études portant sur l'identification du tissu hématopoïétique de l'annélide polychète *Arenicola Marina*, Breton-Gorius (1963) relate la présence d'hématine (hème libre avec un hydroxyle lié au fer) dans des cellules de la paroi du corps, du tissu chloragogène et du corps cardiaque sans lui attribuer de rôle particulier. La même année Patel et Spencer (1963) montre que l'hémoglobine d'Arénicole ne forme pas de sulfohémoglobine mais que l'hématine et un "pigment brun" agiraient comme des catalyseurs de l'oxydation des sulfures. Dans l'épiderme de l'échiurien *Urechis caupo* (phylogénétiquement très proche des annélides (McHugh 1997)), en plus de bactéries symbiotiques sulfo-oxydantes, on trouve de petits organites contenant des sulfures, les SOBs (Sulfo Oxidizing Bodies) qui présentent des capacités d'oxydation des sulfures (Arp et al. 1995). Dans le même article, le fluide de la cavité coelomique chez *Urechis caupo* contient entre autre de l'hémoglobine et de l'hématine qui toutes les deux présentent le pouvoir d'oxyder les sulfures en thiosulfate (Arp et al. 1995). Une telle situation a été décrite également chez *Solemya reidi*, un mollusque bivalve (famille des Solemyidae) qui possède aussi des bactéries sulfo-oxydantes symbiotiques (Powell and Somero 1985). Des entités analogues (inclusions de sulfures) ont été trouvées également au niveau de l'épiderme du nématode marin *Oncholaimus campylocercoides*, qui lui permettrait de stocker et de diminuer les concentrations en H_2S (Thiermann et al. 2000 ; fig. 22). Chez l'annélide oligochète de marécages tropicaux *Alma emini*, des SMBs (sulfide metabolizing bodies analogues des SOBs) et des bactéries sulfo-oxydantes co-existent également. Chez la lucine *Lucina pectinata* des mangroves, deux types d'inclusions ont été identifiés : l'un contient de l'hématine comme (les SOBs de *Solemya reidi* (fig. 22) et *Urechis caupo*) et l'autre pourrait contenir des enzymes lysosomiaux. Chez le bivalve *Codakia orbicularis* (famille des Lucinidae) qui vit dans un milieu oxydé, de telles structures n'ont pas été trouvées

ce qui suggère que la présence de celles-ci représentent une adaptation aux milieux réduits riches en sulfures (Liberge et al. 2001).

1. 

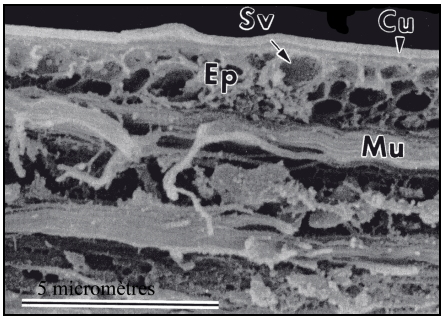
2. 

Figure 22. 1. SOBs colorés en noir au niveau des branchies chez *Solemya reidi* d' apès Powell and Somero (1985)

2. Inclusions de H_2S chez le nématode *Oncholaimus campylocercoides* Cu : cuticule, Mu : cellules musculaires , SV : vacuoles de sulfures, Ep : épithélium. D'après Thiermann et al (2000)

Ainsi le pouvoir oxydant de l'hème libre sous forme d'hématine dans le sang et/ou dans la présence de corps d'inclusion représente un moyen pour contrecarrer l'effet délétère de H_2S dans les tissus de ces invertébrés. Chez *Arenicola marina* le thiosulfate qui est le principal produit d'oxydation de H_2S, non toxique, est excrété ensuite par diffusion passive (Hauschild et al. 1999).

2. Oxydation biologique de H_2S

En conditions réduites, les fortes teneurs en H_2S sont corrélées avec de faibles concentrations en oxygène. Ainsi, quand la pression partielle en oxygène est faible ou lorsque les concentrations en sulfures deviennent inhibitrices, la mitochondrie réduit le fumarate en succinate (fig.23). L'accumulation du succinate traduit un basculement vers un métabolisme anaérobie sulfure-dépendant. Powell and Somero (1986) ont démontré que l'oxydation des sulfures en thiosulfate par des mitochondries tissulaires du bivalve *Solemya reidi* vivant en présence de H_2S dans des sédiments marins profonds, était aussi et à la fois un moyen de lutter contre la toxicité de H_2S et de synthétiser de l'ATP. Cette même caractéristique a été observée chez la moule *Geukensia demissa* (Parrino et al. 2000). Cette dernière vit dans des zones intertidales marécageuse anoxiques et riches en H_2S à proximité des racines de la

monocotylédone *Spartina alterniflora* qui sont capables elles aussi d'oxyder les sulfures (Lee 1999).

Depuis, cette propriété d'oxydation des sulfures couplée à la phosphorilation oxydative a été observée chez des poissons, des annélides polychètes et des priapuliens. La figure 23 montre que les électrons de H_2S peuvent entrer, selon les espèces étudiées, dans la chaîne membranaire à différents niveaux : au niveau du pool des ubiquinones, du complexe 3 ou du cytochrome c.

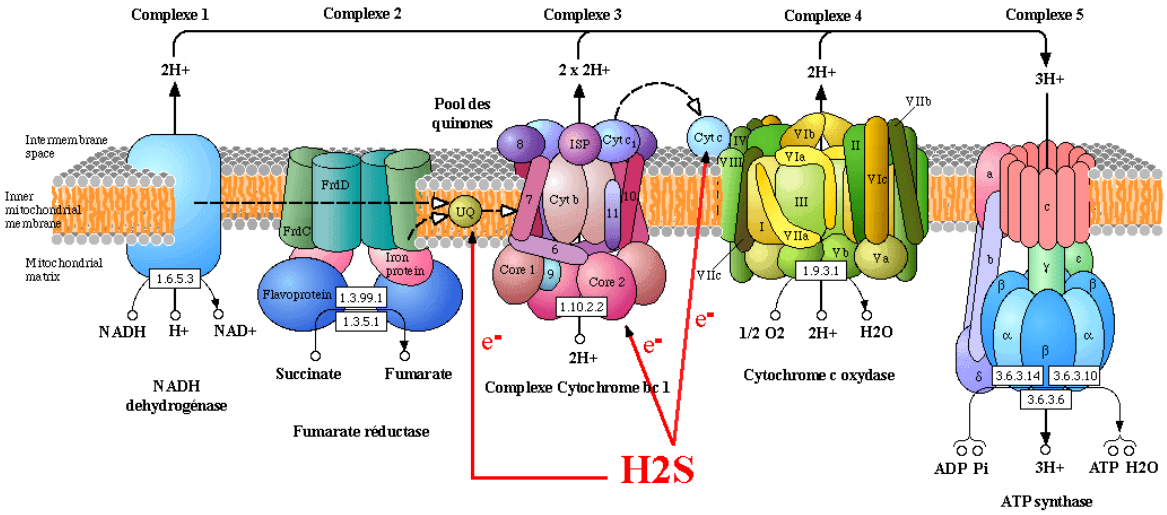


Figure 23. Phosphorylation oxydative et oxydation de H_2S au niveau de la chaine membranaire d'électron

Il faut retenir que les fortes concentrations en H_2S en et les faibles concentrations en O_2 alternent avec les faibles concentrations H_2S et les fortes concentration en O_2. L'ATP produit par le métabolisme anaérobie (concomitante absence d'oxygène et présence de H_2S) compenserait l'inhibition transitoire du cytochrome c.

III. L'environnement hydrothermal

L'hydrothermalisme profond est une manifestation géothermique spectaculaire localisée à l'axe des dorsales océaniques (fig. 24), lieu d'accrétion (déchirement de la croûte océanique accompagnée de remontée de magma) résultant du mouvement des plaques tectoniques.

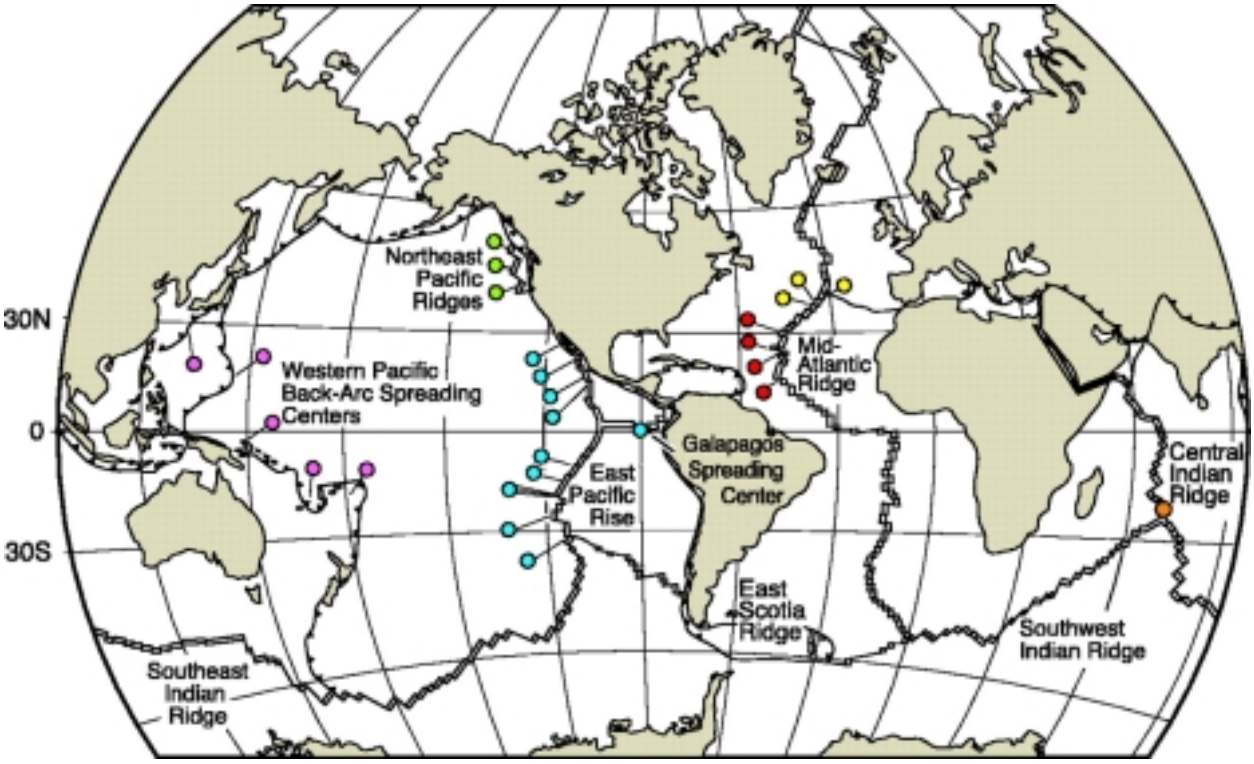


Figure 24. Carte des principales sources hydrothermales explorées. Rose : Pacifique Ouest. Vert : Pacifique Nord Est. Bleu : Pacifique Est. Jaune : Azores. Rouge : dorsale médio-atlantique. Orange : océan indien. Tiré de Van Dover et al. (2002)

Au niveau de la vallée axiale (graben) du rift, l'eau de mer (2 à 3°C) s'infiltre dans le plancher basaltique fissuré et se rapproche de la chambre magmatique (fig.25). Le gradient de température croissant et qui atteint 800 à 900°C permet des réactions chimiques spécifiques. L'eau s'enrichit ainsi en divers composés, et rejaillit sous la forme d'un fluide chaud (350°C), anoxique, acide (2.0<pH<5.9), riche en CO_2, riche en éléments réduits tels que H_2S (jusqu'à 4 mM), SiO2, CH_4 (jusqu'à 50 nM), NH_4^+, H^+ et en éléments métalliques tels que Mn^{2+} (jusqu'à 1 mM), Fe^{2+} (jusqu'à 200 µM), Li^+, Cd^{2+}, Cu^{2+} et Zn^{2+} (Edmond et al. 1982).

Le fluide hydrothermal peut rejaillir à l'interface eau/croûte océanique sans dilution préalable et précipiter sous la forme de sulfures poly-métalliques et de sulfates de calcium pour former les cheminées hydrothermales (150 à 400 °C) (fig.25), ou se mélanger en sub-surface dans des réseaux de fissures pour s'échapper lentement à des températures plus faibles (mélange diffus, 5-50°C). C'est dans cet environnement où les gradients thermochimiques sont accentués, vers 2500 m de profondeur, où la lumière est absente (pas de photosynthèse possible), qu'ont été découvertes en 1977 (Corliss and Ballard 1977) des communautés animales présentant un taux d'endémisme élevé, de fortes biomasses et une faible diversité spécifique.

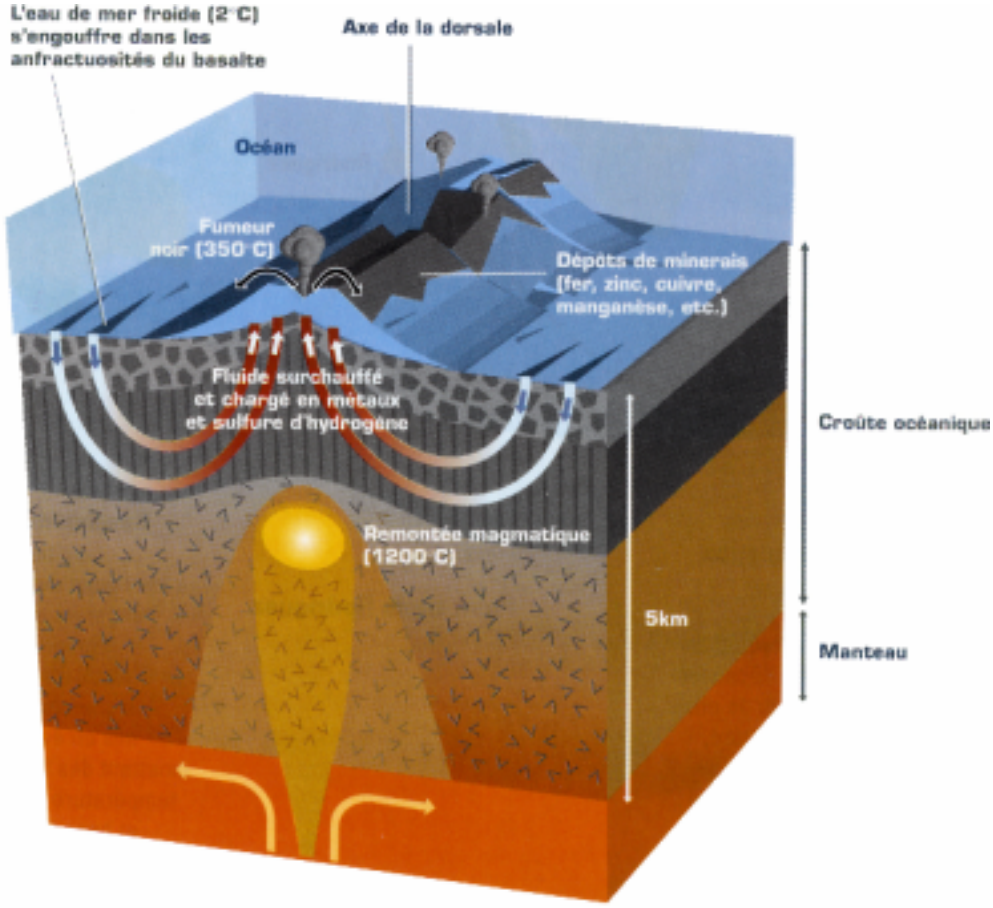


Figure 25. Mécanismes de formation des sources hydrothermales sur les dorsales océaniques.
D'aprés "Plein phare sur les abysses" dans Le journal du CNRS, n°50, juin 2002, Schéma de Vincent Perrotte.

Depuis 1977, 400 espèces des sources hydrothermales, dont une grande majorité d'invertébrés, ont été décrites (Van Dover et al. 2002). L'environnement hydrothermal est un milieu privilégié pour étudier l'effet de la sélection dans l'évolution des invertébrés marins. L'activité hydrothermale est caractérisée par une grande variabilité spatiale et temporelle qui s'illustre par la combinaison d'évènements non périodique (l'activité tectonique au niveau de la dorsale et la convection des fluides chauds à l'intérieur de la croûte océanique poreuse) et d'évènement périodiques (effet des phénomènes de marée en profondeur). Ces espèces sont exposées à des conditions physico-chimiques qualifiées (anthropomorphiquement parlant) d'extrêmes et d'hostiles à la vie car rappelant a bien des égards les pollutions d'origine anthropiques (métaux lourds, H_2S, radioactivité), incompatibles à la survie de la grande majorité des espèces vivant à la surface de la Terre, dont l'espèce humaine.

IV. Sources hydrothermales et suintements froids : réseaux trophiques

Le réseau trophique de l'écosystème des sources hydrothermales est basé sur la chimiosynthèse bactérienne : de nombreuses espèces, 75 à 90% de la biomasse totale des sources hydrothermales, vivent en symbiose avec des bactéries chimioautotrophes (Fustec et al. 1988 ; Cavanaugh et al. 1981). Celles-ci sont capables d'oxyder le sulfure d'hydrogène ou le thiosulfate (Vetter et al. 1987) et de se servir directement de l'énergie d'oxydation pour synthétiser de la matière organique via le cycle de Benson-Calvin et grâce à l'enzyme de fixation du carbone chez les organismes autotrophes, la RuBP carboxylase (Felbeck 1981 ; Felbeck et al. 1981). Ces métabolites sont ensuite transférés à l'hôte par pinocytose ou lyse bactérienne .

Pour assurer leurs besoins nutritifs, ces espèces doivent nécessairement se trouver dans la zone de mélange eau de mer/fluide hydrothermal et par conséquent être adaptées à une alternance constante de conditions aéro-anaérobie, de température, de CO_2, d'H_2S. Les études de ce milieu unique, instable et fragmenté ont mis en évidence des adaptations surprenantes des organismes vivant à proximité de ces sources.

Ainsi, l'écologie des sources hydrothermales est paradoxalement dépendante du sulfure d'hydrogène car sa biodisponibilité (toxique mais énergétique) constante façonne la répartition des espèces. Bien que des stratégies de détoxication vis-à-vis de H_2S soient requises (Zierenberg et al. 2000), il a été clairement montré que la seule abondance de H_2S ne suffit pas à expliquer la répartition spatiale des espèces. En effet, dans les micro-habitats où la température est supérieure à 30 °C, la formation de groupement chimiques solubles fer-sulfure, réduit fortement la disponibilité des formes libres H_2S et HS^- selon l'équilibre $Fe^{2+} + H_2S \leftrightarrow FeS + 2H^+$. Cette spéciation chimique contrôle aussi les aires de répartitions des espèces en fonction du degré de tolérance à H_2S (Luther et al. 2001).

Les suintements froids sont des processus géologiques actifs résultant d'une migration de fluides à travers la couverture sédimentaire par effet de compaction (subduction des plaques ou poids des sédiments eux-mêmes). Ces fluides retrouvés au niveau des marges continentales sont issus de la dégradation de la matière organique dissoute dans l'eau interstitielle des sédiments et sont constitués d'hydrocarbures liquides (huile) ou gazeux (méthane) (Van Dover et al. 2002). Les fluides ont une origine biogénique et non minérale comme pour les sources hydrothermales, et sont également riches en H_2S. Les concentrations en sulfures sont généralement plus faibles mais plus stables temporellement et spatialement (McMullin et al. 2000). La température des émissions est faible, varie très peu au cours du temps et atteint en

moyenne 8°C. L'écosystème est similaire à celui des sources hydrothermales et repose aussi sur la chimiosynthèse bactérienne et principalement sur l'oxydation de H_2S et du méthane (CH_4).

V. ***Riftia pachyptila*** **: Vestimentifère, phylum des annélides**

Riftia pachyptila Jones (Jones 1981) est un ver tubicole sessile, dont le tube (exosquelette chitino-protéique) peut atteindre deux mètres de long (fig. 26).

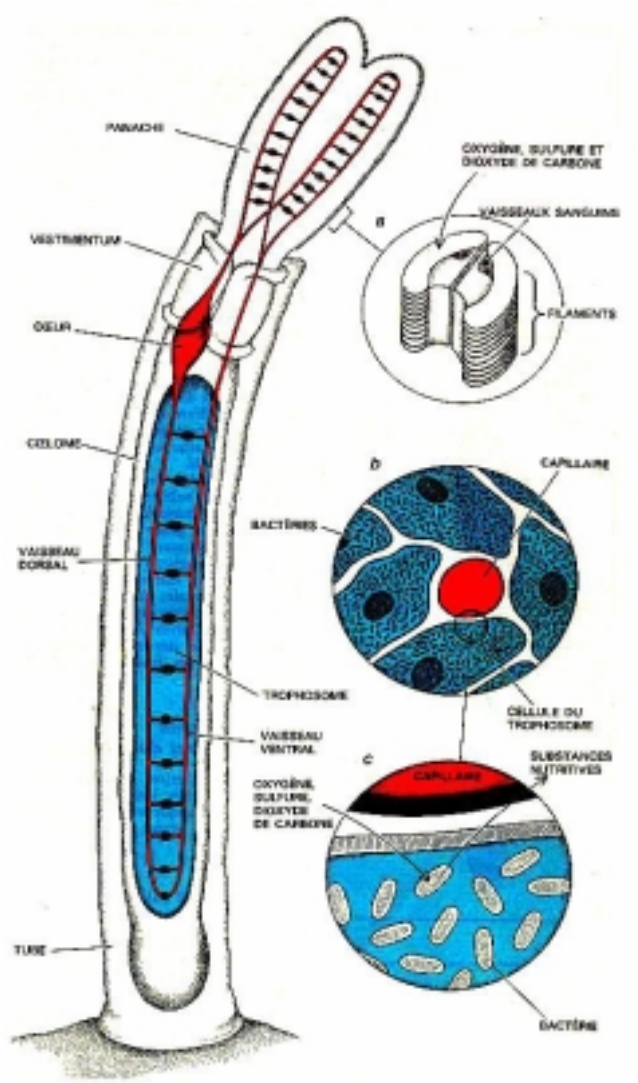




Figure 26. *Riftia pachyptila* : schéma (d'après Pour la Science, Juillet 1987, p 57-63) : Ce ver tubicole capte au niveau de sa branchie O_2, CO_2 et H_2S qui sont véhiculés par le système vasculaire jusqu'aux bactéries symbiotiques hébergées dans le trophosome. Photos *in situ*, vers 2500 mètres sur la dorsale Est-Pacifique.

Il est inféodé aux sources hydrothermales de la dorsale du pacifique Est et présente une biomasse supérieure à la dizaine de kilogrammes par mètre carré (Fustec et al. 1987). Il ne possède pas de tube digestif à l'état adulte, mais un organe richement vascularisé, le trophosome, qui abrite des bactéries chimioautotrophes sulfo-oxydantes (Felbeck et al. 1985). Celles-ci appartiennent au groupe des γ protéobactéries et sont endosymbiotiques car intracellulaires, localisées dans des cellules spécialisées, les bactériocytes (Cavanaugh 1983). La transmission des bactéries est horizontale. Bien que le développement larvaire n'ait été observé que jusqu'au stade trochophore (Marsh et al. 2001), le scénario d'acquisition des bactéries se base sur les observations réalisées sur une espèce proche de Riftia, *Ridgea piscesae*, dont les étapes du développement sont bien décrites (Southward 1988 ; Young et al. 1996). Le recrutement a lieu au cours de la phase larvaire, quand la larve du vestimentifère possède encore une bouche, un tube digestif et un anus, pendant les 20 premiers jours suivant la fécondation. Les bactéries sont approvisionnées en oxygène, en dioxyde de carbone et en sulfure d'hydrogène dissous dans la colonne d'eau. Ces molécules sont puisées à partir du panache branchial de l'animal et parcourent le système vasculaire liées à l'hémoglobine (Arp et al. 1987). Selon la classification du National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) *Riftia pachyptila* appartient à la classe des vestimentifères du phylum des Pogonophores, distinct du phylum des annélides. Les pogonophores (hormis les vestimentifères) sont des vers tubicoles symbiotiques fins, de 2 à 3 millimètres, que l'on ne trouve pas à proximité des sources hydrothermales comme les vestimentifères, mais dans des zones plus ou moins profondes, des estuaires, des fjords, également au niveau des suintements froids, sans aire géographique vraiment précise, mais dans des sédiments riches en sulfure d'hydrogène. De plus en plus de données phylogénétique, anatomiques, ontogéniques et moléculaires s'accordent maintenant à regrouper les pogonophores et les vestimentifères dans le phylum des Annélides. Les pogonophores et les vestimentifères peuvent être donc considérés comme des annélides hautement évolués dans la mesure où ils réalisent des symbioses dont la mise en place se traduit par la perte ou la modification de certains caractères propres aux annélides. En terme d'anatomie comparée, la segmentation typiquement annélidienne est retrouvée sur une petite région basale des vestimentifères et des pogonophores que l'on appelle l'opistosome (partie la plus proche du point d'ancrage du ver sur le substrat rocheux). De plus, sur chacun de ces segments on trouve des soies (fig. 27), comme chez les annélides polychètes (Jones 1985).

En terme d'ontogénie, le tractus digestif est présent chez les larves des vestimentifères et disparaît après l'établissement de la symbiose (Young et al. 1996).

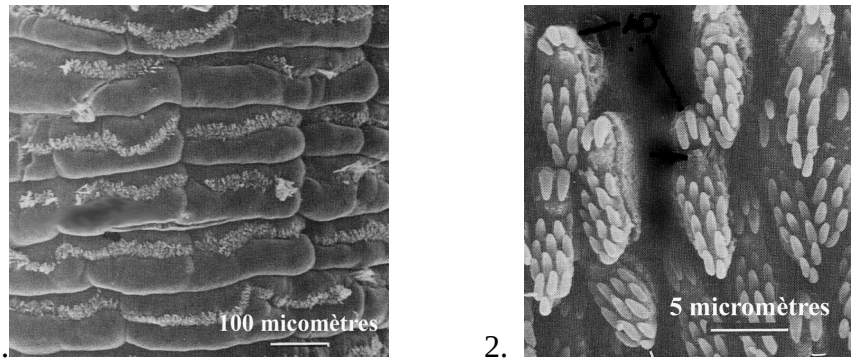


Figure 27. Microscopie électronique de l'opistosome adulte de *Riftia pachyptila* 1. détail des segments de l'opistosome avec les soies (crêtes diffuses blanches). 2. Zoom sur les soies. D'après Jones (1985)

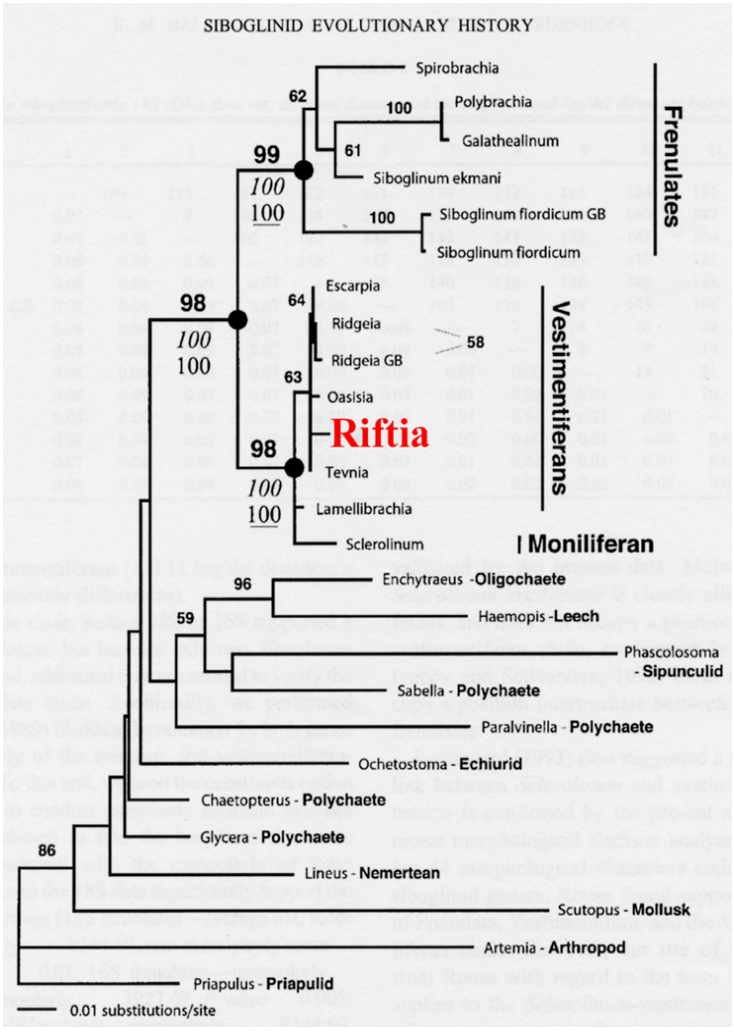


Figure 28. Reconstruction phylogénétique par maximum de vraisemblance à partir de l'ARN 18 S avec 500 itérations de Bootstrap. Seules les valeurs de bootstrap supérieures à 50¨% sont montrées. On peut voir à la racine des siboglinidés (vestimentifères (dont *Riftia pachyptila*), pogonophores (Frenulates) et monolifera) des polychètes et oligochètes. D'après Halanych et al (2001)

A partir de leur analyse cladistique, Rouse et Fauchald (1997 ; 2001) ont proposé de regrouper pogonophores et vestimentifères au sein des Siboglinidés, eux-mêmes compris dans les polychètes. Bien qu'il n'y ait pas de démonstration de la monophylie des Annélides, plusieurs phylogénies moléculaires basées sur le facteur d'élongation 1α montrent que les pogonophores (dont les vestimentifères) dérivent des annélides (Kojima et al. 1993; McHugh 1997) ou que les pogonophores (dont les vestimentifères) ne forment pas un phylum à part entière (Boore and Brown 2000). La figure 28 montre que les vestimentifères et les pogonophores (Frenulates) sont des groupes frères (Halanych et al. 2001).

D'un point de vue moléculaire, la comparaison des hémoglobines d'annélides, de pogonophores et de vestimentifères, montre une proximité structurale difficilement attribuable à une l'évolution convergente.

En effet, de très fortes ressemblances existent, tant sur l'organisation de la famille multigénique codant pour les globines ou de la position des introns dans les gènes de globines, que sur la conservation de la structure quaternaire hautement complexe de ces mêmes hémoglobines. Ce dernier point est détaillé dans les paragraphes suivants.

VI. Les hémoglobines de *Riftia pachyptila* : analyse comparée avec les hémoglobines d'Annélides

1. Les hémoglobines extracellulaires de *Riftia pachyptila*

Riftia pachyptila possède 3 hémoglobines extracellulaires capables de lier réversiblement, sur 2 sites différents, l'oxygène et le sulfure d'hydrogène (Arp et al. 1987). Les hémoglobines V1 (3500 kDa) et V2 (400 kDa) sont dissoutes dans le système vasculaire alors que l'hémoglobine C1 (400 kDa) se trouve dans le liquide coelomique (Zal et al. 1996b). L'hémoglobine V1 forme une structure complexe en bicouche hexagonale HBL-Hb (HBL, pour hexagonal bilayer et Hb pour hémoglobine) composée de 56 chaînes de structure (ou linkers ; fig 29) qui assurent la cohésion de la molécule et d'environ 144 chaînes polypeptidiques de type globine. Les hémoglobines V2 et C1 sont constituées par 24 chaînes de globines uniquement (Zal et al. 1996a).

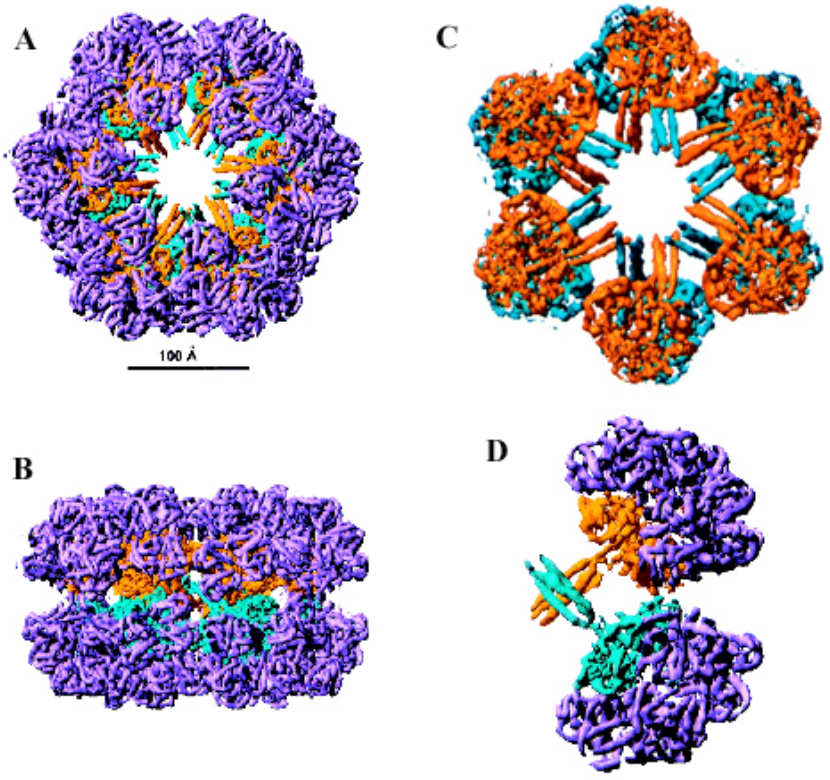


Figure 29.

Seule la structure de l'hémoglobine HBL-Hb de *Lumbricus terrestris* qui est résolue en densité électronique à 5.5 Å permet de voir la position des linkers.

A. vue de dessus de l'HBL-Hb avec les globines (violet) et les linkers (orange et bleu clair).

B. vue latérale de l'HBL-Hb

C. Les linkers oranges sont associés avec les globines de la première couche, les linkers bleu clair avec celles de la seconde couche.

D.Association des globines et des linkers

D'après Royer et al (2000).

Les globines extracellulaires des annélides, des pogonophores et des vestimentifères sont divisées en 2 familles A et B (Gotoh et al. 1987) et subdivisées en 4 sous-familles (A1, A2, B1 et B2) selon la nomenclature de Suzuki et al (1993). Cette nomenclature n'est cependant pas toujours utilisée car le problème majeur est de classer la globine nouvellement séquencée dans le bon groupe de globines orthologues. Certains auteurs ont inscrit leurs séquences dans des banques de données selon des nomenclatures plus personnelles. Dans sa phylogénie des globines d'annélides Dewilde utilise Type 1 pour les globines A et type II pour les globines de la famille B (Dewilde et al. 1996 ; fig. 30).

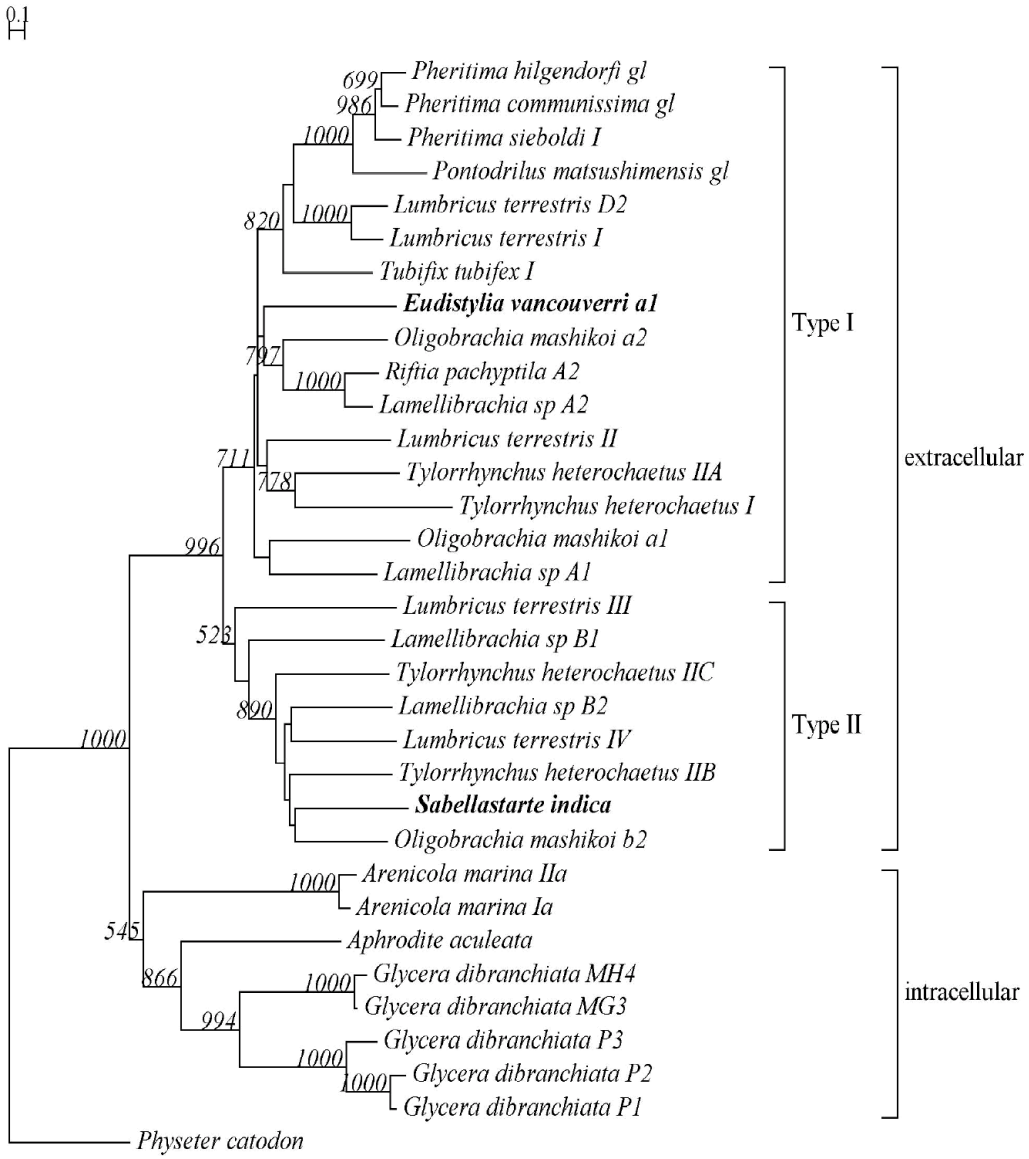


Figure 30. Phylogénie moléculaire des globines d'annélides. Légendes des taxons en annexe 1. D'après Dewilde et al (1996)

2. Cystéines : pont dissulfures intra et inter-chaînes et fixation du sulfure d'hydrogène

La présence de **cystéines libres conservées** (non impliquées dans des ponts intra ou inter-globines (fig. 31)) a été observée chez **les globines A2** de *Riftia pachyptila*, du vestimentifère *Lamellibrachia sp* (Suzuki et al. 1990a), du pogonophore *Oligobrachia mashikoï* (Yuasa et al. 1996) et chez **les globines B2** du vestimentifère *Lamellibrachia sp* (Takagi et al. 1991), du pogonophore *Oligobrachia mashikoï* (Yuasa et al. 1996) et de l'annélide sabelidé *Sabellastarte indica* (Suzuki et al. 1995). **Ces cystéines libres sont absentes des famille A1 et B1**.

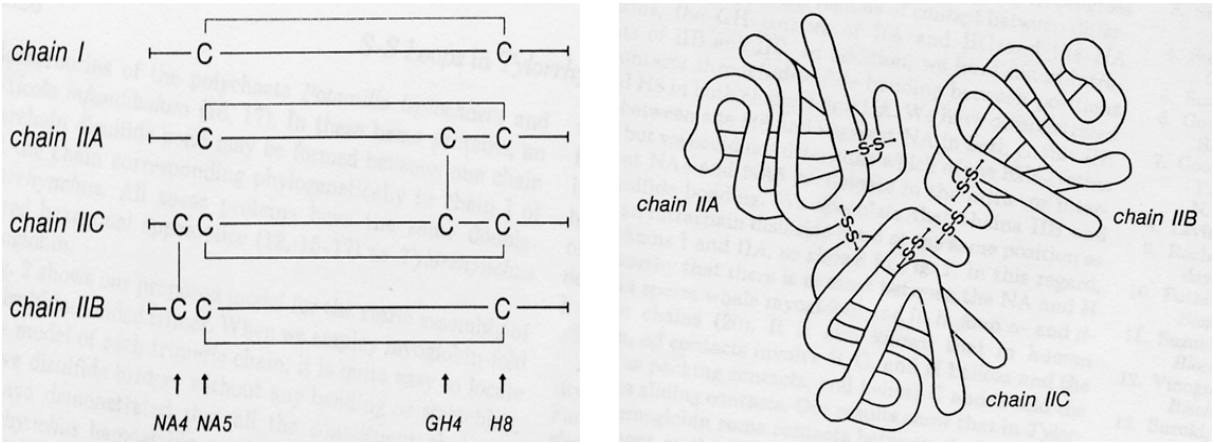


Figure 31. Pont dissulfures intra et inter chaîne chez le polychète *Tylorrhynchus heterochaetus*. 1. Exemple de pont dissulfure intra-chaîne commun à toutes les HBL-Hbs (entre les cystéines NA5 et H8) ou inter-chaîne (entre deux NA4 et entre deux GH4). 2. Dans ce cas, ces ponts dissulfures génèrent un hétérotrimère de globines. D'après Suzuki et al (1988).

Suzuki a été le premier à suggérer que les cystéines libres étaient impliquées dans la fixation du sulfure d'hydrogène. Ses travaux sur le vestimentifère des suintements froids du Japon, *Lamellibrachia sp*., le conduisirent à soupçonner, suite à une simple observation olfactive, que l'hémoglobine de son modèle fixait également le sulfure d'hydrogène. L'hémoglobine purifiée de *Lamellibrachia sp* stockée 6 à 7 jours à 4°C dégageait une odeur de mercaptan (Suzuki et al. 1989). Dans ce même article et dans un article suivant (Suzuki et al. 1990a) ces auteurs proposèrent que les cystéines libres pourraient fixer le sulfure d'hydrogène. Cette hypothèse fut ensuite démontrée grâce au blocage spécifique des résidus cystéines par la N-ethylmaléimide sur des hémoglobines purifiées vasculaires 1, 2 et coelomiques de *Riftia pachyptila* et ce pour la cystéine libre de A2 et celle de B2 (Zal et al. 1998a ; fig. 32).

Le sulfure d'hydrogène se lie donc de manière covalente et réversible sur les cystéines libres des globines A2 et B2 de *Riftia pachyptila*.

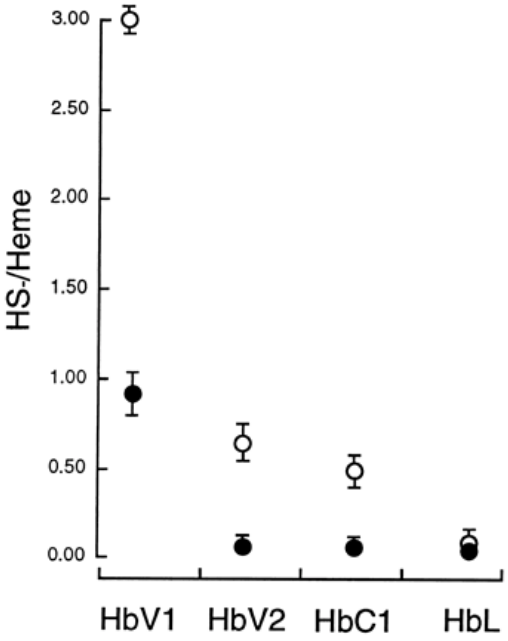


Figure 32. Fixation de H_2S avant (○) et après traitement (●).

Par les trois hémoglobines de *Riftia pachyptila*;
HbV1: hémoglobine vasculaire V1 (HBL-Hb)
HbV2: hémoglobine vasculaire V2
HbC1: hémoglobine coelomique
Par l'hémoglobine de *Lumbricus Terrestris*: HbL

Le NEM mime la fixation de HS^- sur le groupement thiol R-SH, selon la formule suivante:

$$\left(R\text{–}SH + HS^- + H^+ \underset{[2H^+]}{\overset{[2H^+]}{\rightleftarrows}} R\text{–}SSH \right)$$

D'après Zal et al. (1998).

On remarquera sur le graphique de la figure 32 que l'hémoglobine de *Lumbricus terrestris* avant ou après traitement ne fixe pas de H_2S. Ce ver vit en milieu oxydé sans sulfures et ne possède pas de cystéine libre sur aucune de ses globines.

On remarquera également que l'hémoglobine V1 après traitement (la seule qui contienne des chaînes de structure (linker)), conserve un reliquat de fixation de H_2S. Après le séquençage complet du linker de *Lamellibrachia sp* et suite à des expériences préliminaires, Suzuki et al (1990b) ont proposé que des groupements thiols pourraient également fixer H_2S et participer à son transport.

En fait, dans cette démonstration de la fixation de H_2S par les hémoglobines totales de *Riftia pachyptila*, un second mécanisme, associé uniquement aux HBL-Hb et plus particulièrement aux chaînes de structure (linkers), permet également de fixer des sulfures grâce à l'ouverture de ponts dissulfures. Ce mécanisme repose sur la formation de groupements persulfates (Zal et al. 1998a), selon la formule:

$$R\text{–}SS\text{–}R + 2HS^- + 2H^+ \underset{[2H^+]}{\overset{[2H^+]}{\rightleftarrows}} 2R\text{–}SSH$$

Ainsi, les deux mécanismes de fixation des sulfures par les globines (A2 et B2) et les chaînes de structures des hémoglobines de *Riftia pachyptila*, sont respectivement associés aux cystéines libres et à l'ouverture de ponts dissulfures (fig. 33).

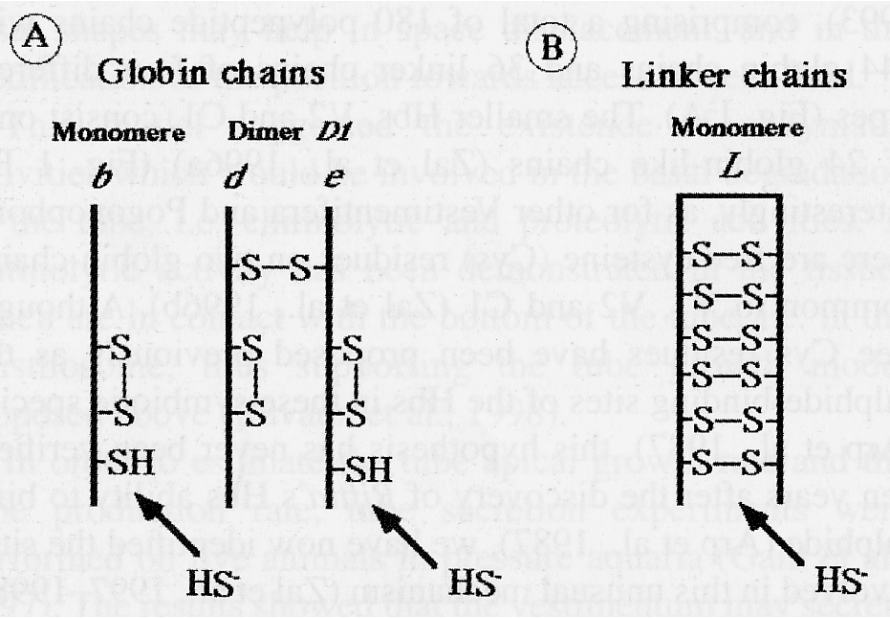


Figure 33. Les deux mécanismes de fixation des sulfures par les hémoglobines de *Riftia pachyptila*. Les chaînes A2 (b) et B2 (e) portent les cystéines libres. L symbolise un des linkers. D'après (Zal 1998).

Chez *Riftia pachyptila*, les HBL-Hbs ne présentent pas de trimères de globines comme pour *Lumbricus terrestris* et *Tylorrhynchus heterochaetus* (fig.31), mais des dimères de globines.
A ce jour, les seules autres séquences protéiques orthologues des globines A2 et B2 d'annélides non symbiotiques vivant en présence de H_2S, sont celles de Sabellidés (polychètes). Il s'agit de la chaîne B2 de *Sabellastarte indica* (Suzuki et al. 1995) et de *Sabella spallanzanii* (Pallavicini et al. 2001), et de la chaîne A2 de *Sabella spallanzanii* (Pallavicini et al. 2001).
Cependant, la caractérisation par spectrométrie de masse des hémoglobines de *Arenicola marina*, une annélide polychète des estrans riches en H_2S, et d'*Alvinella pompejana*, une annélide polychète des sources hydrothermales, toutes deux **non symbiotiques**, a révélée aussi la présence de cystéines libres capable de fixer des sulfures (Zal et al. 1997a ; Zal et al. 1997b). A ce jour, les séquences protéiques comme nucléotidiques de ces hémoglobines ne sont pas disponibles. La localisation des cystéines libres au sein des globines orthologues A2 et B2 de ces deux derniers annélides n'est donc pas encore connue.

VII. conclusion

Riftia pachyptila est un modèle biologique qui offre le double avantage d'être une espèce symbiotique d'une part, et emblématique des milieux réduits instables et fragmentés d'autre part. Ce statut le place en termes d'adaptation comme espèce d'étude de référence. Son métabolisme singulier est basé sur la chimiosynthèse bactérienne qu'il entretient grâce au transport par ses hémoglobines d'un composé énergétique et toxique, le sulfure d'hydrogène. De plus, cette espèce possède comme les autres annélides des milieux oxydés des caractères conservés comme l'organisation de sa famille d'hémoglobines. Ainsi, la nature de l'environnement réduit dans lequel vit *Riftia pachyptila* permet d'accéder à des stratégies d'adaptation originales acquises par cet organisme. Ceci est illustré par la fonction de transport des hémoglobines d'un composé toxique agressif majoritairement létal pour les espèces des milieux oxydés : le sulfure d'hydrogène.

Cette thèse qui porte sur l'étude de l'évolution moléculaire des hémoglobines extracellulaires de *Riftia pachyptila* et leur capacité à lier réversiblement le sulfure d'hydrogène commence avec un jeu de données à deux dimensions, physiologique et biochimique (ce chapitre). L'objectif principal de ce travail est d'accéder aux séquences codantes des globines et des "linkers" de *Riftia pachyptila* afin de comprendre, à travers l'émergence de la famille multigénique, l'apparition et le maintient de la fonction de fixation de l'H_2S. Cette démarche repose majoritairement sur l'identification et la mesure de l'intensité des pressions évolutives. En d'autres termes, décrypter les effets de la sélection naturelle à un niveau moléculaire.

Nous avons ouvert une troisième dimension, celle de l'évolution moléculaire, à partir d'outils de la biologie moléculaire et de concepts en évolution pour retracer l'histoire de la famille multigénique des hémoglobines de *Riftia pachyptila*. Nous avons décrit cette famille multigénique, cherché à localiser les cystéines libres impliquées dans la fixation du sulfure d'hydrogène, déterminé les vitesses évolutives, analysé comparativement les globines de *Riftia pachyptila* avec celles déjà séquencées des autres annélides. Cette première partie qui apparaît dans les chapitres 3 et 4 nous a permis de formuler des scénarii évolutifs sur l'acquisition et le maintient de la fonction de fixation de H_2S au cours de l'évolution des annélides, qu'elles soient symbiotiques ou non.

Nous avons dans un deuxième temps (chapitre 5), séquencé des globines supplémentaires de vestimentifères et, grâce à des séquences nucléotidiques de globines d'un annélide oligochète d'une EST (Expressed Sequence Tag) disponible dans les banques de données, nous avons pu aborder le problème de l'absence des cystéines libres chez les annélides des milieux oxydés.

Cette approche se pose en terme d'adaptation moléculaire. Elle utilise de récents programmes basés sur des algorithmes de maximum de vraisemblance qui nous ont permis de proposer un mécanisme évolutif pouvant expliquer la perte de la fonction de fixation des sulfures chez les annélides des milieux oxydés.

Chapitre 3 :

Analyse de la famille multigénique codant les hémoglobines de *Riftia pachyptila*

Chapitre 3. Analyse de la famille multigénique codant les hémoglobines de *Riftia pachyptila*

I. Introduction

Ce chapitre décrit la famille multigénique de *Riftia pachyptila***,** sa structuration et sa composition. Cette famille multigénique est comparée aux autres familles de globines d'annélides, en terme d'organisation, de sous-unités (nombre et masse moléculaire) et en terme de structure de gène. Ce chapitre tente également de faire le lien entre les données de spectrométries de masse et celles de biologie moléculaire. La première étape de cette thèse a consisté à séquencer et caractériser les séquences de globine correspondant aux six types de chaînes de globines décrites par spectrométrie de masse qui composent des hémoglobines de *Riftia pachyptila*. Les trois hémoglobines V1, V2 et C1 de *Riftia pachyptila* n'ont pas la même structure et composition (Chapitre 2 ; Zal et al. 1996 a et b). Six globines a, b, c, d, e et f ont été dénombrées par spectrométrie de masse (tableau 1) mais seules trois globines b, c, et e sont retrouvées dans les trois hémoglobines. De plus, les sous-unités **d** et **e** d'une part, et **e** et **f** d'autre part, forment des dimères, respectivement nommés **D1** et **D2** (Zal et al. 1996 b).

Chaîne d'hémoglobines	Masse moléculaire (Da)	V1	V2	C1
a	15933,4	○	★	★
b	16133,5	★	★	★
c	16805,9	★	★	★
d *	15578,5	★	★	○
e **	16149,1	★	★	★
f *	16368,1	○	★	★

Tableau 1.Composition du système multi-hémoglobine de *Riftia pachyptila* V1: vasculaire (HBL-Hb) ; V2: vasculaire ; C: coelomique. **○: absent, ★: présent. * d+e forment le dimère D1. * e+f forment le dimère D2** D'après Zal et al. (1996)

Pour expliquer et comprendre l'émergence d'une telle complexité, nous avons entrepris de séquencer ces 6 globines et d'assigner à chacune d'elle son vis-à-vis décrit en spectrométrie de masse. Une phylogénie moléculaire de ces globines, leurs signatures moléculaires et leurs potentiels de dimérisation sont autant d'informations pouvant permettre d'élaborer un scénario évolutif.

