



**HAL**  
open science

# Caractérisation des familles multigéniques des globines et des linkers codant l' hémoglobine extracellulaire de *Arenicola marina*, dans le cadre de la mise au point d'un substitut sanguin humain

Christine Chabasse

## ► To cite this version:

Christine Chabasse. Caractérisation des familles multigéniques des globines et des linkers codant l' hémoglobine extracellulaire de *Arenicola marina*, dans le cadre de la mise au point d'un substitut sanguin humain. Biologie cellulaire. Paris 6, 2005. Français. NNT : . tel-01114957

**HAL Id: tel-01114957**

**<https://hal.sorbonne-universite.fr/tel-01114957>**

Submitted on 10 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'UNIVERSITE  
PIERRE ET MARIE CURIE PARIS VI**

**Spécialité**  
Biologie Moléculaire

Présentée par

Mlle Christine CHABASSE

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITE PARIS VI

**Caractérisation des familles multigéniques des globines et des linkers codant  
l'hémoglobine extracellulaire de *Arenicola marina*, dans le cadre de la mise au  
point d'un substitut sanguin humain**

Soutenue le 16 Décembre 2005

devant le jury composé de :

M. Serge Vinogradov	Wayne State University School of Medicine, USA	Rapporteur
M. Bernard Knoops	Université catholique de Louvain	Rapporteur
M. Jean Delettré	Université Pierre et Marie Curie, Paris VI	Examinateur
M. Didier Jollivet	CNRS, Station Biologique de Roscoff	Examinateur
M. André Toulmond	Université Pierre et Marie Curie, Paris VI	Examinateur
M. Franck Zal	CNRS, Station Biologique de Roscoff	Directeur de Thèse

## REMERCIEMENTS

Je tiens tout d'abord à remercier Franck Zal pour m'avoir accueillie dans l'équipe Ecophysiologie: Adaptation et Evolution Moléculaires et m'avoir permis de réaliser cette expérience enrichissante à la Station Biologique de Roscoff.

Je souhaiterais aussi remercier chaleureusement Xavier Bailly pour l'encadrement scientifique que j'ai reçu au cours de ces trois dernières années, mais aussi pour son dynamisme, sa bonne humeur et son envie perpétuelle d'aller encore plus loin.

Je tiens à exprimer toute ma reconnaissance au Professeur Jean Delettré qui m'a fait l'honneur d'accepter d'être le Président de mon jury de thèse, et qui a toujours su être attentif à chacune de nos rencontres.

Je souhaiterais par ailleurs exprimer toute ma gratitude au Professeur Serge Vinogradov pour avoir accepté d'être rapporteur de cette thèse, mais aussi pour ses commentaires et discussions sur l'ensemble de nos travaux, et sur les globines en général.

Je remercie également le Professeur Bernard Knoops pour avoir participé en tant que rapporteur au jury de cette thèse et pour toutes les suggestions concernant le manuscrit.

Merci au Dr. Didier Jollivet pour avoir accepté d'expertiser cette thèse, et pour son regard sur mon travail et ses encouragements.

Enfin, merci au Professeur André Toulmond d'avoir accepté de faire partie de mon jury de soutenance. Dorénavant, je vérifierai systématiquement l'étymologie des termes scientifiques que j'emploie...

Je souhaiterai aussi remercier Sylvia Dewilde et Luc Moens pour leurs conseils et remarques concernant les articles. Par ailleurs, je tiens à exprimer ma gratitude envers les personnes qui ont contribué à l'obtention des espèces citées dans cette thèse: merci à João Gil pour les spécimens de *Glycera dibranchiata* et à Dave Lincoln pour les spécimens de *Amphitrite ornata*.

Merci à toutes les personnes de l'équipe Ecophysiologie, ainsi que l'ensemble des membres de la station Biologique. Un grand merci à Manue, pour m'avoir soutenue/supportée/encouragée, merci pour nos soirées, nos ballades plus ou moins sportives, tes conseils en tout genre et pour m'avoir permis d'avoir un chat à mi-temps !! Une pensée particulière pour Vincent, qui a sa part de responsabilité dans ma venue à Roscoff: merci pour tes visites dans le bureau et pour les innombrables pauses-café (ou autre) à chaque fois que c'était nécessaire (ou non, d'ailleurs...).

Merci à Virginie pour avoir égayé le bureau, pour nos apéros-cacahuètes et pour nos duos violon/piano.

Merci aussi à tous ceux que j'ai croisé et qui m'ont soutenu pendant ces trois années: Morgan, pour m'avoir toujours trouvé de la place pour le séquençage, Stéphane Egée, pour tes innombrables conseils et nos cafés du dimanche, Claire, Lise, Marie, Delphine et Baptiste, pour avoir ouvert les portes de l'équipe EGPM, Alain pour tes cours de danse bretonne et tout le reste, Ana et Allan pour la découverte de la cuisine brésilienne, Julien, pour avoir dépoussiéré mon piano, Fabrice, Séverine, Laurent, Christophe, Sophie M, Rozenn et tous les autres que je ne peux pas citer...

Merci à JP, mon « petit » frère, d'être toujours là pour moi et pour avoir su trouver les mots qu'il fallait dans les périodes plus difficiles.

Merci aussi à ma mère et à mon père à qui je dois tant, et sans qui je ne serai pas arrivée jusque là.

Enfin, pour ton soutien incessant, même dans les pires périodes de doutes inhérent à une thèse, et pour avoir cru en moi... merci Matt.

## ABREVIATIONS

ADNc: ADN complémentaire

ARNm: ARN messenger

Cygb: Cytoglobine

DHP: Déhaloperoxydase

EDTA: Acide éthylènediaminotétraacétique

ESI-MS: ElectroSpray Ionization Mass Spectrometry

FHbs: Flavohémoglobine

Gb : Globine

GCS: Globin-Coupled Sensor ou capteur sensible de type globine

Hb: Hémoglobine

Hc: Hémocyanine

Hr: Hémérythrine

HBL-Hb: Hexagonal Bilayer Hemoglobin

Mb: Myoglobine

MET: Microscopie électronique à transmission

M-MLV-RT: Moloney Murine Leukemia Virus Reverse Transcriptase

Ngb : Neuroglobine

LDL: Low Density Lipoprotein

LDLR: Low Density Lipoprotein Receptor

LDL-A: Low density lipoprotein Receptor Class A repeat.

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

Pgb: Protoglobine

RACE: Rapid Amplification of cDNAs Ends

SDgb: Hémoglobines Simple Domaine

SOD: Superoxyde dismutase

UTR: UnTranslated Region

# **Préambule**

## PREAMBULE

Les annélides sont présents aussi bien dans les écosystèmes terrestres que marins et expriment trois types de globines (intracellulaires circulantes et non circulantes et extracellulaires). Les hémoglobines extracellulaires d'annélides sont composées à la fois de sous-unités de globines et de chaînes de structure, dont l'association confère une structure quaternaire dite en bicouche hexagonale (HBL). Des travaux antérieurs concernant les HBL-Hb d'annélides ont mis en évidence des propriétés fonctionnelles qui pourraient permettre la mise au point d'un substitut sanguin. Dans le cadre de la production d'un tel substitut sanguin à partir de l'HBL-Hb de l'annélide polychaete *Arenicola marina*, l'alternative à l'extraction et à la purification de cette hémoglobine à partir d'annélides vivantes est d'envisager sa production par génie génétique. Nous nous sommes donc intéressés pendant cette thèse à la caractérisation de la famille multigénique des globines et celle des chaînes de structure qui codent pour cette hémoglobine extracellulaire, et avons analysés les résultats obtenus dans un cadre évolutif et fonctionnel plus large.

Le premier chapitre présente tout d'abord les différents pigments respiratoires, et expose dans un second temps la diversité structurale et fonctionnelle de la superfamille des globines au sein des organismes vivants. Il présente aussi une nouvelle classification de ces globines qui repose à la fois sur leurs caractéristiques structurales et sur les études de phylogénies moléculaires.

Le deuxième chapitre décrit les trois lignées de globines présentes chez les annélides, en mettant l'accent sur les HBL-Hb.

Le troisième chapitre présente l'HBL-Hb de *Arenicola marina* ainsi que la méthodologie employée pour caractériser les deux familles multigéniques des globines et de chaînes de structure.

L'article présenté dans le quatrième chapitre décrit la famille multigénique des globines de *Arenicola marina* en faisant le parallèle entre les résultats obtenus par biologie moléculaire au cours de cette thèse et les précédentes données obtenues par spectrométrie de masse.

L'article présenté dans le cinquième chapitre concerne l'émergence des trois types de globines (extracellulaires et intracellulaires circulantes et non-circulantes) chez les annélides.

Le sixième chapitre présente un troisième article qui concerne l'évolution moléculaire des chaînes de structure d'annélides, les Linkers.

La conclusion (Chapitre 7) rappelle les différents résultats obtenus, replacés dans le contexte plus large de l'émergence des HBL-Hb chez les annélides.



# **Sommaire**

## SOMMAIRE

### Chapitre 1 : Introduction générale : les globines et autres « pigments respiratoires ».. 1

<b>I Les globines et la phylogénie moléculaire.....</b>	<b>1</b>
A- Les trois domaines du vivant et les globines .....	1
B- La famille multigénique des globines .....	2
<b>II Hémocyanines, hémérythrine et globines ne sont pas que des pigments respiratoires transporteurs de O<sub>2</sub>.....</b>	<b>4</b>
A- Les hémocyanines .....	4
B- Les hémérythrine et myohémérythrine .....	5
C- Les globines.....	6
<b>III Les globines de l'arbre universel du vivant.....</b>	<b>8</b>
A- Deux types structuraux de globines : les globin-fold 3/3 et 2/2.....	8
B- Les globines « deux-sur-deux » .....	10
C- Les globines « trois-sur-trois ».....	11
1) Les protoglobines et les capteurs sensitifs de type globine .....	11
2) Les autres 3/3 Hbs.....	13
a) Les Flavohémoglobines .....	13
b) Les globines Mono-domaines .....	14
c) Les globines Symbiotiques et Non Symbiotiques de plantes .....	16
d) Les Neuroglobines .....	17
e) Les Cytoglobines .....	19
f) Les Globines .....	19
D- Conclusion.....	20
<b>IV Evolution des globines au sein de l'arbre universel du vivant.....</b>	<b>21</b>
A- Structure des gènes codant pour les globines.....	21
B- Ubiquité des globines au sein de l'arbre universel.....	23
<b>V Conclusion : la globine, quelle(s) fonction(s) ancestrale(s) ?.....</b>	<b>25</b>

### Chapitre 2 : Les globines d'annélides ..... 27

<b>I Les globines intracellulaires .....</b>	<b>28</b>
A- Les globines intracellulaires non circulantes (ou Myoglobines).....	28
B- Les globines intracellulaires circulantes .....	29
<b>II Les globines extracellulaires.....</b>	<b>30</b>
A- Présentation générale.....	30
B- Structure et composition des HBL-Hbs .....	31
C- Les globines des HBL-Hbs .....	33
1) Caractéristiques.....	33
a) Ponts disulfures intra-chaînes .....	33
b) Peptides signaux .....	34
2) Evolution des familles multigéniques des globines extracellulaires.....	34
a) La famille multigénique .....	34
b) Structure des gènes : positions des introns.....	35
3) Cystéines libres et fixation de l'hydrogène sulfuré.....	36
D- Les chaînes de structure ou Linkers .....	38

1) Les modules LDL-A .....	38
2) Structure des gènes de linkers .....	40
3) Activité superoxyde-dismutase .....	40

**Chapitre 3 : Introduction à la caractérisation des gènes codant pour l'hémoglobine extracellulaire de *Arenicola marina* ..... 41**

<b>I Contexte de l'étude .....</b>	<b>41</b>
A- Les substituts sanguins .....	41
B- Les différentes voies de recherche .....	41
C- Les premiers résultats sur les HBL-Hbs de <i>L. terrestris</i> et <i>A. marina</i> .....	43
<b>II Introduction à la caractérisation des gènes codant pour l'hémoglobine extracellulaire de <i>Arenicola marina</i>.....</b>	<b>44</b>
A- Description de l'annélide polychaete <i>Arenicola marina</i> .....	44
B- L'HBL-Hb de <i>Arenicola marina</i> .....	45
C- Méthodologie .....	46
D- Contexte d'analyse des séquences obtenues .....	47

**Chapitre 4 : Premier article : The multigenic family of the extracellular hemoglobin from the annelid polychaete *Arenicola marina* ..... 49**

**Chapitre 5 : Deuxième article : Evolutionary functional transition from oxygen transport to halometabolites detoxification in the annelids globins family ..... 69**

**Chapitre 6 : Troisième article : Gene structure and molecular phylogeny of the linker chains from the giant annelid hexagonal bilayer hemoglobins ..... 94**

**Chapitre 7 : Conclusion et perspectives ..... 117**

<b>I Résultats et perspectives.....</b>	<b>117</b>
<b>II Emergence des lignées de globines chez les Annélides.....</b>	<b>120</b>
A- Positions des introns.....	120
B- Caractéristiques structurales des HBL-Hbs.....	121
C- Les globines extracellulaires et les Linkers.....	121

**Bibliographie..... 123**

**Annexes..... 147**

<b>Annexe 1 : Premières étapes de la synthèse in vitro d'hémoglobine extracellulaire de haut poids moléculaire, en vue d'élaborer un substitut sanguin .....</b>	<b>147</b>
<b>Annexe 2 : Résultats préliminaires de la surexpression d'un monomère de globine et de linker de <i>Arenicola marina</i> .....</b>	<b>148</b>
<b>Annexe 3 : Matériel et Méthode .....</b>	<b>150</b>

<b>Annexe 4</b> : Séquences nucléotidiques et protéiques des globines et de linkers de <i>Arenicola marina</i> obtenus pendant la thèse .....	152
A- Séquences nucléotidiques des globines et de linkers de <i>Arenicola marina</i> .....	152
B- Séquences protéiques des globines et de linkers de <i>Arenicola marina</i> .....	154
<b>Annexe 5</b> : Séquences nucléotidiques et protéiques des globines intracellulaires non-circulantes et circulantes et des linkers d'annélides obtenues pendant la thèse .....	156
<b>Annexe 6</b> : Séquence des introns de globines extracellulaires d'annélides .....	159
<b>Annexe 7</b> : Séquence des introns des linkers d'annélides .....	163
<b>Annexe 8</b> : Séquence des introns des globines intracellulaires circulantes et non circulantes d'annélides .....	166

# **Chapitre 1**

**Introduction générale : les globines et  
autres « pigments respiratoires »**

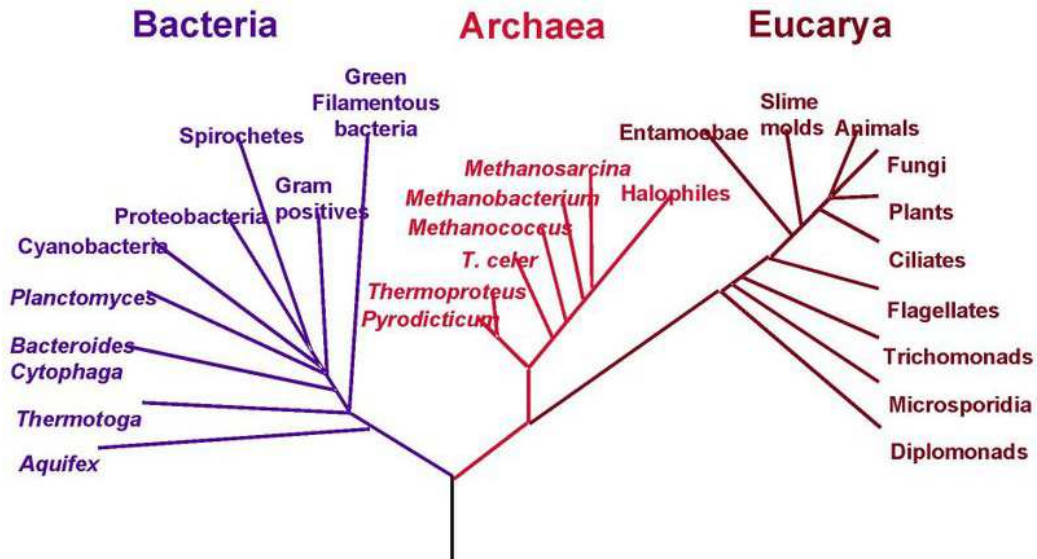
## **Chapitre 1 : INTRODUCTION GENERALE : LES GLOBINES ET AUTRES « PIGMENTS RESPIRATOIRES »**

### **I Les globines et la phylogénie moléculaire.**

#### A- Les trois domaines du vivant et les globines.

Au cours des années 60, des techniques de plus en plus fines de séquençage des protéines (Zuckerlandl, Jones and Pauling 1960; Niall and Edman 1962) et des acides nucléiques (Sanger, Brownlee and Barrell 1965) ont permis d'accéder aux séquences polypeptidiques (séquences primaires) et aux séquences nucléotidiques codantes et non codantes. Ces techniques révolutionnaires ont permis d'avoir accès à un nouveau type d'information : plus que le simple enchaînement d'acides aminés permettant d'accéder à des informations fonctionnelles au sein de la protéine ou que l'enchaînement des nucléotides à partir desquels sera découvert entre autre le code génétique, ces séquences ont constitué les bases d'une discipline émergente, **l'évolution moléculaire**. Les différences observées à partir de la comparaison de séquences de protéines homologues de différents organismes plus ou moins génétiquement éloignés vont permettre de considérer ces informations comme marqueurs de l'évolution : "Molecules as documents of evolutionary history" (Zuckerlandl and Pauling 1965).

Dans un premier temps, des protéines telles que les globines furent utilisées comme marqueur génétique, de même que les histones ou les cytochromes c (Margoliash 1963). Au cours des trois décennies suivantes, des séquences protéiques et nucléiques de plus en plus nombreuses et issues d'organismes de plus en plus variés furent disponibles grâce à des techniques de biologie moléculaire de plus en plus performantes. Puis les séquences codant pour certaines sous-unités des ribosomes furent utilisées pour essayer de résoudre et construire l'arbre universel du vivant (Woese, Kandler and Wheelis 1990). Contrairement aux globines qui présentent de fortes divergences de séquences, les ARNs sont soumis à des contraintes fortes et stables et sont relativement non affectés par les changements environnementaux (Woese, Kandler and Wheelis 1990). A partir de ces séquences, Woese proposa ainsi une classification des organismes vivants en trois grands domaines : Bacteria, Eukaryota et Archaea (Woese, Kandler and Wheelis 1990), représentée dans la Figure 1.



**Figure 1 :** Représentation de l'arbre phylogénétique universel. D'après Woese et al. (Woese, Kandler and Wheelis 1990).

Bien que les globines ne soient pas de bons marqueurs génétiques pour suivre l'évolution des espèces à l'échelle du vivant, nous verrons cependant qu'elles sont exprimées dans les trois domaines décrits par Woese.

### B- La famille multigénique des globines.

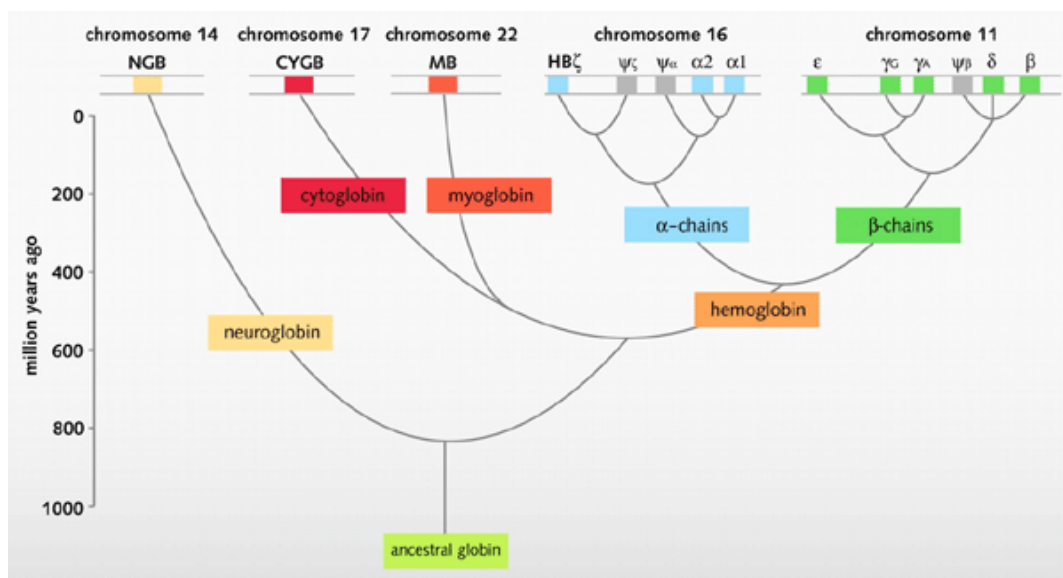
Une des premières protéines ainsi étudiée fut l'hémoglobine des vertébrés. Des anomalies concernant des altérations des propriétés de l'hémoglobine sont apparues comme responsables de plusieurs pathologies, telles que les thalassémies ou l'anémie falciforme. C'est en comparant les profils de mobilité électrophorétique de l'hémoglobine d'individus sains et atteints de drépanocytose, ainsi que le pourcentage d'hémoglobine normale et mutante au sein des érythrocytes de ces individus, que Pauling mit en évidence l'implication d'un gène codant pour l'hémoglobine et de deux allèles distincts expliquant cette pathologie (Pauling, Itano and et al. 1949). Les causes de cette **maladie héréditaire** (aujourd'hui désignée comme **maladie génétique**) furent ensuite décrites par Ingram, qui montra que la différence entre les deux hémoglobines se traduisait par le remplacement d'un acide aminé, un acide glutamique par une valine dans l'hémoglobine mutante, et ceci en raison d'une mutation au niveau du gène codant pour l'hémoglobine (Ingram 1956; Ingram 1957). Au cours de ses travaux sur les anémies méditerranéennes (thalassémies majeures et mineures), Roche étudia les différences

de solubilité entre l'hémoglobine adulte et fœtale et mit en évidence l'existence d'une hétérogénéité des chaînes composant l'hémoglobine, en d'autres termes, une expression concomitante des deux hémoglobines chez l'adulte (Roche, Derrien and Roques 1952). L'ensemble de ces travaux conduisit Ingram en 1961 à proposer un premier schéma pour l'évolution des gènes de globine chez l'Homme (Ingram 1961). Il proposa que l'ancêtre commun à toutes les globines actuelles était une protéine monomérique de type myoglobine. L'hypothèse de copies de ce même gène obtenues à partir de duplications successives fut avancée pour expliquer l'émergence de cet ensemble de gènes codant pour des protéines sensiblement différentes.

Chez l'homme, les différentes chaînes polypeptidiques mises en évidence sont  $\alpha 1$ ,  $\alpha 2$ ,  $\beta$ ,  $\delta$ ,  $\epsilon$ ,  $\gamma A$ ,  $\gamma G$ ,  $\theta$  et  $\zeta$ , chacune étant codée par un gène différent, transcrit à différents stades de développement : embryonnaire ( $\alpha$ ,  $\epsilon$ ,  $\gamma$ ,  $\theta$  et  $\zeta$ ), fœtal ( $\alpha$  et  $\gamma$ ) et adulte ( $\alpha$ ,  $\beta$  et  $\delta$ ). L'ensemble de ces gènes forme une **famille multigénique**.

Aujourd'hui, plusieurs centaines de séquences codant pour des globines sont disponibles dans les banques de données. L'ensemble de ces données a permis d'illustrer la diversité et la complexité de la superfamille des globines, protéines ubiquitaires que l'on retrouve exprimées dans les trois lignées évolutives majeures de l'arbre universel du vivant (Vinogradov et al. 2005).

La figure 2, restreinte aux globines de vertébrés, illustre la diversité des différents types de globines exprimées à différents niveaux cellulaires, tissulaires ou stades de développement (Burmester et al. 2002).



**Figure 2 :** Diversité des globines de vertébrés. D'après Burmester et al. (Burmester et al. 2002).

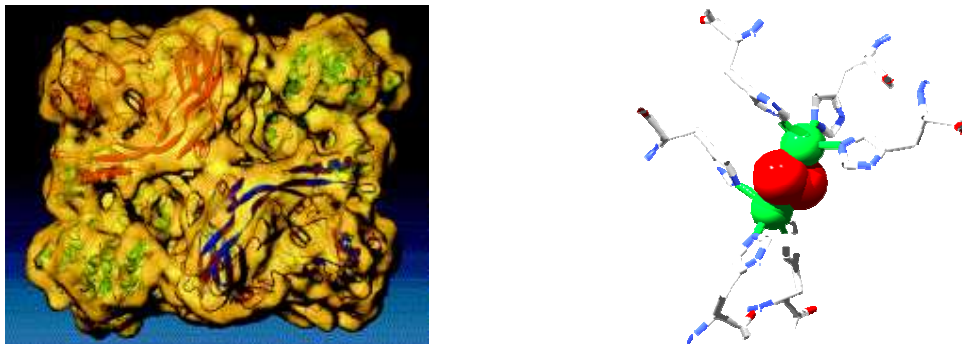


## **II Hémocyanines, hémérythrine et globines ne sont pas que des pigments respiratoires transporteurs de O<sub>2</sub>.**

Historiquement, ces trois métalloprotéines ont été caractérisées par leur couleur (pigments) et leur fonction (fixer et transporter l'oxygène) : les **hémocyanines** (bleues) lient directement l'oxygène grâce à deux atomes de cuivre, les **hémérythrine** (roses) font intervenir deux atomes de fer, et les **globines** (rouges), qui font intervenir un atome de fer et un groupement prosthétique, l'hème.

### A- Les hémocyanines.

Les hémocyanines (Hc) sont dissoutes dans le tissu sanguin d'un grand nombre de mollusques et d'arthropodes (van Holde and Miller 1995). Chez les mollusques, elles sont composées de plusieurs sous-unités de polypeptides multi-domaines qui s'assemblent pour former de larges polymères de 3.000 à 10.000 kDa (van Holde and Miller 1995), alors que chez les arthropodes, l'unité de base est un hexamère de 450 kDa, dont le degré d'assemblage varie en fonction de l'espèce étudiée (Figure 3) (Markl and Decker 1992). Bien que la structure quaternaire et les séquences primaires des Hc soient très différentes entre ces deux phyla, le site de liaison de l'oxygène fait toujours intervenir deux atomes de cuivre (Figure 3).



**Figure 3 :** **A gauche :** Reconstruction 3D de l'unité de base (Hexamère) de l'hémocyanine de l'arthropode *Palinurus elephas* : les différents monomères sont colorés en rouge, bleu et vert (Meissner et al. 2003). **A droite :** Structure du site de liaison à l'oxygène des hémocyanines. Les atomes de cuivre sont représentés en vert, l'oxygène en rouge.

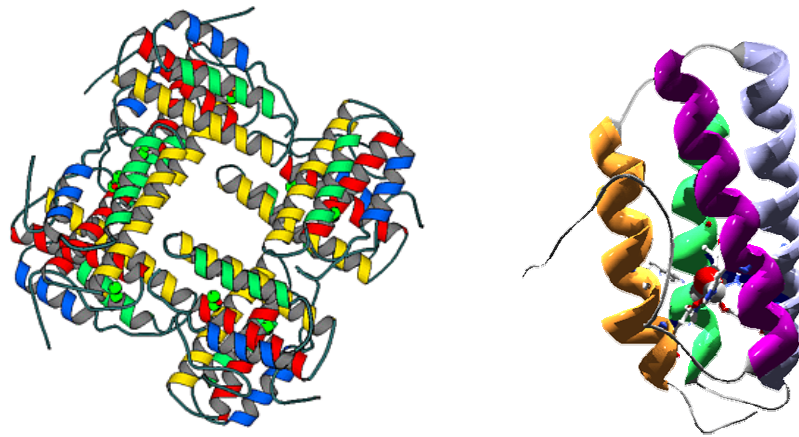
Plusieurs études ont montré que les Hc possédaient, en plus de la capacité à lier l'oxygène, une faible activité phénoloxidase (Salvato et al. 1998; Decker and Terwilliger 2000; Decker and Tuczec 2000) et tyrosinase (van Gelder, Flurkey and Wichers 1997). Les tyrosinases et phénoloxydases représentent une classe de protéines ayant probablement évolué à partir d'une protéine ancestrale dont la fonction était de protéger les organismes primitifs (au métabolisme anaérobie) de l'oxygène toxique issu de la photosynthèse (Canfield and Teske 1996; van Holde, Miller and Decker 2001). Les études réalisées sur les Hc semblent indiquer que la capacité à fixer de l'oxygène ait été développée bien après ces fonctions de phénoloxydase et tyrosinase (van Holde, Miller and Decker 2001; Immesberger and Burmester 2004).

D'autres protéines, notamment les cryptocyanines de crustacés ou les hexamérines d'insectes (Telfer and Kunkel 1991), ont des séquences et des structures très similaires à celles des Hc, sans pour autant contenir d'atomes de cuivre ni posséder la capacité de lier l'oxygène (Terwilliger and O'Brien 1992). L'émergence de telles protéines est expliquée par des événements de duplication, dont la (ou les) copie(s) dupliquée(s) aurai(en)t perdu le site de fixation au cuivre (Terwilliger 1998). Ces protéines de la famille des Hc illustrent les mécanismes évolutifs de diversification de fonctions après duplication d'un gène ancestral suivie d'événements de spéciation (Terwilliger 1998).

#### B- Les hémérythrines et myohémérythrines.

Les hémérythrines (Hr) ont été décrites dans quatre phyla de protostomiens : les siponcles, les brachiopodes, les priapulides et certains annélides (Klotz and Klotz 1955; Demuynck et al. 1991; Terwilliger 1998). La structure tridimensionnelle de ces protéines consiste en 4 hélices  $\alpha$  regroupées autour d'un site actif composé de deux atomes de fer (Figure 4).

Le terme générique d'hémérythrine recouvre en fait deux protéines homologues issues d'une duplication ancestrale: la première, trouvée dans les muscles et appelée myohémérythrine (MHR), se présente sous forme d'une protéine monomérique de 120 acides aminés (Hendrickson and Klippenstein 1974; Takagi and Cox 1991); la seconde, appelée hémérythrine, existe sous forme d'homopolymères de polypeptides de 113 à 117 acides aminés et se trouve dans des cellules en circulation dans le système vasculaire (hémérythrocytes) (Kurtz 1992).

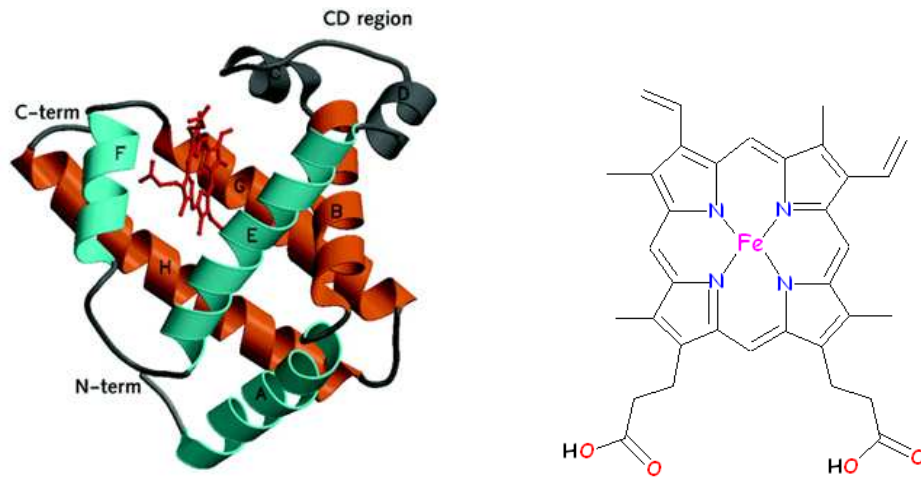


**Figure 4 :** **A gauche :** Représentation 3D de l'hémérythrine octamérique de *Themiste dyscritum* (PDB 1HMD). **A droite :** Représentation 3D d'un des monomères de l'hémérythrine de *T. dyscritum* (PDB 1HMD).

Plusieurs études ont montré que les Hr n'avaient pas pour seule fonction la liaison et le transport de l'oxygène. Des propriétés anti-bactériennes ainsi qu'une capacité de liaison au cadmium ont été mises en évidence chez l'annélide polychaete *Nereis diversicolor* (Nejmeddine et al. 1988; Dhainaut et al. 1989). Par ailleurs, une Hr a récemment été identifiée chez la bactérie méthanotrophique Gram-négative *Methylococcus capsulatus* et aurait un rôle physiologique important dans l'oxydation du méthane (Karlsen et al. 2005).

### C- Les globines.

Le terme générique de globine (Gb) - souvent appelée hémoglobine (Hb) par abus de langage- recouvre une large variété de protéines homologues ayant comme caractéristique commune d'être composées par l'association d'une partie protéique, et d'une partie prosthétique, l'hème (Figure 5). L'hème est un noyau aromatique, la protoporphyrine IX, dont les quatre azotes centraux sont liés à un atome de fer. Les travaux de Kendrew (1960) sur la structure de la myoglobine de cachalot ont montré que la globine est un polypeptide d'environ 150 acides aminés qui possède une structure tridimensionnelle typique appelée « globin-fold », composée par la succession de 6 à 8 hélices  $\alpha$  (Figure 5) (Kendrew et al. 1960). Conventionnellement, on désigne les hélices  $\alpha$  par une lettre, de A à H, de l'extrémité N-terminale à l'extrémité C-terminale (Perutz 1970).



**Figure 5** : A- Structure quaternaire de la myoglobine de cachalot. Les hélices  $\alpha$  sont désignées de A à H et l'hème est représenté en rouge. D'après Pesce et al. (Pesce et al. 2002)). B- Noyau protoporphyrine IX ou hème.

Les Gbs ont été décrites dans les trois grands domaines du vivant, Bacteria, Archaea et Eucaryota (Vinogradov et al. 2005) et de nombreux travaux montrent que ces protéines ont une localisation tissulaire et/ou cellulaire beaucoup plus complexe que ce qui avait tout d'abord été observé, à la fois chez les organismes multicellulaires et unicellulaires.

Les **Gbs intracellulaires cytoplasmiques non circulantes** sont exprimées spécifiquement au sein d'un tissu. On peut citer par exemple la « eye-globin » (GbE) du poulet *Gallus gallus*, exprimée uniquement dans les tissus oculaires (Kugelstadt et al. 2004), les neuroglobines de vertébrés, exprimées dans les neurones du système nerveux central et périphérique (Hankeln et al. 2005), les Hbs symbiotiques et non symbiotiques de plantes (telles que le riz *Oryza sativa* (Arredondo-Peter et al. 1997)), les Hbs de levures (Keilin and Tissieres 1954)) ou encore de bactéries (Wakabayashi, Matsubara and Webster 1986)).

Les **Gbs intracellulaires cytoplasmiques circulantes** ont été caractérisées au sein de cellules circulantes dans un système vasculaire, telles que l'hémoglobine intracellulaire trouvée dans les érythrocytes chez les vertébrés, ou encore de l'annélide *Glycera dibranchiata* (Vinogradov, Machlik and Chao 1970).

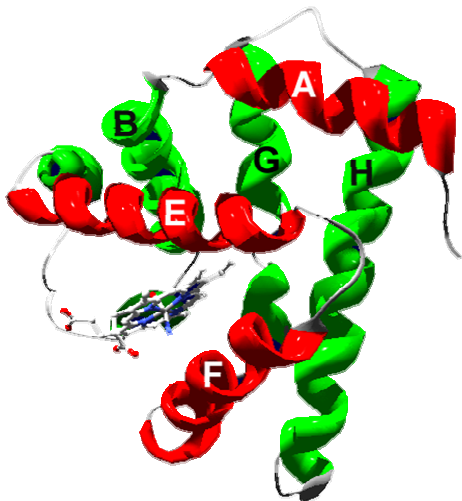
Enfin, on peut trouver des Gbs directement dissoutes dans le compartiment vasculaire, d'où le terme **Gbs extracellulaires**, telles que les Gbs extracellulaires de mollusques ou d'annélides (Weber and Vinogradov 2001).

L'évolution des globines est un problème constamment revisité (Zuckerkindl 1965; Riggs 1991; Hardison 1996; Vinogradov et al. 2005). L'existence de globines aux fonctions et aux structures aussi différentes qu'inattendues remet aujourd'hui en question la fonction unique de transporteur d'oxygène qui leur a été très tôt attribuée. Récemment, la multiplication des données issues de la génomique a permis de mettre en évidence de nouveaux types de globines, et ceci dans les trois domaines du vivant. Ces résultats ont conduit à une nouvelle classification, reposant sur leurs caractéristiques structurales, qui sera décrite dans le paragraphe suivant (Vinogradov et al. 2005).

### III Les globines de l'arbre universel du vivant.

#### A- Deux types structuraux de globines : les globin-fold 3/3 et 2/2.

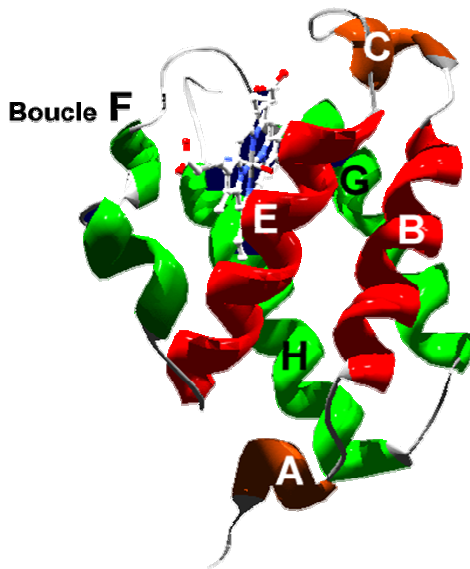
Les premières structures tridimensionnelles des hémoglobines et des myoglobines, obtenues à partir de protéines cristallisées de vertébrés, ont présenté une structure appelée « trois-sur-trois » (ou 3/3), c'est-à-dire trois hélices  $\alpha$  faisant face à trois autres (Figure 6) (Kendrew et al. 1960; Perutz 1960).



**Figure 6 :** Structure 3D de la myoglobine humaine (PDB 2MM1) avec le repliement « trois-sur-trois ». Les hélices A, E et F en rouge font face aux trois autres hélices B, G et H en vert. L'hème au premier plan, est placé entre les hélices E et F.

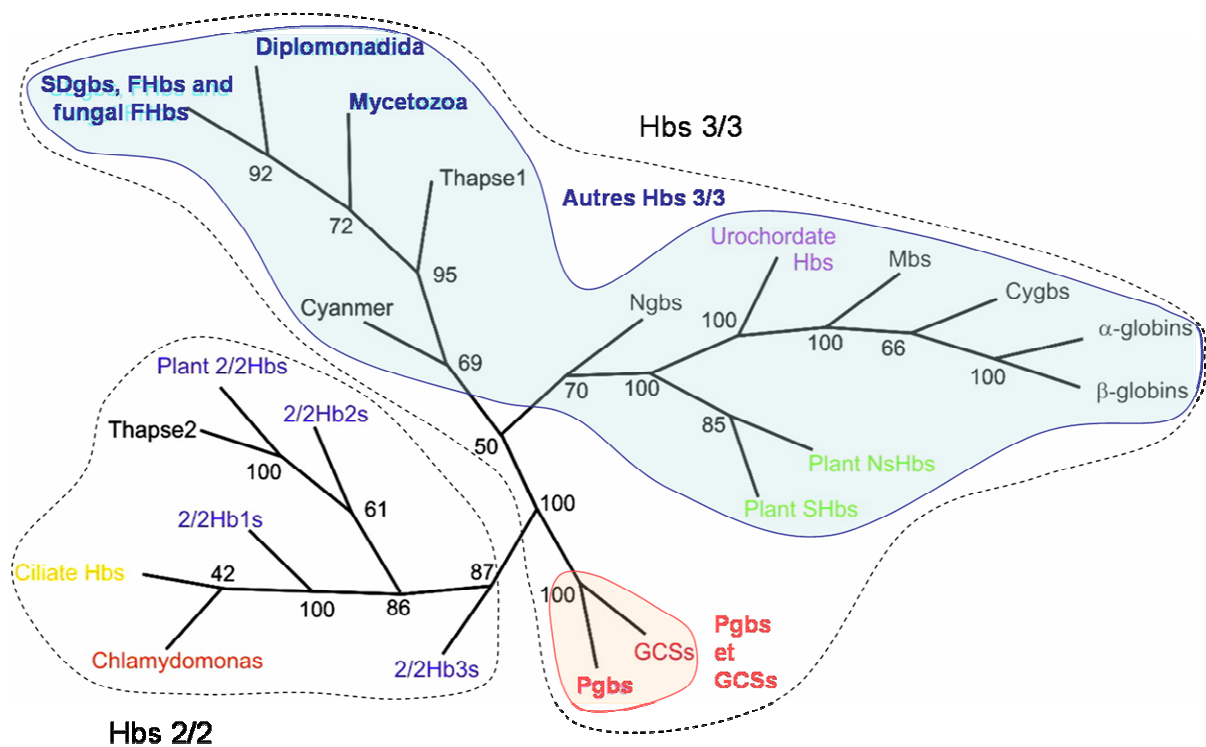
Cette structure était considérée comme universelle chez les globines jusqu'à la découverte d'une autre catégorie de globines présentes chez les bactéries, les plantes et les eucaryotes unicellulaires (Wittenberg et al. 2002; Milani et al. 2005), appelées « hémoglobines tronquées » en raison de leur structure primaire, plus courte de 20 à 40 acides

aminés. Elles présentent une structure tertiaire différente dite «deux-sur-deux » (ou 2/2), où l'hélice A est plus courte, de même que la région inter-hélice CE, et où la plus grande partie de l'hélice F forme une boucle (Figure 7) (Pesce et al. 2000; Milani et al. 2001; Milani et al. 2003).



**Figure 7 :** Structure 3D de la globine de *Thermobifida fusca* (PDB 2BMM) avec le repliement « deux-sur-deux ». Les hélices B et E en rouge font face aux deux autres hélices G et H en vert. L'hème au premier plan, est placé entre les hélices E et F.

D'autres protéines particulières, telles que les flavohémoglobines ou les capteurs sensitifs de type globine sont rencontrées dans certains organismes : ce sont des globines chimériques qui possèdent un domaine additionnel au globin-fold. Les phylogénies moléculaires récentes, réalisées à partir d'alignements multiples des domaines globin-fold des globines des trois grands domaines du vivant révèlent l'existence de trois lignées de globines distinctes : d'une part les globines 2/2, et d'autre part, deux lignées de globines au sein des globines 3/3 (Figure 8) (Vinogradov et al. 2005).



**Figure 8 :** Arbre phylogénétique présentant les groupes majeurs de globines au sein des trois domaines du vivant. D’après Vinogradov et al. 2005 (Vinogradov et al. 2005).

### B- Les globines « deux-sur-deux ».

Les globines 2/2 sont trouvées chez les plantes, les bactéries et les eucaryotes unicellulaires (Wittenberg et al. 2002; Milani et al. 2005). A ce jour, aucune Hb 2/2 n’a été identifiée chez les métazoaires. Au sein de cette lignée de Hb 2/2, trois sous familles ont été décrites : les 2/2 Hb1 (ou HbN), les 2/2 Hb2 (ou HbO) et les 2/2 Hb3. Ces trois groupes diffèrent majoritairement par la composition en acides aminés du site de liaison de l’hème, et par les modalités d’accession à l’hème par les ligands (Wittenberg et al. 2002; Lecomte, Vuletich and Lesk 2005).

La co-expression des trois types de globines de type 2/2 chez *Mycobacterium avium* (Gram–positive) et *Methylococcus capsulatus* (proteobactérie) indiquerait que ces trois lignées d’Hb 2/2 ont divergé de leur ancêtre commun avant la séparation des lignées majeures des Bacteria (Wittenberg et al. 2002).

A ce jour, rien n’est connu sur les fonctions des 2/2 Hb3, ni des 2/2 Hb chez les plantes (Watts et al. 2001). Des études ont montré que les 2/2 Hb1 interviendraient dans la



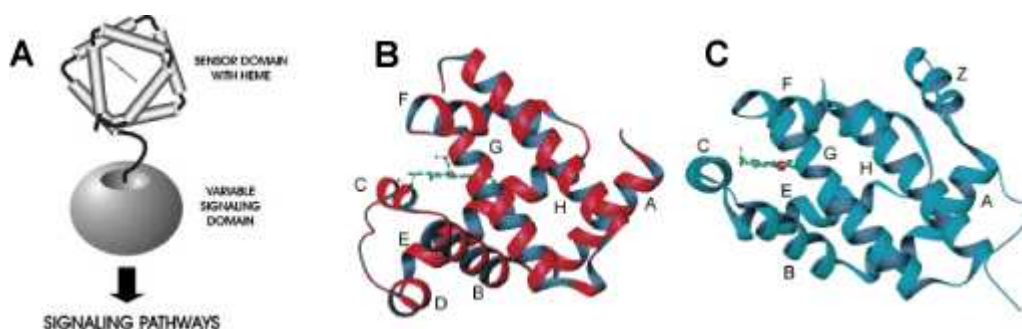
détoxification du monoxyde d'azote (NO) chez les bactéries pathogènes *Mycobacterium bovis* et *M. tuberculosis*, en réponse aux émissions radicalaires de NO par les macrophages de l'hôte lors des processus de défense anti-pathogènes (Ouellet et al. 2002; Pathania et al. 2002a). Les 2/2 Hb2 seraient plutôt exprimées en association avec les membranes cellulaires et stimuleraient significativement la respiration cellulaire, ce qui suggère une interaction avec la chaîne de transport des électrons (Mukai et al. 2002; Pathania et al. 2002b).