II. Matériel et méthodes

Le matériel et les méthodes ont été en grande partie décrits dans les chapitres 4 et 5. Cependant un complément sur le concept d'amorce dégénérée, l'extraction d'ADN génomique, l'extraction des ARN totaux, la purification des ARN messagers et la réalisation de la banque se trouvent en annexe 2. Nous avons conçu quatre jeux d'amorces dégénérées à partir de séquences de globines du vestimentifère *Lamellibrachia sp* des suintements froids du Japon dont une globine de chaque sous-famille A1, A2, B1 et B2 était disponible dans les banques de données. Retenons que la dégénérescence des amorces conjuguée à une faible stringence lors de l'amplification par PCR (faible température d'hybridation) nous a permis de cloner et de séquencer 3 globines dans la famille B1 à partir d'un seul jeu d'amorces dégénérées. Le séquençage partiel des 6 ADNc correspondant aux 6 sous-familles de globines de *Riftia pachyptila* nous a permis de synthétiser des sondes spécifiques et de cribler une banque d'ADNc et d'en récupérer certains complets. Le séquençage partiel d'un des quatre linkers qui composent l'HBL-Hb a également pu être réalisé ainsi que le séquençage des introns des gènes codant les globines A2 et B2. L'ensemble de ces données a été utilisé pour décrire l'évolution de la famille multigénique par rapport à la propriété de fixation et de transport du sulfure d'hydrogène, et pour rechercher les positions des cystéines libres sur les séquences primaires de globines (chapitre 4 et 5).

III. Résultats et Discussion

1. Séquençage des ADNc de globines et organisation de la famille multigénique

Six ADNc différents ont été séquencés soit complètement ou partiellement (annexe 3). Dans les cas de séquençage partiel, 80 % de la séquence codante est séquencée. Nous avons recherché à identifier dans un premier temps les groupes d'orthologie de ces globines à partir d'une phylogénie moléculaire des globines des vestimentifères, pogonophores et annélides des milieux riches en sulfures. Le regroupement distinct au sein de chaque sous-famille et les signatures moléculaires évidentes repérées dans l'alignement multiple (annexe 4 et chapitre 4) au sein de chaque groupe orthologue, nous ont permis d'assigner sans ambiguïté ces nouvelles globines à leur groupe respectif.

A partir de l'alignement des globines de *Riftia pachyptila* , nous avons retrouvé la présence de séquences d'adressage (en orange fig. 34) pour A1, A2, B1a et B2 grâce à une suite de

programmes prédictifs du site de clivage associé au signal peptide (http://psort.nibb.ac.jp/form2.html) (Nakai and Horton 1999). La séquence B2 ne commence pas par une méthionine et suggère que la partie 5' de la séquence d'adressage est incomplète. Seules les séquences codantes A2, A1 et B1a sont complètes. Les autres sont partiellement séquencées mais suffisamment pour contenir les signatures moléculaires permettant d'identifier les différentes globines paralogues de cette famille multigénique.

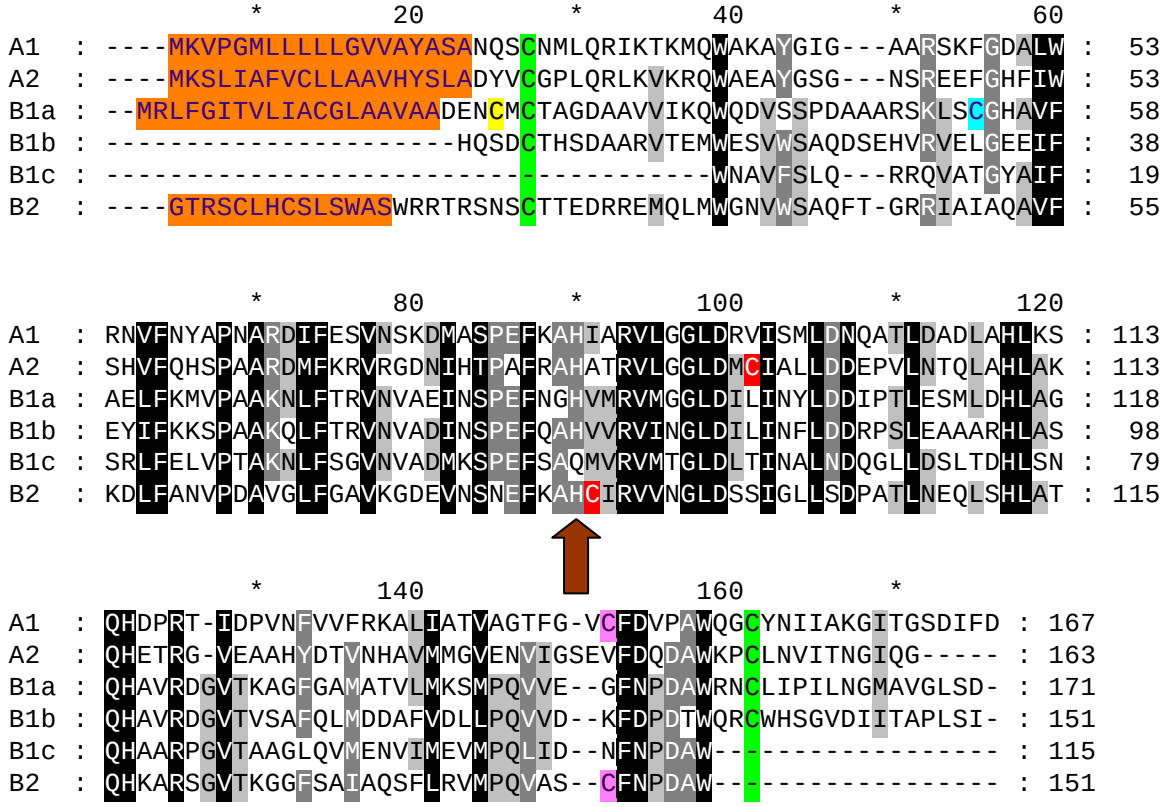


Figure 34. Alignement multiple des 6 globines de *Riftia pachyptila*. **Séquences oranges**: peptide signaux ; **cystéines vertes** pour les ponts disulfures intra-chaîne ; **cystéines rouges** cystéine de liaison de H_2S ; **cystéines rose** et **bleues** pour les ponts disulfures inter-chaînes ; cystéine **jaune** cystéine fonction inconnue. 100% de similarité, 80% de similarité, 60% de similarité. Histidine proximale (flèche)

L'ensemble des positions des cystéines est résumé dans la figure 35 qui ne prend pas en compte les séquences d'adressage. On peut caractériser 3 types de cystéines. Les cystéines qui réalisent un pont **disulfure intra-chaîne** chez toutes les hémoglobines extracellulaires des annélides et qui sont localisées en positions 7 et 143. Les cystéines impliquées dans des ponts **disulfures inter-chaînes** qui sont localisées en position 6, 35 et 133. Puis les deux cystéines **Cys+1 et Cys+11 impliquées dans la fixation des sulfures**, et dont la nomenclature particulière (Bailly et al. 2002) est définie à partir du nombre d'acides aminés les séparant de l'histidine proximale conservée (rarement une glutamine) (flèche marron).

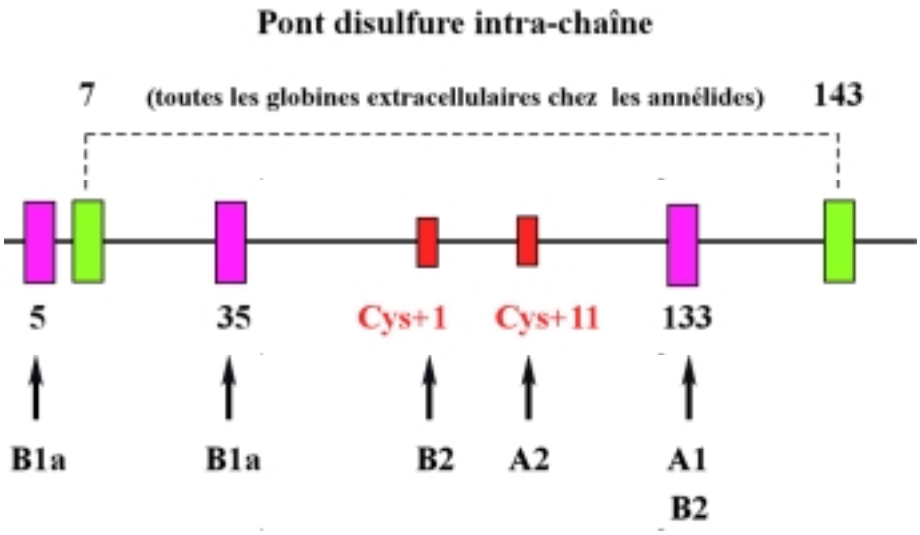


Figure 35. Positions des différentes cystéines chez les globines de *Riftia pachyptila*

2. Comparaison des séquences obtenues et des données de spectrométrie

L'utilisation d'un spectromètre de masse à nébulisation électrique (Zal et al. 1996 a et b) a préalablement permis de calculer le poids moléculaire de chaque chaîne de globine, de détecter le nombre de cystéines dans chacune de celles-ci et d'identifier les globines impliquées dans la formation des dimères (tableau 2).

Chaîne de globine	Masse moléculaire (Da)	Nombre de cystéines
a	15933,4	2
b	16133,5	3
c	16805,9	2
d*	15578,5	3
e**	16149,1	4
f*	16368,1	3

Tableau 2. Masse moléculaire et nombre de cystéines des chaînes de globines estimées par spectrométrie de masse. L'association entre les chaînes * forme le dimère D1 et celle entre les chaînes * constitue le dimère D2.

Dans un premier temps, les masses moléculaires putatives et le nombre de cystéines nous ont permis d'assigner **respectivement la chaîne f et la chaîne b aux sous-familles A1 et A2** (tableau 3).

Pour les chaînes a, c, d et e, soit le séquençage est incomplet, soit la masse moléculaire du polypeptide, déduite de la traduction de l'ADNc (tableau 3), n'est pas la même que la masse obtenue par spectrométrie de masse (tableau 2). Cependant, la chaîne **e** possède 4 cystéines

(tableau 2) comme la globine B2 et B1a (tableau 3). La chaîne **e** possède : (1) deux cystéines pour réaliser un pont intra-chaîne (2) une cystéine pour réaliser un pont disulfure inter-chaîne (dimère) et (3) une cystéine libre (Chapitre 2, fig. 33). Comme chez *Lamellibrachia sp*, La globine B2 est la seule qui présente une cystéine libre (Cys+1) impliquée dans la fixation des sulfures. On retrouve également la cystéine conservée en position 133 de la globine B2 de *Riftia pachyptila* comme celle de *Lamellibrachia sp* et des autres annélides (annexe 4). Par contre la globine B1a de *Riftia pachyptila* ne possède pas de cystéine libre impliquée dans la fixation des sulfures, ni de cystéine en position 133. Chez *Lamellibrachia sp*, la globine A1 (chaîne **f** chez Riftia) qui présente comme celle de *Riftia pachyptila* une cystéine conservée en position 133, forme un dimère avec la globine B2 (Suzuki, 1989). Ces observations sont suffisantes pour affirmer que **la globine B2 est la chaîne e décrite par spectrométrie de masse et fait un dimère avec la chaîne f soit la globine A1 chez *Riftia pachyptila*.**

Bien que la globine B1a présente un poids moléculaire dépassant sensiblement celui calculé par spectrométrie pour la chaîne **d**, et **une cystéine surnuméraire inattendue en position 5** (tableau 3), elle reste l'unique candidate pour former un dimère avec la globine B2 (ou chaîne **e)** entre la cystéine 133 de B2 et la cystéine en position 35 de B1a (fig.35). Les deux autres globines B1b et B1c comme les autres globines orthologues B1 des vestimentifères et pogonophores (annexe 4), ne possèdent que les cystéines du pont disulfure intra-chaîne (positions 7 et 143, fig.35).

La présence de la cystéine surnuméraire de B1a en NH_2 terminal pourrait être expliquée par un problème de séquençage, mais tous les clones la possèdent. Une modification post-transcriptionnelle, post-traductionnelle ou un clivage atypique de la séquence d'adressage ne sont pas à écarter pour expliquer la présence de cette cystéine et la différence de poids moléculaire observée.

Séquence	Nombre de cystéines (après séquençage)	Masse moléculaire théorique après séquençage	Assignation séquençage/spectrométrie
B1 b	2 =	17042 #	a ou c
A2	3 =	16134 =	b
B1 c	2 =	Séquence incomplète	a ou c
B1 a	4 # (3 attendues)	16304 #	d
B2	4 =	Séquence incomplète	e
A1	3 =	16367 =	f

Tableau 3. Comparaison des chaînes de globines obtenues par spectrométrie de masse et séquençage.
= : identiques aux données de spectrométrie, # : différentes.

3. Evolution de la famille multigénique des globines de *Riftia pachyptila*

A partir des alignements (fig. 34 et annexe 4) et de la comparaison des données de spectrométrie avec celles de biologie moléculaire, nous avons réalisé une phylogénie moléculaire des globines de *Riftia pachyptila*. Elle montre que la famille multigénique actuelle a émergé à partir de 5 évènements de duplication, conduisant à deux familles, A et B et quatre sous-familles A1, A2, B1 et B2 et dans la sous-famille B1, aux globines B1a, B1b et B1c (fig. 36).

Deux dimères D1 et D2 sont formés par l'association des globines B2 et B1a d'une part et par B2 et A1 d'autre part (fig. 36).

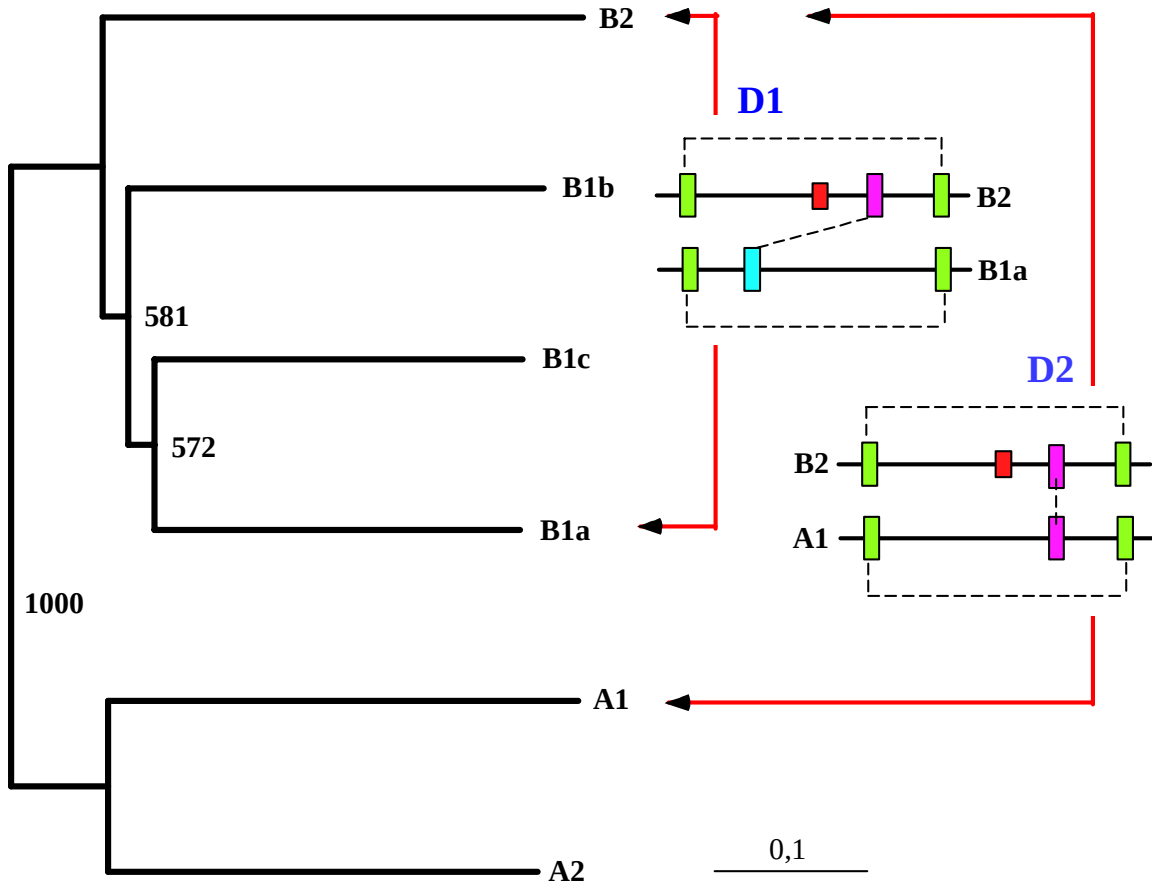


Figure 36. Correspondance entre la phylogénie moléculaire des globines de *Riftia pachyptila* et leur dimérisation. Arbre phylogénétique obtenu par Neighbor Joining avec un test de bootstrap sur 1000 réplicats.

Seule l'hémoglobine V2 est formée par toutes les globines (présentes à différents ratios) et par les deux dimères D1 et D2. L'hémoglobine V1 ne contient pas le dimère D2 car la gobine A1

n'y est pas retrouvée. De même pour l'hémoglobine coelomique C1, le dimère D1 est absent car la globine B1a ne fait pas partie de cet assemblage. Ainsi, le dimère peut être considéré comme une unité fonctionnelle qui pourrait conférer au polymère dans lequel elle se trouve une meilleure stabilité ou permettrent des phénomènes allostériques complexes. Ces dimérisations pourraient permettre la fixation d'autres ligands : dans le cas des trimères de globines de *Lumbricus terrestris*, cette association serait au cœur de la fixation de ions calcium, grâce à un faisceau de charge positives (Fushitani et al. 1988). La fixation du calcium permettrait d'avoir un fort effet Bohr pouvant conférer un avantage sélectif en terme d'affinité pour l'oxygène.
On remarquera que trois globines sont simultanément présentes dans les trois hémoglobines V1, V2 et C1 : les globines A2, B2 et une des deux globine B1b ou B1c qui ne possède que les deux cystéines permettant la formation du pont dissulfure intra-chaîne (tableau 1). Pour B1 (a ou c) on peut penser à un rôle structural clé ou à une capacité de fixation particulière, mais ces hypothèses restent très spéculatives. Par contre, la présence des globines B2 et A2 peut s'expliquer par l'importance de la fixation et du transport du sulfure d'hydrogène via les cystéines libres (Cys+1 et Cys+11).

4. Evolution de la famille multigénique des globines au sein des annélides

La présence et la position des cystéines impliquées dans les ponts disulfures inter et intra-chaînes des hémoglobines ainsi que les possibilités de dimérisations et trimérisations des globines chez les principaux groupes annélidiens (polychètes, oligochètes, achètes, pogonophores et vestimentifères) permettent de dégager un premier schéma évolutif de la famille multigénique des hémoglobines extracellulaires des annélides (fig. 37).

D'une part, les polychètes et oligochètes ont des hémoglobines présentant des trimères qui associent les globines A2 et B2 par leurs cystéines en position 133 et les globines B2 et B1 par leurs cystéines en position 6 (chapitre 2, fig.31 ; Fushitani et al. 1988 ; Suzuki et al. 1988). **La globine A1 reste toujours monomérique**.

D'autre part, les achètes, les pogonophores et les vestimentifères présentent des hémoglobines dimériques. **La globine A2 reste toujours monomérique**.

Chez les vestimentifères *Riftia pachyptila* et *Lamellibrachia sp* et chez le pogonophore *Oligobrachia mashikoi* les globines B2 et A1 sont associées par leurs cystéines en position 133.

Chez *Riftia pachyptila*, un autre dimère est réalisé entre les globines B2 et B1, mais associant deux cystéines en des positions différentes, respectivement 133 et 35. La cystéine 35 est le résidu supposé de la globine B1a permettant de réaliser le pont disulfure décelé par spectrométrie de masse.

Chez le pogonophore *Oligobrachia mashikoi*, il existe un trimère de globines résultant de l'association du dimère B2-A1 et d'une troisième globine. La troisième globine n'est ni A2, B2 ou A1 selon les données de spectrométrie de masse et SDS-page (Yuasa et al. 1996). Il est fort probable que ce soit une globine B1, mais on ne connaît pas la position de la cystéine qui permet le pont disulfure avec celle de la globine B2.

Pour les achètes, seules quatre globines co-existent, deux forment un dimère, et les deux autres sont des monomères (Shishikura et al. 1997). Bien que seules les séquences NH_2 terminale soient disponibles (annexe 4), on peut facilement identifier deux lignées A et B ainsi que deux sous-lignées au sein de chaque famille. Aucun indice ne permet de discriminer les sous-famille en A1, A2, B1 ou B2 à cause du micro-séquencage (par définition court, donc ne fournissant pas une séquence protéique assez longue). **La seule certitude est qu'une des deux globines B forme un dimère avec une des deux globines A** (Shishikura et al. 1997).

Le scénario le plus parcimonieux, eu égard aux positions conservées des cystéines et à celles perdues, repose sur une émergence de la famille multigénique A1, A2, B1 et B2 avant la séparation du groupe annélidien formant des trimères et du groupe annélidien formant des dimères (fig.37). En effet, la présence simultanée des cystéines en positions 133 chez les trois globines A1, A2 et B2 des annélides actuelles suggère que, chez les globines B1, la cystéine en position 133 a été perdue après la duplication de la globine B ancestrale en B1 et B2 (fig. 37). De plus, le probable résidu cystéine en position 133 au sein des globines ancestrales A et B suggère aussi que celles-ci pouvaient théoriquement réaliser des dimères de globines.
On peut raisonnablement penser que la cystéine 133 était initialement présente chez l'ancêtre annélidien mais qu'elle fut également perdue dans la lignée A2 formant des dimères, ainsi que dans la lignée A1 à trimères. La cystéine 6 a dû apparaître avant la duplication B1 et B2, mais perdue dans le groupe formant des dimères. Il est fortement moins probable que la cystéine en position 6 ait pu seulement apparaître dans le groupe formant des trimères car cela suppose un

gain simultané et indépendant à la même position dans la lignée B2 et B1 de ce même groupe.

Les dimérisations impliquant les globines B1 au sein du groupe formant des dimères restent plus difficile à interpréter. Cependant l'exemple de la diversification de la sous-famille B1 de *Riftia pachyptila* (positions marginales des cystéines 5 et 35), des famille de globines de *Lamellibrachia sp* (Suzuki et al. 1989) et d'*Oligobrachia mashikoi* (Yuasa et al. 1996) suggèrent que la radiation des globines B1 est récente et spécifique des vestimentifères et des pogonophores.

Ainsi ces duplications des globines B1 chez les pogonophores et les vestimentifères auraient pu générer des dimérisations originales avec cystéines en positions nouvelles sur des gènes dupliqués de type B1 orphelins (n'ayant pas d'orthologues), permettant même dans le cas de *Oligobrachia mashikoi* de faire un trimère de type B2-A1-B1 qui n'implique pas la globine A2 monomérique (Yuasa et al. 1996).

Bien que le but de cette thèse ne soit pas de résoudre la phylogénie des annélides, ces résultats préliminaires apportent une vision un peu différente des phylogénies classiques et suggèrent une dichotomie entre le groupe à trimères (polychètes et oligochètes) d'une part, et le groupe à dimères (vestimentifères, pogonophores et achètes) d'autre part. La classification selon Rouse et Fauchald (1997) qui place les vestimentifères et les pogonophores au sein des polychètes, et les oligochètes et les achètes au sein des Clitellata à l'extérieur des polychètes, n'est donc pas soutenue par nos résultats. L'ensemble des phylogénies des annélides montre que les annélides polychètes sont à la base des vestimentifères et les pogonophores (chapitre 2). Donc, si les polychètes sont à la base des annélides comme le suggère également Westheide (1997), on peut imaginer d'après la figure 37 qui allie (1) les duplications de gènes de globines, (2) les capacités de polymérisation spécifiques de ces globines et (3) les évènements de spéciation, qu'une lignée de polychète aurait pu être à l'origine du groupe des vestimentifères, des pogonophores et des achètes (ou hirudinés) formant des dimères n'impliquant pas **A2 monomérique**. Une autre lignée de polychètes serait à l'origine des oligochètes émergeant cependant dans le groupe à trimères, dont A1 monomérique.

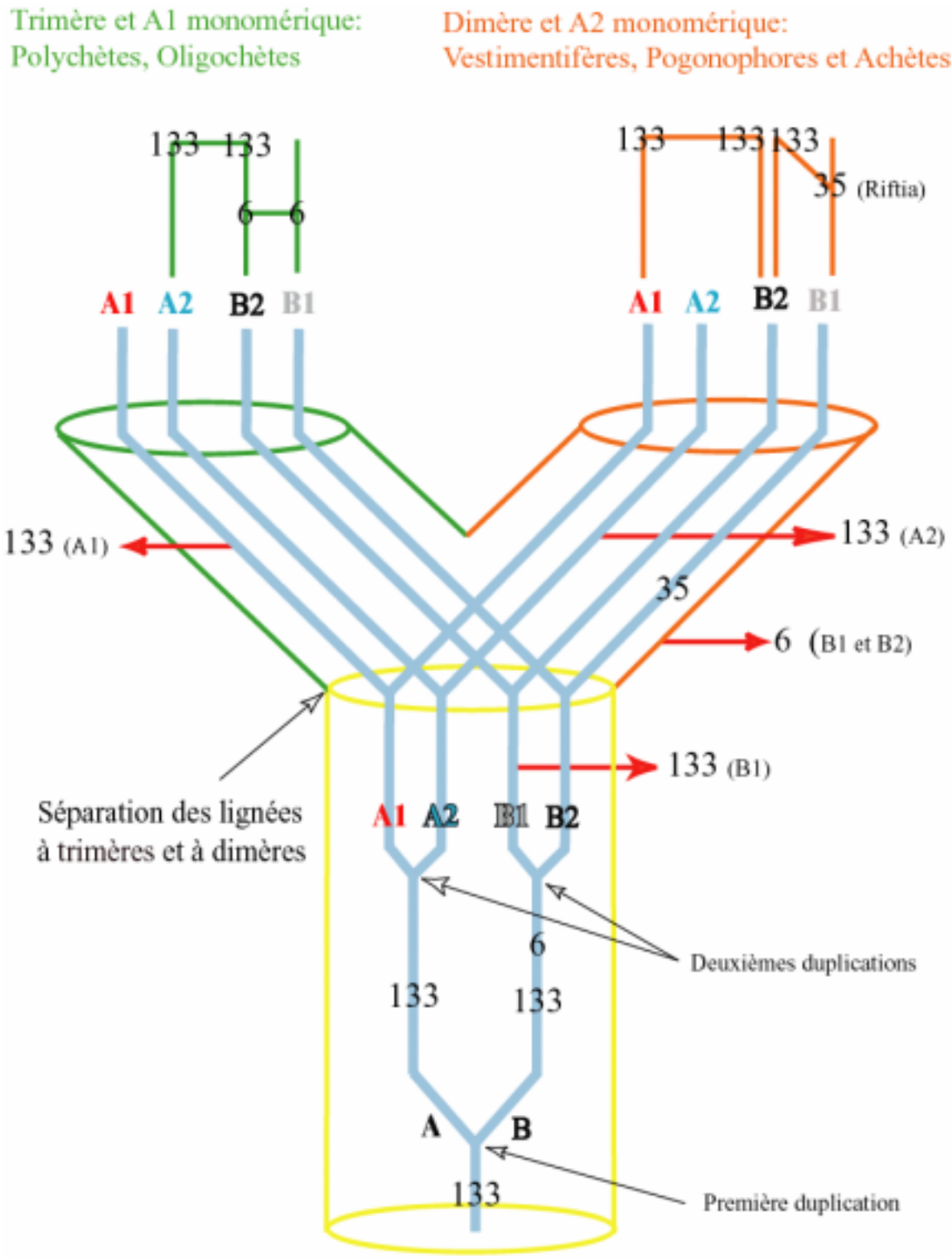


Figure 37. Evolution de la famille multigénique des hémoglobines au sein des Annélides. En orange : les vestimentifères, les pogonophores et les achètes. En vert : les polychètes et les oligochètes. En jaune : ancêtre annélidien avant radiation. Les nombres et chiffres qui apparaissent sur les branches bleues symbolisent la position et le moment (approximatif) de l'acquisition de cystéines. Les flèches rouges symbolisent la perte de cystéines.

5. Chaîne de structure des hémoglobines vasculaires V1 (HBL-Hb) de *Riftia pachyptila*

Une des quatre chaînes de structure (linker) qui composent l'HBL-Hb V1 de *Riftia pachyptila* a été séquencée partiellement. Sa séquence apparaît au sein d'un alignement d'autres "linkers" d'annélides dans l'annexe 5. Comme les autres "linkers" il possède une région homologue des LDLr (low density lipoprotein receptor) riche en cystéines. La famille des gènes LDLr code des protéines membranaires de lysosomes qui sont impliquées dans la fixation de ligands spécifiques à dégrader dans l'organite. Dans l'espèce humaine, cette protéine intervient dans la régulation du cholestérol.

Cette région riche en cystéines est probablement impliquée dans la formation des pont disulfures intra chaînes (Suzuki et Riggs 1993) et dans le deuxième mécanisme de la fixation du sulfure d'hydrogène (Chapitre 2, IV, 2, fig. 33).

Les travaux de Suzuki, basés sur la présence de séquences homologues partagées par les globines et les LDLr, suggèrent que les linkers sont des protéines chimériques qui auraient pu émerger par glissement d' exons ("exon shuffling"), à partir de différents gènes (Suzuki et al. 1990b ; Suzuki et Riggs 1993). Cette hypothèse sur l'origine des linkers, associant un ancien gène de globine et un gène (ancestral) de type LDLr est également supportée par les capacités de certains linkers de lier un hème, donc de présenter une poche hydrophobe similaire à celles des globines et de pouvoir potentiellement transporter de l'oxygène (Ownby et al. 1993).

6. Structure du gène A2 et B2 de globine de *Riftia pachyptila* : séquence d'adressage, position des introns et séquence 3'UTR

a. Séquence d'adressage

Les hémoglobines extracellulaires sont d'abord synthétisées dans le tissu hématopoïétique puis exportées dans le réseau vasculaire ou le coelome. Les séquences d'adressage qui permettent l'excrétion des protéines matures ont été retrouvées chez les clones complets A1, A2 et B1a (partiellement chez B2) et sont constitués de 19 acides aminés contre 16 chez *Lumbricus terrestris*. On retrouve, comme pour la majorité des peptides signaux, une région riche en leucines (fig. 34), ainsi qu'une faible identité de séquence entre eux (Williams et al. 2000).

b. Introns et séquences 3' UTR

Les deux introns du gène codant la globine A2 et le deuxième intron du gène codant la globine B2 ont été séquencés (fig. 38). Comme pour le gène codant la globine B1 de *Lumbricus terrestris* (Jhiang et al. 1988), les gènes codant ces globines ont une structure dite en "deux introns - trois exons", identique à la structure des gènes de globines de vertébrés. Les introns trouvés chez *Riftia pachyptila* sont plus courts que ceux trouvés chez *Lumbricus terrestris* mais plus long que ceux trouvés chez les vertébrés. Leurs positions et séquences respectives sont présentées dans l'annexe 6. On remarquera une différence dans la longueur des introns séquencés.

Les introns qui interrompent le cadre de lecture entre deux codons sont dits "en phase zéro". Ceux qui interrompent le cadre de lecture entre la première et la deuxième paire de bases sont dits "en phase un", ceux entre la deuxième et la troisième paire de bases d'un codon sont dits en "phase deux". L'intron 1 de la globine A2 de *Riftia pachyptila* et de la globine B1 de *Lumbricus terrestris* est en phase 2 et l' intron 2 des gènes A2 et B2 de *Riftia pachyptila* et B1 de *Lumbricus terrestris* est en phase 0.

Nous pourrons retenir que la présence d'introns peut conférer au moins deux avantages : celui de l'épissage alternatif qui augmente les possibilités fonctionnelles d'une protéine ou d'une famille de protéine, et celui qui permet le phénomène de glissement d'exon ("exon shuffling") dont la conséquence peut aboutir à la formation de nouveau gènes (Fedorov et al. 1998).

• Présence de microsatellites dans l'intron 2 de A2 (annexe 6) :

La présence de microsatellites dans des introns d'eucaryotes est répertoriée dans la banque de donnée ISIS (Intron Sequence and Information database) à l'adresse suivante : http://isis.bit.uq.edu.au/. Seuls deux motifs $(CA)_n$ et $(CATA)_n$, sur les trois microsatellites de l'intron 2 ont été trouvés dans tous les microsatellites déjà identifiés dans les groupes suivants: Humain, Mammifères, Invertébrés, Plantes, Primates, Rongeurs, Vertébrés et une classe "autres eucaryotes". Le troisième motif microsatellite $(CGCA)_n$ est beaucoup plus rare puisqu'il représente moins de 1% des microsatellites trouvés dans un des groupes précédents. Les rôles fonctionnels des microsatellites ont été largement décrit. Ils entrent dans la composition de certains promoteurs, servent aussi d"enhancers" et peuvent également lier des protéines de régulation pour l'expression de gènes (Golstein et Schlötterer 1999). Il est également connu que certains introns (sans microsatellites) entrent dans la régulation de

l'expression de certains gènes (Ash et al. 1993). Il faudrait pouvoir accéder à l'intron 2 de globines orthologues A2 pour comparer ces motifs et leur degré de conservation.

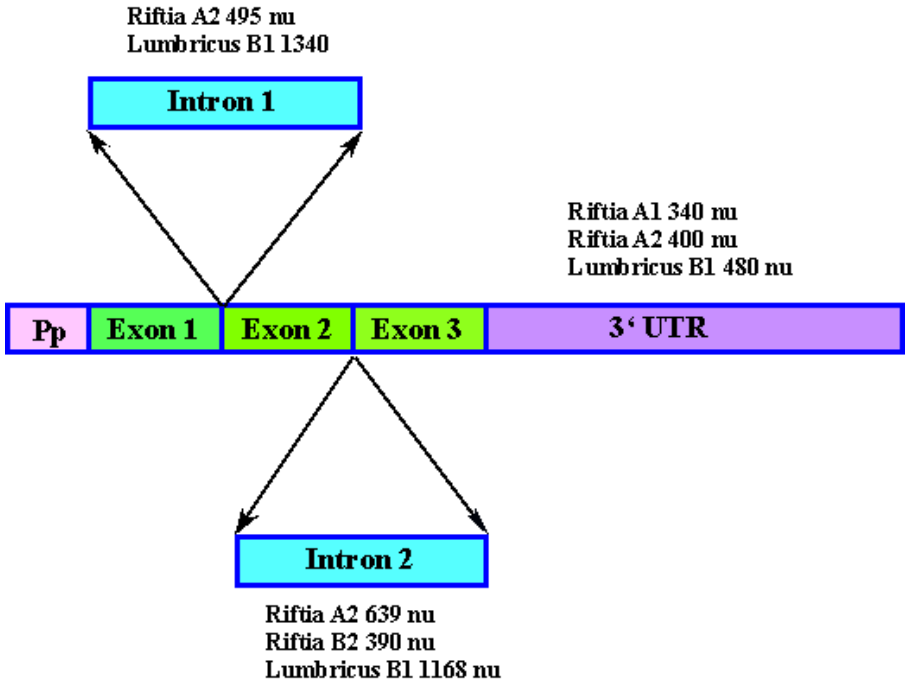


Figure 38. Structure du gène codant les globines A2 et B2 de Riftia, avec la séquence d'adressage Pp, la succession des exons et des introns et la partie 3'UTR. nu : nucléotides

• Les parties transcrites non traduites en 3' (3'UTR) des globines de *Riftia pachyptila* sont longues (340 à 400 nucléotides ; fig. 37 et annexe 3) par rapport à celles des vertébrés (entre 95 et 132) et plus petites que celles des myoglobines de vertébrés (jusqu'à 548 nucléotides pour la myoglobine du phoque). Elles sont presque du même ordre de grandeur que la partie 3'UTR de la globine B1 de *Lumbricus terrestris* (480). Il n'est pas impossible que la taille de la région 3'UTR soit corrélée à la présence d'éléments régulateurs de la transcription, à relier avec le caractère extracellulaire de ces protéines.

7. Conclusion

La famille multigénique des hémoglobines extracellulaires du vestimentifère *Riftia pachyptila* présente la même organisation que les familles multigéniques codant les hémoglobines d'autres espèces d'annélides polychètes et oligochètes.

Les capacités d'hétéro-polymérisation par des ponts disulfures inter-chaînes de ces globines varient selon deux grandes lignées évolutives : la lignée des polychètes et des oligochètes dont les hémoglobines sont constituées de trimères de globines, la sous-unité A1 restant monomérique, et la lignée des achètes, des vestimentifères et des pogonophores dont les hémoglobines sont constituées principalement de dimères de globines, la sous-unité A2 restant monomérique. En plus de l'organisation de la famille multigénique (fig.38), les structures quaternaires en bicouches hexagonales et la structure des gènes de globines (fig.39) confortent les arguments phylogénétiques et ontogéniques qui suggèrent que les vestimentifères et les pogonophores sont des annélides au même titre que les polychètes et les oligochètes, mais qu'ils pourraient dériver de ces derniers. Le scénario évolutif le plus parcimonieux suggère que toutes ces globines dérivent d'une même globine ancestrale, et que la première duplication a probablement permis la première hétéro-polymérisation. Cependant, on peut également penser que la capacité d'homo-polymérisation existait initialement. Cela pose l'intéressante question de la fonctionnalité intrinsèque de la protéine polymérique par rapport à la protéine monomérique : la capacité de polymérisation permet-elle d'acquérir de nouvelles fonctions (radicalement différentes de la fonction originale) ou plutôt d'apporter des niveaux de régulation plus fins et plus performants que la fonction originale comme par exemple la coopérativité ? Cette forte capacité de polymérisation aurait aussi pu permettre à la lignée extracellulaire d'hémoglobine d'émerger en conservant la même fonction que la lignée intracellulaire : Fushitani et al. (1988) ont proposé que l'acquisition de séquence d'adressage serait apparue dans la lignée des globines extracellulaires, elle-même dérivant de la lignée des hémoglobines intracellulaires.

En matière d'évolution moléculaire, la duplication des gènes permet la diversification et l'innovation au sein du génome. Deux sous-familles de globines des annélides des milieux riches en sulfures, A2 et B2, possèdent des cystéines libres capables de lier réversiblement le sulfure d'hydrogène. Leurs paralogues respectifs A1 et B1 n'en possèdent pas. Les hémoglobines des annélides vivant dans des milieux oxydés (sans sulfures) ne possèdent pas de telles cystéines, et ce, sur aucune de leurs globines.

Ces observations, les fonctionnalités et la description de la famille multigénique des hémoglobines de *Riftia pachyptila* nous ont permis d'initier une réflexion et un travail d'identification des mécanismes évolutifs permettant d'expliquer le maintien ou l'absence de cystéines libres, corrélé à la capacité ou non de lier et de transporter le sulfure d'hydrogène. La suite de cette thèse s'intéresse à l'évolution de la fonction de transport du sulfure d'hydrogène au sein de la famille multigénique des hémoglobines de *Riftia pachyptila* et plus généralement des vestimentifères et des pogonophores (chapitre 4), puis à l'évolution de cette fonction au sein des annélides (chapitre 5).

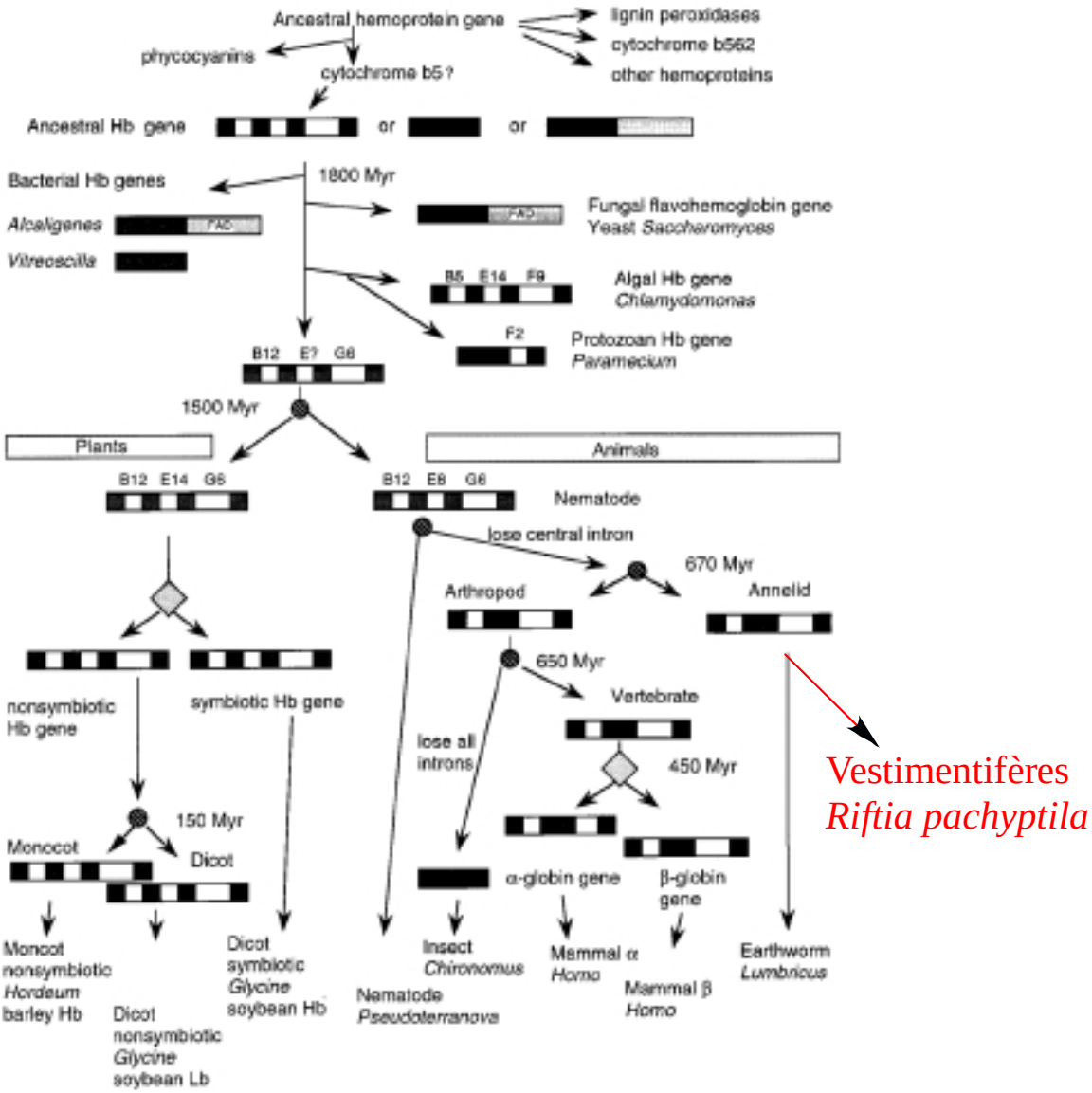


Figure 39. Evolution des gènes codant pour l'hémoglobine : position, perte et gains des introns. Schéma modifié (en rouge) tiré de (Hardison 1996)

Chapitre 4 : Premier article

Evolution de la fonction de fixation des sulfures dans la famille multigénique de globines chez le ver tubicole *Riftia pachyptila* des sources hydrothermales profondes

Article paru dans **Molecular Biology and Evolution**

Vol 19 n°9. Pp 1421-1433. (2002)

Chapitre 4: Premier article

Evolution de la fonction de fixation des sulfures dans la famille multigénique de globines chez le ver tubicole *Riftia pachyptila* des sources hydrothermal profondes.

Riftia pachyptila a la capacité de transporter le sulfure d'hydrogène en fixant ce dernier sur les cystéines libres (non impliquées dans les ponts dissulfures intra et inter-chaînes) associées à son hémoglobine extracellulaire. D'autres annélides, symbiotiques ou non, vivant en milieux réduits, possèdent le même type d'hémoglobine et la même capacité de fixation de H_2S. Chez les annélides symbiotiques comme *Riftia pachyptila*, la liaison du sulfure d'hydrogène à destination des bactéries chimiosynthétiques situées dans le trophosome est assimilée à une fonction de transport de H_2S. Par contre, la capacité de liaison du sulfure d'hydrogène chez les annélides non symbiotiques est assimilée à une fonction de détoxication de H_2S. Parce que les annélides non symbiotiques précèdent, d'un point de vue évolutif, les annélides vivant en symbiose, la fonction de transport de H_2S serait une apomorphie par rapport à la fonction de détoxication de H_2S.

La description de la famille multigénique des globines extracellulaires de *Riftia pachyptila* montre une organisation conforme à celle des autres annélides. Les six ADNc séquencés confirment que cette famille, comme celle de l'oligochète *Lumbricus terrestris*, du polychète *Tylorrhynchus heterochaetus* et du vestimentifère *Lamellibrachia sp*, est divisée en deux familles, A et B. Chacune d'elle est subdivisée en deux sous-familles A1, A2 et B1, B2.

Nous avons cherché à comprendre comment la fonction de fixation des sulfures par ces hémoglobines a évolué chez les annélides des milieux réduits, en prenant la famille multigénique de *Riftia pachyptila* comme modèle. Comment cette fonction s'est-elle maintenue ? Est-elle apparue au cours des duplications ou était-elle présente avant même la diversification de la famille multigénique ? Quel est le jeu des pressions sélectives qui peut expliquer la situation actuelle ?

La position des cystéines libres impliquées dans la fixation et le transport des sulfures est conservée dans toutes les globines orthologues A2 et B2 des annélides vivant en présence de sulfures. Nous avons respectivement appelé Cys+1 et Cys+11 ces cystéines conservées des globines B2 et A2. Bien que **la démonstration de la fixation de H_2S n'ait pas été réalisée pour les hémoglobines des annélides qui possèdent de telles cystéines libres chez A2 et B2 à l'instar des hémoglobines de *Riftia pachyptila*, nous avons admis par analogie**

fonctionnelle et par la présence de H_2S dans leur environnements respectifs, que ces cystéines sont également impliquées dans la fixation de H_2S.

Nous avons également considéré les globines des chlorocruorines des annélides Sabellidés qui possèdent ces mêmes cystéines libres comme des homologues des chaînes de globines des hémoglobines des autres annélides (chapitre 1. I). La seule différence est portée par l'hème qui est légèrement modifié (substitution d'un groupement vinyle par un groupement formyle). Par ailleurs, la structure du gène de la globine du polychète *Eudistylia vancouverii* dont le pigment respiratoire est une chlorocruorine, est identique à celle du gène des globines d'annélides possédant des hémoglobines (même position des introns : Dewilde et al. 2001). Ces mêmes résidus, Cys+1 et Cys+11 sont absents chez toutes les globines des annélides vivants en milieux oxydés.

Une étude détaillée des séquences montre (1) l'absence de ces résidus dans les sous-familles paralogues des globines A1 et B1 des annélides des milieux réduits, (2) des vitesses évolutives plus élevées des globines A1 et B1 par rapport à celles des globines A2 et B2, (3) la conservation de la structure 2D du domaine qui flanque les deux cystéines libres (SBD, Sulfide Binding Domain) entre les globines paralogues A2 et B2 chez les annélides des milieux réduits, (4) la conservation de cette même structure mais sans les cystéines chez les annélides des milieux oxydés.

Ces données suggèrent que les fortes concentrations en sulfures opèrent (ou participent pour une grande partie) en tant que pression de sélection pour le maintien des cystéines libres. Nous proposons un scénario évolutif de la fixation des sulfures par des cystéines libres des hémoglobines extracellulaires des annélides. Nous avons superposé sur la phylogénie moléculaire des globines de *Riftia pachyptila* les évènements putatifs de perte de cystéines tout au long de l'émergence de la famille. L'acquisition (le gain) indépendante de tels résidus dans les lignées A2 et B2 nous semble hautement improbable eu égard aux vitesses évolutives estimées, à la conservation du domaine SBD, à l'étude détaillée des séquences, à l'ensemble des données paléo-géologiques et écologiques. L'exemple du transport des sulfures par les hémoglobines de mollusques des milieux réduits par des seules interactions électrostatiques avec l'hème, renforce l'idée que le transport des sulfures par les hémoglobines extracellulaires chez les annélides par une cystéine est une caractéristique de ce phylum. L'ancêtre annélidien non symbiotique était probablement adapté à la vie dans les sulfures avec de telles cystéines qui exerçaient très probablement une fonction de détoxication en chélatant H_2S. La capacité de fixation chez les annélides symbiotiques aurait évolué vers une fonction de transport de H_2S. Cependant, d'autres séquences de polychètes phylogénétiquement éloignés restent à

acquérir pour entériner ce modèle. Les analyses et les interprétations des données moléculaires sont présentées en détail dans l'article qui suit.

Evolution of the sulfide-binding function within the globin multigenic family of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*.

Xavier Bailly[1], Didier Jollivet[1], Stephano Vanin[2], Jean Deutsch[3], Franck Za1[1], François Lallier[1] and André Toulmond[1]

Laboratoire Ecophysiologie, Station Biologique de Roscoff

[1] Station Biologique de Roscoff, UPR 9042 CNRS-UPMC-INSU, Laboratoire Ecophysiologie, BP. 74, Place Georges Teissier, 29682 Roscoff cedex, France.

[2] Universita di Padova, Department of Biology, 3 Via G. Colombo, 35131 Padova, Italy.

[3] Biologie du Développement, UMR 7622, CNRS et Université P & M CURIE (Paris 6) 9 quai St-BERNARD, case 241, 75252 PARIS Cedex 05 France .

Address for correspondence and reprints: Xavier Bailly, Station Biologique de Roscoff, UPR 9042 CNRS-UPMC-INSU, Laboratoire Ecophysiologie, BP. 74, Place Georges Teissier, 29682 Roscoff cedex, France. phone: 33 2 98 29 23 67, Fax: 33 2 98 29 23 24, email: bailly@sb-roscoff.fr

Key words: Hexagonal bilayer hemoglobin, duplication, selection, H2S, sulfide-binding domain, *Riftia pachyptila*.

Running head: Evolution of a multigenic globin family in annelids

Abstract

The giant extracellular hexagonal bilayer hemoglobin (HBL-Hb) of the deep-sea hydrothermal-vent tube worm *Riftia pachyptila* is able to transport simultaneously O_2 and H_2S in the blood from the gills to a specific organ: the trophosome that harbors sulfide-oxidizing endosymbionts. This vascular HBL-Hb is made of 144 globins from which four globin types (A1, A2, B1 and B2) co-evolve. The hydrogen sulfide is bound at a specific location (not on the heme site) onto two of these globin types. In order to understand how such a function emerged and evolved in vestimentiferans and other related annelids, six partial cDNAs corresponding to the 6 globins known to compose the multigenic family of *R. pachyptila* have been identified and sequenced. These partial sequences (c.a. 120 a.a. i.e. 80% of the entire protein) were used to reconstruct molecular phylogenies in order to trace duplication events that have led to the family organization of these globins and to locate the position of the free cysteine residues known to bind hydrogen sulfide. From these sequences, only two free

cysteine residues occur at positions Cys + 1 (i.e. one a.a. from the well-conserved distal histidine) and Cys + 11 (i.e. 11 a.a. from the same histidine) in globins B2 and A2, respectively. These two positions are well-conserved in annelids, vestimentiferans and pogonophorans which live in sulfidic environments. The structural comparison of the hydrophobic environment that surrounds these cysteine residues (the Sulfide Binding Domain) using HCA-plots, together with the cysteine positions in paralogous strains, suggest that the sulfide-binding function might have emerged prior to the annelid radiation in order to detoxify this toxic compound. Moreover, globin evolutionary rates are highly different between paralogous strains. This suggests that either the two globin subfamilies involved in the sulfide-binding function (A2 and B2) have evolved under strong directional selective constraints (negative selection) and/or that the two other globins (A1 and B1) have accumulated more substitutions through positive selection or have evolved neutrally after a relaxation of selection pressures. A likely scenario on the evolution of this multigenic family is proposed and discussed from this dataset.

Introduction

From the earliest studies of evolutionary processes at a molecular level, hemoglobin has been one of the favourite molecule to model the evolution of multigenic families (Rossi-Fanelli,Cavallini and De Marco 1955; Zuckerkandl, Jones and Pauling 1960). Ingram (1961) stressed the process of gene duplication to explain human hemoglobin subunit's emergence and different levels of selection to maintain such proteins through the course of evolution. Numerous amino-acid and nucleotide-based phylogenies of vertebrate globins have been carried out to reconstruct the most appropriate scenario of the evolution of the paralogous genes encoding these molecules (Goodman et al. 1971; Goodman, Moore and Matsuda 1975). Invertebrate globins has also been investigated. Their progressive identification and characterization at a gene level allowed several authors to relate the huge hemoglobin diversity found in invertebrates to their evolutionary history (Goodman et al. 1988; Hardison 1998).

The present description of the complete multigenic globin family of the vestimentiferan *Riftia pachyptila* Jones (1981) provides an additional original model of evolution in which globins exhibit different ligant specificities, raising questions about the sub-functionalization of duplicated genes. Living in a sulfide-rich environment at hydrothermal vents along the East Pacific Rise (EPR) at a depth of 2500 meters, the tube-worm *R. pachyptila* possesses hemoglobins that are able to bind simultaneously and reversibly O_2 and a highly toxic

molecule, hydrogen sulfide (H_2S) (Arp, Childress and Vetter 1987). Endosymbiotic sulfo-oxidizing bacteria harboured within a specific organ of the worm, the trophosome, use H_2S to synthetise organic compounds which are used by the host (Childress and Fischer 1992). Adaptation to H_2S is of prime importance for this worm which has to cope with its nutritional requirements and the toxicity of this reduced compound, this latter inhibiting the mitochondrial oxidative chain reaction and the transport of oxygen (Nicholls 1975). Life in sulfidic habitats has been widely studied and several kinds of adaptive mechanisms have been selected in organisms to detoxify H_2S (Vetter and Powell 1991; Vismann 1991; Grieshaber and Volkel 1998). One of them is the use of extracellular hemoglobins via the occurrence of free cysteine residues (Zal et al. 1998). Hemoglobins of *R. pachyptila* are extracellular, two are vascular and one is coelomic. One of the two vascular hemoglobins possesses a hexagonal-bilayer quaternary structure (HBL-Hb) (Terwilliger, Terwilliger and Schabtach 1980) which is only found in the Annelida and Vestimentifera (Weber and Vinogradov 2001). This complex multimeric hemoglobin is made of 144 globin chains and structural non-globin chains, the linkers, which enable such a typical shape. Six different globin subunits and their associated molecular weight and free cysteine residues (i.e. cysteine residues not involved in intra- or inter-chain dissulfide bridges) have been previously described from these hemoglobins by mass spectrometry and amino acid sequencing of the A2 chain (Zal et al. 1996a; 1996b; 1997b). The free cysteine residues only occur on two globin strains, one per globin. As suspected by Suzuki et al. (1989), these free cysteine residues are involved in the binding of H_2S (Zal et al. 1997b; 1998). Free cysteine residues were also found in other globins of worms living in transient or permanent sulfide-rich or polluted environments such as the symbiotic vestimentiferan *Lamellibrachia sp.* (Takagi et al. 1991), the symbiotic pogonophoran *Oligobrachia mashikoi* (Yuasa, Green and Takagi 1996) and the non-symbiotic polychaetes *Sabella spallanzanii* (Pallavicini et al. 2001) and *Sabellastarte indica* (Suzuki, Hirao and Vinogradov 1995). One might consider that these species have all the ability to bind sulfides via their free cysteine residues by analogy to the sulfide-binding function of both the vestimentiferans *Lamellibrachia sp.* and *R. pachyptila* simply because such a binding process has been demonstrated in other polychaetes such as the lugworm, *Arenicola marina* (Zal, Gotoh and Toulmond 1999). This paper addresses a presentation of a new and complete multigenic family of extracellular globins in *R. pachyptila* and a possible scenario of duplication events that explains the evolution of such a peculiar respiratory pigment. It also focusses on the sulfide-binding function and its possible role in the evolution of extracellular globins in annelids, vestimentiferans and pogonophorans living in sulfidic environments.

HBL-Hbs carry an original function involved in the binding and transport of hydrogen sulfide which indeed raises two fundamental questions about (1) how selection acts to maintain free cysteine residues solely in two globin subfamilies after several duplication events, at the multigenic family level and (2) what is the likely evolutionary history of such a detoxifying mechanism since its appearance in annelids.

Materials and Methods

Biological Material

Juvenile specimens of the hydrothermal-vent tubeworm *R. pachyptila* (3 to 5 centimeters length) were collected twice at one single vent site from the ridge segment 9°50'N on the East Pacific Rise (Riftia field: 9°50.75'N, 104°17.57'W) at a depth of about 2500 m during the French oceanographic cruise HOT 96 and the American cruise LARVE99.The worms were sampled using the telemanipulated arm of the submersibles 'Nautile' and 'Alvin', brought back alive to the surface inside an insulated basket and immediatly frozen and stored in liquid nitrogen after their recovery on board.

Characterization of extracellular globin genes

Globin-specific primer design - All degenerate forward and reverse primers were designed according to the amino acid sequences of the four extracellular globin sub-families that form the hexagonal bilayer hemoglobin of *Lamellibrachia sp.* (i.e. sub-families A1 (Takagi et al. 1991), A2 (access n°: P15469), B1 and B2 (Takagi et al. 1991)). Primer designs were performed using the Oligo 4.0 software according to the following criteria: (1) the corresponding amino acid domains used must be close to either the NH_2-terminal end (forward primer) or the COOH-terminal end (reverse primer: legend of Fig. 1), (2) they must contain a large proportion of two-fold or three-fold degenerated amino acids, and (3) they must be specific of a single globin sub-family.

Total RNA extraction and cDNA synthesis - Entire juveniles of *R. pachyptila* were crushed in liquid nitrogen. Total RNAs were then recovered and extracted using the RNAble® buffer (Eurobio). The mRNA purification was performed from the total RNAs using the oligodT resin of the mRNA Purification Kit ® (Pharmacia Biotech). Reverse transcription was initiated with an anchored oligodT CTC CTC TCC TCT CCT CTT(16) primer.

Globin amplification and sequencing - Each partial globin cDNA was amplified by PCR using a set of degenerate primers on a Perkin Elmer GenAmp PCR System 2400®. PCR steps were performed identically for each globin sub-family amplification, with the same annealing temperature (50°C). PCR conditions were as follows: initial denaturation at 96°C for 5 min , 35 cycles consisting of 96°C for 50 s , 50°C for 50 s, and 72°C for 50 s.The reaction was completed by an elongation step of 10 min at 72°C. Amplifications were carried out in 25 µl reaction mixtures containing 10-50 ng cDNA target , 50-100 ng of each degenerate primer, 200 µM dNTPs, 2.5 mM $MgCl_2$ and 1 unit of Taq DNA polymerase (Promega). PCR products were visualized on a 1% agarose gel under U.V. Gel slices containing DNA fragments of the expected size were collected and subsequently purified onto Ultrafree®-DA. All PCR products were then cloned using a TOPO-TA Cloning® kit (Invitrogen). Purified plasmids containing the globin insert were used in a dye-primer cycle sequencing reaction using either the labeled Texas Red universal primer T7 (5'- GTA ATA CGA CTC ACT ATA GGG C -3') or the M13 reverse (5'-GGA AAC AGC TAT GAC CAT G -3') and the Thermo Sequenase™ pre-mixed cycle sequencing kit from Amersham™. PCR-products were subsequently run on a Vistra automated DNA Sequencer 725.

Molecular phylogenetic analysis

Analyses were performed on different sets of data: (1) paralogous cDNA sequences of the globin family of *R. pachyptila* and (2) orthologous amino acid sequences of globins from *R. pachyptila* and closely-related annelids sharing similar HBL-Hb extracellular respiratory pigments. These globin sequences belong to *Lumbricus terrestris* (access n°: A29134, A28151, B28151, C28151) in Shishikura et al. (1987) and Fushitani, Matsuura and Riggs (1988), *Tylorrhynchus heterochaetus* (access n°: P02219, P02220, P09966, P13578) in Suzuki and Gotoh (1986), *Lamellibrachia sp.* (access n°: S08284) in Suzuki, Takagi and Ohta (1990), *Oligobrachia mashikoi* (access n°: S72251, S72252, S72253) in Yuasa, Green and Takagi (1996) and *Sabella spallanzanii* (access n°: CAC37411 and CAC37412) in Negrisolo et al. (2001). DNA or amino acid sequences were aligned with the Clustal X software (Thompson et al. 1997) and using the hidden Markov model procedures and the SAM system (Karplus et al. 1997) (http://www.cse.ucsc.edu/research/compbio/) in order to assess the congruence of alignments for both paralogous and orthologous data sets.

Informative sites - The more informative and conserved blocks of amino acids were selected for molecular phylogenetic analysis from a multiple alignment of globin sequences using the Gblock software (Castresana 2000). Final conserved blocks of contiguous amino acids were obtained applying the following settings in option b ("blocks parameters"): minimum number of sequences for a conserved position: 8, minimum number of sequences for a flank position: 12, maximum number of contiguous nonconserved positions: 8, minimum length of an initial block: 15, minimum length of a block after gap cleaning: 10.

Molecular phylogeny - Molecular phylogenetic trees were constructed using Neighbor-Joining (NJ), Maximum Parsimony (MP) and Maximum Likelihood (ML) methods. NJ, MP and ML were respectively computed with the Phylo_win package (Galtier, Gouy and Gautier 1996), PAUP version 3.1.1 (Swofford 1993) and Puzzle version 4.0 (Strimmer and von Haeseler 1996). Phylogenetic distances were computed from observed levels of divergence and from that expected under the Dayhoff's PAM matrix. MP trees were obtained with an heuristic search using the tree bisection-reconnection (TBR) branch swapping option, stepwise addition (closest taxa addition) and the collapse option for zero length branches. ML analysis was performed according to the Jones, Taylor and Thornton (JTT) model of amino acid substitutions (Jones, Taylor and Thornton 1992), with and without a gamma distribution of substitution rates among sites, and using the quartet puzzling method. Bootstrap proportions were obtained from 1000 re-sampling sequence alignments for NJ and parsimony trees and from 100 000 puzzling steps in the particular case of the likelihood method. The "duplicate gene rooting" procedure (Donoghue and Mathews 1998), which was originally proposed as an alternative to root the universal tree of life for which outgroups are not available (Iwabe et al. 1989), was used to root our paralogous globin trees.

Evolutionary rates - To test whether sub-families of vestimentiferan globins may evolve at different evolutionary rates, genetic distances between *R. pachyptila* and *Lamellibrachia sp.* orthologous amino acid sequences of each sub-family, namely A1, A2, B1 and B2 were calculated according to the Jones, Taylor and Thornton (1992) matrix using the Maximum Likelihood procedure of the Puzzle software. In this particular case, the four genetic distances obtained between the two taxa relate directly to the evolutionary rate because the evolutionary time is identical between the four paralogous genes. The occurrence of a global molecular clock was tested on three orthologous sets of globins, namely A1, A2 and B2, using the relative rate test comparing substitution rates between sequences, according

to the absolute difference |d1- d2| of Sarich & Wilson (1973) where d1 and d2 represents the Maximum Likelihood distances of species 1 and species 2 to a third more divergent reference species.

2-D Cys structural domains

A Hydrophobic Cluster Analysis (HCA plots) was performed to predict the secondary structure of primary globin amino acid sequences to assess the molecular environment of free cysteine residues that are involved in the sulfide-binding function. These plots were obtained using the DrawHCA software (Callebaut et al. 1997) (http://www.lmcp.jussieu.fr/~soyer/www-hca/hca-form.html).

Results

Identification and characterization of the complete globin multigenic family of R. pachyptila.

Six specific globin sequences (accession numbers: AJ439732, AJ439733, AJ439734, AJ439735, AJ439736, AJ439737) were identified after cloning and sequencing, in agreement with the number of subunits previously described by mass-spectrometry (Zal et al. 1996a, b). The six globins were then aligned and compared with the four extracellular globins of *Lamellibrachia sp*. designated A1, A2, B1 and B2 following to the nomenclature of Suzuki, Takagi and Ohta (1993) and with the globin sequences of *Oligobrachia mashikoi, Sabella spallanzanii, Lumbricus terrestris and Tylorrhynchus heterochaetus* (Fig. 1). Among the six *R. pachyptila* sequences, three were unambiguously identified as orthologous to the A1, A2 and B2 globins of *Lamellibrachia sp.* The last three *R. pachyptila* sequences were much more variable. However, they share residues with the B family (Fig. 1): a serine (S) at position 6, present in a majority of B globins; and an asparagine (N) at position 66, which is absent from the A globins. In addition they shared two indels with the B globin: an insertion of one residue at position 89, and a deletion of two residues at positions 113-114. Furthermore, a 3-residues motif PQV at position 108-110, specific of the B globins, was well conserved. These characters allowed us to allocate these three sequences to the B family without ambiguity. Furthermore, we propose that these three latter globins fall within the B1 group on the following basis: B1a, B1b and B1c shared a 4-residues motif, VNVA, at position 40-43 and a methionine (M) at position 99 with the globin B1 of *Lamellibrachia sp*. B1a and B1b shared also a valine (V) at position 102 with *Lamellibrachia sp* B1 (Fig. 1).

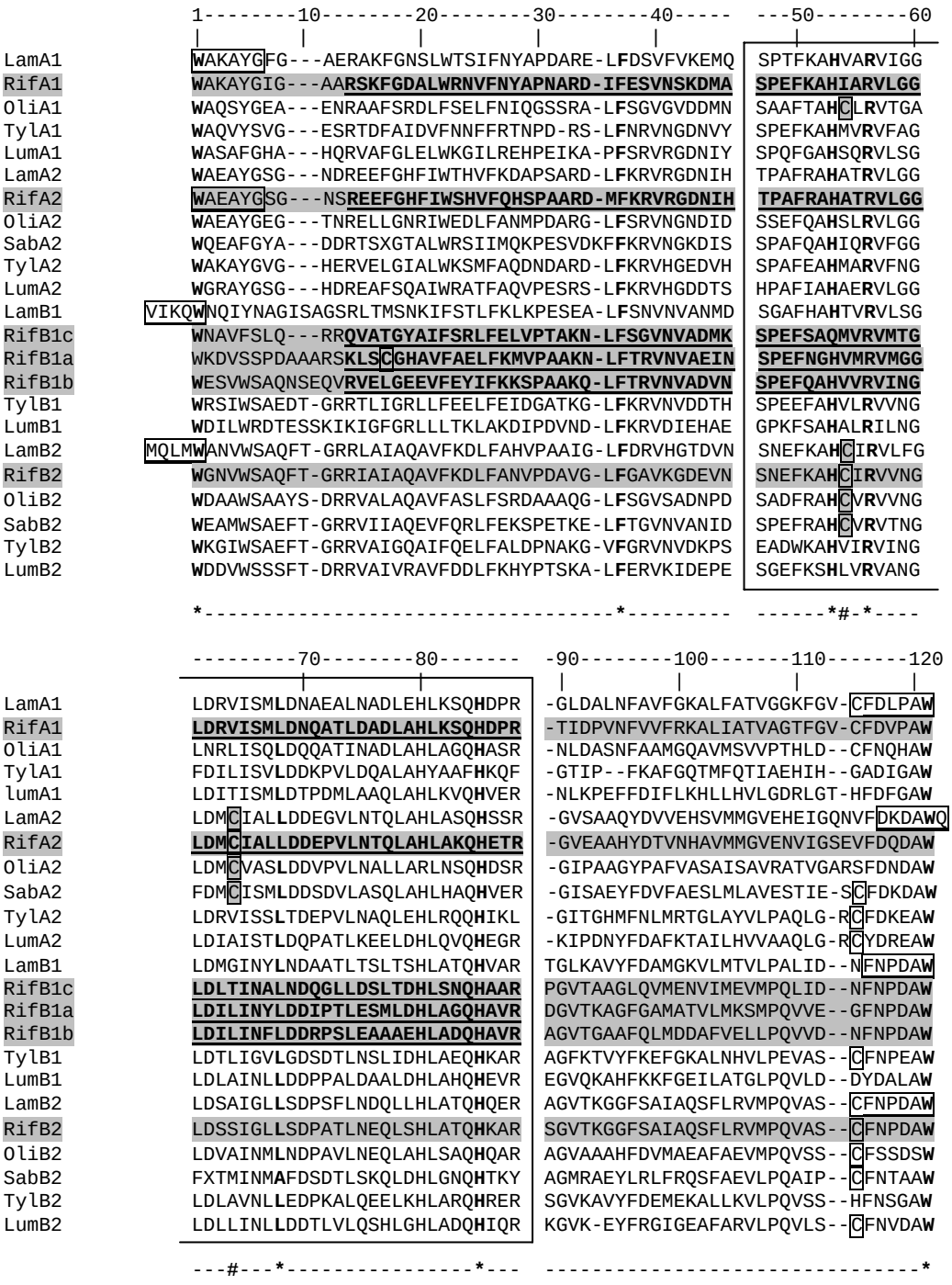


FIG 1. Multiple alignment of partial amino-acid globin sequences of *Sabella spallanzanii* (Sab), *Oligobrachia mashikoi* (Oli), *Lamellibrachia sp.* (Lam), *Tylorrhynchus heterochaetus* (Tyl), *Lumbricus terrestris terrestris* (Lum) and *R. pachyptila*(Rif). ***Amino acids*** in Lam sequences correspond to the portions used to design the following specific sub-family degenerated primers: A1 forward: 5'-TGG GCN RAR GCN TAY GG-3' ; A1 reverse: 5'-CCA NGC NGG NAR RTC RAA-3' ; A2 forward: 5'-TGG GCN RAR GCN TAY GG-3' ; A2 reverse: 5'-YTG CCA NGC RTC YTK RTC-3' ; B1 forward: 5'-GTN AYH AAR CAG TGG VA-3' ; B1 reverse: 5'-CCA NGC RTC NGG NTT RAA-3' ; B2 forward: 5'-RAT GCA GYT NAT GTG GG-3' ; B2 reverse: 5'-NGC RTC NGG RTT RAA RCA-3'. The new globin sequences of *R. pachyptila* are shaded. (*****) indicates the position of the well-conserved amino acids in extracellular globins of annelids according to Vinogradov et al.

(1993), (#) indicates the location of the free cysteine residues (**C**) involved in the sulfide-binding function and **C** indicates the cysteine residues involved in the globin dissulfide bridge intra or inter-globin chain. Bold and underlined *R. pachyptila* globin sequences correspond to the restricted portion of the contiguous most conserved amino acids obtained using the Gblock software between the six extracellular globins of *R. pachyptila*. The boxed multiple-alignment portion corresponds to the amino acid portion of the second exon selected for the phylogeny reconstruction according to Gblock.

The high amino-acid identity between *R. pachyptila* and *Lamellibrachia sp.* globin subfamilies (72%, 81%, 51%, and 82% for the A1, A2, B1 and B2 globin chains respectively) and the occurrence of well-conserved amino acid motifs typifying the extracellular HBL-Hb of annelids (Vinogradov et al. 1993), strongly sustain the orthological relationship of *R. pachyptila* globin sub-families with other annelid globin sub-families so far described. The well-conserved distal histidine at position 53 (Fig. 1), known to be involved in fine-tuning the ligant affinities of hemoglobin (Olson et al. 1988), is substituted by a glutamine in the B1c globin.

Cysteine residues in R. pachyptila *globins*

The *R. pachyptila* globin sequences also display free cysteine residues (i.e. not involved in dissulfide bridges within or between the globin polypeptide chains) whose positions correspond to those previously described in *Lamellibrachia sp.* (Fig. 1). These free cysteine residues are only located in the A2 and B2 sub-families of *R. pachyptila*. These cysteines occur at position one after the distal histidine in B2 and at position eleven after the homologous histidine in A2. Cysteine residue in a similar position are also found in *Lamellibrachia sp*., *Oligobrachia mashikoi* and *Sabella spallanzanii* A2 and B2 globins, species living in sulfide-rich (or sporadically-rich) environments, whereas no such free cysteines are found in A2 and B2 globin of *Lumbricus terrestris* and *Tylorrhynchus heterochaetus*, two species living in free-H_2S environment. Another free cysteine residue is also found in the globin A1 sequence of *Oligobrachia mashikoi* (Fig. 1). We called Cys + 1 and Cys + 11 these two free cysteine residues. Other cysteine residues are also found in *R. pachyptila* globin sequences. However, these residues are always involved in dissulfide bridges in HBL-Hb of both *R. pachyptila* and the other annelids so far investigated. Cysteine residues involved in intra-chain dissulfide bridges were not detected due to the incomplete sequencing of our clones at both COOH and NH2 ending sides. However, cysteine residues involved in dissulfide bridges between globins A1 and B2 occur at positions 115 and 17, respectively.

Phylogenetic reconstruction of the R. pachyptila globin multigenic family

Because most sites of the second and the third codon positions were saturated between the six paralogous sequences (data not shown), a computed selection of conserved blocks of amino acids obtained by the elimination of poorly aligned positions and divergent regions (Gblock software) was used to generate a restricted alignment of 74 amino acids (Fig. 1). Tree were constructed subsequently using NJ, MP and ML methods and compared. All methods gave identical topologies, consequently we only present the ML consensus tree (Fig. 2).

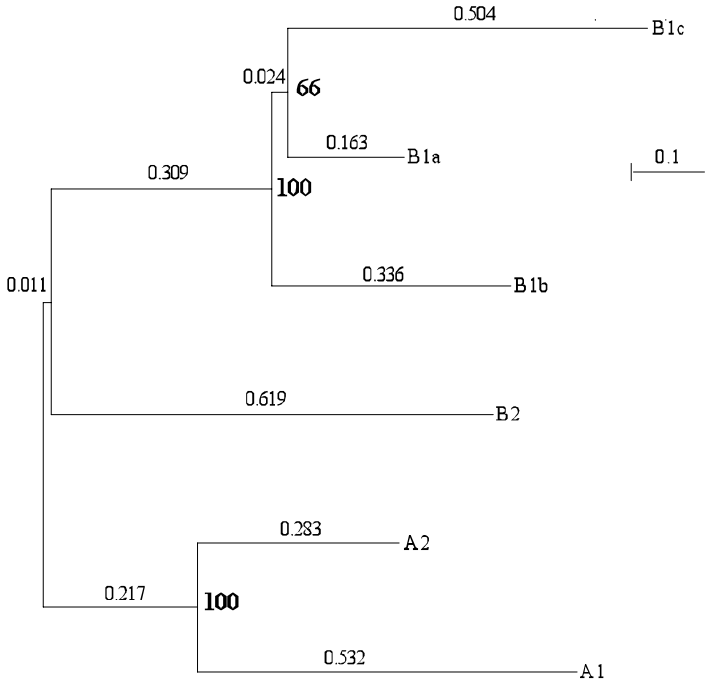


FIG 2. Maximum likelihood consensus tree obtained from the six *R. pachyptila* globins using 100 000 puzzling steps. Clade A was used as a root to refine the topology of clade B. Bold values at each node correspond to bootstrap confidence level and values above the branches to branch lengths.

Tree topologies clearly indicate that this multigenic family emerged from an initial duplication event and then led to four duplication events that produced the actual multigenic family. These duplications appear to be asymmetrical between families A and B and have led to the recent emergence of the three B1 globins

Globin molecular phylogeny of vestimentiferan, pogonophorans and closely-related polychaetes living in sulfidic environments

Globins from several additional closely-related annelid species were added to the *R. pachyptila* globin sequences to test the phylogenetical signal within each set of orthologous sequences, providing that these species possess at least the four main paralogous globin types, namely A1, A2, B1 and B2.

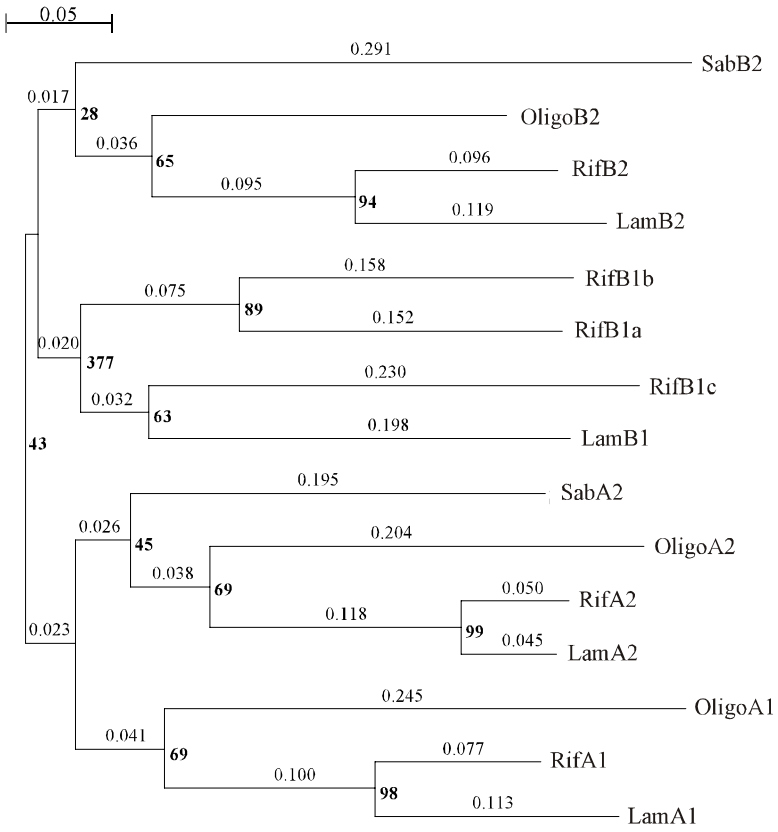


FIG 3. Neighbor-joining consensus tree obtained using the observed distances (branch length above the branches) estimated from the Gblock-selected portion of globin sequences and 1000 bootstrap replicates (values in bold). Rooting is performed according to the 'duplicate rooting procedure'. The first dichotomy separates the A and B families and the second the A1, A2, B1 and B2 sub-families. *Sabella spallanzanii* (Sab), *Oligobrachia mashikoi* (Oli), *Lamellibrachia sp.* (Lam) and *R. pachyptila* (Rif).

Following Gblock and similar tree reconstruction methods as previously described, we constructed a restricted alignment of the 42 most informative contiguous amino-acids that corresponds to a portion of the well-described second exon of the extracellular globin gene (see: boxed alignment in Fig. 1). Irrespective of the method used (NJ, MP and ML), data gave identical tree topologies (Fig. 3).

The ancestral duplications that have led to the emergence of the A and B globin families and then to the A1, A2, B1 and B2 sub-families are clearly identified, *i.e.* sequences from each orthologous set of genes are clustered together in a similar fashion (Fig. 3). Maximum

parsimony gave a slightly different result with two aberrant branching in the tree. To produce a coherent phylogeny with this method, we examined the alignment, character by character, and were able to detect and thus eliminate two homoplasic positions (*i.e.* positions 23 and 40) involved in the unexpected topology. This restored a tree topology similar to that obtained with the NJ and ML methods for the four paralogous sets of globins.

Such superimposable species clustering between the four sets of orthologous sequences led us to test the occurrence of a molecular clock in order to investigate whether our molecular dataset follows a nearly neutral rate of evolution. To test this hypothesis, genetic distances were calculated from pairwise combinations of globins from the same sub-family using Puzzle version 4 (Maximum Likelihood) software, the Jones, Taylor and Thornton matrix and 10 000 puzzling steps (Table 1).

Table 1.

ML estimates of the expected number of amino-acid substitutions per site (below the diagonal) and the observed distances (above the diagonal).

Globin	A2				B2				A1			
Species	Sab	Oli	Lam	Rif	Sab	Oli	Lam	Rif	Sab	Oli	Lam	Rif
Sab	-	0.44	0.40	0.40	-	0.49	0.54	0.52	-	-	-	-
Oli	0.75	-	0.37	0.37	0.79	-	0.38	0.36	-	-	0.46	0.42
Lam	0.60	0.54	-	0.10	1.15	0.58	-	0.21	-	0.61	-	0.19
Rif	0.64	0.56	0.22	-	0.85	0.41	0.20	-	-	0.72	0.22	-

Sab: *Sabella spallanzanii*; Oli: *Oligobrachia mashikoi*; Lam: *Lamellibrachia sp.*; Rif: *R. pachyptila*.

Then, a relative rate test was applied on the A1, A2 and B2 sub-families, but not for B1 for which no ancestral reference has been obtained (*i.e.* a third ancestral species). Results are presented in Table 2. The global mean of distance differences (d1 – d2) is 0.145 +/- 0.113. We used a t-test for paired comparisons in order to test the null hypothesis of identical evolutionary rates *i.e.* if the difference does not significantly depart from zero, then we could not reject the hypothesis of a molecular clock. The analysis indicated that taxa evolved at a same rate within each set of orthologous sequences at a threshold of 1% but not of 5% (ddl = 4, t = 3.85).

Table 2.

Relative-rate test values (r) obtained for the globin sub-families A1, A2 and B2.

Globin	((1, 2), 3)	d1	d2	r
B2	((Oli,Lam),Sab)	0.79	1.15	0.35
	((Oli,Rif),Sab)	0.79	0.85	0.06
	((Rif,Lam),Sab)	0.85	1.15	0.30
	((Rif,Lam),Oli)	0.41	0.58	0.17
Mean ± SD				0.22 ± 0.13
A2	((Oli,Lam),Sab)	0.75	0.60	0.15
	((Oli, Rif),Sab)	0.75	0.64	0.11
	((Rif,Lam),Sab)	0.64	0.60	0.04
	(Rif,Lam),Oli)	0.56	0.54	0.02
Mean ± SD				0.08 ± 0.06
B1	((Rif,Lam),Oli)	0.72	0.61	0.11

r corresponds to the absolute difference between d1 and d2 where d1 represents the maximum-likelihood distance estimated between the reference species (3) and the first species (1) and d2 the distance between the reference species and the second species (2). Sab: *Sabella spallanzanii*; Oli: *Oligobrachia mashikoi*; Lam: *Lamellibrachia sp.*; Rif: *Riftia.*

Evolutionary rates between paralogous globins

Genetic distances were recalculated over the whole globin amino-acid sequences between the vestimentiferans *R. pachyptila* and *Lamellibrachia sp.* for the four paralogous globin types (i.e. amino acids occurring between the conserved COOH- and NH_2-terminal tryptophans: Table 3). First, genetic distances differ between the four sub-families. Second, genetic distances obtained from pairwise comparisons of *R. pachyptila* and *Lamellibrachia sp.* globins from partial sequences generated with Gblock software (74 a.a. and 42 a.a., respectively) produced the same trend. This result clearly indicates that A2 and B2, which possess the free cysteine residues at positions Cys + 11 and Cys + 1, display a lower evolutionary rate as opposed to the A1 and B1 globins.

Table 3.

Maximum likelihood estimates of the expected number of amino-acid substitutions per site (genetic distance) between orthologous sequences of *R. pachyptila* and *Lamellibrachia sp* for each globin sub-families.

	Genetic Distance (Jones, Taylor and Thortorn matrix)
d A1	0.38
d A2 (Cys + 11)	0.22
d B1	0.79
d B2 (Cys + 1)	0.20

Hydrophobic Cluster Analysis of R. pachyptila *globins*

The secondary structure of globins in *R. pachyptila, Lamellibrachia sp., Oligobrachia mashikoi* and *Sabella Spallanzanii* were deduced from HCA plots and compared according to the tree topology of figure 4.

The analysis focused on the restricted sulfide-binding domain (SBD) that flanks the free cysteine residues, i.e. the specific local HCA shape flanking the Cys + 1 and the Cys + 11. It is clear that two different hydrophobic motifs are conserved over annelid species (Fig. 4). Another interesting finding consists of a B2 structure-like in *Oligobrachia mashikoi* A1 with a free cysteine residue and a A2 and B2 structure-like but with no cysteine residue in *Lamellibrachia sp* B1. Hence, there is either a typical SBD structure or a degenerated fingerprint of such a structure in almost all the HBL-Hb globins. This motif persistence may indicate that the SBD is ancestral.

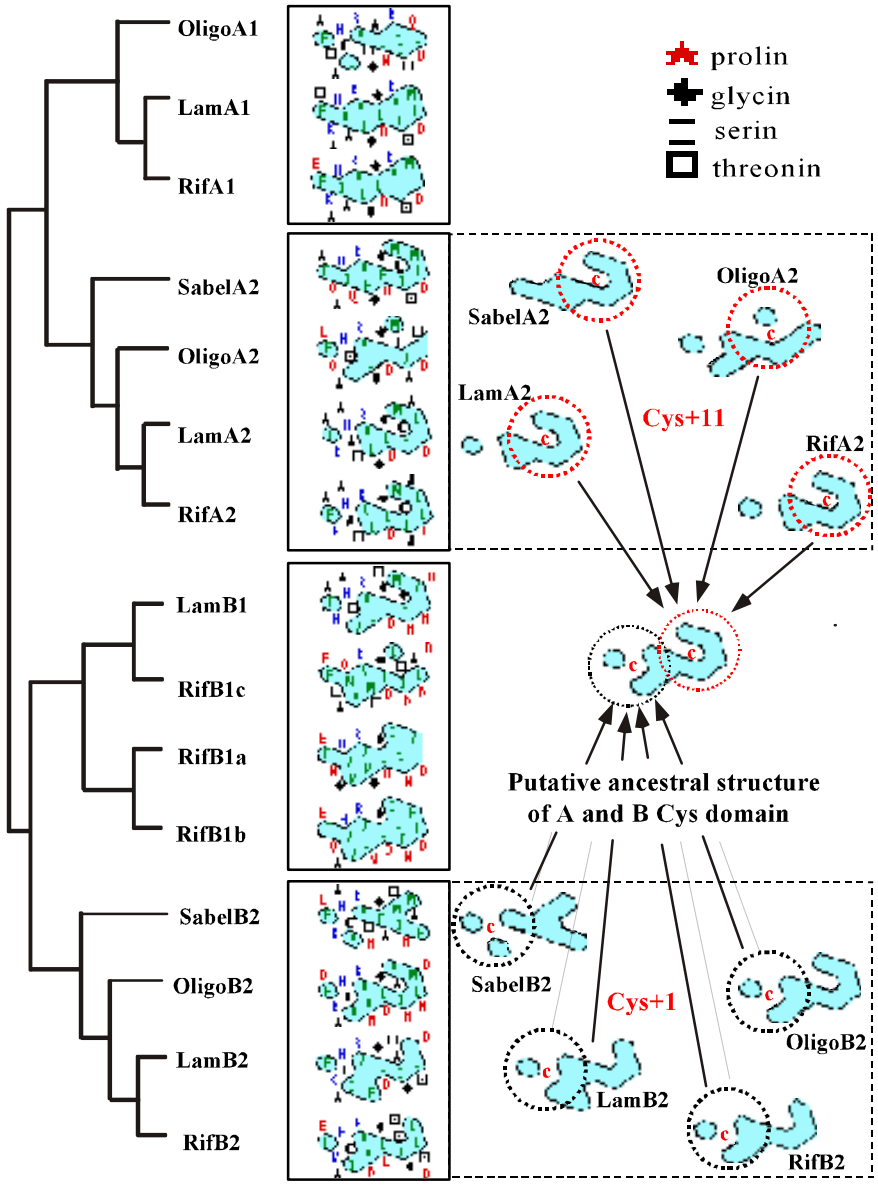


FIG 4. Neighbor-joining tree (similar to FIG.3) showing phylogenetical relationships between the two-dimensional HCA-plots of the SBDs (from position 49 to position 69) from the 4 extracellular globin sub-families. SBDs are similar and well-conserved in shape for both the A2 and B2 sub-families and remain as fingerprints in the A1 and B1 sub-families. The maintenance of such a hydrophobic pattern even for globins that lack free cysteine residues (data not shown) suggests a putative ancestral sulfide-binding function.

Discussion

Rooting the globin family and extracting the most informative amino-acid sequences

Ingram (1961) implicitly suggested that the hemoglobin gene is paralogous to the gene of myoglobin and arose by gene duplication. Although the concept of an universal myoglobin-like ancestor has been recently discussed (Suzuki and Imai 1998), these authors, together with Goodman et al (1988) and Hardison (1998) proposed the existence of a common ancestral

globin gene before the invertebrate/vertebrate divergence which is paralogous to the ancestral myoglobin gene. As a consequence, a closely related annelid myoglobin should have been used to root our extracellular paralogous globin trees. However, the available Annelida myoglobins were too genetically distant from our extracellular annelid globins to provide good outgroups. To overcome this rooting procedure, the *R. pachyptila* multigenic globin tree was rooted by using the globin A family against the globin B family and *vice versa* because the A and B Annelida extracellular globin genes are paralogous and arised from a single ancient duplication event. This allowed us to reconstruct trees that fit the recent annelid phylogeny of Rouse and Fauchald (1995, 1997).

Typically, the non-vertebrate globins exhibit a substantially greater amount of variation than that found in vertebrate globins (Kapp et al. 1995). As such, analyses must be performed on a carefully-selected group of sequences to recover the less biased phylogenetic signal from this new set of globin sequences. We assume that the central exonic region of the extracellular globins from annelids is more appropriate to reconstruct the molecular evolutionary history. Indeed, there is a positive correlation between the gene organization of the globin and its structural features (Go 1981; Blake 1981). The second exon encodes for the well-conserved and ubiquitous heme-binding fold in vertebrates (Craik, Buchman and Beychok 1981) and thus is less prone to differentiate with time. We assume that the second exon of extracellular globins encodes for the same structural features because intron positions are conserved between annelids and the intracellular globins of vertebrates (Jhiang, Garey and Riggs 1988; X. Bailly, unpublished data). Interestingly, the second exon also corresponds to the amino acid portion selected by GBlock and is the only set of data allowing the reconstruction of congruent phylogenies. Moreover, this amino acid portion of the globin sequence also displays the two free cysteine residues involved in the sulfide-binding function, allowing us to parallel the globin phylogeny and the molecular history of the unusual sulfide uptake in annelids living in sulfidic environments.

Globin assignment to A1, A2, B1 and B2 sub-families: are sub-families true paralogous genes?

One important consideration in phylogeny reconstruction from gene families is to discriminate between orthologous and paralogous genes when the time of divergence between some observed species is important, or if gene duplication is ancient (i.e. paralogous genes under saturation). The annelid radiation approximatively took place about 600 millions years ago (McHugh 2000) leading to the emergence polychaetes, achaetes, oligochaetes,

vestimentiferans and pogonophorans. Moreover, Rouse (2001) argued that pogonophorans and vestimentiferans are specialized Siboglinidae polychaetes that separated from Sabellidae even more recently. The extracellular hexagonal bilayer hemoglobin (HBL-Hb) is found in various annelid taxa including achaetes, oligochaetes, polychaetes and vestimentiferans. Because such a molecule is a complex edifice made of at least 144 globin chains and additional linkers (structural chains), it is unlikely that this hemoglobin could have appeared in different lineages independently and thus have evolved subsequently through convergent evolution. As a consequence, because four main types of extracellular globins are present in the oligochaete *Lumbricus terrestris*, the vestimentiferans *Lamellibrachia sp.* and *R. pachyptila* and the polychaete *Tylorrhynchus heterochaetus*, this strongly suggests that HBL-Hb emerged prior to the annelid radiation via two subsequent duplications. At least, it follows that one annelid ancestor possessed the A and B globin families as stated by Gotoh (1987), Suzuki, Takagi and Ohta (1993) and other studies dealing with annelid globin genes organization and evolution. By restricting our analysis to extracellular globins from vestimentiferans and closely-related polychaete relatives (e.g. *Sabella*) and by sorting sequences according to family-specific amino acid motifs and 2-D structure, we were able to demonstrate for the first time an unequivocal discrimination between paralogous and orthologous annelid globins and that the four main globin strains A1, A2, B1 and B2 co-evolved separately. In this respect, our work reveals that previous studies using extracellular globins produced unresolved tree topologies with incorrect sub-family assignment (Dewilde et al. 1996; Yuasa, Green and Takagi 1996; Zal et al. 1997b; Negrisolo et al 2001). Our results therefore confirm the hypothesis that the HBL-Hb arised before the annelid radiation from two ancient subsequent duplications, other duplications being more recent and probably species-specific. The present molecular phylogeny displays an identical species clustering for the four paralogous clades in which the species branching order is in agreement with other annelid phylogenies which involve vestimentiferans (Black et al. 1997; Halanych, Lutz and Vrijenhoek 1998; Rouse 2001). This finding demonstrates that extracellular globins could be used to test molecular clock hypothesis in multigenic families of invertebrates when uninformative portions of the globin sequence are removed.

Different evolutionary rates between paralogous genes despite molecular clock

The evolution of the four main globin paralogous clades, i.e. the similar species branching order (in agreement with other gene phylogenies), allowed us to test the hypothesis of a molecular clock using the relative rate test. Our results suggest the occurrence of a globin-

specific and domain-specific molecular clock between annelid species. Interestingly, this clock seems to be detectable only within a portion of the well-conserved second exon of the globin sequence. This raises questions about how a clock, which translates a neutral or nearly neutral molecular evolutionary rate could be detected in such a highly conserved structural domain (i.e. the second exon is prone to strong selective pressures). One explanation is to consider that only few amino acids (those responsible for the globin fold) are under selection, others (the highest fraction) evolving neutrally, i.e. are free to vary. Indeed, despite the common widespread idea that selective processes mainly act on the active catalytic site, Graur (1985) showed that the lowest rate of amino acid changes is not focused in the active parts of the molecule. An alternative explanation is that extracellular globins fit the covarion model of Fitch & Markovitz (1970) which asserts that only a small fraction of amino acid positions of a protein is free to vary (**co**ncomitant **vari**able cod**on**s) at a given time in its evolutionary history, due to functionnal constraints. The present situation may illustrate “how protein sequences can differ considerably among distant taxa even though only a limited number of positions are free to vary within a particular lineage" (Miyamoto and Fitch 1995). The joint phylogenetic analysis of the four paralogous sets of extracellular globins allows us to discriminate between these two possible explanations. While the relative rate test lacks power to be sure of a molecular clock in extracellular globins of Sabellidae, Vestimentifera and Pogonophora, our results suggest that (1) the four globin sub-families evolve at different rates (i.e. branch lengths differ between paralogous clades) and (2) each rate appears to be constant between species for the four orthologous clades which exhibit the same species branching order. This finding therefore seems to reject the first hypothesis simply because one might expect the four sub-families to evolve at a nearly same rate if neutral (similar branch lengths between clades). Fractions of amino acids free-to-vary are thus probably different between the four sub-families, hence the different evolutionary rates between sub-families, but could explain why species follow the same pattern of clustering, hence the neutral evolution in orthologous globins. Thus, each sub-family displays a specific molecular clock because of natural selection acted differentially on the four globin sub-families. However, the fact that selective pressures acted differently on paralogous strains but homogeneously on orthologous strains suggests that the whole environmental constraints did not change since the radiation of both Siboglinidae and Sabellidae. Such an assumption seems likely since these organisms are all found in sulfidic environments. Thus, our results could be viewed as a strong insight to underline the joint action of selective pressure and (nearly) neutral evolution as claimed by Ohta (2000) to reconcile neutralist and selectionist partisans. However, even if the selected

amino acid portion used here to reconstruct this molecular phylogeny could be refined, it cannot be extended to highly divergent taxa such as the oligochaete *Lumbricus terrestris* or the polychaete *Tylorrhynchus heterochaetus* although they both possess an HBL-Hb and thus the four types of extracellular globins. In this case, one can argue that such a loss of signal is due to saturation. But an alternative explanation would also be to consider that selective pressures may have been altered since these two latter highly divergent species are also known to inhabit very-well oxygenated environments. These observations suggest that extracellular globins are not an informative marker to perform phylogeny analyses over the whole Annelida phylum. However it provides a powerful tool to study molecular evolution of paralogous genes within closely-related species.

*Lower evolutionary rate*s *correlates with the occurrence of free cysteine residues and Sulfide Binding Domain (SBD).*

To test whether the differing evolutionary rates between *R. pachyptila* globin sub-families could be due to the action of selective pressures, genetic distances between the orthologous sequences of two closely-related species of Vestimentifera (*R. pachyptila* and *Lamellibrachia sp.*) were calculated for the four paralogous sub-families, providing that both species are endosymbiotic, live in the same ecological niche and separated less than 50 millions year ago. Indeed, a recent clock calibration has been obtained for the vent annelids (rate = 0.2% per Myr: Chevaldonné et al. in press) together with a phylogenetic tree for the vestimentiferans and the pogonophorans for the mitochondrial COI marker (Black et al. 1995; Halanych et al. 1998). These results allowed us to estimate the divergence time between *Lamellibrachia sp.* and *R. pachyptila* to be around 46 Myr (no divergence time is available between Sabellidae and Siboglinidae). In other words, as genetic distances measure the expected accumulation of mutations since speciation, an attempt was made to establish whether A1, A2, B1 and B2 evolved at the same evolutionary speed. It clearly appeared that A2 and B2 sub-families display the lowest evolutionary rates. Interestingly, the A2 and B2 globin chains are both involved in sulfide binding via their respective free cysteine residues (Cys + 1 and Cys + 11), and one could obviously assume that this rate decrease may be the result of a strong directional selection towards the maintenance of this unusual globin function. Genetic distances from the orthologous *R. pachyptila* and *Lamellibrachia sp.* globin without the portion corresponding to the sulfide binding domain (i.e. 21 amino acids surrounding the free cysteine) provide the same trend. Thus, the sulfide-binding function associated with the A2 and B2 globin chains could have been strongly selected by limitation of amino acid changes

during the course of evolution. In addition, such an assumption is well supported by (1) the maintenance of a typical A2 and B2 hydrophobic secondary structure associated with the free cysteine residues in most lineages (2) the remains of a putative SBD in A1 and B1, these globins being devoid of free cysteines in species living in sulfidic environments and (3) the noticeable absence of homologous free cysteines residues in A2 and B2 globins of annelids living in well-oxygenated environments despite the persistence of a more or less degenerated SBD fingerprint.

A non-exclusive additional explanation might be that A1 and B1 globin sub-families have accumulated more mutations than A2 and B2 just after the duplication event. This is the acceleration of evolution by duplication effect (Hill and Hastie 1987; Kini and Chan 1999) and can be due to relaxation of selective pressures (Kimura 1981) or positive Darwinian selection (Zhang, Rosenberg and Nei 1998). However, one cannot exclude a selective structural effect of the multimeric hemoglobin on monomers A2 and B2 relative to A1 and B1 according to their location (allosteric effect). Further information concerning the hemoglobin structure of *R. pachyptila* is required.

As both *Lumbricus terrestris* and *Tylorrhynchus heterochaetus* do not exhibit free cysteine residues and the ability to bind H_2S, it may be that sulfide-binding is a derived function in the Siboglinodae. However, from our data, we demonstrated that the positions Cys + 1 and Cys + 11 involved in the sulfide-binding function were both well-conserved in the A2 and B2 strains, for at least vestimentiferans, pogonophorans and Sabellid worms. Moreover, a free cysteine residue was also found in the A1 strain in *Oligobrachia mashikoi* at the Cys + 1 position with a similar super-imposable SBD. Because free cysteine residues were found at the same location on different paralogous strains, the cysteine at position Cys + 1 and Cys + 11 might have arisen before the first duplication that led to the emergence of the A and B globin families. These data allowed us to refute the hypothesis that H_2S detoxification/transport via a reactive cysteine acquisition process may derived or gained independently in different annelid lineages. It is more parsimonious to propose that the first annelids inhabiting sulfidic environments lost the ability to bind sulfides in well-oxygenated environments. These findings indicate that free cysteine residues have been conserved since the radiation of the Siboglinidae and even more before the radiation of annelids. This hypothesis is also well supported by the remnants of a putative sulfide binding domain of annelids living in well-oxygenated environment (*Tylorrhynchus heterochaetus* and *Lumbricus terrestris*). A parsimonious evolutionary scenario is presented in Figure 5 and relies on the assumption that the vestimentiferan non-symbiotic ancestors lived in contact with high sulfide

concentrations and had a hemoglobin involved in a H_2S detoxification process. Indeed, if A2 and B2 strains are both involved in sulfide binding via their respective free cysteines, one could assume that this property must have been maintained during the annelid evolution. Therefore, this function has then evolved to be a H_2S detoxification/transport function in vestimentiferans as a means to fuel their endosymbionts during the symbiosis acquisition process (Arp, Childress and Vetter 1987). This evolutionary scheme is well-supported by the occurrence of free cysteine residues also involved in the sulfide-binding function in other non-symbiotic polychaetes living in sulfidic environments.Zal et al.(1997a; 1999) described a HBL-Hb which binds sulfides via free cysteine residues (positions unknown) in the shallow-water polychaete *Arenicola marina* and the hydrothermal vent polychaete *Alvinella pompejana*. Thus, the presence of a degenerated SBD in globins that do not possess free cysteine residues in HBL-Hb-bearing annelids living in well-oxygenated habitats (*Lumbricus terrestris* and *Tylorrhynchus heterochaetus*) is an insight of the probable loss of these free cysteine residues by mutation/drift when selection is relaxed (nearly neutral evolution).

Is Sulfide-binding function via free cysteine a plesiomorphic innovation in annelid evolution?
H_2S has been a widespread molecule on the Earth surface since prebiotic times untill the present and is found at high concentrations in some places such as hydrothermal vents which have been proposed to mimic the primary Earth conditions (Clark, Dowling and Huang 1998). Tunnicliffe (1992) and Tunnicliffe and Fowler (1996) pointed out that hydrothermal vents has existed from Phanerozoic and possibly Archean. In addition, Haymon, Koski and Sinclair (1984), Banks (1985) and Little et al. (1997) have described tube-worm and worm-like fossils in ophiolites of 350 Myr or more (Carboniferous and Silurian). It is therefore likely that hydrothermal-vent species may have been adapted to sulfidic conditions since the beginning of metazoan life. Recent studies regarding organismal sulfide resistance strategies (Zierenberg, Adams and Arp 2000) clearly indicate the fundamental contribution in species ecology of H_2S concentrations. In that respect, one can strongly suspect the necessity of maintenance of the sulfide-binding function by a directional selection processes (negative selection) acting on genes encoding A2 and B2 globins. Despite very few annelid hemoglobins having been studied so far, many non-symbiotic and symbiotic polychaetes and oligochaetes, colonise sulfidic environments such as mangrove swamps, anoxic basins, marine sediments, deep-sea groundwater seeps and hydrothermal vents (Vetter and Powell 1991; Arp, Menon and Julian 1995; Volkel 1995; Hahlbeck, Arndt and Schiedek 2000).

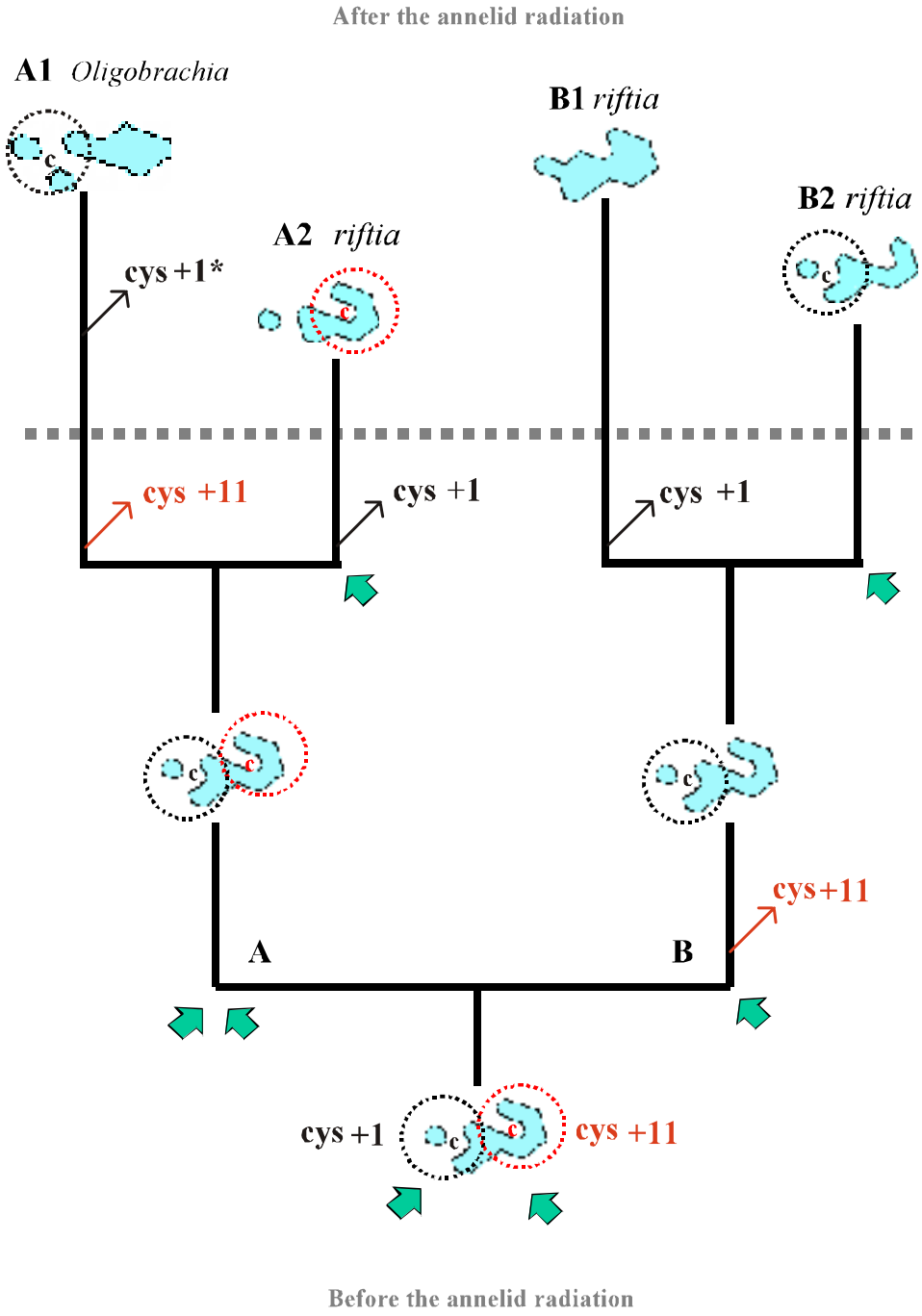


FIG 5. Speculative evolutionary history of the sulfide-binding function through duplications that led to the actual HBL-Hbs in worms living in sulfidic environments. This evolutionary scenario is based on (1) our present knowledge of the positions of free cysteine residues in vestimentiferans, (2) the occurrence of a cysteine residue at position Cys+1 in the A1 sub-family of *Oligobrachia mashikoi* and (3) knowledge of the positions of free cysteine residues in other polychaete species. Arrows symbolize selective pressures acting on extracellular globins. The scenario proposes an asymmetrical loss of the free cysteine residues prior to the annelid radiation, the Cys + 1 residue being maintained on the A strain prior to the annelid radiation and subsequently lost in the A1 sub-family after the annelid radiation, depending on taxa (*: lost in Vestimentifera).