### C- Les globines « trois-sur-trois ».

Les phylogénies moléculaires suggèrent l'existence de deux grandes familles au sein de ce groupe de Hb 3/3 : d'un côté les protoglobines (Pgb) et les capteurs sensitifs de type globine (Globin-coupled sensors ou GCS), et de l'autre, toutes les autres 3/3 Hbs, regroupant les globines Mono-Domaines (Single-domain globins ou SDgbs), les flavohémoglobines (FHbs) ainsi que les globines 3/3 de métozoaires (Vinogradov et al. 2005).

#### 1) Les protoglobines et les capteurs sensitifs de type globine.

Les capteurs sensitifs de type globine (GCS) ont été récemment découverts au sein des domaines Bacteria et Archaea (Hou et al. 2001a; Hou et al. 2001b; Freitas, Hou and Alam 2003). Ces protéines chimériques sont caractérisées par la présence dans leur partie N-terminale, d'un domaine de type globin-fold, et dans leur partie C-terminale, d'un domaine de signalisation cellulaire (Figure 9) (Freitas, Hou and Alam 2003).



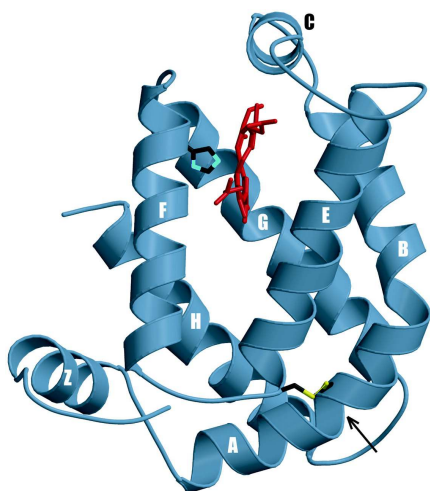
**Figure 9 :** A- Représentation schématique d'une GCS. D'après Hou et al. (Hou et al. 2001b). B-C- Représentation 3D de la myoglobine de cachalot (B) et du domaine globine de la GCS de *Bacillus subtilis* (C). D'après Zhang et al. (Zhang and Phillips 2003).



Chez l'Archaea *Halobacterium salinarum*, la partie C-terminale de ces GCS correspond au domaine de signalisation des MCP (Methyl-accepting Chemotaxis Protein). Les MCP, trouvées chez les Bacteria ou les Archaea, sont des protéines chimioréceptrices transmembranaires capables à la fois de lier divers ligands (ou de réagir à divers stimuli), et d'interagir avec des voies de signalisation intracellulaires (Bourret and Stock 2002). En fonction du stimulus reçu, ces protéines vont produire un signal excitateur permettant la modification de la mobilité de l'organisme (Bibikov et al. 2004).

Au sein des GCS, la fixation de l'oxygène sur l'hème conduit à un changement structural qui modifie l'activité du domaine MCP impliqué dans des voies de signalisation cellulaires. Ce chimiotactisme permet aux bactéries de s'orienter en fonction de la concentration en oxygène dans l'environnement (Hou et al. 2000). Les travaux réalisés sur d'autres GCS ont permis de mettre en évidence différentes fonctions pour ces globines, notamment la régulation de transcription de gène, soit par une interaction directe avec l'ADN, soit en régulant les seconds messagers soit en interagissant directement avec un facteur de transcription (Freitas, Hou and Alam 2003; Freitas et al. 2004; Freitas et al. 2005; Gilles-Gonzalez and Gonzalez 2005).

Des protoglobines (Pgbs) ont été décrites chez les Archaea *Aeropyrum pernix* et *Methanosarcina acetivorans* (Freitas et al. 2004). Elles possèdent la capacité de lier l'oxygène, le monoxyde de carbone et le monoxyde d'azote. Les auteurs ont réalisé un modèle 3D de Pgbs en se basant sur la structure cristallographique du domaine globine de la GCS de *Bacillus subtilis* (Zhang and Phillips 2003). Ce modèle leur a permis de mettre en évidence que l'hème était lié de façon covalente à la globine, et que la tyrosine B10 était orientée de façon parallèle à l'hème et non vers le ligand (Figure 10). Ceci aurait pour conséquence de diminuer la capacité de ces Pgbs à lier l'oxygène et suggérerait que les Pgbs seraient des prototypes des hémoglobines contemporaines (Freitas et al. 2004).

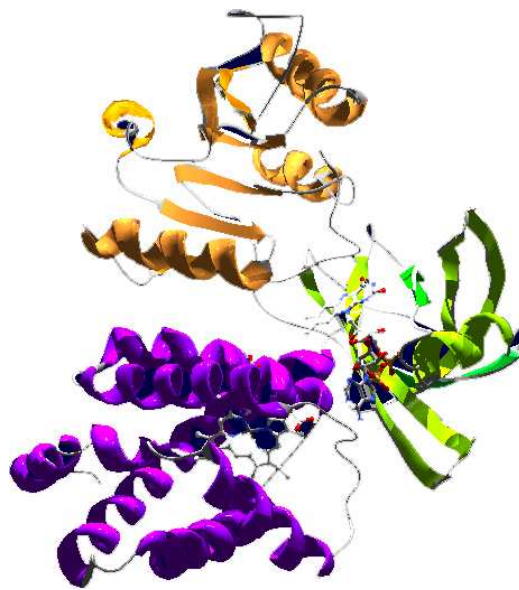


**Figure 10 :** Modèle 3D de la Pgb de *Aeropyrum pernix*. L'histidine proximale est représentée en noir et vert, l'hème en rouge. La présence d'un pont disulfure intra-chaîne (en jaune) est indiquée par une flèche. L'hélice précédant l'hélice A est notée Z. D'après Freitas et al. (Freitas et al. 2004).

2) Les autres Hbs 3/3.

a) Les Flavohémoglobines.

Les flavohémoglobines (FHbs) sont des protéines chimériques comprenant un domaine N-terminal de type globin-fold, et un domaine C-terminal Flavin Adénine Dinucléotide (FAD) réductase, trouvées chez les bactéries et les levures (Figure 11) (Keilin 1953; Keilin and Ryley 1953; Oshino, Oshino and Chance 1973; Wu, Wainwright and Poole 2003).



**Figure 11 :** Représentation 3D de la FHb de *Alcaligenes Eutrophus* (PDB 1CQX). Le domaine N-terminal en globin-fold est représenté en violet. Le domaine C-terminal à flavine est représenté en vert et orange, avec en son centre, le groupement FAD.

Diverses fonctions ont été proposées pour les FHbs bactériennes, en se basant notamment sur leur transcription en réponse soit à une limitation en oxygène soit à un stress oxydatif ou nitrosatif (Poole and Hughes 2000; Farres et al. 2005; Poole 2005). En condition aérobie et en présence de NADH, les FHbs ont une activité nitrique oxyde dioxygénase (ou NOD) qui catalyse l'oxydation des ions radicalaires du NO en nitrate (NO<sub>3</sub>) (Gardner, Costantino and Salzman 1998; Gardner et al. 1998; Gardner 2005), alors qu'en condition

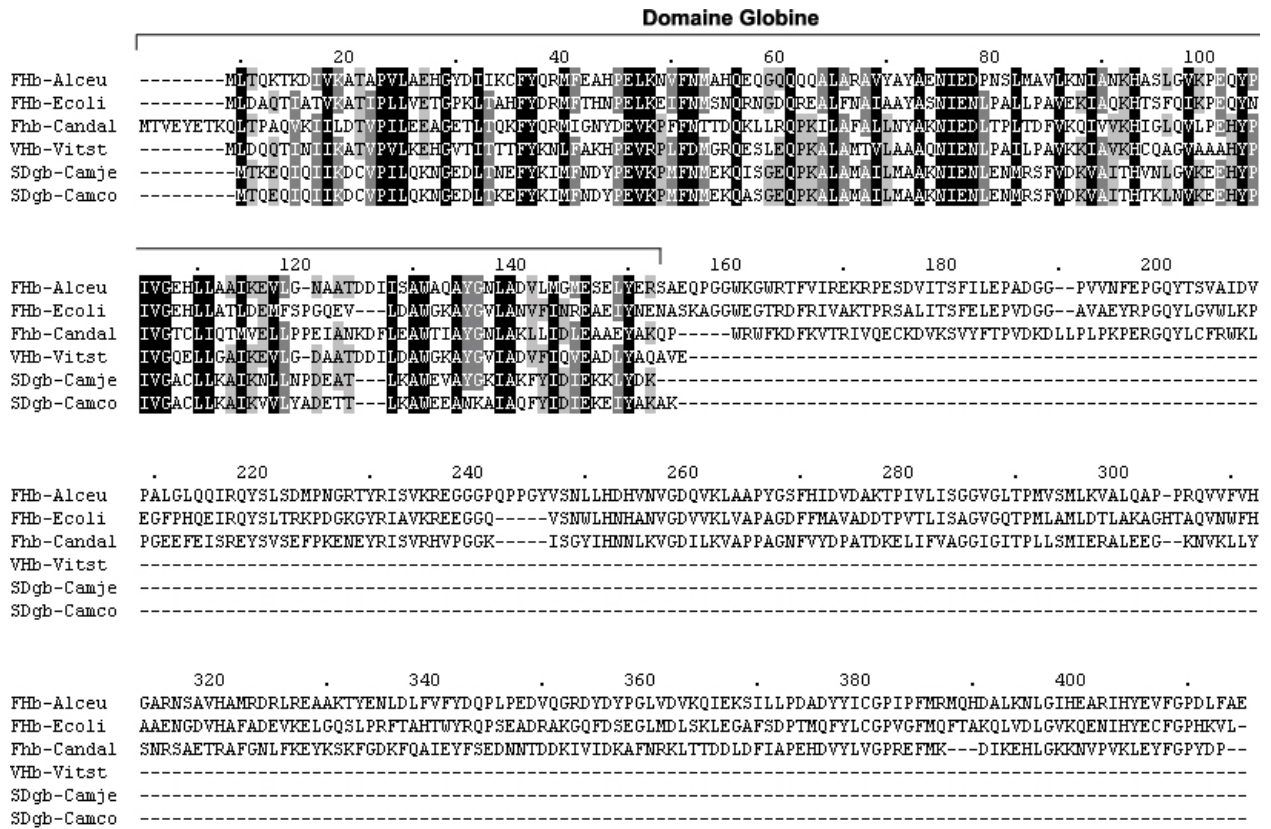
anaérobie, les FHbs possèdent plutôt une activité NO réductase, générant des acides nitreux (N<sub>2</sub>O) (Poole et al. 1997; Farres et al. 2005; Poole 2005).

Les FHbs d'organismes pathogènes tels que les bactéries *Erwinia chrysanthemi* (Boccarda et al. 2005) et *Salmonella enterica* (Stevanin et al. 2002), et les levures *Candida albicans* (Ullmann et al. 2004) et *Cryptococcus neoformans* (de Jesus-Berrios et al. 2003), en interceptant le NO produit par l'hôte, vont permettre l'atténuation de la réaction d'hypersensibilité de l'hôte pendant l'infection. Chez la bactérie Gram-négative *Alcaligenes eutrophus*, la FHb est impliquée dans la diminution de la concentration en NO grâce à son activité de dénitrification qui produit du N<sub>2</sub>O (Cramm, Siddiqui and Friedrich 1994).

La similarité structurale des sites actifs des domaines globin-fold des FHbs et des peroxydases suggère une autre activité pour les FHbs (Ermler et al. 1995; Ilari et al. 2002). Néanmoins, aucune activité peroxydase ni mono-oxygénase n'a été expérimentalement démontrée pour les FHbs (Bonamore et al. 2003a). Récemment, Bonamore et collaborateurs ont suggéré que la FHb de *Escherichia coli* (HMP) possédait une activité alkylhydroperoxyde réductase, catalysant ainsi la réduction des hydroperoxydes de lipides membranaires et permettant la réparation des membranes phospholipidiques suite à un stress oxydatif et/ou nitrosatif (Bonamore et al. 2003a; Bonamore et al. 2003b; D'Angelo et al. 2004).

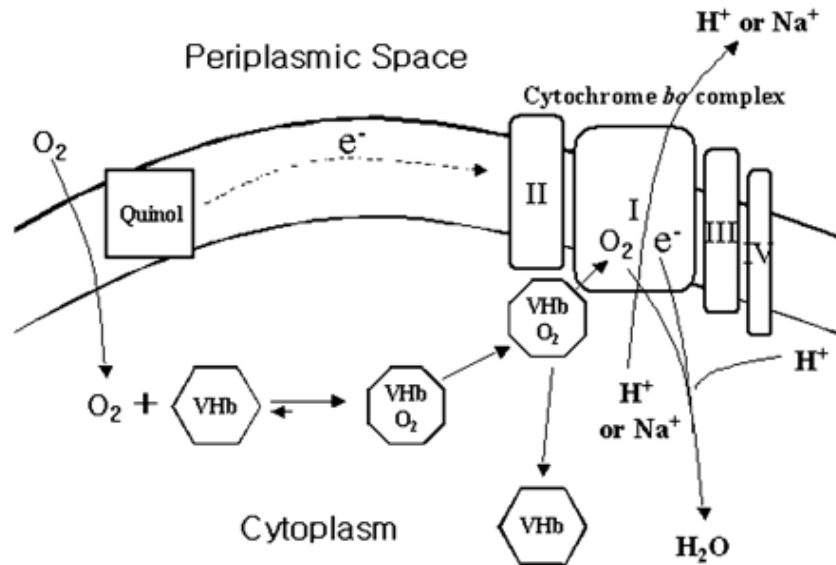
#### b) Les globines Mono-domaines.

Les globines Mono-domaines (Single-domain globins ou SDgbs) sont très proches des FHbs. En effet, leurs séquences s'alignent parfaitement avec le domaine N-terminal globin-fold des FHbs (Frey and Kallio 2003; Wu, Wainwright and Poole 2003). A l'heure actuelle, très peu de SDgbs ont été mises en évidence : elles ont été décrites chez les bactéries *Vitreoscilla stercorari*, sous le terme de Vhb (Figure 12) (Wakabayashi, Matsubara and Webster 1986), ainsi que chez *Campylobacter jejuni* et *Campylobacter coli* (Elvers et al. 2004).



**Figure 12** :\_Alignement multiple réalisé à partir des séquences de FHbs de *Alcaligenes Eutrophus* (P39662) (**Alceu**), de *Escherichia coli* (BAA16460) (**Ecoli**) et de *Candida albicans* (XP\_711060) (**Candal**), de la Vhb de *Vitreoscilla stercoraria* (P04252) (**Vitst**) et des SDgbs de *Campylobacter jejuni* (NP\_282714) (**Camje**) et de *Campylobacter coli* (AAP86201) (**Campcol**).

Les fonctions des SDgbs semblent similaires à celles des FHbs. Chez la bactérie aérobique Gram-négative *Vitreoscilla stercorari*, la Vhb semble impliquée dans la fixation et le transport de l’oxygène et/ou sa réduction en eau, puisque (i) que son expression est augmentée quand la bactérie est soumise à des conditions hypoxiques et (ii) la Vgb se lie à la sous-unité I du cytochrome b ubiquinol oxydase, où l’oxygène est réduit en eau (Figure 13) (Park et al. 2002).

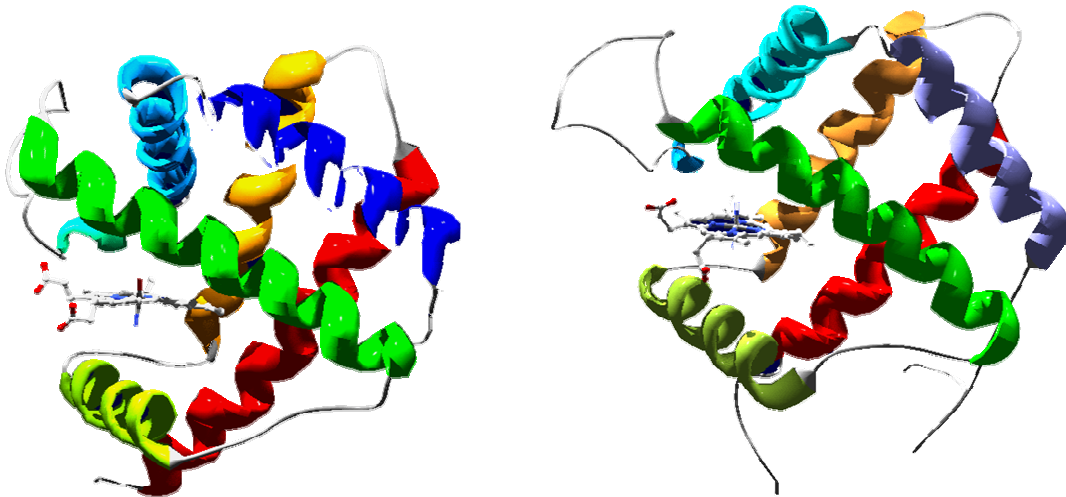


**Figure 13 :** Mécanisme d'interaction possible entre la Vhb et la sous-unité I du cytochrome b ubiquinol oxydase chez *Vitreoscilla stercorari*. D'après Park et al. (Park et al. 2002).

Chez les bactéries *Campylobacter jejuni* et *Campylobacter coli*, les SDgbs ne semblent pas être induites par une limitation de la concentration en oxygène dans le milieu ni intervenir dans la protection contre les ions superoxydes ou peroxydes, mais plutôt jouer un rôle clé dans la détoxification du NO et composés dérivés (Elvers et al. 2004).

### c) Les globines Symbiotiques et Non Symbiotiques de plantes.

Des globines ont été tout d'abord décrites dans les plantes légumineuses (telles que le lupin, *Lupinus luteus* ou le soja, *Glycine max*), capables d'utiliser l'azote atmosphérique comme source d'azote pour leur croissance grâce à leur association symbiotique avec les bactéries du genre rhizobium qui fixent cet azote atmosphérique. Ces bactéries sont trouvées dans un organe appelé nodosité symbiotique, et sont appelées bactéroïdes. Les globines symbiotiques (SHbs) sont exprimées en forte concentration dans les nodules de ces plantes, où elles ont deux fonctions : faciliter la diffusion de l'oxygène jusqu'aux bactéroïdes (Appleby 1992) et maintenir une faible concentration en oxygène indispensable pour la fixation efficace de l'azote sur les nitrogénases bactériennes sensibles à l'oxygène (Ott et al. 2005).

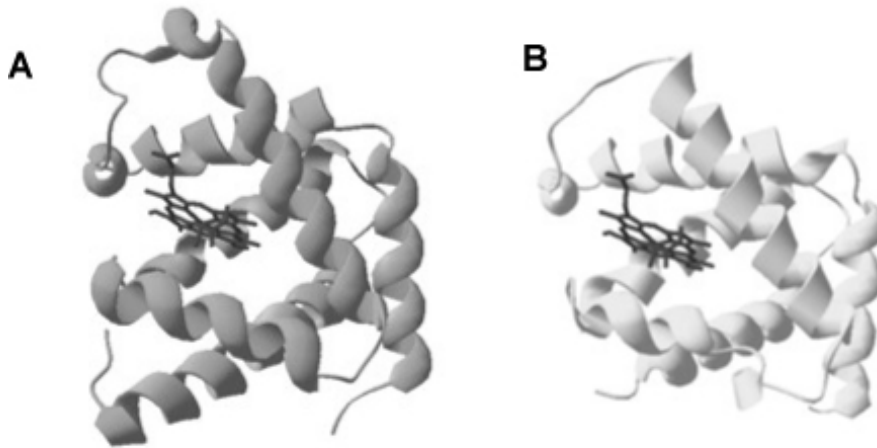


**Figure 14 :** A gauche : Représentation 3D de la SHbs de *Lupinus luteus* (PDB 2GDM). A droite : Représentation 3D de la NsHbs de *Oryza sativa* (PDB 1D8U).

D'autres globines ont été identifiées chez des plantes non-symbiotiques (*Arabidopsis thaliana* ou *Casuarina glauca*). L'expression de ces globines non-symbiotiques (NsHbs) est augmentée au sein de la cellule en réponse à différents stimuli environnementaux (Trevaskis et al. 1997), telles que de faibles concentrations en oxygène (Kundu, Trent and Hargrove 2003) ou une augmentation de la concentration en sucrose (alpha-D-glucopyranosyl-beta-D-fructofuranoside) (Hunt et al. 2001). Des travaux récents ont mis en évidence le rôle des NsHbs dans des voies métaboliques impliquant le monoxyde d'azote (NO), fournissant ainsi un type de respiration alternatif à la chaîne de transport des électrons en conditions hypoxiques (Igamberdiev et al. 2005).

#### d) Les Neuroglobines.

Les Ngbs ont reçu leur nom en accord avec leur expression prédominante dans le tissu nerveux (Burmester et al. 2000). La structure cristallographique de la Ngb humaine révèle la structure caractéristique en « globin-fold » et la présence d'une cavité au sein de la protéine n'existant chez aucune autre globine, et qui pourrait influencer la diffusion du ligand au sein de la molécule (Figure 15) (Pesce et al. 2003).



**Figure 15 :** Représentations 3D de la Mb humaine (A) et de la Ng humaine (B). D'après Burmester et al. (Burmester and Hankeln 2004).

Une fonction de liaison et de transport de l'oxygène, même si elle n'a pas encore été clairement démontrée, est fortement envisagée. On pense que des cellules métaboliquement actives et demandeuses de grande quantité d'oxygène, comme les neurones, peuvent exprimer une protéine spécialisée qui participerait au maintien du métabolisme aérobie en stockant l'oxygène, et ceci même en conditions d'hypoxie aiguë (Burmester et al. 2000; Hankeln et al. 2005). Dans le cas de cellules photoréceptrices de rétines de vertébrés exprimant également des neuroglobines, une relation a été mise en évidence entre le taux d'expression des Ngbs et l'activité métabolique de ces cellules: la localisation cellulaire des Ngbs correspond parfaitement avec les zones de consommation maximale d'oxygène (Schmidt et al. 2003).

Par ailleurs, l'effet neuroprotecteur des Ngbs en cas d'ischémie et de reperfusion des tissus cérébraux laisse supposer que ces protéines pourraient agir comme fixateur des espèces réactives de l'oxygène (ROS) ainsi que des dérivés nitrogénés (Sun et al. 2001; Sun et al. 2003; Fago et al. 2004).

Des globines sont exprimées dans le tissu nerveux chez des annélides (*Aphrodite aculeata*) (Dewilde et al. 1996), des mollusques (*Tellina alternata*) et nématodes (Geuens et al. 2004). Chez les mollusques, les Ngbs joueraient un rôle significatif dans le maintien de l'activité neuronale en cas de stress hypoxique (Kraus and Colacino 1986).

*e) Les Cytoglobines.*

Les cytoglobines (Cygbs) sont exprimées dans tous les tissus de mammifères (Kawada et al. 2001; Burmester et al. 2002; Trent and Hargrove 2002). Elles possèdent environ 190 acides aminés, présentant ainsi une vingtaine d'acides aminés supplémentaires aux deux extrémités N et C-terminales, par rapport aux Hbs 3/3 d'environ 150 acides aminés. Ces extensions pourraient intervenir dans des interactions protéine-protéine particulières (Hankeln et al. 2005).

Les Cygbs ont tout d'abord été décrites comme des protéines fortement exprimées dans les cellules étoilées hépatiques de type fibroblaste, activées à la suite d'une fibrose du foie (Kawada et al. 2001). Par ailleurs, des travaux montrent que les Cygbs sont aussi exprimées dans les cellules étoilées des tissus fibreux du pancréas lors d'une pancréatite et dans les cellules de type fibroblaste des régions nécrotiques du rein après une néphropathie chronique suite à un jeûne prolongé (Kawada et al. 2001; Nakatani et al. 2004). Ces résultats suggèrent que la fonction majeure des Cygbs s'exerce au sein des fibroblastes de différents organes. Ces cellules sont connues non pas pour avoir un métabolisme important, mais plutôt pour permettre la production massive de protéines de matrice extracellulaire, telle que le collagène. Or, la synthèse de collagène est consommatrice d'oxygène au moment de l'hydroxylation des résidus prolines par les prolyl-hydroxylases. De plus, (i) l'expression des Cygbs -comme le collagène- est augmentée en cas d'hypoxie (Schmidt et al. 2004) et (ii) une surexpression de Cygb entraîne une synthèse de collagène (Nakatani et al. 2004). Tout ceci laisse supposer que les Cygbs soit sont directement impliquées dans l'approvisionnement en oxygène des prolyl-hydroxylases, soit participent à une voie métabolique inconnue qui aboutirait à l'augmentation de la synthèse de collagène (Hankeln et al. 2005). Enfin, les Cygbs ont été identifiées dans le noyau de certaines cellules de vertébrés, où elles pourraient être impliquées dans l'interaction avec d'autres protéines jouant un rôle dans la régulation de la transcription suite à un stimulus (Geuens et al. 2003).

*f) Les Globines X.*

Un nouveau type de globine a récemment été décrit chez les poissons et les amphibiens, et n'a pour l'instant pas été identifié chez les amniotes (Roesner et al. 2005). Des travaux préliminaires ont montré que les globines X (GbX) sont exprimées dans beaucoup de tissus (muscles, cœur, rein, foie, rate, branchies et intestin), mais pas dans le cerveau ni dans

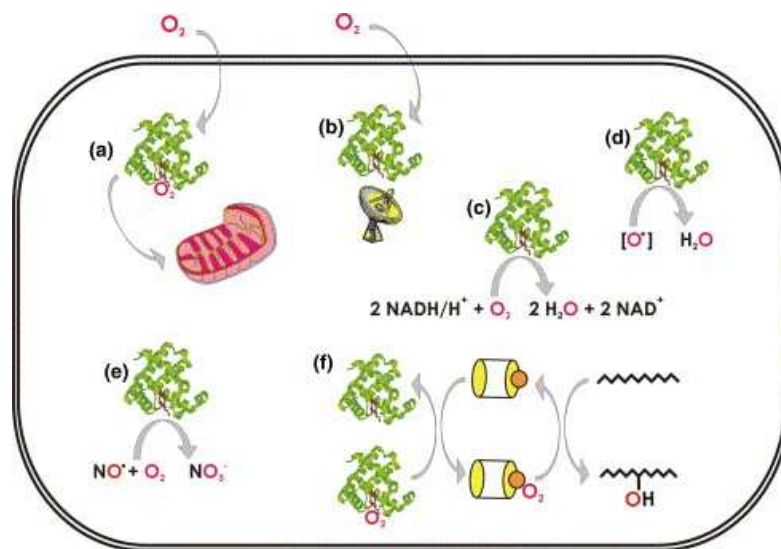


les yeux (Roesner et al. 2005). Les fonctions possibles des GbX sont pour l'instant toujours inconnues.

#### D- Conclusion.

L'accumulation des séquences issues de la génomique a permis la découverte de nouveaux types de globines, et ceci dans des organismes où leur présence n'était pas soupçonnée, tels que chez l'arthropode *Drosophila melanogaster* (Burmester and Hankeln 1999), le nématode *Caenorhabditis elegans* (Hoogewijs et al. 2004) ou l'urochordé *Ciona intestinalis* (Ebner, Burmester and Hankeln 2003). La figure 16 résume les fonctions connues à l'heure actuelle pour les globines intracellulaires circulantes et non circulantes.

La structure des hémoglobines extracellulaires d'Annélides ainsi que leurs fonctions seront développées dans le chapitre suivant.



**Figure 16** : Résumé des différentes fonctions possibles des globines intracellulaires circulantes et non circulantes. D'après Hankeln et al. (Hankeln et al. 2005).

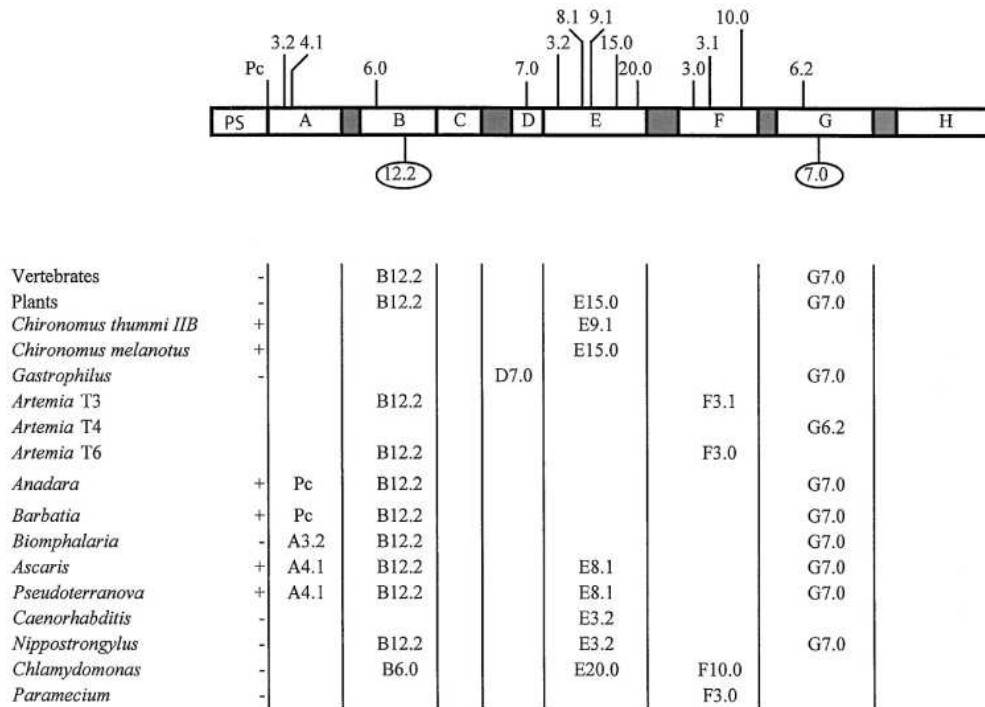
- (a) Rôle dans la liaison, le stockage et/ou la diffusion de l'oxygène vers les mitochondries.
- (b) Rôle chimiotactile, ou implication dans des voies de signalisation intracellulaires.
- (c) Rôle de terminale oxydases, permettant la régénération de NAD<sup>+</sup> pour faciliter la glycolyse et augmenter la production d'ATP en conditions d'hypoxie.
- (d) Rôle dans la détoxification par liaison des espèces réactives de l'oxygène ou de l'azote.
- (e) Rôle de nitrate dioxygénase, permettant la conversion des excès de monoxyde d'azote (NO) en nitrate.
- (f) Rôle dans l'apport direct de molécules de O<sub>2</sub> aux enzymes cytoplasmiques nécessitant du O<sub>2</sub> pour les réactions chimiques qu'elles catalysent.

#### **IV Evolution des globines au sein de l'arbre universel du vivant.**

##### A- Structure des gènes codant pour les globines.

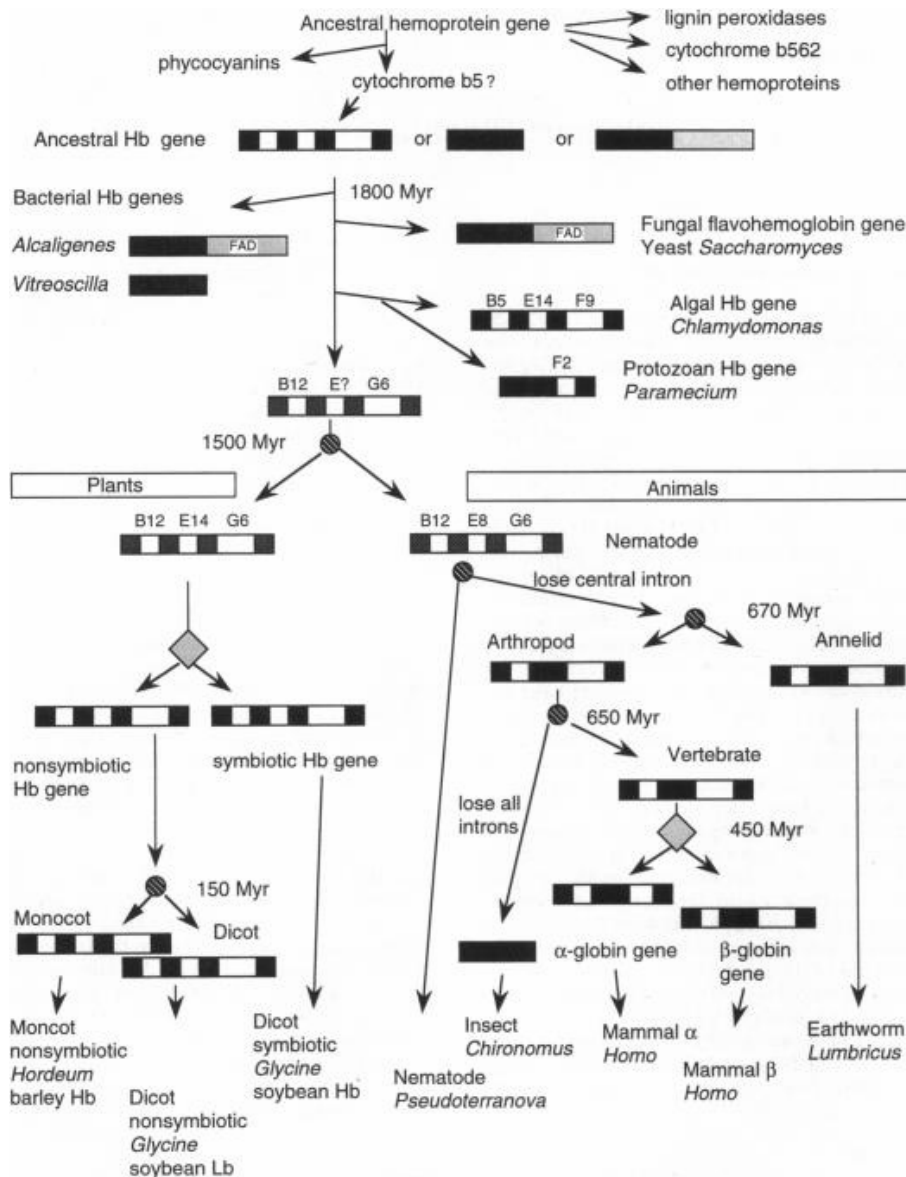
Historiquement, l'accès aux séquences nucléotidiques et particulièrement à l'ADN génomique des organismes, a permis de mettre en évidence la structure « en mosaïque » des gènes eucaryotes. Ceux-ci sont composés de séquences non codantes (introns) insérées entre les séquences codantes (exons) (Breathnach, Mandel and Chambon 1977; Jeffreys and Flavell 1977). Ces découvertes amenèrent de nombreuses questions, notamment celle de savoir quel pouvait être le rôle de ces séquences introniques ou quelle était la relation existant entre la structure du gène, les domaines fonctionnels et la fonction de la protéine (Craik, Buchman and Beychok 1980). La structure des gènes codant pour les globines de vertébrés fut d'abord décrite : ils présentaient une structure de 3 exons/ 2 introns, les positions d'introns correspondant à des insertions au niveau des hélices B et G de la protéine. Ces gènes servirent de modèle pour tester l'hypothèse selon laquelle les exons correspondaient à des domaines fonctionnels de la protéine (Gilbert 1978). Plusieurs études permirent par la suite de déterminer que la séquence peptidique correspondant au seul exon central des gènes des globines (c'est-à-dire le domaine de fixation de l'hème) n'était pas suffisante pour assurer la fonction de liaison à l'oxygène (Craik, Buchman and Beychok 1980; Craik, Buchman and Beychok 1981), remettant ainsi en question l'hypothèse d'une correspondance stricte entre un exon et un domaine fonctionnel.

La présence et la position des introns furent aussi utilisées pour comprendre l'émergence et l'évolution des gènes de globines (Figure 17). Les hypothèses portant sur l'origine des introns amenèrent à l'élaboration de plusieurs théories contradictoires : la théorie « intron-early » (Gilbert 1987; Roy 2003), selon laquelle les premiers introns auraient servi à la mise en place des premiers gènes, la théorie « intron-late » (Palmer and Logsdon 1991; Kersanach et al. 1994), selon laquelle les introns auraient « été rajoutés » au cours de l'évolution pour scinder les gènes contigus originaux, et une théorie plus récente qui suggère qu'une partie des introns étaient déjà présente dans l'ancêtre commun primitif, alors que d'autres introns ont été ajoutés au cours de l'évolution, accordant ainsi les deux premières théories (De Souza et al. 1998).



**Figure 17:** Diversité des positions d'introns au sein des gènes de globines. Les hélices sont désignées de A à H. PS, peptide signal, PC, intron précédant la séquence codante. + et - indiquent la présence ou non d'un peptide signal. B12.2 correspond ainsi à une insertion après le 2<sup>e</sup> nucléotide du 12<sup>e</sup> codon de l'hélice B. D'après Dewilde et al. (Dewilde et al. 1998).

A la suite de ses travaux sur la relation possible entre les exons et les unités structurales de l'hémoglobine, Go suggéra que l'exon central devait être le résultat de la fusion de deux exons à la suite de la perte d'un intron (Go 1981). Cette hypothèse fut confirmée par la découverte dans les gènes codant pour les globines de plantes, d'un troisième intron central, correspondant à une insertion au niveau de l'hélice E (Jensen et al. 1981; Andersson et al. 1996). Par la suite, la présence des deux introns correspondant aux positions conservées dans les hélices B et G fut décrite dans de nombreux gènes de globines trouvées dans des espèces appartenant à différentes lignées évolutives. La mise en évidence de ces deux introns ainsi que d'un troisième intron central, en position différente de celui trouvé chez les plantes, dans le gène de globine des nématodes *Pseudoterranova* et *Ascaris suum* (Dixon and Pohajdak 1992; Sherman et al. 1992), suggéra l'existence d'un ancêtre commun, présentant une structure 4 exons/ 3 introns, aux gènes codant pour les globines de plantes et d'animaux (Figure 18) (Hardison 1996).



**Figure 18 :** Evolution des gènes codant pour les globines. Les introns sont représentés en blanc, les exons en noir. D’après Hardison (Hardison 1996).

### B- Ubiquité des globines au sein de l’arbre universel.

Les globines au sens large regroupent toutes les globines appartenant aux deux grands groupes décrits précédemment, c’est-à-dire les Hbs 2/2 et les Hbs 3/3. L’existence de globines bien que structurellement et fonctionnellement différentes au sein des trois grands domaines du vivant suggère que ces protéines ont évolué à partir d’un ancêtre commun, qui possédait un

domaine de type « globin-fold » et qui aurait pu agir comme protéine redox (Moens et al. 1996; Vinogradov et al. 2005). En se basant sur les similarités de séquences entre les globines et les cytochromes *b2* et *b5*, Runnegar a émis l'hypothèse que ces protéines auraient pu évoluer à partir d'un précurseur de cytochrome *b*-like (Runnegar 1984). La découverte récente des globines 2/2 a relancé le débat sur l'histoire évolutive et l'origine de la superfamille des globines.

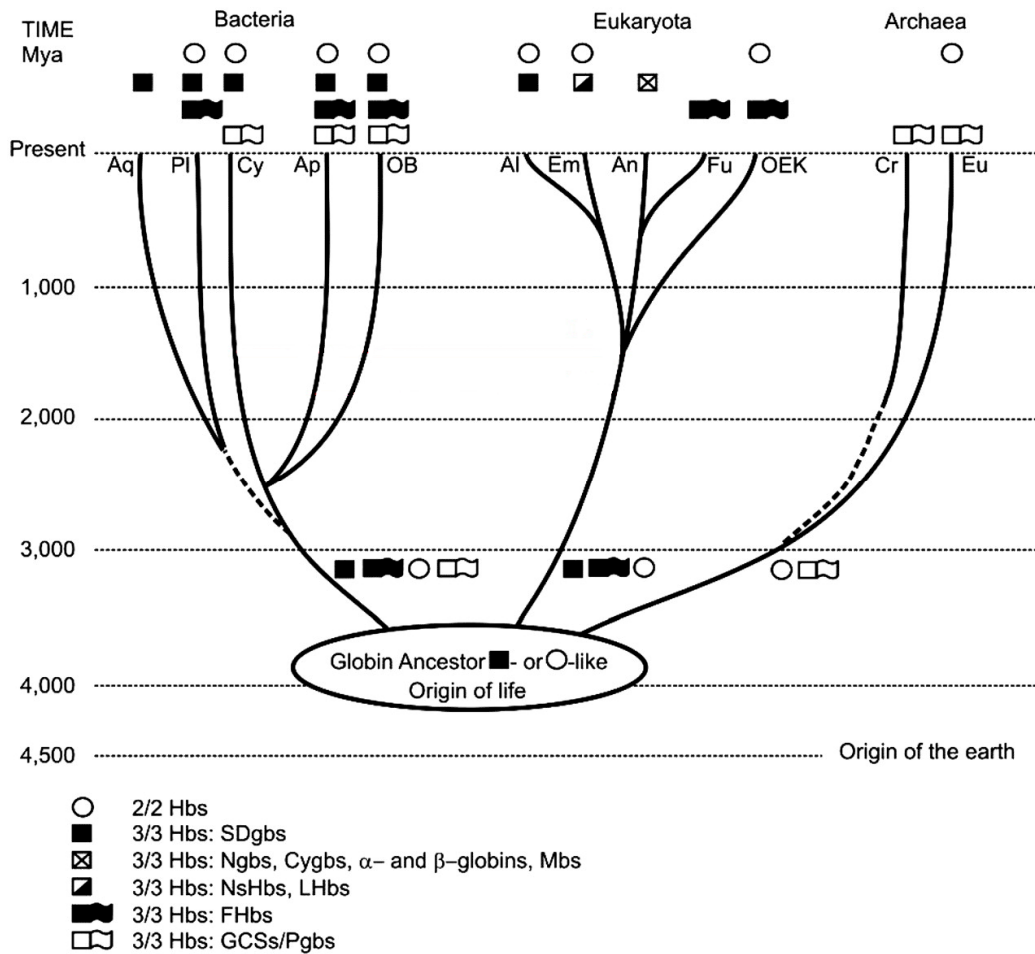
Les globines 2/2 et 3/3 sont toutes les deux retrouvées au sein des trois grands domaines du vivant (Vinogradov et al. 2005). Les informations connues à l'heure actuelle ne permettent pas de savoir quelle structure, entre celle des globines 2/2 et celle des globines 3/3, est la plus ancestrale.

La figure 19 illustre l'émergence des différentes lignées de globines au sein des Archaea, Bacteria et Eucaryota.

Pour expliquer l'émergence des FHbs et des GCS, deux événements indépendants ont du avoir lieu, impliquant la fusion de gènes codant pour une SDgbs-like - possédant une structure 3/3 - avec des gènes codant pour des protéines enzymatiques : d'une part des précurseurs des domaines FAD pour les FHbs, et d'autre part des précurseurs des domaines de signalisation intracellulaire actuels pour les GCS (Vinogradov et al. 2005).

On estime que la divergence entre les Archaea et les Eucaryota s'est produite il y a 4 milliards d'années, et celle entre les Eucaryota et les Bacteria il y a 2.7 milliards d'années (Hedges 2002). La présence respective des FHbs chez les Bacteria et Eucaryota et des GCS chez les Archaea et Bacteria suggère que ces recombinaisons génétiques ont eu lieu avant l'émergence de ces trois domaines : Archaea, Bacteria et Eucaryota. Par conséquent, les GCS et Pgbs doivent avoir été perdues chez les Eucaryota, de même que les FHbs et Sdgs chez les Archaea (Vinogradov et al. 2005).

Des duplications successives au sein de chaque lignée permettraient d'expliquer la grande diversité actuelle des globines. De même, la présence de FHbs et de Hbs 2/2 dans certaines lignées d'Eucaryota pourrait résulter de plusieurs transferts horizontaux de gènes (Vinogradov et al. 2005).



**Figure 19 :** Distribution des trois lignées de globines au sein des trois domaines du vivant (Archaea, Bacteria et Eukaryota). Les lignes pointillées indiquent des évènements dont la date est incertaine. Aq, Aquificales, Cy, Cyanobactérie, Ap,  $\alpha$ -protéobactérie, OB, autres bactéries, Al, Algues, Em, Embryophyte, An, Animaux, Fu, Champignons, OEK, Autres Eucaryotes, Cr, Crenarchaeota, Eu, Euryarchaeota. D'après Vinogradov et al. (Vinogradov et al. 2005).

### V Conclusion : la globine, quelle(s) fonction(s) ancestrale(s) ?

A la suite de ses travaux sur les propriétés de liaison du monoxyde d'azote par les hémoglobines, Bonaventura posait déjà le problème de la fonction universelle et ubiquitaire de transport de l'oxygène des globines en écrivant ironiquement : « l'hémoglobine fixe **également** l'oxygène ! » (Bonaventura and Lance 2001).

Les derniers travaux en date compilant l'ensemble des données fonctionnelles et structurales sur les Gbs du vivant tendent vers un nouveau paradigme concernant l'évolution des globines : Les Gbs auraient initialement été des protéines enzymatiques dont la fonction de transport de l'oxygène et du monoxyde d'azote au sein du domaine Eucaryota serait une

fonction dérivée, en d'autres termes acquise plus tardivement au cours de l'évolution. Vinogradov écrit et propose ainsi: « **Because of the original enzymatic functions, greatly expanded and diversified, remain the predominant functions of most extant globins, O<sub>2</sub> and NO transport should be viewed as specialized developments that accompanied the emergence and evolution of metazoans subsequent to the rise in atmospheric O<sub>2</sub> content to present-day levels.** » (Vinogradov et al. 2005).

## **Chapitre 2**

### **Les globines d'annélides**



## Chapitre 2 : LES GLOBINES D'ANNELIDES.

Les hémoglobines d'invertébrés présentent une plus grande diversité au niveau des structures quaternaires que les hémoglobines des autres eucaryotes : on trouve des Hbs allant du monomère au polymère géant composé de plusieurs sous-unités, ainsi que des polypeptides possédant plusieurs domaines de type globin-fold reliés entre eux par de courtes séquences, appelés multi-domaines. Les annélides (mais aussi les arthropodes et les mollusques) expriment des globines au sein de différents tissus ou types cellulaires (décrits dans le chapitre précédent), mais aussi directement en solution dans le compartiment sanguin ou cœlomique : elles sont dites extracellulaires.

La figure 20 illustre la diversité de structure quaternaire mise en évidence pour les globines intracellulaires ou extracellulaires.

Intracellular Hbs			Mw (kD)
Monomeric	○	Mb, nerve & coelomic Hb, <i>Chironomus</i> *	17
Dimeric	∞	Bivalve RCs, <i>Chironomus</i> *	34
Tetrameric	⊕	Bivalve RCs, Insect tracheal cells Vertebrate RCs	68
Polymeric	⊗	<i>Glycera dibranchiata</i>	
Multi-subunit Hbs			
Hexagonal bilayer 12 dodecamers		<b>Annelida, Vestimentifera</b> 3 trimers, 3 monomers, 3 linkers (144 heme-chains, 36 linkers) <i>Lumbricus, Arenicola, Riftia</i>	3600
2 dodecamers		<b>Vestimentifera, Pogonophora</b>	~400
Pentagonal bilayer 10 dodecamers		<b>Pulmonate Molluscs</b> 5 dimers <i>Helisoma, Planorbis, Biomphalaria</i>	1700
Multi-domain, Multi-subunit Hbs			
2 or 3 subassemblies		<b>Polychaeta</b> 4-domain subunits <i>Branchiopolynoe</i>	115 or 174
Quadrangular bilayer		<b>Nematoda</b> 8 (2-domain) subunits <i>Ascaris, Parascaris</i>	328
Penta-/Octagonal bilayers		<b>Branchiopod crustaceans</b>	
	2 (9-domain) subunits	<i>Artemia</i> (Anostracan)	250
	10 (2-domain) "	<i>Cyzicus</i> (Conchostracan)	300
	16 (2-domain) "	<i>Daphnia</i> (Cladoceran)	490
	24 (2-domain) "	<i>Lepidurus</i> (Notostracan)	800
"Rods" 14 - 24 domains		<b>Bivalve molluscs</b> <i>Astarte</i> <i>Cardita</i>	800 - 12000

\*Extracellular in *Chironomus* larvae

**Figure 20:** Diversité des globines d'invertébrés. D'après Weber et Vinogradov (Weber and Vinogradov 2001).

Les globines d'annélides décrites jusqu'à présent sont toutes de type Hbs 3/3. Elles peuvent être classées en trois catégories distinctes en fonction de leur localisation tissulaire : les globines intracellulaires non circulantes (de type myoglobine), les globines intracellulaires circulantes et les globines extracellulaires. Trois familles d'annélides : les Terebellidae, les Alvinellidae et les Opheliidae co-expriment les trois types de globines (Weber 1978 ; Hourdez et al. 2000).

## **I les globines intracellulaires.**

### **A- Les globines intracellulaires non circulantes (ou Myoglobines).**

On connaît relativement peu de choses sur les myoglobines d'Annélides. Ces protéines sont trouvées sous forme de monomère, dans le cas de *Arenicola marina* (Weber and Pauptit 1972) ou de *Glycera robusta* (Terwilliger, Garlick and Terwilliger 1976), et très rarement sous forme de dimères, dans le cas de *Travisia foetida* (Terwilliger, Garlick and Terwilliger 1980) ou de *Aphrodite aculeata* (Wittenberg, Briehl and Wittenberg 1965). De même que chez les vertébrés, les Mbs d'annélides ont une affinité pour l'oxygène intermédiaire entre celle des globines circulantes (intracellulaires ou extracellulaires) et la cytochrome oxydase mitochondriale, ce qui suggère qu'elles pourraient agir comme « relais » au niveau du transfert de l'oxygène (Weber and Vinogradov 2001).

L'espèce la plus étudiée est *Arenicola marina*, qui exprime dans la paroi du corps cinq isoformes de myoglobines de deux types différents : MbI et MbII (Kleinschmidt and Weber 1998), possédant une affinité pour l'oxygène sensiblement différentes (Weber and Pauptit 1972; Kreutzer and Jue 1996). Ceci suggérerait que le transfert d'oxygène depuis le compartiment vasculaire jusqu'aux mitochondries ne se ferait pas par l'intermédiaire d'un seul composant, mais grâce à la succession de plusieurs protéines possédant une affinité croissante pour l'oxygène (Weber and Pauptit 1972). Récemment, la Mb du polychaete *Ophelia bicornis* a été purifiée. Ses caractéristiques fonctionnelles, très proches de celles de *Arenicola marina* ou *Glycera dibranchiata*, suggère une capacité de liaison et de stockage de l'oxygène (Sanna et al. 2005).

Une myoglobine homodimérique, assez particulière, a été trouvée chez le polychaete *Aphrodite aculeata* (Wittenberg, Briehl and Wittenberg 1965; Dewilde et al. 1996). Son expression dans le système nerveux lui a valu son nom de Mb neurale ou « Neuro-myoglobine » (Wittenberg, Briehl and Wittenberg 1965). Des travaux supplémentaires ont

montré que cette protéine est en fait trouvée dans un certain nombre d'autres tissus dont les muscles, le pharynx et l'intestin, et que sa fonction est très semblable à celles des autres myoglobines, suggérant fortement que la protéine exprimée chez *Aphrodire aculeata* est bien une myoglobine (Dewilde et al. 1996).

A noter : une protéine possédant une activité déhaloperoxydase a été décrite chez le polychaete *Amphitrite ornata* (Chen et al. 1996; Lebioda et al. 1999). Cet enzyme permet de déhalogéner les dérivés bromophénols sécrétés par d'autres polychaetes co-habitant avec *Amphitrite ornata* dans les zones de sédiments sablo-vaseux, tels que *Thelepus crispus* et *Notomastus lobatus* (Lebioda et al. 1999). La structure cristallographique de la déhaloperoxydase montre qu'elle possède la structure caractéristique en « globin-fold », et sa séquence peptidique présente une grande similarité avec la myoglobine du mollusque *Aplysia limacina* (LaCount et al. 2000). La localisation tissulaire de la déhaloperoxydase (intracellulaire ou extracellulaire) reste inconnue.

#### B- Les globines intracellulaires circulantes.

Parmi les Annélides, la présence de cellules circulant dans la cavité coelomique est restreinte à certaines familles de polychaete : les Capitellidae, Glyceridae, Alvinellidae, Ophelidae, Terebellidae et Scalibregmidae (Weber and Vinogradov 2001).

Les globines intracellulaires circulantes peuvent être monomériques (Terebellidae), dimériques (Capitellidae et les Ophelidae), tétramériques (Glyceridae et Scalibregmidae) ou avoir plusieurs autres états d'agrégation (Mangum 1985; Weber and Vinogradov 2001).

Le polychaete *Glycera dibranchiata* exprime deux types de globines intracellulaires dans des erythrocytes nucléés: une globine monomérique et une globine polymérique ayant plusieurs degrés d'assemblage (Hoffmann and Mangum 1970; Seamonds, Forster and Gottlieb 1971; Zafar et al. 1990b), formées à partir de différents monomères (Vinogradov, Machlik and Chao 1970). Ces deux globines sont synthétisées dans des quantités similaires, et regroupent chacune un certain nombre de chaînes polypeptidiques (Shafie et al. 1976; Kandler, Constantinidis and Satterlee 1985; Zafar et al. 1990b). Des travaux ont montré que la globine monomérique a une affinité pour l'oxygène légèrement supérieure à celle de la globine polymérique (Mizukami and Vinogradov 1972).

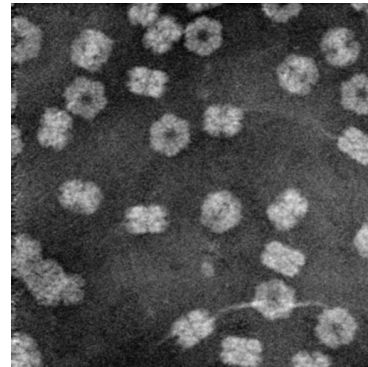
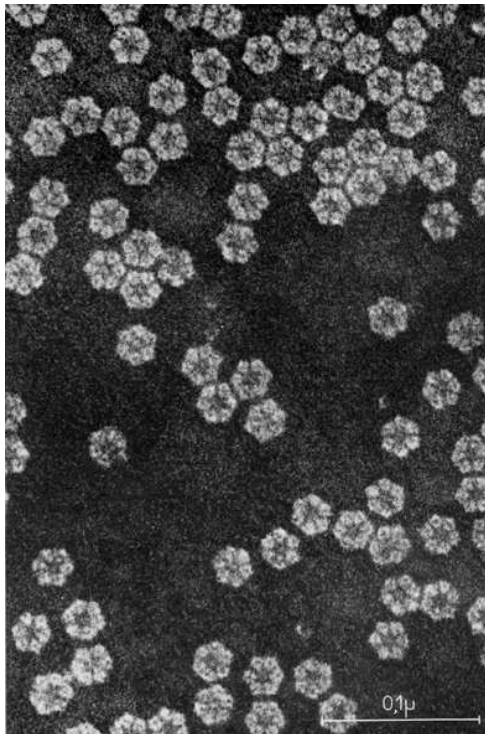
Plusieurs études ont permis de mettre en évidence la fonction des globines intracellulaires chez les Annélides (Weber 1978; Hourdez et al. 2000) qui expriment aussi de façon concomitante des myoglobines et des hémoglobines extracellulaires. Chez les annélides

*Neoamphitrite figulus*, *Pista pacifica*, *Terebella lapidaria*, *Amphitrite ornata* et *Alvinella pompejana*, l'affinité de l'hémoglobine extracellulaire pour l'oxygène est plus faible que celle de l'hémoglobine intracellulaire circulante, permettant ainsi un flot d'oxygène de l'environnement extérieur vers le système vasculaire, puis vers le système coelomique et enfin vers les myoglobines (Terwilliger 1974; Weber et al. 1977; Wells and Warren 1982; Hourdez et al. 2000).

## **II les globines extracellulaires.**

### **A- Présentation générale**

Les hémoglobines extracellulaires d'annélides sont caractérisées par un haut poids moléculaire, pouvant atteindre 4000 kDa (Lamy et al. 1996; Weber and Vinogradov 2001). Ces protéines sont hautement hétéropolymériques, et apparaissent composées, en microscopie électronique, de deux hexagones superposés (Figure 21), d'où le nom de structure quaternaire en bicouche hexagonale (**Hexagonal Bilayer hemoglobin : HBL-Hb**).



**Figure 21 :** Microscopies électroniques de l'hémoglobine extracellulaire de *Arenicola marina*. D'après Breton-Gorius (Breton-Gorius 1963).

Au sein des Annélides, les Siboglinidae (vestmentifères et pogonophores), tels que *Riftia pachyptila* ou *Oligobranchia mashikoi*, possèdent une hémoglobine extracellulaire de 400 kDa qui correspond à des degrés de polymérisation moindre (Yuasa et al. 1996; Zal et al. 1996a).

Par ailleurs, on trouve dans quatre familles de polychaete : les Sabellidae, les Serpulidae, les Chlorohaemidae et les Ampharetidae (Terwilliger 1992), une hémoglobine extracellulaire qui ne diffère des autres HBL-Hbs que par la nature du groupement prosthétique (l'hème): un des groupements vinyle est remplacé par un groupement formyle. Ces hémoglobines furent appelées chlorocruorines, mais les travaux de Suzuki ont montré que les globines des hémoglobines extracellulaires et des chlorocruorines sont en fait des protéines homologues (Suzuki, Hirao and Vinogradov 1995).

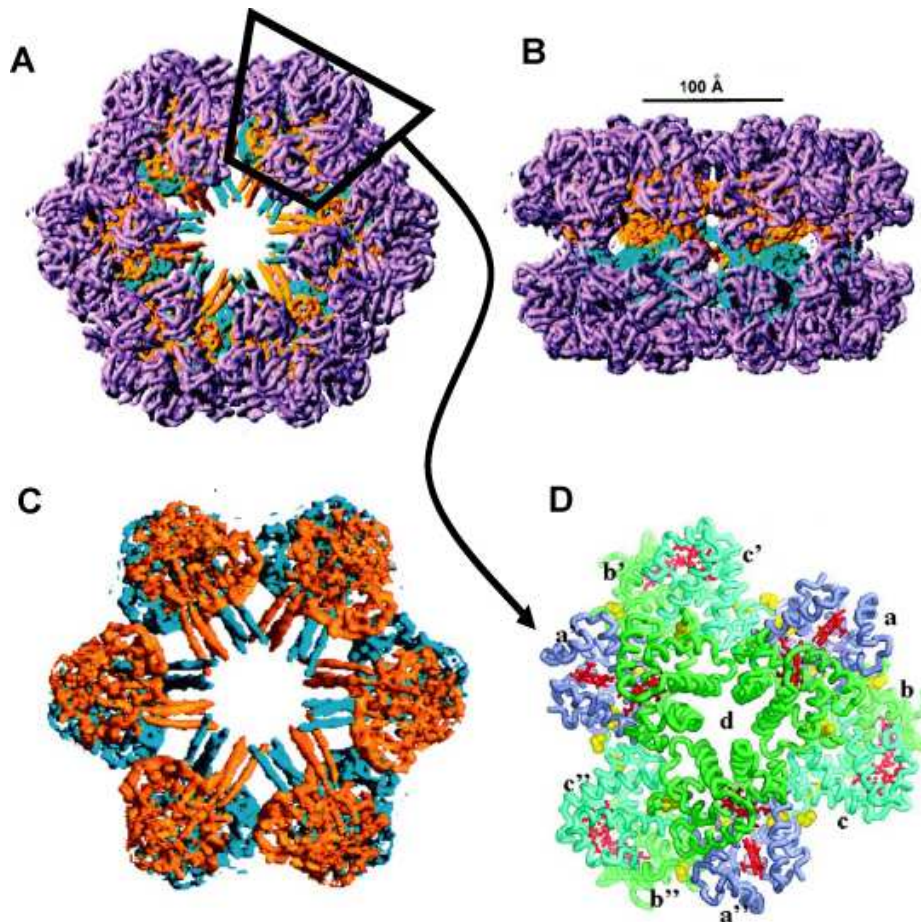
Des études ont révélé que les hémoglobines extracellulaires des Annélides sont synthétisées au niveau des cellules hématopoïétiques, puis excrétées dans le système vasculaire (Breton-Gorius 1963; Boilly and Sean 1979). Pourtant, les mécanismes permettant l'assemblage *in vivo* de telles protéines, que ce soit par des phénomènes d'assemblage spontané ou grâce à l'intervention de protéines chaperonnes, sont encore inconnus.

#### B- Structure et composition des HBL-Hbs.

Les nombreux travaux réalisés sur les HBL-Hb ont montré que ces protéines sont composées par deux types de chaînes polypeptidiques : **les globines**, qui possèdent la structure caractéristique de type globin-fold, et **les linkers**, qui jouent un rôle dans le maintien de la structure quaternaire (Shlom and Vinogradov 1973; Vinogradov 1985; Lamy et al. 1996). Les globines et les linkers ne possèdent ni les mêmes structures primaires et tertiaires, ni les mêmes propriétés fonctionnelles.

La structure la plus étudiée à ce jour est celle de l'HBL-Hb de *Lumbricus terrestris*. Cette protéine est constituée de 144 chaînes de globines et de 36 chaînes de linkers. Les globines sont de quatre types différents (a, b, c, d) et peuvent s'assembler en trimère (abc) ou rester sous forme monomérique (d). La formation des trimères est réalisée par des ponts-disulfures inter-chaînes. Les trimères et monomères de globines s'assemblent ensuite pour former un dodécamère, qui correspond au 1/12<sup>ème</sup> de la structure en bicouche hexagonale (Figure 22D). La structure cristallographique de l'HBL-Hb de l'oligochaete *Lumbricus terrestris* a été déterminée avec une résolution à 5.5Å (Royer et al. 2000). La figure 24 permet de voir les positions exactes des globines et des linkers au sein de l'HBL-Hb (Royer et al. 2000). Plus récemment, la structure cristallographique d'un des dodécamère a été obtenue, confirmant le modèle proposé pour la structure de cette HBL-Hb (Strand et al. 2004). Récemment, les structures des hémoglobines extracellulaires de 400kDa de *Riftia pachyptila*

et *Oligobranchia mashikoi* ont également été déterminées (Flores et al. 2005; Numoto et al. 2005).



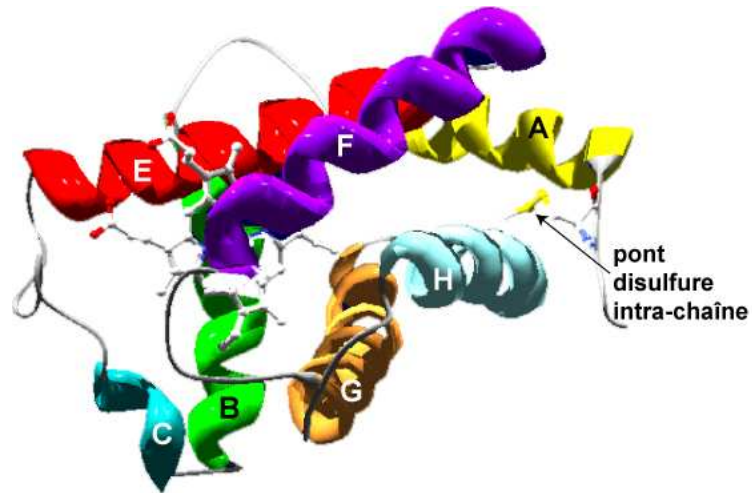
**Figure 22 :** Structure de l'HBL-Hb de *Lumbricus terrestris* (résolution à 5.5 Å). **A-** Vue de dessus de la protéine entière. **B-** Vue latérale de la protéine entière. **C-** Vue des linkers au centre de la protéine. Les globines sont représentées en violet et les linkers en orange et bleu. D'après Royer et al. (Royer et al. 2000). **D-** Agrandissement du dodécamère de *Lumbricus terrestris*. (Les trimères sont appelés abc, a'b'c' et a''b''c'' et les monomères d). D'après Strand et al. (Strand et al. 2004).

C- Les globines des HBL-Hbs.

1) Caractéristiques.

a) *Ponts disulfures intra-chaînes.*

Une caractéristique structurale singulière des globines extracellulaires d'annélides est la présence de deux cystéines conservées impliquées dans un pont disulfure intrachaîne (Figure 23) (Suzuki, Kapp and Gotoh 1988).



**Figure 23 :** Représentation 3D d'une des globines de *Lumbricus terrestris*. Le pont disulfure intrachaîne est désigné par une flèche (PDB 1X9F).

Ni les globines intracellulaires circulantes de *Glycera dibranchiata* (Imamura, Baldwin and Riggs 1972), ni les myoglobines de *Arenicola marina* (Kleinschmidt and Weber 1998), de *Aphrodite aculeata* (Dewilde et al. 1996) ou de *Ophelia bicornis* (Sanna et al. 2005) ne possèdent ces cystéines aux positions conservées capables de former un pont disulfure. Ces observations suggèrent que l'acquisition de ces cystéines pourrait être corrélée à la localisation extracellulaire de ces globines.

b) *Peptides signaux.*

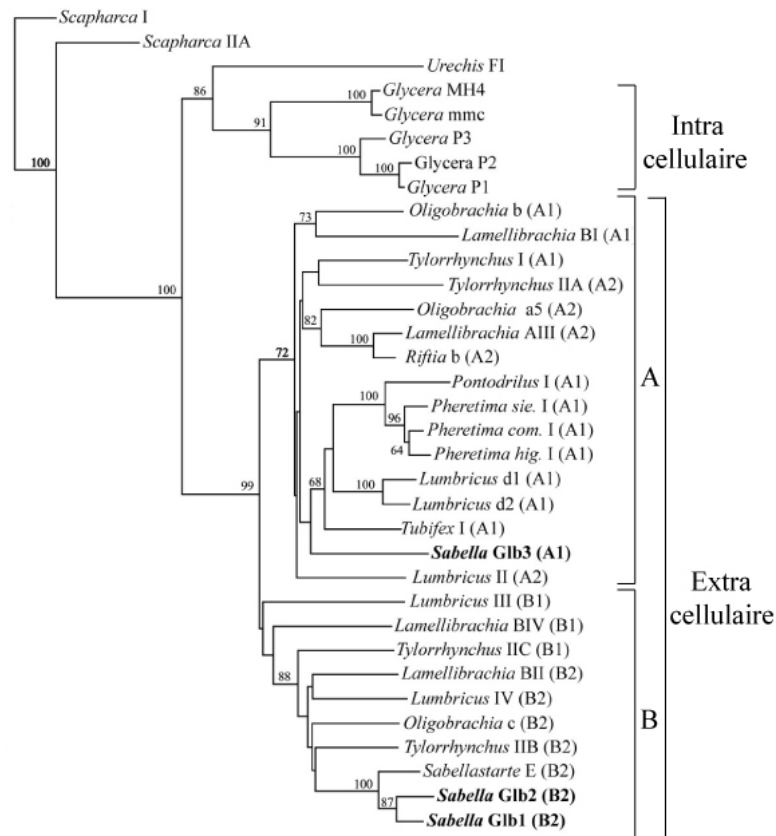
Une autre particularité des globines composant les HBL-Hb concerne la présence de peptide signaux permettant l'externalisation (adressage) des globines dans le milieu vasculaire ou coelomique. Des travaux ont suggéré que ces peptides signaux auraient été insérés au niveau des gènes codant pour les globines par des événements de recombinaisons génétiques à partir d'autres gènes codant pour des protéines sécrétées (Riggs 1991; Williams, Pal and Hurst 2000).

2) Evolution des familles multigéniques des globines extracellulaires.

a) *La famille multigénique.*

Les gènes codant pour les globines qui composent l'HBL-Hb forment une famille multigénique. L'étude des différentes séquences primaires des globines d'annélides a permis de mettre en évidence une organisation en deux familles A et B (Gotoh et al. 1987), subdivisées en quatre sous familles (A1, A2, B1 et B2) (Suzuki, Takagi and Ohta 1993). Cette organisation en quatre sous-familles clairement identifiées est toutefois à nuancer. Les premières études ont en effet été réalisées sur des globines provenant de polychaete et d'oligochaete (Gotoh et al. 1987; Suzuki, Takagi and Ohta 1993). Les phylogénies moléculaires réalisées après l'obtention de séquences protéiques de globines provenant de Siboglinidae (vestmentifères et pogonophores) et d'Hirudinae suggèrent que la classification en quatre sous-familles A1, A2, B1 et B2 pourrait ne pas refléter l'organisation réelle de ces familles multigéniques, suggérant ainsi une organisation plus complexe (Figure 24) (Negrisolo et al. 2001; Suzuki and Vinogradov 2003).





**Figure 24 :** Analyse phylogénétique des globines extracellulaires d'Annélides. D'après Negrisolo et al. (Negrisolo et al. 2001). Les abréviations entre parenthèse correspondent à la nomenclature proposée par Suzuki et al (Suzuki, Takagi and Ohta 1993).

b) Structure des gènes : positions des introns.

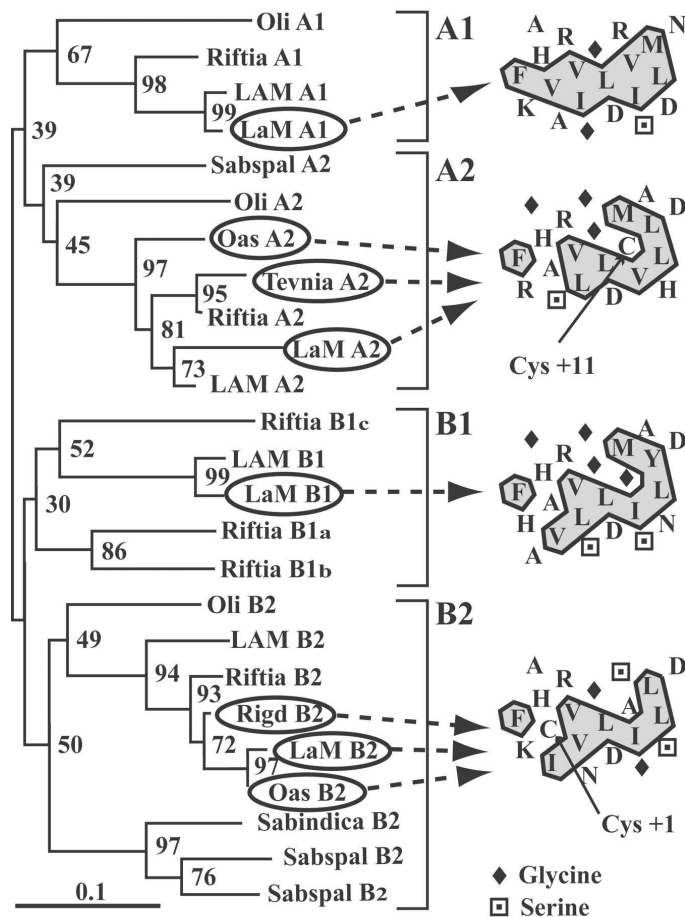
Les structures de gènes de globines extracellulaires d'annélides connues avant les travaux de cette thèse sont celles des globines B1 de *Lumbricus terrestris* (Jhiang and Riggs 1989), A1 de *Eudystilia vancouverii* (Dewilde et al. 2001) et A2 et B2 de *Riftia pachyptila* (Bailly 2003). Ces gènes possèdent tous deux introns, aux positions conservées B12.2 et G7.0 (c'est-à-dire respectivement après le 2ème nucléotide du 12ème codon dans l'hélice B et entre le 6ème et le 7ème codon dans l'hélice G), retrouvées aussi dans de nombreux gènes de globines d'autres organismes vivants (Hardison 1996). Ces résultats confirment l'existence d'un ancêtre commun à tous les gènes codant pour des globines extracellulaires.

### 3) Cystéines libres et fixation de l'hydrogène sulfuré.

Certaines globines présentent, en plus de ces cystéines impliquées dans la formation de ponts disulfures intra- ou inter-chaînes, des cystéines libres impliquées dans la liaison de l'hydrogène sulfuré (Zal 1996; Zal et al. 1998). La plus étudiée, l'hémoglobine de *Riftia pachyptila*, le vestimentifère des sources hydrothermales de la dorsale du Pacifique Est, possède en plus d'une forte affinité pour l'oxygène (Arp et al. 1990), une capacité à lier réversiblement l'hydrogène sulfuré (H<sub>2</sub>S) de façon indépendante de la liaison à l'oxygène (Arp, Childress and Vetter 1987). Les travaux de Suzuki sur l'hémoglobine du vestimentifère *Lamellibrachia sp.* le conduisirent à supposer que des cystéines libres présentes dans les globines étaient impliquées dans la liaison de l'hydrogène sulfuré (Suzuki, Takagi and Ohta 1990a). Par ailleurs, il a aussi proposé que les linkers possèderaient aussi la capacité de lier le sulfure d'hydrogène sur les groupements thiols présents dans la chaîne polypeptidique (Suzuki, Takagi and Ohta 1990b). Ces hypothèses furent par la suite confirmées par le blocage spécifique des résidus cystéines par la N-ethylmaleimide sur les globines de *R. pachyptila*, entraînant une perte de capacité à fixer le sulfure d'hydrogène (Zal et al. 1997c; Zal et al. 1998).

Les analyses des séquences primaires de globines de plusieurs annélides vivant dans des zones riches en hydrogène sulfuré, tels que *Riftia pachyptila* ou *Lamellibrachia sp.*, confirmèrent la présence de cystéines libres en positions conservées sur certaines chaînes de globines : les A2 et les B2. Les globines des annélides vivant dans des milieux oxydés, c'est-à-dire sans hydrogène sulfuré, tels que *Lumbricus terrestris* et *Tylorrhynchus heterochaetus* ne possèdent pas cette capacité de liaison du H<sub>2</sub>S (Arp, Childress and Vetter 1987), et ne présentent pas non plus les cystéines conservées au niveau de leurs séquences primaires (Garlick and Riggs 1982; Suzuki, Takagi and Gotoh 1982; Suzuki et al. 1985; Fushitani, Matsuura and Riggs 1988). Les cystéines sont trouvées dans toutes les globines A2 et B2 des annélides vivant dans des environnements riches en hydrogène sulfuré, aux positions Cys+1 et Cys+11, soit 1 acide aminé après l'histidine proximale conservée pour B2, soit 11 acides aminés après l'histidine proximale conservée pour A2 (Bailly et al. 2002). A l'exception de la globine A1 de *Oligobrachia mashikoi*, aucune globine A1 ou B1 ne possède ces résidus conservés, que ce soit chez les annélides vivant en milieux réduits (riches en hydrogène sulfuré) ou oxydés (sans hydrogène sulfuré). De plus, aucune globine A2 ou B2 provenant d'annélides vivant en milieux oxydés ne présente ces cystéines conservées (Bailly et al. 2002).

Bailly et al. ont également montré la conservation des structures 2D des domaines qui incluent les deux cystéines libres, à la fois entre les globines A2 et B2 des annélides vivant en milieux réduits, mais aussi avec les globines A2 et B2 d'annélides vivant en milieux oxydés, malgré l'absence des cystéines libres chez ces dernières (Figure 25) (Bailly et al. 2002).



**Figure 25** : Arbre phylogénétique réalisé à partir des séquences d'acides aminés des globines d'annélides vivant en milieux réduits. La structure 2D (Hydrophobic cluster analysis HCA) de la séquence flanquant les cystéines libres est représenté pour chaque sous-famille (A1/A2 et B1/B2). Les cystéines libres (Cys+1) et (Cys+11) sont désignées par une flèche. D'après Bailly et al. (Bailly et al. 2003).

La présence de ces cystéines libres en positions conservées ainsi que la conservation du domaine incluant ces cystéines suggèrent que ces cystéines libres étaient présentes dans l'ancêtre commun à toutes les globines extracellulaires d'annélides, et qu'elles ont été perdues

après duplications dans les lignées A1 et B1 et uniquement maintenues dans les globines A2 et B2 impliquées dans la liaison de l'hydrogène sulfuré (Figure 25) (Bailly et al. 2002).

Récemment, deux articles concernant les hémoglobines extracellulaires des deux annélides vivant en milieux réduits, *Riftia pachyptila* (Flores et al. 2005) et *Oligobranchia mashikoi* (Numoto et al. 2005) ont été publiés : la structure cristallographique d'une des hémoglobines extracellulaires de *Riftia pachyptila* a en effet révélé que la liaison de l'hydrogène sulfuré ne se ferait pas par l'intermédiaire des cystéines libres, mais plutôt grâce à des interactions entre l'hydrogène sulfuré et des ions zinc ( $Zn^{2+}$ ) (Flores et al. 2005). Parallèlement, la structure cristallographique de l'hémoglobine extracellulaire de *Oligobranchia mashikoi* (Numoto et al. 2005) n'a pas mis en évidence la présence d'ions  $Zn^{2+}$ . De plus, des expériences permettant de mesurer la quantité de méthylmercure fixé sur les cystéines libres, représentatifs de la liaison de l'hydrogène sulfuré, ont suggéré que les cystéines libres présentent dans l'hémoglobine extracellulaire de *Oligobranchia mashikoi* jouaient un rôle actif dans la liaison de l'hydrogène sulfuré (Numoto et al. 2005).

#### D- Les chaînes de structure ou Linkers.

##### 1) Les modules LDL-A.

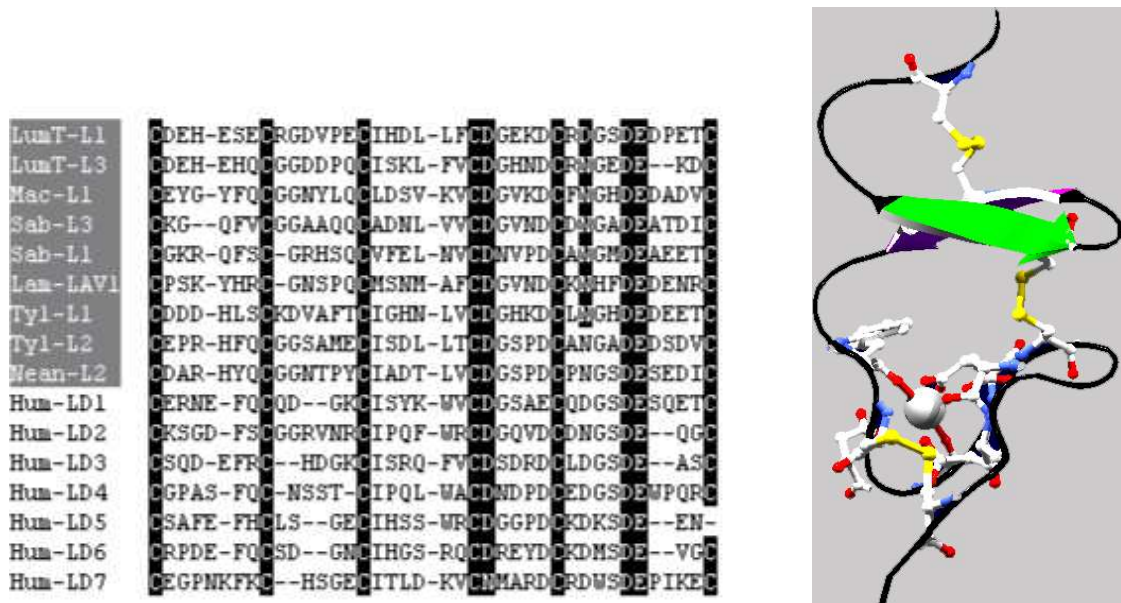
Les chaînes polypeptidiques impliquées dans le maintien de la structure quaternaire des HBL-Hb, sont appelées Linkers ou chaînes de structure. Ces protéines sont spécifiques du phylum des Annélides, puisqu'à l'heure actuelle, aucune protéine homologue aux linkers n'a été trouvée dans d'autres phyla.

A l'inverse des globines d'Annélides, peu de séquences primaires de linkers sont connues. Ces polypeptidiques sont composés de 185 à 225 acides aminés et présentent dans leur partie N-terminale un domaine particulier de 39 acides aminés, comprenant six cystéines invariables ainsi que plusieurs acides aminés conservés (Figure 28) (Suzuki and Riggs 1993). Ce domaine correspond exactement aux répétitions de classe A des récepteurs aux LDL (Low Density Lipoprotein) trouvées dans le domaine de liaison au ligand (Sudhof et al. 1985a; Suzuki and Riggs 1993). Ce domaine est commun à de nombreuses autres protéines de la superfamille des LDLR telles que les Very Low Density Lipoprotein receptor (VLDLR) (Takahashi et al. 1992), les LDLR- Related Protein (LRP), ainsi que dans d'autres protéines aux fonctions plus diverses, comme les facteurs de complément C9 or C8 (DiScipio et al.

1984), la glycoprotéine rénale gp330 (Raychowdhury et al. 1989; Saito et al. 1994) ou une entérokinase (Kitamoto et al. 1994).

De nombreuses études ont été réalisées sur les propriétés fonctionnelles et structurales des domaines LDL-A, notamment chez les LDLR. D'une part, des interactions ioniques ont été mises en évidence entre les acides aminés acides (ac. Glutamique et ac. Aspartique) des domaines LDL-A et certains acides aminés basiques du ligand (Sudhof et al. 1985b; Brown and Goldstein 1986; Guo et al. 2004). D'autre part, ces modules LDL-A lient du calcium ( $Ca^{2+}$ ), essentiel au maintien de la conformation structurale indispensable pour leur activité biologique (van Driel et al. 1987; Dirlam et al. 1996; Guo et al. 2004).

Bien que le  $Ca^{2+}$  semble aussi avoir un rôle déterminant dans la mise en place d'une conformation permettant la liaison entre les linkers et les dodécamères de globines (Kuchumov et al. 1999; Kuchumov, Loo and Vinogradov 2000), le rôle exact de ce domaine au sein des linkers reste à caractériser davantage (Fig. 26).



**Figure 26 :** A gauche : Alignement des domaines LDL-A des linkers d'annélides (grisé) et des 7 domaines LDL-A du LDLR humain. A droite : Représentation 3D du cinquième module LDL-A du domaine de liaison au ligand du LDLR humain (1AJJ). Les ponts disulfures sont représentés en jaune et l'atome de calcium en blanc.

## 2) Structure des gènes de linkers.

Avant le début de cette thèse, seule la structure du gène codant pour le linker L1 de l'oligochaete *Lumbricus terrestris* était connue : la séquence codante est interrompue par deux introns (figure 27). Le premier intron se situe juste avant le domaine LDL-A, et le deuxième se trouve après le codon stop (Suzuki and Riggs 1993).



**Figure 27 :** Position des introns dans le gène de Linker L1 chez *Lumbricus terrestris*.

L'origine évolutive des linkers chez les annélides est incertaine. Néanmoins, la présence de l'intron 1 juste avant le domaine LDL-A en position identique à celles des introns connus pour plusieurs récepteurs LDL (Sudhof et al. 1985a; Mehta et al. 1991), a conduit Suzuki à suggérer que les linkers proviendraient de la duplication d'un gène d'hémoglobine suivi d'un évènement de recombinaison génétique (exon shuffling) qui aurait permis l'insertion de la séquence codant pour le domaine LDL-A dans le gène de Linker (Suzuki, Takagi and Gotoh 1990; Suzuki and Riggs 1993). Nous avons apporté de nouveaux éléments de réponse (Chapitre 6) nous permettant de proposer une nouvelle structure de gène et de compléter le discours évolutif des linkers d'annélides.

## 3) Activité superoxyde-dismutase.

Une propriété particulière des HBL-Hb est l'activité superoxyde dismutase (SOD), qui intervient dans la liaison des ions radicalaires  $O_2^{\cdot -}$  (Liochev et al. 1996). Cette activité, découverte chez *Lumbricus terrestris* (Liochev et al. 1996), est connue pour être liée à la présence d'ions cuivre ( $Cu^{2+}$ ) et zinc ( $Zn^{2+}$ ) (Standley et al. 1988). Des mesures d'activité enzymatiques ont montré que le dodécamère, composé de chaînes de globines mais d'aucun linker, avait très peu d'activité SOD par rapport à la molécule entière (Liochev et al. 1996). Ceci suggère que l'activité superoxyde dismutase résiderait plutôt dans les linkers (Liochev et al. 1996).

## **Chapitre 3**

**Introduction à la caractérisation des gènes  
codant pour l'hémoglobine extracellulaire  
de *Arenicola marina***

## **Chapitre 3 : INTRODUCTION A LA CARACTERISATION DES GENES CODANT L'HEMOGLOBINE EXTRACELLULAIRE DE *ARENICOLA MARINA*.**

### **I Contexte de l'étude.**

#### **A- Les substituts sanguins.**

La majorité des opérations chirurgicales ne nécessitent pas de transfusion sanguine. Mais les greffes d'organes et certaines interventions orthopédiques, vasculaires, cardiaques et gynécologiques peuvent entraîner d'importants saignements. Pendant longtemps, pour compenser ces pertes, on a presque exclusivement eu recours aux banques de sang. Peu de patients s'inquiétaient des risques éventuels de la transfusion, qui était alors considérée comme l'un des traitements les plus fiables. Dans les années 1980, de nombreux rapports scientifiques révélèrent la possibilité de transmission du virus VIH (Virus d'Immuno-déficience Humain) aboutissant au SIDA (Syndrome d'Immuno-Déficience Acquise), ainsi que d'autres virus par le biais de la transfusion sanguine.

Différentes solutions furent alors étudiées pour développer un substitut sanguin capable de remplacer le sang issu des donneurs. Un tel produit doit répondre à de nombreux critères : il doit être capable de lier l'oxygène, de le transporter dans les capillaires et de le délivrer à tous les organes. Il doit aussi avoir un temps de demi-vie assez long. De plus, un substitut sanguin « idéal » devra être universel, c'est-à-dire compatible avec tous les types sanguins. La solution transfusée doit être dénuée de tout agent pathogène, ne pas avoir d'effet antigénique, ne pas interférer avec les mécanismes de la coagulation, le système immunitaire ou tout autre voie métabolique. Concernant son utilisation, un substitut sanguin doit être facilement transportable, prêt à l'emploi et immédiatement efficace. Enfin, d'un point de vue économique, ce produit doit avoir un prix proche de celui des poches de sang utilisées actuellement (Riess 2001).

#### **B- Les différentes voies de recherche.**

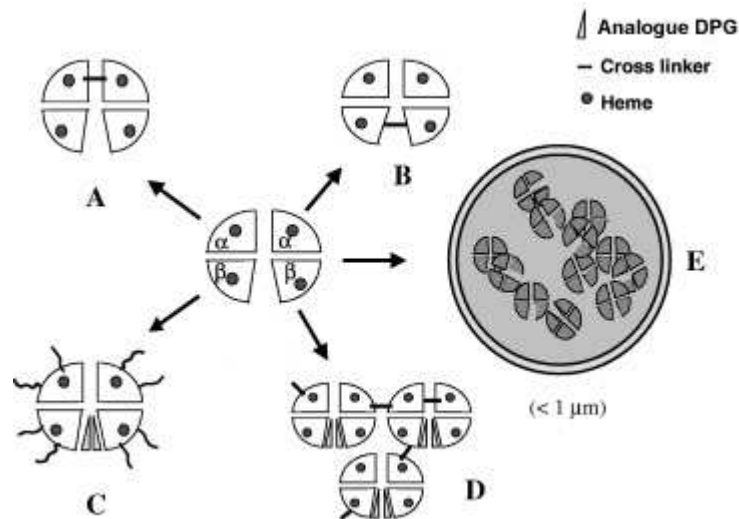
A l'heure actuelle, deux axes principaux sont étudiés pour développer un substitut sanguin humain : les **émulsions de perfluorocarbones (PFC)** et les transporteurs d'oxygène à base d'hémoglobine ou **hemoglobin based oxygen carriers (HBOCs)**.



Les PFCs sont des hydrocarbures dont les atomes d'hydrogènes ont été remplacés par des fluorines. Ces solutions inertes permettent la dissolution et le transport passif de l'oxygène et du dioxyde de carbone. Des solutions de PFCs sont actuellement en phase III d'essais cliniques (*Oxygen<sup>TM</sup>*), et certains produits (*Perftoran<sup>TM</sup>*) sont déjà utilisés en Russie (Maevsky et al. 2005).

Les travaux concernant les HBOCs ont pour idée initiale d'utiliser des hémoglobines de mammifères purifiées mais sont confrontés au problème de la toxicité rénale de ces hémoglobines qui, une fois en dehors de l'hématie, se dissocient en quatre monomères capables de passer au niveau des glomérules rénaux. Pour remédier à cela, deux possibilités ont été envisagées: d'une part l'encapsulation des hémoglobines, d'autre part, la stabilisation des hémoglobines tétramériques par polymérisation ou pontage chimique.

L'encapsulation des Hbs est basée sur l'idée de recréer les conditions naturelles des hématies, tout en s'affranchissant de la présence des antigènes marqueurs des groupe sanguins localisés à la surface des globules rouges. Ce procédé implique l'encapsulation des hémoglobines dans une vésicule composée d'une membrane phospholipidique (liposome) (Figure 28E).



**Figure 28:** Différentes approches des HBOCs à partir d'hémoglobine humaine purifiée. **A-B-** Stabilisation du tétramère par pontage intramolécule (Cross-link). **C-** Augmentation de la taille effective de l'hémoglobine par conjugaison avec du polyéthylène glycol. **D-** Polymérisation des hémoglobines par utilisation d'agents couplant. **E-** Encapsulation des hémoglobines dans un liposome. D'après Kim et al. (Kim and Greenburg 2004).