So far, all the studied worms living in sulfidic environments possess (1) an intracellular hemoglobins with free cysteine residues (Garey and Riggs 1986), or (2) an extracellular hemoglobin with free cysteine residues (Zal et al. 1997b). Both intra- and extracellular hemoglobins are thus involved in sulfide-binding. Chlorocruorins of some of these worms possess free cysteine residues as well, but no evidence for a sulfide-binding function has yet to be provided. In conclusion, the SBD resemblance and the conserved positions of free cysteine residues strongly suggest that the sulfide-binding function is a widespread adaptation in annelid, vestimentiferan and pogonophoran extracellular globins which emerged prior to the annelid radiation and may also have concerned some intracellular globins in the same phylum. It is interesting that annelids display a H_2S covalent mechanism, whereas molluscs living in sulfide-rich environment bind H_2S via amino acid electrostatic interaction (Nguyen 1998). This suggests a functional convergence that involves different binding processes and that H_2S binding via free cysteine is an annelid innovation.

Acknowledgements

We gratefully acknowledge the captains and crews of the NO L'Atalante and the RV Atlantis II, the pilots and teams of the submersibles Nautile and Alvin, and F. Gaill (UPMC, Paris, France), L. S. Mullineaux (WHOI, Woods Hole, USA) and H. Felbeck (SCRIPPS Institute, San Diego, USA), chief-scientists of the HOT96 and LARVE99 cruises. We wish to thank Dr. Dugald McGlashan for helping us with the correction of this manuscript and two anonymous referees for their useful comments. This work was supported by the Conseil régional de Bretagne and the Ministère de l'Education Nationale et de la Recherche (ACC-SV3).

Literature Cited

Arp, A. J., J. J. Childress, and R. D. Vetter. 1987. The sulfide-binding protein in the blood of the vestimentiferan tube-worm *Riftia pachyptila*, is the extracellular hemoglobin. J. Exp. Biol. 128:139-158.

Arp, A. J., J. G. Menon, and D. Julian. 1995. Multiple mechanisms provide tolerance to environmental sulfide in *Urechis caupo*. Amer. Zool. 35:132-144.

Banks, D. A. 1985. A fossil hydrothermal worm assemblage from the Tynagh lead-zinc deposit in Ireland. Nature 313:128-131.

Black, M. B., K. M. Halanych, P. A. Y. Maas, W. R. Hoeh, J. Hashimoto, D. Desbruyères, R. A. Lutz, and R. C. Vrijenhoek. 1997. Molecular systematics of vestimentiferan tubeworms from hydrothermal vents and cold-water seeps. Mar. Biol. 130:141-149.

Blake, C. C. 1981. Exons and the structure, function and evolution of haemoglobin. Nature 291:616.

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

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17:540-552.

Chevaldonné, P., D. Jollivet, D. Desbruyères, R. D. Lutz, and R.C. Vrijenhoek. 2002. Sibling species of eastern Pacific hydrothermal-vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial *COI* clock calibration. Cah. Biol. Mar. in press.

Childress, J. J., and C. R. Fischer. 1992. The biology of hydrothermal vent animals: physiology, biochemistry and autotrophic symbioses. Oceanogr. Mar. Biol. Annu. Rev. 30:337-441.

Clark, P. D., N. I. Dowling, and M. Huang. 1998. Comments on the role of H_2S in the chemistry of Earth's early atmosphere and in prebiotic synthesis. J. Mol. Evol. 47:127-132.

Craik, C. S., S. R. Buchman, and S. Beychok. 1981. O_2 binding properties of the product of the central exon of beta-globin gene. Nature 291:87-90.

Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt. 1978. A model of evolutionary change in protein. Pp. 354-352 in M. O., Dayhoff, ed. Atlas of protein sequence structure, Vol. 5, Suppl. 3, National Biomedical Research Foundation, Washington DC.

Dewilde, S., M. Blaxter, M. L. Van Hauwaert, J. Vanfleteren, E. L. Esmans, M. Marden, N. Griffon, and L. Moens. 1996. Globin and globin gene structure of the nerve myoglobin of *Aphrodite aculeata*. J. Biol. Chem. 271:19865-19870.

Donoghue, M. J., and S. Mathews. 1998. Duplicate genes and the root of angiosperms, with an example using phytochrome sequences. Mol. Phylogenet. Evol. 9:489-500.

Fitch, W. M., and E. Markowitz. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem. Genet. 4:579-593.

Fushitani, K., M.S. Matsuura, and A.F. Riggs. 1988. The amino acid sequences of chains a, b and c that form the trimer subunit of the extracellular hemoglobin from *Lumbricus terrestris terrestris*. J. Biol. Chem. 263:6502-6517.

Galtier, N., M. Gouy, and C. Gautier. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. 12:543-548.

Garey, J. R., and Riggs A. F. 1986. The hemoglobin of *Urechis caupo*. The cDNA-derived amino acid sequence. J. Biol. Chem. 261:16446-16450

Go, M. 1981. Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90-92.

Goodman, M., J. Barnabas, G. Matsuda, and G. W. Moore. 1971. Molecular evolution in the descent of man. Nature 233:604-613.

Goodman, M., G. W. Moore, and G. Matsuda. 1975. Darwinian evolution in the genealogy of hemoglobin. Nature 253:603-608.

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

Gotoh, T., F. Shishikura, J. W. Snow, K. I. Ereifej, S. N. Vinogradov, and D. A. Walz. 1987. Two globin strains in the giant annelid extracellular haemoglobins. Biochem. J. 241:441-445.

Graur, D. 1985. Amino acid composition and the evolutionary rates of protein-coding genes. J. Mol. Evol. 22:53-62.

Grieshaber, M. K., and S. Volkel. 1998. Animal adaptations for tolerance and exploitation of poisonous sulfide. Annu. Rev. Physiol. 60:33-53.

Hahlbeck, E., C. Arndt, and D. Schiedek. 2000. Sulfide detoxification in *Hediste diversicolor* and *Marenzelleria viridis*, two dominant polychaete worms within the shallow coastal waters of the southern Baltic Sea. Comp. Biochem. Physiol. B 125:457-471.

Halanych, K. M., R. A. Lutz, and R. C. Vrijenhoek. 1998. Evolutionary origins and age of vestimentiferan tube-worms. Cah. Biol. Mar. 39:355-358.

Hardison, R. 1998. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. J. Exp. Biol. 201:1099-1117.

Haymon, R. M., R. A. Koski, and C. Sinclair. 1984. Fossils of hydrothermal vent worms from cretaceous sulfide ores of the Samail Ophiolites, Oman. Science 223:1407-1409.

Hill, R. E., and N. D. Hastie. 1987. Accelerated evolution in the reactive centre regions of serine protease inhibitors. Nature 326:96-99.

Ingram, V. 1961. Gene volution and the haemoglobin. Nature 189:704-708.

Iwabe, N., K. Kuma, M. Hasegawa, S. Osawa, and T. Miyata. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA 86:9355-9359.

Jhiang, M. S., J. R. Garey, and A. F. Riggs. 1988. Exon-intron organization in genes of earthworm and vertebrate globins. Science 240:334-336.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. CABIOS 8:275-282.

Jones, M. L. 1981. *Riftia pachyptila* Jones: observations on the vestimentiferan worm from the Galapagos Rift. Science 213:333-336.

Kapp, O. H., L. Moens, J. Vanfleteren, C. N. A. Trotman, T. Suzuki, and S. N. Vinogradov. 1995. Alignment of 700 globin sequences: extent of amino acid substitution and its correlation with variation in volume. Protein Sci. 4:2179-2190.

Karplus, K., K. Sjolander, C. Barrett, M. Cline, D. Haussler, R. Hughey, L. Holm, and C. Sander. 1997. Predicting protein structure using hidden Markov models. Proteins Suppl. :134-139.

Kimura, M. 1981. Was globin evolution very rapid in its early stages? a dubious case against the rate-constancy hypothesis. J. Mol. Evol. 17:110-113.

Kini, R. M., and Y. M. Chan. 1999. Accelerated evolution and molecular surface of venom phospholipase A2 enzymes. J. Mol. Evol. 48:125-132.

Little, C. T. S., R. J. Herrington, V. V. Maslennikov, N. J. Morris, and V. V. Zaykov. 1997. Silurian hydrothermal-vent community from the southern Urals, Russia. Nature 385:146-148.

McHugh, D. 2000. Molecular phylogeny of the Annelida. Can. J. Zool. 78:1873-1884.

Miyamoto, M. M., and W. M. Fitch. 1995. Testing the covarion hypothesis of molecular evolution. Mol. Biol. Evol. 12:503-513.

Negrisolo, E., A. Pallavicini, R. Barbato, S. Dewilde, A. Ghiretti-Magaldi, L. Moens, and G. Lanfranchi. 2001. The evolution of extracellular hemoglobins of annelids, vestimentiferans, and pogonophorans. J. Biol. Chem. 276:26391-26397.

Nguyen BD, Zhao X, Vyas K, La Mar GN, Lile RA, Brucker EA, Phillips GN Jr, Olson JS, Wittenberg JB. 1998. Solution and crystal structures of a sperm whale myoglobin triple mutant that mimics the sulfide-binding hemoglobin from Lucina pectinata J Biol Chem Apr 17;273(16):9517-26

Nicholls, P. 1975. The effect of sulfide on cytochrome aa3. Isosteric and allosteric shifts of the reduced alpha-peak. Biochim. Biophys. Acta 396:24-35.

Ohta, T. 2000. Mechanisms of molecular evolution. Philos. Trans. R. Soc. Lond. B 355:1623-1626.

Olson, J. S., A. J. Mathews, R. J. Rohlfs, B. A. Springer, K. D. Egeberg, S. G. Sligar, J. Tame, J. P. Renaud, and K. Nagai. 1988. The role of the distal histidine in myoglobin and haemoglobin. Nature 336:265-266.

Pallavicini, A., E. Negrisolo, R. Barbato, S. Dewilde, A. Ghiretti-Magaldi, L. Moens, and G. Lanfranchi. 2001. The primary structure of globin and linker chains from the chlorocruorin of the polychaete *Sabella spallanzanii*. J. Biol. Chem. 276:26384-26390.

Rossi-Fanelli, A., D. Cavallini, and C. De Marco. 1955. Amino-acid composition of human crystallized myoglobin and haemoglobin. Nature 17:377-381.

Rouse, G. W. 2001. A cladistic analysis of Siboglinidae Caullrey, 1914 (Polychaeta, Annelida): formerly the phyla Pogonophora and Vestimentifera. Zool. J. Linn. Soc. 132:55-80.

Rouse, G. W., and K. Fauchald. 1995. The articulation of annelids. Zool. Scripta 24: 269-301.

Rouse, G. W., and K. Fauchald. 1997. Cladistics and polychaetes. Zool. Scripta 26: 139-204.

Sarich, V. M., and A. C. Wilson. 1973. Generation time and genomic evolution in primates. Science 179:1144-1147.

Shishikura, F., J. W. Snow, T. Gotoh, S. M. Vinogradov, and D. A. Waltz. 1987. Amino acid sequence of the monomer subunit of the extracellular hemoglobin of *Lumbricus terrestris terrestris*. J. Biol. Chem. 262:3123-3131.

Strimmer, K., and A. von Haeseler. 1996. Quartet Puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964-969.

Suzuki, T., and T. Gotoh. 1986. The complete amino acid sequence of giant multisubunit hemoglobin from the polychaete *Tylorrhynchus heterochaetus* . J. Biol. Chem. 261: 9257-9267.

Suzuki, T., Y. Hirao, and S. N. Vinogradov. 1995. Primary structure of a constituent polypeptide chain of the chlorocruorin from *Sabellastarte indica*. Biochim. Biophys. Acta 1252:189-93.

Suzuki, T., and K. Imai. 1998. Evolution of myoglobin. Cell Mol. Life Sci. 54:979-1004.

Suzuki, T., T. Takagi, and S. Ohta. 1990. Primary structure of a constituent polypeptide chain (AIII) of the giant haemoglobin from the deep-sea tube worm *Lamellibrachia sp*. A possible H_2S-binding site. Biochem. J. 266:221-225.

Suzuki, T., T. Takagi, and S. Ohta. 1993. N-terminal amino acid sequence of 440 kDa haemoglobins of deep-sea tube worms, *Lamellibrachia sp.*1, *Lamellibrachia sp.*2 and slender vestimentifera gen.sp.1. Evolutionary relationship with annelid hemoglobins. Zool. Sci. 10:141-146.

Suzuki, T., T. Takagi, K. Okuda, T. Furukohri, and S. Ohta. 1989. The deep-sea tube worm hemoglobin: subunit structure and phylogenetic relationship with annelid hemoglobin. Zool. Sci. 6:915-926.

Swofford, D. L. 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign.

Takagi, T., H. Iwaasa, S. Ohta, and T. Suzuki. 1991. Primary structure of 440 kDa hemoglobin from the deep sea tubeworm *Lamellibrachia.* Pp. 245-249 in S. N. Vinogradov and O. H. Kapp, eds. Structure and function of invertebrate oxygen carriers, Springer, New York.

Terwilliger, R. C., N. B. Terwilliger, and E. Schabtach. 1980. The structure of hemoglobin from an unusual deep sea worm (Vestimentifera). Comp. Biochem. Physiol. 65B:531-535.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882.

Tunnicliffe, V. 1992. The nature and origin of the modern hydrothermal vent fauna. Palaios 7:338-350.

Tunnicliffe, V., and C. M. R. Fowler. 1996. Influence of sea-floor spreading on the global hydrothermal vent fauna. Nature 379:531-533.

Vetter, R. D., and G. N. Powell. 1991. Metazoan adaptation to hydrogen sulfide. Pp. 109-128 in C. Bryant, ed. Metazoan life without oxygen. Chapman and Hall, London.

Vinogradov, S. N., D. A. Walz, B. Pohajdak, L. Moens, O. H. Kapp, T. Suzuki, and C. N. A. Trotman. 1993. Adventitious variability? the amino acid sequences of nonvertebrate globins. Comp. Biochem. Physiol. 106B:1-26.

Vismann, B. 1991. Sulfide tolerance: physiological mechanisms and ecological implications. Ophelia 34:1-27.

Volkel, S. 1995. Sulfide tolerance and detoxification in *Arenicola marina* and *Sipunculus nudus*. Amer. Zool. 35:145-153.

Weber, R. E., and S. N. Vinogradov. 2001. Nonvertebrate hemoglobins: functions and molecular adaptations. Physiol. Rev. 81:569-628.

Yuasa, H. J., B. N. Green, and T. Takagi. 1996. Electrospray ionization mass spectrometic composition of the 400 kDa hemoglobin from the pogonophoran *Oligobrachia mashikoi* and the primary structures of the three major globin chain. Biochim. Biophys. Acta 1296:235-244.

Zal, F., T. Gotoh, and A. Toulmond. 1999. The novel function of giant hemoglobins from tubeworms and annelids. in Pp. S16, Fifth International Congress of Comparative Physiology and Biochemistry, 124A, Calgary, Alberta, Canada.

Zal, F., B. N. Green, F. H. Lallier, S. N. Vinogradov, and A. Toulmond. 1997a. Quaternary structure of the extracellular haemoglobin of the lugworm *Arenicola marina*: a multi-angle-laser-light-scattering and electrospray- ionisation-mass-spectrometry analysis. Eur. J. Biochem. 243:85-92.

Zal, F., F. H. Lallier, B. N. Green, S. N. Vinogradov, and A. Toulmond. 1996a. The multi-hemoglobin system of hydrothermal vent tube worm *Riftia pachyptila*: II- Complete polypeptide chain composition investigated by Maximum Entropy analysis of mass spectra. J. Biol. Chem. 271:8875-8881.

Zal, F., F. H. Lallier, J. S. Wall, S. N. Vinogradov, and A. Toulmond. 1996b. The multi-hemoglobin system of hydrothermal vent tube worm *Riftia pachyptila*: I- Reexamination of the number and masses of its constituents. J. Biol. Chem. 271:8869-8874.

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

Zal, F., T. Suzuki, Y. Kawasaki, J. J. Childress, L. F., and A. Toulmond. 1997b. Primary structure of the common polypeptide chain *b* from the multi-haemoglobin system of the hydrothermal vent tube worm *Riftia pachyptila*: an insight on the sulfide binding-site. Proteins Struct. Funct. Genet. 29:562-574.

Zhang, J., H. F. Rosenberg, and M. Nei. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA 95:3708-3713.

Zierenberg, R. A., M. W. Adams, and A. J. Arp. 2000 Life in extreme environments: hydrothermal vents. Proc. Natl. Acad. Sci. USA 97:12961-12962.

Zuckerkandl, E., R. T. Jones, and L. Pauling. 1960. A comparison of animal hemoglobin by triptic peptide pattern analysis. Proc. Natl. Acad. Sci. USA 46:1349-1360.

Chapitre 5 : Deuxième article

La perte de la fonction de fixation du sulfure d'hydrogène chez les annélides des milieux oxydés s'explique par une adaptation moléculaire, résultat de la sélection darwinienne positive

Article paru dans **Proceedings of the National Academy of Sciences of the United States of America**

Vol. 100 n° 10 Pp 5885–5890 (2003)

Chapitre 5: Deuxième article

La perte de la fonction de fixation du sulfure d'hydrogène chez les annélides des milieux oxydés est le résultat de la sélection darwinienne positive

Pour comprendre comment la fonction de fixation de H_2S est apparue et a été maintenue chez les hémoglobines extracellulaires d'annélides des milieux réduits au cours de l'évolution, nous avons caractérisé (cloné et séquencé) la famille multigénique des hémoglobines de *Riftia pachyptila* (Bailly et al, 2002). L'étude de l'évolution moléculaire de ces globines a montré que les paralogues de cette famille évoluent à des vitesses différentes : les deux sous-unités capables de lier réversiblement H_2S, grâce à la présence de résidus cystéines libres, présentent des vitesses évolutives plus lentes que les autres sous-unités impliquées dans le seul transport de l'oxygène. Nous suggérons que la relation évidente entre les annélides des milieux riches en sulfures et le sulfure d'hydrogène, relation basée essentiellement sur le statut toxique et énergétique de H_2S, pouvait expliquer le maintien de la fonction de fixation de H_2S. En d'autres termes, que les concentrations en H_2S constituent une forte pression de sélection (Bailly et al, 2002). Nous proposons que les milieux oxydés pourraient avoir été colonisés par des annélides vivant originellement dans des milieux réduits qui auraient perdu la fonction de fixation de H_2S dans les milieux oxydés (Bailly et al, 2002). **L'émergence des annélides dans des milieux réduits et la colonisation des milieux oxydés par la suite, est le postulat qui a servi de base pour tester l'hypothèse de la perte de la fonction de fixation du sulfure d'hydrogène chez les annélides des milieux oxydés**

Nous avons voulu savoir si l'absence de cette fonction et des cystéines libres chez les annélides des milieux sans sulfures pouvait s'expliquer autrement que par un relâchement des contraintes sélectives (évolution neutre). Parce que les cystéines sont des acides aminés très réactifs et impliqués dans de nombreuses réactions métaboliques, nous nous sommes demandés si la conservation d'un tel résidu au sein des protéines ne pouvait pas conférer un désavantage sélectif pour l'espèce vivant dans un milieu dépourvu de sulfures.

Des reconstructions putatives de structures 3D des globines A2 et B2 de *Riftia pachyptila* montrent que les groupements thiols des cystéines fixatrices de H_2S sont différemment orientés dans l'espace et que leur localisation par rapport à l'hème pourrait leur conférer des rôles différents à l'échelle moléculaire. Ces observations renforcent l'idée que la liaison des sulfures est un mécanisme complexe. Le séquençage de nouvelles globines A1, B1, A2 et B2 d'espèces proches de vestimentifères associées soit aux suintements froids, soit aux sources

hydrothermales, le séquençage d'un linker (qui fixe également les sulfures) de *Riftia pachyptila* et les séquences nucléotidiques des globines A2 et B2 de l'oligochète *Lumbricus rubellus* des milieux oxydés, nous ont permis d'accéder à un autre niveau d'information que les séquences primaires seules. En effet, à partir de ce jeu de données supplémentaire nous avons pu préciser l'action de la sélection naturelle opérant par les fortes concentrations en H_2S sur le maintien de la fonction de fixation des sulfures par les hémoglobines.

L'estimation des substitutions synonymes (dS) et non synonymes (dN) repose sur l'utilisation du programme PAML (Phylogenetic Analysis by Maximum Likelihood, (Yang 1997)) qui permet de calculer la proportion dN/dS qui est l'indicateur du type de sélection qui opère sur les protéines. Quand le rapport dN/dS >1 la sélection positive se traduit par l'accumulation de substitutions non synonymes et permet de modifier la nature de certains acides aminés. Quand dN/dS < 1 la sélection est négative et l'accumulation de substitutions synonymes s'oppose à des remplacement d'acides aminés. L'évolution neutre se traduit par dN/dS = 1.

Le principe de ce programme est de comparer et tester statistiquement différentes hypothèses évolutives. Trois modèles successifs ont permis d'estimer par maximum de vraisemblance le dN/dS à partir d'alignements nucléotidiques de séquences codantes et d'un arbre phylogénétique correspondant. Pour chaque programme, il s'agit de tester un scénario évolutif contre un autre.

Le premier programme permet de tester si la sélection a opéré de manière indépendante sur une lignée évolutive (une branche dans la phylogénie) ou si la sélection est identique dans toutes les lignées (toutes les branches de l'arbre) (annexe 7.1.1). Le second programme permet de tester la présence de sites accumulant préférentiellement des substitutions non synonymes chez l'ensemble des taxons étudiés (annexe 7.1.2). Le troisième programme permet de détecter des sites sous sélection positive dans certaines lignées uniquement parmi l'ensemble des lignées évolutives (annexe 7.1.3).

En confrontant ces résultats aux contraintes sélectives (présence de H2S) des milieux intertidaux, hydrothermaux et terrestres, nous confirmons un scénario évolutif qui suggère que la fixation de H2S via une cystéine libre est une innovation des annélides. Cette fonction semble avoir été perdue par un processus d'adaptation moléculaire chez les annélides modernes des milieux sans H2S (relaxation des contraintes) et indique que cette adaptation aux milieux sans sulfures s'est produite par perte des cystéines libres fixant H2S.

The loss of the hemoglobin H_2S-binding function in annelids from sulfide-free habitats reveals molecular adaptation driven by Darwinian positive selection.

Xavier Bailly[1-2], Riwanon Leroy[2], Susan Carney[3], Olivier Collin[4], Franck Zal[2], André Toulmond[2] and Didier Jollivet[1]

1. Station Biologique de Roscoff, Equipe EGPM, CNRS UMR 7127
2. Station Biologique de Roscoff, Equipe Ecophysiologie, CNRS UMR 7127
3. The Pennsylvania State University, Department of Biology, 208 Mueller Lab, University Park, PA 16802
4. Station Biologique de Roscoff, Service Informatique, CNRS FR2424

Introduction

Emergence of new functions caused by a high evolutionary rate after gene duplication has been heavily debated from the standpoint of neutralist and selectionist partisans. They respectively opposed relaxed constraint (1) and Darwinian positive selection (2) as the likely evolutionary forces for the acquisition of new functions. A consensual theory favored for the concomitant but temporally different action of these two forces, (3) diversifying selection, first acts by the accumulation of favorable mutations providing new functionality or sub-functionality, which then evolves in a neutral way. However, molecular evolution does not necessarily occur in concert with duplication events, but it may also be the result of diversifying selection driven by heterogeneous habitats (4). The estimation of the fixation rate of non-synonymous and synonymous substitutions along orthologous coding sequences from a cluster of evolutionary-related taxa appears to be one of the most powerful tools to detect molecular adaptation (5) (6) (7) (8) (9). However, the signature of molecular adaptation can be cryptic and difficult to extract because an adaptive change (advantageous mutation) may only affect a few number of lineages and only a subset of sites according to their phylogenetic history (10). Ancient adaptive mutation accumulation is typically the situation encountered in globins, which are a widespread and conserved molecule on Earth in members of all living kingdoms, including annelids.

The actual spatial and environmental distribution of annelids, from deep-sea hydrothermal vents to terrestrial habitats, is the consequence of a long history of adaptive strategies since their radiation (11). One of these adaptations concerns the way by which annelids living in sulfide-rich environments protect themselves against or use hydrogen sulfide (H_2S). Such a

process mainly relies on the occurrence of extracellular hemoglobins that bind and transport this toxic compound. H_2S is toxic to aerobic metabolism, particularly to metalloproteins such as cytochrome c oxidase and hemoglobin (Hb) (12). This unusual sulfide-binding function of some annelid hemoglobins was first discovered in the vestimentiferan *Riftia pachyptila*, a mouthless and gutless organism harboring intracellular chemolithoautotrophic sulfide-oxidizing bacterial symbionts; *Riftia* is found living close to the deep-sea eastern Pacific hydrothermal vents (13). Sulfide binding is enabled by the presence of two highly reactive free cysteine residues which covalently bind H_2S (14) (15) (16), each one localized on two different globin subunits that partly form the extracellular hexagonal-bilayer hemoglobin found in annelids (HBL-Hb) (see review (17)). The annelid HBL-Hb is made of 144 globin chains subdivided into two main gene families, A and B, and four subfamilies. A1, A2, B1 and B2, that emerged via at least three duplication events (18) (19). These latter authors found that the free cysteine residues involved in H_2S binding are located at the same positions, Cys+1 and Cys+11 (one and eleven amino acids after the well conserved distal histidine), on globin chains within the B2 and A2 subfamilies respectively for a set of various annelids living in sulfide-rich habitats. Moreover, other non-symbiotic annelid polychaetes living in sulfide-rich habitats such as *Alvinella pompejana* and *Arenicola marina* also possess HBL-Hbs that display a H_2S-binding capability via free cysteines residues (20) (21) (22) for which positions are unknown (no sequence available).

A lot of habitats that display high sulfide concentrations are known to occur on the Earth's surface. Environmental sulfide may have either geothermal (hydrothermal vents, sulphurous springs) or biogenic (cold seeps, marine sediments, mangroves) origins, including anthropogenic deposits in polluted marine or brackish areas. We postulated that species from sulfide rich environment exhibiting free cysteine residues at positions Cys+1 and Cys +11 are able to bind sulfide by analogy to the mechanism utilized by both of the vestimentiferans *Lamellibrachia sp.* and *R. pachyptila.* Such H_2S-binding function appears to be absent in annelids from sulfide-free environments such as the oligochaete *Lumbricus terrestris* (earthworm) and the polychaete *Tylorrhynchus heterochaetus* which lack these residues. The inability of *Lumbricus* hemoglobin to bind sulfide was confirmed by Zal et al. (1997) using specific cysteine inhibitors. As a consequence, Bailly et al. (2002) proposed an evolutionary scenario regarding the evolution of the hemoglobin H_2S-binding function in symbiotic and non-symbiotic annelids living in sulfide-rich habitats and suggested that the H_2S-binding function via a free cysteine residue was (1) an innovation in Phylum Annelida and (2) lost by

the relaxation of selective constraint (neutral evolution) in the annelid ancestors that colonized the newly emerging sulfide-free habitats.

Starting with these assumptions, we focused our attention here on the A2 and B2 homologous free cysteine sites that are located in a well-conserved secondary structure region called the Sulfide Binding Domain (SBD) (19). A composite set of tools, including ancestral sequence reconstruction, evolutionary rate calculation, synonymous and non-synonymous substitution estimation and 3-D globin modeling, were used to investigate the functional evolution of free cysteines in annelids from sulfide-rich and sulfide-free habitats. A new sequence of a linker chain which is involved in both the maintenance of the hemoglobin quaternary structure of *R. pachyptila* and the H_2S binding function (16) was added to the dataset in order to get a complete evolutionary picture of this unusual molecule. Its ability to bind H_2S, as do the A2 and B2 globin chains, enabled us to better describe and understand how H_2S concentrations may have played a central selective role in the evolution of respiratory pigments.

To our knowledge we present and demonstrate here the first case of molecular adaptation (replacement of the free cysteines) due to the relaxation of selective constraint (decline in H_2S concentrations). In other words, the loss of a function (H_2S binding function in sulfide-free habitats) can be also driven by positive selection. The dramatic changes of environmental conditions during the course of evolution and the associated physiological modifications in annelids from well-oxygenated emerging habitats are stressed to explain such a loss of function by molecular adaptation.

Materials and Methods

Biological Materials

Juvenile specimens of *Lamellibrachia nov sp*. were collected around cold-seeps from mud volcanoes in the Mediterranean Sea during the French oceanographic cruise Medinaut (Kazan site: 35°25.88'N, 24°33.56'E) at a depth of about 1705 m. Juvenile specimens of the hydrothermal-vent tubeworms *Riftia pachyptila, Oasisia alvinae* and *Tevnia jerichonana* (up to 3-5 centimeters length) were collected at one single vent site from the ridge segment 9°50'N on the East Pacific Rise (Riftia field: 9°50.75'N, 104°17.57'W) at a depth of 2500 m during the French oceanographic cruise HOT 96 and the American cruise LARVE98**.** Hydrothermal vent tubeworms *Ridgeia piscesae* were collected at the Endeavour Segment of the Juan de Fuca Ridge (Clam Bed, 47º 57'N, 129º 05'W) at a depth of 2197 m during the

Canadian-American Hi-Rise cruise 2001. Worms were sampled using the telemanipulated arms of the submersibles 'Nautile', 'Alvin', and 'ROPOS', brought back alive to the surface inside an insulated box, and immediately frozen and stored in liquid nitrogen after their recovery on board.

Identification and characterization of novel extracellular globin and linker genes.

Bailly et al (2002) described the whole set of paralogous extracellular globins in the hydrothermal vent tubeworm *R. pachyptila* following various molecular techniques. In the present paper, globin and linker primer design, total RNA extraction, cDNA synthesis, RT-PCR amplification, PCR-product cloning and clone sequencing were performed as described in Bailly et al (2002). This protocol was subsequently applied to frozen tissues from *R. pachyptila* in order to recover the linker gene, to the three hydrothermal vent species (*Ridgeia*, *Oasisia* and *Tevnia*) to recover new orthologous A2 and B2 globin genes, and to the Mediterranean cold seep species *Lamellibrachia nov sp* to recover the whole set of paralogous globin genes, namely A1, A2, B1 and B2. In addition, partial sequences of extracellular globins from a terrestrial sulfide-free environment of the earthworm *Lumbricus rubellus* were obtained from an Expressed Sequence Tag library (23). Paralogous globin sequences of *R. pachyptila* and *Lamellibrachia nov sp.* were used to estimate neutral evolutionary rates within each globin subfamily, whereas additional orthologous globin genes obtained from closely-related tubeworms and other annelids were used to study the evolution of H_2S binding function in different sulfide-free and sulfide-rich environments.

Multiple alignments and tree reconstructions

Two sets of partial sequences corresponding to the A2 and B2 extracellular globin subfamilies were compiled and aligned by hand and by using the CLUSTALX software (24). Unrooted tree topologies from multiple alignments of the A2 and B2 amino acid globin sequences were obtained using the Neighbor Joining (NJ) method computed using PHYLIP program (25) with 1000 bootstrap resamplings of the data.

The orthologous A2 globin set was composed of 213 bp of DNA sequence (71 codons) from the terrestrial (sulfide-free) oligochaete *Lumbricus rubellus* (BF422675), the littoral transitory (sulfide-rich) polychaete *Sabella spallanzanii* (AJ131285), the Mediterranean cold-seep (sulfide-rich) vestimentiferan *Lamellibrachia nov sp* **(Accession number)** and the three East Pacific Rise hydrothermal vent (sulfide-rich) vestimentiferans *Oasisia alvinae* **(Accession number)**, *Tevnia jerichonana* **(Accession number)** and *Riftia pachyptila*

(AJ439733). The orthologous B2 globin set was composed of 213 bp of DNA sequence (71 codons) from *Lumbricus rubellus* (BF422540), the two littoral polychaetes *Sabella spallanzanii* (AJ131283) and *Sabellastarte indica* (D58418), the Mediterranean cold-seep vestimentiferan *Lamellibrachia nov sp* **(Accession number)**, the East Pacific Rise hydrothermal vent vestimentiferan *Riftia pachyptila* (AJ439737), and the Juan de Fuca Ridge hydrothermal vent vestimentiferan *Ridgeia piscesae* **(Accession number).** Multiple alignment of linkers from *R. pachyptila*, *Lamellibrachia sp.* (P16222) *Nereis diversicolor* (Q25542) *Tylorrhynchus heterochaetus* (P18208 and P18207) *Lumbricus terrestris* (Q9TXF2 and Q9GV76) *Sabella spallanzanii* (Q9BHK0 and Q9BHA3) were obtained with the CLUSTALX software (24) and presented according to the GeneDoc software (http://www.psc.edu/biomed/genedoc/).

Ancestral globin sequence reconstruction

A globin sequence reconstruction for extinct ancestral nodes was performed from both A2 and B2 unrooted tree topologies, using an enlarged set of A2 and B2 globin amino acid sequences. These additional globin sequences belong to *Lumbricus terrestris* A2 (B28151) and B2 (A28151), *Tylorrhynchus heterochaetus* A2 (A24811) and B2 (P13578), *Lamellibrachia sp.*A2 and B2 (26), *Oligobrachia mashikoi* A2 (S72251) and B2 (S72253) and *Sabella spallanzanii* B2 (AJ131284). The distance-based Bayesian method of (27) (software ANCESTOR available at http://mep.bio.psu.edu/ancseq.html) was used to infer ancestral amino acid sequences and thus to follow the emergence of crucial amino acid residues (i.e. free cysteines) within topologies connecting orthologous globins from annelids living in different environments.

2-D and 3-D globin reconstructions

The secondary structure of the molecular domain surrounding the free cysteine residues involved in the sulfide-binding function was predicted using a Hydrophobic Cluster Analysis (HCA plots) for ancestral amino acid globin sequences and plotted according to the DrawHCA software ((28) http://www.lmcp.jussieu.fr/~soyer/www-hca/hca-form.html). These plots were of prime importance to deduce the level of conservation of the sulfide-binding domain in annelids during the course of evolution. The three-dimensional structure of the A2 and B2 subunits was also crucial to visualize the orientation of the thiol group for each free cysteine (i.e. Cys +1 and Cys +11). As no 3-D globin model is available so far for extracellular globins of annelids, two chimeric human-*R. pachyptila* A2 and B2 globins were

constructed. Because the second exonic region encodes the core of the globin fold (29, 30) and the intron positions are conserved between annelid and human globins, it could be easily replaced (central position) in human alpha (P01922) and beta (P02070) globins by those of the A2 and B2 globins (via an artificial exon-shuffling phenomenon). Chimeric globins made of the first human exon, the second *R. pachyptila* exon and the third human exon were successfully submitted to the automated comparative protein modeling server Swiss Model at http://www.expasy.ch/swissmod/SWISS-MODEL.html (31) in order to get 3-D representations of partial A2 and B2 globin subunits from *R. pachyptila*.

Evolutionary rates from amino acid and nucleotide substitutions

In their first paper, Bailly et al. (2002) showed the occurrence of significant differences between evolutionary rates of the four globin subfamilies, namely A1, A2, B1 and B2, using pairwise comparisons of amino acid sequences between two closely-related vestimentiferan tubeworms. In the present paper, evolutionary rates were estimated between amino acid sequences of the same two genera for the hexagonal-bilayer hemoglobin linker chain as this protein is suspected to have coevolved with the A2 and B2 globin chains. Genetic distances between *R. pachyptila* and *Lamellibrachia nov sp* orthologous linker chains (**accession number)** were calculated according to the Jones, Taylor and Thornton (1992) matrix using the Maximum Likelihood procedure of the Puzzle software (32) . In parallel, neutral evolutionary rates of A1, A2, B1 and B2 paralogous globin subfamilies were estimated using nonsynonymous amino acid substitutions between *Lamellibrachia nov sp* and *R. pachyptila* nucleotide sequences. The likelihood expected substitution rate per site was calculated for each globin subfamily according to the (33) model of substitutions using the Puzzle software. In this case, genetic distances directly relate to the evolutionary rate as evolutionary divergence time is identical between paralogous strains of the same pair of species (34).

Search for Darwinian positive selection from dN/dS ratios

In order to detect Darwinian positive selection acting on extracellular globins (i.e. acceleration of the fixation rate of non-synonymous mutations in proteins), an approach based on the maximum likelihood estimation of the non-synonymous/synonymous substitution rate ratio (dN/dS = ω (35)) was applied using our sets of coding sequences. The ratio ω provides a sensitive measure of selective pressures acting at the protein level, with ω values of <1, =1, and >1 indicating purifying selection, neutral evolution, and diversifying selection, respectively.

Three hierarchical steps relying on codon-based likelihood models were considered in order to conduct selection pressure analysis for both the A2 and B2 subfamilies. All of these models use Maximum Likelihood (ML) methods for estimating the parameters of a transition matrix describing the substitution rate between pairs of codons, including dN/dS ratios (ω), transition/transversion ratios and branch lengths: (1) the lineage-specific model (5) (36) (2) the site-specific model (6) (37) and (3) the branch-site model (38). For each step we compared nested models (the null and the alternative hypothesis) with the Likelihood Ratio Test (LRT) following the formula: $2\delta L = 2(L_1 - L_0)$ where L_1 is the alternative hypothesis and L_0 is the null hypothesis. The log likelihood difference between the two models is expected to be χ^2 distributed with the number of degrees of freedom equal to the difference in the number of parameters between the models.

The lineage-specific model

Three models allowing ω ratios to vary between the branches of the tree but not to vary between amino acid sites were implemented. Model A (one-ratio model) assumes and estimates a single average ω ratio for all branches and can be compared with Model B using LRT. This latter model assumes in addition to the average ω ratio a second independent ω ratio for one of the lineages. As opposed to these two first models, model C (free-ratio model) assumes a different ω ratio for all branches of the tree to detect putative positive selection among lineages or internal nodes. This latter model can be compared to the nested model B using a LRT.

The site-specific model

Five models (M1, M2, M3, M7 and M8) allowing ω ratios to vary between amino acid sites but not between lineages were used to seek for the occurrence of sites under positive selection. The M1 model (neutral) only assumes two site classes in which $\omega_0 = 0$ (any mutation is deleterious at a given site) and $\omega_1 = 1$ (any mutation is neutral at a given site) and for which the proportions p_0 and p_1 could be estimated over the whole protein. This first model could be compared with the M2 model (selection) using a LRT since the latter adds a third site class with an estimated free ω_2 ratio. The model M3 (discrete) uses a general discrete distribution of the ω ratios among sites with three site classes for which the proportions p_0, p_1 and p_2 are estimated. This model can be tested against model A (one-ratio) using a LRT. The model M7 (beta) assumes a beta distribution of the ω ratios among sites within the interval 0- 1 and can only be compared with the model M8 (beta & gamma) which adds a new site class of an estimated free ω ratio outside the upper boundary of the interval. M2, M3 and M8 correspond to alternative models allowing the occurrence of sites with $\omega > 1$,

and which are respectively nested to the null hypothesis models M1 (neutral), M0 (one-ratio) and M7 (beta).

The branch site-specific model

Lineage-specific models (with an *a priori* prespecified branch that is suspected to be under positive selection (model B)) combined with site-specific models (models M2 or M3), provide new interesting tools to test for positive selection at each amino acid site within a prespecified lineage (foreground branch) as opposed to the rest of the lineages (background branches). These models, Bm2 and Bm3, can be respectively compared for LRT with models M1 (neutral) and M3 (discrete). For site-specific and branch site-specific models and especially models M2, M3 and M8, the Bayes theorem is (automatically) used to calculate posterior probabilities of site classes for each site. Sites displaying $\omega > 1$ and a high posterior probability can be suspected to be under diversifying selection.

All analyses were performed using the CODEML program of the PAML package (39) available at http://abacus.gene.ucl.ac.uk/software/paml.html. Files containing the aligned sequences are available from the author upon request.

Results

Thiol positions in the 3-D structure of the A2 and B2 globins

In their previous paper, Bailly et al. (2002) suggested that the sulfide-binding domains (SBD) of the extracellular globins A2 and B2 from *R. pachyptila* may be involved in two different binding functions, namely sulfide detoxification and sulfide transport to the endosymbionts located in the trophosome. Slight differences in their sulfide affinity constants could thus be reflected in both the molecular environment of the free cysteine residues and the thiol orientation in relation to the heme. In order to test such a hypothesis, chimeric 3-D globin reconstructions of A2 and B2 were performed. The replacement of the second exon (71 amino acids) of the human beta globin (half the protein) by the second exonic region of the A2 and B2 extracellular globins of *R. pachyptila*, allowed us to obtain a putative crystal structure of these sulfide-binding subunits (Fig. 1).

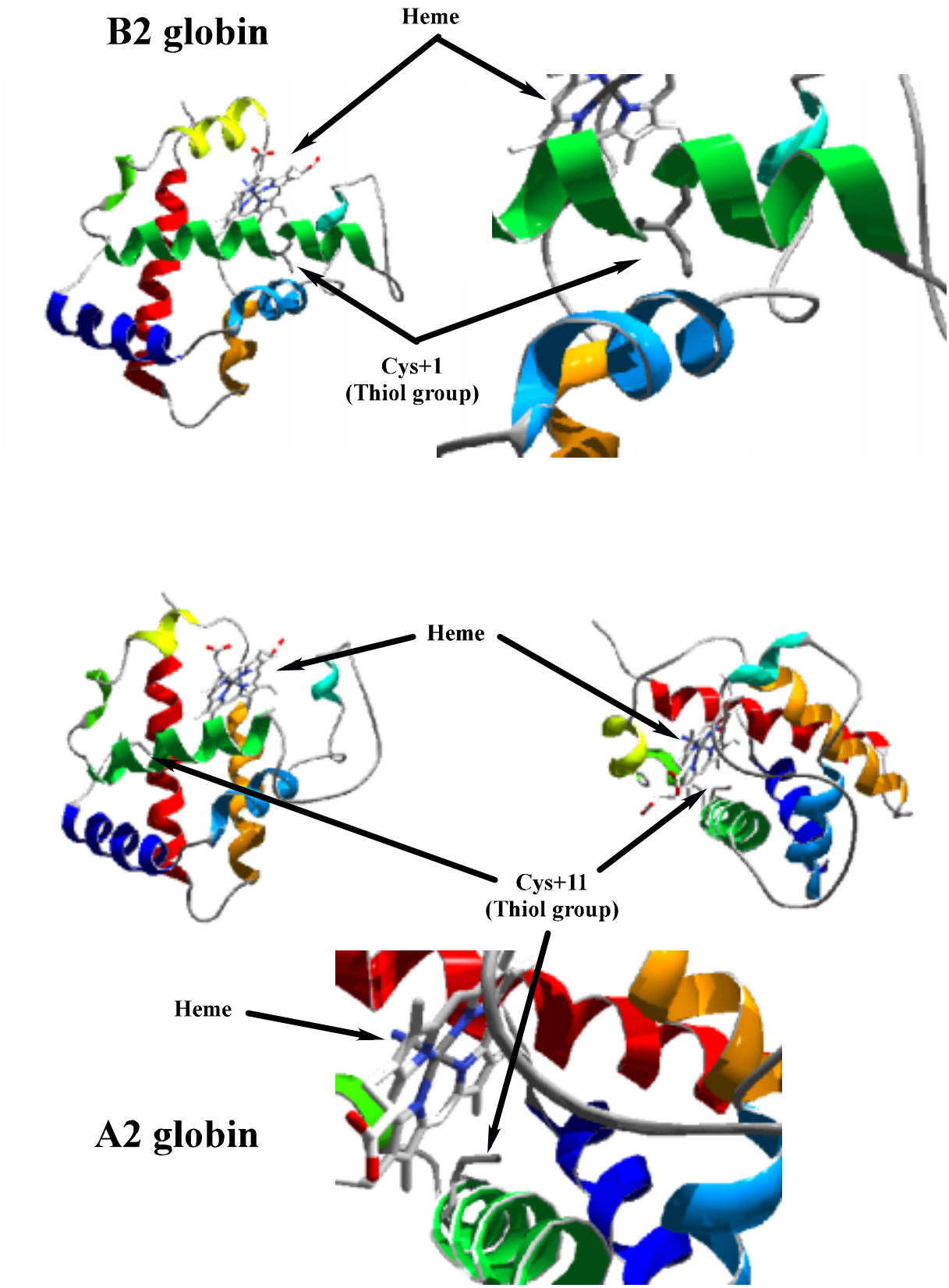


FIG.1. Putative three-dimensional representation of the *R. pachyptila* B2 and A2 globin fold. A global view of the entire subunits and a close-up of the thiol group region allow visualization of the spatial orientation of the cysteines in the molecule and their position relative in relation to the heme.

Even if such a process is surprising, results are in agreement with a vast number of previous studies showing a high level of conservation of the globin fold, even between unrelated species. This chimeric globin allowed us to localize the free cysteines and to orient the thiol group in opposition to the heme pocket: Cys +1 and Cys +11 present two different molecular orientations, and thus accessibilities.

New globin sequences from hydrothermal vent and cold seep species

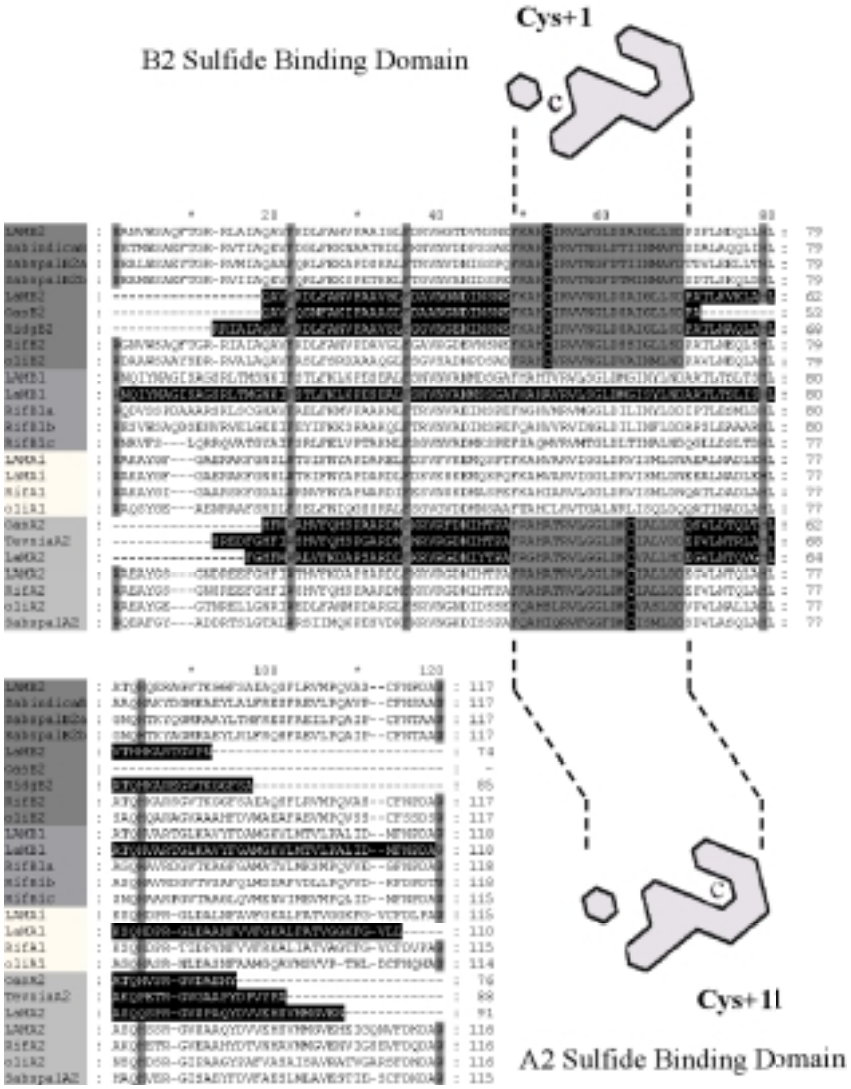


FIG. 2. Multiple alignment of partial globin amino acid sequences from *Riftia pachyptila* (Rif), *Sabella spallanzanii* (Sabspal), *Oligobrachia mashikoi* (Oli), *Lamellibrachia sp.* (LAM), *Lamellibrachia nov sp.* (LaM), *Oasisia alvinae* (Oas), *Ridgeia piscesae* (Ridg) and *Tevnia jerichonana* (Tev). The new globin sequences from *Lamellibrachia nov sp.* (LaM), *Oasisia alvinae* (Oas), *Ridgeia piscesae* (Ridg) and *Tevnia jerichonana* (Tev) are colored in black. The two shaded portions containing the B2 globin cys+1 and the A2 globin cys+11 sites represent the two putative Sulfide Binding Domains we previously defined in Bailly (2002)

Five partial sequences of both the A2 and B2 globins have been obtained for *Tevnia jerichonana* (A2: 88 a.a.), *Oasisia alvinae* (A2: 76 a.a. and B2: 53 a.a.) and *Ridgeia piscesae* (B2: 71 a.a.). Globin assignment was based on sequence homology, specific amino acid patterns and the presence of free cysteine residues Cys +1 and Cys +11 (see (19)) without ambiguity (see Fig. 2). A complete set of the globin subfamilies A1, A2, B1 and B2 was also sequenced from *Lamellibrachia nov sp*, a tubeworm inhabiting Mediterranean mud volcanoes. As above, globin sequences were unambiguously assigned to the right paralogous subfamily by the reconstruction of molecular phylogenies in which the well-defined globin subfamilies of *R. pachyptila* and *Lamellibrachia sp* from Japan (14) (26) (19) were inserted (Fig. 3).

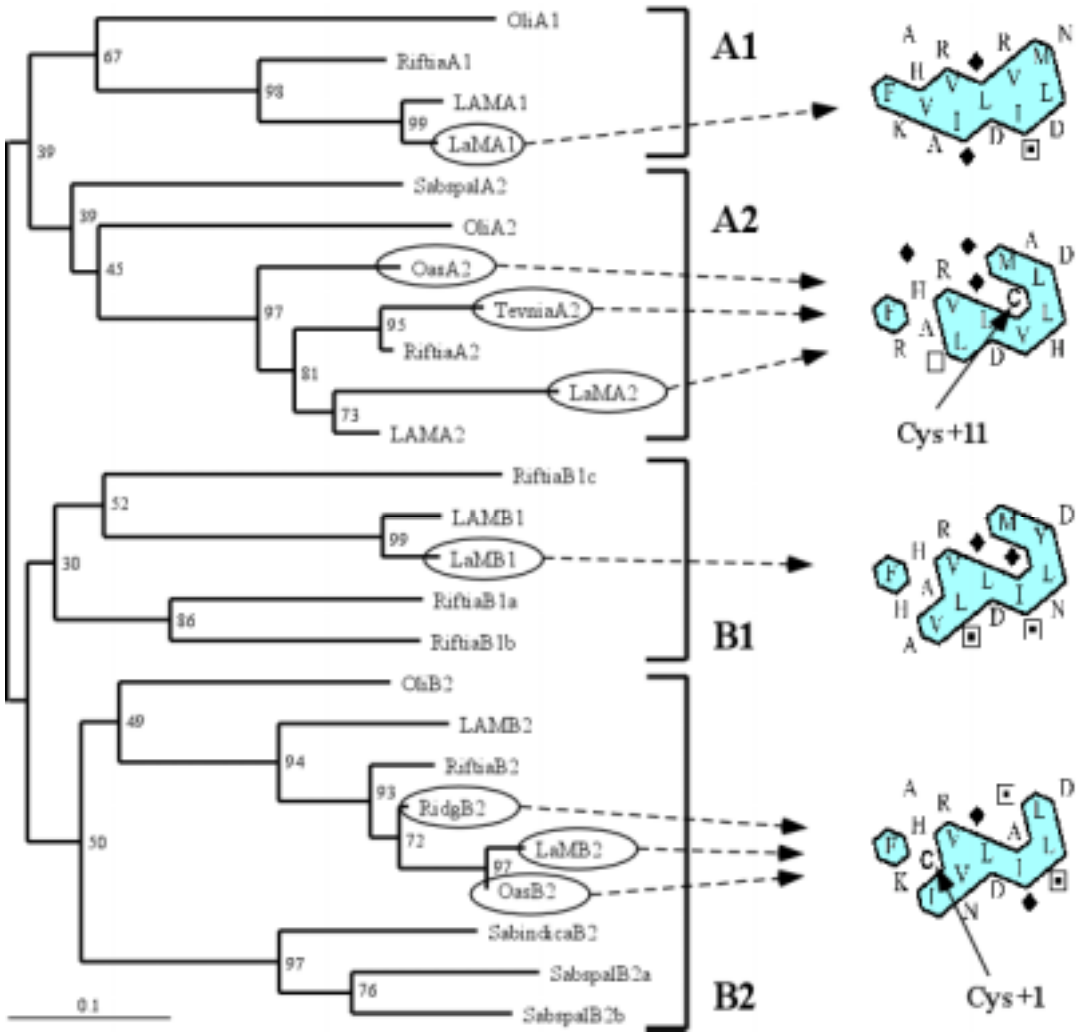


FIG. 3. Neighbor-joining consensus tree with percentage bootstrap support (1000 replicates). Rooting is performed according to the 'duplicate rooting procedure' (65) using clade A versus clade B. The first dichotomy separates the A and B families and the second distinguishes the A1, A2, B1 and B2 subfamilies. Typical

subfamily HCA plots (secondary structure) were performed from each new *Lamellibrachia m sp.* Free cysteines (Cys+11) in A2 and (Cys+1) in B2 globins are shown by an arrow.