Les hémoglobines pontées peuvent être obtenues par différentes techniques. De façon à éviter la dissociation des tétramères, les monomères  $\alpha$  ou  $\beta$  peuvent être pontés entre eux grâce à des agents chimiques tels que le bis (3,5-dibromosalicyl) fumarate (DBBF) ou le nor-2-formylpyridoxal 5-phosphate (NFPLP) : on parle alors de pontage intramoléculaire (Figure 28A et 28B). Une autre approche est le pontage intermoléculaire, ou polymérisation des hémoglobines, grâce à des agents couplant tels que le glutaraldéhyde ou le glycoaldéhyde, qui vont permettre aux hémoglobines de rester plus longtemps dans la circulation sanguine (Figure 28D). Enfin, toujours dans l'optique d'augmenter la taille des hémoglobines, celles-ci peuvent être conjuguées à des polymères biocompatibles comme les polysaccharides (ex : polyéthylène glycol PEG) (Figure 28C).

C'est dans ce contexte d'utilisation d'hémoglobines naturelles acellulaires polymérisées que l'idée d'utiliser les hémoglobines extracellulaires d'Annélides, et plus particulièrement de *Arenicola marina*, comme substitut sanguin s'est développée.

#### C- Les premiers résultats sur les HBL-Hbs de *Lumbricus terrestris* et *Arenicola marina*.

Des travaux antérieurs, réalisés sur l'hémoglobine extracellulaire de l'oligochaete *Lumbricus terrestris* avaient en effet montré que les HBL-Hb possèdent des propriétés structurales et fonctionnelles compatibles avec les exigences concernant le développement d'un substitut sanguin (Hirsch et al. 1997). Les HBL-Hb sont naturellement polymérisées et présentent des résistances à l'auto-oxydation et à la dissociation importantes (Hirsch et al. 1997). De plus, les HBL-Hb sont facilement purifiables et la solution obtenue est homogène, contrairement aux Hb polymérisées grâce à des agents chimiques (Hirsch et al. 1997). Les expériences de transfusions menées sur des souris d'une part avec l'hémoglobine extracellulaire de *Lumbricus terrestris* (Hirsch et al. 1997) et d'autre part avec l'HBL-Hb de *Arenicola marina* (Zal, Lallier and Toulmond 2000) ont montré que ces hémoglobines n'étaient pas toxiques et n'entraînaient pas d'effets physiopathologiques chez l'animal. Enfin, la fonction de transporteur d'oxygène chez l'animal a été testée grâce à l'observation du maintien de la balance énergétique dans un cœur de cobaye perfusé avec une solution contenant de l'HBL-Hb (Hirsch et al. 1997).

Tous ces travaux abondent dans le sens d'une utilisation potentielle des HBL-Hbs d'Annélides comme substitut sanguin. A terme, et dans le cadre de la production d'un substitut sanguin, l'alternative à l'extraction et à la purification de cette hémoglobine à partir

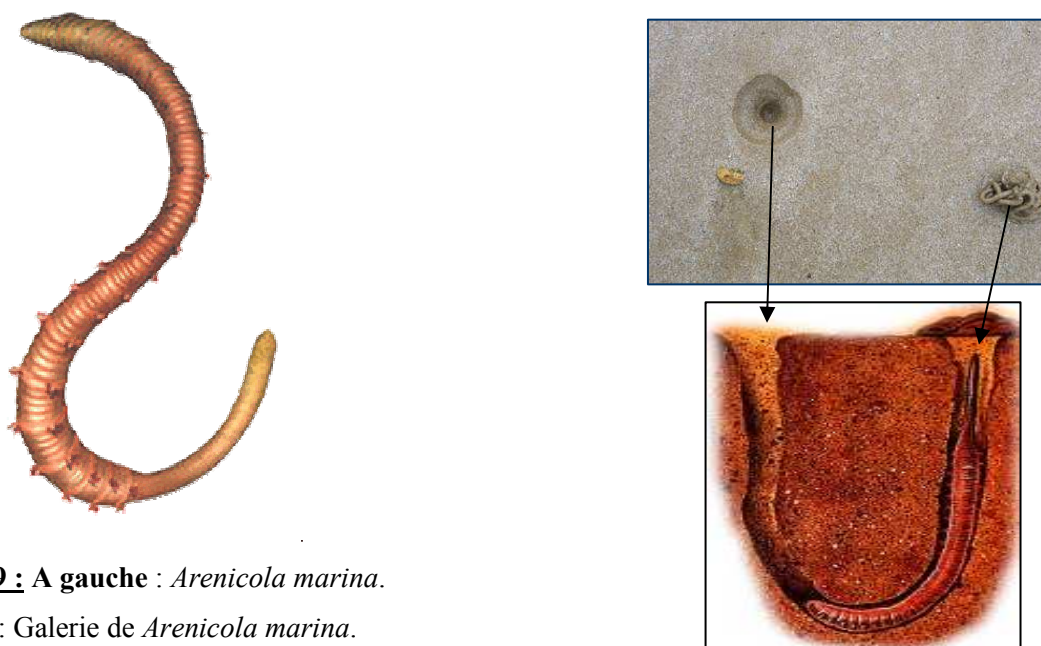
d'annélides vivantes élevées en ferme aquacoles, est d'envisager la production de cette hémoglobine par génie génétique (Zal et al. 2003). L'étape initiale de cette approche biotechnologique consiste donc à caractériser les différents gènes des familles multigéniques qui codent pour cette hémoglobine extracellulaire.

## **II Introduction à la caractérisation des gènes codant pour l'hémoglobine extracellulaire de *Arenicola marina*.**

### **A- Description de l'annélide polychaete *Arenicola marina*.**

*Arenicola marina* est une annélide polychaete vivant dans les sédiments sablo-vaseux, riches en sulfures, des zones intertidales des côtes de l'Atlantique Nord, de la mer Noire et de l'Adriatique. Les adultes peuvent atteindre une taille allant de 12 à 20 cm et varient en couleur du rose au noir, en passant par le rouge, le vert et le marron foncé (Figure 29). Ce polychaete sédentaire possède un corps séparé en une région thoracique et une région abdominale. La région thoracique consiste en 19 parapodes (avec des soies) dont seulement 13 exhibent des branchies. La région abdominale est plus étroite et consiste en plusieurs segments dénués de soies et de branchies (Figure 29) (Ashworth 1904).

*Arenicola marina* vit dans une galerie en forme de J de 20 à 40 cm de profondeur aisément repérable de la surface par la présence de dépressions à la tête de la galerie et de tortillons de sable déféqué à la sortie de la galerie (Figure 29).



**Figure 29 :** A gauche : *Arenicola marina*.

A droite : Galerie de *Arenicola marina*.

B- L'HBL-Hb de *Arenicola marina*.

Les propriétés fonctionnelles et structurales de l'HBL-Hb de *Arenicola marina* ont été très étudiées (Toulmond 1970a; Weber 1970; Toulmond 1977; Toulmond 1979; Weber 1981).

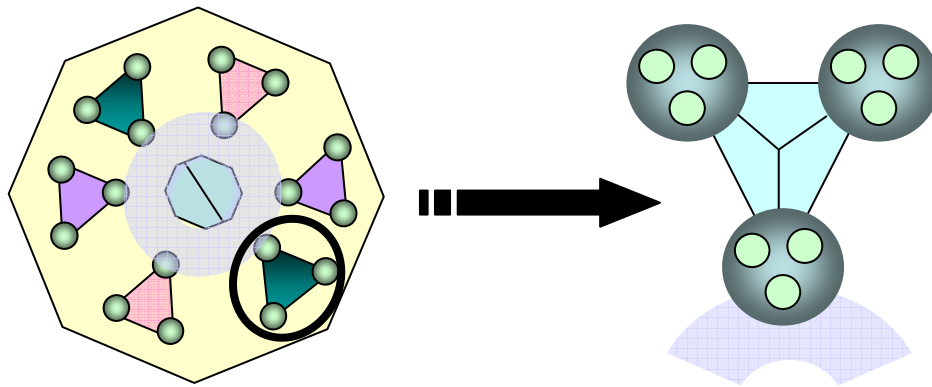
Les travaux de Zal et al. en 1997 permirent de déterminer le nombre de chaînes polypeptidiques composant l'HBL-Hb: 156 sous-unités de globines et 42 chaînes de linkers (Zal et al. 1997b).

La figure 30 résume la structure de l'HBL-Hb de *A. marina* déterminée par spectrométrie de masse.

Chaîne	Nombre de copies	Masse (Da)	Cys totales	Cys Libres
<b>Monomère</b>				
a1	36	15954.4	3	1
a2	72	15976.8	3	1
<b>Monomère impliqué dans un trimère</b>				
b1	4	15920.9	3	0
b2	8	16020.1	3	0
b3	4	16036.2	3	0
c	16	16664.8	3	0
d1	6	16983.2	5	1
d2	10	17033.1	5	1
<b>Linkers</b>				
L1	35	25174.1	12	0
L2	7	26829.7	12	0

**Figure 30 :** Composition de l'HBL-Hb de *A. marina*. Cys, Cystéines. D'après Zal et al. (Zal et al. 1997b). Les cystéines libres correspondent au nombre de cystéines non impliquées dans la formation de ponts disulfure intra ou inter-chaîne.

Les données de spectrométrie de masse ont aussi permis de proposer un modèle pour la structure quaternaire de l'hémoglobine de *Arenicola marina* (Figure 31) (Zal et al. 1997b). Une partie des globines de *Arenicola marina* forme des hétéro-trimères, les autres restent sous forme monomériques. Les linkers forment des homo et hétéro-dimères. Le docédamère est formé d'un trimère et de neuf monomères (Figure 31).



**Figure 31 :** Modèle de l'HBL-Hb de *A. marina*. **A gauche :** Vue de dessus de l'HBL-Hb. Chaque hexagone est composé de six dodécamères, chacun constitué d'un trimère (triangle en vert, rose ou violet) et de neuf monomères (cercle en vert clair). Les linkers apparaissent dans la partie centrale (Cercle mauve). **A droite :** Représentation d'un dodécamère. D'après Zal et al. (Zal et al. 1997b).

### C- Méthodologie.

L'objectif principal de cette thèse est de caractériser les gènes codant pour les globines et les linkers de *Arenicola marina* afin de pouvoir entreprendre la sur-expression de chacune des chaînes polypeptidiques et de développer des protocoles pour obtenir une HBL-Hb stable et fonctionnelle à partir des monomères sur-exprimés.

Les travaux antérieurs portant sur la structure de l'hémoglobine extracellulaire de *Arenicola marina* ont révélé l'existence de huit sous-unités différentes de globines et deux sous-unités différentes pour les chaînes de structure (linkers) (Zal et al. 1997b). L'hypothèse de travail que nous avons retenue est celle de huit gènes codant pour les globines et deux gènes codant pour les linkers.

Avant le commencement de cette thèse, aucune séquence protéique ou nucléotidique de globines ou de linkers de *Arenicola marina* n'était connue. Les globines extracellulaires d'annélides sont codées par une famille multigénique organisée en deux familles distinctes A et B, sub-divisées en quatre sous-familles A1/A2 et B1/B2. Nous avons tout d'abord cherché dans les banques de données des séquences de globines d'autres annélides appartenant aux quatre sous-familles A1, A2, B1 et B2. A partir d'alignements multiples, nous avons identifié des régions d'acides aminés conservés afin de concevoir des amorces dégénérées spécifiques de chacune des sous-familles A1/A2 et B1/B2. Après extraction des ARN totaux de *Arenicola marina* et purification des ARN messagers, des expériences de RT-PCR (Reverse-

Transcription Polymerase Chain Reaction) utilisant les amorces dégénérées nous ont permis d'obtenir des séquences d'ADNc de globines (Annexe 3). Parce que les régions conservées n'incluaient pas les acides aminés des extrémités N et C-terminales, les séquences d'ADNc obtenues étaient partielles. Pour obtenir les séquences complètes, incluant au moins les codons initiateur et stop ainsi que les séquences 5'UTR (Untranslated region) et 3'UTR, nous avons eu recours à la RACE PCR (Rapid Amplification of cDNA Ends Polymerase Chain Reaction) (Annexe 3).

Toutes les expériences de RACE PCR ne nous ont pas permis d'obtenir la séquence d'ADNc complète. Une alternative à cette technique pour séquencer les extrémités manquantes fut la « Marche sur le chromosome » (ou « Chromosome walking ») à partir d'ADN génomique (Annexe 3). Cette technique permet l'amplification de séquences dans les régions 5' et/ou 3' d'une séquence connue, à partir d'une amorce spécifique conçue dans la séquence d'intérêt et d'une amorce dégénérée pouvant s'hybrider à différents endroits sur le génome.

Pour déterminer la présence ou l'absence d'introns et leur position au niveau des gènes codant pour les globines extracellulaires et intracellulaires, nous avons conçu des amorces spécifiques, flanquant les sites putatifs des introns, à partir des séquences de chacun des ADNc préalablement identifié.

#### D- Contexte d'analyse des séquences obtenues.

Dans un premier temps (Chapitre 4), nous présentons les cinq globines de *Arenicola marina* que nous avons caractérisées, replacées dans le cadre de l'évolution des familles multigéniques des globines chez les annélides.

Dans un second temps (Chapitre 5), nous présentons les séquences d'ADNc des globines A2 et B2 de *Arenicola marina* et de la globine intracellulaire circulante de *Alvinella pompejana*. Nous avons aussi déterminé les séquences et positions d'introns des globines A2 et B2 de *Arenicola marina*, des globines A1, B1a, B1b et B1c de *Riftia pachyptila*, de la myoglobine MbIIa de *Arenicola marina*, de la globine intracellulaire circulante de *Alvinella pompejana*, de la globine intracellulaire circulante mIV de *Glycera dibranchiata* et de la déhaloperoxydase (DHP) de *Amphitrite ornata*. L'analyse de la position des introns au sein des globines intracellulaires circulantes et non-circulantes et extracellulaires d'annélides (séquences présentées dans les annexes 5, 6 et 8) nous a permis d'étudier l'évolution moléculaire de ces trois types de globines.

Enfin, dans un troisième temps (Chapitre 6), nous avons pu aborder l'évolution moléculaire des linkers d'annélides, grâce 1) au séquençage d'un des linker de *Arenicola marina*, 2) à l'analyse de séquences obtenues à partir d'EST (Expressed Sequence Tag) d'autres annélides (Annexe 5), et 3) par l'étude des positions d'introns dans un certains nombres de ces linkers (Annexe 7).

## **Chapitre 4**

### **Premier article:**

**The multigenic family of the extracellular  
hemoglobin from the annelid polychaete**

*Arenicola marina*



## **Chapitre 4: Premier article**

### **The multigenic family of the extracellular hemoglobin from the annelid polychaete *Arenicola marina*.**

Soumis le 24/10//2005 à Comparative Biochemistry and Physiology- Part B: Biochemistry & Molecular Biology

L'annélide polychaete *Arenicola marina* vit dans les sédiments sablo-vaseux des zones intertidales des côtes de l'Atlantique Nord, de la mer Noire et de l'Adriatique. Elle possède une hémoglobine extracellulaire dont la structure quaternaire dite en bicouche hexagonale (HBL-Hb) est composée de globines, dont la fonction est d'assurer le transport en oxygène, et de linkers, qui assurent le maintien de la structure quaternaire.

Dans le contexte du développement d'un substitut sanguin à partir de l'hémoglobine extracellulaire de *Arenicola marina*, nous avons envisagé la production de cette hémoglobine *in vitro*, impliquant pour cela la caractérisation des différents gènes codant pour cette HBL-Hb. Des travaux antérieurs sur la structure de l'hémoglobine extracellulaire de *Arenicola marina*, réalisés en spectrométrie de masse par électrobulisaison (ESI-MS), ont révélé l'existence de huit sous-unités différentes (Zal et al. 1997). L'hypothèse de travail la plus simple que nous avons retenue est celle de huit gènes codant pour ces huit sous-unités. A ce jour, nous avons séquencé cinq ADNc codant pour des globines.

La description de la famille multigénique partielle des globines extracellulaires de *A. marina* montre une organisation similaire à celle décrite pour d'autres annélides telles que *Riftia pachyptila* ou *Lumbricus terrestris*, c'est-à-dire une organisation en deux familles distinctes A et B, sub-divisées en quatre sous-familles A1/A2 et B1/B2 (Gotoh et al. 1987; Gotoh et al. 1991; Suzuki et al. 1993). Néanmoins, les phylogénies moléculaires obtenues après la caractérisation de séquences de globines d'Hirudinae, de Siboglinidae et des globines de *A. marina* issues de nos travaux suggèrent que l'organisation des familles multigéniques de globines extracellulaires pourrait être plus complexe que les quatre sous-familles précédemment décrites.

Parmi les séquences de globines caractérisées chez *A. marina*, nous avons mis en évidence la présence d'une globine B1, d'une globine B2, ainsi que trois globines A2. Des travaux antérieurs ont montré que les annélides vivant en milieux réduits (tels que *Riftia*

*pachyptila* ou *Lamellibrachia sp.*) expriment des globines A2 et B2 qui possèdent des cystéines libres, non impliquées ni dans le pont disulfure intrachaîne ni dans des ponts disulfures interchaînes (Bailly et al. 2002). Ces cystéines libres sont trouvées en positions conservées Cys+1 (dans le cas des hémoglobines B2) et Cys+11 (dans le cas des hémoglobines A2) (Bailly et al. 2002). Ces cystéines libres sont absentes des globines A1 et B1 des annélides vivant dans de milieux réduits, mais aussi de toutes les globines appartenant à des annélides vivant en milieux oxydés, tels que *Lumbricus terrestris* ou *Tylorrhynchus heterochaetus* (Bailly et al. 2002). Ces cystéines libres semblent directement impliquées dans la fonction de liaison de l'hydrogène sulfuré par les hémoglobines extracellulaires (Zal et al. 1998).

L'environnement de *A. marina* est riche en hydrogène sulfuré, et son HBL-Hb est capable de fixer et de transporter de l'hydrogène sulfuré (Zal et al. 1999). Parmi les trois globines A2 exprimées chez *A. marina*, deux d'entre elles ne possèdent pas la cystéine en position Cys+11 pourtant attendue. L'analyse des structures 2D de la région incluant cette cystéine dans les trois globines A2 de *A. marina* montre qu'elles sont identiques entre elles mais aussi avec les structures 2D déterminées pour d'autres globines A2 provenant d'annélides vivant en milieux réduits.

L'absence de la cystéine +11, remplacée par une alanine dans deux des trois globines A2 de *A. marina*, pourrait s'expliquer par une adaptation moléculaire limitant les désordres métaboliques dus à la présence d'un trop grand nombre de résidus cystéines libres aux groupements thiols réactifs menaçant l'homéostasie en fixant excessivement de l'hydrogène sulfuré ou d'autres composés.

## **The multigenic family of the extracellular hemoglobin from the annelid polychaete *Arenicola marina*.**

Christine Chabasse<sup>1</sup>, Xavier Bailly<sup>1</sup>, Morgane Rousselot<sup>1</sup> and Franck Zal<sup>1</sup>.

<sup>1</sup>Equipe Ecophysiologie: Adaptation et Evolution Moléculaires, UPMC – CNRS UMR 7144, Station Biologique, BP 74, 29682 Roscoff cedex, France

Running title: New surprising members in the extracellular hemoglobin multigenic family from the annelid polychete *Arenicola marina*.

Key words: annelid, HBL-Hb, hemoglobin, H<sub>2</sub>S, sulfide-binding function.

Abbreviations: Hb, hemoglobin; HBL-Hb; Hexagonal Bilayer Hemoglobin

### **Abstract**

The extracellular hemoglobin of the lugworm *Arenicola marina* which inhabits on the intertidal area, a sulfide-rich environment, comprises eight globin chains previously determined by mass spectrometry. We have cloned and sequenced five of the globins components. The deduced amino-acids sequences exhibit an extracellular signal peptide and two cysteine residues involved in an internal disulfide bond. The molecular weights calculated from the globin primary structures obtained from complete cDNA sequences are in good agreement with the mass spectrometry values obtained with the native hemoglobin. Phylogenetic trees constructed using the new *Arenicola* globin sequences and those from other annelid hemoglobins, indicate that the commonly accepted division of the annelid extracellular hemoglobins into four distinct subfamilies may not reflect the real multigenic family organization. Two of the globins were found to be A2 globin chains lacking the cysteine residues proposed to be involved in the binding of hydrogen sulfide by such hemoglobin. We discuss the unusual absence of these cysteines in the light of their invariant

occurrence in the A2 subfamily of hemoglobins from annelids inhabiting sulfide-rich environments.

## **Introduction**

The respiratory properties of the polychaete *Arenicola marina*, a lugworm found in the marine tidal areas, have been investigated extensively (Toulmond 1970a; Toulmond 1970b; Weber 1981; Zal et al. 1997; Wohlgemuth et al. 2000; Jouan et al. 2001). *A. marina* has a giant, extracellular hemoglobin (Hb), typical of annelid hemoglobins present in the vascular system. These Hbs, with a molecular mass of ~3.5 MDa, exhibit a characteristic quaternary structure, consisting of two hexagonal arrays and called Hexagonal Bilayer structure (HBL) (Lamy et al. 1996). Based on the crystal structure of *Lumbricus terrestris* HBL-Hb (Royer et al. 2000) and the cryomicroscopy data of *A. marina* HBL-Hb (Jouan et al. 2001), the two hexagonal arrays consist of twelve dodecamers of ~17 kDa globin chains (156 chains) and forty-two nonglobin, 24-32 kDa linker chains. The globin chains bind oxygen and others gaseous ligands reversibly (Vinogradov et al. 1993; Imai 1999; Weber and Vinogradov 2001) whereas the linker chains are required for the calcium-dependent assembly of the native HBL structure as shown for the *L. terrestris* HBL-Hb (Vinogradov et al. 1986; Gotoh et al. 1998; Kuchumov et al. 1999; Lamy et al. 2000).

Previous phylogenetic analyses of HBL-Hb globin sequences revealed a multigenic organization based on two distinct families A and B (Gotoh et al. 1987), subdivided in four sub-families: A1/A2 and B1/B2 (Gotoh et al. 1991; Suzuki et al. 1993; Suzuki and Vinogradov 2003). However, analyses employing recently acquired amino-acids sequences of globin chains from HBL-Hbs, especially those from the Hirudinae (*Macrobdella decora* and *Haemadipsa zeylanica*), have questioned this simple classification (Negrisolo et al. 2001; Shishikura 2004). Thus, it remains questionable whether the HBL-Hb multigenic family encoding for globin chains is strictly clustered in four sub-families (A1/A2 and B1/B2) or exhibits a more complex subdivision.

In 1997, based on the analyses of electrospray ionization mass spectrometry (ESI-MS), Zal and collaborators reported that the *A. marina* HLB-Hb was composed of eight different globin chains and two linker chains (Zal et al. 1997). Recently, two globin and one linker genes have been sequenced, encoding for A2a, B2 and L2 chains respectively. In this present work, we report the cDNA sequences of three new globin genes A2b, A2c and B1 of the *A. marina* HLB-Hb, as well as the correspondence with the ESI-MS data.

Our phylogenetic analyses of the *A. marina* globin chains confirm that annelid extracellular globin multigenic families organisation may be more complex than previously considered. An interesting additional feature, the absence of an expected free cysteine residue involved in sulfide binding in the A2 subfamily is also discussed in respect to the evolution of the hemoglobin sulfide binding function in annelids inhabiting sulfide-rich environments.

## **Material and Methods**

### ***Collection of biological material***

Juveniles specimens of the lugworm *A. marina* were collected at low tide from a sandy shore near Roscoff (Penpoull beach), Nord Finistère, France, and kept in local running sea water for 24h.

### ***Total RNA extraction and cDNA synthesis***

Entire juvenile specimens of *A. marina* were crushed into liquid nitrogen. Total RNAs were extracted using the RNable® buffer (Eurobio) and poly-A RNAs were then isolated using mRNA Purification Kit® (Amersham). A reverse transcriptase PCR was carried out using an anchor 5'-CTC CTC TCC TCT CCT CTT CCT oligo(dT)<sub>17</sub> primer.

### ***Amplification of cDNA***

The PCR reaction was carried out in a 25µL volume containing 10-50 ng cDNA/gDNA template, 100 ng of each degenerate primer, 200µM dNTPs, 2.5mM MgCl<sub>2</sub> and 1 Unit DNA polymerase (Uptima, Interchim). PCR conditions were as follows: an initial denaturation step at 95°C for 5 min, 35 cycles consisting of denaturation at 95°C for 40s, annealing at 58°C for 40s, extension at 72°C for 50s, and a final elongation step at 72°C for 10 min. Primers are available upon request to the authors.

### ***Cloning and sequencing***

The PCR products were cloned using a TOPO®-TA cloning Kit (Invitrogen). The positive recombinant clones were isolated and plasmid DNA was prepared by the FlexiPrep Kit (Amersham). Purified plasmids containing the putative globin insert were used in a dye-primer cycle sequencing reaction, using either the labeled Texas Red universal primer T3 or the primer T7 and the Big Dye® Terminator V3.1 Cycle Sequencing kit (Applied

Biosystems). PCR products were subsequently run on 3100 Genetic Analyser (Applied Biosystems).

### ***Rapid amplification of cDNA ends***

cDNAs ends were obtained by PCR using the 5' and 3' RACE kit from Roche according to the manufacturer's instructions. Buffer, reagents and other conditions for the nested PCR were as described by the manufacturer. The RACE products were purified, cloned with the TOPO-TA cloning kit (Amersham) and sequenced as described above.

### ***Sequences analysis***

The peptide signal cleavage site was determined by the SignalP 3.0 Server (<http://www.cbs.dtu.dk/services/SignalP/>) (Bendtsen et al. 2004).

### ***Hydrophobic cluster analysis***

The two-dimensional Hydrophobic Cluster Analysis (HCA) (<http://www.lmcp.jussieu.fr/~soyer/www-hca/hca-form.html>) (Callebaut et al. 1997) was used to detect and establish the similarity between the different globin chains of *A. marina*.

### ***Globin multiple alignment construction***

The globin sequences used in the multiple alignment are listed in table 1. Amino acid sequences were aligned with MUSCLE program (Edgar 2004) ([http://phylogenomics.berkeley.edu/cgi-bin/muscle/input\\_muscle.py](http://phylogenomics.berkeley.edu/cgi-bin/muscle/input_muscle.py)) and verified manually.

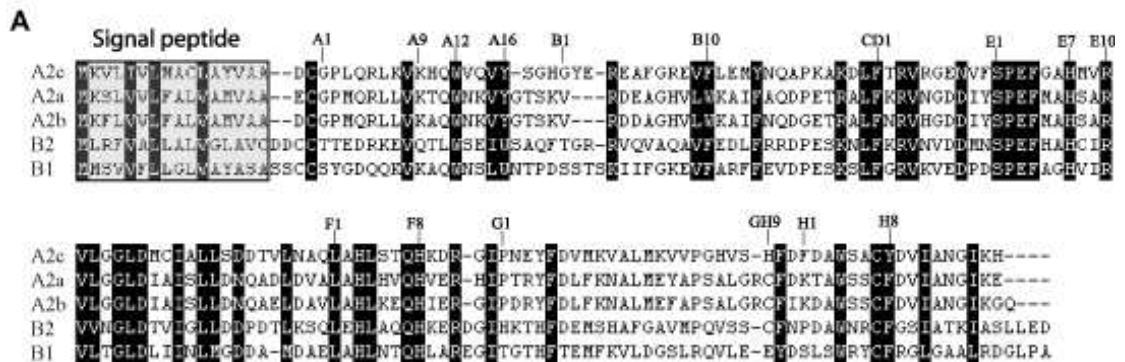
### ***Molecular phylogeny***

A phylogenetic tree was constructed using the Neighbor-Joining (NJ) method from modified multiple alignment in which gapped domains ambiguously aligned were manually deleted (Fig.2). The tree was computed using the PHYLIP programs package version 3.63 (J. Felsenstein, Phylogeny Inference package, version 3.63; Department of Genetics, University of Washington, Seattle) with 1,000 bootstrap resamplings and the JTT transition matrix (Jones et al. 1992). Bayesian analysis was carried out using MrBayes 3.1.1 (<http://mrbayes.csit.fsu.edu/index.php>) and the JTT transition matrix (Jones et al. 1992). Four chains were run simultaneously for  $10^6$  generations and trees were sampled every 100 generations, producing a total of  $10^4$  trees.

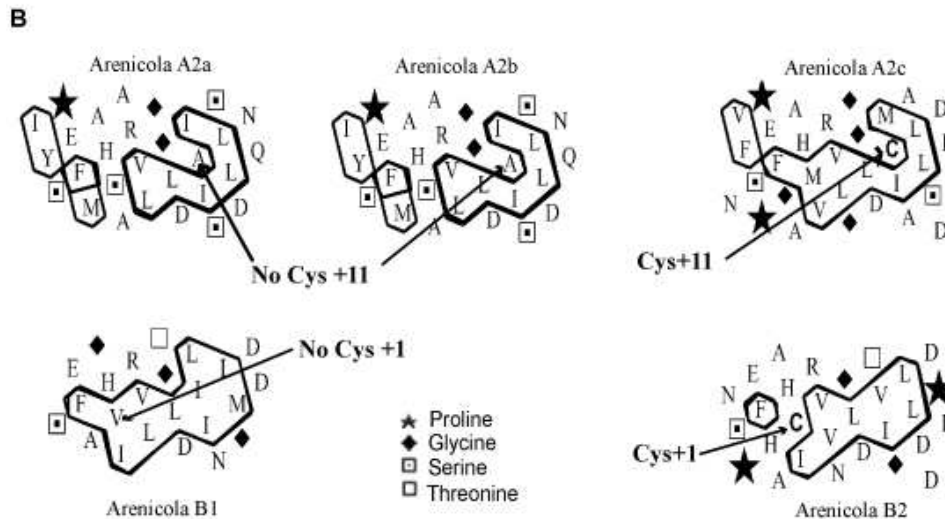
## Results and Discussion

### Globin chains description and assignation to sub-families.

We succeeded in sequencing three *Arenicola marina* globin cDNAs in full length, including part of 5'UTR, entire coding sequence and full 3'UTR. The amino-acid sequences deduced from the nucleotides sequences comprised 158 amino-acids for A2b (AM109949), 157 amino-acids for A2c (AM109950) and 164 amino-acids for B1 (AM109951). For each one of these sequences, the 16 first amino-acids have been predicted to be a signal peptide, which is in agreement with the extracellular localization of HBL-Hbs. Figure 1A shows the alignment of the deduced complete amino-acids sequences of these three globin chains with two others *A. marina* globin chains already described, A2a (CAI56308) and B2 (CAI56309). These new globin chains were unambiguously designated as A2b, A2c and B1. Indeed, the presence of A9-Lys, B10-Trp, and both deletions of two-residue between the A and B helices and one-residue between F and G helices indicate that A2b and A2c belong to the A family. Similarly, the presence of the A4-Asp, A16-Trp, B10-Phe, as well as the insertion of two residues between A and B helices, and one residue between F and G helices clearly indicate that B1 belongs to the B family. However, the assignment to either the A1/A2 or B1/B2 sub-families has only been possible thanks to the phylogenetic tree.



**Figure 1:** A- Multiple alignment of the five globin amino-acids sequences from *Arenicola marina* HBL-Hb. The 16 first amino-acids residues have been predicted to be a signal peptide, and are indicated by a grey box. The conserved amino-acids residues are indicated in black. Letters above the sequence indicate the helical designation, based on the *Lumbricus terrestris* helical structure (Strand et al. 2004).



**Figure 1: B-** The 2-D HCA plots of the sulfide binding domain are indicated for the five globin chains from *A. marina*.

The cDNA-derived amino-acids sequences of the three new globin chains were aligned with annelids HBL-Hb globin chains in the multiple alignment shown in Fig.2. The sequence of the extracellular globin from *Lumbricus terrestris* (Strand et al. 2004) were used to determine the helix assignment (Fig.2). Only five amino-acids are conserved between all extracellular globin chains (Fig.2): NA2-Cys, A12-Trp, E10-Arg, F8-His, and H7-Cys. However, in all *A. marina* globin sequences, the distal Histidine E7-His is invariant, and not replaced by a Phenylalanine residue (as in *Macrobodella decora* A2, *Haemadipsa zeylanica* A2 and *Eudistylia vancouverii* A1) or by a Glutamine residue (such as in *Oligobrachia mashikoi* B1 and *Riftia pachyptila* B1c).

The deduced amino-acids sequences of A2b, A2c and B1 globin chains have a molecular mass of 15921.1 Da, 15977.3 Da and 16660.8 Da respectively. In a previous study leading on *A. marina* extracellular Hb (Zal et al. 1997), polypeptidic chain composition of HBL-Hb was determined by Electrospray Ionization Mass Spectrometry (ESI-MS). The correspondence between the calculated masses of the globin chains found in *A. marina* and the molecular masses previously determined by ESI-MS is shown in Table 2. A difference of approximately 88 Da appears in the case of A2a chain, whereas the molecular masses of the other ones are nearly identical. Such a discrepancy could be explained either by the sequencing of an allelic form because *A. marina* specimens used in this article were from a different batch from those used for the ESI-MS study, or to post-translation modifications, such as carboxylation of two glutamic or aspartic acid residues (44 Da each).

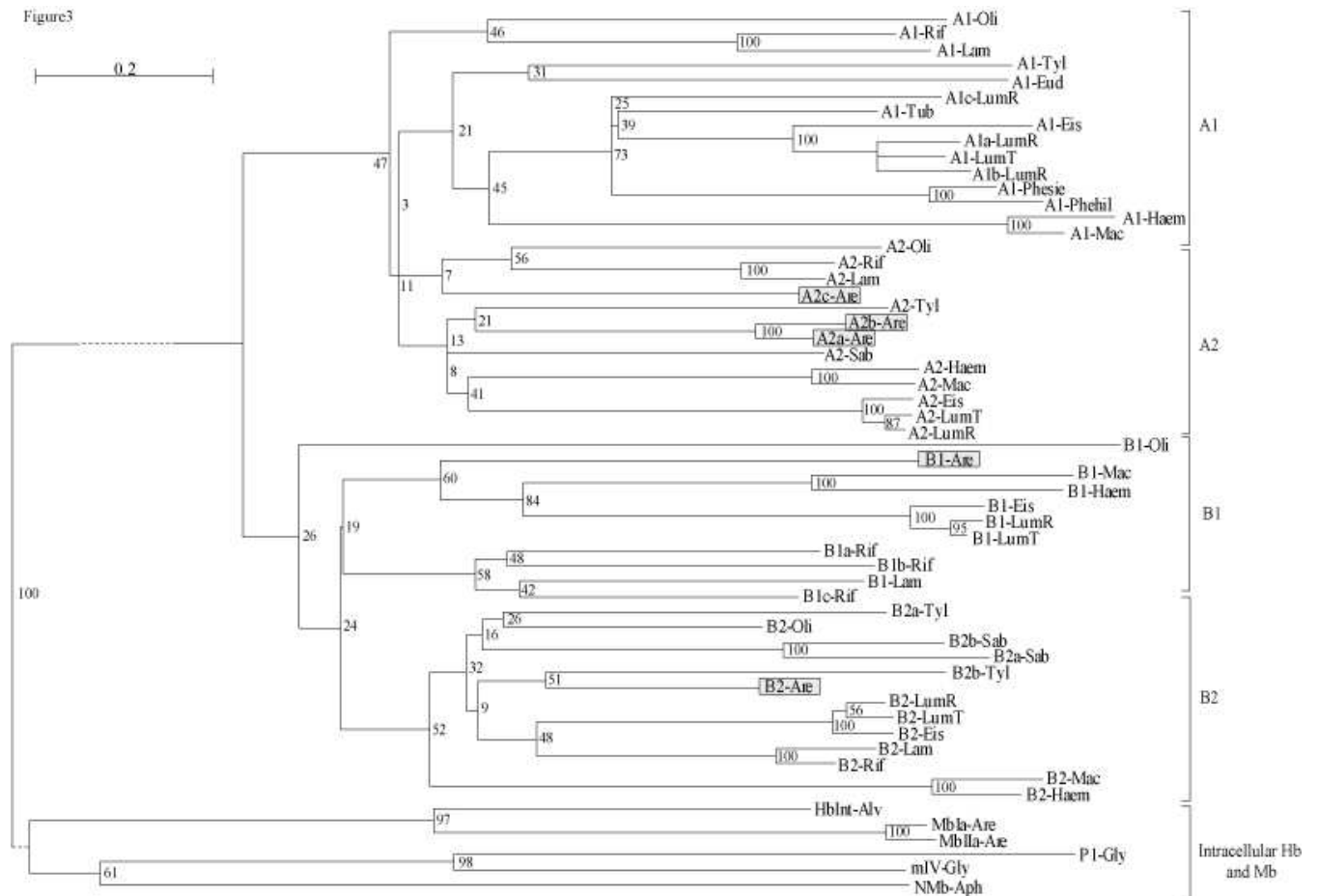




**Figure 2:** Amino-acid sequence alignment of *Arenicola marina* globin chains with other annelid hexagonal bilayer hemoglobins (E), intracellular hemoglobin and myoglobin (I). Abbreviations are shown in table 1. Letters above the sequence indicate the helical designation for key residues, based on the *L. terrestris* helical structure (Strand et al. 2004). The five extracellular hemoglobin conserved amino-acids are indicated by black triangle.

*Globin is a poor phylogenetic marker.*

In this study, we have constructed a phylogenetic tree from annelid intracellular hemoglobin, myoglobin and extracellular hemoglobin, including the new *A. marina* globin chains (Fig.3). The trees obtained with the Neighbor-Joining (NJ) and the Bayesian methods were congruent (*i.e* exhibiting the same topology). As shown in Fig.3, the extracellular hemoglobins cluster independently from the intracellular ones (including Hb contained in free circulating cells and Mb) with high bootstraps values. Moreover, two main groups, A and B, clearly appear in the extracellular hemoglobin cluster, in good agreement with the classification of annelids extracellular globins already suggested by Gotoh and collaborators (Gotoh et al. 1987). The phylogenetic tree (Fig. 3) allows us to assign the three new globin chains from *A. marina* to either B1 or A2 sub-families. However, even though the B family is separated into two sub-families B1 and B2, we can not clearly demonstrate the occurrence of two distinct sub-families in the A cluster, reinforcing the assumption that such a classification into four distinct sub-families would be an oversimplification of the real situation (Negrisolo et al. 2001). Globin chains in the three annelid groups, the polychaetes, hirudinae and oligochaetes, evolved obviously under different selection pressures, given the variety of ecological niches occupied by annelids and the phylogenetic complexity of this phylum. Indeed, no studies have demonstrated the monophyly of annelids: it is not surprising to observe such a divergence between orthologous and paralogous globin chains, whose corollary is the saturation of the phylogenetic signatures and then the poor resolution of globin subfamilies in molecular phylogenetic analyses.



**Figure 3:** Neighbor-joining tree for amino-acid sequences of annelid extracellular HBL-Hb (A1, A2, B1 and B2), intracellular hemoglobin and myoglobin. Bootstraps values are shown at the branching point. See Table 1 for abbreviations.

Occurrence of cysteine residues involved in intra- and inter-globin chain disulfide bridge.

The cysteine residues in HBL-Hbs globins play an important role in the subunit structure through the formation of intra- and inter globin chain disulfide bridges. As described in Zal and collaborators (Zal et al. 1997), and confirmed by the primary sequences shown here, the *A. marina* globin chains exhibit either three or five cysteine residues (Fig.1A). The crystal structure of *Lumbricus terrestris* HBL-Hb dodecamer have been recently reported by Strand and collaborators (Strand et al. 2004) and confirms the occurrence of intra-chain disulfide bridge between globin chains thank to two cysteine residues NA2-Cys and H7-Cys (Fig.2). These residues are present in *A. marina* globin chains and more widely to all annelid extracellular Hbs. Neither the annelid intracellular hemoglobins from *Glycera dibranchiata* and *Alvinella pompejana*, nor the myoglobins from *A. marina* and *Ophelia bicornis*, have these cysteine residues. Therefore, this intra-chain disulfide bridge (NA2-Cys and H7-Cys)

seems to be characteristic of HBL-Hbs and implies that the occurrence of these cysteine residues may be correlated with the extracellular nature of the HBL-Hbs.

In addition to the two cysteine residues involved in the foregoing intra-chain disulfide bridge, annelid HBL-Hb globin chains have several additional cysteine residues implicated in interchain disulfide bonding (Fig.2). Indeed, based on the three-dimensional structure of the *L. terrestris* dodecamer (Strand et al. 2004), the NA1-Cys and GH9-Cys residues in the *A. marina* globin chains A2a, A2b, B1 and B2, would form inter-chain disulfide bonds in the globin hetero-trimer subunit.

#### Presence or absence of free cysteine residues involved in sulfide-binding.

Free cysteine residues (*i.e* not involved in disulfide bridge) have been proposed to be implicated in the binding of hydrogen sulfide (H<sub>2</sub>S) (Zal et al. 1998). These free cysteine residues, found in positions E8 (Cys+1) and E19 (Cys+11) after the conserved distal E7-His (Bailly et al. 2002), only occur in all A2 and B2 sub-families of the extracellular Hbs of annelids living in sulfide-rich environment (Bailly et al. 2003).

Recently, two contradictories articles have been published concerning two sulfide-rich living annelids: *R. pachyptila* (Flores et al. 2005) and *Oligobranchia mashikoi* (Numoto et al. 2005). The crystal structure of *R. pachyptila* hemoglobin has revealed that the sulfide-binding activity would rather be mediated through interaction between H<sub>2</sub>S and zinc ions (Zn<sup>2+</sup>) more than through interaction with free cysteine residues (Flores et al. 2005). Paradoxically, the crystal structure of the extracellular hemoglobin from *O. mashikoi* does not show any Zn<sup>2+</sup>, and experiments leading on the sulfide binding activity of Hb, illustrated by the binding of mercury (Hg) derivatives on free cysteine residues, suggest that *O. mashikoi* Hb binds as many sulfide molecules as free cysteine residues in the Hb (Numoto et al. 2005). In addition, occurrence of Zn<sup>2+</sup> has also been reported in hemoglobins from annelids living in sulfide-free environments, such as *Lumbricus terrestris*, *Tylorrhynchus heterochaetus* (Standley et al. 1988) and *Pheretima hilgendorfi* HLB-Hbs (Ochiai et al. 1993), in which no sulfide-binding function has been detected (Arp et al. 1987) and with A2 globin chains which do not exhibit the specific free cysteine residues (Fig.2).

As a consequence, it strongly suggests that free cysteine residues (and not Zn<sup>2+</sup>) are necessary for the sulfide-binding activity. However, Zn<sup>2+</sup> may play an additional role in the sulfide-binding function in some cases but only with the presence of free cysteine residues. A formal and ultimate experiment necessary to clearly demonstrate that cysteine residues are essential for the sulfide-binding would consist on either realize crystallography of Hb with

sulfide or test the sulfide-binding capacity of the Hb after site-directed mutagenesis of the cysteine residues.

*A. marina* lives in the reduced layer of sandy beaches, naturally enriched in H<sub>2</sub>S, and possesses a HBL-Hb which exhibits a sulfide-binding function (Zal et al. 1999). In fig.2, the multiple alignment reveals that a free cysteine residue is found in position (Cys+1) in the B2 globin chain, and in position (Cys+11) in the A2c chain, as expected. In the A2a and A2b globin chains, an alanine residue was found in position (Cys+11), which corresponds to the non-polar amino-acids found in the A2 sequences in the HBL-Hbs from *Lumbricus terrestris* or *Tylorrhynchus heterochaetus* living in sulfide-free environments (Bailly et al. 2003). The secondary structures of the domain flanking the free cysteine residues of *A. marina* globins were deduced from Hydrophobic Cluster analysis (HCA) plots (Fig. 1B) and compared with the ancestral sulfide-binding domain (SBD) described by Bailly and collaborators (Bailly et al. 2002; Bailly et al. 2003). As shown in Fig.1B, the molecular environment of these free cysteine residues is identical for the A2a, A2b and A2c globin chains. Given the occurrence of cysteine residue at position (Cys+11) in all A2 globin chains from annelids living in sulfide-rich environments, the absence of cysteine at position (Cys+11) in the A2a and A2b chains from *A. marina*, could be explained by the replacement of these free cysteine by alanine residues, resulting from a duplication event, followed by positive Darwinian selection (Bailly et al. 2003). Indeed, it has been established that the unexpected occurrence, by spontaneous mutation event, of a highly reactive amino-acid such as free cysteine could induce metabolic disorders or have a deleterious effect (Concha et al. 1997; Leder et al. 1999). In this case, the multiplication of H<sub>2</sub>S-binding sites due to the duplication events of the A2 globin genes could have led to a disruption of sulfide metabolic pathways. In addition, it is highly improbable that the cysteine residue at position Cys+11 only appeared in *A. marina* A2c chain because this residue is widespread in A2 globin chains from annelids living in sulfide-rich environments. The replacement of the cysteine residue at position Cys+11 by an alanine residue in the A2a and A2b chains in the *A. marina* HBL-Hb illustrates the selective pressure acting on the sulfide-binding function of annelid HBL-Hbs.

### **Acknowledgments:**

We are indebted to Serge Vinogradov for his critical reading and comments on the manuscript. This work was supported by CNRS, European grant (FEDER n° presage 3814)

and the Conseil Régional de Bretagne (contrat n°A2C809). The sequencing was realized at the Roscoff sequencing core facility Ouest genopole® Plateform.

**Table 1:** Globin sequences shown in the multiple alignment and molecular phylogeny.

Specie	Globin chain	Abbreviation	Accession number
<i>Arenicola marina</i>	A2a	A2a-Are	CAI56308
	A2b	A2b-Are	AM109949
	A2c	A2c-Are	AM109950
	B1	B1-Are	AM109951
	B2	B2-Are	CAI56309
	Mbla Mbla	Mbla-Are Mbla-Are	Kleinschmidt 1998 Kleinschmidt 1998
<i>Alvinella pompejana</i>	Hbl	Hblnt-Alv	CAI56311
<i>Aphrodite aculeata</i>	nerve Mb	NMb-Aph	AAC47259
<i>Eiseinia andrei</i>	A1	A1-Eis	BP524381
	A2	A2-Eis	BP524621
	B1	B1-Eis	BP524383
	B2	B2-Eis	BP524384
<i>Eudystilia vancouverii</i>	A1	A1-Eud	JC7606
<i>Glycera dibranchiata</i>	monomer IV	mIV-Gly	P02216
	P1	P1-Gly	CAA37995
<i>Haemadipsa zeylanica</i>	A1	A1-Haem	BAC84989
	A2	A2-Haem	BAC84991
	B1	B1-Haem	BAC84990
	B2	B2-Haem	BAC84992
<i>Lamellibrachia sp.</i>	A1	A1-Lam	AAP40327
	A2	A2-Lam	P15469
	B1	B1-Lam	AAP40328
	B2	B2-Lam	AAP04529
<i>Lumbricus rubellus</i>	A1a	A1a-LumR	DR009556
	A1b	A1b-LumR	CF415947
	A1c	A1c-LumR	CO046661
	A2	A2-LumR	DR009004
	B1	B1-LumR	CAA09958
	B2	B2-LumR	BF422540
<i>Lumbricus terrestris</i>	A1	A1-LumT	P08924
	A2	A2-LumT	P02218
	B1	B1-LumT	P11069
	B2	B2-LumT	P13579
<i>Macrobdeella decora</i>	A1	A1-Mac	BAC82447
	A2	A2-Mac	BAC82445
	B1	B1-Mac	BAC82448
	B2	B2-Mac	BAC82446
<i>Oligabrachia mashikoi</i>	A1	A1-Oli	Q7M419
	A2	A2-Oli	Q7M413
	B1	B1-Oli	BAD86545
	B2	B2-Oli	Q7M418
<i>Pheretima hilgendorfi</i>	A1	A1-Phehil	P83122
<i>Pheretima sieboldi</i>	A1	A1-Phesie	P11740
<i>Riftia pachyptila</i>	A1	A1-Rif	CAD29154
	A2	A2-Rif	P80592
	B1a	B1a-Rif	CAD29156
	B1b	B1b-Rif	CAD29157
	B1c	B1c-Rif	CAD29158
	B2	B2-Rif	CAD29159
<i>Sabella spallanzanii</i>	A2	A2-Sab	CAC37412
	B2a	B2a-Sab	CAC37410
	B2b	B2b-Sab	CAC37411
<i>Tubifex tubifex</i>	A1	A1-Tub	P18202
<i>Tylorrhynchus heterochaetus</i>	A1	A1-Tyl	P02219
	A2	A2-Tyl	P09966
	B2a	B2a-Tyl	P13578
	B2b	B2b-Tyl	P02220

**Table 2:** Correspondence between *Arenicola marina* globin chain molecular masses determined by Electrospray Mass Spectrometry (ESI-MS) (Zal et al. 1997) and calculated from the cDNA deduced amino-acid sequence.

ESI-MS			cDNA-derived amino-acids			$\Delta$ Mass (Da)
Chain	Mass (Da)	Total Cys	Chain	Mass (Da)	Total Cys	
<b>Monomer</b>						
a1	15954.4	3	×	×	×	×
a2	15976.8	3	A2c	15977.3	3	0.5
<b>Monomer involved in Trimer</b>						
b1	15920.9	3	A2b	15921.1	3	0.2
b2	16020.1	3	A2a	15932.2	3	87.9
b3	16036.2	3	×	×	×	×
c	16664.8	3	B1	16660.8	3	4.0
d1	16983.2	5	×	×	×	×
d2	17033.1	5	B2	17033.0	5	0.1

× : data not available.



## References

- Arp, A. J., Childress, J. J. and Vetter, R. D., 1987. The sulphide-binding protein in the blood of the vestimentiferan tube-worm, *Riftia pachyptila*, is the extracellular haemoglobin. *J Exp Biol.* 128, 139-158.
- Bailly, X., Jollivet, D., Vanin, S., Deutsch, J., Zal, F., Lallier, F. and 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.
- Bailly, X., Leroy, R., Carney, S., Collin, O., Zal, F., Toulmond, A. and Jollivet, D., 2003. The loss of the hemoglobin H<sub>2</sub>S-binding function in annelids from sulfide-free habitats reveals molecular adaptation driven by Darwinian positive selection. *Proc Natl Acad Sci U S A.* 100, 5885-5890.
- Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* 340, 783-795.
- Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B. and Mornon, J. P., 1997. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci.* 53, 621-645.
- Concha, N. O., Rasmussen, B. A., Bush, K. and Herzberg, O., 1997. Crystal structures of the cadmium- and mercury-substituted metallo-beta-lactamase from *Bacteroides fragilis*. *Protein Sci.* 6, 2671-2676.
- Edgar, R. C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792-1797.
- Flores, J. F., Fisher, C. R., Carney, S. L., Green, B. N., Freytag, J. K., Schaeffer, S. W. and Royer, W. E., Jr., 2005. Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. *Proc Natl Acad Sci U S A.* 102, 2713-2718.
- Gotoh, O., Suzuki, T. and Takagi, T., 1991. Nomenclature of the Major Constituent Chains Common to Annelid Extracellular Hemoglobins. 279-283,
- Gotoh, T., Shishikura, F., Snow, J. W., Ereifej, K. I., Vinogradov, S. N. and Walz, D. A., 1987. Two globin strains in the giant annelid extracellular haemoglobins. *Biochem J.* 241, 441-445.

- Gotoh, T., Sano, T., Shibuya, A., Yamaki, M., Imai, K. and Ebina, S., 1998. Hexagonal bilayer structuring activity of linker chains of an annelid giant hemoglobin from the polychaete *Perinereis aibuhitensis*. Arch Biochem Biophys. 360, 75-84.
- Imai, K., 1999. The haemoglobin enzyme. Nature. 401, 437, 439.
- Jones, D. T., Taylor, W. R. and Thornton, J. M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 8, 275-282.
- Jouan, L., Taveau, J. C., Marco, S., Lallier, F. H. and Lamy, J. N., 2001. Occurrence of two architectural types of hexagonal bilayer hemoglobin in annelids: comparison of 3D reconstruction volumes of *Arenicola marina* and *Lumbricus terrestris* hemoglobins. J Mol Biol. 305, 757-771.
- Kuchumov, A. R., Taveau, J. C., Lamy, J. N., Wall, J. S., Weber, R. E. and Vinogradov, S. N., 1999. The role of linkers in the reassembly of the 3.6 MDa hexagonal bilayer hemoglobin from *Lumbricus terrestris*. J Mol Biol. 289, 1361-1374.
- Lamy, J., Kuchumov, A., Taveau, J. C., Vinogradov, S. N. and Lamy, J. N., 2000. Reassembly of *Lumbricus terrestris* hemoglobin: a study by matrix-assisted laser desorption/ionization mass spectrometry and 3D reconstruction from frozen-hydrated specimens. J Mol Biol. 298, 633-647.
- Lamy, J. N., Green, B. N., Toulmond, A., Wall, J. S., Weber, R. E. and Vinogradov, S. N., 1996. Giant Hexagonal Bilayer Hemoglobins. Chem Rev. 96, 3113-3124.
- Leder, A., Wiener, E., Lee, M. J., Wickramasinghe, S. N. and Leder, P., 1999. A normal beta-globin allele as a modifier gene ameliorating the severity of alpha-thalassemia in mice. Proc Natl Acad Sci U S A. 96, 6291-6295.
- Negrisol, E., Pallavicini, A., Barbato, R., Dewilde, S., Ghiretti-Magaldi, A., Moens, L. and Lanfranchi, G., 2001. The evolution of extracellular hemoglobins of annelids, vestimentiferans, and pogonophorans. J Biol Chem. 276, 26391-26397.
- Numoto, N., Nakagawa, T., Kita, A., Sasayama, Y., Fukumori, Y. and Miki, K., 2005. Structure of an extracellular giant hemoglobin of the gutless beard worm *Oligobranchia mashikoi*. Proc Natl Acad Sci U S A.
- Ochiai, T., Hoshina, S. and Usuki, I., 1993. Zinc as modulator of oxygenation function and stabilizer of quaternary structure in earthworm hemoglobin. Biochim Biophys Acta. 1203, 310-314.
- Royer, W. E., Jr., Strand, K., van Heel, M. and Hendrickson, W. A., 2000. Structural hierarchy in erythrocrucorin, the giant respiratory assemblage of annelids. Proc Natl Acad Sci U S A. 97, 7107-7111.

- Shishikura, F., 2004. The complete amino acid sequences of four globins from the land leech *Haemadipsa zeylanica* var. *japonica*. *Comp Biochem Physiol B Biochem Mol Biol.* 137, 443-454.
- Standley, P. R., Mainwaring, M. G., Gotoh, T. and Vinogradov, S. N., 1988. The calcium, copper and zinc content of some annelid extracellular haemoglobins. *Biochem J.* 249, 915-916.
- Strand, K., Knapp, J. E., Bhyravbhatla, B. and Royer, W. E., Jr., 2004. Crystal structure of the hemoglobin dodecamer from *Lumbricus* erythrocytes: allosteric core of giant annelid respiratory complexes. *J Mol Biol.* 344, 119-134.
- Suzuki, T., Takagi, T. and 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, 141-146.
- Suzuki, T. and Vinogradov, S. N., 2003. Globin and linker sequences of the giant extracellular hemoglobin from the leech *Macrobdella decora*. *J Protein Chem.* 22, 231-242.
- Toulmond, A., 1970a. [Oxygen fixation by blood in the arenicola (*Arenicola marina*(L.), Annelida, Polychaeta)]. *C R Acad Sci Hebd Seances Acad Sci D.* 270, 1368-1371.
- Toulmond, A., 1970b. [Fixation of carbon dioxide by the blood in the arenicola (*Arenicola marina* (L.), Annelida, Polychaeta)]. *C R Acad Sci Hebd Seances Acad Sci D.* 270, 1487-1490.
- Vinogradov, S. N., Lugo, S. D., Mainwaring, M. G., Kapp, O. H. and Crewe, A. V., 1986. Bracelet protein: a quaternary structure proposed for the giant extracellular hemoglobin of *Lumbricus terrestris*. *Proc Natl Acad Sci U S A.* 83, 8034-8038.
- Vinogradov, S. N., Walz, D. A., Pohajdak, B., Moens, L., Kapp, O. H., Suzuki, T. and Trotman, C. N., 1993. Adventitious variability? The amino acid sequences of nonvertebrate globins. *Comp Biochem Physiol B.* 106, 1-26.
- Weber, R. E., 1981. Cationic control of O<sub>2</sub> affinity in lugworm erythrocytes. *Nature.* 292, 386 - 387.
- Weber, R. E. and Vinogradov, S. N., 2001. Nonvertebrate hemoglobins: functions and molecular adaptations. *Physiol Rev.* 81, 569-628.
- Wohlgemuth, S. E., Taylor, A. C. and Grieshaber, M. K., 2000. Ventilatory and metabolic responses to hypoxia and sulphide in the lugworm *Arenicola marina* (L.). *J Exp Biol.* 203 Pt 20, 3177-3188.

- Zal, F., Green, B. N., Lallier, F. H., Vinogradov, S. N. and Toulmond, A., 1997. 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., Leize, E., Lallier, F. H., Toulmond, A., Van Dorsselaer, A. and Childress, J. J., 1998. S-Sulfohemoglobin and disulfide exchange: the mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins. *Proc Natl Acad Sci U S A.* 95, 8997-9002.
- Zal, F., Gotoh, T. and Toulmond, A., 1999. Fifth International Congress of Comparative Physiology and Biochemistry. *Comp Biochem Physiol.* 124A,

## **Chapitre 5**

**Deuxième article:**

**Evolutionary functional transition from  
oxygen transport to halometabolites  
detoxification in the annelids globins family**

## **Chapitre 5: Deuxième article**

### **Evolutionary functional transition from oxygen transport to halometabolites detoxification in the annelids globins family.**

Les **globines intracellulaires non circulantes**, aussi appelées myoglobines, sont exprimées au sein de certains tissus tels que les muscles ; Les **globines intracellulaires circulantes**, sont exprimées au sein de cellules en circulation dans le système vasculaire ou coelomique des annélides, et les **globines extracellulaires** sont directement dissoutes dans le système vasculaire. Seules trois familles d'Annélides, les Terebellidae, les Alvinellidae et les Opheliidae, co-expriment ces trois types de globines.

Pour comprendre les relations phylogénétiques existant entre ces trois différentes lignées de globines, nous nous sommes intéressés à la structure de leurs gènes. Seules quatre structures de gènes codant pour des **globines extracellulaires** étaient connues avant ce travail: celles de la B1 de *Lumbricus terrestris* (Jhiang, Garey and Riggs 1988), de la A1 de *Eudystilia vancouverii* (Dewilde et al. 2001) et des A2 et B2 de *Riftia pachyptila* (Bailly 2003). Tous ces gènes présentaient le motif classique trois exons/ deux introns, avec les introns aux positions conservées B12.2 et G7.0. Nous avons recherché la présence de ces deux introns chez d'autres globines extracellulaires : A2 et B2 de *Arenicola marina*, ainsi que A1, B1a, B1b et B1c chez *Riftia pachyptila*. Ces résultats confirment l'existence d'un ancêtre commun à toutes les globines extracellulaires d'annélides, dont la structure du gène possédait probablement le motif trois exons/ deux introns.

Par ailleurs, aucune structure de gène codant pour une globine intracellulaire circulante n'avait été décrite chez les annélides. Concernant les globines intracellulaires non circulantes, seule la structure du gène de la neuro-myoglobine de *Aphrodite aculeata* avait été déterminé (Dewilde et al. 1996). Le gène codant pour cette neuro-myoglobine ne présentait qu'un seul intron en position G7.0. Cette globine, tout d'abord présentée comme une neuroglobine, est en fait surexprimée dans le tissu nerveux mais est aussi présente dans un certain nombre d'autres tissus comme le pharynx ou les muscles, ce qui permet de la considérer comme une globine intracellulaire non-circulante et non comme un nouveau type de globine.

La caractérisation de la structure des gènes codant pour une **globine intracellulaire non-circulante** de *Arenicola marina*, ainsi que des gènes codant pour les **globines intracellulaires circulantes** de *Glycera dibranchiata* et de *Alvinella pompejana* a permis de rejeter l'hypothèse d'une convergence structurale et fonctionnelle entre les globines intracellulaires circulantes et non-circulantes et les globines extracellulaires, puisque ces gènes possèdent tous le même motif avec les deux introns aux positions conservées B12.2 et G7.0. **Ces résultats confirment donc l'existence d'un ancêtre commun de type globine à ces trois lignées de globines d'annélides, dont le gène possédait déjà le motif trois exons/deux introns.**

Les analyses réalisées sur la séquence de la globine intracellulaire circulante de *Alvinella pompejana* ont révélé une forte similarité entre cette séquence et celle de la déhaloperoxydase (DHP) de *Amphitrite ornata*. Cet enzyme a pour fonction la détoxification des composés halogénés présents dans l'environnement de cette espèce (Chen et al. 1996). La structure cristallographique de la DHP a montré que cet enzyme possède une structure de type « globin-fold » ainsi qu'un hème (LaCount et al. 2000), et a gardé sa capacité à lier l'oxygène (Roach et al. 1997).

Nous avons cherché la structure du gène de la DHP : celui-ci présente les deux introns aux positions conservées B12.2 et G7.0. Les similarités de séquence et de structure entre la DHP et les globines intracellulaires, de même que les positions d'introns conservées suggèrent que le gène codant pour la DHP est probablement issu d'une duplication ancestrale à partir d'un gène de type globine, en d'autres termes, la DHP est une globine possédant une activité déhaloperoxydase.

*Amphitrite ornata* est un polychaete qui appartient à la famille des Terebellidae et co-exprime les trois types de globines. La globine intracellulaire circulante de *Amphitrite ornata* a été décrite dans les coelomocytes et est fonctionnelle à l'état monomérique (Mangum et al. 1975). La DHP est composée d'un homodimère, mais sa localisation tissulaire reste inconnue. On peut toutefois se demander si les monomères composant la DHP sont ou non identiques à la globine intracellulaire circulante. Ces résultats ouvrent de nouvelles perspectives dans la possibilité de décrire de nouvelles fonctions au sein des globines intracellulaires de certains Terebellidae, tels que *Amphitrite ornata* ou *Alvinella pompejana* qui étaient jusqu'à présent cantonnées à un rôle strict de transporteur d'oxygène.

## **Evolutionary functional transition from oxygen transport to halometabolites detoxification in the annelids globins family**

Chabasse<sup>1\*</sup>, C., Bailly<sup>1\*</sup>, X., Hourdez<sup>1</sup>, S., Dewilde<sup>2</sup>, S., Martial<sup>1</sup>, S., Rousselot<sup>1</sup>, M., Moens<sup>2</sup>, L., and Zal<sup>1</sup>, F.

\* **Have contributed equally to this work**

<sup>1</sup> Equipe Ecophysiologie: Adaptation et Evolution Moléculaires, UPMC – CNRS UMR 7144, Station Biologique, BP 74, 29682 Roscoff cedex, France

<sup>2</sup> University of Antwerp - Campus Drie Eiken, Department of Biomedical Sciences, Universiteitsplein 1, B-2610 Antwerp, Belgium

Running title: Molecular evolution of the globin family in annelids

Key words: gene structure, evolution, intracellular globin, extracellular globin, myoglobin, dehaloperoxydase, annelid

Abbreviations: Hb, hemoglobin; Mb, myoglobin; HBL-Hb; Hexagonal Bilayer Hemoglobin; DHP, dehaloperoxidase

### **Abstract**

We show that circulating intracellular Hb (*Alvinella pompejana* and *Glycera dibranchiata*), non-circulating intracellular Hb (*Arenicola marina* myoglobin) and extracellular Hb from various annelids all share a similar gene structure with two conserved intron position in B12.2 and G7.0.

This strongly suggests an ancestral Hb-like gene common to these three Hb types and some duplication or gene co-option events to explain such a diversification of hemoglobin entities and derived functions. In addition, there is a clear similarity and close phylogenetic relationships between some intracellular Hb and *A. ornata* dehaloperoxidase (DHP). We have characterized the gene structure of *A. ornata* in order to confirm assumptions regarding its



homology with hemoglobins. It appears that DHP share the same intron position than hemoglobin in annelids showing that DHP is a derived hemoglobin with a peculiar function.

## **Introduction**

Globins are heme-containing proteins that reversibly bind oxygen and other gaseous ligands (Imai 1999), and are widespread in the three major kingdoms of life (Hardison 1998; Vinogradov et al. 2005). Despite their great diversity of structures, the basic unit involved in the quaternary formula is a monomeric globin sharing a specific and highly conserved fold referred to as “globin-fold”. It usually comprises 6 to 8  $\alpha$ -helices (named A through H) separated by inter-helical regions. Based on this conserved basic structure and its prevalence in living organisms, it has been suggested that globin genes have evolved from a common ancestral gene which, after successive duplications and speciation events, led to the globin genes encoding for the widespread globin superfamily (Goodman et al. 1988; Moens et al. 1996; Hardison 1998; Suzuki and Imai 1998; Vinogradov et al. 2005).

Globins are among the best-studied proteins, especially in vertebrates where four types have been isolated: hemoglobin (Hb), myoglobin (Mb), cytoglobin (Burmester et al. 2002; Pesce et al. 2002) and neuroglobin (Burmester et al. 2000; Pesce et al. 2002). To date, three types of hemoglobins have been described in annelids: (1) Non circulating intracellular hemoglobin (Mb found in the cytoplasm of muscle cells) (Wittenberg 1970; Suzuki and Imai 1998), (2) circulating intracellular Hb (Vinogradov et al. 1993) and (3) extracellular Hb dissolved in circulating fluids (Vinogradov et al. 1993; Weber and Vinogradov 2001). These three types of Hb display diversity in sequence, quaternary structure and functions such as binding and transport of oxygen and hydrogen sulfide, superoxide dismutase and monooxygenase activities (Weber and Vinogradov 2001).

Annelid non-circulating intracellular Hb are generally encountered as monomers (and rarely found as dimers) (Wittenberg, Briehl and Wittenberg 1965; Terwilliger, Garlick and Terwilliger 1980)), and are thought to function as oxygen scavengers and storage (Suzuki and Imai 1998) as it is the case in the annelid *Arenicola marina* (Weber and Pauptit 1972; Kleinschmidt and Weber 1998).

Annelid circulating intracellular Hb can be found as monomers, dimers, tetramers and larger aggregational states (Mangum 1985; Weber and Vinogradov 2001). To date, we only

know the sequences (cDNA and amino acid sequences) of the intracellular Hb from the marine polychaete *Glycera dibranchiata* (Zafar et al. 1990a; Zafar et al. 1990b). *G. dibranchiata* erythrocytes contain two different types of Hbs, one is monomeric and the other one is polymeric (Seamonds, Forster and Gottlieb 1971; Harrington et al. 1978; Zafar et al. 1990b).

Annelid extracellular HBL-Hbs (Weber and Vinogradov 2001) contain two kinds of polypeptides chains, globin chains and linker chains. The latter one, by interaction with globin chains, allows the assemblage into large multi-subunit structure with a molecular weight of 3-4.10<sup>6</sup> Da. Extracellular globin chains are encoded by a multigenic family, which organization has previously been described (Gotoh et al. 1987; Bailly et al. 2002).

Only three annelids families are known to express simultaneously the three types of globin: the Terebellidae, the Alvinellidae and the Opheliidae (Weber 1978; Hourdez et al. 2000). The sporadic co-occurrence of the three globin types throughout three annelid families can be generalized to the whole annelid phylum, assuming that they were probably all present in a common ancestor. They may not be reported together more often because 1) they were not studied, or 2) they are expressed in very small amount (not easily detectable) or at a specific developmental stage or 3) are not expressed in the studied species (*i.e* loss of one of these Hb types). Despite studies regarding evolution of non-circulating intracellular Hb (Mbs) (Suzuki and Imai 1998) or extracellular Hbs (Gotoh and Suzuki 1990; Negrisolo et al. 2001; Bailly et al. 2002), the phylogenetic relationships between these different globins in annelids remains unclear. It has also been early claimed that the considerable differences in term of Hb amino acid sequence in annelids could “possibly herald the existence of a multiple origin of hemoglobins within annelids, and advocate caution in the practice of deducing phylogenetic relationships from protein sequence homologies” (Weber 1978).

In order to understand the emergence and evolution of these hemoglobins, we have sequenced new annelid globins polypeptides, cDNAs and genes. The study of gene structure (intron positions) provides an obvious opportunity to explore gene evolution.

The typical pattern of two introns/ three exons (with intron positions in B12.2 and G7.0) already found in numerous eukaryotic globin genes (Hardison 1998) has already been reported in four annelid extracellular globin genes from *Lumbricus terrestris* (Jhiang and Riggs 1989), from *Eudystilia vancouverii* (Dewilde et al. 2001) and from *Riftia pachyptila* (Bailly 2003). Except for the nerve Mb of *Aphrodite aculeata* where the first intron is missing (Dewilde et al. 1996), neither intracellular circulating nor non-circulating Hbs gene structures are known. We have characterized the gene structure of these Hbs, and the results have

confirmed an intuitive assumption: the common origin of the genes encoding for the three hemoglobin types.

Interestingly Blast interrogations revealed a strong amino-acid similarity between *Alvinella pompejana* intracellular Hbs and the dehaloperoxidase (DHP) enzyme from the marine annelid polychaete *Amphitrite ornata* involved in halometabolites detoxication (Chen et al. 1996). In addition molecular phylogenies analyses of Hb including DHP have shown the clustering of DHP with both circulating and non-circulating Hbs.

*Amphitrite ornata* possesses DHP which function is to convert halogenated phenols into quinones in presence of hydrogen peroxide (Chen et al. 1996). These heme-containing enzymes exhibit two conserved distal and proximal histidines found in most of globin sequences (Roach et al. 1997). Moreover, the crystal structure of native DHP has revealed a globin fold (LaCount et al. 2000). These data suggest that DHP activity may have arisen by a globin gene duplication and divergence (LaCount et al. 2000). Nevertheless, the gene structure of this DHP remained unknown and the evolutionary link between DHP and globin was not supported from a molecular genetic point of view: it is reasonable to rule out the homology between DHP and Hb given a possibility of structural convergence.

In this paper, we report 1) the sequence and gene structures of two HBL-Hbs from the polychaete *A. marina*, 2) the gene structures of four HBL-Hbs from the vestimentiferan *R. pachyptila*, 3) the nucleotide sequence and gene structure of an *A. marina* non-circulating Mb, 4) the sequence and gene structure of the polychaete *A. pompejana* intracellular circulating Hb, 5) the sequence and gene structure of the polychaete *G. dibranchiata* intracellular circulating Hb and 6) the sequence and gene structure of the polychaete *A. ornata* DHP A.

Position of introns, molecular signature features, phylogenetic relationships and quaternary structure of these Hb allowed us to suggest an evolutionary functional innovation from a Hb involved in oxygen transport to an enzymatic Hb involved in the detoxication of halometabolite in annelids. After all, functional inferences based on the compilation of *A. ornata* DHP functional properties, *A. ornata* DHP structure, multiple alignment and molecular motifs found from multiple alignment and molecular phylogenetic analysis, also enable to challenge the unique function of intracellular Hb previously described as oxygen carrier in some annelid such as Terebellidae.

## **Materials and Methods**

### ***Collection of biological material.***

Small specimens of the lugworm *Arenicola marina* were collected at low tide from a sandy shore near Roscoff (Penpoull beach), Nord Finistère, France, and kept in local running sea water for 24h.

Specimens of *Alvinella pompejana* were collected at 2500 m depth on the East Pacific Rise (9°50'N) by the manned submersible Nautille during the HOPE'99 cruise. Once on board, the animals were kept in chilled sea-water (10°C) until used for tissue collection.

Specimens of the hydrothermal vent tube worm *Riftia pachyptila* were collected at the vent site from the ridge segment 9°50'N on the EPR (Riftia field: 9°50.75'N, 104°17.57'W) at a depth of about 2.500m, during the French oceanographic cruise *HOT 96*. The worms were sampled using the telemanipulated arms of the submersibles *Nautille* and *Alvin*, brought back alive to the surface inside an insulated basket, and immediately frozen and stored in liquid nitrogen after their recovery on board.

Specimens of *Amphitrite ornata* were collected at Debidue flats, in the North Inlet estuary (Georgetown, SC, 33°20'N, 79°10'W) and immediately placed in 70% alcohol until used for DNA extraction.

Specimens of *Glycera dibranchiata* were imported from Maine, USA and immediately preserved in 70% alcohol until used for DNA extraction.

### ***Preparation of Alvinella pompejana intracellular hemoglobin.***

The coelomic fluid was collected by carefully opening the dorsal body wall in the middle part of the body. The coelomic fluid was centrifuged at 1000 g for 3 minutes at 4°C, and the cells were washed twice with filtered sea-water. After the last centrifugation, three volumes of distilled water were added to the obtained pellet of cells, provoking cell lysis. The obtained suspension was then centrifuged at 10,000 g for 5 minutes at 4°C. The supernatant, containing the cell extract, was then separated and frozen in liquid nitrogen. In order to prevent hydrolysis by proteases, PMSF (phenyl methyl sulfonyl fluoride) was added to a final concentration of 1 µmol/l before freezing. Intracellular Hb was prepared as described previously by Hourdez et al. (Hourdez et al. 2000).

***A. pompejana intracellular hemoglobin: protein sequencing.***

Heme extraction was performed by acid acetone precipitation. Dehemed Hb was pyridylethylated according to Allen (Allen 1989) and subsequently dialysed against 0.1% TFA. The protein was modified with maleic anhydride and cleaved with trypsin and CNBr. An Asp-Pro cleavage was performed as described in Allen (Allen 1989). The tryptic peptides were separated by liquid chromatography on a reversed phase column Vydac C4 developed with 0.1% TFA/CH<sub>3</sub>CN. The CNBr and Asp-Pro peptides were separated on an SDS-PAGE (Swank and Munkres 1971), and subjected to electroblotting. The peptides were sequenced in an ABI 471-B sequencer (Applied Biosystems) operated as recommended by the manufacturer. Amino-terminal sequence was obtained by subjecting intact Hb by Edman degradation.

***Total RNA extraction and cDNA synthesis.***

Entire juvenile specimens of *A. marina* and *A. pompejana* were crushed into liquid nitrogen. Total RNAs were extracted using the RNable® buffer (Eurobio) and poly-A RNAs were then isolated using mRNA Purification Kit® (Amersham). A reverse transcriptase PCR was carried out using an anchor 5'-CTC CTC TCC TCT CCT CTT CC(T)<sub>17</sub> primer.

***Isolation of Genomic DNA.***

Entire specimens of *A. marina*, *A. pompejana*, *R. pachyptila*, *G. dibranchiata* and *A. ornata* were washed in de-ionized H<sub>2</sub>O, and then incubated in 700µL of a PK Buffer (50mM Tris/HCl, 100mM NaCl, 25mM EDTA and 1% SDS, pH 8) with 15 µL Proteinase K (10µg/µl) at 65°C for 1h. The supernatant was separated by centrifugation. The DNA was separated by a classic phenol/chloroform extraction. The resulting DNA was precipitated by addition of isopropanol, placed at -20°C overnight, and washed once with 75% Ethanol. DNA was resuspended in 100µl TE buffer (10mM Tris/HCl, 0.1mM EDTA, pH 8) and stored at 4°C until used.

***Amplification of cDNA and genomic DNA.***

The PCR reaction was carried out in a 25µL volume containing 10-50 ng cDNA/gDNA template, 100 ng of each degenerate primer, 200µM dNTPs, 2.5mM MgCl<sub>2</sub> and 1 Unit DNA polymerase (Uptima, Interchim). PCR conditions were as follows: an initial denaturation step at 95°C for 5 min, 35 cycles consisting of denaturation at 95°C for 30s,

annealing for 30s, extension at 72°C for 40s, and a final elongation step at 72°C for 10 min. Primers used are available upon request to the authors.

### ***Cloning and sequencing.***

The PCR products were cloned using a TOPO<sup>®</sup>-TA cloning Kit (Invitrogen). The positive recombinant clones were isolated and plasmid DNA was prepared by the FlexiPrep Kit (Amersham). Purified plasmids containing the putative globin insert were used in a dye-primer cycle sequencing reaction, using either the labelled Texas Red universal primer T3 or the primer T7 and the Big Dye<sup>®</sup> Terminator V3.1 Cycle Sequencing kit (Applied Biosystems). PCR products were subsequently run on a 3100 Genetic Analyser (Applied Biosystems).

### ***Rapid amplification of cDNA ends.***

cDNAs ends were obtained by PCR using the 5' and 3'RACE kit (Roche) according to the manufacturer's instructions. Buffer, reagents and other conditions for the nested PCR were as described by the manufacturer. The RACE products were purified, cloned with the TOPO-TA cloning kit (Amersham) and sequenced as described above.

### ***Sequences analyses.***

#### ***Signal peptide***

The peptide signal cleavage site was predicted by the SignalP 3.0 Server (Bendtsen et al. 2004) (<http://www.cbs.dtu.dk/services/SignalP>).

#### ***Database analysis***

The tBlastn and tBlastx search algorithms (Altschul et al. 1997) were employed to search data at Uniprot (<http://www.ebi.ac.uk>).

### ***Hemoglobin multiple alignment***

The hemoglobin sequences used in the multiple alignment are listed in table 1. Amino acid sequences were aligned with MUSCLE program (Edgar 2004) ([http://phylogenomics.berkeley.edu/cgi-bin/muscle/input\\_muscle.py](http://phylogenomics.berkeley.edu/cgi-bin/muscle/input_muscle.py)) and adjusted manually.

### ***Molecular phylogeny***

An unrooted phylogenetic tree was constructed using the Neighbour-Joining (NJ) method from modified multiple alignments in which gapped domains ambiguously aligned were manually deleted (Fig.1). The tree was computed using the PHYLIP programs package version 3.63 (J. Felsenstein, Phylogeny Inference package, version 3.63; Department of Genetics, University of Washington, Seattle) with 1,000 bootstrap resamplings and the Jones, Taylor and Thornton (JTT) transition matrix (Jones, Taylor and Thornton 1992).

Bayesian analysis was carried out using MrBayes 3.1.1 (<http://mrbayes.csit.fsu.edu/index.php>) and the JTT transition matrix (Jones, Taylor and Thornton 1992). Four chains were run simultaneously for 10<sup>6</sup> generations and trees were sampled every 100 generations, producing a total of 10<sup>4</sup> trees.

### **Results**

#### *Identification of A2 and B2 chains from A. marina extracellular Hb*

A2 and B2 hemoglobin sequences (AJ880690 and AJ880691) were aligned and compared with other globins in the multiple alignment (Fig.1). The sequence of the extracellular globin from *Lumbricus terrestris* (Strand et al. 2004) was used to determine the helix assignment (Fig.1). These two new globin chains possess the well-conserved globins amino acids: Pro-C2, Phe-CD1, His-F8 and Trp-H4, as well as the two cysteines NA2 and H7 known to be involved in intra-chain disulfide bridge (Fushitani, Matsuura and Riggs 1988; Suzuki, Kapp and Gotoh 1988). Obvious molecular signatures and phylogenetic analyses allowed to unambiguously assign these two new globin chains to A2 and B2 extracellular Hb clusters according to the classification previously proposed by Suzuki et al (Suzuki, Takagi and Ohta 1993): A and B families are respectively subdivided in A1, A2 and B1, B2 sub-families.

- *A2 chain*: This sequence contains an open reading frame of 157 codons (including the initiation ATG codon). As in other annelid extracellular globins, there is a signal peptide which comprises residues 1 to 16. This signal peptide was removed in the alignment presented in Fig.1.

This sequence clearly belongs to the A family, as evidenced by typical molecular features: Lys-A9, Trp-B10 and also a deletion of three residues between A and B helices, and a deletion of one residue between F and G helices. Moreover, the two-residue Gly-Pro (at position A1-A2) indicates that this sequence belongs to the A2 group (Fig.1).

- *B2 chain*: This sequence contains an open reading frame of 165 codons (including the initiation ATG codon). A signal peptide was found and comprises residues 1 to 16. This signal peptide is not included in the alignment (Fig.1).

This sequence shares specific amino acid with the B family: Asp-A4, Trp-A16, Phe-B10 and Leu-B13. Furthermore, it shows an insertion of three residues between A and B helices, an insertion of one residue between F and G helices and a three-residue motif Pro-Gln-Val at position G17-19. Moreover, the three-residue motif Thr-Gly-Arg between A and B helices indicates that this sequence belongs to the B2 group (Fig.1).

#### *A. pompejana intracellular hemoglobin*

The partial cDNA sequence (AJ880693) was obtained using degenerated primers designed from Edman protein sequencing in order to look for gene intron position. This sequence displays the well-conserved amino acids of hemoglobins: Pro-C2, Phe-CD1, His-F8 and the Trp-H4 (Fig.1). However, the conserved Trp-A12 is replaced by an Ile residue, as in *A. marina* MbIIa.

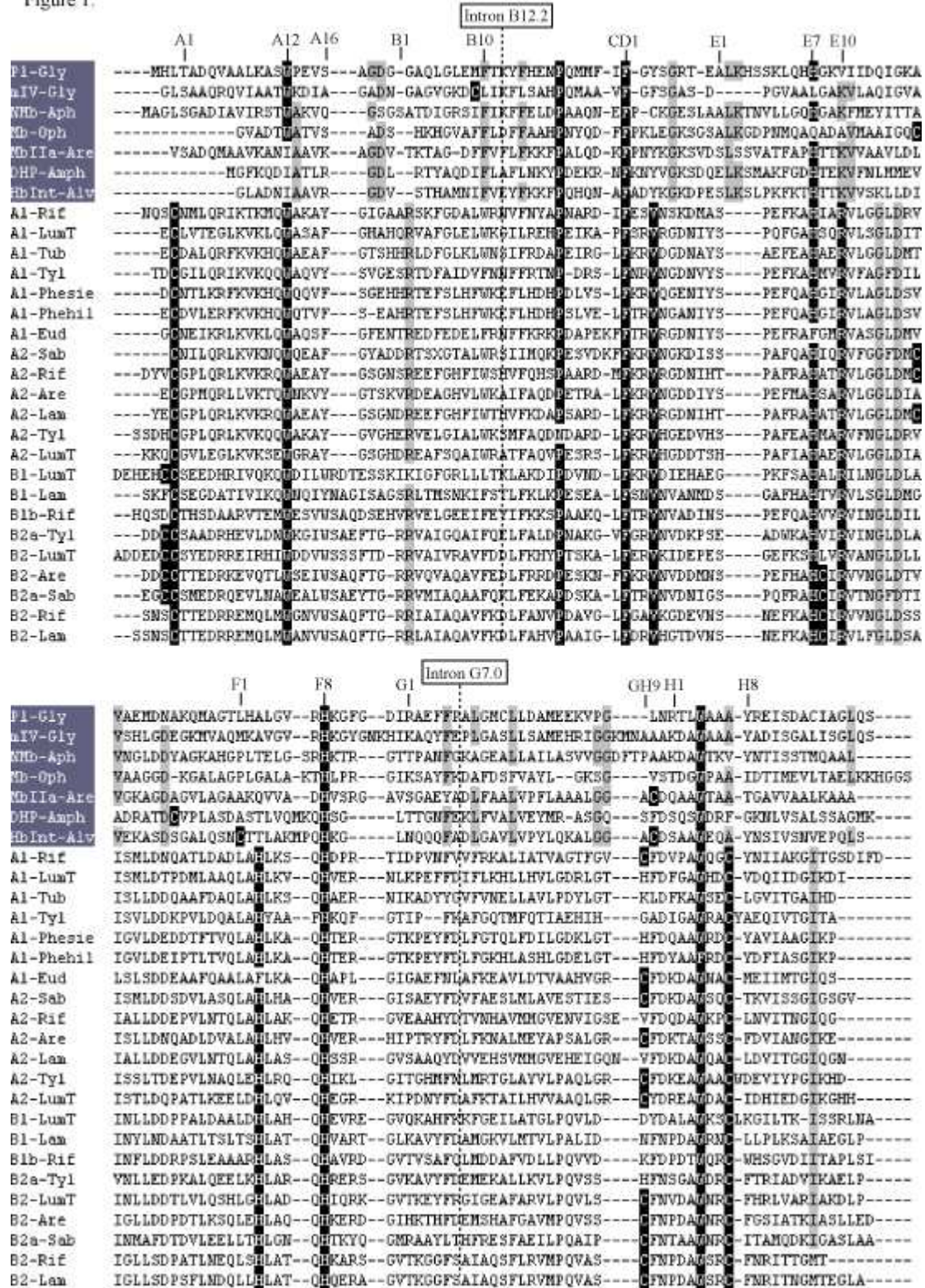
#### *Gene structure*

Introns were sequenced and characterized for *A. marina* A2, B2 extracellular Hbs, myoglobin MbIIa (AJ880692), the extracellular globin chains from *Riftia pachyptila*: A1, B1a, B1b and B1c, *A. pompejana* intracellular Hb, *G. dibranchiata* Hb mIV and for the *A. ornata* dehaloperoxidase A. The intron localization of *R. pachyptila* A2 and B2 was previously reported by Bailly et al (Bailly 2003). The position and length of each intron are summarized in Table.2. For all the sequences, including the dehaloperoxidase, the insertion positions of the two introns correspond to the usual B12.2 and G7.0 positions already described for many others hemoglobins sequences including *L. terrestris* (Jhiang, Garey and Riggs 1988) and *E. vancouverii* hemoglobins (Dewilde et al. 2001) (Hardison 1998) (Fig.1).

The splice sites have also been analyzed: the 5' splice donor is marked by a 8-nucleotide conserved sequence, the 3' acceptor site corresponds to a pyrimidine-rich region of 11-nucleotide followed by (C/T)AG, and the typical branch point signal corresponds to a 5-nucleotide sequence which functions as a signal for the spliceosome (Keller and Noon 1984; Harris and Senapathy 1990). In all introns, splicing donor and acceptor sequences are conform to the consensus sequences (Table 2).



Figure 1:



**Figure 1:** Multiple alignment of annelids DHP, extracellular and intracellular hemoglobins (circulating and non-circulating) amino acid sequences. Positions of intron 1 (B12.2) and 2 (G7.0) are indicated by dashed lines. The conserved amino-acids residues are indicated in black. Letters above

the sequence indicate the helical designation, based on the *Lumbricus terrestris* helical structure (Strand et al. 2004). Signal peptides, when present, have been removed. See table 1 for abbreviations.

### *Phylogenetic relationships*

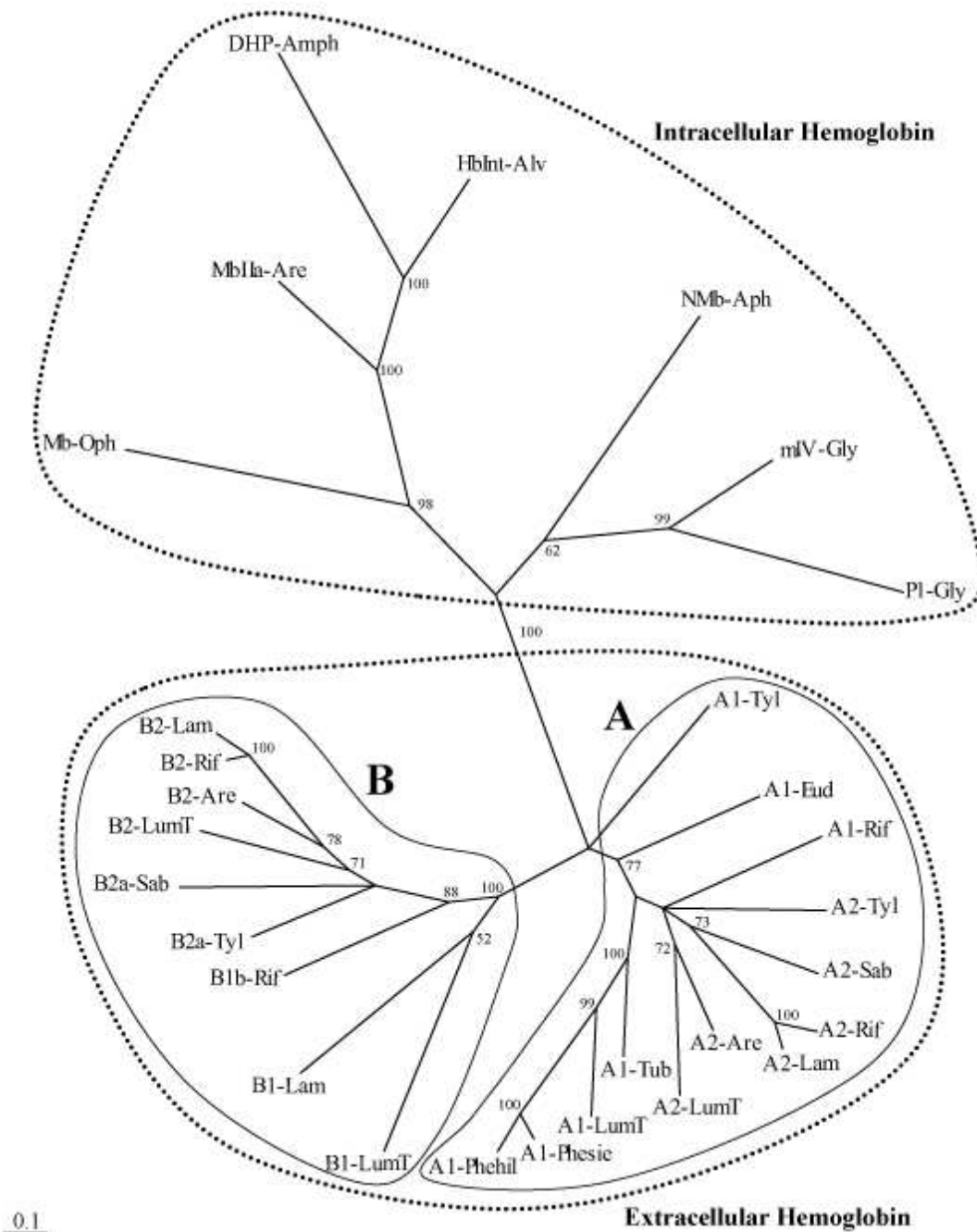
The Bayesian tree is shown in Fig.2. Two main clusters can be identified: one comprises all intracellular hemoglobins and the other, all extracellular hemoglobins. The intracellular cluster comprises both Mbs and intracellular Hbs. The extracellular cluster is divided into two groups: the A and B families (Gotoh et al. 1987), as expected. Newly sequenced *A. marina* A2 and B2 globin chains clearly cluster with A2 and B2 extracellular Hb chains. *A. pompejana* intracellular Hb and *A. ornata* DHP obviously belong to the intracellular cluster, different from a second cluster of intracellular hemoglobin including *G. dibranchiata* and *A. aculeata*.