All of the A2 and B2 globins of vestimentiferan tubeworms displayed free cysteine residues at positions Cys +11 and Cys +1 respectively and similar Sulfide-Binding Domain amino acid patterns. In addition, one of the four linker chains of the hexagonal bilayer hemoglobin of *R. pachyptila* was partially sequenced (101 a.a.). This sequence is orthologous to the *Lamellibrachia sp* linker chain **(accession number)** with a high level of sequence identity (90%) and similarity (83%). A multiple alignment was performed with other linker chains from coastal and terrestrial annelids (see Fig. 4).

```
                  *        20         *        40         *        60         *
Riftia    : ------------------------------------------------------------------- :   -
Lamelli   : ---------AAV----QPLSVSDAMGARVDAQAWRVDR-LTKQFQAISDAADTSIGAAKSGGDIARH------ :  53
Nereis    : -------DDCVC----PGGRQWAGIASKADAQEARINR-LAGRVEAVAEKLRVGGEKVKHFLGDF-------- :  53
Tylor1    : -------DDCVC----PGGREWASVASKADSQEARVNR-LAGRVEALAENLRAGGDRLSHYKSEF-------- :  53
Lumterr1  : ----------DS----PPAQSHDEIIDKLIERTNKITT-SISHVESLLDD-RLDPKXIRKAG---------- :  46
Tylor2    : GALGDRNGDCACDRPSPRGYWGGGMTGRSAFADPHIAEGRLANQDARLGTLEEEVDKLQHKYDDFIAGKTARR :  73
Sabspa1   : -------MRSLFLLLVVGIAFGNGALSRTRRSSSWGSSLTTARVDSQEMRLDQLAKQLDSLEGSLEGGDLGIP :  66
Lumterr2  : -------MWYVLGLMLVGLAAGASDPYQERRFQYLVKN-QNLHIDYLAKKLHDIEEEYNKLTHDVDK------ :  59
Sabspa2   : --------MRALLLMCIGAALVAAAVSDNCPECDKPDPATMXKLDENAARLNRLVARAVDVSDTIKRRQSEYE :  65

                       80         *       100         *       120         *       140
Riftia    : ----------------------------------------CDGVNDCKNHYDEDSVRCVVPMV-ANSTWIG- :  30
Lamelli   : MLN-------SHLDDHW---CPSKYHRC-GNSPQCMSNMAFCDGVNDCKNHFDEDENRCVVPVT-ANSTWIG- : 113
Nereis    : ---RELEFRVDELEGNG---CDARHYQCGGNTPYCIADTLVCDGSPDCPNGSDESEDICHIPIP-AGTVLIGH : 119
Tylor1    : ---RELEFRVDELEGNG---CEPRHFQCGGSAMECISDLLTCDGSPDCANGADEDSDVCHIPPI-AGTLLVGH : 119
Lumterr1  : ----SLRHRVEELEDPS---CDEHEHQCGGDDPQCISKLFVCDGHNDCRNGEDEKD--CTLPTK-ADKFIG-- : 107
Tylor2    : ERFAQLKDRVWGLEAHH---CDDDHLSCKDVAFTCIGHNLVCDGHKDCLNGHDEDEETCSIAAS-VGSSFEG- : 141
Sabspa1   : ASMNSLSLRLDKIEGND---CGKRQFSCGRHSGQCVFELNVCDNVPDCANGMDEAEETCSNPVP-AGRSWNG- : 134
Lumterr2  : KTIRQLKARISNLEEHH---CDEHESECRGDVPECIHDLLFCDGEKDCRDGSDEDPETCSLNITHVGSSYTG- : 128
Sabspa2   : DAEVQAQNMLQRLEKRNGVLCKG-QFVCGGAAQQCADNLVVCDGVNDCDNGADEATDICNIKAK-VGDSWVG- : 135

                    *       160         *       180         *       200         *       22
Riftia    : -YPAYDHCTQRRPYEMIISVTSAPSSSVYKVHQPLKVQV--DLFSKNHGVKQSASLTGDAYYCKGSQRLLHCP : 100
Lamelli   : -YPAYDHCTQRRPYEMVISITSAPSDIVYKVHQPLKVQV--DLFSKKGGLKQSASLHGDAVYCKGSQRLIVAP : 183
Nereis    : LNTDHDFCTRRKPTEMDLIITSVTRSKYFQSRLQVKANL--DLKYTAEGAEQEDVLPVEGYYNFGTHQLVILP : 190
Tylor1    : LNTDHDFCTKRKPNEFDLFISSVQRSSYFQSRLKVKGNL--QIKYTAEGRDQEDVLQVKGYYNFGTHQLVILP : 190
Lumterr1  : -DVVFDFCTKRRPEHMTLAFESS----------------------PADGE-----------YSFADHRLTIHP : 146
Tylor2    : QIRQLDTCTKRKPSAFRFIITNVDVPKYFPQEPHVKATI--LMTSSKDGHETQSSLAVDGVYDFTHRKVILYS : 212
Sabspa1   : -VAHWDSCLVRSNNKMNVLITGNKRYDYFPSRLWVDAVI--IESYTVSGEEQTNTWSISGFYNYASRMVKFNS : 204
Lumterr2  : -LATWTSCEDLNPDHAIVTITAAHRKSFFPNRVWLRATLSYELDEHDHTVSTT---QLRGFYNFGKRELLLAP : 197
Sabspa2   : -SFVYHNCSSAKFSNLRFELKKFERKPNMMSLADVSAML-FWDDNTDADVTFIDSLPYDGTYKFYSRSAHLRS : 206
```

```
          0         *       240         *       260
Riftia   : P---------------------------------------------- : 101
Lamelli  : P--EDDRLEIIGQFDGVSNDRFKGYIVREMSGDKCAEFRFFKQ----- : 224
Nereis   : P--ENDRLGLVCSFRAGNDNRCSAAIVHEASLAHCGDDFVFVRKDDH- : 235
Tylor1   : P--EDDRLGIVCNFRAGNDDRCRAHIVHEASLEHCGDDFVFVKEDDHH : 236
Lumterr1 : P--EEDGLGLVGEFDGYNFDRFVGHIVHELSEEVC-AEFIFHR----- : 186
Tylor2   : P--DKDNLIFECTFPRHDNNHCKGVMKHSGG-DVCLTFTLERID---- : 253
Sabspa1  : P--KEGTMGIGLTCSDAGSDVAQCDFVRATDQHVCGSVTLTRGA---- : 246
Lumterr2 : LKGQSEGYGVICDFNLGDDDHADCKIVVPSSLFVCAHFNAQRY----- : 240
Sabspa2  : IMVVNSSRKCSTSYPER-LTSCITMSTLLTAKRLC------------ : 240
```

FIG.4. Multiple alignment of linker amino acid sequences from *Riftia pachyptila* (Rif) in black, *Sabella spallanzanii* (Sabspa1: Q9BHK0; Sabspa2: Q9BHA3), *Lamellibrachia sp.* (Lamelli: P16222), *Tylorrhynchus heterochaetus* (Tylor1: P18208; Tylor2: P18207), *Lumbricus terrestris* (Lumterr1: Q9TXF2; Lumterr2: Q9GV76) and *Nereis diversicolor* (Nereis: Q25542). Cysteines colored in black are included in a LDL (Low Density Lipoprotein receptor) cysteine-rich conserved domain colored in grey (66).

Ancestral sequence reconstruction from A2 and B2 tree topologies.

The structural evolution of the SBD and especially the presence/absence of the free cysteines at positions Cys +1 and Cys +11 were followed along topologies of the orthologous sets of A2 (Fig. 5) and B2 (topology not shown as the evolution of the SBD parallels that of the A2 strains) for an enlarged set of terrestrial annelid amino acid sequences.

Ancestral sequence reconstruction and therefore the ancestral 2-D structure of the SBD predicted by the HCA-plots allowed us to show that free cysteines and their molecular environment have been maintained in ancestors inhabiting sulfide-rich habitats but were not present after the node preceding the evolutionary strain leading to marine sulfide-free polychaetes (*Tylorrhynchus heterochaetus*) and terrestrial sulfide-free oligochaetes (*Lumbricus terrestris* and *Lumbricus rubellus*).

Evolutionary rates

One genetic distance based on the linker amino acid sequences of *R. pachyptila* and *Lamellibrachia sp*. under the Dayhoff matrix (40) was added to those previously calculated for globins of the same pair of species in Bailly et al. (2002). This distance is quite similar (**0.20**) to those obtained for A2 (**0.20**) and B2 (**0.22**) globins between the same species. The structural (linker) and the functional (A2 and B2) globins of the hexagonal bilayer hemoglobin appeared to similarly evolve at the lowest evolutionary rate as opposed to the other two (A1 and B1) globin subfamilies (Table 1). This seems to indicate that these two functional and structural proteins may have coevolved since the formation of this macromolecule.

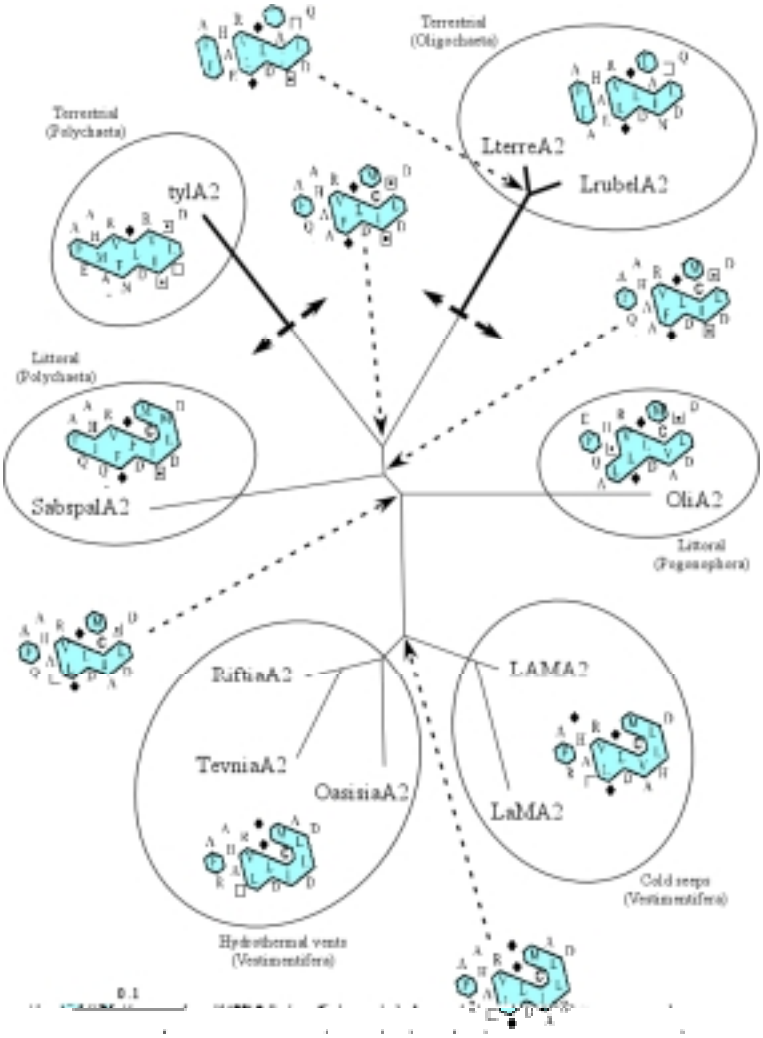


FIG.5. Ancestral amino acid sequence reconstruction from A2 orthologous globin cluster topology. *Riftia pachyptila* (Riftia), *Sabella spallanzanii* (Sabspal), *Oligobrachia mashikoi* (Oli), *Lamellibrachia sp.* (LAM), *Lamellibrachia nov sp.* (LaM), *Lumbricus terrestris* (Lterre), *Lumbricus rubellus* (Lrubel), and *Tylorrhynchus heterochaetus* (tyl). HCA representations of partial reconstructed sequences (Sulfide Binding Domain with free cysteine site +11) have been plotted for each node (dotted arrows). Modern A2 globin HCA patterns are plotted for each environmental habitat. **Bolded topology** from dotted arrows in terrestrial polychaete and oligochaete lineages indicates the absence of free cysteine +11.

A second set of genetic distances has also been calculated from synonymous substitutions for globin subfamilies A1, A2, B1 and B2 between *Riftia pachyptila* and *Lamellibrachia nov sp* (Table 1). It is worth noting that differences observed between distances and thus between evolutionary rates when comparing the four paralogous amino acid sequences (19) have been greatly attenuated when using synonymous substitutions and range between 0.30 and 0.54

substitutions per codon. Differences between A1 and A2 are slight, while those between B1 and B2 are more drastic.

Table 1. Maximum likelihood estimates of the expected number of mutations per site (genetic distance) between orthologous nucleotide and amino acid sequences of globins and between amino acid sequences of a linker chain of *R. pachyptila* and *Lamellibrachia sp*.

Gene	Genetic distance (synonymous substitutions)*	Genetic distance (amino acids)**
A1 Rif / LaM	**0.45**	**0.38****
A2 Rif / LaM	**0.33**	**0.22****
B1 Rif / LaM	**0.54**	**0.79****
B2 Rif / LaM	**0.30**	**0.20****
Linker Rif / LaM	no data	**0.20**

* estimates based on the number of codons to allow for comparison with distances based on amino acids.
** from Bailly *et al*.2002

dN/dS ratio analyses

The dN/dS ratio analyses were performed solely on the two globin subfamilies A2 and B2 as they are both involved in the sulfide-binding function via the free cysteine residues Cys +11 and Cys +1 in annelids living in sulfide-rich habitats. All model parameters (fixed or estimated), LRTs and the putative positively selected sites are reported in tables 2 and 3 **(annexe 7)**. The small number of lineages analyzed and the estimation of the frequency of codon utilization in vestimentiferan globins prevented us from using the model of equal codon frequencies (the f61 model in PAML package). For all models, the equilibrium frequencies were estimated from the products of the average observed nucleotide frequencies at the three codon positions (the f3x4 option in PAML package).

1) the lineage-specific models

Lineage-specific models have been compared using LRTs. Model C does not differ significantly from model A. Moreover, the intermediate model, (model B) in which an independent ω_1 is estimated for a specific lineage (or branch: e.g. the *Lumbricus rubellus* lineage) and another uniform one, ω_0, is estimated for the remaining lineages (or branches: e.g. hydrothermal vents, cold seeps and littoral annelids), is not significantly different from the two other extreme models (i.e. models A and C). These tests do not support the hypothesis

that the neutral model (model A) or the free-ratio model (model C) fit the tree topologies of both A2 and B2 globins better than the *a priori* more adjustable one (model B), for which each lineage could be tested separately. However, all of these models use ω values averaged over all sites of each pairwise comparison of globin sequences and thus are not sensitive enough to detect a slight selective effect on a given amino acid site.

2) the site-specific models

Models that include variable dN/dS ratios among sites but constant dN/dS ratios among lineages indicate a tendency towards a non-neutral evolutionary rate along the A2 and B2 globins sequences among lineages. This phenomenon of selective pressures affecting the A2 and B2 subfamilies differentially is showed by the distribution of posterior probability plots that belong to the ω_2 class (from the m2 model) for all amino acid sites in the sulfide binding domain sequences as seen in figure 6.

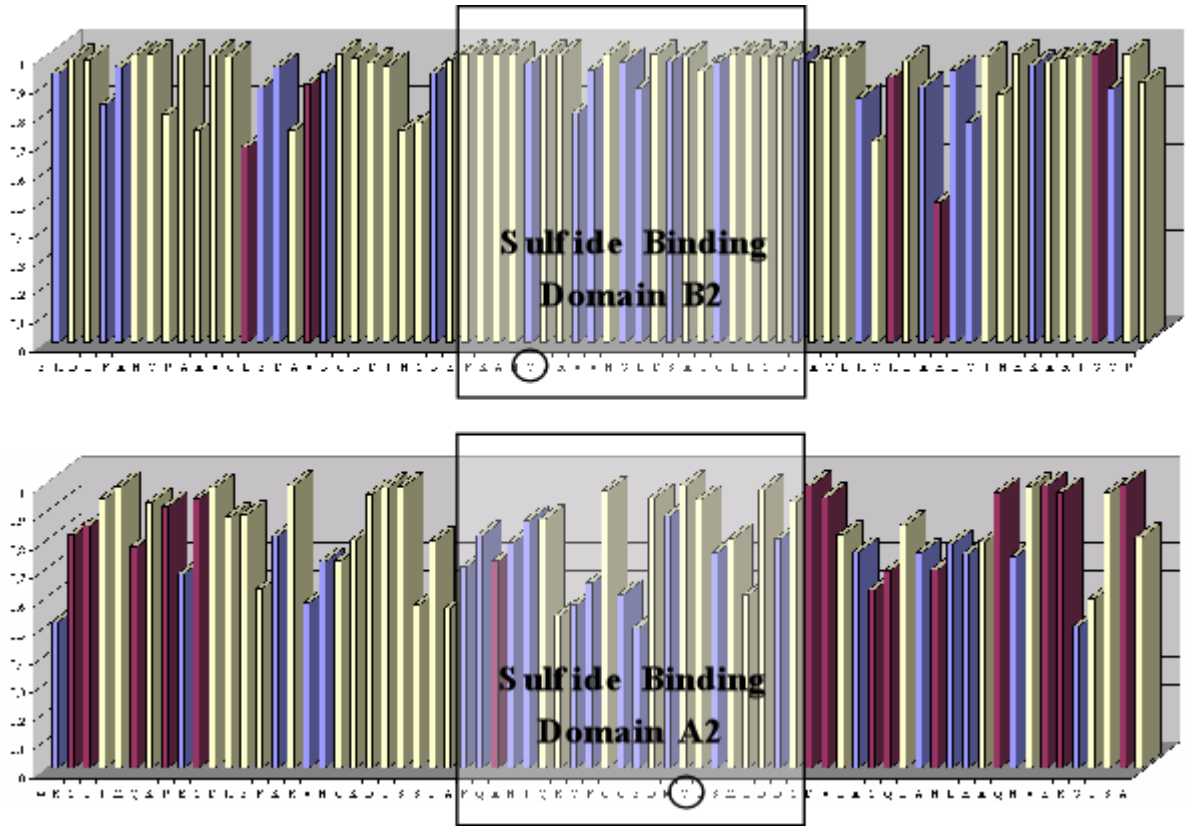


FIG.6. Distribution along the B2 and A2 globin sequences of the probabilities of observing an amino acid site under 1) purifying selection (dN/dS ratio = ω = 0) in white or 2) strong directed selection in grey (ω = 0.14 for B2 and ω =0.21 for A2) and 3) neutral selection in black (ω = 1). This pattern has been obtained according to the m2 model (analysis site by site from the compilation of all B2 and A2 globins). X-axis represents 70 sites flanking A2 and B2 SBD globins (*Lamellibrachia m sp* sequence as reference) and Y-axis represents the values of posterior probabilities (between 0 and 1).

This figure provides two clearly different patterns of ω ratios, with the B2 globin subfamily showing more evidence of selective constraint than the A2 subfamily.

The models M2 and m3 both allow the occurrence of variable selective pressures at a given amino acid site among lineages and present significant likelihood values when compared to their nested null models m1 and A (neutral model), but they do not display positively selected sites. The m8 model (beta and gamma distributions) fits the globin topologies significantly better than the m7 model in the A2 lineages but not in B2, and it was likely to detect a few positively selected sites outside the SBD boundaries.

3) the branch-site models

These two more realistic models that combine lineage-specific and site-specific models seem more sensitive to accurately define the selective state of each amino acid site within a given lineage compared to the others. Results obtained from the Bm2 and Bm3 models indicate that both the A2 and B2 globin subfamilies of the *Lumbricus rubellus* (foreground) lineage, representing annelids from terrestrial sulfide-free habitats, had undergone selection in a different way than (background) those lineages representing annelids living in various sulfide-rich habitats (see Tables 2 and 3 **(annexe 7)** and Fig. 7).

The models Bm2 and Bm3 provide significantly better likelihood values than either lineage-specific models (Bm2 versus A) or site-specific models (Bm3 versus m2), and they were likely to detect positively-selected sites in the foreground lineage within the SBD. It is worth noting that five positively-selected sites out of six (83 %) and 50 % of the positively-selected sites fall into the SBD for the *Lumbricus* B2 and A2 globin subfamilies, respectively. The Bm3 model detects more positively-selected sites than the Bm2 model for both A2 and B2 globins, suggesting that this former model is more appropriate to detect positive selection. Both the Bm2 and Bm3 models also detected more positively-selected sites within the A2 strain than within the B2 strain, supporting our assumption that selection acted differentially on these two subfamilies. In both models and both subfamilies, the key sites for HS- binding (i.e. the free cysteine 32C in B2 and 42C in A2) are subjected to positive selection with a posterior probability of 99% in A2 (with the Bm3 model) and B2 (with the Bm2 and Bm3 models) and 88% in A2 (with the Bm2 model).

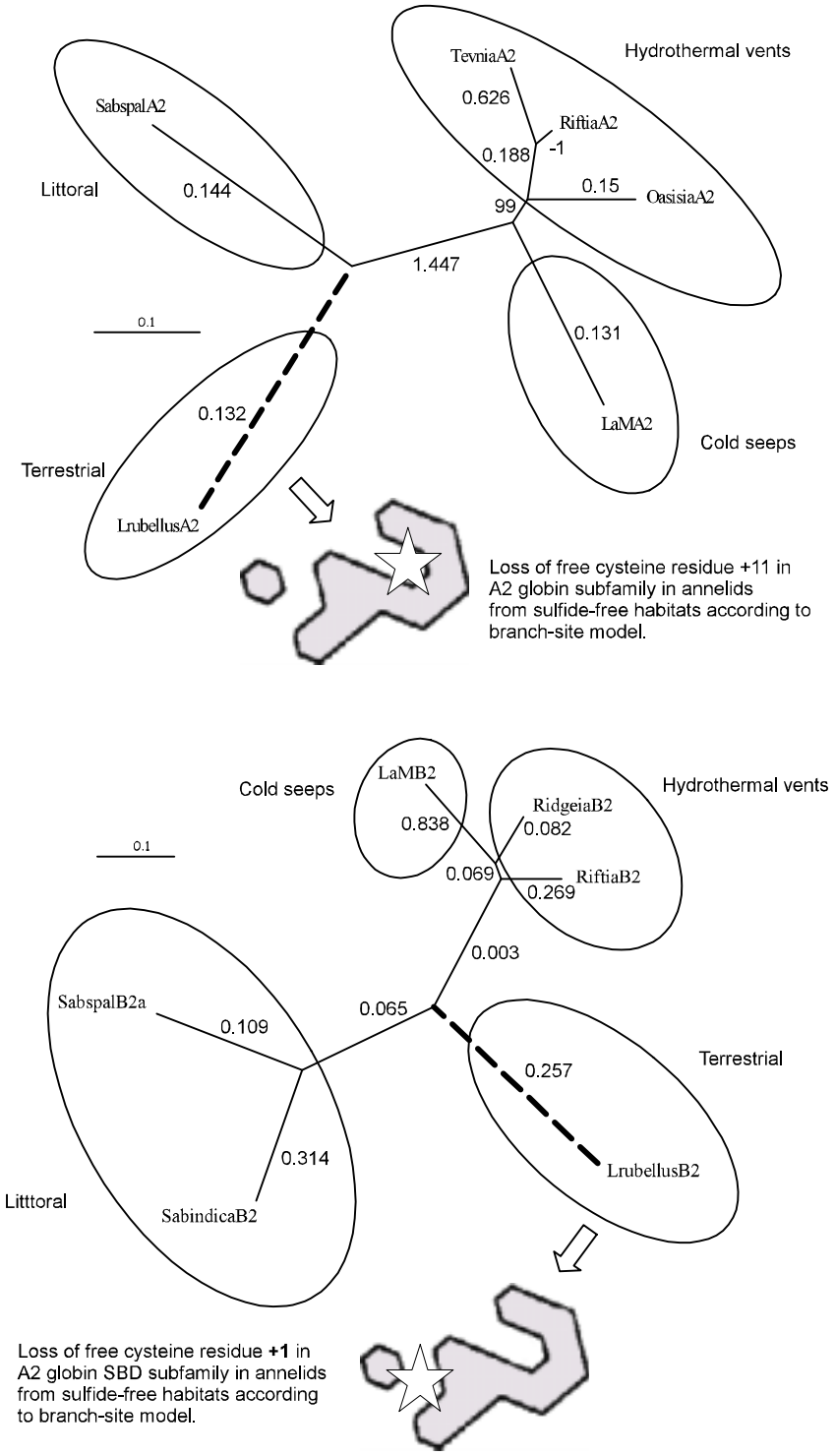


FIG.7. Orthologous globin A2 and B2 topologies from nucleotide sequences of *Riftia pachyptila* (Riftia), *Sabella spallanzanii* (Sabspal), *Sabellastarte indica* (Sabindica), *Lamellibrachia nov sp.* (LaM), *Lumbricus rubellus* (Lrubel) *Tevnia jerichonana* , *Oasisia alvinae* and *Ridgeia piscesae*. Dotted line represents the oligochaete lineage (sulfide-free habitats) where Cys+11 and Cys+1 sites are under positive selection (stars in Sulfide Binding Domain HCA representation).

Other analyses using Bm2 and Bm3 with different fixed lineages were also performed but they yielded low and insignificant likelihood scores and no positively-selected amino acid residues within the SBD (data not shown). All of these models were also run with $\omega = 1$ fixed in the remaining lineages (i.e not estimated) yielding similar results.

Discussion

Sulfide binding function: a widespread function in annelids living sulfide-rich habitats that may have been lost in annelids from sulfide-free habitats

The major problem encountered in the study of the evolutionary relationships of paralogous genes of multigenic families, and especially between highly divergent taxa such as invertebrates, is mainly due to the high level of divergence between these genes arisen by ancient gene duplication (41) (42) and their consequent subfunctionalization (43). Bailly et al. (2002) suspected early that the A2 and B2 globin subfamilies of the complex hexagonal bilayer hemoglobin of annelids have undergone strong directional selective constraints driven by high levels of H_2S concentrations in taxa living in sulfide-rich habitats. This would explain their low evolutionary rate, the maintenance of the free cysteine residues at the conserved positions Cys + 1 and Cys + 11, and the maintenance of a conserved SBD secondary structure. The presence of homologous free cysteine residues at positions Cys+1 and Cys+11 in A2 and B2 globins from additional vestimentiferans such as the Mediterranean *Lamellibrachia nov sp*, and the eastern Pacific *Oasisia alvinae*, *Tevnia jerichonana* and *Ridgeia piscesae,* as well as their absence in the other A1 and B1globin subfamilies from *Lamellibrachia nov sp* (as it has already been described in *R. pachyptila*, *Lamellibrachia sp*. from Japan and other annelid species) demonstrate the widespread occurrence and the level of conservation of these residues in annelids inhabiting sulfide-rich environments. The hydrophobic cluster analysis plots also showed that the Sulfide-Binding Domain associated with these residues is virtually identical between these lineages. These new results therefore strengthen our hypothesis that all of the vestimentiferans share the same ancestral ability to bind H_2S even if the mechanism of H_2S acquisition from the environment could be different from one species to another: *R. pachyptila* uses exclusively its plume for sulfide acquisition (44) whereas basal roots serve this purpose in the seep tubeworms (45). Moreover, Bailly et al. (2002) proposed that the occurrence of free cysteines at positions Cys +1 and Cys +11 is a

plesiomorphic state rather than an apomorphic one occasionally acquired over the course of annelid evolution: it indeed seems rather difficult to explain sulfide adaptation from a single molecular gain of cysteines on globins as sulfide metabolism requires many physiological and molecular modifications, for both symbiotic (*R. pachyptila*) and non-symbiotic (*Alvinella pompejana)* worms. It is thus more reasonable to support the hypothesis of the loss of sulfide-binding function in worms in sulfide-free environments rather than the gain of this function in sulfidic worms.

In order to better characterize the kind and the molecular area of selection acting on A2 and B2 subfamilies, especially on the sulfide-binding function, this present study was performed on a set of A2 orthologous globin genes on one hand and on a second set of B2 orthologous globin genes on the other hand for endosymbiotic and non-endosymbiotic annelids living in sulfide-rich and sulfide-free environments. However, it was not possible to discriminate between the loss and the gain of the free cysteine residues from the ancestral amino acid sequence reconstruction as these sequences have been estimated from an unrooted tree which does not provide any information about the sense of evolution. Nonetheless, in both the A2 and B2 globin strains, the node leading to the polychaete and oligochaete lineages inhabiting sulfide-free habitats, which thus represents the ancestral globin, possesses the free cysteine residues and the associated SBD. Together with the fact that other polychaetes from sulfidic environments such as *Alvinella pompejana* and *Arenicola marina* also possess free cysteine residues (22) , these results suggest that the SBD may have been lost in annelids during the emergence of sulfide-free habitats.

Both linker and the A2 and B2 globin chains display similar low evolutionary rates compared to A1 and B1 globins: co-evolution via similar selective patterns and protein –protein interaction.

Bailly et al. (2002) previously reported that the evolutionary rates of amino acids of the four globin subfamilies of *R. pachyptila* greatly differed from each other and proposed that such differing rates may result from both different selective pressures and a molecular clock-like behavior. Interestingly, the structural chain which allows the cohesion of the hexagonal bilayer structure, the so-called linker, also displays an amino acid evolutionary rate as low as that of A2 and B2 globins. From these data, one can suspect a simple coevolution effect due to structural constraints imposed by the globin-linker interaction (46) (47). However, it is important to underline that the *R. pachyptila* linkers have a functional role of sulfide fixation

in forming persulfide groups (R-SSH) by the cleavage of the linker disulfide bridges that acts as an additional sulfide-binding mechanism (16) : this property enables *Riftia* to bind more sulfides in the vascular blood and the coelomic fluid than just via the S-sulfohemoglobin formation that uses the free cysteines Cys +1 and Cys +11. This different but functional sulfide-binding site could thus be also highly selected for its contribution to either the transport or the immobilization of sulfides in vestimentiferans. It follows that the sulfide-binding ability alone could be sufficient to explain these highly different evolutionary rates between the set of paralogous globins and thus sustain the role of selective forces associated with the presence of H_2S for the conservation of free cysteines. This assumption in favor of differential selective pressures acting on each paralogous globin clade, which probably hides a paradoxical clock-like behavior, is reinforced by the calculation of new rates based on synonymous codons from pairwise comparisons of orthologous sequences of *R. pachyptila* and *Lamellibrachia nov sp*. for the four globin subfamilies. These comparisons showed that background evolutionary rates are similar between the four paralogous clades and that a neutral rate occurred throughout the whole globin family. As a consequence, in addition to their common oxygen transport function (heme), each globin subfamily exhibits different but subtle functional constraints which are probably imposed by (1) H_2S immobilization and/or transport to the endosymbionts, (2) the location of globin subunits in the multimeric assemblage of the hexagonal bilayer hemoglobin or (3) some intrinsic globin features such as dimerization, allosteric effects of some amino acids, and maybe unknown cryptic functions not yet discovered.

Seeking for positively-selected sites in well-conserved orthologous sequences

Hemoglobin is an ubiquitous molecule found “from bacteria to man” (48) (49) for which the three-dimensional structure (globin fold) (50) is well-conserved among highly divergent evolutionary lineages. This structural universality demonstrates that globin is a strongly selected molecule, due in part to its ability to transport and bind oxygen. By restricting our observation to the A1, A2, B1 and B2 globin subfamilies of annelids living in sulfidic conditions, one can expect under a model of neutral evolution or under homogeneous selective pressures a higher similarity between two sister-duplicates (.i.e. A1 and A2 or B1 and B2) than between A1 and B1 or A2 and B2, for instance. Despite their great evolutionary distance in terms of common ancestry (two duplication events), the A2 and B2 subunits share an obvious secondary structure conservation of the Sulfide-Binding Domain and the

maintenance of functional free cysteines in annelids living in sulfide-rich conditions, two noteworthy insights of a functional coevolution between these duplicated genes. Thus, it was more appropriate to seek for selected sites in such a well-conserved structure than to try to detect a cryptic adaptive evolution over the whole set of globin sequences. However, the lineage-specific models were not able to discriminate significantly between either a homogeneous (A) or a different (C) dN/dS distribution (neutral, positive or negative selection) acting between lineages, even if a different rate of nonsynonymous substitution accumulation is assumed in the sulfide-free lineages. Nevertheless, the m2 site-specific model exhibited putative positively-selected sites toward the periphery of the Sulfide-Binding Domain for both the A2 and B2 globin subfamilies. This lack of apparent positive selection in the A2 and B2 globin strains may reflect either that these globins are highly negatively selected even in sulfide-free lineages or that the accumulation of positively selected amino acid sites occurred in ancient times before the annelid radiation, leading to this unusual, complex, extracellular hemoglobin. The lineage-specific and site-specific models indeed average the dN/dS ratio over all sites of a given sequence and all lineages respectively and are not sensitive enough to detect selective pressures at a specific amino acid of a given lineage. This lack of power was already pointed out by (10) when episodic diversifying selection has only occurred on a few amino acids of a strongly negatively-selected molecule. This may explain why most examples of diversifying selection come from xenobiotic recognition molecules or genes associated with male reproduction that have undergone recent adaptive changes such as primate lyzozymes (51) (5), HIV membrane proteins (6) (7) or the salmon gamete recognition system (9). Interestingly, patterns of amino acid sites under purifying, deleterious or neutral selection are different between the A2 and B2 subunits (considering that this pattern has been obtained from the site-specific model m2). In the B2 sequences, all sites are strongly negatively selected whereas in the A2 sequences the selection pressures are less homogeneous and showed weak posterior probabilities. This strongly indicates different levels of selection pressures between paralogous globin strains and thus reinforces the subfunctionality consequences of duplication events, which in this case may be explained by subtle differences in the free cysteine properties. This last point is reinforced by the different location of the free cysteines in A2 and B2 globin chains and the spatial orientation of their respective functional thiol group.

Positively-selected sites in sulfide-free lineages: Why the loss of a function could also be considered as a molecular adaptation rather than a relaxation of constraints (neutral evolution)

Bailly et al. (2002) proposed a possible relaxation of selective constraint on the annelid globins from sulfide-free environments to explain the absence of free cysteine residues on the A2 and B2 subunits. H_2S was one of the most abundant molecules on the Earth's surface during the prebiotic era and its concentration greatly decreased during the course of evolution, remaining only at some specific places (e.g. hydrothermal vents) as a relic of the primitive conditions (52) (53). Thus, the acquisition and the maintenance of H_2S binding functions were of a prime importance during the beginning of life and were probably subjected to purifying selection in the ancient annelid lineages living in sulfide-rich habitats. To test whether the Sulfide-Binding Domain and especially the free cysteine residues in globins from annelids in sulfide-free environments were subjected to ancient diversifying selection, we used the new branch-site set of models implemented in PAML (38). These more intricate and realistic models simultaneously combine different ω ratios among sites and lineages and *a priori* evolutionary hypotheses within a prespecified lineage: in our case, the hypothetical loss (versus the gain) of free cysteine residues in sulfide-free lineages was combined with the way that they were lost during evolution (neutral relaxation or Darwinian positive selection). For both the A2 and B2 globins, the significantly lowest likelihood values were obtained when we assumed a different dN/dS ratio in the sulfide-free lineages as opposed to lineages from the other habitats (hydrothermal vents, cold-seeps, coastal or intertidal hypoxic sediments. In only this case (all branch combinations have been tested: results not shown), the sites Cys +1 and Cys +11 were clearly subjected to positive selection with a strong posterior probability together with other residues of the Sulfide-Binding Domain. These positively-selected sites, found only within the SBD, are in agreement with the loss or the degeneracy of the hydrophobic secondary structure surrounding the free cysteine residues in annelids from terrestrial sulfide-free habitats (see (19)).

The high posterior probabilities obtained for both the Cys +1 and Cys +11 sites are associated with a dN/dS >1. This result rules out the assumption that the loss of the free cysteine residues followed a neutral pattern of evolution. Free cysteine residues are key amino-acids in biochemical reactions because of their well-known highly reactive lateral chain. To our knowledge, all free cysteines reported from proteins display an active and diversified functional role. These residues are involved in the binding of some catalytic co-factors (54),

pH-dependant oligomerization (55), macromolecular complexation, and many other molecular detoxifying functions. However, in the case of specific genetic diseases or metabolic disorders, free cysteines can bind atypical ligands. Examples of such ligands include benzoquinone, a known carcinogenic metabolite, in rodent globin, (56), mercury ions inhibiting enzymatic reactions (57), or oligomeric complexation by abnormal aggregation with subsequent physiological modification in mice mutant hemoglobin (58). These three examples among others illustrate the deleterious role of unexpected free cysteines in mutant organisms and indicate that the disappearance of a specific ligand (e.g. H2S) can drastically reduce organismal fitness when conditions have changed.

We assumed that when sulfide concentrations decreased or when annelid ancestors moved from sulfide-rich to sulfide-free habitats, the presence of free cysteines residues could have provoked or induced irreversible troubles and homeostatic disorders. Thus, it is not surprising to find such sites to have undergone molecular adaptation. In addition, the loss of free cysteines on both the A2 and B2 globin subunits in annelids from sulfur free environments confirms their implication in sulfide binding in annelids from sulfur rich environments. This event has probably occurred during a short period of time as, because of their biochemical reactivity, free cysteines would have been rapidly recruited for alternative biochemical uses via a cooperative process in a manner similar to the occurrence of San Marco spandrels in Venice (59). The loss of globin genes in human and apes (60) and their consequent loss of function in Antarctic fishes (61) (62) (63) have already been reported, but our study is the first to demonstrate that the loss of a function could be driven by positive Darwinian selection.

The sulfide-binding function: a long and intricate physiological and molecular acquisition process

An argument in favor of the loss of free cysteine residues is that adaptation to sulfide did not only require cysteine acquisition on respiratory pigments but was a long biochemical pathway establishment by which animals have acquired various adaptive physiological detoxification mechanisms, at the levels of molecules, cells, tissues, and up to the whole organism. It is unlikely that this complicated metabolic process was the result of a preadapted pathway that abruptly shifted to sulfide utilization from using a molecule that possessed similar detoxifying and energy carrier potentials for the purpose of oxidizing a different metabolic compound.

The 3-D reconstruction of the A2 and B2 subunits showed the orientation of the cysteine thiol group in respect to the heme group. Their localization and their molecular environment differ greatly, thus suggesting a different cysteine reactivity. A well-known chemical rule regarding

molecular interaction and covalent bonds is that the protein-ligand association greatly relies on the steric environment imposed by the flanking lateral chains.

One must keep in mind that the symbiotic worms from sulfidic environments use H_2S to fuel their endosymbionts, a situation not encountered in non-symbiotic worms from similar habitats. It is this detoxification function in non-symbiotic worms (a plesiomorphic trait) that could have evolved in symbiotic worms into a H_2S transport function (apomorphic trait). This means that the H_2S-binding capacity is probably an intricate mechanism that acts differentially according to the location of the cysteines and with regard to the heterotrophic or autotrophic status of the annelid. We mentioned earlier that the vestimentiferan linker chains have the ability to bind H_2S via the formation of persulfide groups (16) as a complementary process to the S-sulfohemoglobin formation via the free cysteines. These H_2S fixation sites are probably interconnected within the hemoglobin structure. However, persulfide group formation does not occur in non-symbiotic polychaetes from sulfide-rich habitats (16), and thus it could be more related to H_2S transport.

A precise localization of the 144 globin subunits with regard to the linker chains as well as additional physiological tests are required prior to assigning each H_2S-binding site (molecular environment) to a specific function.

All of these new molecular data together with those from Bailly et al. (2002) form a set of arguments that allow us to argue that annelid ancestors initially lived in sulfide-rich conditions. The use of dN/dS ratio analyses shows that the sulfide-binding function has been secondarily lost by diversifying selection in annelid lineages from sulfide-free environments. In other words, we propose that the loss of sulfide-binding function is a disadaptation process (64) that was a prerequisite to avoid the detrimental physiological effects associated with the collateral activity of free cysteine residues regarding other blood compounds.

1. Kimura, M. (1981) *J Mol Evol* **17,** 110-3.
2. Goodman, M., Moore, G. W. & Matsuda, G. (1975) *Nature* **253,** 603-608.
3. Ohta, T. (1988) *Evolution* **42,** 375-386.
4. Hedrick, P. W., Ginevan, M. E. & Ewing, E. P. (1976) *Ann. Rev. Ecol. Syst.* **7,** 1-32.
5. Yang, Z. (1998) *Mol Biol Evol* **15,** 568-73.
6. Nielsen, R. & Yang, Z. (1998) *Genetics* **148,** 929-36.
7. Zanotto, P. M., Kallas, E. G., de Souza, R. F. & Holmes, E. C. (1999) *Genetics* **153,** 1077-89.
8. Yang, Z., Swanson, W. J. & V.D., V. (2000) *Mol Biol Evol* **17,** 1446-1455.
9. Ford, M. J. (2001) *Mol Biol Evol* **18,** 639-47.
10. Bielawski, J. P. & Yang, Z. (2001) *Mol Biol Evol* **18,** 523-9.
11. McHugh, D. (2000) *Can J Zool* **78,** 1873-1884.
12. Nicholls, P. (1975) *Biochim Biophys Acta* **396,** 24-35.
13. Arp, A. J., Childress, J. J. & Vetter, R. D. (1987) *J. exp. Biol.* **128,** 139-158.
14. Suzuki, T., Takagi, T. & Ohta, S. (1990) *Biochem J* **266,** 221-5.
15. Zal, F., Suzuki, T., Kawasaki, Y., Childress, J. J., Lallier, F. H. & Toulmond, A. (1997) *Proteins* **29,** 562-74.
16. Zal, F., Leize, E., Lallier, F. H., Toulmond, A., Van Dorsselaer, A. & Childress, J. J. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8997-9002.
17. Weber, R. E. & Vinogradov, S. N. (2001) *Physiological Reviews* **81,** 569-628.
18. Gotoh, T., Shishikura, F., Snow, J. W., Ereifej, K. I., Vinogradov, S. N. & Walz, D. A. (1987) *Biochem J* **241,** 441-5.
19. Bailly, X., Jollivet, D., Vanin, S., Deutsch, J., Zal, F., Lallier, F. H. & Toulmond, A. (2002) *Mol Biol Evol* **19,** 1421-1433.
20. Zal, F., Green, B. N., Lallier, F. H., Vinogradov, S. N. & Toulmond, A. (1997) *Eur J Biochem* **243,** 85-92.
21. Zal, F., Green, B. N., Lallier, F. H. & Toulmond, A. (1997) *Biochemistry* **36,** 11777-11786.
22. Zal, F., Gotoh, T. & Toulmond, A. (1999) in *Fifth International Congress of Comparative Physiology and Biochemistry* (Comp. Biochem. Physiol., Calgary, Alberta, Canada), Vol. 124A, pp. S16.
23. Sturzenbaum, S., Parkinson, J., Blaxter, M., Morgan, J., Kille, P., Schaffner, W. & Georgiev, O. (2000) *unpublished* accession numbers (BF422675)(BF422540).
24. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res* **25,** 4876-82.
25. Felsenstein, J. (1989) *Cladistics* **5,** 164-166.
26. Takagi, T., Iwaasa, H., Ohta, S. & Suzuki, T. (1991) in *Structure and Function of Invertebrate Oxygen Carriers*, eds. Vinogradov, S. N. & Kapp, O. H. (Springer, New York), pp. 245-249.
27. Zhang, J. & Nei, M. (1997) *J Mol Evol* **44(Suppl 1),** S139-S146.
28. Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. & Mornon, J. P. (1997) *Cell Mol Life Sci* **53,** 621-45.
29. Craik, C. S., Buchman, S. R. & Beychok, S. (1980) *Proc Natl Acad Sci U S A* **77,** 1384-8.
30. Craik, C. S., Buchman, S. R. & Beychok, S. (1981) *Nature* **291,** 87-90.
31. Guex, N. & Peitsch, M. C. (1997) *Electrophoresis* **18,** 2714-23.
32. Strimmer, K. & von Haeseler, A. (1996) *Mol Biol Evol* **13,** 964-969.
33. Hasegawa, M., Kishino, H. & Yano, T. (1985) *J Mol Evol* **22,** 160-74.
34. Halanych, K. M., Lutz, R. A. & Vrijenhoek, R. C. (1998) *Cah Biol Mar* **39,** 355-358.
35. Yang, Z. & Bielawski, J. P. (2000) *Tree* **15,** 496-502.

36. Yang, Z. & Nielsen, R. (1998) *J Mol Evol* **46,** 409-18.
37. Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. (2000) *Genetics* **155,** 431-49.
38. Yang, Z. & Nielsen, R. (2002) *Mol Biol Evol* **19,** 908-917.
39. Yang, Z. (1997) *CABIOS* **13,** 555-556.
40. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) *In Atlas of protein sequence structure, Vol. 5, Suppl. 3 National Biomedical Research Foundation, Washington DC. (ed: Dayoff, M.O.),* 354-352.
41. Ohno, S. (1970) *Springler-Verlag, Berlin.*
42. Ohta, T. (1980) *(Lecture. Notes in Bioinformatics, Vol.37) Springer-Verlag, Berlin.*
43. Lynch, M. & Force, A. (2000) *Genetics* **154,** 459-73.
44. Childress, J. J. & Fischer, C. R. (1992) *Oceanogr. Mar. Biol. Annu. Rev.* **30,** 337-341.
45. Freytag, J. K., Girguis, P. R., Bergquist, D. C., Andras, J. P., Childress, J. J. & Fisher, C. R. (2001) *Proceedings of the National Academy of Sciences of the United States of America* **98,** 13408-13413.
46. Suzuki, T., Takagi, T. & Gotoh, T. (1990) *J Biol Chem* **265,** 12168-77.
47. Royer, W. E., Jr., Strand, K., van Heel, M. & Hendrickson, W. A. (2000) *Proc Natl Acad Sci U S A* **97,** 7107-11.
48. Hardison, R. C. (1996) *Proc Natl Acad Sci U S A* **93,** 5675-9.
49. Hardison, R. (1998) *J Exp Biol* **201,** 1099-117.
50. Perutz, M. F. (1979) *Annu Rev Biochem* **48,** 327-86.
51. Messier, S. & Stewart, C. (1997) *Nature* **385,** 151-154.
52. Clark, P. D., Dowling, N. I. & Huang, M. (1998) *J Mol Evol* **47,** 127-32.
53. Tunnicliffe, V. (1992) *Palaios* **7,** 338-350.
54. Branden, R., Malmstrom, B. G. & Vanngard, T. (1973) *Eur J Biochem* **36,** 195-200.
55. Sakai, K., Sakurai, K., Sakai, M., Hoshino, M. & Goto, Y. (2000) *Protein Sci* **9,** 1719-29.
56. Miranda, J. J. (2000) *Biochem Biophys Res Commun* **275,** 517-23.
57. Concha, N. O., Rasmussen, B. A., Bush, K. & Herzberg, O. (1997) *Protein Sci* **6,** 2671-6.
58. Leder, A., Wiener, E., Lee, M. J., Wickramasinghe, S. N. & Leder, P. (1999) *Proc Natl Acad Sci U S A* **96,** 6291-5.
59. Gould, S. J. & Lewontin, R. C. (1979) *Proc R Soc Lond B Biol Sci* **205,** 581-98.
60. Zimmer, E. A., Martin, S. L., Beverley, S. M., Kan, Y. W. & Wilson, A. C. (1980) *Proc Natl Acad Sci U S A* **77,** 2158-62.
61. Cocca, E., Ratnayake-Lecamwasam, M., Parker, S. K., Camardella, L., Ciaramella, M., di Prisco, G. & Detrich, H. W., 3rd (1995) *Proc Natl Acad Sci U S A* **92,** 1817-21.
62. Zhao, Y., Ratnayake-Lecamwasam, M., Parker, S. K., Cocca, E., Camardella, L., di Prisco, G. & Detrich, H. W., 3rd (1998) *J Biol Chem* **273,** 14745-52.
63. Bargelloni, L., Marcato, S. & Patarnello, T. (1998) *Proc Natl Acad Sci U S A* **95,** 8670-5.
64. Baum, A. D. & Larson (1991) *Syst. Zool* **40,** 1-18.
65. Donoghue, M. J. & Mathews, S. (1998) *Mol Phylogenet Evol* **9,** 489-500.
66. Suzuki, T. & Riggs, A. F. (1993) *J. Biol. Chem.* **268,** 13548-13555.

Chapitre 6 :

Discussion, Conclusion générale et Perspectives

Chapitre 6 : Discussion, Conclusion générale et Perspectives

I. Emergence des hémoglobines extracellulaires chez les annélides

Les hémoglobines extracellulaires des annélides ont la caractéristique unique d'être constituées de globines présentant un pont disulfure intra-chaîne (Chapitre 3). Les globines constitutives des hémoglobines intracellulaires du polychète *Glycera dibranchiata* (Imamura et al. 1972) et la neuroglobine intracellulaire du polychète *Aphrodite aculeata* (Dewilde et al. 1996) ne possèdent pas de résidus cystéines susceptibles de former un tel pont disulfure. De la même façon les myoglobines cytoplasmiques du polychète *Arenicola marina* (Kleinschmidt et Weber 1998) ne possèdent pas de tels résidus contrairement aux globines constitutives de l'hémoglobine extracellulaire de la même espèce (El Idrissi Slitine 1991) qui possèdent des cystéines dans les positions invariables impliquées dans les ponts disulfures intra-chaînes. Ces observations suggèrent que l'acquisition du pont disulfure intra-chaîne est un caractère apomorphe pour les hémoglobines extracellulaires d'annélides. En effet, sachant (1) que l'hémoglobine et la myoglobine proviennent d'un évènement de duplication ancestral, et (2) qu'a l'instar des myoglobines cytoplasmiques, les hémoglobines intracellulaires ne présentent pas de cystéines aux positions conservées du pont disulfure, il est probable que l'acquisition de ce caractère (pont disulfure) soit secondaire et corrélée à la localisation extracellulaire de ces hémoglobines.

La divergence précoce entres hémoglobines extra et intracellulaires a été confirmée par les travaux effectués chez le polychète *Travisia japonica*. En effet, les hémoglobines extracellulaires de *Travisia japonica* ne sont pas reconnues par les anticorps anti-hémoglobines intracellulaires de la même espèce (Fushitani et al. 1982, 1988). Par ailleurs il est manifeste que les hémoglobines intracellulaires de *Glycera dibranchiata* ne possèdent manifestement pas d'orthologues parmi les globines extracellulaires comme le suggère la phylogénie moléculaire d'hémoglobines intra et extracellulaires présentée dans la figure 39. La présence de séquences d'adressage dans les polypeptides des hémoglobines extracellulaires qui conduit à leur externalisation dans le milieu vasculaire ou coelomique, peut être considérée comme une innovation par rapport aux hémoglobines intracellulaires des érythrocytes ou cœlomocytes circulants des annélides.. La masse moléculaire importante des hémoglobines extracellulaires circulantes pourrait permettre également d'échapper à une élimination rapide au niveau des organes de filtration, les néphridies (analogues des reins chez les animaux à coelome) (Weber et Vinogradov 2001). De plus, l'organisation polymérique

complexe, à la fois en terme d'acquisition des ponts disulfures inter-chaînes et en terme d'interaction avec les chaînes de structure ("linkers") renforce le scénario évolutif d'une émergence tardive des hémoglobines extracellulaires. Ceci suggère que les hémoglobines intracellulaires sont probablement plus primitives que les hémoglobines extracellulaires et que les deux lignées de globines constitutives des hémoglobines intra et extracellulaires sont probablement issues d'une duplication très ancestrale, précédant la radiation des annélides.

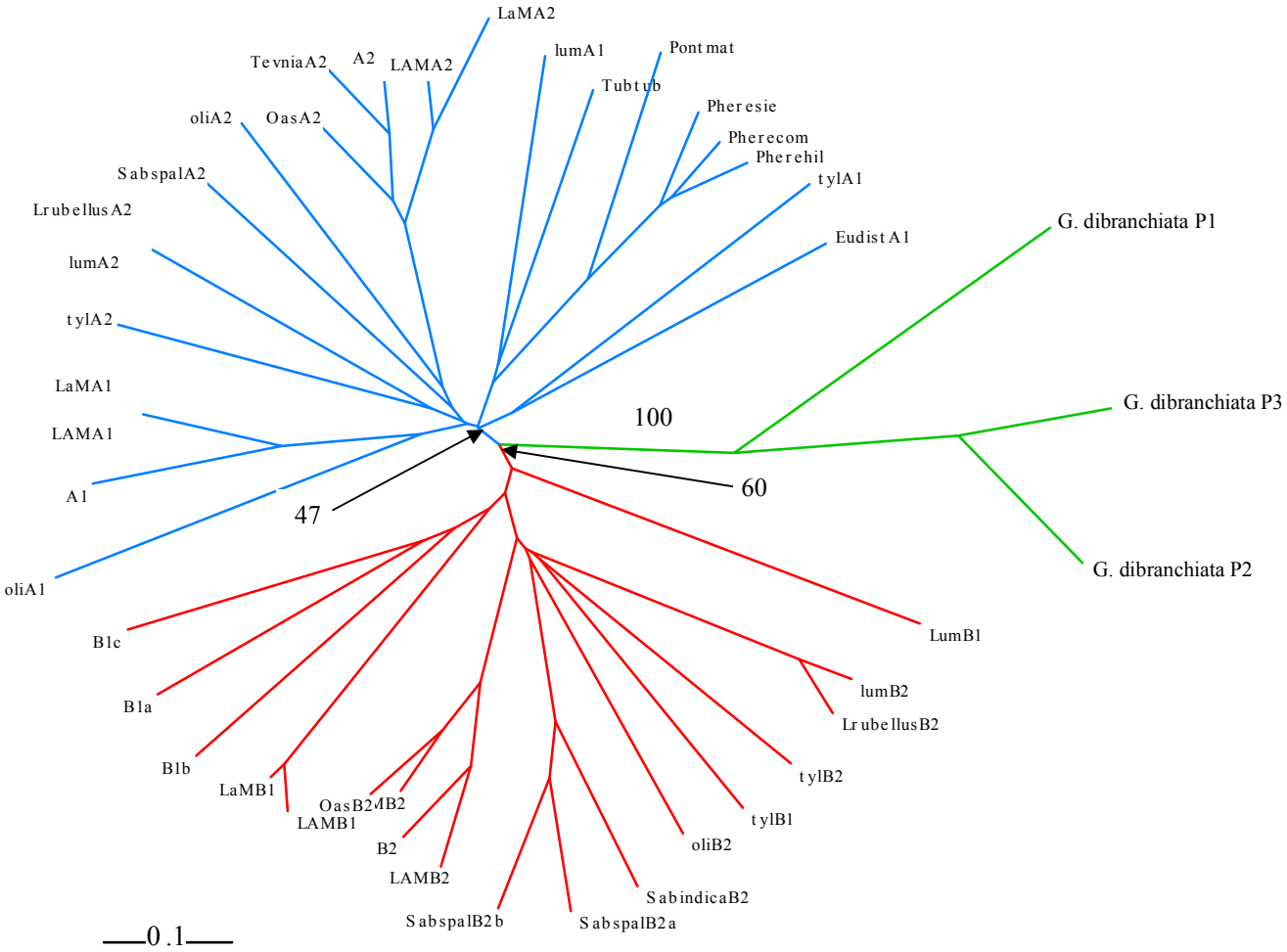


Figure 39. Phylogénie moléculaire non racinée d'hémoglobines extracellulaires (**familles A** et **famille B**) et **intracellulaires** (*Glycera dibranchiata*) réalisée par Neighbor Joining et un bootstrap de 1000 réplicats. Seules les valeurs de bootstrap des branches conduisant aux familles A et B et aux globines extracellulaires sont indiquées.