## **Discussion**

### *Annelid globin gene structure*

All introns described in this article exhibit the splicing donor GT and acceptor AG sites (Table 2), whereas in the *Lumbricus terrestris* B1 globin gene, the donor GT is replaced by GC (Jhiang, Garey and Riggs 1988). The typical branch point sequence, involved in spliceosome binding, sometimes diverges from the standard consensus sequence (CTRAY), but studies have shown that this consensus sequence may not reflect the majority of branch point signals (Harris and Senapathy 1990).

We used the conservation of intron insertion position pattern in order to follow gene evolution relationships, assuming that conservation of intron position is a strong argument to reject evolutionary convergence and to state on the homology between considered genes. We have shown that *Riftia pachyptila* and *Arenicola marina* extracellular Hbs (A2 and B2), *Arenicola marina* Mb and *Alvinella pompejana* and *Glycera dibranchiata* intracellular Hb all share the same classical two introns/ three exons pattern (Table 2). These results allow ruling out the possibility of a structural and functional convergence between circulating and non-circulating intracellular hemoglobin and extracellular Hbs genes in annelids: it is unlikely that these globin genes have emerged independently from unrelated genes but rather from a common globin-like gene ancestor. In the *Aphrodite aculeata* nerve Mb evolutionary lineage, the gene structure might be explained by the loss of the first intron (Dewilde et al. 1996).



**Figure 2:** Bayesian phylogenetic tree based on annelids extracellular and intracellular hemoglobins (circulating and non-circulating) and DHP amino acid sequences. Bootstraps values are indicated above branches.

*Singular evolutionary history of extracellular Hb in respect to intracellular Hb in annelids*

Working in the field of comparative biochemistry and especially on intracellular hemerythrins and hemoglobins, Manwell early claimed a neglected problem (still unsolved) in annelids Hb evolution (Manwell and Baker 1988): “The transition between intracellular and extracellular respiratory proteins represents a profound evolutionary accomplishment. It is much more than placing an appropriate signal polypeptide portion, labeling a protein for extracellular export.”

Molecular signatures (Fig.1) clearly indicate that circulating and non-circulating intracellular Hbs share more common features among them than with extracellular Hbs. Phylogenetic analyses (Fig.2) show that these intracellular Hbs cluster independently from extracellular Hbs, with high bootstrap values. This strongly suggests that extracellular Hbs lineages have an independent evolutionary history. In the future, it will remain to state whether extracellular hemoglobin derived or not from a duplication event from one of the two intracellular hemoglobin genes. In addition, experiments realized on the polychaete *Travisia japonica* by Fushitani et al (Fushitani, Ochi and Morimoto 1982) have shown that antibody against the extracellular Hbs did not cross-react with the intracellular Hbs. This reinforces an early or rapid divergence between intra and extracellular Hbs lineages.

Globins are found in many unicellular organisms such as archae, eubacteria and lower eukaryote (Zhu and Riggs 1992; Andersson et al. 2003; Freitas et al. 2004; Vinogradov et al. 2005). Extracellular globins targeting via signal peptide acquisition is likely to be an apomorphic characteristic in multicellular organisms in respect to unicellular ones. It has been proposed that the secretory peptide could have been acquired by insertion of some other secreted protein gene by genetic recombination (Riggs 1991; Williams, Pal and Hurst 2000). Therefore, the secretory peptide found in all annelid extracellular Hbs must have been acquired before the extracellular Hbs multigenic family emergence.

Even though acquisition of a signal peptide must have been a determinant step, other features are specific to annelid extracellular Hbs. Extracellular state seems to be correlated to the formation of high molecular mass complex involved in the total colloid osmotic pressure of body fluids (vascular and coelomic bloods) (Toulmond 1992) and in minimizing rapid Hbs loss by excretion. HBL-Hbs are huge hetero-complex of hundred polypeptides chains composed of both globins and linkers, which association allow the structural cohesion of the whole protein. It is clear that the evolutionary elaboration of the HBL-Hbs structure have necessitated fine evolutionary mechanisms which have implied at the same time the possibility for globins and linkers to interact and to bind into dimers or other aggregationnal

states while conserving a functional hemoglobin. Indeed, linkers are supposed to result from duplication of globins and re-arrangement of exons (Riggs 1991). It has been suggested that these “non-oxygen-binding subunits” could also intrinsically possess superoxide dismutase and/or methemoglobin reductase activities necessary to keep the hemoglobin functional (Manwell and Baker 1988). Evolution of HBL-Hbs implies the co-evolution between both extracellular globins and linkers, which also represents a unique feature in annelids among the rest of the living organisms, including mollusks and arthropods which sometimes also possess extracellular Hbs but not HBL-Hb quaternary structure.

Given the similarity of both circulating and non-circulating intracellular Hbs with respect to extracellular Hbs and the occurrence of the same gene structure pattern in the three globins, these observations together suggest a complex molecular genetic scenario to explain the emergence of extracellular Hb in annelids from a common ancestral gene encoding for a Hb-like protein. It is important to note that arthropods extracellular Hbs sequences also exhibit a signal peptide, but do not possess neither internal disulfide bridge nor HBL-Hbs quaternary structure, which suggests a different evolutionary history compared with extracellular Hbs in annelids. This supports the emergence (innovation) of phylum specific extracellular Hb type.

#### *Dehaloperoxidase is an hemoglobin*

We have found that in Blast procedures, *Alvinella pompejana* intracellular hemoglobin primary sequence respectively blasts at  $2e-18$  (tblastn) and  $4e-19$  (tblastx) with *Amphitrite ornata* DHP, just after the *Arenicola marina* myoglobin  $2e-21$ (tblastn) and  $6e-22$  (tblastx) but before the rest of hemoglobins including intracellular (*Glycera dibranchiata*) and extracellular ones ( $\sim e-7$ ).

In addition, Blast procedures using *Amphitrite ornata* DHP as query respectively blasts with *Alvinella pompejana* intracellular Hb and *Arenicola marina* Mb at  $2e-18$  and  $4e-9$  (tblastn) and  $1e-18$  and  $2e-5$  (tblastx) before all hemoglobins and others proteins.

Moreover, *A. ornata* DHP, *A. pompejana* intracellular Hb and *A. marina* Mb exhibit high similarity and identity in term of molecular signatures between each other and with intracellular monomeric hemoglobin NH<sub>2</sub>-terminal sequence of the polychaete *Enoplobranchus sanguineus* (Weber et al. 1977).

These results, suggesting a close relationship between DHP and intracellular Hb, have been reinforced by molecular phylogenies analysis showing that *A. ornata* DHP, *A. pompejana* intracellular Hb, *A. marina* Mb and *Ophelia bicornis* Mb cluster together with a

high bootstrap values, in respect to other intracellular Hb (*G. dibranchiata* and *A. aculeata* Mb) and extracellular hemoglobins. Interestingly we have found that *A. ornata* DHP encoding gene exhibit the same gene structure as extracellular globin, *A. marina* myoglobin and intracellular hemoglobin from *A. pompejana* and *G. dibranchiata*. As another classical Hb, *A. ornata* DHP is able to bind oxygen: it is found in oxyferrous state when natively purified (Roach et al. 1997) and also exhibit a globin fold, a heme and distal histidine (Roach et al. 1997; LaCount et al. 2000).

The conserved intron positions between the three annelid hemoglobin types and DHP, the DHP globin fold and the amino-acid similarities between protein sequences including Hb and DHP allow to rule out some structural convergence evolution and to claim for homology (*i.e* common ancestry): in other words, we have confirmed through a molecular genetic approach that the so-called dehaloperoxidase is a hemoglobin with a derived function, using its heme to bind peroxide ligand in order to catalyze the oxidative dehalogenation of polyhalogenated phenols.

To date, the appearance of this peculiar DHP function in annelids cannot be explained by the sole fact of a functional shift (mutational or duplication events) which have lead to the sub-functionalization into DHP of a gene encoding for classical hemoglobin initially involved in oxygen transport. The tissue expression localization of DHP remains unknown: the DHP homodimeric enzyme has been extracted from crude extract of *A. ornata* (Chen et al. 1996). It is also important to keep in mind that a monomeric circulating intracellular Hb in *A. ornata* has been described in coelomic circulating cells (Mangum et al. 1975) which amino acid sequence is still unknown. It raises the question whether the monomeric Hb is identical to the two monomers of DHP. Indeed, we could reasonably suspect that the dimerisation of the monomeric intracellular Hb have lead to an emerging property of halometabolite detoxication. In this case it would not be a matter of a functional shift driving by a genetic modification event but rather a functional innovation using pre-existing structures.

Even though it remains to demonstrate that monomeric Hb is or is not identical to DHP monomers, it is clear that DHP function derived from a classical hemoglobin and could have occurred like an adaptation driven by selection action based on territorial war with annelids excreting halogenous compounds (Chen et al. 1996).

The corollary of this findings raise the question about the possibility to detect other or additional function(s) of the intracellular circulating hemoglobins in annelids such as *A. ornata* (Mangum et al. 1975) and *A. pompejana* (Hourdez et al. 2000). It is possible to challenge the single oxygen transport function of intracellular Hb in these annelids thanks to

molecular phylogenetic and gene structure analyses which support and illustrate the possibilities of (hidden) unexpected functions for intracellular Hb. For instance, in the case of *A. pompejana*, the harsh environment of hydrothermal vents is well-known to exhibit large amount of a wide range of toxic compounds which have to be neutralized or avoided.

**Acknowledgments:**

We gratefully thank Dr. João Gil for collecting specimens of *Glycera dibranchiata*, and Dr. David Lincoln for collecting specimens of *Amphitrite ornata*. We acknowledge the captain and crew of the NO L'Atalante, the pilots and team of the submersibles Nautile and Alvin, and the *Hot96* and *HOPE'99* cruises. This work was supported by CNRS, European grant (FEDER n° presage 3814) and the Conseil Régional de Bretagne (contrat n°A2C809). Sylvia Dewilde is a postdoctoral fellow of the FWO (Fund for scientific research flanders).

**Table 1:** Globin sequences shown in the multiple alignment and molecular phylogeny.

<b>Specie</b>	<b>Globin chain</b>	<b>Abbreviation</b>	<b>Accession number</b>
<i>Amphitrite ornata</i>	DHP A	DHP-Amph	AAF97245
<i>Arenicola marina</i>	A2a* B2* MbIIa	A2a-Are B2-Are MbIIa-Are	CAI56308 CAI56309 Kleinschmidt 1998
<i>Alvinella pompejana</i>	HbI*	HbInt-Alv	CAI56311
<i>Aphrodite aculeata</i>	nerve Mb	NMb-Aph	AAC47259
<i>Eudystilia vancouverii</i>	A1	A1-Eud	JC7606
<i>Glycera dibranchiata</i>	monomer IV P1	mIV-Gly P1-Gly	P02216 CAA37995
<i>Lamellibrachia sp.</i>	A2 B1 B2	A2-Lam B1-Lam B2-Lam	P15469 AAP40328 AAP04529
<i>Lumbricus terrestris</i>	A1 A2 B1 B2	A1-LumT A2-LumT B1-LumT B2-LumT	P08924 P02218 P11069 P13579
<i>Ophelia bicornis</i>	Mb	Mb-Oph	AAX73248
<i>Pheretima hilgendorfi</i>	A1	A1-Phehil	P83122
<i>Pheretima sieboldi</i>	A1	A1-Phesie	P11740
<i>Riftia pachyptila</i>	A1 A2 B1b B2	A1-Rif A2-Rif B1b-Rif B2-Rif	CAD29154 P80592 CAD29157 CAD29159
<i>Sabella spallanzanii</i>	A2 B2a	A2-Sab B2a-Sab	CAC37412 CAC37410
<i>Tubifex tubifex</i>	A1	A1-Tub	P18202
<i>Tylorrhynchus heterochaetus</i>	A1 A2 B2a	A1-Tyl A2-Tyl B2a-Tyl	P02219 P09966 P13578

\* indicates sequences presented in this work.



**Table 2:** Position, length of introns, exon/ intron splice junctions and possible branch points in *A. marina*, *A. pompejana*, *R. pachyptila*, *G. dibranchiata* and *A. ornata* Hb and DHP genes.

		Intron	Position	Size (bp)	Splicing donor	BP	Splicing acceptor
<b>Consensus</b>					HAG <b>gt</b> ragt	ctray	yyny <b>ag</b> GK
<i>A. marina</i>	A2	Intron 1	B12.2	370	GGC <b>gt</b> aagt	N.F	ctgt <b>ag</b> TA
		Intron 2	G7.0	773	GAT <b>gt</b> aagt	ctaaa	ctat <b>ag</b> CT
	B2	Intron 1	B12.2	487	GGA <b>gt</b> aagt	ctaac	ttac <b>ag</b> CC
		Intron 2	G7.0	495	GAC <b>gt</b> aagt	ataac	ctgc <b>ag</b> GA
<i>R. pachyptila</i>	A2	Intron 1	B12.2	495	CCA <b>gt</b> gagt	ctaata	ttgc <b>ag</b> TG
		Intron 2	G7.0	639	GAC <b>gt</b> aagc	cttat	cctc <b>ag</b> AC
	B2	Intron 1	B12.2	531	GAC <b>gt</b> aagc	cttaa	ttgc <b>ag</b> CC
		Intron 2	G7.0	390	TCT <b>gt</b> gagt	ctgac	ttgc <b>ag</b> CC
	A1	Intron 1	B12.2	505	AAA <b>gt</b> aagt	cttat	tgac <b>ag</b> CG
		Intron 2	G7.0	473	GTT <b>gt</b> aagt	gtcat	ttgc <b>ag</b> GT
	B1a	Intron 1	B12.2	1183	CGA <b>gt</b> aagt	ctcac	catc <b>ag</b> GC
		Intron 2	G7.0	947	GGA <b>gt</b> aagt	ctaata	tttc <b>ag</b> GC
	B1b	Intron 1	B12.2	601	GTA <b>gt</b> aagt	ctgac	tcac <b>ag</b> CA
		Intron 2	G7.0	655	CAG <b>gt</b> agag	ttaat	ttgc <b>ag</b> CT
	B1c	Intron 1	B12.2	1029	CAG <b>gt</b> ttgt	ctgac	ttgc <b>ag</b> AT
		Intron 2	G7.0	489	CAG <b>gt</b> aaag	ctaac	ttgc <b>ag</b> GT
<i>A. pompejana</i>	Hb	Intron 1	B12.2	135	GGA <b>gt</b> aagt	ctaac	accc <b>ag</b> GT
	Intra	Intron 2	G7.0	241	GCT <b>gt</b> aagt	ctaac	tttc <b>ag</b> GA
<i>A. marina</i>	Mb	Intron 1	B12.2	301	CTT <b>gt</b> aagt	ctcaa	ccac <b>ag</b> CC
		Intron 2	G7.0	500	ACT <b>gt</b> aagt	ctgac	cgcc <b>ag</b> GA
<i>G. dibranchiata</i>	mIV	Intron 1	B12.2	978	CAA <b>gt</b> aagt	ttgat	tttc <b>ag</b> GT
		Intron 2	G7.0	912	GAG <b>gt</b> aggt	ataac	tttc <b>ag</b> CC
<i>A. ornata</i>	DHP	Intron 1	B12.2	110	CGC <b>gt</b> aagc	ctcat	ctgc <b>ag</b> AT
		Intron 2	G7.0	2424	GAG <b>gt</b> gaat	atgac	cttc <b>ag</b> AA

**Note:** The consensus sequence presented are from Mount et al. (Mount 1982; Keller and Noon 1984). Sequences in italic correspond to branch points where the well conserved C or T or A is not found. **BP**, Branch Point, **N.F**, Not Found.

### Literatures Cited

- Allen, G. 1989. Sequencing of proteins and peptides. Pp 73-104 in R.H Burdon and P.H Van Knippenberg, eds. Laboratory techniques in biochemistry and molecular biology. Vol. 9. Elsevier North Holland, Amsterdam.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.
- Andersson, J. O., A. M. Sjogren, L. A. Davis, T. M. Embley and A. J. Roger. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr Biol.* **13**: 94-104.
- Bailly, X., D. Jollivet, S. Vanin, J. Deutsch, F. Zal, F. Lallier and A. Toulmond. 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.
- Bailly, X. 2003. Evolution moléculaire de la fonction de liaison du sulfure d'hydrogène par les hémoglobines de *Riftia pachyptila*, annélide tubicole des sources hydrothermales profondes. Thèse de doctorat. École Pratique des Hautes Etudes et Université Pierre et Marie Curie (Paris VI), France.
- Bendtsen, J. D., H. Nielsen, G. von Heijne and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* **340**: 783-795.
- Burmester, T., B. Weich, S. Reinhardt and T. Hankeln. 2000. A vertebrate globin expressed in the brain. *Nature.* **407**: 520-523.
- Burmester, T., B. Ebner, B. Weich and T. Hankeln. 2002. Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. *Mol Biol Evol.* **19**: 416-421.
- Chen, Y. P., S. A. Woodin, D. E. Lincoln and C. R. Lovell. 1996. An unusual dehalogenating peroxidase from the marine terebellid polychaete *Amphitrite ornata*. *J Biol Chem.* **271**: 4609-4612.
- 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.
- Dewilde, S., M. L. Van Hauwaert, S. Vinogradov, A. Vierstraete, J. Vanfleteren and L. Moens. 2001. Protein and gene structure of a chlorocruorin chain of *Eudistylia vancouverii*. *Biochem Biophys Res Commun.* **281**: 18-24.

- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792-1797.
- Freitas, T. A., S. Hou, E. M. Dioum, J. A. Saito, J. Newhouse, G. Gonzalez, M. A. Gilles-Gonzalez and M. Alam. 2004. Ancestral hemoglobins in Archaea. *Proc Natl Acad Sci U S A.* **101**: 6675-6680.
- Fushitani, K., O. Ochi and H. Morimoto. 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**: 267-273.
- 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*. *J Biol Chem.* **263**: 6502-6517.
- 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, O. and T. Suzuki. 1990. Molecular Assembly and Evolution of Multi-Subunit Extracellular Annelid Hemoglobins. *Zoolog Sci.* **7**: 1-16.
- 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.
- Hardison, R. 1998. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J Exp Biol.* **201 (Pt 8)**: 1099-1117.
- Harrington, J. P., G. Suarez, T. A. Borgese and R. L. Nagel. 1978. Subunit interactions of *Glycera dibranchiata* hemoglobin. *J Biol Chem.* **253**: 6820-6825.
- Harris, N. L. and P. Senapathy. 1990. Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. *Nucleic Acids Res.* **18**: 3015-3019.
- Hourdez, S., F. H. Lallier, M. C. De Cian, B. N. Green, R. E. Weber and A. Toulmond. 2000. Gas transfer system in *Alvinella pompejana* (Annelida polychaeta, Terebellida): functional properties of intracellular and extracellular hemoglobins. *Physiol Biochem Zool.* **73**: 365-373.
- Imai, K. 1999. The haemoglobin enzyme. *Nature.* **401**: 437, 439.
- Jhiang, S. M., J. R. Garey and A. F. Riggs. 1988. Exon-intron organization in genes of earthworm and vertebrate globins. *Science.* **240**: 334-336.

- Jhiang, S. M. and A. F. Riggs. 1989. The structure of the gene encoding chain c of the hemoglobin of the earthworm, *Lumbricus terrestris*. J Biol Chem. **264**: 19003-19008.
- Jones, D. T., W. R. Taylor and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. **8**: 275-282.
- Keller, E. B. and W. A. Noon. 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. Proc Natl Acad Sci U S A. **81**: 7417-7420.
- Kleinschmidt, T. and R. E. Weber. 1998. Primary structures of *Arenicola marina* isomyoglobins: molecular basis for functional heterogeneity. Biochim Biophys Acta. **1383**: 55-62.
- LaCount, M. W., E. Zhang, Y. P. Chen, K. Han, M. M. Whitton, D. E. Lincoln, S. A. Woodin and L. Lebiada. 2000. The crystal structure and amino acid sequence of dehaloperoxidase from *Amphitrite ornata* indicate common ancestry with globins. J Biol Chem. **275**: 18712-18716.
- Mangum, C. P., B. R. Woodin, C. Bonaventura, B. Sullivan and J. Bonaventura. 1975. The role of coelomic and vascular hemoglobin in the annelid family terebellidae. Comp Biochem Physiol. **51A**: 281-294.
- Mangum, C. P. 1985. Oxygen transport in invertebrates. Am J Physiol. **248**: R505-514.
- Manwell, C. and C. M. A. Baker. 1988. *Magelona* haemerythrin: tissue specificity, molecular weights and oxygen equilibria. Comp Biochem Physiol. **89B**: 453-463.
- Moens, L., J. Vanfleteren, Y. Van de Peer, K. Peeters, O. Kapp, J. Czeluzniak, M. Goodman, M. Blaxter and S. Vinogradov. 1996. Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationships. Mol Biol Evol. **13**: 324-333.
- Mount, S. M. 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**: 459-472.
- Negrisolò, 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.
- Pesce, A., M. Bolognesi, A. Bocedi, P. Ascenzi, S. Dewilde, L. Moens, T. Hankeln and T. Burmester. 2002. Neuroglobin and cytoglobin. Fresh blood for the vertebrate globin family. EMBO Rep. **3**: 1146-1151.
- Riggs, A. F. 1991. Aspects of the Origin and Evolution of Non-Vertebrate Hemoglobins. Amer. Zool. **31**: 535-545.

- Roach, M. P., Y. P. Chen, S. A. Woodin, D. E. Lincoln, C. R. Lovell and J. H. Dawson. 1997. Notomastus lobatus chloroperoxidase and Amphitrite ornata dehaloperoxidase both contain histidine as their proximal heme iron ligand. *Biochemistry*. **36**: 2197-2202.
- Seamonds, B., R. E. Forster and A. J. Gottlieb. 1971. Heterogeneity of the hemoglobin from the common bloodworm *Glycera dibranchiata*. *J Biol Chem*. **246**: 1700-1705.
- Strand, K., J. E. Knapp, B. Bhyravbhatla and W. E. Royer, Jr. 2004. Crystal structure of the hemoglobin dodecamer from *Lumbricus erythrocruorin*: allosteric core of giant annelid respiratory complexes. *J Mol Biol*. **344**: 119-134.
- Suzuki, T., O. H. Kapp and T. Gotoh. 1988. Novel S-S loops in the giant hemoglobin of *Tylorrhynchus heterochaetus*. *J Biol Chem*. **263**: 18524-18529.
- Suzuki, T., T. Takagi and S. Ohta. 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**: 141-146.
- Suzuki, T. and K. Imai. 1998. Evolution of myoglobin. *Cell Mol Life Sci*. **54**: 979-1004.
- Swank, R. T. and K. D. Munkres. 1971. Molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with sodium dodecyl sulfate. *Anal Biochem*. **39**: 462-477.
- Terwilliger, R. C., R. L. Garlick and N. B. Terwilliger. 1980. Characterization of hemoglobins and myoglobin of *Travisia foetida*. *Comp Biochem Physiol*. **66B**: 261-266.
- Toulmond, A. 1992. Properties and functions of extracellular heme pigments. In *Advance in comparative and environmental physiology* (Mangum, C.P., ed.), Vol 13. 231-256.
- Vinogradov, S. N., D. A. Walz, B. Pohajdak, L. Moens, O. H. Kapp, T. Suzuki and C. N. Trotman. 1993. Adventitious variability? The amino acid sequences of nonvertebrate globins. *Comp Biochem Physiol B*. **106**: 1-26.
- Vinogradov, S. N., D. Hoogewijs, X. Bailly, R. Arredondo-Peter, M. Guertin, J. Gough, S. Dewilde, L. Moens and J. R. Vanfleteren. 2005. Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc Natl Acad Sci U S A*. **102**: 11385-11389.
- Weber, R. E. and E. Pauptit. 1972. Molecular and functional heterogeneity in myoglobin from the polychaete *Arenicola marina* L. *Arch Biochem Biophys*. **148**: 322-324.

- Weber, R. E., C. Mangum, H. Steinman, C. Bonaventura, B. Sullivan and J. Bonaventura. 1977. Hemoglobins of two terebellid polychaetes: *Enoplobranchus sanguineus* and *Amphitrite ornata*. *Comp Biochem Physiol A*. **56**: 179-187.
- Weber, R. E. and S. N. Vinogradov. 2001. Nonvertebrate hemoglobins: functions and molecular adaptations. *Physiol Rev*. **81**: 569-628.
- Williams, E. J., C. Pal and L. D. Hurst. 2000. The molecular evolution of signal peptides. *Gene*. **253**: 313-322.
- Wittenberg, B. A., R. W. Briehl and J. B. Wittenberg. 1965. Haemoglobins of invertebrate tissues. Nerve haemoglobins of *Aphrodite*, *Aplysia* and *Halosydna*. *Biochem J*. **96**: 363-371.
- Wittenberg, J. B. 1970. Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. *Physiol Rev*. **50**: 559-636.
- Zafar, R. S., L. H. Chow, M. S. Stern, J. S. Scully, P. R. Sharma, S. N. Vinogradov and D. A. Walz. 1990a. The cDNA sequences encoding two components of the polymeric fraction of the intracellular hemoglobin of *Glycera dibranchiata*. *J Biol Chem*. **265**: 21843-21851.
- Zafar, R. S., L. H. Chow, M. S. Stern, S. N. Vinogradov and D. A. Walz. 1990b. The heterogeneity of the polymeric intracellular hemoglobin of *Glycera dibranchiata* and the cDNA-derived amino acid sequence of one component. *Biochim Biophys Acta*. **1041**: 117-122.
- Zhu, H. and A. F. Riggs. 1992. Yeast flavohemoglobin is an ancient protein related to globins and a reductase family. *Proc Natl Acad Sci U S A*. **89**: 5015-5019.

## **Chapitre 6**

**Troisième article:**

**Gene structure and molecular phylogeny of  
the linker chains from the giant annelid  
hexagonal bilayer hemoglobins**

## **Chapitre 6: Troisième article**

### **Gene structure and molecular phylogeny of the linker chains from the giant annelid hexagonal bilayer hemoglobins.**

Soumis le 17/08/2005 à Journal of Molecular Evolution

Les HBL-Hb d'annélides sont composées de deux types de chaînes polypeptidiques : les globines, dont les fonctions ont été détaillées dans les chapitres précédents, et les Linkers. Les linkers sont impliqués dans la mise en place et le maintien de la structure quaternaire des HBL-Hbs. La présence d'ions zinc ( $Zn^{2+}$ ) et cuivre ( $Cu^{2+}$ ) a aussi permis de suggérer que les linkers pouvaient présenter une activité superoxyde dismutase (SOD), assurant la protection de l'hémoglobine extracellulaire contre les ions superoxydes  $O_2^-$  (Liochev et al. 1996).

L'analyse des séquences primaires de linkers a révélé la présence d'un domaine particulier, similaire à celui rencontré dans le domaine de liaison aux ligands des récepteurs aux lipoprotéines de faible densité (Low density Lipoprotein receptor ou LDLR), appelé LDL-A (répétition de type A des LDLR) (Sudhof et al. 1985a; Suzuki and Riggs 1993). Ce domaine est composé d'un motif conservé présentant six cystéines ainsi que plusieurs résidus acides (acide glutamique et acide aspartique) invariables. Des travaux réalisés sur les LDL-A des LDLR humain ont mis en évidence l'implication de ces résidus acides dans la liaison d'ions calcium ( $Ca^{2+}$ ) nécessaires pour assurer un repliement correct pour la fonction des LDL-A au sein des LDLR (Fass et al. 1997; Rudenko et al. 2002; Simonovic et al. 2001). Ces interactions pourraient expliquer le besoin d'ions  $Ca^{2+}$  au cours des mécanismes d'assemblage des HBL-Hbs, ceux-ci permettant aux linkers d'avoir une structure capables de lier les dodécamères de globines (Kuchumov et al. 2000; Kuchumov et al. 1999).

De plus, les similarités de séquences entre les domaines LDL-A des linkers et ceux des autres membres de la famille des LDLR laissent supposer que les six cystéines présentes dans les modules LDL-A peuvent être impliquées dans les mêmes ponts disulfures que ceux retrouvés dans les modules LDL-A des LDLR humains (Bieri et al. 1995).

Dans un premier temps, nous avons obtenu, soit par séquençage nucléotidique soit par recherche dans les banques de données, de nouvelles séquences de linkers chez les annélides *Arenicola marina* (polychaete), *Riftia pachyptila* (vestmentifère), *Alvinella pompejana* (polychaete), *Lumbricus rubellus* (oligochaete) et *Eiseinia andrei* (oligochaete). Les



phylogénies moléculaires obtenues à partir de ces séquences nous ont amené à envisager l'existence de deux familles distinctes au sein des linkers.

Dans un second temps, nous nous sommes intéressés à la structure des gènes codant les linkers. Seules deux structures partielles avaient été déterminées, chez *Lumbricus terrestris* et *Neanthes diversicolor* (Suzuki et al. 1994; Suzuki and Riggs 1993). Ces gènes présentaient un intron précédant le module LDL-A et un intron dans la partie 3'UTR. Nous avons déterminé que la position de ce deuxième intron n'était pas conservée entre toutes les séquences étudiées. Nous avons aussi révélé la présence de deux introns supplémentaires. Ainsi, le module LDL-A est flanqué par deux introns, dont les positions sont identiques à celles trouvées dans les domaines de liaison aux ligands des LDLR.

Ces résultats remettent en question l'hypothèse selon laquelle les gènes codant pour les linkers sont issus de la duplication d'un gène de globine suivie par l'insertion de la séquence codant pour le domaine LDL-A (mécanisme d'exon shuffling) (Suzuki and Riggs 1993).

## Gene structure and molecular phylogeny of the linker chains from the giant annelid hexagonal bilayer hemoglobins

Christine Chabasse<sup>1</sup>, Xavier Bailly<sup>1</sup>, Sophie Sanchez<sup>1</sup>, Morgane Rousselot<sup>1</sup> and Franck Zal<sup>1</sup>.

<sup>1</sup>Equipe Ecophysiologie: Adaptation et Evolution Moléculaires, UPMC – CNRS UMR 7144, Station Biologique, BP 74, 29682 Roscoff cedex, France

Key words: gene structure, evolution, linker, annelid, LDL receptor, LDL-A.

Abbreviations: Hb, hemoglobin; HBL-Hb; Hexagonal Bilayer Hemoglobin; LDL, Low Density Lipoprotein; LDLR, Low Density Lipoprotein Receptor; LDL-A, Low density lipoprotein Receptor Class A repeat.

### Abstract

The giant extracellular hexagonal bilayer hemoglobin (HBL-Hb) found only in annelids, is a ~3500 kDa heteropolymeric structure, involved in both oxygen and hydrogen sulfide transport. The HBL-Hbs are comprised of globin and linker chains, the latter being required for the assembly of the quaternary structure. The linker chains, varying in size from 225 to 283 amino-acids, have a conserved cysteine-rich domain within their N-terminal moiety that is homologous to the cysteine rich modules constituting the ligand binding domain of the Low Density Lipoprotein Receptor (LDLR) protein family found in many metazoans. We have investigated the gene structure of linkers from *Arenicola marina*, *Alvinella pompejana*, *Nereis diversicolor*, *Lumbricus terrestris* and *Riftia pachyptila*. We found, contrary to the results obtained earlier with linker genes from *N. diversicolor* and *L. terrestris*, that in all of the foregoing cases, the linker LDL-A module is flanked by two phase 1 introns, as in the human LDLR gene, with two more introns in the 3' side whose positions varied with the species. In addition, we obtained thirteen linker cDNAs that have been determined experimentally or found in the EST database LumbriBASE. A molecular phylogenetic analysis of the linker primary sequences demonstrated that they cluster into two distinct families of linker proteins.

We propose that the common gene ancestor to annelid linker genes exhibited a 4 introns and 5 exons structure and gave rise to the two families subsequent to a duplication event.

## **Introduction**

The giant extracellular hemoglobins (Hbs) of annelids, called hexagonal bilayer hemoglobins (HBL-Hbs) because of their characteristic quaternary structure (Lamy et al. 1996; Vinogradov 1985; Weber and Vinogradov 2001), have a molecular mass of ~3.5 MDa, and are comprised of globin and linker chains. Based on the crystallographic structure of *Lumbricus terrestris* HBL-Hb, there are 144 globin chains (~17 kDa) that reversibly bind oxygen and others gaseous ligands (Imai 1999; Vinogradov et al. 1993; Weber and Vinogradov 2001), and 36 linker chains (24-32 kDa) that are devoid of heme and are required, together with the presence of  $\text{Ca}^{2+}$ , to form the HBL complex (Gotoh et al. 1998; Kuchumov et al. 1999; Lamy et al. 2000; Vinogradov et al. 1986).

In contrast to the substantial number of annelid extracellular globin sequences that have become available over the last 20 years (Bailly et al. 2002; Gotoh et al. 1987), only ten linker amino-acid sequences are known from the polychaetes *Neanthes diversicolor* (Suzuki et al. 1994), *Tylorrhynchus heterochaetus* (Suzuki et al. 1990a) and *Sabella spallanzanii* (Pallavicini et al. 2001), the vestimentiferan *Lamellibrachia sp.* (Suzuki et al. 1990b), the oligochaete *Lumbricus terrestris* (Fushitani et al. 1996; Suzuki and Riggs 1993) and the hirudinean *Macrobodella decora* (Suzuki and Vinogradov 2003). The linker chains vary from 185 to 255 amino-acids and share a 39-residue cysteine-rich module within their N-terminal moiety, that is very similar to Low Density Lipoprotein Receptor class A repeats (LDL-A, also called complement-type repeats). These LDL-A modules are found in the N-terminal region of the Low Density Lipoprotein Receptor (LDLR) in many metazoans, such as *Homo sapiens* (Sudhof et al. 1985a), *Xenopus laevis* (Mehta et al. 1991) and *Caenorhabditis elegans* (Chen et al. (31 co-authors). 2005). LDL-A modules are also found in other members of the LDLR superfamily, including Very Low Density Lipoprotein receptor (VLDLR) (Takahashi et al. 1992), LDLR- Related Protein (LRP), and in diverse LDLR unrelated functional proteins such as C9 or C8 complementary factor (DiScipio et al. 1984), the renal glycoprotein gp330 (Raychowdhury et al. 1989; Saito et al. 1994), the receptor for subgroup A Rous sarcoma virus (Bates et al. 1993) and enterokinase (Kitamoto et al. 1994). The LDL-A modules of these proteins exhibit six invariant cysteine residues and highly conserved acidic

residues (aspartatic acid and glutamatic acid). It has been suggested that these acidic residues are involved in the ligand binding through ionic interactions with the basic residues of the ligand (Brown and Goldstein 1986; Guo et al. 2004; Mahley 1988; Sudhof et al. 1985b). Moreover, it has been shown that several of these LDL-A modules bind  $\text{Ca}^{2+}$  which is essential for the maintenance of their biologically active conformations (Dirlam et al. 1996; Guo et al. 2004; van Driel et al. 1987). Based on structural similarities, it has been inferred that such interactions involving  $\text{Ca}^{2+}$  and acidic interactions would explain the molecular mechanisms for the subunit assembly in HBL-Hbs (Suzuki and Riggs 1993). In the case of *Lumbricus terrestris* HBL-Hb, it is likely that the binding of  $\text{Ca}^{2+}$  to the linker module acidic groups makes the linkers conformationally competent to bind the dodecamer globin subunits to form the HBL structure (Kuchumov et al. 2000; Kuchumov et al. 1999).

The evolutionary origin of linkers in annelids remains unclear. Suzuki et al. have proposed that functional linkers could result from globin gene duplication and exon shuffling processes, including the insertion of an LDL-A module (Suzuki and Riggs 1993; Suzuki et al. 1990b). In this paper we report the cDNA sequence and gene structure of a linker chain from the polychaete *Arenicola marina*, two partial cDNA linker sequences and gene structures from the vestimentiferan *Riftia pachyptila*, three cDNA linker sequences and partial gene structures from a deep-sea hydrothermal vent polychaete *Alvinella pompejana*, the revised gene structures of *N. diversicolor* linker L2 (Suzuki et al. 1994) and *L. terrestris* L1 (Suzuki and Riggs 1993), and six linker sequences found in expressed sequence tag (EST) database (LumbriBASE: <http://www.earthworms.org/>) from two oligochaetes *Lumbricus rubellus* and *Eiseinia andrei* (Lee et al. 2004). Our results indicate the presence of additional introns in the linker genes from *A. marina*, *R. pachyptila*, *A. pompejana*, *N. diversicolor* and *L. terrestris* which were not described in an earlier studies of *L. terrestris* and *N. diversicolor* linker gene structure (Suzuki and Riggs 1993). Furthermore, a molecular phylogenetic analysis of the 22 annelid linker sequences demonstrates that they cluster into two groups of linkers, which have probably evolved via duplication of a single ancestral linker gene. Since no LDLR or VLDLR related genes were found in the databank of annelids EST other than the LDL-A module-like in the HBL-Hb linker genes, it appears that the latter represent a rare case of a phylum-specific gene, whose recruitment was a functional and structural innovation restricted to annelids.

## **Material and Methods**

### *Collection of biological material*

Specimens of the polychaetes *Arenicola marina* and *Neanthes diversicolor* were collected at low tide from a sandy shore near Roscoff (Penpoull beach), Nord Finistère, France, and kept in local running sea water for 24h. Juvenile specimens of the hydrothermal vent tube worm *Riftia pachyptila* were collected at the vent site from the ridge segment 9°50'N on the EPR (Riftia field: 9°50.75'N, 104°17.57'W) at a depth of about 2.500m, during the French *HOT 96* and the American *LARVE99* oceanographic cruises. Additional individuals were collected on the South EPR during the Franc oceanographic cruise *BIOSPEEDO* (Riftia field: 17°25.47'S, 113°12.281'W). The worms were sampled using the telemanipulated arms of the submersibles *Nautile* and *Alvin*, brought back alive to the surface in a thermally-insulated basket, and immediately frozen and stored in liquid nitrogen after their recovery on board.

Specimens of *Alvinella pompejana* were collected at 2.500 m depth on the East Pacific Rise (9°50'N) by the manned submersible *Nautile* during the *HOPE'99* cruise. Once on board, the animals were immediately frozen and stored in liquid nitrogen until used.

Specimens of *Lumbricus terrestris* were collected at St Kirio (Ploujean, Nord Finistère, France) and frozen at -80°C until used.

### *Total RNA extraction and cDNA synthesis*

Entire specimens of *A. marina*, *A. pompejana* and *R. pachyptila* were crushed while submerged in liquid nitrogen. Total RNAs were extracted using the RNAble® buffer (Eurobio) and poly-A RNAs were then isolated using mRNA Purification Kit® (Amersham). A reverse transcriptase PCR was carried out using an anchor 5'-CTC CTC TCC TCT CCT CTT CCT oligo(dT)<sub>17</sub> primer.

### *Isolation of Genomic DNA*

Specimens of *A. marina*, *A. pompejana*, *N. diversicolor*, *R. pachyptila* and *L. terrestris* were incubated in 700µL of a PK Buffer (50mM Tris/HCl containing 100mM NaCl, 25mM EDTA and 1% SDS, pH 8) with 15 µL Proteinase K (10µg/µl) at 65°C for 1h. The supernatant was recovered after a centrifugation at 12 000g for 5 min at 4°C, and extracted with 700µL phenol. The material was extracted once with phenol/ chloroform/ isoamyl alcohol (25/24/1), and once with chloroform. The resulting DNA was precipitated with isopropanol at -20°C

overnight, and washed once with 75% Ethanol. DNA was finally resuspended in 100µl TE buffer (10mM Tris/HCl, 0.1mM EDTA, pH 8) and stored at 4°C.

#### *Amplification of cDNA and genomic DNA*

The PCR reaction was carried out in a 25µL volume containing 10-50 ng cDNA/ gDNA template, 100 ng of each degenerate primer, 200µM dNTPs, 2.5mM MgCl<sub>2</sub> and 1 Unit DNA polymerase (Uptima, Interchim). PCR conditions were as follows: an initial denaturation step at 95°C for 5 min, 35 cycles consisting of denaturation at 95°C for 40s, annealing at 58°C for 40s, extension at 72°C for 50s, and a final elongation step at 72°C for 10 min. Primers used are available upon request.

#### *Cloning and sequencing*

The PCR products were the cloned using a TOPO<sup>®</sup>-TA cloning Kit (Invitrogen). The positive recombinant clones were isolated and plasmid DNA was prepared by the FlexiPrep Kit (Amersham). Purified plasmids containing the putative globin insert were used in a dye-primer cycle sequencing reaction, using the Big Dye<sup>®</sup> Terminator V3.1 Cycle Sequencing kit (Applied Biosystems). PCR products were subsequently run on 3100 Genetic Analyser (Applied Biosystems) at Roscoff sequencing core facility Ouest genopole<sup>®</sup> Plateform.

#### *Rapid amplification of cDNA ends*

cDNAs ends were obtained by PCR using the 5' and 3' RACE kit from Roche according to the manufacturer's instructions. Buffer, reagents and other conditions for the nested PCR were as described by the manufacturer. The RACE products were purified, cloned with the TOPO-TA cloning kit (Amersham) and sequenced as described above.

#### *Sequence analysis*

The peptide signal cleavage site was predicted by the SignalP 3.0 Server (Bendtsen et al. 2004) (<http://www.cbs.dtu.dk/services/SignalP/>).

#### *Linker multiple alignment construction*

Analyses were performed on the linker chains of *Lumbricus terrestris* L1 (AAF99389) and L3 (S65723), *Tylorrhynchus heterochaetus* L1 (P18207) and L2 (P18208), *Lamellibrachia sp.* LAV-1 (A35012), *Sabella spallanzanii* L1 (CAB38536) and L3 (CAC37413), *Macrobdella decora* L1 (BAC82449) and *Neanthes diversicolor* L2 (BAA09580).

Others sequences have been found in LumbriBASE (<http://www.earthworms.org/>): *Lumbricus rubellus* L1 (CO046524), L2 (CO046717), L3 (CF839233) and L4 (BF422664 and BG269978), and *Eisenia andrei* L1 (BP524390), L2 (BP524387) and L3 (BP524672).

Amino-acid sequences were aligned with MUSCLE program (Edgar 2004) ([http://phylogenomics.berkeley.edu/cgi-bin/muscle/input\\_muscle.py](http://phylogenomics.berkeley.edu/cgi-bin/muscle/input_muscle.py)) and verified manually.

### *Molecular phylogeny*

An unrooted phylogenetic tree was constructed using the Neighbor-Joining (NJ) method from the linker amino-acids sequences multiple alignment. Gapped domains ambiguously aligned were ignored (Fig.1). The tree was computed using the PHYLIP programs package version 3.63 (J. Felsenstein, Phylogeny Inference package, version 3.63; Department of Genetics, University of Washington, Seattle) with 1,000 bootstrap reiterations and the JTT transition matrix (Jones et al. 1992).

## **Results**

### *Characterization of new linker sequences*

#### *Arenicola marina*

There are two different linker chains, L1 and L2, involved in homo- and heterodimer formation, with a mass of 25,174.1 and 26,829.7 Da respectively (Zal et al. 1997b). The *A. marina* linker presented here (AM000028) contains an open reading frame of 256 codons including a signal peptide (residues 1 to 16). The cDNA-derived amino-acid sequence without the signal peptide has a calculated mass of 26,740.7 Da. We assigned this amino-acid sequence to linker L2 even though it exhibits a difference of 89 Da with this polypeptide. This difference may be explained either by the sequencing of an allelic form due to the fact that *A. marina* specimens were harvested in different areas, or to post-translation modifications, such as carboxylation of two glutamic or aspartic acid residues. The cDNA-derived amino-acid sequence exhibits 12 cysteine residues, which is in agreement with previous studies on *A. marina* HBL-Hb (Zal et al. 1997b).

#### *Riftia pachyptila*

The vestimentiferan *Riftia pachyptila* possesses four types of linker chains (Zal et al. 1996). Two partial linker sequences (AM000032 and AM000033) have been identified and

arbitrarily called LX and LY. The signal peptide of the LY sequences includes residues 1 to 19.

*Alvinella pompejana*:

The extracellular HBL-Hb of the polychaete *Alvinella pompejana* involves four type of linker chains (L1 to L4) (Zal et al. 1997a). Three linker sequences have been sequenced here (AM000029, AM000030 and AM000031). The first sequence contains an open reading frame of 225 amino-acids, with the first 19 amino-acids characterized as a signal peptide. The molecular mass calculated for the polypeptide deduced from the cDNA is 22,895 Da. The amino-acid sequence exhibits 10 cysteine residues, which is in agreement with previous studies on *A. pompejana* HBL-Hb, and corresponds to linker L1 (*Mr* 22,889 Da) (Zal et al. 1997a). Two other partial cDNA sequences have been found and arbitrarily called L2 and L3.

*Eisenia andrei*

Three linker sequences have been found in the EST database LumbriBASE and have been called L1 (BP524390), L2 (BP524387) and L3 (BP524672) (Lee et al. 2004). The amino-acid sequence derived from the L3 cDNA is composed of 240 amino-acids, with the first 19 amino-acids constituting the signal peptide. The L1 and L2 sequences are not complete, with the signal peptides comprising the first 20 and 16 residues, respectively.

*Lumbricus rubellus*

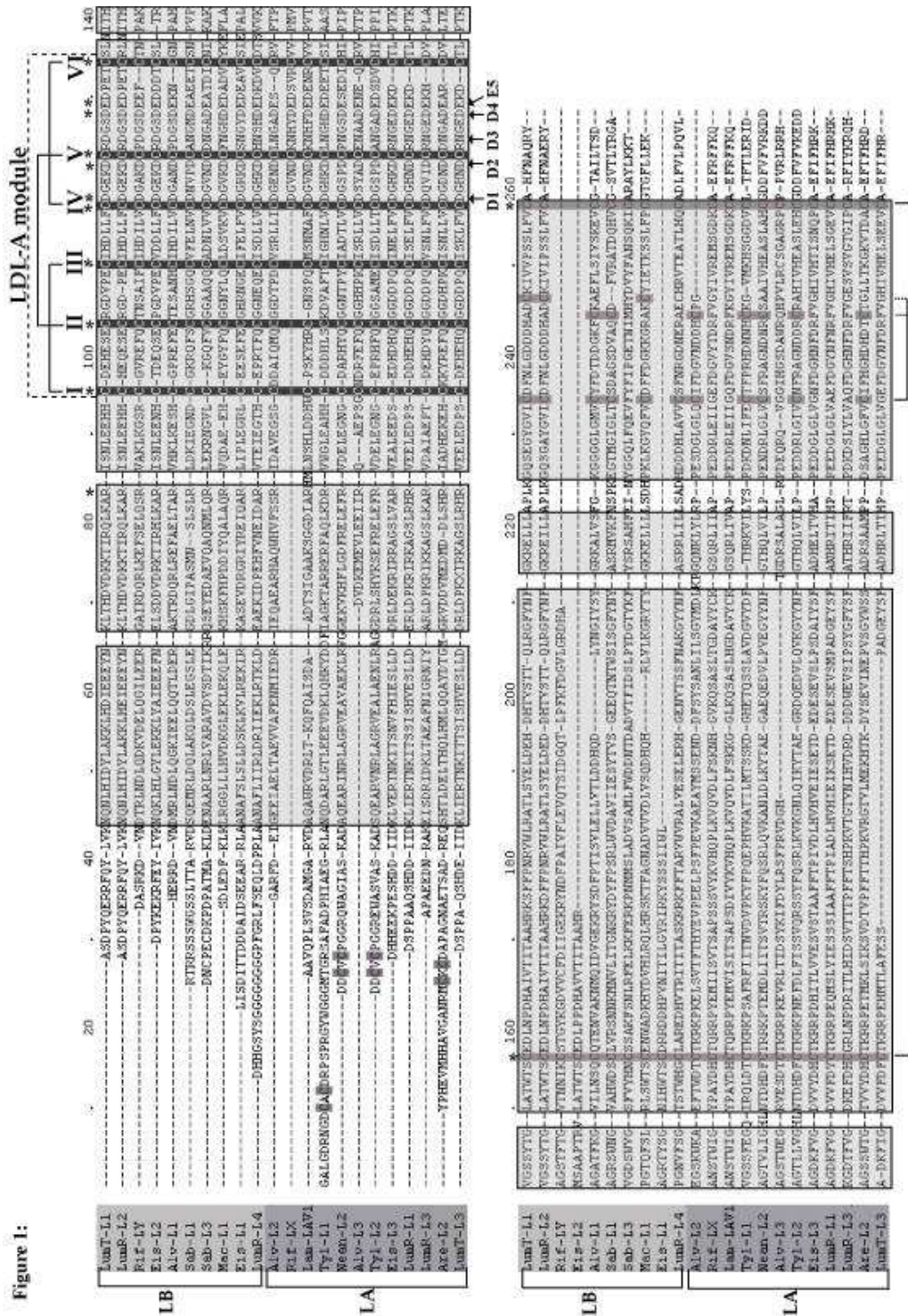
Four complete linker sequences: L1 (CO046524), L2 (CO046717), L3 (CF839233) and L4 (BF422664 and BG269978) have been found in the EST database LumbriBASE. The sequences contain an open reading frame of 236, 241, 243 and 283 amino-acids respectively. For each sequence, the signal peptide has been identified and corresponds to the residues 1 to 15 for L1, 1 to 16 for L2, 1 to 26 for L3 and 1 to 16 for L4. The calculated masses are 24,950.5 Da, 25,0905.8 Da, 24,455.2 Da and 29,983.4 Da, respectively.

*Multiple alignment*

A multiple alignment of the linker amino-acid sequences with the signal peptide removed is shown in Fig.1. Thirteen residues appear to be invariant: Arg-83, Cys-95, Cys-103, Cys-110, Cys-117, Asp-118, Asp-122, Cys-123, Asp-128, Glu-129, Cys-134, Cys-156 and Cys-257, in agreement with previous results (Suzuki, et al, 1990, 1994; (Suzuki, 2003 #238). A remarkable feature of the linker chain is the conservation of a 39-residue cysteine-rich domain (at positions 95-134), comprised of a repeating pattern of cysteinyl residues: C-X<sub>5-7</sub>-C-X<sub>5-6</sub>-C-X<sub>6</sub>-C-D-X<sub>3</sub>-D-C-X<sub>4</sub>-D-E-X<sub>2-4</sub>-C.



The linker sequences tend to form two clusters: one shares an Asp154, a 5-residue Tyr-Lys-Arg-Lys/Arg-Pro motif at 157-161, a Ser170 and a 2-residue motif Pro-Glu/Asp at positions 225-226, while the second one shares linkers with the Trp-Thr-Ser motif at position 153-155. *A. marina* linker L2, *R. pachyptila* linker LX, *A. pompejana* L2 and L3, *L. rubellus* L1 and L3 and *E. andrei* L3 belong to the first group, and *R. pachyptila* linker LY, *A. pompejana* L1, *L. rubellus* L2 and L4, and *E. andrei* L1 and L2 belong to the second cluster (Fig.1).



**Figure 1:** Multiple alignment of linker amino-acid sequences of *Lumbricus terrestris* (LumT), *Lumbricus rubellus* (LumR), *Eiseinia andrei* (Eis), *Alvinella pompejana* (Alv), *Riftia pachyptila* (Rif), *Sabella spallanzanii* (Sab), *Macrobdella decora* (Mac), *Lamelibrachia sp.* (Hou et al.), *Tylorrhynchus heterochaetus* (Tyl), *Neanthes diversicolor* (Nean) and *Arenicola marina* (Are). Signal peptides, when present, have been removed. The conserved residues are marked by an asterisk. The LDL-A module is indicated by dashed lines. The Roman numerals refer to the order of appearance of the cysteine residues in the LDL-A module. Solid lines indicated the possible intra-chain disulphide bonds. Grey boxes correspond to the restricted alignment used for the phylogeny reconstruction.

### *Linker gene structure*

The complete gene structure of *A. marina* linker and partial gene structures of *R. pachyptila* LX and *A. pompejana* L1 have been determined, as well as the revised gene structure from *N. diversicolor* L2 and *L. terrestris* L1 (Fig.2a and 2b).

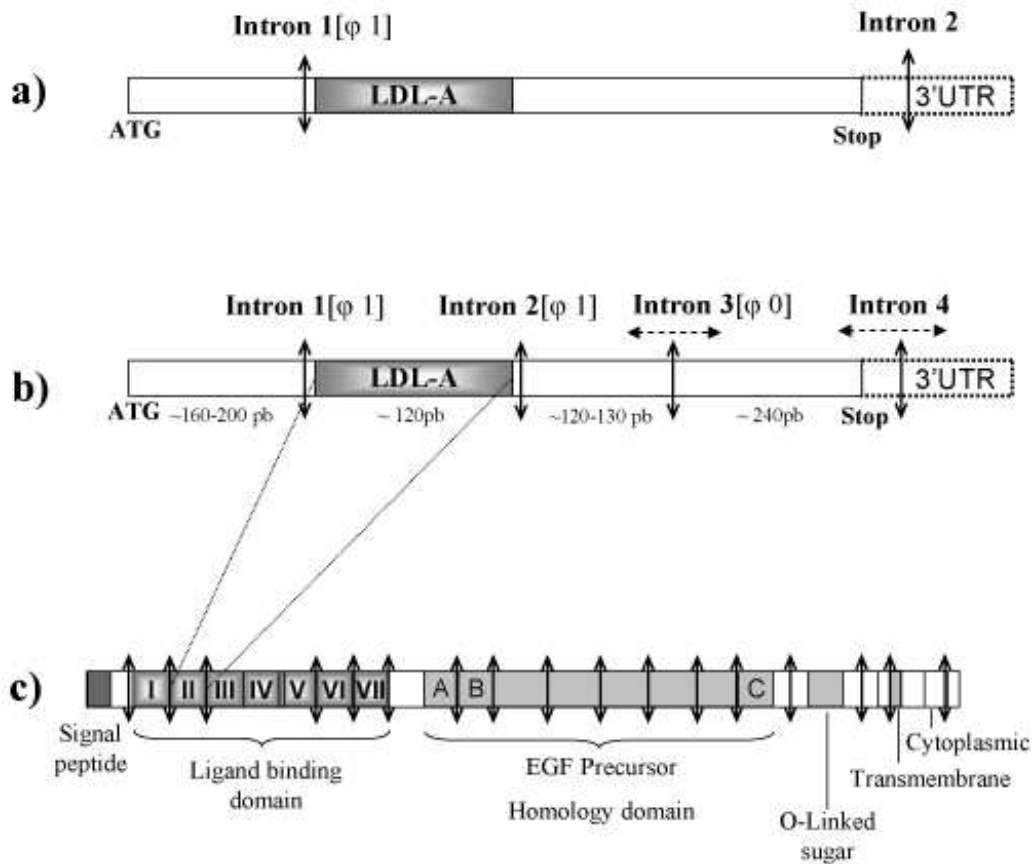
In *A. marina* linker sequence, a first intron (434 pb) splits the codon at position 91 in phase 1 (Fig.2).

A second intron position has been identified in *A. marina* linker (348 pb), *N. diversicolor linker L2* (347 pb) and *L. terrestris L1* (951 pb) and splits the codon in phase 1 at position 135 in the alignment (Fig.1). In *R. pachyptila* linker, the splice site is at the same position, but the intron sequence is not complete.

A third intron position, variable among the species studied here, has been identified between codon 182 and 183 in *A. marina* linker (562 pb), between codon 178 and 179 in *N. diversicolor L2* (391 pb), between codon 193 and 194 in *A. pompejana L1* (382 pb) and between codon 181 and 182 in *R. pachyptila linker LX* (673 pb). They all split the codon in phase 0. No intron has been found in this region in *L. terrestris L1*.

A fourth intron position, also variable among the species studied here, has been found at position 266 in phase 1 in *A. marina linker L2* (757 pb). In *N. diversicolor*, the fourth intron (359 pb) is found 4 nucleotides downstream from the stop codon. In *R. pachyptila linker LX*, the fourth intron (657 pb) is found just after the stop codon.

The splice junctions have been analyzed and are in agreement with consensus sequences (*i.e.* the splicing donor GT and acceptor AG are present).



**Figure 2:** a) Linker gene structure as described by Suzuki et al. (Suzuki and Riggs 1993). b) Present revised linker gene structure. c) Human LDLR gene structure (Sudhof et al. 1985a). Black arrows indicate intron positions.

*Phylogenetic analysis of annelid linker chains*

The neighbor-joining tree is shown in Fig.3. Two main clusters separated by a node with a bootstrap-value of 50 can be identified: cluster LA and cluster LB. These two clusters also correspond to the two group identified from the multiple alignment (Fig.1).

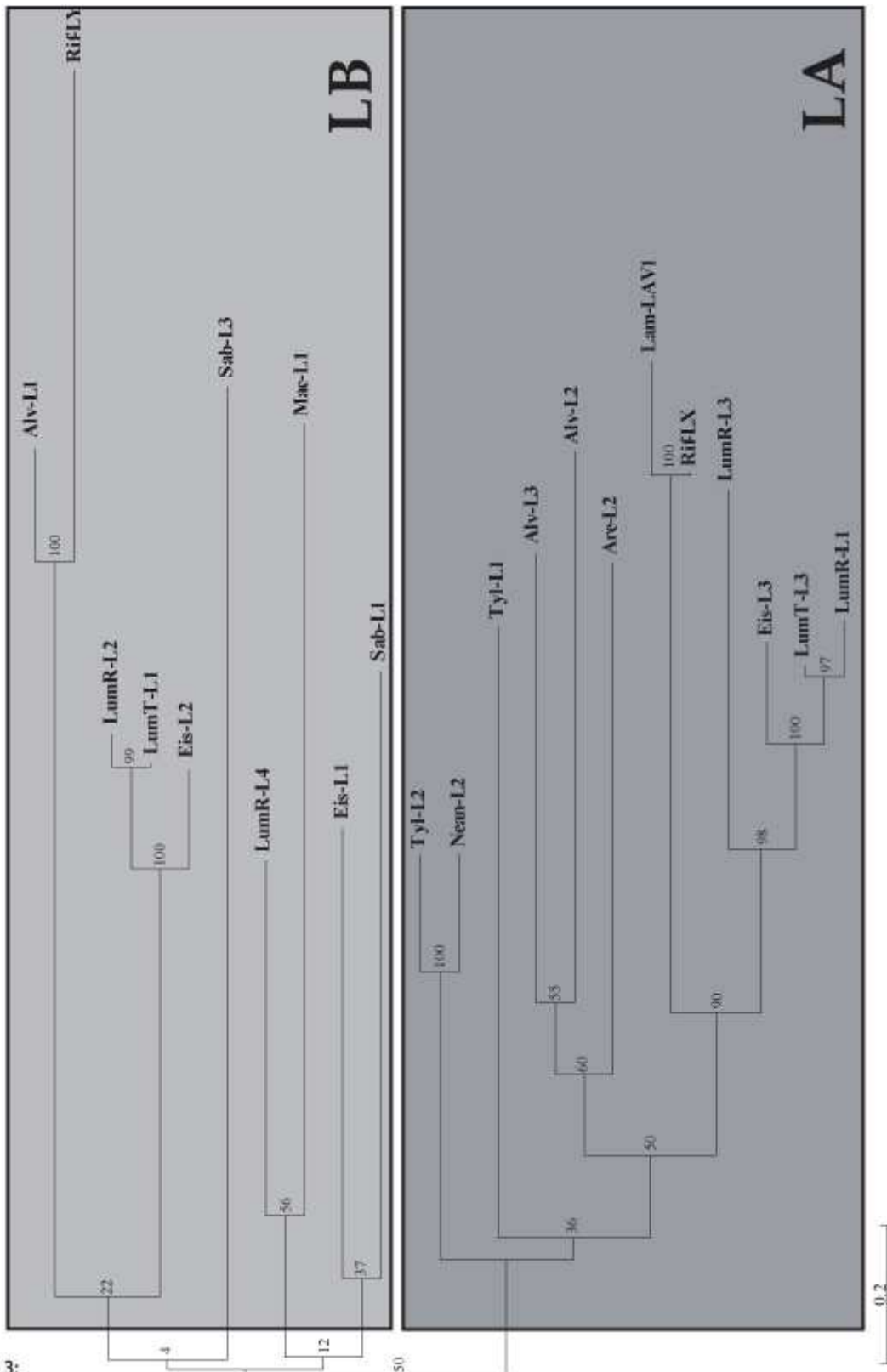


Figure 3:

**Figure 3:** Neighbor-joining tree of linker amino-acids sequences obtained with 1,000 bootstraps reiterations. The two clusters are called by LA and LB. Bootstraps values are indicated. See Figure 1 for abbreviations.

## **Discussion**

### *Functional and structural inferences from linker sequences*

#### *Cysteine residues motif in LDL-A modules*

Six of the ten to twelve cysteine residues found in linker sequences are part of the cysteine-rich module common to all linkers and the LDL-A modules comprising the ligand binding domains of the LDLR protein superfamily and similar modules in several unrelated proteins. The structure of the mammalian LDL-A module has been determined (Bieri et al. 1995), showing that the connectivity of the three disulfide bridges is Cys(I)-Cys(III), Cys(II)-Cys(V) and Cys(IV)-Cys(VI), with the Roman numerals referring to the order of the cysteine residues in the LDL-A module (Fig.1). In view of the high similarity between the primary sequences of the linker cysteine-rich module and the LDL-A module, it appears reasonable to expect that the annelid linkers should exhibit the same disulfide connectivity as in the LDL-A module

#### *Aspartic and glutamic acidic residues in LDL-A modules*

Several studies have shown that five conserved aspartic and glutamic acidic residues interacting with calcium (referred as D<sup>1</sup>, D<sup>2</sup>, D<sup>3</sup>, D<sup>4</sup> and E<sup>5</sup>) (Fig.1) play a role in the conformation and function of the LDL-A module in the LDLR superfamily (Fass et al. 1997; Rudenko et al. 2002; Simonovic et al. 2001). The side chains of four of the conserved acidic residues (D<sup>1</sup>, D<sup>2</sup>, D<sup>4</sup> and E<sup>5</sup>) as well as the carbonyl oxygen group of two non-conserved residues are involved in the calcium coordination of the LDL-A modules of LDLR (Fass et al. 1997). Indeed, Ca<sup>2+</sup> is required for correct folding but also for maintaining the structural integrity of LDL-A module in LDLR superfamily (Atkins et al. 1998; Guo et al. 2004; van Driel et al. 1987). The residues D<sup>4</sup> and E<sup>5</sup> are present in all of the modules studied (Guo et al. 2004), and are also found in all the linkers sequences shown in Fig.1. Hence, D<sup>4</sup> and E<sup>5</sup> can be considered as invariant residues, like the six conserved cysteine residues. Interestingly, Kuchumov et al. have proposed that Ca<sup>2+</sup> is necessary for the linkers to adopt an assembly-competent conformation in order to allow the re-assembly of the HBL structure subsequent to experimental dissociation of *L. terrestris* HBL-Hb in solution (Kuchumov et al. 2000; Kuchumov et al. 1999).

In addition, the acidic residues (except D<sup>2</sup> and D<sup>3</sup>) are supposed to be also involved in the ligand binding in the LDL-A modules of LDLR (Fass et al. 1997). Surprisingly, in 17 linker sequences among the 22 shown in Fig.1, the aspartic acid D<sup>3</sup> is substituted by an asparagine residue (Fig.1), which may be functionally significant for the LDL-A module in annelid linker

chains relative to the lipoprotein receptor LDL-A modules in the other organisms. Furthermore, although the LDL-A module has been proposed to be the domain responsible for the binding between linker chains and globin dodecamer subunits (Kuchumov et al. 1999), the role of the different acidic residues as well as of the conserved D<sup>3</sup> asparagine residue in linker sequences remains to be elucidated.

#### *Cysteine residues signatures outside the LDL-A module*

Four or six cysteine residues are found outside the LDL-A module in linkers either in conserved or variable positions (Fig.1). From data described for *T. heterochaetus* L1 in Suzuki et al. (Suzuki et al. 1994), we assume that Cys-156 and Cys-257, and Cys-234 and Cys-244/246 would form intrachain disulfide bridge. As a consequence, the two cysteine residues present in the N-terminal region of the sequence are those probably involved in interchain disulfide bridge (Suzuki et al. 1990a). Indeed, as suggested by Suzuki et al. (Suzuki et al. 1990a) and confirmed by the new linker sequences, the two cysteine residues (before position 40 in the alignment Fig.1) are present in linker chains known to form dimers or trimers (such as *T. heterochaetus*, *S. spallanzanii*, *N. diversicolor* and *A. marina*) and are absent in linker chains present only as monomers (*L. terrestris*, *A. pompejana*, *M. decora* and *Lamellibrachia. sp.*).

#### Revised linker gene structure

##### *Intron positions in agreement with previous studies*

The earlier studies on the gene structures of linker L1 from *L. terrestris* (Suzuki and Riggs 1993), and linker L2 from *N. diversicolor* (Suzuki et al. 1994) have shown a two introns/ three exons gene pattern (Fig. 2a).

The first intron occurs on the 5' flank of the LDL-A module, *i.e.* in position 91 (Fig.1) and is also found at that position in the linker L2 sequence of *A. marina*.

The second intron described in *L. terrestris* linker L1 is in the 3'UTR region (Suzuki and Riggs 1993). Even though the position is not strictly conserved in *R. pachyptila* LX and *N. diversicolor* L2, the intron position is also in the 3'UTR region, whereas in *A. marina* L2 the intron position is in the coding sequence (Fig.2b). Introns in the 3'UTR sequence occur rarely (Nagy and Maquat 1998) and could play a role in nonsense-mediated reduction of mRNA abundance (Zhang et al. 1998a). Intron position polymorphism between two adjacent codons (phase 0) or inside a codon (phase 1 or 2) can occur, as in the case of hemoglobin gene (Hardison 1998), but to our knowledge, such an amplitude around the stop codon has not been

frequently reported in literature. Therefore, it is difficult to speculate on a possible intron sliding event or on the biological significance with respect to mRNA transcription or regulation.

#### *Discovery of new intron positions*

Surprisingly, we found an additional intron in *A. marina*, *N. diversicolor*, *R. pachyptila* and *L. terrestris* linker sequences flanking 3' to the LDL-A module (Fig. 2b). The LDL-A modules of annelids are in fact flanked by two introns, similar to the human LDL-A modules (Fig. 2c). The ligand binding domain of the human LDLR is composed of seven collinear LDL-A modules (Sudhof et al. 1985a). The gene structure shows that the different LDL-A modules are encoded by distinct exons flanked by phase 1 inserted introns, except for domain III-IV-V, encoded by a single exon (Fig.2c). Interestingly, the LDL-A module of linker sequences is also flanked by two phase 1 introns. Patthy has shown that most of the identified cases of exon shuffling leading to new chimeric protein (for example interleukin-2 receptor, follistatin or factor XIIIb subunit) exhibit a typical phase 1 intron pattern (Patthy 1987; Patthy 1991; Patthy 1995). Suzuki et al. have proposed that the linker sequence would result from the duplication of a globin gene followed by the loss of the first exon and insertion of an LDL-A module by an exon shuffling process (Suzuki and Riggs 1993). The occurrence of these flanking introns supports the assumption of an exon shuffling event to explain the presence of the LDL-A module in linker genes.

A fourth intron has also been found in *A. marina*, *N. diversicolor*, *R. pachyptila* and *A. pompejana* linker sequences but not in *L. terrestris* linker L1. Although they all are phase 0 introns, the intron positions are not conserved and are distributed over a region of 16 amino-acids. The absence of the fourth intron in the *L. terrestris* L1 appears to be due to a loss of this intron in this species.

#### Linker structure

Suzuki et al. have proposed that linker sequences can be separated into three domains (Suzuki et al. 1994): N-terminal domain 1 (position 1 to 94), the cysteine-rich domain (position 95 to 136) and the C-terminal domain (position 137 to 268). The new gene structures presented here suggest rather that linker sequences should be separated into four domains, corresponding to the four exons, which suggest a more complicated emergence.

The hypothesis of a common origin for the linker and globin genes has been proposed by Suzuki et al. (Suzuki and Riggs 1993), involving particularly the N-terminal domain of linker

sequences. Furthermore, this common origin is supported by the recognition of both globin and linker chains by monoclonal antibodies against *L. terrestris* HBL-Hb (Lightbody et al. 1988) which implies that they share some common epitopes.

On the other hand, this hypothesis of a globin-like N-terminal domain is challenged by the structure of *L. terrestris* HBL-Hb (Royer et al. 2000), which shows the linker chains as long coiled-coil helices attached to a globular domain. Furthermore, this globular domain is composed of both the LDL-A module and the C-terminal domain, and is in contact with the globin subunits, and the N-terminal domain does not appear as a globin-like structure, and is involved only in linker-linker interactions (Royer et al. 2000). A higher resolution crystal structure will be required to determine the structure of the linker chains in HBL Hbs.

#### Linker evolutionary history

The phylogenetic analysis (Fig.3) suggests that the linker sequences cluster in two groups. We assume that the common ancestor gene to all linker genes had a 4 introns/5 exons structure and gave rise to the two families of linker genes after a duplication event.

Historically, linkers have been named according to their chronological discovery (purification and characterization) such as L1 for the first one, L2 for the second one, etc... This numbering has no sense in a molecular phylogenetic framework taking into account paralogy and orthology relationships between genes or genes products (*i.e* proteins). We propose to name arbitrarily LA and LB the two families of linkers, *i.e* the two distinct clusters which appear in Fig.3.

The LDL-A modules from the LDLR family and the annelid linker sequences exhibit the same conserved amino-acids (cysteine and acidic residues) and are similarly flanked by two introns. Based on these observations, all the modules found in the proteins from *Homo sapiens*, *Xenopus laevi*, *Drosophila melanogaster*, *Caenorhabditis elegans* and annelids could be considered as homologous and resulting from a common ancestor. Moreover, as described above, linker genes are specific to the annelid phylum and the LDL-A modules in annelids may have been recruited to contribute to the elaboration of linker chains. Phylogenetic analyses of LDL-A modules from both LDLR and linker sequences (unpublished data) show that LDL-A from linker cluster independently in respect to the LDL-A modules from LDLR family (despite low bootstraps-value). This supports a distinct and different evolutionary history.

So far, there have been no reports in the literature of either LDLR or VLDLR genes in Annelids and none were found *via* BLAST searches in EST's from *L. rubellus* and *E. andrei*.



In contrast, some LDLR related sequences have been identified from the molluscs *Crassostrea gigas* (CAD82865) and *Lymnaea stagnalis* (CAA80651) (Tensen et al. 1994), the arthropod *Drosophila melanogaster* (AAM50824) (Stapleton et al. 2002) and the nematode *Caenorhabditis elegans* (Q04833) (Yochem and Greenwald 1993). This strongly suggests that an LDLR gene was present in the ancestor common to both protostomes and deuterostomes, and would have been lost in the annelid phylum. On the other hand, it is also possible that the LDLR gene may have existed in the annelid ancestor but was co-opted into a different function, explaining the absence of LDLR-like proteins in annelids.

In this paper, we have presented new linker sequences as well as a revised linker gene structure. These results indicate the presence of two linker families: LA and LB, similar to the presence of two families of globin chains in annelids (Gotoh et al. 1987) and provides a new hypothesis regarding the evolutionary history of linkers. However, more data regarding both primary sequences and structural features are necessary to really understand the molecular events that gave rise to the linker genes, and ultimately, to the HBL-Hbs.

### **Acknowledgments:**

We are indebted to Serge Vinogradov for his critical reading and comments on the manuscript. We thank the captains and crews of the NO L'Atalante, the pilots, cruises and team of the submersibles Nautile and Alvin of the *HOT 96*, *HOPE'99*, *LARVE99* and *BIOSPEEDO*. This work was supported by CNRS, European grant (FEDER n° presage 3814) and the Conseil Régional de Bretagne (contrat n°A2C809).

## **Literatures Cited**

- Atkins AR, Brereton IM, Kroon PA, Lee HT, Smith R (1998) Calcium is essential for the structural integrity of the cysteine-rich, ligand-binding repeat of the low-density lipoprotein receptor. *Biochemistry* 37:1662-70
- Bailly X, Jollivet D, Vanin S, Deutsch J, Zal F, Lallier F, 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* 19:1421-33
- Bates P, Young JA, Varmus HE (1993) A receptor for subgroup A Rous sarcoma virus is related to the low density lipoprotein receptor. *Cell* 74:1043-51
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783-95
- Bieri S, Djordjevic JT, Daly NL, Smith R, Kroon PA (1995) Disulfide bridges of a cysteine-rich repeat of the LDL receptor ligand-binding domain. *Biochemistry* 34:13059-65
- Brown MS, Goldstein JL (1986) A receptor-mediated pathway for cholesterol homeostasis. *Science* 232:34-47
- Chen N, Harris TW, Antoshechkin I, Bastiani C, Bieri T, Blasiar D, Bradnam K, Canaran P, Chan J, Chen CK, Chen WJ, Cunningham F, Davis P, Kenny E, Kishore R, Lawson D, Lee R, Muller HM, Nakamura C, Pai S, Ozersky P, Petcherski A, Rogers A, Sabo A, Schwarz EM, Van Auken K, Wang Q, Durbin R, Spieth J, Sternberg PW, Stein LD ((31 co-authors). 2005) WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Res* 33:D383-9
- Dirlam KA, Gretch DG, LaCount DJ, Sturley SL, Attie AD (1996) Expression and characterization of a truncated, soluble, low-density lipoprotein receptor. *Protein Expr Purif* 8:489-500
- DiScipio RG, Gehring MR, Podack ER, Kan CC, Hugli TE, Fey GH (1984) Nucleotide sequence of cDNA and derived amino acid sequence of human complement component C9. *Proc Natl Acad Sci U S A* 81:7298-302
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-7
- Fass D, Blacklow S, Kim PS, Berger JM (1997) Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. *Nature* 388:691-3

- Fushitani K, Higashiyama K, Asao M, Hosokawa K (1996) Characterization of the constituent polypeptides of the extracellular hemoglobin from *Lumbricus terrestris*: heterogeneity and discovery of a new linker chain L4. *Biochim Biophys Acta* 1292:273-80
- Gotoh T, Sano T, Shibuya A, Yamaki M, Imai K, Ebina S (1998) Hexagonal bilayer structuring activity of linker chains of an annelid giant hemoglobin from the polychaete *Perinereis aibuhitensis*. *Arch Biochem Biophys* 360:75-84
- 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
- Guo Y, Yu X, Rihani K, Wang QY, Rong L (2004) The role of a conserved acidic residue in calcium-dependent protein folding for a low density lipoprotein (LDL)-A module: implications in structure and function for the LDL receptor superfamily. *J Biol Chem* 279:16629-37
- Hardison R (1998) Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J Exp Biol* 201 (Pt 8):1099-117
- 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
- Imai K (1999) The haemoglobin enzyme. *Nature* 401:437, 439
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8:275-82
- Kitamoto Y, Yuan X, Wu Q, McCourt DW, Sadler JE (1994) Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains. *Proc Natl Acad Sci U S A* 91:7588-92
- Kuchumov AR, Loo JA, Vinogradov SN (2000) Subunit distribution of calcium-binding sites in *Lumbricus terrestris* hemoglobin. *J Protein Chem* 19:139-49
- Kuchumov AR, Taveau JC, Lamy JN, Wall JS, Weber RE, Vinogradov SN (1999) The role of linkers in the reassembly of the 3.6 MDa hexagonal bilayer hemoglobin from *Lumbricus terrestris*. *J Mol Biol* 289:1361-74
- Lamy J, Kuchumov A, Taveau JC, Vinogradov SN, Lamy JN (2000) Reassembly of *Lumbricus terrestris* hemoglobin: a study by matrix-assisted laser desorption/ionization mass spectrometry and 3D reconstruction from frozen-hydrated specimens. *J Mol Biol* 298:633-47
- Lamy JN, Green BN, Toulmond A, Wall JS, Weber RE, Vinogradov SN (1996) Giant Hexagonal Bilayer Hemoglobins. *Chem Rev* 96:3113-3124

- Lee MS, Cho SJ, Lee JA, Park BJ, Cho HJ, Moon JS, Kim SK, Choo JK, Park SC (2004) Transcriptome analysis in the midgut of the earthworm (*Eisenia andrei*) using expressed sequence tags. Unpublished
- Lightbody JJ, Quabar AN, Mainwaring MG, Young JS, Walz DA, Vinogradov SN, Gotoh T (1988) Immunological relatedness of annelid extracellular hemoglobins and chlorocruorins. *Comp Biochem Physiol B* 90:301-5
- Liochev SI, Kuchumov AR, Vinogradov SN, Fridovich I (1996) Superoxide dismutase activity in the giant hemoglobin of the earthworm, *Lumbricus terrestris*. *Arch Biochem Biophys* 330:281-4
- Mahley RW (1988) Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science* 240:622-30
- Mehta KD, Chen WJ, Goldstein JL, Brown MS (1991) The low density lipoprotein receptor in *Xenopus laevis*. I. Five domains that resemble the human receptor. *J Biol Chem* 266:10406-14
- Nagy E, Maquat LE (1998) A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci* 23:198-9
- 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
- Patthy L (1987) Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett* 214:1-7
- Patthy L (1991) Exons--original building blocks of proteins? *Bioessays* 13:187-92
- Patthy L (1995) Protein evolution by exon shuffling.
- Raychowdhury R, Niles JL, McCluskey RT, Smith JA (1989) Autoimmune target in Heymann nephritis is a glycoprotein with homology to the LDL receptor. *Science* 244:1163-5
- Royer WE, Jr., Strand K, van Heel M, Hendrickson WA (2000) Structural hierarchy in erythrocrucorin, the giant respiratory assemblage of annelids. *Proc Natl Acad Sci U S A* 97:7107-11
- Rudenko G, Henry L, Henderson K, Ichtchenko K, Brown MS, Goldstein JL, Deisenhofer J (2002) Structure of the LDL receptor extracellular domain at endosomal pH. *Science* 298:2353-8

- Saito A, Pietromonaco S, Loo AK, Farquhar MG (1994) Complete cloning and sequencing of rat gp330/"megalin," a distinctive member of the low density lipoprotein receptor gene family. *Proc Natl Acad Sci U S A* 91:9725-9
- Simonovic M, Dolmer K, Huang W, Strickland DK, Volz K, Gettins PG (2001) Calcium coordination and pH dependence of the calcium affinity of ligand-binding repeat CR7 from the LRP. Comparison with related domains from the LRP and the LDL receptor. *Biochemistry* 40:15127-34
- Stapleton M, Carlson J, Brokstein P, Yu C, Champe M, George R, Guarin H, Kronmiller B, Pacleb J, Park S, Wan K, Rubin GM, Celniker SE (2002) A *Drosophila* full-length cDNA resource. *Genome Biol* 3:RESEARCH0080
- Sudhof TC, Goldstein JL, Brown MS, Russell DW (1985a) The LDL receptor gene: a mosaic of exons shared with different proteins. *Science* 228:815-22
- Sudhof TC, Russell DW, Goldstein JL, Brown MS, Sanchez-Pescador R, Bell GI (1985b) Cassette of eight exons shared by genes for LDL receptor and EGF precursor. *Science* 228:893-5
- Suzuki T, Ohta T, Yuasa HJ, Takagi T (1994) The giant extracellular hemoglobin from the polychaete *Neanthes diversicolor*. The cDNA-derived amino acid sequence of linker chain L2 and the exon/intron boundary conserved in linker genes. *Biochim Biophys Acta* 1217:291-6
- 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, Takagi T, Gotoh T (1990a) Primary structure of two linker chains of the extracellular hemoglobin from the polychaete *Tylorrhynchus heterochaetus*. *J Biol Chem* 265:12168-77
- 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, Vinogradov SN (2003) Globin and linker sequences of the giant extracellular hemoglobin from the leech *Macrobdella decora*. *J Protein Chem* 22:231-42
- Takahashi S, Kawarabayasi Y, Nakai T, Sakai J, Yamamoto T (1992) Rabbit very low density lipoprotein receptor: a low density lipoprotein receptor-like protein with distinct ligand specificity. *Proc Natl Acad Sci U S A* 89:9252-6
- Tensen CP, Van Kesteren ER, Planta RJ, Cox KJ, Burke JF, van Heerikhuizen H, Vreugdenhil E (1994) A G protein-coupled receptor with low density lipoprotein-

- binding motifs suggests a role for lipoproteins in G-linked signal transduction. Proc Natl Acad Sci U S A 91:4816-20
- van Driel IR, Goldstein JL, Sudhof TC, Brown MS (1987) First cysteine-rich repeat in ligand-binding domain of low density lipoprotein receptor binds Ca<sup>2+</sup> and monoclonal antibodies, but not lipoproteins. J Biol Chem 262:17443-9
- Vinogradov SN (1985) The structure of invertebrate extracellular hemoglobins (erythrocrucorins and chlorocruorins). Comp Biochem Physiol B 82:1-15
- Vinogradov SN, Lugo SD, Mainwaring MG, Kapp OH, Crewe AV (1986) Bracelet protein: a quaternary structure proposed for the giant extracellular hemoglobin of *Lumbricus terrestris*. Proc Natl Acad Sci U S A 83:8034-8
- Vinogradov SN, Walz DA, Pohajdak B, Moens L, Kapp OH, Suzuki T, Trotman CN (1993) Adventitious variability? The amino acid sequences of nonvertebrate globins. Comp Biochem Physiol B 106:1-26
- Weber RE, Vinogradov SN (2001) Nonvertebrate hemoglobins: functions and molecular adaptations. Physiol Rev 81:569-628
- Yochem J, Greenwald I (1993) A gene for a low density lipoprotein receptor-related protein in the nematode *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 90:4572-6
- Zal F, Green BN, Lallier FH, Toulmond A (1997a) Investigation by electrospray ionization mass spectrometry of the extracellular hemoglobin from the polychaete annelid *Alvinella pompejana*: an unusual hexagonal bilayer hemoglobin. Biochemistry 36:11777-86
- 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, Lallier FH, Wall JS, Vinogradov SN, Toulmond A (1996) The multi-hemoglobin system of the hydrothermal vent tube worm *Riftia pachyptila*. I. Reexamination of the number and masses of its constituents. J Biol Chem 271:8869-74
- Zhang J, Sun X, Qian Y, Maquat LE (1998a) Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. RNA 4:801-15

## **Chapitre 7**

### **Conclusion et perspectives**

## **Chapitre 7: Conclusions et perspectives.**

Les deux premiers chapitres de ce manuscrit ont mis en évidence la diversité structurale et fonctionnelle des globines au sein des trois grands domaines du vivant Bacteria, Eucaryota et Archaea (Chapitre 1), et de façon plus spécifique au sein des annélides (Chapitre 2).

L'objectif initial de ce travail de thèse était de caractériser la famille multigénique codant les globines ainsi que celle codant les linkers de *Arenicola marina* dans le cadre de la mise au point d'un substitut sanguin humain à partir de l'HBL-Hb de *Arenicola marina*. L'ensemble des résultats obtenus nous a permis d'analyser ces nouvelles données dans un cadre évolutif et fonctionnel plus large.

Nous avons pu ainsi apporter de nouveaux éléments concernant :

- a) l'évolution de la famille multigénique des globines extracellulaires de *Arenicola marina* (Chapitre 4)

- b) l'émergence des trois lignées de globines chez les annélides (intracellulaires non circulantes (myoglobines), intracellulaires circulantes et globines extracellulaires (Chapitre 5))

- c) l'évolution moléculaire des gènes codant pour les chaînes de structure (Linkers) des HBL-Hb d'annélides (Chapitre 6).

Dans ce chapitre de conclusion (Chapitre 7), nous rappellerons les résultats obtenus et suggérerons des approches méthodologiques possibles pour continuer la caractérisation des deux familles multigéniques de globines et de linkers codant l'hémoglobine extracellulaire de *Arenicola marina*. Enfin, nous intégrerons l'ensemble des informations et résultats issus des chapitres 4, 5 et 6 pour les replacer dans le contexte plus large de l'émergence des HBL-Hb chez les annélides.

### **I Résultats et perspectives.**

Nous avons réussi à caractériser :

- cinq ADNc de globines extracellulaires de *Arenicola marina*
- un ADNc de linker de *Arenicola marina*
- un ADNc de globine intracellulaire de *Alvinella pompejana*
- trois ADNc de linkers de *Alvinella pompejana*



Nous avons aussi caractérisé la structure des gènes (séquences et positions des introns) :

- des globines A2 et B2 de *Arenicola marina*
- des globines A1, B1a, B1b et B1c de *Riftia pachyptila*
- de la myoglobine MbIIa de *Arenicola marina*
- de la globine intracellulaire circulante de *Alvinella pompejana*
- de la globine intracellulaire circulante mIV de *Glycera dibranchiata*
- de linker L2 de *Arenicola marina*
- du linker LX de *Riftia pachyptila*
- du linker L1 de *Lumbricus terrestris*
- du linker L2 de *Neanthes diversicolor*.

Parmi les globines extracellulaires de *Arenicola marina*, nous avons identifié une globine B1, une globine B2 et trois globines A2, mais aucune globine A1 (Figure 32). L'absence d'une ou plusieurs globines A1 peut s'expliquer par le fait que les amorces dégénérées conçues à partir des régions conservées de la sous-famille A1 étaient trop différentes de la séquence recherchée pour permettre une amplification : la ou les séquences de globines A1 de *Arenicola marina* sont probablement divergentes par rapport aux globines A1 disponibles dans les banques de données.

L'abondance de chaque sous-unité au sein de l'HBL-Hb de *Arenicola marina* décrites par spectrométrie de masse montre que certaines globines non identifiées sont les moins représentées dans la molécule (Chapitre 4). On peut supposer que la faible représentation d'une sous-unité dans la molécule est proportionnelle aux taux de transcription des ARNm de la sous-unité concernée. Ainsi, la faible quantité d'ARNm correspondant à ces sous-unités représentait trop peu de matrice pour la Reverse Transcription et l'amplification.

Pour obtenir les séquences d'ADNc manquants, on pourrait envisager de séparer et de purifier les différents monomères de globines et de linkers composant l'hémoglobine extracellulaire de *Arenicola marina*, et de réaliser un séquençage *de novo* et/ou un séquençage d'Edman. Les peptides ainsi obtenus permettraient de concevoir des amorces dégénérées spécifiques des sous-unités non identifiées de *Arenicola marina*.

Chaîne	ESI-MS		Masse déduite du cDNA			$\Delta$ Masse (Da)
	Masse (Da)	Cys Tot	Chaîne	Masse (Da)	Cys Tot	
<b>Monomère</b>						
a1	15954.4	3	×	×	×	×
a2	15976.8	3	A2c	15977.3	3	0.5
<b>Monomère impliqué dans un trimère</b>						
b1	15920.9	3	A2b	15921.1	3	0.2
b2	16020.1	3	A2a	15932.2	3	87.9
b3	16036.2	3	×	×	×	×
c	16664.8	3	B1	16660.8	3	4.0
d1	16983.2	5	×	×	×	×
d2	17033.1	5	B2	17033.0	5	0.1
<b>Linker</b>						
L1	25174.1	12	×	×	×	×
L2	26829.7	12	L2	26740.7	12	89

**Figure 32 :** Correspondance entre les masses des chaînes de globines et de linkers obtenues par spectrométries de masses (ESI-MS) (Zal et al. 1997b) et celles des séquences protéiques déduites à partir des cDNAs caractérisés par biologie moléculaire. × non déterminé.

Nous pouvons supposer que les gènes codant pour les globines extracellulaires d'annélides sont regroupés comme les gènes de globines des mammifères au sein de cluster sur le chromosome. Des expériences répétées de « marche sur le chromosome » réalisées à partir d'un des gènes de globines de *Arenicola marina* déjà identifiés permettraient peut-être de séquencer un ou plusieurs autres gènes de globines, à condition que les régions inter-géniques soient relativement courtes.

D'autres outils de biologie moléculaires pourraient permettre l'obtention des ADNc manquants des deux familles multigéniques. Il serait en effet possible de réaliser une banque d'ADNc de *Arenicola marina* puis de cribler cette banque grâce à des sondes spécifiques synthétisées à partir des ADNc déjà obtenus, mais en utilisant des températures d'hybridations plus faibles (faible stringence) afin de favoriser des mésappariements lors des étapes d'hybridation et de caractériser des séquences différentes.

## **II Emergence des lignées de globines chez les Annélides.**

### **A- Positions des introns.**

Pour comprendre les relations phylogénétiques existant entre les trois lignées de globines d'annélides (globine intracellulaire circulante et non circulante et globine extracellulaire) nous avons déterminé la structure de leur gène. En effet, les mêmes positions d'introns trouvées dans des gènes codant pour des protéines homologues montrent que ces gènes ont évolué à partir d'un ancêtre commun. L'étude de la structure des gènes des globines intracellulaires circulantes et non circulantes et des globines extracellulaires d'annélides a mis en évidence et confirmé la présence d'un motif conservé de trois exons/deux introns. Ces positions d'introns conservées confirment que ces trois lignées de globines sont issues d'un gène ancestral de type globine qui devait posséder au moins cette structure de gène (Chapitre 5).