II. La famille multigénique des hémoglobines de *Riftia pachyptila*

L'organisation de la famille multigénique des hémoglobines de *Riftia pachyptila* est conforme à celles des familles multigéniques d'annélides déjà décrites. Des évènements de duplications successifs et la conservation de résidus cystéines impliqués dans la fixation du sulfure d'hydrogène, permettent d' élaborer un scénario plausible permettant de retracer l'évolution de cette famille multigénique et, plus généralement, celle de la fonction de fixation des sulfures chez les annélides. Des indices de structure secondaire et de vitesse évolutive des gènes de la famille multigénique (Chapitre 4) nous ont permis de comprendre pourquoi (pression de sélection) et comment (duplications et conservation de l'environnement moléculaire flanquant les cystéines) la fonction de fixation du sulfure d'hydrogène n'est portée que par les deux globines paralogues A2 et B2. Nous avons caractérisé le domaine de fixation des sulfures (SBD pour Sulfide Binding Site, Chapitre 4) en tant que structure conservée flanquant la cystéine libre. L'absence de ces cystéines au sein des sous familles A1 et B1 et leurs vitesses évolutives plus grandes que celles des globines A2 et B2, suggèrent qu'au moment des duplications des relâchement de contraintes sélectives au niveau des dupliquats ont pu occasionner la perte des cystéines libres. Ainsi les globines A2 et B2 seraient codées par des **plésiogènes** (dont la fonction associée est identique à la fonction ancestrale) et les globines A1 et B1 par des **apogènes.** En spéculant, on peut penser que la multiplication par duplication des cystéines libres fixatrices de H_2S aurait pu poser un problème de stœchiométrie entraînant un potentiel supplémentaire de liaison du substrat H_2S déjà en équilibre avec sa protéine, la conséquence d'un tel bouleversement pouvant occasionner une modification de l'homéostasie.
En comparant les structures secondaires SBD (Chapitre 4) de *Riftia pachyptila* et d'autres annélides des milieux oxydés, nous nous sommes aperçus que le domaine de fixation des sulfures était conservé malgré l'absence des résidus cystéines et l'incapacité à fixer les sulfures. Cette observation nous a conduit à émettre l'hypothèse que les cystéines Cys+1 et Cys+11 avaient été perdues chez les annélides des milieux sans sulfures. Nous avons rejeté l'hypothèse alternative d'une colonisation des milieux réduits par les annélides à partir des milieux oxydés

1. Rejet de l'hypothèse alternative : colonisation des milieux riches en H_2S par des annélides venant des milieux oxydés

Il semble difficilement concevable que la seule acquisition d'une cystéine en +1 ou +11 sur une globine puisse permettre à une annélide des milieux oxydés de coloniser un milieu riche en H_2S. En effet, la fixation et le transport de ce composé nécessite, **au minimum,** une prise en charge du composé soit par des bactéries chimiosynthétiques soit par un système enzymatique oxydant des sulfures (Chapitre 2, Chapitre 5). Acquérir autant d'adaptations moléculaires et physiologiques requiert manifestement **trop d'évènements en trop peu de temps pour survivre dans un environnement réduit sachant que l'effet toxique du sulfure d'hydrogène est très rapide sinon immédiat.** Cependant on ne peut théoriquement pas exclure une adaptation graduelle à l'H_2S, mais le manque d'arguments convainquant allant dans ce sens nous a fait rejeter cette hypothèse.

On retrouve un grand nombre d'annélides très différentes (polychètes, oligochètes, achètes, vestimentifères et pogonophores) dans les milieux réduits. Le séquençage partiel des hémoglobines extracellulaires multi-domaines de *Branchipolynoe sp*, annélides polychètes commensales des moules des sources hydrothermales, a aussi montré la présence d'une cystéine libre en position Cys+11 (Hourdez 2000). Dans l'hypothèse d'une colonisation récente des milieux réduits par un ancêtre annélidien migrant des milieux oxydés il faudrait concevoir une annélide pré-adaptée à la fixation d'un composé proche de H_2S en terme d'encombrement stérique et de réactivité (lequel ?). A la lumière de nos connaissances et en postulant que le mécanisme de fixation de H_2S soit le même pour toutes ces annélides (ce qui semble le cas) il faudrait expliquer la situation actuelle par de nombreuses séries répétées de convergences adaptatives d'un mécanisme fonctionnel complexe. C'est hautement improbable.

2. Les sources hydrothermales et les suintements froids ont-ils été colonisés par des annélides littoraux ou sub-littoraux équipées de telles adaptations ?

Des fossiles assimilés à des polychètes alvinellidés et à des vestimentifères ont été retrouvés dans des ophiolites (roches magmatiques originaires de la croûte océanique remontées sur la croûte continentale par la tectonique des plaques) datées du Silurien (- 420 millions d'années) au Cénozoïque (- 65 millions d'années ; Little et al. 1998). Pour ces auteurs, les fossiles et leur endémicité aux sources hydrothermales et aux suintements froids contredisent les

hypothèses de Newman (1985) au moins en ce qui concerne les annélides. Selon ces hypothèses, les résurgences de fluides réduits auraient été des refuges ayant permis à certaines espèces de résister aux extinctions massives du Crétacé (-65 millions d'années). En ce qui concerne les alvinélidés et les vestimentifères (ou leurs ancêtres) les notions de refuge et d'endémicité s'excluent, car en assimilant ces fossiles aux vestimentifères modernes, Little et al (1998) suggèrent que ces annélides ont été associés dès l'origine aux sources hydrothermales et aux suintements froids. Implicitement, une telle origine suggère aussi une adaptation ancienne aux sulfures puisque les ophiolites contiennent en majorité des sulfures poly-métalliques.

Cependant, une autre théorie suggère, à partir des données moléculaires, que les vestimentifères des sources hydrothermales seraient issus d'une radiation récente, inférieure à 100 millions d'années (Cénozoïque) (Halanych et al. 1998). L'hypothèse d'une invasion récente des sources hydrothermales à partir des vestimentifères des suintements froids a également été émise car certains **vestimentifères opportunistes** originaires des suintements froids se retrouvent également à proximité de sources hydrothermales (Black et al. 1997). Signalons cependant que des périodes d'anoxie globale qui incluent le fond des océans seraient survenues, selon les auteurs, à la transition Permien-Triass (250 millions d'années), au Jurassique (Toarcien : -180 millions d'années), ou encore au Paléocène (- 65 millions d'années). Les communautés des sources hydrothermales profondes n'auraient pas, selon un consensus général, résisté à de telles conditions. La conséquence serait l'extinction de ces communautés. Dans ce cas il faut envisager une recolonisation entre -250 et -65 millions d'années de ces milieux réduits à partir d'espèces habitant hors du champ hydrothermal, puis un évènement de radiation récent pour être en accord avec les données moléculaire. Cette recolonisation n'a cependant pu se faire qu'à partir d'espèces annélidiennes proches des milieux plus continentaux mais tout aussi réduits.

Mais dans quelle mesure certaines espèces n'auraient-elles pas pu résister à cette anoxie ? Les scénarii évolutifs reposant sur des évènements d'extinctions, de recolonisations et de radiations sont trop hypothétiques pour être débattus ici, mais cependant mitigent les deux hypothèses de l'origine des communautés hydrothermales.

Souvenons nous que le fluide des sources hydrothermales est hautement radioactif (Cherry et al. 1982). A ce titre, s'adapter à des conditions aussi mutagènes est, d'un point de vue de l'adaptation moléculaire, difficile à concevoir. Nous pensons que les sources hydrothermales pourraient être une image plausible d'un environnement primitif assez répandu à la surface de

la terre et dans lequel le ou les ancêtres des annélides modernes auraient émergé avec une panoplie d'adaptations, dont celle leur permettant de vivre en présence de sulfures.

3. Fonction de densité des hémoglobines

Un parallèle intéressant peut être fait entre les faibles vitesses évolutives observées chez les globines A2 et B2 (par rapport au globine A1 et B1) et le concept de **fonction de densité des protéines**, développé par (Zuckerkandl (1976a). La densité fonctionnelle d'une protéine est définie comme la proportion de résidus impliqués dans la (les) fonction(s) spécifique(s) de la protéine en question, le reste des résidus n' intervenant que dans la structure et le repliement de la protéine. Dans cette publication, Zuckerkandl ajoute : " Since the evolutionary variation in the primary structure of a protein is slowed down by its specific function, there should be a simple relation between evolutionary rates and functional density of different kinds of proteins". La double fonctionnalité (cumul de deux fonctions vitales : le co-transport de composés énergétiques H_2S et O_2) par les globines A2 et B2, la présence des cystéines libres et leur vitesse évolutive réduite qui suggère une forte sélection négative (chapitre 4), reflètent un état de forte densité fonctionnelle. Ainsi, les hémoglobines se révèlent être un excellent exemple d'une lignée moléculaire dont on peut considérer que la structure primaire est "gelée" : paradoxalement, bien que les hémoglobines aient une structure primaire très variable, leur structure tertiaire est toujours du type "globin fold". Dans cette structure "gelée" viennent se greffer, en plus de la capacité à lier O_2, des fonctions plus spécifiques telles que la capacité de lier NO chez l'hémoglobine humaine ou H_2S chez les globines extracellulaires des annélides. Pour Zuckerkandl (1976 b) la notion de densité fonctionnelle maximale serait approchée de façon asymptotique au cours de l'évolution des protéines. Cette notion "du jamais atteint mais s'en rapprochant" est également bien illustrée par les hémoglobines en général et ressort de ce travail sur la famille des hémoglobines de *Riftia pachytila*. Selon Zuckerkandl, les deux forces principales qui opèrent sur la densité fonctionnelle des protéines sont des contraintes énergétiques et organisationnelles. L'auteur suggère qu'avoir des polypeptides fonctionnels de faible longueur constitue un avantage économique pour les cellules primitives, car plus rapidement synthétisées (il souligne ici la contrainte énergétique). On observe dans tous les organismes vivants, procaryotes comme eucaryotes, que le repliement en "globin-fold" est universel : cette structure acquise très tôt dans l'évolution a donc probablement atteint rapidement un haut niveau de densité fonctionnelle. De plus les exemples de fonctions spécifiques des hémoglobines du règne du vivant (Chapitre 1) montrent que cette molécule est

capable d'innover. On peut ainsi penser qu'une fois établie, la structure de base ("le globin fold") fut au gré des lignées évolutives une aire potentielle et privilégiée à un "bricolage" (dans le sens de François Jacob) limité, donc une aire d'adaptation moléculaire putative. Zuckerkandl considère aussi que l'interaction moléculaire devient un avantage en terme de performance des fonctions (par exemple la coopérativité pour l'oxygène ou les linkers des HBL-Hbs d'annélides) et de régulation de cette performance (on n'observe pas d'agglomération anarchique, mais des motifs répétés : cet ordre structural confère en soi une régulation dans les capacités d'interaction avec les ligands).

III. Détecter de l' adaptation moléculaire sans estimer les substitutions nucléotidiques

A partir de la comparaison des différentes structures d'hémoglobines chez différentes espèces, Perutz (1983) arrive à la conclusion suivante : " **New chemical functions appear to have evolved only by a few amino acid substitutions in key positions … The structural evidence suggests that most of amino acid replacements between species are neutral or nearly so, caused by random drift of selectively equivalent mutant genes, and that adaptative mechanisms generally operate by a few replacement in key position** ". Les propos de Perutz sont bien illustrés par l'exemple de l'effet Root, caractéristique des hémoglobines de certains poissons téléostéens capables de réguler le volume de leur vessie natatoire en y injectant à partir de leur oxyhémoglobine de l'O_2 sous haute pression hydrostatique. L'effet Root qui représente une version extrême de l'effet Bohr est dû au remplacement d'un seul acide aminé, la cystéine F9 de la chaîne β des mammifères par une sérine chez ces poissons (Perutz 1983). Dans le même ordre d'idée, mais concernant l'adaptation des hémoglobines des crocodiles à fixer deux ions bicarbonates pour réduire l'affinité de l'oxygène et rester ainsi jusqu'à une heure en apnée, Perutz (1983) suggérait que seuls 3 acides aminés mutés dans l'hémoglobine humaine suffisaient pour acquérir cette fonction. Douze années plus tard Komiyama et al. (1995), à partir d'expériences de mutagenèse dirigée, ont montré qu'il fallait remplacer 7 acides aminés pour obtenir cet effet. Dans un très didactique et intéressant article intitulé "The structural basis of molecular adaptation" Golding et Dean (1998) citent l'exemple de l'hémoglobine à forte affinité pour l'oxygène d'une oie migratrice, *Anser indicus*, capable de survoler l'Everest à plus de 9 kilomètres d'altitude et soumise à des pressions partielles d'oxygène représentant 30% de celle du niveau de la mer. L'oie domestique *Anser anser*, proche parent de l'espèce précédente, incapable de voler à ces altitudes, possède une hémoglobine d'une affinité normale. Seuls trois

acides aminés de la chaîne α et un acide aminé de la chaîne β permettent de distinguer les hémoglobines de ces deux espèces. Ces substitutions éliminent des contacts de type Van der Waals entre les sous unités α et β et permettent l'adaptation au vol à haute altitude.

La fixation du sulfure d'hydrogène par les hémoglobines de mollusques résulte d'interactions électrostatiques sans implications de cystéines libres. Cette singularité n'est observée que chez ces animaux. Des travaux de mutagénèse dirigée effectués sur l'hémoglobine de cachalot ont montré *in vitro* que trois substitutions d'acides aminés suffisent à conférer cette même propriété de fixation de H_2S (Nguyen et al. 1998).

D'autres exemples d'adaptation moléculaire concernant des protéines comme l'opsine des pigments visuels ou l'isocitrate déhydrogénase, enzyme qui catalyse l'oxydation du citrate en α-cetoglutarate, sont basés sur des données structurales comparatives et intègrent des données de physiologie, d'écologie, de fonctionnalité (mutagenèse dirigée) et de phylogénie des espèces. Ils suggèrent tous des mécanismes adaptatifs n'impliquant qu'une modification d'un nombre restreint de mutations à partir d'une protéine ancestrale.

En d'autres termes, cette approche multidisciplinaire montre une action de la sélection naturelle innovante, et particulièrement de la sélection positive, à un niveau moléculaire et localisée en des points clés de la protéine. Cette démarche confirme aussi que d'autres positions de la molécule évoluent de manière neutre et confirme qu'interprétation neutraliste et sélectionniste sont complémentaires et ne s'excluent pas.

Cependant, cette approche peut trouver une limite qui serait l'effet Dykhuizen : une ou quelques substitutions neutres établies dans un environnement donné peuvent conduire à une nouvelle fonction ou subfonctionalité et conférer un avantage sélectif pour l'espèce confrontée à un nouvel environnement (Dykhuizen et Hartl 1980 ; Bielawski et Yang 2001). Ainsi des substitutions neutres peuvent leurrer un évènement d'adaptation moléculaire. Pour écarter cette possibilité, il faut en plus estimer le taux de fixation des substitutions synonymes et non synonymes de manière à quantifier l'impact de la sélection naturelle (positive, neutre ou négative) sur les protéines.

IV. Détecter de l'adaptation moléculaire en estimant les substitutions nucléotidiques

Des données expérimentales décrites dans le paragraphe précédent procurent nombre d'indices d'adaptation moléculaire sans pour autant permettre de réfuter un effet Dykhuizen.

Notre démarche et les outils que nous avons utilisés pour explorer les effets de la sélection naturelle sur la fixation des mutations non synonymes au sein des hémoglobines de *Riftia pachyptila* reposent sur des modèles statistiques et des algorithmes de maximum de vraisemblance. Ils permettent d'estimer les substitutions synonymes et non synonymes entre séquences codantes d'hémoglobines apparentées, à partir d'un alignement de séquences, d'une phylogénie moléculaire de référence (chapitre 5 et annexe 7) et avec les programmes de Yang (Yang 1997). Les premiers exemples d'adaptation évolutive par sélection positive ont d'abord été démontrés à l'échelle des lignées évolutive en moyennant les substitutions synonymes et non synonymes sur l'ensemble de la protéine (Yang 1998) (annexe 7.1). Mais les exemples résumés dans le paragraphe précédent montrent qu'une approche locale, à l'échelle de l'acide aminé, est nécessaire. Cette approche a été rendue possible par le second et le troisième modèle de Yang (Yang et al. 2000 ; Yang et Nielsen 2002 ; annexes 7.2 et 7.3). Ces 2 derniers modèles de type RAS (pour Rate-Across-Sites) permettent d'estimer les substitutions synonymes et non synonymes à partir de distributions statistiques de différents types et simulent la possibilité d'un taux évolutif différent entre les sites d'une même protéine **(**Chapitre 5**)**. Ces deux modèles permettent de détecter des sites sous sélection positive, mais c'est le troisième modèle (Yang et Nielsen 2002) qui permet de révéler des sites sous sélection positive dans une seule lignée évolutive parmi l'ensemble des lignées considérées. Les cystéines impliquées dans la liaison du sulfure d'hydrogène par les hémoglobines de *Riftia pachyptila* ont pu ainsi être caractérisées comme étant sous sélection darwinienne positive dans les lignées d'annélides vivant dans les environnements sans sulfures. L'ensemble des résultats, dont ceux obtenus et présentés dans le chapitre 4, suggèrent que cette perte de cystéine, donc de la fonction de liaison réversible de H_2S, est le résultat d'une adaptation moléculaire par un relâchement de la contrainte sélective constituée par la présence de forte concentration en sulfure d'hydrogène. Pourtant, le relâchement des contraintes sélectives et la perte de fonction qui s'en suit est généralement expliquée par de l'évolution moléculaire neutre et non à de l'adaptation moléculaire.

V. Un concept marginal : la perte de fonction peut refléter de l'adaptation moléculaire et non un simple effet de la dérive génétique

Le relâchement des pressions de sélection est généralement mis en avant par la théorie neutraliste de l'évolution pour expliquer la perte de fonction par accumulation de mutations neutres. Cette situation permet aussi de révéler des mutations neutres qui seraient délétères et

contre-sélectionnées sans ce relâchement des contraintes sélectives. Pour la théorie sélectionniste, l'absence de contraintes sélectives permet d'expliquer l'apparition de nouvelles fonctions (source d'innovation moléculaire ou adaptation moléculaire) en permettant l'apparition de substitutions non synonymes qui changent la nature de l'acide aminé (la sélection positive).

Nos résultats et interprétations indiquent qu'un relâchement des contraintes sélectives peut entraîner **une perte de fonction par adaptation moléculaire** (i.e. par sélection positive). Cette situation n'est concevable que lorsque le maintien de la fonction présente un coût ou lorsque dans un environnement nouveau cette fonction confère un désavantage sélectif. C'est ce second point que nous avons mis en avant (Chapitre 5) pour expliquer que l'absence de cystéines libres +1 et +11 dans globines B2 et A2 des annélides vivant en milieux oxydés et dépourvus de sulfures. En effet la présence d'une cystéine dans une structure protéique est semble-t-il toujours corrélée à une réaction ou un mécanisme moléculaire particulier. De nombreux travaux portant sur l'apparition ou la disparition d'une cystéine par mutagenèse dirigée montrent que de telles modifications conduisent à l'inhibition de certaines réactions métaboliques. Par exemple (cf exemples supplémentaires dans le chapitre 5), la disparition de ponts disulfures au sein de la sous-unité β de la pompe à proton (ATPase) gastrique provoque la perte de la fonction remplie par cette protéine (Kimura et al. 2002). La substitution d'une cystéine pour une alanine dans le centre catalytique de l'enzyme D2 thyroïdienne qui catalyse l'activation de T4 (principal produit sécrété par la thyroïde) en T3, modifie les interactions avec le ligand et donc les vitesses de réaction enzymatique (Kuiper et al. 2002). De tels désordres entraînent des effets délétères sinon létaux à l'échelle de l'organisme.

La perte d'un gène ou de son expression et sa conséquence ont déjà été caractérisées et expliquées par la présence de mutations altérant le cadre de lecture, ou de mutations non sens dans les séquences codantes, mais aussi dans les séquences régulatrices en amont ("upstream"). De telles mutations expliquent la perte, chez le moustique *Culex pipiens,* d'un récepteur cellulaire à une toxine bactérienne qui a permis à l'insecte d'acquérir un mécanisme de résistance à cette bactérie utilisée comme bio-insecticide (Darboux et al. 2002). Chez les primates supérieurs (grands singes et humains) la perte de l'enzyme qui permet la synthèse d'un épitope retrouvé également chez des micro-organismes a permis à ces primates de développer des anticorps contre ces mêmes micro-organismes pathogènes potentiels (Koike et al. 2002). A notre connaissance, aucun de ces travaux ne démontrent formellement que de la sélection positive est la cause de ces pertes de gènes et de ces fonctions.

Sur la base de notre analyse du rôle joué par les cystéines, de l'écologie des annélides vivant en présence ou en absence de sulfures, de données paléo-géologiques et phylogénétiques, nous proposons que les cystéines +1 et +11 ont été perdues par sélection positive. Cette perte constitue une adaptation moléculaire, conséquence, soit de la disparition de H_2S dans le milieu, soit de la colonisation des milieux oxydés.

VI. Limites expérimentales

1. Pseudogènes et famille multigénique

L'évolution des familles multigéniques ne conduit pas seulement à des gènes dupliqués fonctionnels. Parfois le gène dupliqué peut évoluer en pseudogène. Ce gène peut ne pas être transcrit ou, s'il l'est, peut ne pas être traduit. Dans le deuxième cas, une mutation affecte la traduction et induit soit un décalage du cadre de lecture soit des mutations non-sens. Ainsi, bien que la probabilité soit très faible, il est donc possible de synthétiser des ADNc qui correspondent à des gènes non-exprimés (Cappai et al. 1992). Tant que l'ensemble des ADNc ne sera pas complètement séquencé et confronté aux données de spectrométrie de masse (Chapitre 3) cette situation doit être envisagée même si elle est peu probable.
De la même façon, bien qu'en terme de nombre de globines les approches de spectrométrie de masse et de biologie moléculaire convergent vers un même résultat et confortent le nombre de 6 sous-unités distinctes pour les hémoglobines de *Riftia pachyptila*, il faut envisager que cette famille puisse être plus complexe.

2. Séquences nucléotidiques des globines disponibles : sous-échantillonnage ?

Une seule séquence nucléotidique d'annélide oligochète, la globine A2 ou B2 de *Lumbricus rubellus*, suffit-elle pour étudier l'effet de la sélection (positive, neutre ou négative) sur les sites Cys+1 et Cys+11 étendu à l'ensemble des annélides des milieux oxydés (Chapitre 5) ?
Le manque de séquences nucléotidiques des globines A2 et B2 d'annélides des milieux oxydés à mettre en face des globines des annélides vivant en présence de H_2S, est fâcheux.
Cependant, les séquences primaires orthologues de globines A2 et B2 (annexe 4) d'annélides des milieux sans sulfures (*Lumbricus terrestris, Tylorrhynchus heterochaetus*) montrent que les substitutions d'acides aminés aux sites respectifs Cys+11 et Cys+1, sont restreintes à des acides aminés non polaires (leucine et valine pour B2, alanine et valine

pour A2) (annexe 4). Chez la globine B2 de *Lumbricus rubellus* le site Cys+1 est occupé par une phénylalanine (non polaire). Nous avons donc simulé les situations où le nœud phylogénétique conduisant à la lignée des globines d'annélides des milieux oxydés (figure 7 du chapitre 5) pourrait être une globine ancestrale dont le site occupé par une cystéine en milieu réduit pourrait être l'un de ces 3 acides aminés en milieu oxydé. Nous avons donc remplacé successivement la phénylalanine de *Lumbricus rubellus* par une alanine, une valine et une leucine.

Il faut se rappeler et retenir que le troisième modèle "branch and site" de Yang (Chapitre 5) est capable de détecter des sites sous sélection à chaque nœud d'une branche interne de l'arbre phylogénétique, en reconstruisant la séquence ancestrale putative à ces mêmes nœuds. La figure 5 du chapitre 4, réalisée à partir de séquences primaires de globines orthologues B2, montre un regroupement des annélides des milieux oxydés. Sur la base de cette phylogénie moléculaire on peut donc, eu égard aux acides aminés trouvés au site Cys+1 des globines B2 des annélides des milieux oxydés, considérer qu'un des 3 acides aminés pourrait occuper ce site dans le premier nœud de la lignée conduisant aux globines de ces milieux. **Chaque simulation montre que le site reste toujours sous sélection positive quelle que soit la substitution par un des trois acides aminés.**

Le site Cys+1 de *Lumbricus rubellus* a été remplacé avec chaque codon du code génétique. On peut ainsi voir quels sont les codons qui produisent un acide aminé sous sélection positive au site Cys+1. Dans le tableau obtenu (fig. 40) la couleur rose indique les codons dont la présence ne permet pas au modèle de détecter de la sélection positive. La couleur jaune indique des codons pour lesquels le modèle détecte de la sélection positive mais avec une probabilité inférieure à 0.99. La couleur bleue montre que quand la cystéine est substituée par l'un des acides aminés correspondant aux codons, le site Cys +1 est sous sélection positive (avec $P>0.99$).

Chez les globines paralogues A1 et B1 des annélides des milieux riches ou non en sulfures et pour les sites Cys+11 et Cys+1 respectifs, on observe une majorité de substitutions d'acides aminés non polaires et d'un acide aminé polaire non ionique, la thréonine (annexe 4). Toutes ces substitutions (fig. 40) traduisent un évènement de sélection positive. Ainsi toutes les substitutions observées et testées suggèrent que les paralogues A1 et B1 des annélides des milieux riches en sulfures ou non ont probablement perdu après duplication par sélection positive la cystéine présente chez les globines ancestrales A et B (Chapitre 4, figure 5).

Ce travail conforte également l'hypothèse selon laquelle, après un évènement de duplication, le relâchement des contraintes sélectives sur le gène dupliqué permet l'accumulation de

substitutions non synonymes pouvant altérer la nature du polypeptide. Dans le cas qui nous occupe, l'accumulation de substitutions non synonymes conduit à la perte de la cystéine pendant l'émergence de la famille multigénique des globines au sein de l'ancêtre annélidien, avant radiation du phylum.

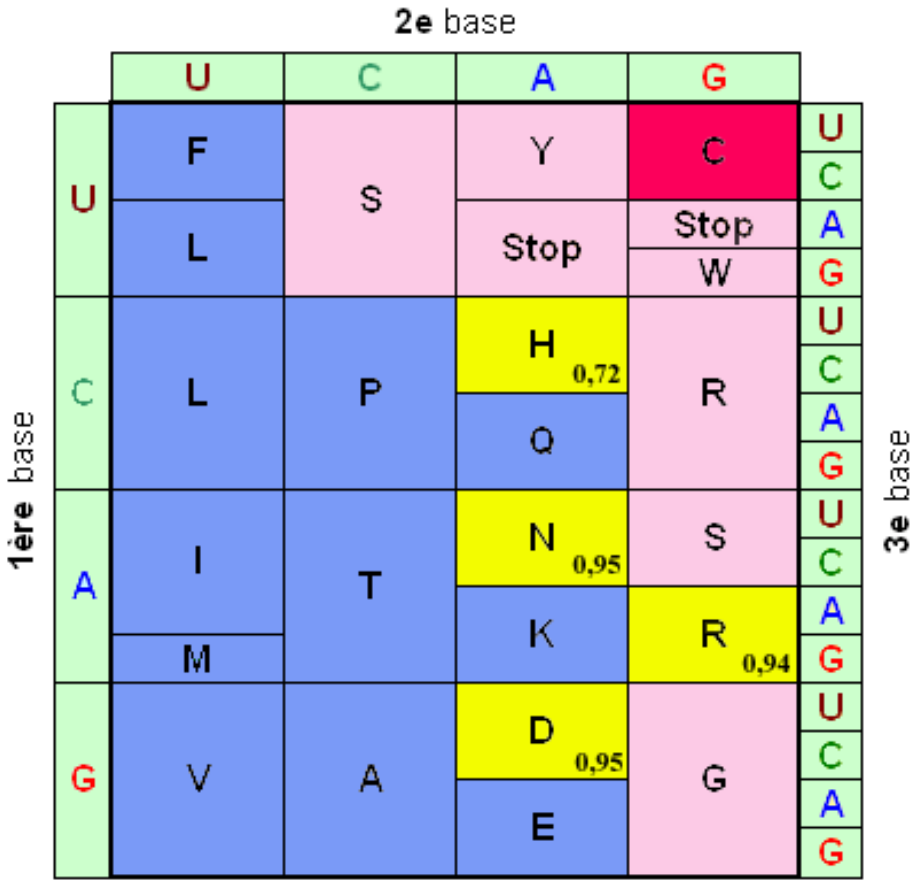


Figure 40. Tableau du code génétique et probabilité (P>0,99 P = 0 P < 0.99) pour que le site Cys+1 de la globine B2 soit ou non sous sélection positive quand il est occupé par l'acide aminé correspondant (à l'exception de la cystéine et des codons "stop").

VII. Conclusion

Nous avons caractérisé la famille multigénique des hémoglobines de *Riftia pachyptila* et proposé un scénario évolutif centré sur la fonction de fixation des sulfures par des cystéines libres. Ce travail sur les gènes paralogues apparus par duplication et les gènes orthologues apparus par spéciation de cette famille multigénique, s'ajoutant aux données écologiques et physiologiques, nous ont permis de montrer que les concentrations en sulfure d'hydrogène

opèrent une forte pression sélective sur les globines. L'intensité de cette pression de sélection est plus importante sur les gènes codant les globines qui fixent H_2S chez les annélides vivant en milieu réduit. La phylogénie moléculaire des globines, la conservation des cystéines, la conservation des structures secondaires ainsi que les différentes vitesses évolutives entre certains gènes paralogues des globines extracellulaires des annélides vivant en présence de H_2S, suggèrent que la présence des résidus cystéine est ancestrale. L'absence des cystéines libres dans des globines d'annélides des milieux oxydés dépourvus de sulfures, nous a conduit à tester l'hypothèse évolutive de la perte de fonction de fixation des sulfures chez ces annélides (via la perte des cystéines). Un *a priori* conceptuel nous a dans un premier temps fait interpréter cette perte comme un simple relâchement des contraintes. Mais la réactivité chimique de la cystéine a orienté nos recherches vers une interprétation différente et particulièrement vers un effet de pression sélective connu sous le nom de sélection positive, qui aurait pu opérer sur les globines porteuses de ces cystéines. Cette pression a pour conséquence d'augmenter les substitutions non-synonymes par rapport aux substitutions synonymes avec un corollaire, la capacité d'acquérir de nouveaux acides aminés pouvant conférer certaines propriétés et éventuellement un avantage sélectif.

Nous avons montré en utilisant des programmes permettant de tester statistiquement des hypothèses évolutives complexes que les sites occupés par les cystéines libres chez les annélides des milieux réduits ont évolué sous sélection positive. Cette sélection positive chez les annélides des milieux oxydés, a conduit à leur remplacement par des résidus neutres, non réactifs. En effet, le maintien de cystéines libres dans un environnement privé de son ligand naturel a dû être sévèrement contre-sélectionné au vu du potentiel d'interaction que peut avoir la cystéine avec d'autres composés circulant dans le sang d'une part, et des désordres métaboliques délétères ou sub-létaux qui peuvent en découler d'autre part. Le coût du maintien de la cystéine dans ces conditions aurait pu conférer un désavantage sélectif certain.

A la lumière de ces résultats et interprétations, nous pensons que les annélides modernes descendent d'ancêtres vivant dans des milieux réduits et qu'un ou plusieurs évènements d'adaptation moléculaire expliquent la perte de la fonction de liaison des sulfures dans les milieux oxydés. Ce cas constitue un exemple de **désadaptation** au sens de Baum et Larson (1991). Si cette hypothèse est confirmée par des travaux de mutagenèse dirigée sur les cystéines et leur proche environnement (cf perspectives), ces résultats conforterons le paradigme qui se profile de plus en plus, à savoir que l'évolution adaptative nécessite que peu de modifications moléculaires (Golding et Dean, 1998).

Il existe cependant des exemples d'adaptations moléculaires impliquant conjointement l'évolution des protéines, l'adaptation des organismes et les conditions environnementales, **mais qui nécessitent d'importants remaniements génomiques**. On peut citer l'exemple des glycoprotéines antigel des poissons Notothenioidés de l'Antarctique. C'est le gène du trypsinogène transformé par des recrutements des régions régulatrices et des amplifications de répétitions nucléotidiques, qui est à l'origine du nouveau gène codant des protéines antigel permettant l'adaptation au froid (Chen et al. 1997). De même, chez les procaryotes l'acquisition de nouvelles fonctions enzymatiques par recombinaison est également un exemple d'adaptation moléculaire mettant en jeu de grands remaniements (Hall et Zuzel 1980). Enfin, les gènes mutateurs bactériens qui augmentent le taux de mutations et de recombinaisons génèrent aussi des adaptations moléculaires en terme de résistance aux antibiotiques et de virulence (LeClerc et al. 1996).

A la lumière de ce travail et de ces exemples d'adaptation il conviendrait de discerner trois types d'évènements d'adaptation moléculaire :

-Le recrutement de gènes pour d'autres fonctions (expressions spatiales et temporelles différentes à partir de mutations dans des régions régulatrices).

-Des remaniements de séquences nucléotidiques importantes au niveau du génome.

-Des innovations moléculaires ponctuelles dans des lignées moléculaires immédiatement après duplication ou après l'établissement de cette lignée établie (exemple de la perte de fonction de fixation des sulfures par perte des cystéines).

Nous voudrions pour terminer, relativiser la toxicité du sulfure d'hydrogène telle qu'elle a pu être perçue au cours de ce manuscrit, et spéculer sur une spécificité possible de la neuroglobine du cerveau humain. Certes H_2S est un composé toxique mais il est produit de façon endogène dans le cerveau où il joue le rôle de neuromodulateur. Un article récent montre que dans les maladies à déclin cognitif et particulièrement dans la maladie d'Alzheimer les niveaux en H_2S dans le cerveau diminuent considérablement (Eto et al. 2002). Il faut aussi remarquer que les neuroglobines humaines, murines et des poisson téléostéens *Danio rerio* et *Tetraodon nigroviridis* (fig. 41) possèdent selon les cas deux ou trois cystéines que l'on ne retrouve pas dans les hémoglobines α et β humaines (Burmester et al. 2000). Il serait donc vraiment intéressant de pouvoir tester les capacités de transport ou de stockage transitoire de H_2S par ces neuroglobines à l'aide de ces cystéines et qui sait, discuter d'une éventuelle convergence évolutive entre fonction de fixation de H_2S via des cystéines libres chez les hémoglobines extracellulaires d'annélides et les neuroglobines de vertébrés.

```
                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  10         20         30         40         50
Mus musculus              MERP---ESE LIRQSWRVVS RSPLEHGTVL FARLFALEPS LLPLFQYNGR
Rattus norvegicus         MERL---ESE LIRQSWRAVS RSPLEHGTVL FSRLFALEPS LLPLFQYNGR
Homo sapiens              MERP---EPE LIRQSWRAVS RSPLEHGTVL FARLFALEPD LLPLFQYNCR
Tetraodon nigroviridis    MEKLSSKDKE LIRGSWDSLG KNKVPHGVIL FSRLFELDPE LLNLFHYT-T
Danio rerio               MEKLSEKDKG LIRDSWESLG KNKVPHGIVL FTRLFELDPA LLTLFSYS-T

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                  60         70         80         90        100
Mus musculus              QFSSPEDCLS SPEFLDHIRK VMLVIDAAVT NVEDLSSLEE YLTSLGRKHR
Rattus norvegicus         QFSSPEDCLS SPEFLDHIRK VMLVIDAAVT NVEDLSSLEE YLATLGRKHR
Homo sapiens              QFSSPEDCLS SPEFLDHIRK VMLVIDAAVT NVEDLSSLEE YLASLGRKHR
Tetraodon nigroviridis    NCGSTQDCLS SPEFLEHVTK VMLVIDAAVS HLDDLHSLED FLLNLGRKHQ
Danio rerio               NCGDAPECLS SPEFLEHVTK VMLVIDAAVS HLDDLHTLED FLLNLGRKHQ

                          ....|....| ....|....| ....|....| ....|....| ....|....|
                                 110        120        130        140        150
Mus musculus              AVGVRLSSFS TVGESLLYML EKCLGPDFTP ATRTAWSRLY GAVVQAMSRG
Rattus norvegicus         AVGVRLSSFS TVGESLLYML EKCLGPDFTP ATRTAWSQLY GAVVQAMSRG
Homo sapiens              AVGVKLSSFS TVGESLLYML EKCLGPAFTP ATRAAWSQLY GAVVQAMSRG
Tetraodon nigroviridis    AVGVKPQSFA MVGESLLYML QCSLGQAYTA SLRQAWLNMY SVVVASMSRG
Danio rerio               AVGVNTQSFA LVGESLLYML QSSLGPAYTT SLRQAWLTMY SIVVSAMTRG

                          ....|....|
                                 160
Mus musculus              WD--GE----
Rattus norvegicus         WD--GE----
Homo sapiens              WD--GE----
Tetraodon nigroviridis    WAKNGEDKAD
Danio rerio               WAKNGEHKSN
```

Figure 41. Alignement multiple des neuroglobines humaines (*Homo sapiens* (AAH32509)), murines (*Mus musculus* (AAH24263) et *Rattus norvegicus* (AAL47568)) et de poissons téléostéens (*Tetraodon nigroviridis* (CAC59975) et *Danio rerio* (CAC59947)) avec leurs cystéines absentes dans les hémoglobines et les myoglobines de ces organismes.

VIII. Perspectives

(1) Des travaux de caryotypage de *Riftia pachyptila* ont été initié (Bailly et Dixon, non publié) mais n'ont pas abouti et mériteraient d'être poursuivis pour localiser les deux familles sur les chromosomes et retracer l'histoire des duplications au sein du génome entier de *Riftia pachyptila*.

(2) Le séquençage de l'hémoglobine intracellulaire d'*Alvinella pompejana* ainsi que celui de ses hémoglobines extracellulaires et celles d' *Arenicola marina* permettraient de confirmer les hypothèses d'ancestralité sur la position des cystéines libres à partir d'autres lignées de polychètes. En fait, un plus grand nombre de séquences nucléiques d'hémoglobines d'annélides non symbiotiques et symbiotiques vivant ou non dans des milieux riches en H_2S

est nécessaire pour renforcer les analyses d'estimation des substitutions synonymes et non synonyme. De nouvelles séquences permettraient aussi de détecter d'autres sites sous sélection positive dans d'autres lignées.

(3) Les modèles RAS (Chapitre 6, paragraphe 4) sont dits homotachiques car ils reposent sur le postulat que chaque acide aminé orthologue (i.e. dans chaque lignée évolutive) évolue à la même vitesse. (Lopez et al. 2002) ont montré que la réalité de l'évolution des protéines est en fait hétérotachique : la vitesse évolutive d'une position peut varier au cours du temps, et donc dans des lignées. Le modèle de covarion de Fitch utilisé dans nos travaux (chapitre 4) et qui stipule qu'à un moment donné seule une petite fraction des positions (les covarions) accepte des substitutions, est le seul modèle hétérotachique. Cependant ce modèle théorique ne permet pas de reproduire les observations faites sur un pool de 2700 séquences protéiques de cytochrome b (Lopez et al. 2002). Au début des années 90 l'introduction des modèles RAS avaient déjà permis d'améliorer les reconstructions des phylogénies quand les taxons étaient très divergents et en limitant l'artefact bien connu de l'attraction des longues branches. Il faut espérer que les prochains modèles d'estimation des substitutions non synonymes et synonymes pourront intégrer de nouveaux modèles générant de l'hétérotachie afin de rendre encore plus performante la détection de sites cryptiques sous sélection positive.

(4) Pour comprendre les mécanismes moléculaires fins de fixation et de transport de H_2S il serait intéressant de maîtriser et continuer la surexpression des hémoglobines de *Riftia pachyptila* que nous avons commencé pendant ma thèse (fig. 42) ainsi qu'un test fonctionnel de fixation des sulfures à partir d'hémoglobine recombinante. Ce sont les pré-requis pour faire de la mutagenèse dirigée sur les cystéines : l'idée serait de mettre une cystéine en +11 quand elle se trouve naturellement en +1 et vice et versa. La seconde étape serait de surexprimer une hémoglobine de *Lumbricus terrestris* ayant une cystéine en + 11 sur A2 et +1 sur B2 pour restaurer l'éventuelle fonctionnalité (perdue) de fixation des sulfures.

(5) Bien sur la surexpression de chacune des globines, soit séparément soit sur un même vecteur d'expression permettrait la reconstruction *in vitro* et donc de comprendre les étapes d'association de ces édifices. La fin du séquençage des chaînes de structure de *Riftia pachyptila* (Chapitre 5) permettrait également de reconstruire l'hémoglobine en bicouches hexagonale à partir d'une surexpression de tous les gènes codant pour cette structure (Chapitre 2).

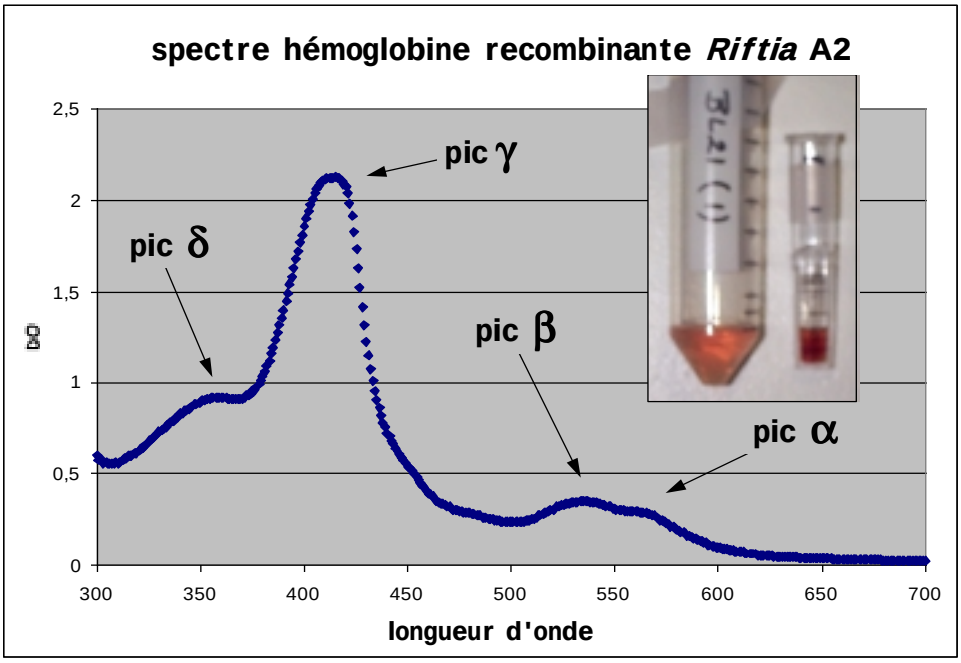


Figure 42. Spectre de l'hémoglobine recombinante où l'on retrouve les quatre pics caractéristiques des hémoglobines (après ultra-sonication et lyse des bactéries (orange) et concentrée après ultra-filtration (rouge) dans l'encadré).

(6) L'étude de la force des promoteurs des globines de *Riftia pachyptila* couplé avec un gène rapporteur serait intéressante dans la mesure où la production d'hémoglobine dans ce ver géant est proportionnelle à sa taille. Cette production peut s'expliquer soit parce que le tissu hématopoïétique occupe une large place (la localisation *in situ* à partir de sonde ARN de séquences codantes et de la région 3'UTR de la globine A2 est en cours), soit parce que les niveaux de transcription sont particulièrement élevés.

(7) Enfin, le séquençage d'autres globines de polychètes, d'oligochètes et d'achètes devrait permettre de concevoir des amorces "universelles" poly- oligo- et achète spécifique de chaque familles A1, A2, B1 et B2. En effet, comme le montrent les alignements de l'annexe 4, des zones conservées aux environs des parties NH2 et COOH terminal devraient permettrent de concevoir des amorces dégénérées suffisamment spécifiques. Un projet de grande envergure en collaborations consisterait à récolter des annélides dans des environnements réduits terrestres et marins aussi divers que les mangroves, les estuaires, les sources hydrothermales, les suintements froids, les grottes riches en sulfures (Mexique, Italie, Roumanie, Etats-Unis) mais aussi des milieux oxydés terrestres et marins de l'ensemble du globe. Les annélides sont de bons indicateurs biogéographiques et sont adaptés à de multiples conditions

environnementales. Leur hémoglobine est donc une bonne cible pour détecter des évènements d'adaptation moléculaire avce l'utilisation d'outils d'analyses de plus en plus puissant. Un tel travail permettrait de combiner une approche de biodiversité écologique avec de la "génétique adaptative" afin de comprendre les directions évolutives d'une protéine au sein d'un phylum aussi vaste que celui des annélides. L'hémoglobine, parce qu'elle est universelle, est une source d'exemples d'adaptation et de ce fait un modèle théorique adéquat pour modéliser ces processus moléculaires évolutifs (adaptatifs). Rappelons nous qu'elle est à la fois protéine fonctionnelle et enzymatique (Chapitre 1). Cette approche toujours couplée aux modèles permettant de détecter les sites sous sélection positive pourrait permettre de mieux formaliser la démarche de l'ingénierie génétique (étendue à l'ensemble des protéines) et particulièrement de **l'évolution dirigée *in vitro*** : elle pourrait permettre de mutagénéiser des acides aminés bien ciblés et diminuer ainsi les constructions inutiles ou éviter d'avoir recours à la puissante et remarquable technique du DNA shuffling qui est assez aléatoire et dont les protocoles sont complexes surtout sur le plan du criblage. A titre d'exemple (parmi d'autres), cette technique a permis d'augmenter significativement l'activité péroxidase de la myoglobine du cœur de cheval (Wan et al. 1998).

Bibliographie

BIBLIOGRAPHIE

Andersson C, Rostergaard JE, Llewellyn DJ, Dennis ES, Peacock WJ (1996) A new hemoglobin gene from soybean: A role for hemoglobin in all plants. P.N.A.S. 93:5682-5687

Appleby CA (1992) The origin and functions of haemoglobin in plants. Sc. Progress. 76:365-398

Arp AJ, Childress JJ, Vetter RD (1987) The sulphide-binding protein in the blood of the vestimentiferan tube-worm *Riftia pachyptila*, is the extracellular hemoglobin. J. exp. Biol. 128 : 139-158

Arp AJ, Menon JG, Julian D (1995) Multiple mechanisms provide tolerance to environmental sulfide in *Urechis caupo*. Amer. Zool. 35 : 132-144

Arredondo-Peter R, Hargrove MS, Moran JF, Sarath G, Klucas RV (1998) Plant hemoglobins. Plant. Physiol. 118 : 1121-5

Ash J, Ke Y, Korb M, Johnson LF (1993) Introns are essential for growth-regulated expression of the mouse thymidylate synthase gene. Mol. Cell. Biol. 13 : 1565-71

Awenius C, Hankeln T, Burmester T (2001) Neuroglobins from the zebrafish *Danio rerio* and the pufferfish *Tetraodon nigroviridis*. Biochem. Biophys. Res. Commun. 287 : 418-21

Bailly X, Jollivet D, Vanin S, Deutsch J, Zal F, Lallier FH, Toulmond A (2002) Evolution of the sulfide binding function within the globin multigenic family of the deep-sea hydrothermal vent tubeworm *Riftia pachyptila*. Mol. Biol. Evol. 19 : 1421-1433

Baum AD, Larson (1991) Adaptation reviewed: a phylogenetic methodology for studying character macroevolution. Syst. Zool 40 : 1-18

Beale D, Lehmann H (1965) Abnormal haemoglobin and the genetic code. Nature 207 : 259-261

Bielawski JP, Yang Z (2001) Positive and negative selection in the DAZ gene family. Mol. Biol. Evol. 18 : 523-9

Bogusz D, Appleby CA, Landsmann J, Dennis ES, Trinick MJ, Peacock WJ (1988) Functioning haemoglobin genes non-modulating plants. Nature 331 : 178-180

Bonaventura J, Lance VP (2001) Nitric oxide, Invertebrates and hemoglobin. Am. Zool. 41 : 346-359

Boore JL, Brown WM (2000) Mitochondrial genomes of *Galathealinum, Helobdella*, and *Platynereis*: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol. Biol. Evol. 17 : 87-106

Breton-Gorius J (1963) Etude au microscope électronique des cellules chloragogènes d'*Arenicola marina* L. Annal. Sc. Nat. Zool. 5 : 211-272

Burmester T, Hankeln T (1999) A globin gene of Drosophila melanogaster. Mol. Biol. Evol. 16 : 1809-11

Burmester T, Weich B, Reinhardt S, Hankeln T (2000) A vertebrate globin expressed in the brain. Nature 407 : 520-3

Burmester T (2001) Molecular evolution of the arthropod hemocyanin superfamily. Mol. Biol. Evol. 18 : 184-95

Burmester T, Ebner B, Weich B, Hankeln T (2002) Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. Mol. Biol. Evol. 19 : 416-21

Burr AH, Hunt P, Wagar DR, Dewilde S, Blaxter ML, Vanfleteren JR, Moens L (2000) A hemoglobin with an optical function. J. Biol. Chem. 275 : 4810-5

Cappai R, Kaslow DC, Peterson MG, Cowman AF, Anders RF, Kemp DJ (1992) Cloning and analysis of the RESA-2 gene: a DNA homologue of the ring-infected erythrocyte surface antigen gene of Plasmodium falciparum. Mol. Biochem. Parasitol. 54 : 213-21

Cavanaugh C (1983) Symbiotic chemoautotrophic bacteria in marine invertebrates from sulphide-rich habitats. Nature 302 : 58-61

Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB (1981) Prokaryotic cells in the hydrothermal vent tube-worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. Science 213 : 340-342

Chen L, DeVries AL, Cheng CH (1997) Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. P.N.A.S.94 : 3811-6

Corliss J, Ballard RD (1977) Oasis of life in the cold abyss. Natl. Geogr. 152 : 441-454

Couture M, Chamberland H, St-Pierre B, Lafontaine J, Guertin M (1994) Nuclear genes encoding chloroplast hemoglobins in the unicellular green alga *Chlamydomonas eugametos*. Mol. Gen. Genet. 243 : 185-97

Couture M, Das TK, Lee HC, Peisach J, Rousseau DL, Wittenberg BA, Wittenberg JB, Guertin M (1999a) *Chlamydomonas* chloroplast ferrous hemoglobin. Heme pocket structure and reactions with ligands. J. Biol. Chem. 274 : 6898-910

Couture M, Yeh SR, Wittenberg BA, Wittenberg JB, Ouellet Y, Rousseau DL, Guertin M (1999b) A cooperative oxygen-binding hemoglobin from *Mycobacterium tuberculosis*. P.N.A.S.96 : 11223-8

Craik CS, Buchman SR, Beychok S (1980) Characterization of globin domains: heme binding to the central exon product. P.N.A.S.77 : 1384-8.

Craik CS, Buchman SR, Beychok S (1981) O2 binding properties of the product of the central exon of beta-globin gene. Nature 291 : 87-90.

Darboux I, Pauchet Y, Castella C, Silva-Filha MH, Nielsen-LeRoux C, Charles JF, Pauron D (2002) Loss of the membrane anchor of the target receptor is a mechanism of bioinsecticide resistance. P.N.A.S.99 : 5830-5

Dewilde S, Blaxter M, Van Hauwaert ML, Vanfleteren J, Esmans EL, Marden M, Griffon N, Moens L (1996) Globin and globin gene structure of the nerve myoglobin of *Aphrodite aculeata*. J. Biol. Chem. 271 : 19865-70

Dewilde S, Blaxter M, Van Hauwaert ML, Van Houte K, Pesce A, Griffon N, Kiger L, Marden MC, Vermeire S, Vanfleteren J, Esmans E, Moens L (1998a) Structural, functional, and genetic characterization of *Gastrophilus* hemoglobin. J. Biol. Chem. 273 : 32467-74

Dewilde S, Winnepenninckx B, Arndt MH, Nascimento DG, Santoro MM, Knight M, Miller AN, Kerlavage AR, Geoghagen N, Van Marck E, Liu LX, Weber RE, Moens L (1998b) Characterization of the myoglobin and its coding gene of the mollusc *Biomphalaria glabrata*. J. Biol. Chem. 273 : 13583-92

Dewilde S, Van Hauwaert ML, Vinogradov S, Vierstraete A, Vanfleteren J, Moens L (2001) Protein and gene structure of a chlorocruorin chain of *Eudistylia vancouverii*. Biochem. Biophys. Res. Commun. 281 : 18-24.

Durner J, Gow AJ, Stamler JS, Glazebrook J (1999) Ancient origins of nitric oxide signaling in biological systems. P.N.A.S.96 : 14206-7

Dykhuizen D, Hartl DL (1980) Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background. Genetics 96 : 801-17

Edmond JM, Van Damm KL, McDuff RE, Measures CI (1982) Chemistry of hot springs on the EPR and their effluent dispersal. Nature 297 : 187-191

El Idrissi Slitine F (1991) Thèse de Doctorat: Etude structurale et fonctionnelle d'hémoglobines extracellulaires d'annélides (erythrocruorines) Université Pierre et Marie Curie (Paris 6)

Eto K, Asada T, Arima K, Makifuchi T, Kimura H (2002) Brain hydrogen sulfide is severely decreased in Alzheimer's disease. Biochem. Biophys. Res. Commun. 293 : 1485-8

Fedeli D, Tiano L, Gabbianelli R, Caulini GC, Wozniak M, Falcioni G (2001) Hemoglobin components from trout (*Salmo irideus*): determination of their peroxidative activity. Comp. Biochem. Physiol. B. 130 : 559-64

Fedorov A, Fedorova L, Starshenko V, Filatov V, Grigor'ev E (1998) Influence of exon duplication on intron and exon phase distribution. J. Mol. Evol.46 : 263-71

Felbeck H (1981) Chemoautotrophic potential of the hydrothermal vent tubeworm, *Riftia pachyptila* Jones (Vestimentifera). Science 213 : 336-338

Felbeck H, Somero GN, Childress JJ (1981) Calvin-Benson cycle and sulfide oxidation enzymes in animals from sulfide-rich habitats. Nature 293 : 291-293

Felbeck H, Powell MA, Hand SC, Somero GN (1985) Metabolic adaptations of hydrothermal vent animals. In: Jones ML (ed) Hydrothermal vents of the Eastern Pacific : an overview. Bulletin of the Biological Society of Washington, p 261-272

Fitch DHA, Bailey WJ, Tagle DA, Goodman M, Sieu L, Slightom JL (1991) Duplication of the gamma Gene Mediated by L1 Long Interspersed Repetitive Elements in an Early Ancestor of Simian Primates. P.N.A.S. 7396-7400

Fitch WM (1966) An improved method of testing for evolutionary homology. J. Mol. Biol. 16 : 9-16

Fitch WM (1967) Evidence suggesting a non-random character to nucleotide replacement in naturally occuring mutations. J. Mol. Biol. 26 : 499-507

Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst. Zool. 19 : 99-113

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155 : 279-84.

Fushitani K, Ochi O, Morimoto H (1982) Purification, characterization, and comparison of extracellular hemoglobins in coelomic fluid and blood vessel plasma of *Travisia japonica* (Polychaeta, Annelid). Comp. Biochem. Physiol. B 72 : 267-73

Fushitani K, Matsuura S, Riggs A (1988) The amino acid sequences of chains a, b, and c that form the trimer subunit of the extracellular hemoglobin from *Lumbricus terrestris*. J. Biol. Chem. 263 : 6502-6517

Fustec A, Desbruyères D, Juniper SK (1987) Deep-sea hydrothermal vent communities at 13° N on the East Pacific Rise: microdistribution and temporal variations. Biol. Océanogr. 4 : 121-164.

Fustec A, Desbruyeres D, Laubier L (1988) Biomass estimation of animals communities associated with deep-sea hydrothermal vents near 13°N/EPR. Oceanol. Acta 8 n°SP : 15-22

Gardner PR, Gardner AM, Martin LA, Salzman AL (1998) Nitric oxide dioxygenase: an enzymic function for flavohemoglobin. P.N.A.S. 95: 10378-83

Garry DJ, Ordway GA, Lorenz JN, Radford NB, Chin ER, Grange RW, Bassel-Duby R, Williams RS (1998) Mice without myoglobin. Nature 395 : 905-8

Gilbert W (1978) Why genes in pieces? Nature 271 : 501.

Gilbert W, de Souza SJ, Long M (1997) Origin of genes. P.N.A.S.94 : 7698-703

Go M (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291 : 90-2.

Golding GB, Dean AM (1998) The structural basis of molecular adaptation. Mol. Biol. Evol. 15 : 355-69

Golstein DB, Schlötterer C (1999) Microsatellites : Evolution and Applications. Oxford university press

Goodman M, Barnabas J, Matsuda G, Moore GW (1971) Molecular evolution in the descent of man. Nature 233 : 604-13.

Goodman M, Moore GW, Matsuda G (1975) Darwinian evolution in the genealogy of hemoglobin. Nature 253 : 603-608

Gotoh T, Shishikura F, Snow JW, Ereifej KI, Vinogradov SN, Walz DA (1987) Two globin strains in the giant annelid extracellular haemoglobins. Biochem. J. 241 : 441-5

Grieshaber MK, Volkel S (1998) Animal adaptations for tolerance and exploitation of poisonous sulfide. Annu Rev Physiol 60 : 33-53

Halanych KM, Feldman RA, Vrijenhoek RC (2001) Molecular evidence that *Sclerolinum brattstromi* is closely related to vestimentiferans, not to frenulate pogonophorans (Siboglinidae, Annelida). Biological Bulletin 201 : 65-75

Hall BG, Zuzel T (1980) Evolution of a new enzymatic function by recombination within a gene. P.N.A.S.77 : 3529-33

Hardison RC (1996) A brief history of hemoglobins: plant, animal, protist, and bacteria. P.N.A.S.93 : 5675-9

Hauschild K, Weber WM, Clauss W, Grieshaber MK (1999) Excretion of thiosulphate, the main detoxification product of sulphide, by the lugworm *Arenicola marina* . J. Exp. Biol. 202 : 855-866

Holland SK, Blake CC (1987) Proteins, exons and molecular evolution. Biosystems 20 : 181-206

Hou S, Larsen RW, Boudko D, Riley CW, Karatan E, Zimmer M, Ordal GW, Alam M (2000) Myoglobin-like aerotaxis transducers in Archaea and Bacteria. Nature 403 : 540-4.

Hou S, Belisle C, Lam S, Piatibratov M, Sivozhelezov V, Takami H, Alam M (2001) A globin-coupled oxygen sensor from the facultatively alkaliphilic *Bacillus halodurans* C-125. Extremophiles 5 : 351-4

Hou S, Freitas T, Larsen RW, Piatibratov M, Sivozhelezov V, Yamamoto A, Meleshkevitch EA, Zimmer M, Ordal GW, Alam M (2001) Globin-coupled sensors: a class of heme-containing sensors in Archaea and Bacteria. P.N.A.S.98 : 9353-8

Hourdez S (2000) Adaptations respiratoires des annélides polychètes de milieux abyssaux hypoxiques. Thèse de Doctorat de l'Université Pierre et Marie Curie, Paris VI

Hvitved AN, Trent JT, 3rd, Premer SA, Hargrove MS (2001) Ligand binding and hexacoordination in *Synechocystis* hemoglobin. J. Biol. Chem. 276 : 34714-21

Imamura T, Baldwin T, Riggs A (1972) The amino acid sequence of the monomeric hemoglobin component from the bloodworm, Glyat liver. J.Biol.Chem. 247 : 2785-2797

Ingram V (1956) A specific chemical difference between the globins of normal human and sickled-cell anemia haemogobin. Nature 178 : 792-794

Ingram V (1957) Gene mutations in human haemoglobin: the chemical difference between normal and sickled-cell haemogobin. Nature 180 : 326-328

Ingram V (1961) Gene volution and the haemoglobin. Nature 189 : 704-708

Itano HA (1957) The humans hemoglobins: their properties and genetic control. Adv. Protein. Chem 12 : 215-268

Iwaasa H, Takagi T, Shikama K (1989) Protozoan myoglobin from *Paramecium caudatum*. Its unusual amino acid sequence. J. Mol. Biol. 208 : 355-8

Iwaasa H, Takagi T, Shikama K (1990) Protozoan hemoglobin from *Tetrahymena pyriformis*. Isolation, characterization, and amino acid sequence. J. Biol. Chem. 265 : 8603-9

Jensen EO, Paludan K, Hyldig-Nielsen JJ, Jorgensen P, Marcker KA (1981) The structure of a chromosomal leghaemoglobin gene from soybean. Nature 291 : 677-679

Jhiang MS, Garey JR, Riggs AF (1988) Exon-intron organization in genes of earthworm and vertebrate globins. Science 240 : 334-336

Jia L, Bonaventura C, Bonaventura L, Stamier JS (1996) S-nitrosohaemoglobin: a dynamic activity of blood involved in vascular control. Nature 380 : 221-226

Jones ML (1981) *Riftia pachyptila* Jones: observations on the Vestimentiferan worm from the Galapagos rift. Science 213 : 333-336

Jones ML (1985) On the vestimentifera, new phylum: six new species, and other taxa, from hydrothermal vents and elsewhere. Biol. Soc. Wash. Bull. 6 : 117-158

Jukes TH (2000) The Neutral Theory of Molecular Evolution. Genetics 154 : 956-958

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S (1996) Sequence analysis of the genome of the unicellular *cyanobacterium Synechocystis sp.* strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions (supplement). DNA Res 3 : 185-209

Kawada N, Kristensen DB, Asahina K, Nakatani K, Minamiyama Y, Seki S, Yoshizato K (2001) Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells. J. Biol. Chem. 276 : 25318-23

Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, Shore VC (1960) Structure of myoglobin. Nature 185 : 422-427

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217 : 624-6.

Kimura M (1981) Was globin evolution very rapid in its early stages?: a dubious case against the rate-constancy hypothesis. J. Mol. Evol.17 : 110-3

Kimura T, Tabuchi Y, Takeguchi N, Asano S (2002) Mutational study on the roles of disulfide bonds in the beta-subunit of gastric H+,K+-ATPase. J. Biol. Chem. 277 : 20671-7

King JL, Jukes TH (1969) Non-Darwinian evolution. Science 164 : 788-98.

Kleinschmidt T, Weber RE (1998) Primary structures of *Arenicola marina* isomyoglobins: molecular basis for functional heterogeneity. Biochim Biophys Acta 1383 : 55-62

Koike C, Fung JJ, Geller DA, Kannagi R, Libert T, Luppi P, Nakashima I, Profozich J, Rudert W, Sharma SB, Starzl TE, Trucco M (2002) Molecular basis of evolutionary loss of the alpha 1,3-galactosyltransferase gene in higher primates. J. Biol. Chem. 277 : 10114-20

Kojima S, Hashimoto T, Hasegawa M, Murata S, Ohta S, Seki H, Okada N (1993) Close phylogenetic relationship between Vestimentifera (tube worms) and Annelida revealed by the amino acid sequence of elongation factor-1 alpha. J. Mol. Evol.37 : 66-70

Komiyama NH, Miyazaki G, Tame J, Nagai K (1995) Transplanting a unique allosteric effect from crocodile into human haemoglobin. Nature 373 : 244-6

Kraus DW, Wittenberg JB (1990) Hemoglobins of the *Lucina pectinata*/bacteria symbiosis. I. Molecular properties, kinetics and equilibria of reactions with ligands. J. Biol. Chem. 265 : 16043-53

Kuiper GG, Klootwijk W, Visser TJ (2002) Substitution of cysteine for a conserved alanine residue in the catalytic center of type II iodothyronine deiodinase alters interaction with reducing cofactor. Endocrinology 143 : 1190-8

LaCount MW, Zhang E, Chen YP, Han K, Whitton MM, Lincoln DE, Woodin SA, Lebioda L (2000) The crystal structure and amino acid sequence of dehaloperoxidase from *Amphitrite ornata* indicate common ancestry with globins. J. Biol. Chem. 275 : 18712-6.

LeClerc JE, Li B, Payne WL, Cebula TA (1996) High mutation frequencies among *Escherichia coli* and *Salmonella pathogens*. Science 274 : 1208-11

Lee RW (1999) Oxidation of sulfide by *spartina alterniflora* roots. Limnology and Oceanography 44 : 1155-1159

Lewin R (1981) Evolutionary history written in globin genes. Science 214 : 426-7, 429.

Liberge M, Gros O, Frenkiel L (2001) Lysosomes and sulfide-oxidising bodies in the bacteriocytes of *Lucina pectinata*, a cytochemical and microanalysis approach. Marine Biology 139 : 401-409

Lonsdale PF (1977) Clustering of suspension-feeding macrobenthos near abyssal hydrothermal vents at oceanic spreading center. Deep Sea Research 24 : 857-863

Lopez P, Casane D, Philippe H (2002) Heterotachy, an important process of protein evolution. Mol. Biol. Evol. 19 : 1-7

Luther GW, Rozan TF, Taillefert M, Nuzzio DB, Di Meo C, Shank TM, Lutz RA, Cary SC (2001) Chemical speciation drives hydrothermal vent ecology. Nature 410 : 813-816

Margoliash E (1963) Primary structure and evolution of cytochrome C. P.N.A.S. 50 : 672-679

Marsh AG, Mullineaux LS, Young CM, Manahan DT (2001) Larval dispersal potential of the tubeworm *Riftia pachyptila* at deep-sea hydrothermal vents. Nature 411 : 77-80

McHugh D (1997) Molecular evidence that echiurans and pogonophorans are derived annelids. P.N.A.S.94 : 8006-9

McMullin ER, Bergquist DC, Fischer CR (2000) metazoans in extreme environments: adaptation of hydrothermal vent and hydrocarbon seep fauna. Gravitational and Space Biology Bulletin 13 : 13-23

Minning DM, Gow AJ, Bonaventura J, Braun R, Dewhirst M, Goldberg DE, Stamler JS (1999) *Ascaris* haemoglobin is a nitric oxide-activated 'deoxygenase'. Nature 401 : 497-502

Moens L, Dewilde S (2000) Globins in the brain. Nature 407 : 461-2

Moens L, Vanfleteren J, Vandepeer Y, Peeters K, Kapp OH, Czeluzniak J, Goodman M, Blaxter M, Vinogradov SN (1996) Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationships. Mol. Biol. Evol. 13 : 324-333

Nakai K, Horton P (1999) PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization. Trends Biochem. Sci. 24 : 34-6

Nguyen BD, Zhao X, Vyas K, La Mar GN, Lile RA, Brucker EA, Phillips GN, Jr., Olson JS, Wittenberg JB (1998) Solution and crystal structures of a sperm whale myoglobin triple mutant that mimics the sulfide-binding hemoglobin from *Lucina pectinata*. J. Biol. Chem. 273 : 9517-26

Nicholls P (1975) The effect of sulphide on cytochrome aa3. Isosteric and allosteric shifts of the reduced alpha-peak. Biochim. Biophys. Acta 396 : 24-35.

Nirenberg M, Leder P, Bernfield M, Brimacombe R, Trupin J, Rottman F, O'Neal C (1965) RNA codewords and protein synthesis, VII. On the general nature of the RNA code. P.N.A.S. 53

Otha T (1988) Further simulation studies on evolution by gene duplication. Evolution 42 : 375-386

Otha T (1993) Pattern of nucleotide substitutions in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. Genetics 134 : 1271-1276

Otha T (1994) Further examples of evolution by gene duplication revealed through DNA sequence comparisons. Genetics 138 : 1331-1337

Ownby DW, Zhu H, Schneider K, Beavis RC, Chait BT, Riggs AF (1993) The extracellular hemoglobin of the earthworm, *Lumbricus terrestris*. Determination of subunit stoichiometry. J. Biol. Chem. 268 : 13539-47

Pallavicini A, Negrisolo E, Barbato R, Dewilde S, Ghiretti-Magaldi A, Moens L, Lanfranchi G (2001) The primary structure of globin and linker chains from the chlorocruorin of the polychaete *Sabella spallanzanii*. J. Biol. Chem. 276 : 26384-90.

Parrino V, Kraus DW, Doeller JE (2000) ATP production from the oxidation of sulfide in gill mitochondria of the ribbed mussel *Geukensia demissa*. J. Exp. Biol. 203 : 2209-2218

Pastore A, Lesk AM (1990) Comparison of the structures of globins and phycocyanins: evidence for evolutionary relationship. Proteins 8 : 133-55

Patel S, Spencer CP (1963) The oxydation of sulfide by the heam compounds from the blood of *Arenicola marina*. J.mar. biol. Ass. U.K. 43 : 167-175

Pauling L, Itano HA, Singer SJ, Wells IC (1949) Sickle cell anemia, a molecular disease. Science 110 : 543

Perutz MF (1983) Species adaptation in a protein molecule. Mol. Biol. Evol. 1 : 1-28

Perutz MF, Rossman MG, Cullis AF, Muirhead H, Will G, North ACT (1960) Structure of haemoglobin. Nature 185 : 416-422

Pesce A, Couture M, Dewilde S, Guertin M, Yamauchi K, Ascenzi P, Moens L, Bolognesi M (2000) A novel two-over-two alpha-helical sandwich fold is characteristic of the truncated hemoglobin family. Embo J 19 : 2424-34

Poole RK, Hughes MN (2000) New functions for the ancient globin family: bacterial responses to nitric oxide and nitrosative stress. Mol Microbiol 36 : 775-83

Potts M, Angeloni SV, Ebel RE, Bassam D (1992) Myoglobin in a cyanobacterium. Science 256 : 1690-1

Powell MA, Somero GN (1985) Sulfide oxidation occurs in the animal tissue of the gutless clam, *Solemya reidi.* Biol. Bull. 169 : 164-181

Powell MA, Somero GN (1986) Hydrogen sulfide oxidation is coupled to oxidative phosphorilation in mitochondria of *Solemya reidi*. Science 233 : 563-566

Reichert ET, Brown AP (1907) The crystallography of hemoglobins. Proc. Soc. Exp. Biol. Med 5 : 66-68

Reichert ET, Brown AP (1909) The differentiation and specificity of corresponding proteins and other vital substances in relation to biological classification and organic evolution: the crystallography of hemoglobin. Carnegie Institution of Washington, Pub. No. 116.

Riggs A (1959) Molecular adaptation in haemoglobins: nature of the bohr effect. P.N.A.S. 183 : 1037-1038

Rouse GW (2001) A cladistic analysis of Siboglinidae Caullery, 1914 (Polychaeta, Annelida) : formerly the phyla Pogonophora and Vestimentifera. Zoological Journal of the Linnean Society 132 : 55-80

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

Royer WE, Jr., Strand K, van Heel M, Hendrickson WA (2000) Structural hierarchy in erythrocruorin, the giant respiratory assemblage of annelids. P.N.A.S.97 : 7107-11

Runnegar B (1984) Derivation of the globins from type b cytochromes. J. Mol. Evol.21 : 33-41

Schirmer T, Bode W, Huber R, Sidler W, Zuber H (1985) X-ray crystallographic structure of the light-harvesting biliprotein C-phycocyanin from the thermophilic cyanobacterium *Mastigocladus laminosus* and its resemblance to globin structures. J Mol Biol 184 : 257-77

Schopf JW, Walter MR (1982) Origin and early evolution of cyanobacteria: The geological evidence. In: Carr NG, Whitton BA, eds. The biology of cyanobacteria. Oxford, UK: Blackwell Scientific. 543-564

Sherman DR, Guinn B, Perdok MM, Goldberg DE (1992) Components of sterol biosynthesis assembled on the oxygen-avid hemoglobin of *Ascaris*. Science 258 : 1930-2.

Shishikura F (1996) Amino acid sequence of the monomer subunit of the extracellular hemoglobin of the earthworm, *Pheretima hilgendorfi*. Zoolog Sci 13 : 551-8

Shishikura F, Nakamura M (1996) A comparative study on earthworm hemoglobins: an amino acid sequence comparison of monomer globin chains of two species, *Pontodrilus matsushimensis* and *Pheretima communissima* that belong to the family Megascolecidae. Zoolog Sci 13 : 849-56

Shishikura F, Ochiai T, Yamanaka I (1997) Leech extracellular hemoglobin: two globin strains that are akin to vertebrate hemoglobin alpha and beta chains. Zoolog Sci 14 : 923-30

Southward EC (1988) Development of the gut and segmentation of newly settled stages of *Ridgea* (Vestimentifera): implications for relationship between vestimentifera and pogonophora. J. Mar. Biol. Ass. U.K. 68 : 465-487

Sowa AW, Guy PA, Sowa S, Hill RD (1999) Nonsymbiotic haemoglobins in plants. Acta Biochim. Pol. 46 : 431-45

Sun Y, Jin K, Mao XO, Zhu Y, Greenberg DA (2001) Neuroglobin is up-regulated by and protects neurons from hypoxic-ischemic injury. P.N.A.S. 98 : 15306-11

Suzuki T, Kapp OH, Gotoh T (1988) Novel S-S loops in the giant hemoglobin of *Tylorrhynchus heterochaetus*. J. Biol. Chem. 263 : 18524-9

Suzuki T, Takagi T, Okuda K, Furukohri T, Ohta S (1989) The deep-sea tube worm hemoglobin: subunit structure and phylogenetic relationship with annelid hemoglobin. Zool. Sci. 6 : 915-926

Suzuki T, Takagi T, Ohta S (1990a) Primary structure of a constituent polypeptide chain (AIII) of the giant haemoglobin from the deep-sea tube worm *Lamellibrachia*. A possible H2S-binding site. Biochem J 266 : 221-5

Suzuki T, Takagi T, Ohta S (1990b) Primary structure of a linker subunit of the tube worm 3000-kDa hemoglobin. J. Biol. Chem. 265 : 1551-5

Suzuki T, Takagi T, Ohta S (1993) N-terminal amino acid sequences of 440 kDa hemoglobins of the deep-sea tube worms, *Lamellibrachia sp.*1, *Lamellibrachia sp.*2 and slender vestimentifera gen. sp.1 evolutionary relationship with annelid hemoglobins. Zoolog. Sci. 10 : 141-6

Suzuki T, Riggs AF (1993) Linker chain L1 of earthworm hemoglobin. Structure of gene and protein: homology with low density lipoprotein receptor. J. Biol. Chem. 268 : 13548-55

Suzuki T, Hirao Y, Vinogradov SN (1995) Primary structure of a constituent polypeptide chain of the chlorocruorin from *Sabellastarte indica*. Biochim. Biophys. Acta 1252 : 189-93

Suzuki T, Yuasa H, Imai K (1996) Convergent evolution. The gene structure of *Sulculus* 41 kDa myoglobin is homologous with that of human indoleamine dioxygenase. Biochim. Biophys. Acta 1308 : 41-8

Takagi T, Iwaasa H, Ohta S, Suzuki T (1991) Primary structure of 440 kDa hemoglobin from the deep sea tubeworm *Lamellibrachia.* In: Vinogradov SN, Kapp OH (eds) Structure and Function of Invertebrate Oxygen Carriers. Springer, New York, p 245-249

Tarricone C, Galizzi A, Coda A, Ascenzi P, Bolognesi M (1997) Unusual structure of the oxygen-binding site in the dimeric bacterial hemoglobin from *Vitreoscilla sp.* Structure 5 : 497-507

Thiermann F, Vismann B, Giere O (2000) Sulphide tolerance of the marine nematode Oncholaimus campylocercoides-a result of internal sulphur formation? Mar. Ecol. Progr. Ser. 193 : 251-259

Tjepkema JD, Cashon RE, Beckwith J, Schwintzer CR (2002) Hemoglobin in *Frankia,* a nitrogen-fixing actinomycete. Appl. Environ. Microbiol. 68 : 2629-31

Trent JT, 3rd, Hargrove MS (2002) A ubiquitously expressed human hexacoordinate hemoglobin. J. Biol. Chem. 277 : 19538-45

Trevaskis B, Watts RA, Andersson CR, Llewellyn DJ, Hargrove MS, Olson JS, Dennis ES, Peacock WJ (1997) Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins. P.N.A.S. 94 : 12230-4

Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC (2002) Evolution and biogeography of deep-sea vent and seep invertebrates. Science 295 : 1253-7

Vetter RD, Wells ME, Kurtsman AL, Somero GN (1987) Sulfide detoxication by the hydrothermal vent crab *Bythogrea thermydron* and other decapod crustaceans. Physiological Zoology 60 : 121-137

Vogel F, Röhrborn G (1966) Amino-acid substitution in haemoglobins and the mutation process. Nature 210 : 116-117

Wan L, Twitchett MB, Eltis LD, Mauk AG, Smith M (1998) In vitro evolution of horse heart myoglobin to increase peroxidase activity. P.N.A.S.95 : 12825-31

Watson HC, Kendrew JC (1961) Comparison between the amino-acid sequences of sperm-whale myoglobin and human myoglobin. Nature 190 : 670-672

Watson JD, Crick FHC (1953) Molecular structure of nucleic acids. Nature 171 : 737-738

Watts RA, Hunt PW, Hvitved AN, Hargrove MS, Peacock WJ, Dennis ES (2001) A hemoglobin from plants homologous to truncated hemoglobins of microorganisms. P.N.A.S. 98 : 10119-24

Weber RE, Vinogradov SN (2001) Nonvertebrate hemoglobins: Functions and molecular adaptations. Physiological Reviews 81 : 569-628

Westheide W (1997) Direction of evolution within polychaeta. Journal of Natural History 31 : 1-15

Williams EJ, Pal C, Hurst LD (2000) The molecular evolution of signal peptides. Gene 253 : 313-22

Wittenberg JB, Bolognesi M, Wittenberg BA, Guertin M (2002) Truncated hemoglobins: a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants. J. Biol. Chem. 277 : 871-4

Woese CR, Kandler O, Wheelis ML (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. P.N.A.S.87 : 4576-9

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood CABIOS. CABIOS 13 : 555-556

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. 15 : 568-73

Yang Z, Bielawski JP (2000) Statistical mehods for detecting molecular adaptation. Tree 15 : 496-502

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155 : 431-49

Yang Z, Nielsen R (2002) Codon-substitution model for detecting molecular adaptation at individual sites along specific lineage. Mol. Biol. Evol. 19 : 908-917

Young CM, Vazquez E, Metaxas A, Tyler PA (1996) Embryology of vestimentiferan tube worms from deep-sea methane/sulphide seeps. Nature 381 : 514-516

Yuasa HJ, Green BN, Takagi T (1996) Electrospray ionization mass spectrometic composition of the 400 kDa hemoglobin from the pogonophoran *Oligobrachia mashikoi* and the primary structures of the three major globin chain. Biochim. Biophys. Acta (submitted)

Zal F, Lallier FH, Green BN, Vinogradov SN, Toulmond A (1996a) The multi-hemoglobin system of the hydrothermal vent tube worm *Riftia pachyptila* .2. Complete polypeptide chain composition investigated by maximum entropy analysis of mass spectra. J. Biol. Chem. 271 : 8875-8881

Zal F, Lallier FH, Wall JS, Vinogradov SN, Toulmond A (1996b) The multi-hemoglobin system of the hydrothermal vent tube worm *Riftia pachyptila* .1. Reexamination of the number and masses of its constituents. J. Biol. Chem. 271 : 8869-8874

Zal F, Green BN, Lallier FH, Toulmond A (1997a) Investigation by electrospray ionisation mass spectrometry of the extracellular hemoglobin from the polychaete annelid *Alvinella pompejana* : a unusual hexagonal bilayer hemoglobin. Biochemistry 36 : 11777-11786

Zal F, Green BN, Lallier FH, Vinogradov SN, Toulmond A (1997b) Quaternary structure of the extracellular haemoglobin of the lugworm *Arenicola marina*: a multi-angle-laser-light-scattering and electrospray- ionisation-mass-spectrometry analysis. Eur J Biochem 243 : 85-92.

Zal F, Suzuki T, Kawasaki Y, Childress JJ, Lallier FH, Toulmond A (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 29 : 562-74

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

Zal F (1998b) Sulphide binding processes of *Riftia pachyptila* haemoglobins. Cahiers de Biologie Marine 39 : 327-328

Zal F, Leize E, Oros DR, Hourdez S, Van Dorsselaer A, Childress JJ (2000) Haemoglobin structure and biochemical characteristics of the sulphide-binding component from the deep-sea clam *Calyptogena magnifica*. Cahiers De Biologie Marine 41 : 413-423

Zhang C, Wang C, Deng M, Li L, Wang H, Fan M, Xu W, Meng F, Qian L, He F (2002) Full-length cDNA cloning of human neuroglobin and tissue expression of rat neuroglobin. Biochem. Biophys. Res. Commun. 290 : 1411-9

Zhu H, Riggs AF (1992) Yeast flavohemoglobin is an ancient protein related to globins and a reductase family. P.N.A.S.89 : 5015-9

Zierenberg RA, Adams MW, Arp AJ (2000) Life in extreme environments: hydrothermal vents. P.N.A.S.97 : 12961-2.

Zuckerkandle E, Jones RT, Pauling L (1960) A comparison of animal hemoglobin by triptic peptide pattern analysis. P.N.A.S. 46 : 1349-1360

Zuckerkandle E, Pauling L (1962) Molecular disease, Evolution, and Genic Heterogeneity. In Horizons in Biochemistry ed M. Kasha and B. Pullman (New York: Academic Press, 1962)

Zuckerkandle E, Pauling L (1965) Evolutionary Divergence and Convergence in Proteins. In Evolvin genes and proteins. ed: V. Bryson, H.J. Vogel (New York: Academic Press)

Zuckerkandle E (1965) The evolution of hemoglobin. Scientific American 212 : 110-118

Zuckerkandl E (1972) Some aspects of protein evolution. Biochimie 54 : 1095-1102

Zuckerkandl E (1976) Evolutionary processes and evolutionary noise at the molecular level. I. Functional density in proteins. J. Mol. Evol.7 : 167-83

Zuckerkandl E (1976) Evolutionary processes and evolutionary noise at the molecular level. II. A selectionist model for random fixations in proteins. J. Mol. Evol.7 : 269-311

Annexes

Annexe 1 : Légendes complètes : noms des espèces correspondant aux codes utilisés dans les phylogénies des figures 17, 18 et 29

Figure 17. Phylogénie moléculaire de globines

Vitreoscilla stercoraria (VstHb) ; *Lupinus luteu* (LluLegHb); *Oryza sativa* (OsaHb2).*Barbatia lima* (BliHbB);*Barbatia virescens* (BviHb); *Aphrodite aculeata* (AacHb); *Glycera dibranchiata* P3 (GdiHbP3); *Lumbricus terrestris* globine B (LteGbB) et globine E (LteGbD); *Mermis nigrescens* "eye globin" (MniGlbE); *Daphnia magna* premier (Dma1.1) et second domaine (Dma1.2); *Artemia salina* globines E-1 (AsaE-1) and E-7 (AsaE7); *Gasterophilus intestinalis* (GinHb); *Drosophila melanogaster* (DmeGlob1);*Chironomus thummi* Ctt-III and Ctt-X. *Myxine glutinosa* (MglHb); *Physeter catodon* (PcaMb); *Homosapiens* hemoglobines γ (HsaHBG), δ (HsaHBD), ε (HsaHBE) and ζ (HsaHBZ). Neuroglobines humaine (HsaNGB) et murine (MmuNgb); hémoglobines humaines α et β (HsaHBA; HsaHBB); myglobines humaine (HsaMB) et murine (MmuMb,); Hémoglobines murines a et b (MmuHba; MmuHbb)

D'après (Burmester et al. 2000)

Figure 18. Phylogénie moléculaire de globines de vertébrés

Cytoglobine humaine, (HsaCYGB) Cytoglobine de la souris, (MmuCygb); cytoglobine Danio rerio (DreCygb) myoglobine humaine (HsaMB) hémoglobines α (HsaHBA) et β (HsaHBB) hémoglobin lamproie (LzaHbA). hemoglobines 1 ,2 et 3 *Myxine glutinosa* (MglHb1; MglHb2; MglHb3), hemoglobine B *Lampetra zanandreai* (LzaHbB), hemoglobine III *Petromyzon marinus* (PmaHbIII), myoglobine et hémoglobines α et β *Heterodontus portusjacksoni* (HpoMb; HpoHba; HpoHbb), myoglobine *Mustelus antarcticus* (ManMb), hémoglobines α et β *Squalus acanthias* (SacHba; SacHbb), myoglobine et hémoglobines α et β *Mus musculus* (MmuMb; MmuHba; MmuHbb).

D'après (Burmester et al. 2002)

Figure 29. Phylogénie moléculaire de globines d'annélide.

Aphrodite aculeata (Aph,neuroglobine) *Glycera dibranchiata* (Glyc MII (GLB1), MIV, P1, P2, P3); *Urechis caupo* (Urec FI); *Tylorrhynchus heterochaetus*(Tyl I, IIA, IIB, IIC); *Lumbricus terrestris* (Lum I, II, III, IV), *Pheretima sieboldii* (Pher I); T. tubifex (Tub I); *Lamellibrachia* (Lam BI, BII, BIII, BIV)); *Oligobrachia* (Oli 1, 2, 3). myoglobine de *Physeter catodon* (Phys,)

D'après (Dewilde et al. 1996)

Annexe 2 : Matériel et Méthodes

A. Qu'est-ce qu'une amorce dégénérée ?

Le problème majeur qui se pose lorsque l'on veut, à partir d'une séquence protéique, obtenir la séquence nucléotidique correspondante, est **la dégénérescence du code génétique**.
Un acide aminé peut-être codé par plusieurs codons, mais il n'existe qu'une seule séquence sur l'ADN génomique qui code pour cet acide aminé.
Dans ces conditions, concevoir une amorce susceptible de s'hybrider spécifiquement avec cette séquence génomique unique, consiste à prendre en compte l'ensemble des amorces putatives complémentaires de celle-ci.
Exemple :

soit la séquence protéique :	P	Y	L	W
et les codons correspondants :	**CC**T	**TC**T	T**T**A	**TGG**
	CCC	**TC**C	T**T**G	
	CCA		C**T**A	
	CCG		C**T**T	
			C**T**C	
			C**T**G	

Pour une telle séquence protéique il y a **48 séquences nucléotidiques putatives** (4x2x6x1) qui codent pour celle-ci, et dans l'absolu, seulement une séquence peut servir d'amorce pour s'hybrider avec la séquence génomique recherchée.
L'amorce dégénérée s'écrira **symboliquement** : **CCN TCY YTN TGG**
où conventionnellement Y = T ou C et N = A,T,C ou G
Une telle amorce n'a pas de réalité physique, mais le savoir faire des chimistes permet de synthétiser en proportion équimolaire l'ensemble des 48 séquences en solution.
Pendant la PCR, l'amorce la plus fidèle, ou les amorces autorisant une hybridation malgré quelques mésappariements permet en théorie et en pratique aussi d'amplifier le bon fragment.

B. Extraction ADN génomique

1- laver 2 fois le tissu dans 0.5 ml de tampon Protéinase K (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8, 100 mM NaCl) vortex 30s, centrifugation (12000 g), 3-5 min en éliminant à chaque fois le surnageant.

2- mettre le tissu dans 0.5 ml de PK + 10 µl de protéinase K (10 mg/ml) et laisser incuber toute la nuit (ou au moins 5 heures) à 55°C.

3- Des que le tissu est digère, ajouter 10 µl de RNAseA (10 mg/ml) dans la mixture et incuber a 37°C pendant 1-2 h (NB : la RNAse doit etre au préalable mise pendant 10 minutes dans l'eau bouillante pour supprimer toute trace de DNAse).

4- Ensuite effectuer 1 extraction phénol (pH 8), puis 2 extractions phénol/Chloroforme et enfin 1 extraction chloroforme.

5- ajouter 1 ml d'alcool absolu glacé au surnageant récupéré pour précipiter l'ADN (mélange 1:2) et mettre a -20°C pendant 2 à 3 heures.

6- Centrifuger pendant 15 minutes et récupérer le culot d'ADN pour lavage à l'alcool 70, puis lavage alcool absolu et séchage sous vide pendant 15 minutes.

7- L'ADN est ensuite re-dissous dans un tampon TE (10 mM Tris-HCl, 1 mM EDTA, pH 7,4).

C. Extraction des ARN Totaux avec le kit RNAble ® (Eurobio).

Le principe de ce kit repose sur l'utilisation d'une solution à base de phénol et d'isothiocyanate de guanidine qui permet la formation de complexes entre l'ARN et ce dernier. L'ARN reste dans la phase aqueuse et le reste des protéines et de l'ADN dans la phase organique.
L'extraction classique continue avec l'ajout de 100 µl de chloroforme pour 1 ml de tampon d'extraction. Puis l'ARN est précipité à l'éthanol. Les ARN totaux sont repris dans de l'eau ou dans un tampon TE 10 mM Tris-HCl, 1 mM EDTA, pH 7,4**.** CONSERVATION à - 80 °C

D. Purification des ARN messagers (mRNA purification kit ® (Pharmacia Biotech))

Le pool des ARN totaux est passé sur une colonne de chromatographie d'affinité, à travers de la cellulose ayant des extensions oligo$(dT)_{12}$. En théorie, seuls les ARNm polyadénilés s'accrochent et sont secondairement élués, après les ARN de transfert et ribosomaux.

Après élution, si les ARNm ne sont pas utilisés immédiatement, ils doivent être conservés congelés à -80°C, de préférence sous forme précipitée à l'éthanol.

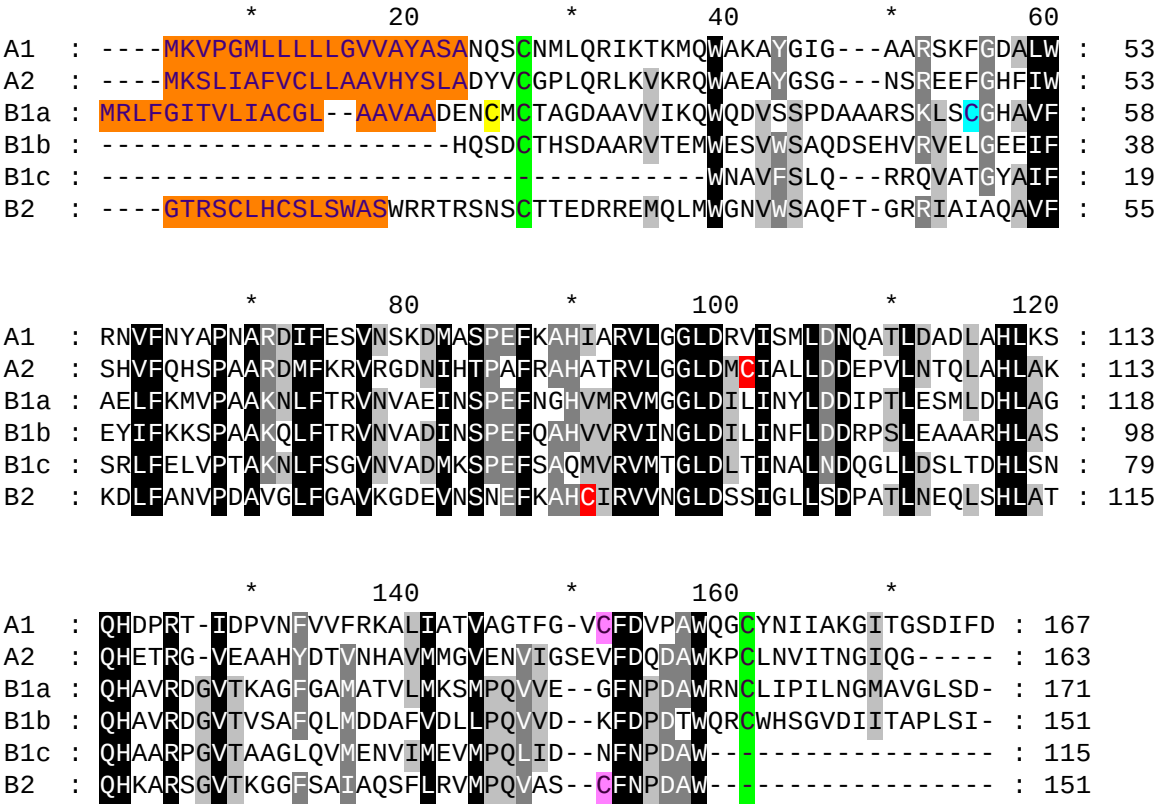
E. Reverse transcription

Réalisée à partir du mélange suivant : 2µg d'ARNm, 5 mM de $MgCl_2$, 2.5 mM de dNTP, 10 U d'inhibiteur d'ARNase , 20 U d'AMV Reverse Transcriptase, 50 pmoles d'oligo dT dans un volume final de 40 µl. Le mélange réactionnel sans l'AMV-RT est d'abord chauffé à 80 °C pendant 5 minutes pour détruire les structures secondaires des ARNm, puis incubé à 42 °C avec l'AMV-RT pendant 30 minutes. Un choc thermique stoppe la réaction en détruisant l'AMV-RT. L'ADNc simple brin est conservé à -20°C jusqu' à son utilisation.

F. Banque ADNc de *Riftia pachyptila*

Réalisée à partir du kit cDNA Synthesis Kit, ZAP-cDNA® Synthesis Kit, and ZAP-cDNA® Gigapack® III Gold Cloning Kit (Stratagene). L'intégralité du protocole est décrit avec précision dans le manuel d'instruction fourni avec le kit, aussi, seules les principales étapes seront décrites. A partir du pool total d'ADNc double brin seule la fraction comprise entre 500 et 2000 pb est sélectionnée pour être clonée dans un vecteur phagemide. L'originalité de ce kit est de pouvoir orienter le clonage de tous les ADNc dans le vecteur et de privilégier en fin de criblage, le séquençage d'une extrémité ou d'une autre de l'ADNc. En effet, après excision du plasmide d'intérêt, le choix des amorces universelles associées au plasmide permet de séquencer à souhait de la partie 3'UTR ou 5'UTR. Cette banque avait un titre de 7. 10^6 pfu/ml au moment des premiers criblages.

Annexe 3 : Séquences nucléotidiques complètes ou partielles des clones contenant les globines, obtenues soit par amplification par PCR à partir d' ADNc soit par criblage de la banque.



Globine A1 **(En bleu : séquence codant le polypeptide mature, en orange : séquence d'adressage, en vert : partie 3' UTR)**

ATGAAGGTTCCCGGCATGCTGCTCCTGCTGCTGGGCGTTGTGGCCTATGCCAGCGCCAACCA
AAGTTGCAACATGCTGCAGCGGATTAAGACGAAGATGCAGTGGGCCAAGGCGTACGGTATCG
GTGCTGCCCGCTCCAAATTCGGCGATGCGCTCTGGAGAAACGTCTTCAACTACGCCCCGAAC
GCTCGTGACATCTTCGAAAGTGTCAACTCCAAAGACATGGCATCGCCCGAGTTCAAGGCTCA
CATTGCCCGAGTGTTGGGAGGTCTTGACCGTGTCATCTCCATGCTCGACAACCAGGCGACGC
TCGATGCTGATCTCGCCCACCTGAAGTCTCAGCACGACCCACGAACCATCGACCCCGTTAAC
TTCGTTGTGTTCCGTAAGGCACTCATCGCGACGGTCGCTGGAACGTTCGGCGTCTGCTTCGA
CGTGCCAGCGTGGCAAGGTTGCTACAACATCATCGCCAAGGGCATCACCGGCTCAGACATAT
TCGACTAAAATGTNTCATTCGCAAGCAGTGCCTTGGCCCCCAATCTCTGTAAACTACACTGA
CAATAGTCTATAAACTCCATTACCAGTAACATTAGTCGCCTCTGTTGGCCTTCAGCGGTAGT
TCTGTCGAGCGCGGTTTGGTTTTCCNCTACTTTTGTCCCATGAGCAATGACAAGAAGAACAA
CAACAAGGGTGGTTGAANAANAAAAACCGACTTAGCNCCTCGCCGTCTATAGGCAAATGGCG
GCCGACGAAATCTNTGAGATGTTGTGTGCGCGGCTGCAACTGTTGTTATAGTGTGTTCGCTA
CCGGCTATCTAACTCTCGTGTCGGAGGACCTTCGTGACAAATAAACCTGTTGGTACAAAAAA
AAAAAAAAAAAAAAAAAAAAAA

Globine A2

ATGAAATCCCTGATCGCCTTCGTGTGCCTTTTGGCCGCTGTACATTACTCCCTTGCCGACTA
CGTGTGCGGACCCCTTCAGCGCCTGAAGGTCAAGCGCCAGTGGGCTGAGGCTTATGGAAGCG
GAAACAGCAGGGAGGAATTCGGACACTTCATCTGGTCCCATGTCTTCCAGCACTCGCCTGCT
GCCCGCGACATGTTCAAGCGCGTCCGCGGTGACAACATCCACACCCCAGCATTCAGGGCCCA
CGCCACCCGTGTGCTCGGTGGACTCGACATGTGCATTGCCCTTCTCGATGATGAACCCGTTC
TGAACACGCAGCTCGCTCATCTTGCCAAGCAGCACGAGACCCGTGGTGTTGAGGCCGCTCAC
TACGACACCGTTAACCACGCCGTCATGATGGGTGTTGAGAACGTCATCGGCTCGGAAGTATT
CGACCAGGATGCGTGGAAGCCTTGCCTCAATGTCATCACCAACGGCATCCAGGGGCTAGNAA
TNTTCTCTCTANCCNCCTCGCTACGTAGCNCAACCGAACTATGGCNCTCGTNGACTTNGTCG
AGATGTCTCCATAAAAATAAACNCNCGAGGANAGTCGTGGCGAGCGGACGAACTTATCTAGC
GGGACACGACTTGTATTCCNCCTCAGGGCTGTTTTAAATCAAAGGGAGGCTGCCGATTGCAA
TTTGGTTTTCCNCTGTTTTTTGTCCCATGAGCGCTCGTCCCNCAGTGGCGACACCCATCGTT
CGTTCAACNCCAAACNAAAAAAAANAAATTCAAATGGCAACCGACTACCATCTCTAACNCATT
ATCAGGTAGTTCCTGTTGGATCGAGTTCGAATGCTTGCATGCTATGTTACNCCCAGTAGCCG
ACTACTGAAGCGGTTGTAAANAATCGAATAAATGTGTGAAAAATGCAAAAAAAAAAAAAAAA

Globine B1a

ATGCGTCTATTCGGCATCACCGTATTGATCGCCTGCGGATTGGCGGCTGTCGCAGCCGACGA
GAACTGCATGTGCACCGCCGGCGACGCTGCCGTTGTCATCAAACAGTGGCAGGACGTGAGCA
GTCCGGACGCCGCAGCAAGGAGCAAGCTCTCTTGTGGACACGCCGTTTTCGCCGAGCTCTTC
AAAATGGTTCCCGCTGCCAAGAATCTGTTCACCCGCGTCAACGTCGCTGAGATCAACAGTCC
CGAATTCAACGGCCACGTGATGCGCGTGATGGGCGGTCTCGATATACTCATCAACTACCTCG
ATGACATTCCTACCCTTGAGAGTATGCTAGATCATCTCGCCGGGCAGCACGCAGTACGTGAC
GGCGTTACCAAAGCAGGCTTCGGAGCGATGGCGACGGTGTTGATGAAGAGTATGCCGCAGGT
GGTTGAGGGATTCAACCCCGACGCCTGGCGCAACTGTCTCATTCCTATCCTCAACGGCATGG
CTGTTGGTCTGTCCGACTAG

Globine B1b

CACCAGAGCGATTGCACCCACAGCGACGCCGCCCGTGTCACAGAGATGTGGGAGAGCGTGTG
GAGTGCGCAGGACTCGGAACACGTGAGAGTGGAGCTAGGAGAGGAGATCTTCGAATACATCT
TCAAGAAGAGTCCGGCCGCGAAACAATTGTTTACCCGAGTCAACGTTGCTGACATCAACAGC
CCAGAGTTCCAGGCTCATGTCGTGCGAGTCATCAACGGTCTTGATATCCTCATCAACTTCCT
CGATGACCGTCCGTCGCTGGAAGCCGCTGCCAGACATCTCGCCAGCCAACATGCCGTTCGAG
ATGGTGTCACAGTTAGCGCTTTCCAGCTGATGGACGACGCCTTCGTCGATTTGCTTCCTCAG
GTCGTCGACAAGTTCGATCCTGACACGTGGCAGCGATGCTGGCACTCGGGCGTCGACATCAT
CACCGCCCCTTTGTCCATCTAA

Globine B1c

TGGAACGCCGTCTTCAGCCTCCAACGGCGGCAAGTTGCCACCGGATACGCCATCTTCTCCAG
ATTGTTCGAGCTCGTGCCCACTGCCAAGAACCTCTTCAGCGGCGTCAACGTTGCTGACATGA
AGAGCCCCGAATTCAGCGCCCAGATGGTTCGTGTGATGACCGGCCTTGACCTCACCATCAAC
GCCCTCAACGACCAGGGTCTGCTCGACTCTCTGACCGACCATCTGTCTAACCAGCATGCCGC
ACGTCCCGGCGTCACGGCCGCTGGCTTGCAGGTCATGGAGAACGTCATCATGGAGGTGATGC
CTCAGCTTATCGACAACTTCAACCCCGACGCCTGG

Globine B2

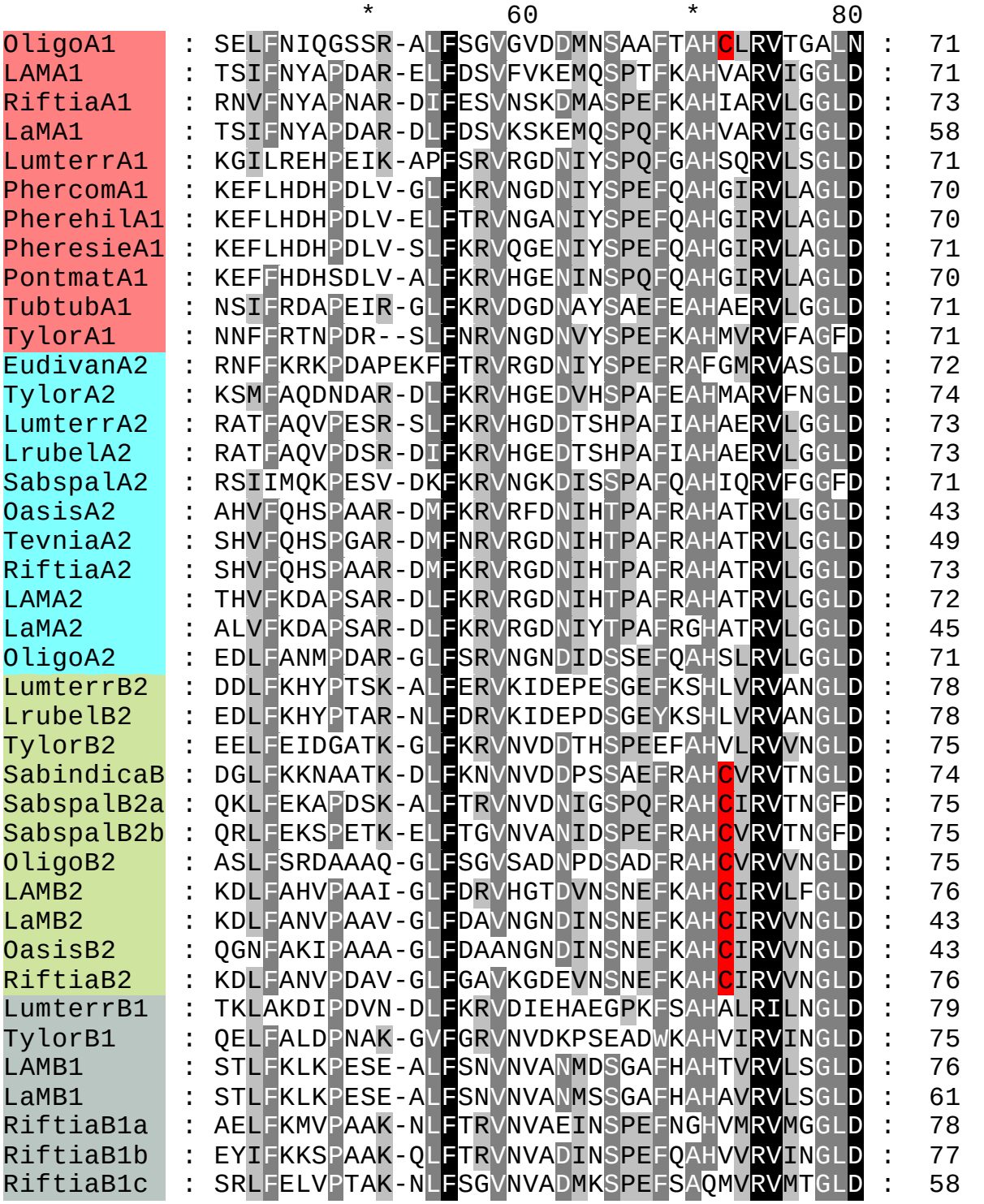
GGCACCAGATCATGCTTGCACTGCTCCTTGTCATGGGCCTCGTGGCGGCGGACTCGCTCCAA
CAGCTGTACGACGGAGGATCGTAGGGAGATGCAGCTCATGTGGGGCAACGTGTGGAGCGCCC
AGTTCACCGGACGTCGTATCGCCATTGCCCAGGCTGTCTTCAAGGACCTCTTCGCCAACGTA
CCAGATGCCGTTGGTCTCTTCGGTGCAGTCAAAGGTGACGAAGTCAACAGCAACGAGTTCAA
GGCGCACTGTATCCGTGTCGTGAACGGTCTCGACTCCTCCATCGGTCTTCTGTCCGACCCTG
CCACCCTCAACGAGCAGCTCAGTCACTTGGCCACTCAGCACAAGGCAAGGTCCGGCGTGACG
AAGGGAGGCTTCTCTGCCATCGCCCAGTCATTCCTGCGAGTCATGCCCCAGGTCGCCAGCTG
CTTCAACCCCGACGCATGG

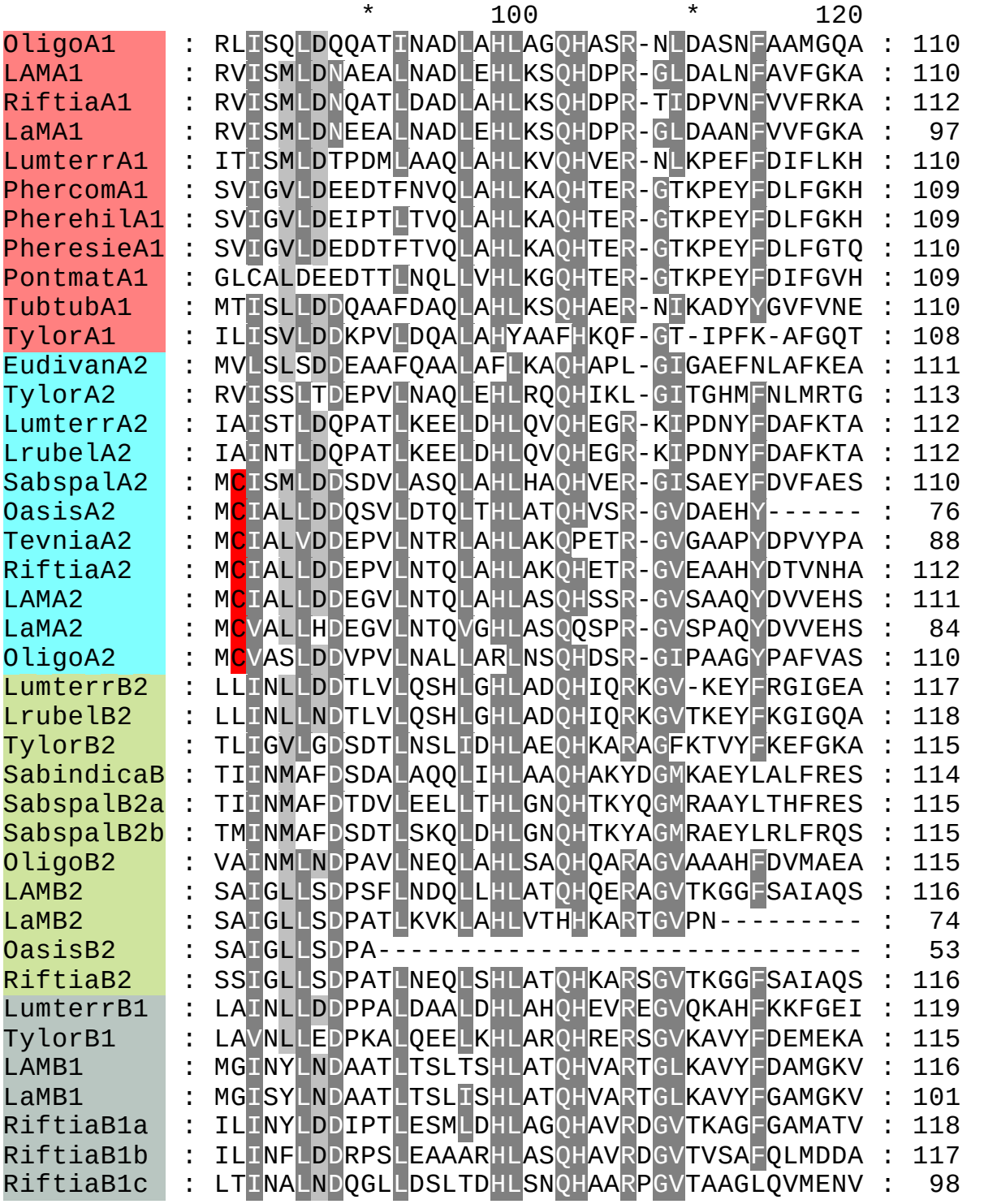
Annexe 4 : Alignements multiples

1. Alignement global des globines extracellulaires d'Annélides

Les noms complets (et les numéros d'accession) des espèces se trouvent à la fin de l'alignement.

```
                        *         20         *         40
OligoA1     :  -----VCNRLEQILVKTQWAQSYGE---AENRAAFSRDLF  :  32
LAMA1       :  -----DCNILQRLKVKMQWAKAYGF---GAERAKFGNSLW  :  32
RiftiaA1    :  ---NQSCNMLQRIKTKMQWAKAYGI---GAARSKFGDALW  :  34
LaMA1       :  ------------------WAKAYGF---GAERAKFGNSLW  :  19
LumterrA1   :  -----ECLVTEGLKVKLQWASAFGH---AHQRVAFGLELW  :  32
PhercomA1   :  -----DCDVLERFKVKHQWQAVFS----EAHRTDFSLHFW  :  31
PherehilA1  :  -----ECDVLERFKVKHQWQTVFS----EAHRTEFSLHFW  :  31
PheresieA1  :  -----DCNTLKRFKVKHQWQQVFSG---EHHRTEFSLHFW  :  32
PontmatA1   :  -----DCDVLARFRVKHQWQEVFL----GSNRMEFSHDLW  :  31
TubtubA1    :  -----ECDALQRFKVKHQWAEAFGT---SHHRLDFGLKLW  :  32
TylorA1     :  ----TDCGILQRIKVKQQWAQVYSV---GESRTDFAIDVF  :  33
EudivanA2   :  -----GCNEIKRLKVKLQWAQSFGF---ENTREDFEDELF  :  32
TylorA2     :  --SSDHCGPLQRLKVKQQWAKAYGV---GHERVELGIALW  :  35
LumterrA2   :  ---KKQCGVLEGLKVKSEWGRAYGS---GHDREAFSQAIW  :  34
LrubelA2    :  ---KKQCGVLEGLKVKSEWGRAYSS---GHDREAFAQAIW  :  34
SabspalA2   :  -----GCNILQRLKVKNQWQEAFGY---ADDRTSLGTALW  :  32
OasisA2     :  ------------------------------------HFMW  :   4
TevniaA2    :  ------------------------SREDFGHFIW  :  10
RiftiaA2    :  ---DYVCGPLQRLKVKRQWAEAYGS---GNSREEFGHFIW  :  34
LAMA2       :  ----YECGPLQRLKVKRQWAEAYGS---GNDREEFGHFIW  :  33
LaMA2       :  ------------------------------FGHFMW  :   6
OligoA2     :  -----DCTSLNRLLVKRQWAEAYGE---GTNRELLGNRIW  :  32
LumterrB2   :  ADDEDCCSYEDRREIRHIWDDVWSSSFTDR-RVAIVRAVF  :  39
LrubelB2    :  AAEEDCCSYEDRREIRHIWDDVWSSSFTDR-RVAIVRAVF  :  39
TylorB2     :  ---DTCCSIEDRREVQALWRSIWSAEDTGR-RTLIGRLLF  :  36
SabindicaB  :  ----SCCSMEDRQEVLKAWETMWSAEFTGR-RVTIAQEVF  :  35
SabspalB2a  :  ---EGCCSMEDRQEVLNAWEALWSAEYTGR-RVMIAQAAF  :  36
SabspalB2b  :  ---EACCSMEDRQEVLKAWEAMWSAEFTGR-RVIIAQEVF  :  36
OligoB2     :  ---SSCCSNEDRANVMHNWDAAWSAAYSDR-RVALAQAVF  :  36
LAMB2       :  --SSNSCTTEDRREMQLMWANVWSAQFTGR-RLAIAQAVF  :  37
LaMB2       :  ------------------------------QAVF  :   4
OasisB2     :  ------------------------------QAVF  :   4
RiftiaB2    :  --RSNSCTTEDRREMQLMWGNVWSAQFTGR-RIAIAQAVF  :  37
LumterrB1   :  DEHEHCCSEEDHRIVQKQWDILWRDTESSKIKIGFGRLLL  :  40
TylorB1     :  ---DDCCSAADRHEVLDNWKGIWSAEFTGR-RVAIGQAIF  :  36
LAMB1       :  ---SKFCSEGDATIVIKQWNQIYNAGISAGSRLTMSNKIF  :  37
LaMB1       :  ------------------WNQIYNAGISAGSRLTMGNKIF  :  22
RiftiaB1a   :  -DENCMCTAGDAAVVIKQWQDVSSPDAAARSKLSCGHAVF  :  39
RiftiaB1b   :  --HQSDCTHSDAARVTEMWESVWSAQDSEHVRVELGEEIF  :  38
RiftiaB1c   :  ------------------WNAVFS---LQRRQVATGYAIF  :  19
```


* 60 * 80
OligoA1 : SELFNIQGSSR-ALFSGVGVDDMNSAAFTAHCLRVTGALN : 71
LAMA1 : TSIFNYAPDAR-ELFDSVFVKEMQSPTFKAHVARVIGGLD : 71
RiftiaA1 : RNVFNYAPNAR-DIFESVNSKDMASPEFKAHIARVLGGLD : 73
LaMA1 : TSIFNYAPDAR-DLFDSVKSKEMQSPQFKAHVARVIGGLD : 58
LumterrA1 : KGILREHPEIK-APFSRVRGDNIYSPQFGAHSQRVLSGLD : 71
PhercomA1 : KEFLHDHPDLV-GLFKRVNGDNIYSPEFQAHGIRVLAGLD : 70
PherehilA1 : KEFLHDHPDLV-ELFTRVNGANIYSPEFQAHGIRVLAGLD : 70
PheresieA1 : KEFLHDHPDLV-SLFKRVQGENIYSPEFQAHGIRVLAGLD : 71
PontmatA1 : KEFFHDHSDLV-ALFKRVHGENINSPQFQAHGIRVLAGLD : 70
TubtubA1 : NSIFRDAPEIR-GLFKRVDGDNAYSAEFEAHAERVLGGLD : 71
TylorA1 : NNFFRTNPDR--SLFNRVNGDNVYSPEFKAHMVRVFAGFD : 71
EudivanA2 : RNFFKRKPDAPEKFFTRVRGDNIYSPEFRAFGMRVASGLD : 72
TylorA2 : KSMFAQDNDAR-DLFKRVHGEDVHSPAFEAHMARVFNGLD : 74
LumterrA2 : RATFAQVPESR-SLFKRVHGDDTSHPAFIAHAERVLGGLD : 73
LrubelA2 : RATFAQVPDSR-DIFKRVHGEDTSHPAFIAHAERVLGGLD : 73
SabspalA2 : RSIIMQKPESV-DKFKRVNGKDISSPAFQAHIQRVFGGFD : 71
OasisA2 : AHVFQHSPAAR-DMFKRVRFDNIHTPAFRAHATRVLGGLD : 43
TevniaA2 : SHVFQHSPGAR-DMFNRVRGDNIHTPAFRAHATRVLGGLD : 49
RiftiaA2 : SHVFQHSPAAR-DMFKRVRGDNIHTPAFRAHATRVLGGLD : 73
LAMA2 : THVFKDAPSAR-DLFKRVRGDNIHTPAFRAHATRVLGGLD : 72
LaMA2 : ALVFKDAPSAR-DLFKRVRGDNIYTPAFRGHATRVLGGLD : 45
OligoA2 : EDLFANMPDAR-GLFSRVNGNDIDSSEFQAHSLRVLGGLD : 71
LumterrB2 : DDLFKHYPTSK-ALFERVKIDEPESGEFKSHLVRVANGLD : 78
LrubelB2 : EDLFKHYPTAR-NLFDRVKIDEPDSGEYKSHLVRVANGLD : 78
TylorB2 : EELFEIDGATK-GLFKRVNVDDTHSPEEFAHVLRVVNGLD : 75
SabindicaB : DGLFKKNAATK-DLFKNVNVDDPSSAEFRAHCVRVTNGLD : 74
SabspalB2a : QKLFEKAPDSK-ALFTRVNVDNIGSPQFRAHCIRVTNGFD : 75
SabspalB2b : QRLFEKSPETK-ELFTGVNVANIDSPEFRAHCVRVTNGFD : 75
OligoB2 : ASLFSRDAAAQ-GLFSGVSADNPDSADFRAHCVRVVNGLD : 75
LAMB2 : KDLFAHVPAAI-GLFDRVHGTDVNSNEFKAHCIRVLFGLD : 76
LaMB2 : KDLFANVPAAV-GLFDAVNGNDINSNEFKAHCIRVVNGLD : 43
OasisB2 : QGNFAKIPAAA-GLFDAANGNDINSNEFKAHCIRVVNGLD : 43
RiftiaB2 : KDLFANVPDAV-GLFGAVKGDEVNSNEFKAHCIRVVNGLD : 76
LumterrB1 : TKLAKDIPDVN-DLFKRVDIEHAEGPKFSAHALRILNGLD : 79
TylorB1 : QELFALDPNAK-GVFGRVNVDKPSEADWKAHVIRVINGLD : 75
LAMB1 : STLFKLKPESE-ALFSNVNVANMDSGAFHAHTVRVLSGLD : 76
LaMB1 : STLFKLKPESE-ALFSNVNVANMSSGAFHAHAVRVLSGLD : 61
RiftiaB1a : AELFKMVPAAK-NLFTRVNVAEINSPEFNGHVMRVMGGLD : 78
RiftiaB1b : EYIFKKSPAAK-QLFTRVNVADINSPEFQAHVVRVINGLD : 77
RiftiaB1c : SRLFELVPTAK-NLFSGVNVADMKSPEFSAQMVRVMTGLD : 58

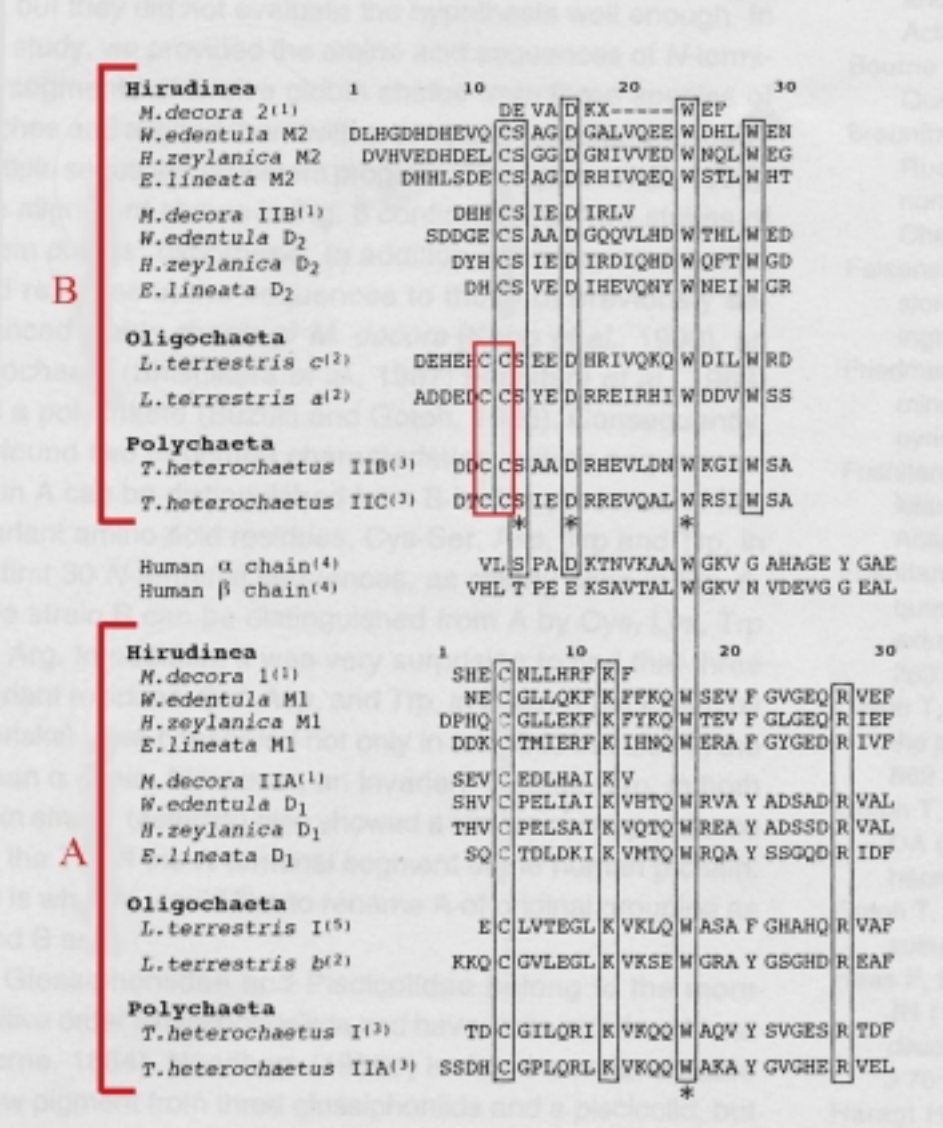

* 100 * 120
OligoA1 : RLISQLDQQATINADLAHLAGQHASR-NLDASNFAAMGQA : 110
LAMA1 : RVISMLDNAEALNADLEHLKSQHDPR-GLDALNFAVFGKA : 110
RiftiaA1 : RVISMLDNQATLDADLAHLKSQHDPR-TIDPVNFVVFRKA : 112
LaMA1 : RVISMLDNEEALNADLEHLKSQHDPR-GLDAANFVVFGKA : 97
LumterrA1 : ITISMLDTPDMLAAQLAHLKVQHVER-NLKPEFFDIFLKH : 110
PhercomA1 : SVIGVLDEEDTFNVQLAHLKAQHTER-GTKPEYFDLFGKH : 109
PherehilA1 : SVIGVLDEIPTLTVQLAHLKAQHTER-GTKPEYFDLFGKH : 109
PheresieA1 : SVIGVLDEDDTFTVQLAHLKAQHTER-GTKPEYFDLFGTQ : 110
PontmatA1 : GLCALDEEDTTLNQLLVHLKGQHTER-GTKPEYFDIFGVH : 109
TubtubA1 : MTISLLDDQAAFDAQLAHLKSQHAER-NIKADYYGVFVNE : 110
TylorA1 : ILISVLDDKPVLDQALAHYAAFHKQF-GT-IPFK-AFGQT : 108
EudivanA2 : MVLSLSDDEAAFQAALAFLKAQHAPL-GIGAEFNLAFKEA : 111
TylorA2 : RVISSLTDEPVLNAQLEHLRQQHIKL-GITGHMFNLMRTG : 113
LumterrA2 : IAISTLDQPATLKEELDHLQVQHEGR-KIPDNYFDAFKTA : 112
LrubelA2 : IAINTLDQPATLKEELDHLQVQHEGR-KIPDNYFDAFKTA : 112
SabspalA2 : MCISMLDDSDVLASQLAHLHAQHVER-GISAEYFDVFAES : 110
OasisA2 : MCIALLDDQSVLDTQLTHLATQHVSR-GVDAEHY------ : 76
TevniaA2 : MCIALVDDEPVLNTRLAHLAKQPETR-GVGAAPYDPVYPA : 88
RiftiaA2 : MCIALLDDEPVLNTQLAHLAKQHETR-GVEAAHYDTVNHA : 112
LAMA2 : MCIALLDDEGVLNTQLAHLASQHSSR-GVSAAQYDVVEHS : 111
LaMA2 : MCVALLHDEGVLNTQVGHLASQQSPR-GVSPAQYDVVEHS : 84
OligoA2 : MCVASLDDVPVLNALLARLNSQHDSR-GIPAAGYPAFVAS : 110
LumterrB2 : LLINLLDDTLVLQSHLGHLADQHIQRKGV-KEYFRGIGEA : 117
LrubelB2 : LLINLLNDTLVLQSHLGHLADQHIQRKGVTKEYFKGIGQA : 118
TylorB2 : TLIGVLGDSDTLNSLIDHLAEQHKARAGFKTVYFKEFGKA : 115
SabindicaB : TIINMAFDSDALAQQLIHLAAQHAKYDGMKAEYLALFRES : 114
SabspalB2a : TIINMAFDTDVLEELLTHLGNQHTKYQGMRAAYLTHFRES : 115
SabspalB2b : TMINMAFDSDTLSKQLDHLGNQHTKYAGMRAEYLRLFRQS : 115
OligoB2 : VAINMLNDPAVLNEQLAHLSAQHQARAGVAAAHFDVMAEA : 115
LAMB2 : SAIGLLSDPSFLNDQLLHLATQHQERAGVTKGGFSAIAQS : 116
LaMB2 : SAIGLLSDPATLKVKLAHLVTHHKARTGVPN--------- : 74
OasisB2 : SAIGLLSDPA---------- : 53
RiftiaB2 : SSIGLLSDPATLNEQLSHLATQHKARSGVTKGGFSAIAQS : 116
LumterrB1 : LAINLLDDPPALDAALDHLAHQHEVREGVQKAHFKKFGEI : 119
TylorB1 : LAVNLLEDPKALQEELKHLARQHRERSGVKAVYFDEMEKA : 115
LAMB1 : MGINYLNDAATLTSLTSHLATQHVARTGLKAVYFDAMGKV : 116
LaMB1 : MGISYLNDAATLTSLISHLATQHVARTGLKAVYFGAMGKV : 101
RiftiaB1a : ILINYLDDIPTLESMLDHLAGQHAVRDGVTKAGFGAMATV : 118
RiftiaB1b : ILINFLDDRPSLEAAARHLASQHAVRDGVTVSAFQLMDDA : 117
RiftiaB1c : LTINALNDQGLLDSLTDHLSNQHAARPGVTAAGLQVMENV : 98

```
                      *       140         *
OligoA1     :  VMSVVP-THL-DCFNQHAWGECYERIASGISG-----  : 140
LAMA1       :  LFATVGGKFG-VCFDLPAWESCYKVIAKGITGNDMFN  : 146
RiftiaA1    :  LIATVAGTFG-VCFDVPAWQGCYNIIAKGITGSDIFD  : 148
LaMA1       :  LFATVGGKFG-V-------------------------  : 108
LumterrA1   :  LLHVLGDRLG-THFDFGAWHDCVDQIIDGIKDI----  : 142
PhercomA1   :  LADHLGDKLG-THFDFGAFRDCYAVIAAGIKP-----  : 140
PherehilA1  :  LASHLGDELG-THFDYAAFRDCYDFIASGIKP-----  : 140
PheresieA1  :  LFDILGDKLG-THFDQAAWRDCYAVIAAGIKP-----  : 141
PontmatA1   :  LLRILDDHLGKTYFARQEWQDCYAVIAAGIKP-----  : 141
TubtubA1    :  LLAVLPDYLG-TKLDFKAWSECLGVITGAIHD-----  : 141
TylorA1     :  MFQTIAEHIH-G-ADIGAWRACYEQIVTGITA-----  : 138
EudivanA2   :  VLDTVAAHVG-RCFDKDAWNACMEIIMTGIQSPAR--  : 145
TylorA2     :  LAYVLPAQLG-RCFDKEAWAACWEVIYPGIK-HD---  : 145
LumterrA2   :  ILHVVAAQLG-RCYDREAWDACIDHIEDGIKGHH---  : 145
LrubelA2    :  ILHVVAAQLG-RCYDREAWDACIDYIEDGIKGHH---  : 145
SabspalA2   :  LMLAVESTIES-CFDKDAWSQCTKVISSGIGSGV---  : 143
OasisA2     :  ----------------------------------------  :   -
TevniaA2    :  ----------------------------------------  :   -
RiftiaA2    :  VMMGVENVIGSEVFDQDAWKPCLNVITNGIQG-----  : 144
LAMA2       :  VMMGVEHEIGQNVFDKDAWQACLDVITGGIQGN----  : 144
LaMA2       :  VMMGVEN-------------------------------  :  91
OligoA2     :  AISAVRATVGARSFDNDAWNSCMNQIVSGISG-----  : 142
LumterrB2   :  FARVLPQVLS--CFNVDAWNRCFHRLVARIAKDLP--  : 150
LrubelB2    :  FARVLPQVLS--CFNVDAWNRCFNRLTARIAKDLP--  : 151
TylorB2     :  LNHVLPEVAS--CFNPEAWNHCFDGLVDVISHRIDG-  : 149
SabindicaB  :  FAEVLPQAVP--CFNSAAWNRCIAAMHEEIGSQLAA-  : 148
SabspalB2a  :  FAEILPQAIP--CFNTAAWNRCITAMQDKIGASLAA-  : 149
SabspalB2b  :  FAEVLPQAIP--CFNTAAWNRCITAMQDVMGASLAA-  : 149
OligoB2     :  FAEVMPQVSS--CFSSDSWNRCFARIANGISAGL---  : 147
LAMB2       :  FLRVMPQVAS--CFNPDAWSRCFNRITNGMTEGLA--  : 149
LaMB2       :  ------------------------------------  :   -
OasisB2     :  ------------------------------------  :   -
RiftiaB2    :  FLRVMPQVAS--CFNPDAW-----------------  : 133
LumterrB1   :  LATGLPQVLD--DYDALAWKSCLKGILTKISSRLNA-  : 153
TylorB1     :  LLKVLPQVSS--HFNSGAWDRCFTRIADVIKAELP--  : 148
LAMB1       :  LMTVLPALID--NFNPDAWRNCLLPLKSAIAEGLP--  : 149
LaMB1       :  LMTVLPALID--NFNPDAW-----------------  : 118
RiftiaB1a   :  LMKSMPQVVE--GFNPDAWRNCLIPILNGMAVGLSD-  : 152
RiftiaB1b   :  FVDLLPQVVD--KFDPDTWQRCWHSGVDIITAPLSI-  : 151
RiftiaB1c   :  IMEVMPQLID--NFNPDAW-----------------  : 115
```

Lumterr : *Lumbricus terrestris* (A29134, A28151, B28151, C28151), Lrubell : *Lumbricus rubellus* (U55073, AJ012205, BF422540, BF422675) Tylor : *Tylorrhynchus heterochaetus* (P02219, P02220, P09966, P13578), LAM : *Lamellibrachia sp.* (S08284 et dans (Suzuki et al. 1993)), Oligo : *Oligobrachia mashikoi* (S72251, S72252, S72253), Sabspal : *Sabella spallanzanii* (CAC37411, CAC37412), Sabindica : *Sabellastarte indica* (BAA09589), Pheresie : *Pheretima sieboldi* (P11740), Pherehil : *Pheretima hilgendorfi* (dans (Shishikura 1996)), Phercom : *Pheretima communissima* (dans (Shishikura and Nakamura 1996)), Pontmat : *Pontodrilus*

matsushimensis (dans (Shishikura and Nakamura 1996)), Tubtub : *Tubifex tubifex*(P18202), Eudivan : *Eudistylia vancouverii* (AF257469), LaM : nouvelle espèce de Lamellibrachia de Méditerranée (cf Chapitre 3, deuxième article). Oasis : Oasisia (cf Chapitre 3, deuxième article), Tevnia : Tevnia (cf Chapitre 3, deuxième article). **cystéines vertes** : pour les ponts dissulfures intra-chaîne; **cystéines rouges** : cystéine de liaison de H_2S; **cystéines rose** et **bleues** : pour les ponts dissulfures inter-chaînes;. 100% d'identité, 80% d'identité, 60% d'identité

2. Alignements multiples de la partie NH_2 terminal de globine d'Achètes (Hirudinea dans l'alignement suivant), d'Oligochètes et Polychètes. Classification en famille A et B sur le critère des doubles cystéines, du motif CCS et des autres résidus qui signent sans ambiguïté l'appartenance de ces séquences dans leur famille respective.
Avec les genres M : *Macrobdella*, W : *Whitmania*, H : *Haemadipsa* et E : *Erpobdella*
D'après (et modifié en rouge) de (Shishikura et al. 1997)



Annexe 5 :

1. Alignement protéique de linkers de HB-LHb. En grisé la région homologue des LDLr (Low Density Lipoprotein receptor)

```
                  *        20         *        40         *
Riftia    : ---------------------------------------------------- :   -
Lamelli   : ---------AAV----QPLSVSDAMGARVDAQAWRVDR-LTKQFQAISDA :  36
Nereis    : -------DDCVC----PGGRQWAGIASKADAQEARINR-LAGRVEAVAEK :  38
Tylor1    : -------DDCVC----PGGREWASVASKADSQEARVNR-LAGRVEALAEN :  38
Lumterr1  : ----------DS----PPAQSHDEIIDKLIERTNKITT-SISHVESLLDD :  35
Tylor2    : GALGDRNGDCACDRPSPRGYWGGGMTGRSAFADPHIAEGRLANQDARLGT :  50
Sabspa1   : -------MRSLFLLLVVGIAFGNGALSRTRRSSSWGSSLTTARVDSQEMR :  43
Lumterr2  : -------MWYVLGLMLVGLAAGASDPYQERRFQYLVKN-QNLHIDYLAKK :  42
Sabspa2   : --------MRALLLMCIGAALVAAAVSDNCPECDKPDPATMXKLDENAAR :  42

                 60         *        80         *       100
Riftia    : ---------------------------------------------------- :   -
Lamelli   : ADTSIGAAKSGGDIARH------MLN-------SHLDDHW---CPSKYHR :  70
Nereis    : LRVGGEKVKHFLGDF-----------RELEFRVDELEGNG---CDARHYQ :  74
Tylor1    : LRAGGDRLSHYKSEF-----------RELEFRVDELEGNG---CEPRHFQ :  74
Lumterr1  : -RLDPKXIRKAG---------------SLRHRVEELEDPS---CDEHEHQ :  66
Tylor2    : LEEEVDKLQHKYDDFIAGKTARRERFAQLKDRVWGLEAHH---CDDDHLS :  97
Sabspa1   : LDQLAKQLDSLEGSLEGGDLGIPASMNSLSLRLDKIEGND---CGKRQFS :  90
Lumterr2  : LHDIEEEYNKLTHDVDK------KTIRQLKARISNLEEHH---CDEHESE :  83
Sabspa2   : LNRLVARAVDVSDTIKRRQSEYEDAEVQAQNMLQRLEKRNGVLCKG-QFV :  91

                  *       120         *       140         *
Riftia    : --------------CDGVNDCKNHYDEDSVRCVVPMV-ANSTWIG--YPA :  33
Lamelli   : C-GNSPQCMSNMAFCDGVNDCKNHFDEDENRCVVPVT-ANSTWIG--YPA : 116
Nereis    : CGGNTPYCIADTLVCDGSPDCPNGSDESEDICHIPIP-AGTVLIGHLNTD : 123
Tylor1    : CGGSAMECISDLLTCDGSPDCANGADEDSDVCHIPPI-AGTLLVGHLNTD : 123
Lumterr1  : CGGDDPQCISKLFVCDGHNDCRNGEDEKD--CTLPTK-ADKFIG---DVV : 110
Tylor2    : CKDVAFTCIGHNLVCDGHKDCLNGHDEDEETCSIAAS-VGSSFEG-QIRQ : 145
Sabspa1   : CGRHSGQCVFELNVCDNVPDCANGMDEAEETCSNPVP-AGRSWNG--VAH : 137
Lumterr2  : CRGDVPECIHDLLFCDGEKDCRDGSDEDPETCSLNITHVGSSYTG--LAT : 131
Sabspa2   : CGGAAQQCADNLVVCDGVNDCDNGADEATDICNIKAK-VGDSWVG--SFV : 138

                 160         *       180         *       200
Riftia    : YDHCTQRRPYEMIISVTSAPSSSVYKVHQPLKVQV--DLFSKNHGVKQSA :  81
Lamelli   : YDHCTQRRPYEMVISITSAPSDIVYKVHQPLKVQV--DLFSKKGGLKQSA : 164
Nereis    : HDFCTRRKPTEMDLIITSVTRSKYFQSRLQVKANL--DLKYTAEGAEQED : 171
Tylor1    : HDFCTKRKPNEFDLFISSVQRSSYFQSRLKVKGNL--QIKYTAEGRDQED : 171
Lumterr1  : FDFCTKRRPEHMTLAFESS----------------------PADGE---- : 134
Tylor2    : LDTCTKRKPSAFRFIITNVDVPKYFPQEPHVKATI--LMTSSKDGHETQS : 193
Sabspa1   : WDSCLVRSNNKMNVLITGNKRYDYFPSRLWVDAVI--IESYTVSGEEQTN : 185
Lumterr2  : WTSCEDLNPDHAIVTITAAHRKSFFPNRVWLRATLSYELDEHDHTVSTT- : 180
Sabspa2   : YHNCSSAKFSNLRFELKKFERKPNMMSLADVSAML-FWDDNTDADVTFID : 187
```

```
                    *       220         *       240         *
Riftia    : SLTGDAYYCKGSQRLLHCPP------------------------------ : 101
Lamelli   : SLHGDAVYCKGSQRLIVAPP--EDDRLEIIGQFDGVSNDRFKGYIVREMS : 212
Nereis    : VLPVEGYYNFGTHQLVILPP--ENDRLGLVCSFRAGNDNRCSAAIVHEAS : 219
Tylor1    : VLQVKGYYNFGTHQLVILPP--EDDRLGIVCNFRAGNDDRCRAHIVHEAS : 219
Lumterr1  : -------YSFADHRLTIHPP--EEDGLGLVGEFDGYNFDRFVGHIVHELS : 175
Tylor2    : SLAVDGVYDFTHRKVILYSP--DKDNLIFECTFPRHDNNHCKGVMKHSGG : 241
Sabspa1   : TWSISGFYNYASRMVKFNSP--KEGTMGIGLTCSDAGSDVAQCDFVRATD : 233
Lumterr2  : --QLRGFYNFGKRELLLAPLKGQSEGYGVICDFNLGDDDHADCKIVVPSS : 228
Sabspa2   : SLPYDGTYKFYSRSAHLRSIMVVNSSRKCSTSYPER-LTSCITMSTLLTA : 236


                  260
Riftia    : ----------------- :   -
Lamelli   : GDKCAEFRFFKQ----- : 224
Nereis    : LAHCGDDFVFVRKDDH- : 235
Tylor1    : LEHCGDDFVFVKEDDHH : 236
Lumterr1  : EEVC-AEFIFHR----- : 186
Tylor2    : -DVCLTFTLERID---- : 253
Sabspa1   : QHVCGSVTLTRGA---- : 246
Lumterr2  : LFVCAHFNAQRY----- : 240
Sabspa2   : KRLC------------- : 240
```

2. Séquence nucléotidique et séquence protéique (partielles) d'un linker de *Riftia pachyptila*

Linker Riftia

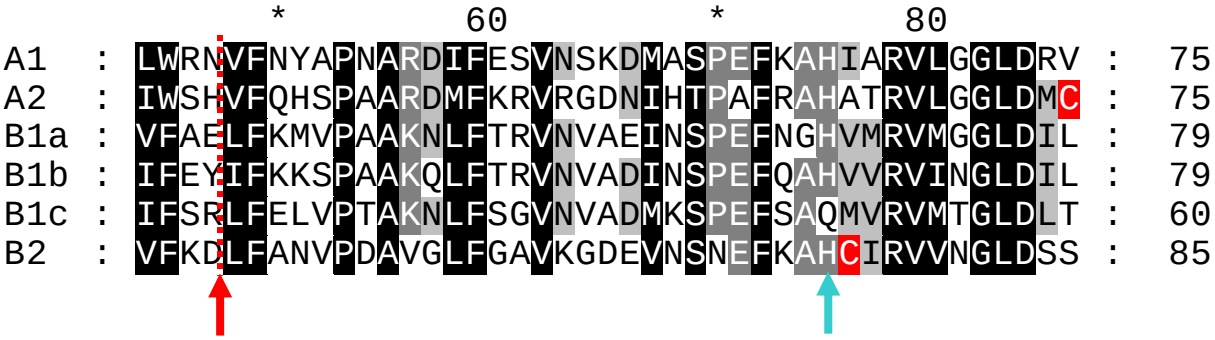
```
TGTGACGGGGTGAATGACTGCAAGAACCACTACGACGAGGACTCAGTTCGATGCGTTGTTCC
CATGGTTGCCAACTCAACATGGATCGGTTACCCCGCCTACGACCACTGCACGCAGCGCCGCC
CGTACGAAATGATCATCAGCGTCACCAGCGCGCCATCCTCCTCTGTCTACAAGGTTCACCAG
CCCCTGAAGGTGCAGGTGGATCTGTTCTCAAAGAACCACGGTGTCAAGCAATCTGCCAGTCT
TACCGGAGATGCCTACTACTGCAAAGGAAGCCAGAGGCTGCTTCATTGCCCCCCC
```

Linker Riftia avec Région LDL

```
CDGVNDCKNHYDEDSVRCVVPMVANSTWIGYPAYDHCTQRRPYEMIISVTSAPSSSVYKVHQ
PLKVQVDLFSKNHGVKQSASLTGDAYYCKGSQRLLHCPP
```

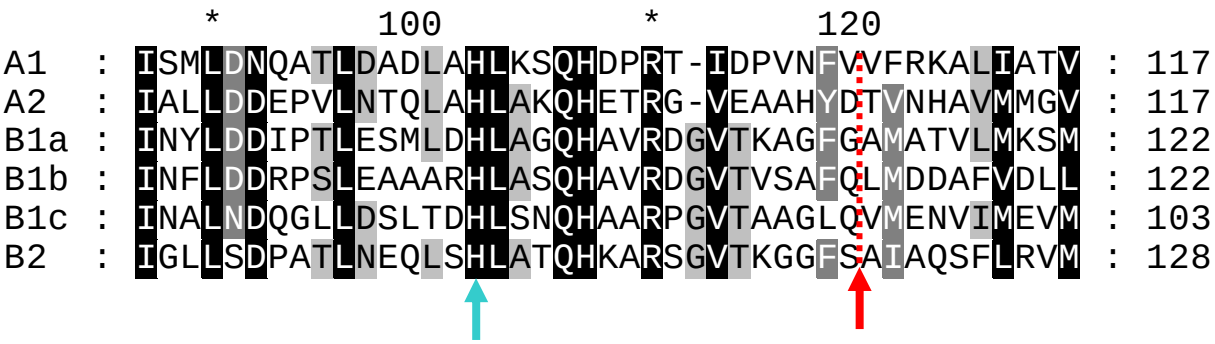
Annexe 6 : Séquences et positions relatives des deux introns au sein de la séquence protéique

Les deux flèches rouges indiquent l'insertion des codons. Les deux flèches bleues indiquent respectivement les histidines proximale et distale qui permettent dans la poche hydrophobe de maintenir l'hème et de fixer l'oxygène sur l'atome de fer



Intron 1 A2

5'GTGAGTATGCTCTTTTCTATATGTCGTCATCTTTGTTTGTTACGTTTTATTATCACGACTGTCGCCTCGTTAT
TGTTATCGATATCGTTGCACTGTGGCGGTGTGTGAAGCTTTAAAGAATTTCGTTTCGTTGACATTTTATGCAGTT
GATAGTGATACATTGTTGCGTTACATGGCGCATTAATTGGGCTTCATCATGCCAGTTTTGGTGTTCATGCGATTG
TTCATTATTATTTTCCATTGCTAAGCAATAGTGGCCCTCCTTGATGGTTTACACTGGCACGTACTTACCACATAC
GGCGGTAAGGTTTCCGATAATTCTCATAGTTTCGCTTAGACATAACTGGTGAACGATAAATGATACGCTAATTAC
AATTAGTATGCCAATATGTATTGCAGCAAATCTACTGTATAATAAGACTTGTATTCGGCTTTCGATTAAATGCGT
ATGTAGTACGACACGATGTTCTTTGAGCACGACGTCTCTCCTTGCAG 3'



Intron 2 A2

5'GTAAGCGTCATTTCTGTTTCACGTTCAATTTCCTTCTCGTCTTCTTCCGCCTCTCGACTTTCTGTTTTGCTTT
TGTTTTCAAATGATCACCATTTTGTTGCGGTTTATTGATGCAATGGCTTTCTGCTCTCTGCCTGTTCCTTCAATA
AGTCGTTTTTCGTAGTTTTTCATAGTTTTGGACAACGTCTACCGCTATGTAAAAGATATTATATACATCAAACAA
TAACTGTTGAAGTTTTGTAGGAATCTAATTTAGAATCATATGTATTTACCGCTGGACGCAATTTGGTTGCTATCG
AATATTACTTGATACAGGCTTAGGGCCATTGTTAAAGATATAATGCACTATACATACATACATACATTCATACAT
ACATACATACATACGCACGCACGCACGCATACGCACATGTACACGCACGCACACACACACACACACACACACACA
CACACACAACACACACACATATGTCAACATGTTGCCAGCTTTCCCCCTTTTTTTGTCGTTGTGAAATCAGTGCTCT
ATTTAATGTACCGCATGAAGCGTGACATGTGTTTTAATACCACCTTATTGGTTCAAATATTTGCATCCATCTGTT
AAATTGCTGTTTAATCACGCCTGTCGTTTGTTTACCCTCAG 3'

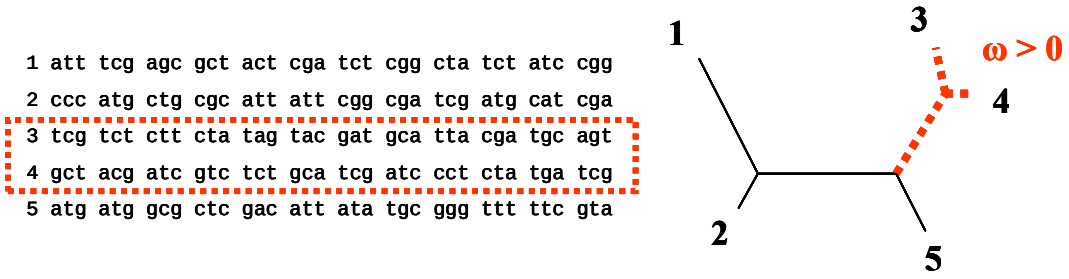
Intron 2 B2

5'GTGAGTACCCGCTTATCATTATGCCGTGTCCTATCATCGAGGCGTAGTTCATACAGTTGCATTCGTAACACTA
GGGTAAATGTGTTGGCGGGCCCGCATAATCCAAAAATTGCTTATACTATCTTGAAAACGGCATGGTCAAATGGCA
AACAAAACATCAGGACTTTCAGCAGGTTTTCTTGGACATTACATAAGTATCTCCAAATTAGGTTGTCGAAATCAT
CAGGCGAGTTATGTGTCTCCAAATTGTGTTCGCAGTCGGTCTTTCGCGTGCGTGTTTTCCTCATTCACGGGTTGT
AGACTTGTGGTACATAATGGGAAACAATTATAGTTTCCCTGACGTATACAAATGTAGGCAATCAACTGACGGAGT
CGTCTCGTCTTTTCCAG 3'

Annexe 7 :

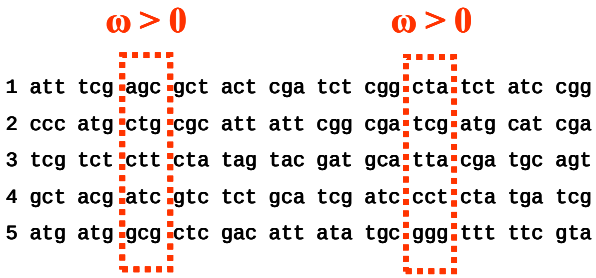
1. les 3 modèles d'estimation des substitutions synonymes et non-synonymes

1. Détection de sélection positive au niveau d'une lignée évolutive



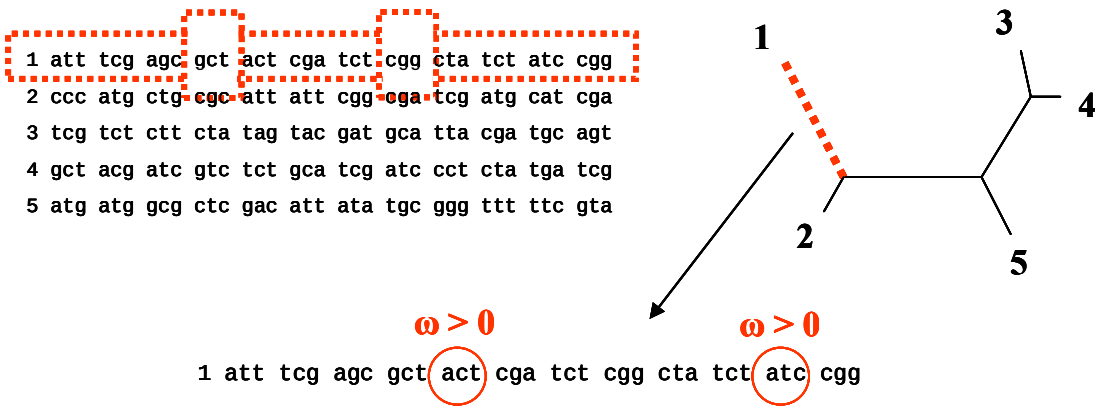
moyenne des dN et dS sur l'ensemble des sites pour chacune des lignées (1)

2. Détection de sites sous sélection positive



moyenne des dN et dS d'un site sur l'ensemble des séquences (2)

3. Détection de sites sous sélection positive dans une lignée évolutive (1+2)



2. Tableaux 2 et 3 du chapitre 4 (article 1)

2. Tableaux 2 et 3 du chapitre 4 (article 1)

Table 2

Log Likelihood Scores and Parameter estimates for the lineage-specific, the site-specific and the branch-site models from globin A2 cluster

Models for A2	L	p	Estimates of parameters	LRT (df)	Positively selected sites	
lineage-specific						
A	-980.73	1	$\omega = 0.208$	B vs. A (1): no significant	None	
B	-980.47	2	$\omega_0 = 0.257$, $\omega_0 = 0.416$		Not allowed	
C	-976.79	9	**Reported in Fig. 7a**	B vs. C (7): no sign	Not allowed	
site-specific						
m1	-970.45	1	$p_0 = 0.290$ $\omega_0 = 0$ $p_1 = 0.709$ $\omega_1 = 1$	m2 vs. m1 (2): **	Not allowed	
m2	-956.75	3	$p_0 = 0.226$ $\omega_0 = 0$ $p_1 = 0.252$ $\omega_1 = 1$ $p_2 = 0.520$ $\omega_2 = 0.209$		None	
m3 (K=2)	-959.34	3	$p_0 = 0.588$ $\omega_0 = 0.411$ $p_1 = 0.089$ $\omega_1 = 0.788$	m3 vs. A (2): **	None	
m7	-959.49	2	$p = 0.445$, $q = 0.729$		Not allowed	
m8	-956.10	4	$p_0 = 0.885$, $p_1 = 0.114$ $p = 0.506$, $q = 1.457$ $\omega = 2.53$	m8 vs. m7 (2):**	49S 50D 65E 61A <u>E10</u> *69S 64V*	
branch-site					Foreground species	Background species
Bm2	-967.78	3	$p_0 = 0.286$ $\omega_0 = 0$ $p_1 = 0.512$ $\omega_1 = 1$ $p_2 + p_3 = 0.201$ $\omega_2 = 2.405$	Bm2 vs. m1:**	51V 67G 41M *23I 42C 48D*	None
Bm3	-952.79	5	$p_0 = 0.493$ $\omega_0 = 0.089$ $p_1 = 0.512$ $\omega_1 = 1.328$ $p_2 + p_3 = 0.374$ $\omega_2 = \infty$	Bm3 vs. m3:**	6M 17R 29Q 45M 49S 53A 60H 69S 70A *4I 41M 24S 54S 55Q* **23I 25S 44S 51V** 42C 48 D 67G	R2 S3 K 8 E10 D50 A57 61A 64V 65E

Note. p is the number of free parameters for the ω distribution. Sites potentially under positive selection are identified using the *Lamellibrachia m sp* sequence as the reference see Fig.2b. Grey sites have a posterior probability of 99%, **bold style >95%,** <u>underlined</u> >90%, 80 %> *italic style* >90%, normal style < 80%. **: P> 0.95

Table 3

Log Likelihood Scores and Parameter estimates for the lineage-specific, the site-specific and the branch-site models from globin B2 cluster

Models for B2	L	p	Estimates of parameters	LRT (df)	Positively selected sites	
lineage-specific						
A	-1146.66	1	$\omega = 0.122$	B vs. A (1): no significant	None	
B	-1146.35	2	$\omega_0 = 0.175$, $\omega_1 = 0.292$		Not allowed	
C	-1142.26	9	**Reported in Fig.7b**	B vs. C (7): no sign	Not allowed	
site-specific						
m1	-1146.86	1	$p_0 = 0.289$ $\omega_0 = 0$ $p_1 = 0.710$ $\omega_1 = 1$	m2 vs. m1 (2): **	Not allowed	
m2	-1113.31	3	$p_0 = 0.270$ $\omega_0 = 0$ $p_1 = 0.086$ $\omega_1 = 1$ $p_2 = 0.642$ $\omega_2 = 0.146$		None	
m3 (K=2)	-1119.24	3	$p_0 = 0.312$ $\omega_0 = 0.004$ $p_1 = 0.687$ $\omega_1 = 0.213$	m3 vs. A (2): **	None	
m7	-1120.60	2	$p = 0.492$, $q = 2.35$		Not allowed	
m8	-1114.40	4	$p_0 = 0.938$, $p_1 = 0.061$ $p = 0.469$, $q = 2.977$ $\omega = \infty$	m8 vs. m7 (2): no sign	None	
branch-site					Foreground species	Background species
Bm2	-1143.17	3	$p_0 = 0.290$ $\omega_0 = 0$ $p_1 = 0.594$ $\omega_1 = 1$ $p_2 + p_3 = 0.114$ $\omega_2 = 85.75$	Bm2 vs. m1 (2):**	28F *30A* **32C**	none
Bm3	-1117.11	5	$p_0 = 0.349$ $\omega_0 = 0.013$ $p_1 = 0.459$ $\omega_1 = 0.252$ $p_2 + p_3 = 0.191$ $\omega_2 = \infty$	Bm3 vs. m3 (2):no sign	N7 A17 S41 S47 A50 V54 28F 30A 32C	none

Note. p is the number of free parameters for the ω distribution. Sites potentially under positive selection are identified using the *Lamellibrachia m sp* sequence as the reference. See Fig.2b. Grey sites have a posterior probability of 99%, **bold style >95%,** underlined >90%, 80 %> *italic style* >90%, normal style < 80%. **: P>0.95

Résumé : Les hémoglobines en bicouches hexagonales du vestimentifère *Riftia pachyptila* des sources hydrothermales profondes de la dorsale du Pacifique Est, ont la propriété de lier et transporter simultanément de l'oxygène et du sulfure d'hydrogène (H_2S) des branchies du ver aux bactéries endosymbiotiques sulfo-oxydantes qu'il héberge. Ces dernières relarguent des composés organiques qu'elles synthétisent par chimiosynthèse à partir de H_2S, O_2 et CO_2. Ces composés sont directement assimilables par leur hôte. Pour comprendre comment la fonction de fixation de H_2S est apparue et a été maintenue au cours de l'évolution nous avons caractérisé (cloné et séquencé) la famille multigénique des hémoglobines de *Riftia pachyptila*. L'étude de l'évolution moléculaire de ces globines montre que les paralogues de cette famille ont des vitesses évolutives différentes: les deux globines impliquées dans la fixation de H_2S grâce à la présence de résidus cystéine libre présentent des vitesses évolutives plus lentes (forte pression de sélection) que les autres globines impliquées dans le seul transport de l'oxygène. L'environnement moléculaire des cystéines est fortement conservé dans les globines orthologues des espèces d'annélides fixant ou non le sulfure d'hydrogène. L'analyse des substitutions synonymes et non-synonymes entre globines orthologues montre que le site occupé par la cystéine chez les annélides des milieux réduits et occupé par d'autres acides aminés chez les annélides des milieux oxydés, est sous sélection positive chez les espèces vivant dans des milieux sans H_2S. En confrontant ces résultats aux contraintes sélectives (présence de H_2S) des milieux intertidaux, hydrothermaux et terrestres, nous avons proposé un scénario évolutif qui suggère que la fonction de fixation de H_2S via une cystéine libre a été perdue chez les annélides modernes des milieux sans H_2S (relaxation des contraintes suivie de sélection positive). La perte des cystéines libres suggère que cette adaptation aux milieux réduits (riches en H_2S) est un caractère plésiomorphe, et que par conséquent l'ancêtre des annélides aurait pu vivre dans un environnement réduit.

Abstract : The hexagonal-bilayer hemoglobin of the vestimentifera *Riftia pachyptila*, a tube-worm living close to hydrothermal vents on the East Pacific Ridge, have the capacity to simultaneously bind and transport hydrogen sulfide (H_2S) and oxygen (O_2) from the tube worm gills to its endosymbiotic sulfo-oxidising bacteria. These bacteria synthesize from H_2S, CO_2 and O_2 some organic compounds they provide to their host. In order to understand how the sulfide binding function has appeared and has been maintained, we cloned and sequenced the *Riftia pachyptila* hemoglobin multigenic family. Molecular evolution studies show that paralogous globin genes in this family have different evolutionary rates. Two globin strains are involved in H_2S binding, and use for that free cysteine residues. These two globin strains exhibit slower evolutionary rates (strong selection pressure) than globin strains solely involved in oxygen transport. Free cysteines molecular environment is strongly conserved in orthologous globins from both annelid species binding hydrogen sulfide or not. Synonymous and non-synonymous analysis between orthologous globins revealed that the site occupied by the free cysteine residue in annelids living in sulfide-rich environments and occupied by other amino acids in annelids from sulfide-free environments, has undergone positive selection in species from sulfide-free environment. From these results and the selective constrains (H_2S occurrence) found in intertidal, hydrothermal and terrestrial environments, we proposed a scenario suggesting that H_2S binding via a free cysteine residue has been lost in modern annelids living in free-sulfide environments (constraint relaxation followed by positive selection). Our work suggests that this adaptation to sulfide-rich environments is a plesiomorphic feature, and that the annelid ancestor could have live in a sulfide-rich environment.