Les analyses réalisées à partir de la séquence de la globine intracellulaire circulante de *Alvinella pompejana* ont révélé une forte similarité entre cette séquence et celle de la déhaloperoxydase (DHP) de l'annélide polychaete *Amphitrite ornata*. De précédents travaux réalisés sur cet enzyme suggéraient une origine commune entre la DHP et les globines. Pour vérifier cette hypothèse nous avons caractérisé la structure du gène codant la DHP. La présence des deux introns B12.2 et G7.0 confirme que la DHP est une globine. Ces résultats suggèrent que le gène codant la DHP est soit identique à celui de la globine intracellulaire circulante de *Amphitrite ornata*, soit issu d'une duplication ancestrale à partir d'un gène de type globine (Chapitre 5). Les données que nous possédons actuellement ne nous permettent pas de trancher pour l'une ou l'autre de ces deux possibilités.

Il serait nécessaire de déterminer quel(s) sont le ou les tissus exprimant la DHP pour savoir si cette globine est ou non exprimée au sein des coelomocytes de *Amphitrite ornata* (lieu d'expression de la globine intracellulaire circulante). Il faudrait aussi purifier la globine intracellulaire circulante de *Amphitrite ornata* et réaliser un séquençage *de novo* et/ou un séquençage d'Edman, afin de pouvoir concevoir des amorces dégénérées pour amplifier la séquence de cette globine, et ainsi la comparer avec la séquence de la DHP déjà connue. Il serait également intéressant de mesurer l'activité déhaloperoxydase d'autres globines (intracellulaires circulantes ou non circulantes) d'annélides pour déterminer si la DHP de *Amphitrite ornata* est une globine possédant une fonction particulière, ou si cette fonction est commune aux autres globines.

### B- Caractéristiques structurales des HBL-Hbs.

La présence de globines dans les trois grands domaines du vivant renforce l'idée d'un ancêtre commun de type intracellulaire. Ainsi, la présence de séquences d'adressage dans la partie N-terminale de la protéine chez les hémoglobines extracellulaires qui permet l'externalisation des hémoglobines dans le système vasculaire ou coelomique, peut donc être considéré comme une innovation par rapport aux globines intracellulaires circulantes et non circulantes. De plus, les globines constitutives de ces hémoglobines extracellulaires possèdent deux résidus cystéines impliqués dans la formation d'un pont disulfure intrachaîne. L'absence de ces résidus dans les séquences de globines intracellulaires circulantes et non circulantes des annélides suggère que le pont disulfure pourrait être corrélé au caractère extracellulaire de ces hémoglobines.

Des expériences réalisées par Fushitani (Fushitani, Ochi and Morimoto 1982) sur le polychaete *Travisia japonica* ont montré que des anticorps anti-globines extracellulaires ne réagissaient pas avec les globines intracellulaires de la même espèce. Ces résultats ainsi que des caractéristiques restreintes aux globines extracellulaires (peptide signal, pont disulfure intrachaîne) renforcent l'idée d'une émergence probablement tardive des globines extracellulaires par rapport aux globines intracellulaires.

### C- Les globines extracellulaires et les Linkers.

La polymérisation des HBL-Hbs en complexes de hauts poids moléculaires minimiserait l'élimination rapide au niveau des organes de filtration. Les Siboglinidae (Vestimentifères et Pogonophores) expriment des hémoglobines extracellulaires de 400 kDa qui correspondent à l'assemblage de deux dodécamères de globines (Chapitre 2). Ces protéines, ne font pas intervenir de chaînes de structure (linkers). Ainsi, l'extracellularité des hémoglobines n'implique pas forcément une structure en HBL-Hb et suggère une émergence secondaire de la structure en HBL, impliquant la présence des linkers.

Dans le cas des HBL-Hb, le fort degré d'hétéro-polymérisation suppose des interactions globines/globines, linkers/linkers mais aussi globines/linkers, tout en conservant une protéine fonctionnelle. Ainsi, l'évolution des HBL-Hbs implique une étroite co-évolution des globines et des linkers chez les annélides.

L'évolution des linkers au sein des annélides reste encore peu étudiée. Cependant, des modules LDL-A similaires à ceux trouvées dans la superfamille des LDLR eucaryotes ont été

décrit au sein des linkers d'annélides (Sudhof et al. 1985a; Suzuki and Riggs 1993). Les similarités de séquences ainsi que la caractérisation de deux introns flanquants le domaine LDL-A en positions conservées entre les linkers d'annélides et les autres protéines possédant des domaines LDL-A (Chapitre 6) renforce l'hypothèse d'une origine commune entre tous ces modules LDL-A de linkers et de LDLR. Aucun gène codant pour une protéine de la famille des LDLR n'a pour le moment été décrit chez les annélides, ni retrouvé dans les banques d'ESTs d'annélides. En se basant sur la présence de gènes de la famille des LDLR dans des phyla de protostomiens tels que les mollusques ou les arthropodes, on peut supposer que l'ancêtre commun aux annélides, arthropodes et mollusques possédait déjà un gène de type LDLR. Dans le cas des annélides, un tel gène aurait pu être recruté pour une autre fonction, (ici, principalement le maintien de la structure quaternaire dans les HBL-Hbs).

D'autres phyla expriment des hémoglobines extracellulaires. Il est toutefois important de garder à l'esprit que même si les globines extracellulaires d'arthropodes possèdent aussi un peptide signal, elles ne possèdent ni les deux résidus cystéines impliqués dans la formation du ponts disulfures intrachaîne des globines extracellulaires d'annélides, ni une structure quaternaire en bicouche hexagonale et n'expriment pas de linkers. Ceci suggère une histoire évolutive différente des hémoglobines extracellulaires d'annélides par rapport aux hémoglobines extracellulaires présentes dans d'autres phyla.

## **Bibliographie**

## Bibliographie

- Allen, G. 1989. Sequencing of proteins and peptides. Pp 73-104 in R.H Burdon and P.H Van Knippenberg, eds. Laboratory techniques in biochemistry and molecular biology. Vol. 9. Elsevier North Holland, Amsterdam.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.
- Andersson, C. R., J. E. Rostergaard, D. J. Llewellyn, E. S. Dennis and W. J. Peacock. 1996. A new hemoglobin gene from soybean: A role for hemoglobin in all plants. *Proc Natl Acad Sci U S A.* **93**: 5682-5687.
- Andersson, J. O., A. M. Sjogren, L. A. Davis, T. M. Embley and A. J. Roger. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. *Curr Biol.* **13**: 94-104.
- Appleby, C. A. 1992. The origin and functions of haemoglobin in plants. *Sci Prog.* **76**: 365-398.
- Arp, A. J., J. J. Childress and R. D. Vetter. 1987. The sulphide-binding protein in the blood of the vestimentiferan tube-worm, *Riftia pachyptila*, is the extracellular haemoglobin. *J Exp Biol.* **128**: 139-158.
- Arp, A. J., M. L. Doyle, E. Di Cera and S. J. Gill. 1990. Oxygenation properties of the two co-occurring hemoglobins of the tube worm *Riftia pachyptila*. *Respir Physiol.* **80**: 323-334.
- Arredondo-Peter, R., M. S. Hargrove, G. Sarath, J. F. Moran, J. Lohrman, J. S. Olson and R. V. Klucas. 1997. Rice hemoglobins. Gene cloning, analysis, and O<sub>2</sub>-binding kinetics of a recombinant protein synthesized in *Escherichia coli*. *Plant Physiol.* **115** (3) : 1259-66.
- Atkins, A. R., I. M. Brereton, P. A. Kroon, H. T. Lee and R. Smith. 1998. Calcium is essential for the structural integrity of the cysteine-rich, ligand-binding repeat of the low-density lipoprotein receptor. *Biochemistry.* **37**: 1662-1670.
- Ashworth, J.H. 1904. *Arenicola* (The Lug-worm). L.M.B.C. memoirs, London.
- Bailly, X., D. Jollivet, S. Vanin, J. Deutsch, F. Zal, F. Lallier and A. Toulmond. 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.

- Bailly, X. 2003a. Evolution moléculaire de la fonction de liaison du sulfure d'hydrogène par les hémoglobines de *Riftia pachyptila*, annélide tubicole des sources hydrothermales profondes. Thèse de doctorat. École Pratique des Hautes Etudes et Université Pierre et Marie Curie (Paris VI), France.
- Bailly, X., R. Leroy, S. Carney, O. Collin, F. Zal, A. Toulmond and D. Jollivet. 2003b. The loss of the hemoglobin H<sub>2</sub>S-binding function in annelids from sulfide-free habitats reveals molecular adaptation driven by Darwinian positive selection. Proc Natl Acad Sci U S A. **100**: 5885-5890.
- Bates, P., J. A. Young and H. E. Varmus. 1993. A receptor for subgroup A Rous sarcoma virus is related to the low density lipoprotein receptor. Cell. **74**: 1043-1051.
- Bendtsen, J. D., H. Nielsen, G. von Heijne and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. **340**: 783-795.
- Bibikov, S. I., A. C. Miller, K. K. Gosink and J. S. Parkinson. 2004. Methylation-independent aerotaxis mediated by the Escherichia coli Aer protein. J Bacteriol. **186**: 3730-3737.
- Bieri, S., J. T. Djordjevic, N. L. Daly, R. Smith and P. A. Kroon. 1995. Disulfide bridges of a cysteine-rich repeat of the LDL receptor ligand-binding domain. Biochemistry. **34**: 13059-13065.
- Boccaro, M., C. E. Mills, J. Zeier, C. Anzi, C. Lamb, R. K. Poole and M. Delledonne. 2005. Flavohaemoglobin HmpX from Erwinia chrysanthemi confers nitrosative stress tolerance and affects the plant hypersensitive reaction by intercepting nitric oxide produced by the host. Plant J. **43**: 226-237.
- Boilly, B. and K. Sean. 1979. Localisation du tissu hématopoïétique chez les Annélides Polychètes. Arch. Anat. Microsc. **68**: 257-270.
- Bonamore, A., A. Farina, M. Gattoni, M. E. Schinina, A. Bellelli and A. Boffi. 2003a. Interaction with membrane lipids and heme ligand binding properties of Escherichia coli flavohemoglobin. Biochemistry. **42**: 5792-5801.
- Bonamore, A., P. Gentili, A. Ilari, M. E. Schinina and A. Boffi. 2003b. Escherichia coli flavohemoglobin is an efficient alkylhydroperoxide reductase. J Biol Chem. **278**: 22272-22277.
- Bonaventura, J. and V. P. Lance. 2001. Nitric oxide, invertebrates and hemoglobin. Am Zool. **41**: 346-359.
- Bourret, R. B. and A. M. Stock. 2002. Molecular information processing: lessons from bacterial chemotaxis. J Biol Chem. **277**: 9625-9628.



- Breathnach, R., J. L. Mandel and P. Chambon. 1977. Ovalbumin gene is split in chicken DNA. *Nature*. **270**: 314-319.
- Breton-Gorius, J. 1963. Etude au microscope électronique des cellules chloragènes d'*Arenicola marina* L. Leur rôle dans la synthèse de l'hémoglobine. Faculté des Sciences de l'Université de Paris.
- Brown, M. S. and J. L. Goldstein. 1986. A receptor-mediated pathway for cholesterol homeostasis. *Science*. **232**: 34-47.
- Burmester, T. and T. Hankeln. 1999. A globin gene of *Drosophila melanogaster*. *Mol Biol Evol*. **16**: 1809-1811.
- Burmester, T., B. Weich, S. Reinhardt and T. Hankeln. 2000. A vertebrate globin expressed in the brain. *Nature*. **407**: 520-523.
- Burmester, T., B. Ebner, B. Weich and T. Hankeln. 2002. Cytoglobin: a novel globin type ubiquitously expressed in vertebrate tissues. *Mol Biol Evol*. **19**: 416-421.
- Burmester, T. and T. Hankeln. 2004. Neuroglobin: a respiratory protein of the nervous system. *News Physiol Sci*. **19**: 110-113.
- Callebaut, I., G. Labesse, P. Durand, A. Poupon, L. Canard, J. Chomilier, B. Henrissat and J. P. Morion. 1997. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci*. **53**: 621-645.
- Canfield, D. E. and A. Teske. 1996. Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. *Nature*. **382**: 127-132.
- Chen, N., T. W. Harris, I. Antoshechkin, et al. (31 co-authors). 2005. WormBase: a comprehensive data resource for *Caenorhabditis* biology and genomics. *Nucleic Acids Res*. **33**: D383-389.
- Chen, Y. P., S. A. Woodin, D. E. Lincoln and C. R. Lovell. 1996. An unusual dehalogenating peroxidase from the marine terebellid polychaete *Amphitrite ornata*. *J Biol Chem*. **271**: 4609-4612.
- Concha, N. O., B. A. Rasmussen, K. Bush and O. Herzberg. 1997. Crystal structures of the cadmium- and mercury-substituted metallo-beta-lactamase from *Bacteroides fragilis*. *Protein Sci*. **6**: 2671-2676.
- Craik, C. S., S. R. Buchman and S. Beychok. 1980. Characterization of globin domains: heme binding to the central exon product. *Proc Natl Acad Sci U S A*. **77**: 1384-1388.

- Craik, C. S., S. R. Buchman and S. Beychok. 1981. O<sub>2</sub> binding properties of the product of the central exon of beta-globin gene. *Nature*. **291**: 87-90.
- Cramm, R., R. A. Siddiqui and B. Friedrich. 1994. Primary sequence and evidence for a physiological function of the flavohemoprotein of *Alcaligenes eutrophus*. *J Biol Chem*. **269**: 7349-7354.
- D'Angelo, P., D. Lucarelli, S. della Longa, et al. 2004. Unusual heme iron-lipid acyl chain coordination in *Escherichia coli* flavohemoglobin. *Biophys J*. **86**: 3882-3892.
- de Jesus-Berrios, M., L. Liu, J. C. Nussbaum, G. M. Cox, J. S. Stamler and J. Heitman. 2003. Enzymes that counteract nitrosative stress promote fungal virulence. *Curr Biol*. **13**: 1963-1968.
- De Souza, S. J., M. Long, R. J. Klein, S. Roy, S. Lin and W. Gilbert. 1998. Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. *Proc Natl Acad Sci U S A*. **95**: 5094-5099.
- Decker, H. and N. Terwilliger. 2000. Cops and robbers: putative evolution of copper oxygen-binding proteins. *J Exp Biol*. **203**: 1777-1782.
- Decker, H. and F. Tuczek. 2000. Tyrosinase/catecholoxidase activity of hemocyanins: structural basis and molecular mechanism. *Trends Biochem Sci*. **25**: 392-397.
- Demuyneck, S., P. Sautiere, J. van Beeumen and N. Dhainaut-Courtois. 1991. [Homologies between hemerythrins of sipunculids and cadmium-binding metalloprotein (MP II) from a polychaete annelid, *Nereis diversicolor*]. *C R Acad Sci III*. **312**: 317-322.
- 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.
- Dewilde, S., M. Blaxter, M. L. Van Hauwaert, et al. 1998. Structural, functional, and genetic characterization of *Gastrophilus* hemoglobin. *J Biol Chem*. **273**: 32467-32474.
- Dewilde, S., M. L. Van Hauwaert, S. Vinogradov, A. Vierstraete, J. Vanfleteren and L. Moens. 2001. Protein and gene structure of a chlorocruorin chain of *Eudistylia vancouverii*. *Biochem Biophys Res Commun*. **281**: 18-24.
- Dhainaut, A., B. Raveillon, M. M'Berri, E. Porchet-Hennere and S. Demuyneck. 1989. Purification of an antibacterial protein in the coelomic fluid of *Nereis diversicolor* (Annelida, Polychaeta). Similitude with a cadmium-binding protein. *Comp Biochem Physiol C*. **94**: 555-560.

- Dirlam, K. A., D. G. Gretch, D. J. LaCount, S. L. Sturley and A. D. Attie. 1996. Expression and characterization of a truncated, soluble, low-density lipoprotein receptor. *Protein Expr Purif.* **8**: 489-500.
- DiScipio, R. G., M. R. Gehring, E. R. Podack, C. C. Kan, T. E. Hugli and G. H. Fey. 1984. Nucleotide sequence of cDNA and derived amino acid sequence of human complement component C9. *Proc Natl Acad Sci U S A.* **81**: 7298-7302.
- Dixon, B. and B. Pohajdak. 1992. Did the ancestral globin gene of plants and animals contain only two introns? *Trends Biochem Sci.* **17**: 486-488.
- Ebner, B., T. Burmester and T. Hankeln. 2003. Globin genes are present in *Ciona intestinalis*. *Mol Biol Evol.* **20**: 1521-1525.
- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792-1797.
- Elvers, K. T., G. Wu, N. J. Gilberthorpe, R. K. Poole and S. F. Park. 2004. Role of an inducible single-domain hemoglobin in mediating resistance to nitric oxide and nitrosative stress in *Campylobacter jejuni* and *Campylobacter coli*. *J Bacteriol.* **186**: 5332-5341.
- Ermler, U., R. A. Siddiqui, R. Cramm and B. Friedrich. 1995. Crystal structure of the flavohemoglobin from *Alcaligenes eutrophus* at 1.75 Å resolution. *Embo J.* **14**: 6067-6077.
- Fago, A., C. Hundahl, H. Malte and R. E. Weber. 2004. Functional properties of neuroglobin and cytoglobin. Insights into the ancestral physiological roles of globins. *IUBMB Life.* **56**: 689-696.
- Farres, J., M. P. Rechsteiner, S. Herold, A. D. Frey and P. T. Kallio. 2005. Ligand binding properties of bacterial hemoglobins and flavohemoglobins. *Biochemistry.* **44**: 4125-4134.
- Fass, D., S. Blacklow, P. S. Kim and J. M. Berger. 1997. Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. *Nature.* **388**: 691-693.
- Flores, J. F., C. R. Fisher, S. L. Carney, B. N. Green, J. K. Freytag, S. W. Schaeffer and W. E. Royer, Jr. 2005. Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin. *Proc Natl Acad Sci U S A.* **102**: 2713-2718.
- Freitas, T. A., S. Hou and M. Alam. 2003. The diversity of globin-coupled sensors. *FEBS Lett.* **552**: 99-104.

- Freitas, T. A., S. Hou, E. M. Dioum, J. A. Saito, J. Newhouse, G. Gonzalez, M. A. Gilles-Gonzalez and M. Alam. 2004. Ancestral hemoglobins in Archaea. *Proc Natl Acad Sci U S A.* **101**: 6675-6680.
- Freitas, T. A., J. A. Saito, S. Hou and M. Alam. 2005. Globin-coupled sensors, protoglobins, and the last universal common ancestor. *J Inorg Biochem.* **99**: 23-33.
- Frey, A. D. and P. T. Kallio. 2003. Bacterial hemoglobins and flavohemoglobins: versatile proteins and their impact on microbiology and biotechnology. *FEMS Microbiol Rev.* **27**: 525-545.
- Fushitani, K., O. Ochi and H. Morimoto. 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**: 267-273.
- 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*. *J Biol Chem.* **263**: 6502-6517.
- Fushitani, K., K. Higashiyama, M. Asao and K. Hosokawa. 1996. Characterization of the constituent polypeptides of the extracellular hemoglobin from *Lumbricus terrestris*: heterogeneity and discovery of a new linker chain L4. *Biochim Biophys Acta.* **1292**: 273-280.
- Gardner, P. R., G. Costantino and A. L. Salzman. 1998. Constitutive and adaptive detoxification of nitric oxide in *Escherichia coli*. Role of nitric-oxide dioxygenase in the protection of aconitase. *J Biol Chem.* **273**: 26528-26533.
- Gardner, P. R., A. M. Gardner, L. A. Martin and A. L. Salzman. 1998. Nitric oxide dioxygenase: an enzymic function for flavohemoglobin. *Proc Natl Acad Sci U S A.* **95**: 10378-10383.
- Gardner, P. R. 2005. Nitric oxide dioxygenase function and mechanism of flavohemoglobin, hemoglobin, myoglobin and their associated reductases. *J Inorg Biochem.* **99**: 247-266.
- Garlick, R. L. and A. F. Riggs. 1982. The amino acid sequence of a major polypeptide chain of earthworm hemoglobin. *J Biol Chem.* **257**: 9005-9015.
- Geuens, E., I. Brouns, D. Flamez, S. Dewilde, J. P. Timmermans and L. Moens. 2003. A globin in the nucleus! *J Biol Chem.* **278**: 30417-30420.
- Geuens, E., S. Dewilde, D. Hoogewijs, et al. 2004. Nerve globins in invertebrates. *IUBMB Life.* **56**: 653-656.
- Gilbert, W. 1978. Why genes in pieces? *Nature.* **271**: 501.

- Gilbert, W. 1987. The exon theory of genes. *Cold Spring Harb Symp Quant Biol.* **52**: 901-905.
- Gilles-Gonzalez, M. A. and G. Gonzalez. 2005. Heme-based sensors: defining characteristics, recent developments, and regulatory hypotheses. *J Inorg Biochem.* **99**: 1-22.
- Go, M. 1981. Correlation of DNA exonic regions with protein structural units in haemoglobin. *Nature.* **291**: 90-92.
- 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, O. and T. Suzuki. 1990. Molecular Assembly and Evolution of Multi-Subunit Extracellular Annelid Hemoglobins. *Zoolog Sci.* **7**: 1-16.
- Gotoh, O., T. Suzuki and T. Takagi. 1991. Nomenclature of the Major Constituent Chains Common to Annelid Extracellular Hemoglobins. Pp 279-283 in S.N. Vinogradov and O.H Kapp, eds. Springer-Verlag, Berlin.
- 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.
- Gotoh, T., T. Sano, A. Shibuya, M. Yamaki, K. Imai and S. Ebina. 1998. Hexagonal bilayer structuring activity of linker chains of an annelid giant hemoglobin from the polychaete *Perinereis aibuhitensis*. *Arch Biochem Biophys.* **360**: 75-84.
- Guo, Y., X. Yu, K. Rihani, Q. Y. Wang and L. Rong. 2004. The role of a conserved acidic residue in calcium-dependent protein folding for a low density lipoprotein (LDL)-A module: implications in structure and function for the LDL receptor superfamily. *J Biol Chem.* **279**: 16629-16637.
- Hankeln, T., B. Ebner, C. Fuchs, et al. 2005. Neuroglobin and cytoglobin in search of their role in the vertebrate globin family. *J Inorg Biochem.* **99**: 110-119.
- Hardison, R. 1998. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J Exp Biol.* **201 (Pt 8)**: 1099-1117.
- Hardison, R. C. 1996. A brief history of hemoglobins: plant, animal, protist, and bacteria. *Proc Natl Acad Sci U S A.* **93**: 5675-5679.
- Harrington, J. P., G. Suarez, T. A. Borgese and R. L. Nagel. 1978. Subunit interactions of *Glycera dibranchiata* hemoglobin. *J Biol Chem.* **253**: 6820-6825.

- Harris, N. L. and P. Senapathy. 1990. Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. *Nucleic Acids Res.* **18**: 3015-3019.
- Hedges, S. B. 2002. The origin and evolution of model organisms. *Nat Rev Genet.* **3**: 838-849.
- Hendrickson, W. A. and G. L. Klippenstein. 1974. Crystals of myohemerythrin. *J Mol Biol.* **87**: 147-149.
- Hirsch, R. E., L. A. Jelicks, B. A. Wittenberg, D. K. Kaul, H. L. Shear and J. P. Harrington. 1997. A first evaluation of the natural high molecular weight polymeric *Lumbricus terrestris* hemoglobin as an oxygen carrier. *Artif Cells Blood Substit Immobil Biotechnol.* **25**: 429-444.
- Hoffmann, R. J. and C. P. Mangum. 1970. The function of coelomic cell hemoglobin in the polychaete *Glycera dibranchiata*. *Comp Biochem Physiol.* **36**: 211-228.
- Hoogewijs, D., E. Geuens, S. Dewilde, L. Moens, A. Vierstraete, S. Vinogradov and J. Vanfleteren. 2004. Genome-wide analysis of the globin gene family of *C. elegans*. *IUBMB Life.* **56**: 697-702.
- Hou, S., R. W. Larsen, D. Boudko, C. W. Riley, E. Karatan, M. Zimmer, G. W. Ordal and M. Alam. 2000. Myoglobin-like aerotaxis transducers in Archaea and Bacteria. *Nature.* **403**: 540-544.
- Hou, S., C. Belisle, S. Lam, M. Piatibratov, V. Sivozhelezov, H. Takami and M. Alam. 2001a. A globin-coupled oxygen sensor from the facultatively alkaliphilic *Bacillus halodurans* C-125. *Extremophiles.* **5**: 351-354.
- Hou, S., T. Freitas, R. W. Larsen, et al. 2001b. Globin-coupled sensors: a class of heme-containing sensors in Archaea and Bacteria. *Proc Natl Acad Sci U S A.* **98**: 9353-9358.
- Hourdez, S., F. H. Lallier, M. C. De Cian, B. N. Green, R. E. Weber and A. Toulmond. 2000. Gas transfer system in *Alvinella pompejana* (Annelida polychaeta, Terebellida): functional properties of intracellular and extracellular hemoglobins. *Physiol Biochem Zool.* **73**: 365-373.
- Hunt, P. W., R. A. Watts, B. Trevaskis, D. J. Llewelyn, J. Burnell, E. S. Dennis and W. J. Peacock. 2001. Expression and evolution of functionally distinct haemoglobin genes in plants. *Plant Mol Biol.* **47**: 677-692.

- Igamberdiev, A. U., K. Baron, N. Manac'h-Little, M. Stoimenova and R. D. Hill. 2005. The haemoglobin/nitric oxide cycle: involvement in flooding stress and effects on hormone signalling. *Ann Bot (Lond)*. **96**: 557-564.
- Ilari, A., A. Bonamore, A. Farina, K. A. Johnson and A. Boffi. 2002. The X-ray structure of ferric *Escherichia coli* flavohemoglobin reveals an unexpected geometry of the distal heme pocket. *J Biol Chem*. **277**: 23725-23732.
- Imai, K. 1999. The haemoglobin enzyme. *Nature*. **401**: 437, 439.
- Imamura, T., T. O. Baldwin and A. Riggs. 1972. The amino acid sequence of the monomeric hemoglobin component from the bloodworm, *Glycera dibranchiata*. *J Biol Chem*. **247**: 2785-2797.
- Immesberger, A. and T. Burmester. 2004. Putative phenoloxidases in the tunicate *Ciona intestinalis* and the origin of the arthropod hemocyanin superfamily. *J Comp Physiol [B]*. **174**: 169-180.
- Ingram, V. M. 1956. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature*. **178**: 792-794.
- Ingram, V. M. 1957. Gene mutations in human haemoglobin: the chemical difference between normal and sickle cell haemoglobin. *Nature*. **180**: 326-328.
- Ingram, V. M. 1961. Gene evolution and the haemoglobins. *Nature*. **189**: 704-708.
- Jeffreys, A. J. and R. A. Flavell. 1977. The rabbit beta-globin gene contains a large large insert in the coding sequence. *Cell*. **12**: 1097-1108.
- Jensen, E. O., K. Paludan, J. J. Hyldig-Nielsen, J. E. Jorgensen and K. A. Marcker. 1981. The structure of a chromosomal leghaemoglobin gene from soybean. *Nature*. **291**: 677-379.
- Jhiang, S. M., J. R. Garey and A. F. Riggs. 1988. Exon-intron organization in genes of earthworm and vertebrate globins. *Science*. **240**: 334-336.
- Jhiang, S. M. and A. F. Riggs. 1989. The structure of the gene encoding chain c of the hemoglobin of the earthworm, *Lumbricus terrestris*. *J Biol Chem*. **264**: 19003-19008.
- Jones, D. T., W. R. Taylor and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci*. **8**: 275-282.
- Jouan, L., J. C. Taveau, S. Marco, F. H. Lallier and J. N. Lamy. 2001. Occurrence of two architectural types of hexagonal bilayer hemoglobin in annelids: comparison of 3D reconstruction volumes of *Arenicola marina* and *Lumbricus terrestris* hemoglobins. *J Mol Biol*. **305**: 757-771.

- Kandler, R. L., I. Constantinidis and J. D. Satterlee. 1985. Evaluation of the extent of heterogeneity in the *Glycera dibranchiata* monomer haemoglobin fraction by the use of n.m.r. and ion-exchange chromatography. *Biochem J.* **226**: 131-138.
- Karlsen, O. A., L. Ramsevik, L. J. Bruseth, O. Larsen, A. Brenner, F. S. Berven, H. B. Jensen and J. R. Lillehaug. 2005. Characterization of a prokaryotic haemerythrin from the methanotrophic bacterium *Methylococcus capsulatus* (Bath). *Febs J.* **272**: 2428-2440.
- Kawada, N., D. B. Kristensen, K. Asahina, K. Nakatani, Y. Minamiyama, S. Seki and K. Yoshizato. 2001. Characterization of a stellate cell activation-associated protein (STAP) with peroxidase activity found in rat hepatic stellate cells. *J Biol Chem.* **276**: 25318-25323.
- Keilin, D. 1953. Haemoglobin in fungi; occurrence of haemoglobin in yeast and the supposed stabilization of the oxygenated cytochrome oxidase. *Nature.* **172**: 390-393.
- Keilin, D. and J. F. Ryley. 1953. Haemoglobin in Protozoa. *Nature.* **172**: 451.
- Keilin, D. and A. Tissieres. 1954. Haemoglobin in certain strains of yeast (*Saccharomyces cerevisiae*). *Process Biochem.* **57**(329th Meeting): xxix-xxx.
- Keller, E. B. and W. A. Noon. 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. *Proc Natl Acad Sci U S A.* **81**: 7417-7420.
- Kendrew, J. C., R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips and V. C. Shore. 1960. Structure of myoglobin. *Nature.* **185**: 422-427.
- Kersanach, R., H. Brinkmann, M. F. Liaud, D. X. Zhang, W. Martin and R. Cerff. 1994. Five identical intron positions in ancient duplicated genes of eubacterial origin. *Nature.* **367**: 387-389.
- Kim, H. W. and A. G. Greenburg. 2004. Artificial oxygen carriers as red blood cell substitutes: a selected review and current status. *Artif Organs.* **28**: 813-828.
- Kitamoto, Y., X. Yuan, Q. Wu, D. W. McCourt and J. E. Sadler. 1994. Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains. *Proc Natl Acad Sci U S A.* **91**: 7588-7592.
- Kleinschmidt, T. and R. E. Weber. 1998. Primary structures of *Arenicola marina* isomyoglobins: molecular basis for functional heterogeneity. *Biochim Biophys Acta.* **1383**: 55-62.
- Klotz, I. M. and T. A. Klotz. 1955. Oxygen-carrying proteins: a comparison of the oxygenation reaction in hemocyanin and hemerythrin with that in hemoglobin. *Science.* **121**: 477-480.
- Kraus, D. M. and J. M. Colacino. 1986. *Science.* **232**: 90-92.



- Kreutzer, U. and T. Jue. 1996. H-NMR signal of *Arenicola marina* myoglobin in vivo as an index of tissue oxygenation. *Eur J Biochem.* **235**: 622-628.
- Kuchumov, A. R., J. C. Taveau, J. N. Lamy, J. S. Wall, R. E. Weber and S. N. Vinogradov. 1999. The role of linkers in the reassembly of the 3.6 MDa hexagonal bilayer hemoglobin from *Lumbricus terrestris*. *J Mol Biol.* **289**: 1361-1374.
- Kuchumov, A. R., J. A. Loo and S. N. Vinogradov. 2000. Subunit distribution of calcium-binding sites in *Lumbricus terrestris* hemoglobin. *J Protein Chem.* **19**: 139-149.
- Kugelstadt, D., M. Haberkamp, T. Hankeln and T. Burmester. 2004. Neuroglobin, cytoglobin, and a novel, eye-specific globin from chicken. *Biochem Biophys Res Commun.* **325**: 719-725.
- Kundu, S., J. T. Trent, 3rd and M. S. Hargrove. 2003. Plants, humans and hemoglobins. *Trends Plant Sci.* **8**: 387-393.
- Kurtz, D. M. J. 1992. Advances in COmparative and Environmental Physiology. Pp. 151-171 *in*: C. P. Mangum (Ed.) Springer Verlag, Heidelberg.
- LaCount, M. W., E. Zhang, Y. P. Chen, K. Han, M. M. Whitton, D. E. Lincoln, S. A. Woodin and L. Lebioda. 2000. The crystal structure and amino acid sequence of dehaloperoxidase from *Amphitrite ornata* indicate common ancestry with globins. *J Biol Chem.* **275**: 18712-18716.
- Lamy, J., A. Kuchumov, J. C. Taveau, S. N. Vinogradov and J. N. Lamy. 2000. Reassembly of *Lumbricus terrestris* hemoglobin: a study by matrix-assisted laser desorption/ionization mass spectrometry and 3D reconstruction from frozen-hydrated specimens. *J Mol Biol.* **298**: 633-647.
- Lamy, J. N., B. N. Green, A. Toulmond, J. S. Wall, R. E. Weber and S. N. Vinogradov. 1996. Giant Hexagonal Bilayer Hemoglobins. *Chem Rev.* **96**: 3113-3124.
- Lebioda, L., M. W. LaCount, E. Zhang, Y. P. Chen, K. Han, M. M. Whitton, D. E. Lincoln and S. A. Woodin. 1999. An enzymatic globin from a marine worm. *Nature.* **401**: 445.
- Lecomte, J. T., D. A. Vuletich and A. M. Lesk. 2005. Structural divergence and distant relationships in proteins: evolution of the globins. *Curr Opin Struct Biol.* **15**: 290-301.
- Leder, A., E. Wiener, M. J. Lee, S. N. Wickramasinghe and P. Leder. 1999. A normal beta-globin allele as a modifier gene ameliorating the severity of alpha-thalassemia in mice. *Proc Natl Acad Sci U S A.* **96**: 6291-6295.
- Lee, M. S., S. J. Cho, J. A. Lee, B. J. Park, H. J. Cho, J. S. Moon, S. K. Kim, J. K. Choo and S. C. Park. 2004. Transcriptome analysis in the midgut of the earthworm (*Eisenia andrei*) using expressed sequence tags. Unpublished.

- Lightbody, J. J., A. N. Quabar, M. G. Mainwaring, J. S. Young, D. A. Walz, S. N. Vinogradov and T. Gotoh. 1988. Immunological relatedness of annelid extracellular hemoglobins and chlorocruorins. *Comp Biochem Physiol B*. **90**: 301-305.
- Liochev, S. I., A. R. Kuchumov, S. N. Vinogradov and I. Fridovich. 1996. Superoxide dismutase activity in the giant hemoglobin of the earthworm, *Lumbricus terrestris*. *Arch Biochem Biophys*. **330**: 281-284.
- Maevsky, E., G. Ivanitsky, L. Bogdanova, et al. 2005. Clinical results of Perftoran application: present and future. *Artif Cells Blood Substit Immobil Biotechnol*. **33**: 37-46.
- Mahley, R. W. 1988. Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. *Science*. **240**: 622-630.
- Mangum, C. P., B. R. Woodin, C. Bonaventura, B. Sullivan and J. Bonaventura. 1975. The role of coelomic and vascular hemoglobin in the annelid family terebellidae. *Comp Biochem Physiol*. **51A**: 281-294.
- Mangum, C. P. 1985. Oxygen transport in invertebrates. *Am J Physiol*. **248**: R505-514.
- Manwell, C. and C. M. A. Baker. 1988. *Magelona* haemerythrin: tissue specificity, molecular weights and oxygen equilibria. *Comp Biochem Physiol*. **89B**: 453-463.
- Margoliash, E. 1963. Primary Structure and Evolution of Cytochrome C. *Proc Natl Acad Sci U S A*. **50**: 672-679.
- Markl, J. and H. Decker. 1992. Advances in Comparative and Environmental Physiology. Pp. 325-376 *in*: C. P. Mangum (Ed.) Springer Verlag, Heildeberg.
- Mehta, K. D., W. J. Chen, J. L. Goldstein and M. S. Brown. 1991. The low density lipoprotein receptor in *Xenopus laevis*. I. Five domains that resemble the human receptor. *J Biol Chem*. **266**: 10406-10414.
- Meissner, U., M. Stohr, K. Kusche, T. Burmester, H. Stark, J. R. Harris, E. V. Orlova and J. Markl. 2003. Quaternary structure of the European spiny lobster (*Palinurus elephas*) 1x6-mer hemocyanin from cryoEM and amino acid sequence data. *J Mol Biol*. **325**: 99-109.
- Milani, M., A. Pesce, Y. Ouellet, P. Ascenzi, M. Guertin and M. Bolognesi. 2001. Mycobacterium tuberculosis hemoglobin N displays a protein tunnel suited for O<sub>2</sub> diffusion to the heme. *Embo J*. **20**: 3902-3909.
- Milani, M., P. Y. Savard, H. Ouellet, P. Ascenzi, M. Guertin and M. Bolognesi. 2003. A TyrCD1/TrpG8 hydrogen bond network and a TyrB10TyrCD1 covalent link shape the

- heme distal site of Mycobacterium tuberculosis hemoglobin O. Proc Natl Acad Sci U S A. **100**: 5766-5771.
- Milani, M., A. Pesce, M. Nardini, et al. 2005. Structural bases for heme binding and diatomic ligand recognition in truncated hemoglobins. J Inorg Biochem. **99**: 97-109.
- Mizukami, H. and S. N. Vinogradov. 1972. Oxygen association equilibria of Glycera hemoglobins. Biochim Biophys Acta. **285**: 314-319.
- Moens, L., J. Vanfleteren, Y. Van de Peer, K. Peeters, O. Kapp, J. Czeluzniak, M. Goodman, M. Blaxter and S. Vinogradov. 1996. Globins in nonvertebrate species: dispersal by horizontal gene transfer and evolution of the structure-function relationships. Mol Biol Evol. **13**: 324-333.
- Mount, S. M. 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**: 459-472.
- Mukai, M., P. Y. Savard, H. Ouellet, M. Guertin and S. R. Yeh. 2002. Unique ligand-protein interactions in a new truncated hemoglobin from Mycobacterium tuberculosis. Biochemistry. **41**: 3897-3905.
- Nagy, E. and L. E. Maquat. 1998. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. Trends Biochem Sci. **23**: 198-199.
- Nakatani, K., H. Okuyama, Y. Shimahara, S. Saeki, D. H. Kim, Y. Nakajima, S. Seki, N. Kawada and K. Yoshizato. 2004. Cytoglobin/STAP, its unique localization in splanchnic fibroblast-like cells and function in organ fibrogenesis. Lab Invest. **84**: 91-101.
- Negrisoló, 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.
- Nejmeddine, A., A. Dhainaut, J. L. Baert, P. Sautiere, B. Founet and P. Boulenguer. 1988. Purification and characterization of a cadmium-binding protein from *Nereis diversicolor* (annelida, polychaeta). Comp Biochem Physiol C. **89**: 321-326.
- Niall, H. and P. Edman. 1962. The N-terminal amino acids of human plasma proteins. J Gen Physiol. **45(4)Pt 2**: 185-193.
- Numoto, N., T. Nakagawa, A. Kita, Y. Sasayama, Y. Fukumori and K. Miki. 2005. Structure of an extracellular giant hemoglobin of the gutless beard worm *Oligobranchia mashikoi*. Proc Natl Acad Sci U S A.

- Ochiai, T., S. Hoshina and I. Usuki. 1993. Zinc as modulator of oxygenation function and stabilizer of quaternary structure in earthworm hemoglobin. *Biochim Biophys Acta*. **1203**: 310-314.
- Oshino, R., N. Oshino and B. Chance. 1973. Studies on yeast hemoglobin. The properties of yeast hemoglobin and its physiological function in the cell. *Eur J Biochem*. **35**: 23-33.
- Ott, T., J. T. van Dongen, C. Gunther, et al. 2005. Symbiotic leghemoglobins are crucial for nitrogen fixation in legume root nodules but not for general plant growth and development. *Curr Biol*. **15**: 531-535.
- Ouellet, H., Y. Ouellet, C. Richard, M. Labarre, B. Wittenberg, J. Wittenberg and M. Guertin. 2002. Truncated hemoglobin HbN protects *Mycobacterium bovis* from nitric oxide. *Proc Natl Acad Sci U S A*. **99**: 5902-5907.
- 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.
- Palmer, J. D. and J. M. Logsdon, Jr. 1991. The recent origins of introns. *Curr Opin Genet Dev*. **1**: 470-477.
- Park, K. W., K. J. Kim, A. J. Howard, B. C. Stark and D. A. Webster. 2002. Vitreoscilla hemoglobin binds to subunit I of cytochrome bo ubiquinol oxidases. *J Biol Chem*. **277**: 33334-33337.
- Pathania, R., N. K. Navani, A. M. Gardner, P. R. Gardner and K. L. Dikshit. 2002a. Nitric oxide scavenging and detoxification by the *Mycobacterium tuberculosis* haemoglobin, HbN in *Escherichia coli*. *Mol Microbiol*. **45**: 1303-1314.
- Pathania, R., N. K. Navani, G. Rajamohan and K. L. Dikshit. 2002b. *Mycobacterium tuberculosis* hemoglobin HbO associates with membranes and stimulates cellular respiration of recombinant *Escherichia coli*. *J Biol Chem*. **277**: 15293-15302.
- Patthy, L. 1987. Intron-dependent evolution: preferred types of exons and introns. *FEBS Lett*. **214**: 1-7.
- Patthy, L. 1991. Exons--original building blocks of proteins? *Bioessays*. **13**: 187-192.
- Pauling, L., H. A. Itano and et al. 1949. Sickle cell anemia a molecular disease. *Science*. **110**: 543-548.
- Perutz, M. F. 1960. Structure of hemoglobin. *Brookhaven Symp Biol*. **13**: 165-183.
- Perutz, M. F. 1970. Stereochemistry of cooperative effects in haemoglobin. *Nature*. **228**: 726-739.

- Pesce, A., M. Couture, S. Dewilde, M. Guertin, K. Yamauchi, P. Ascenzi, L. Moens and M. Bolognesi. 2000. A novel two-over-two alpha-helical sandwich fold is characteristic of the truncated hemoglobin family. *Embo J.* **19**: 2424-2434.
- Pesce, A., M. Bolognesi, A. Bocedi, P. Ascenzi, S. Dewilde, L. Moens, T. Hankeln and T. Burmester. 2002. Neuroglobin and cytoglobin. Fresh blood for the vertebrate globin family. *EMBO Rep.* **3**: 1146-1151.
- Pesce, A., S. Dewilde, M. Nardini, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester and M. Bolognesi. 2003. Human brain neuroglobin structure reveals a distinct mode of controlling oxygen affinity. *Structure (Camb).* **11**: 1087-1095.
- Poole, R. K., N. J. Rogers, A. D'Mello R, M. N. Hughes and Y. Orii. 1997. Escherichia coli flavohaemoglobin (Hmp) reduces cytochrome c and Fe(III)-hydroxamate K by electron transfer from NADH via FAD: sensitivity of oxidoreductase activity to haem-bound dioxygen. *Microbiology.* **143 ( Pt 5)**: 1557-1565.
- Poole, R. K. and M. N. Hughes. 2000. New functions for the ancient globin family: bacterial responses to nitric oxide and nitrosative stress. *Mol Microbiol.* **36**: 775-783.
- Poole, R. K. 2005. Nitric oxide and nitrosative stress tolerance in bacteria. *Biochem Soc Trans.* **33**: 176-180.
- Raychowdhury, R., J. L. Niles, R. T. McCluskey and J. A. Smith. 1989. Autoimmune target in Heymann nephritis is a glycoprotein with homology to the LDL receptor. *Science.* **244**: 1163-1165.
- Riess, J. G. 2001. Oxygen carriers ("blood substitutes")--raison d'etre, chemistry, and some physiology. *Chem Rev.* **101**: 2797-2920.
- Riggs, A. F. 1991. Aspects of the Origin and Evolution of Non-Vertebrate Hemoglobins. *Amer. Zool.* **31**: 535-545.
- Roach, M. P., Y. P. Chen, S. A. Woodin, D. E. Lincoln, C. R. Lovell and J. H. Dawson. 1997. Notomastus lobatus chloroperoxidase and Amphitrite ornata dehaloperoxidase both contain histidine as their proximal heme iron ligand. *Biochemistry.* **36**: 2197-2202.
- Roche, J., Y. Derrien and M. Roques. 1952. [The heterogeneity of human hemoglobins in the adult and the fetus.]. *C R Seances Soc Biol Fil.* **146**: 689-693.
- Roesner, A., C. Fuchs, T. Hankeln and T. Burmester. 2005. A Globin Gene of Ancient Evolutionary Origin in Lower Vertebrates: Evidence for Two Distinct Globin Families in Animals. *Mol Biol Evol.* **22**: 12-20.
- Roy, S. W. 2003. Recent evidence for the exon theory of genes. *Genetica.* **118**: 251-266.

- Royer, W. E., Jr., K. Strand, M. van Heel and W. A. Hendrickson. 2000. Structural hierarchy in erythrocrucorin, the giant respiratory assemblage of annelids. *Proc Natl Acad Sci U S A*. **97**: 7107-7111.
- Rudenko, G., L. Henry, K. Henderson, K. Ichtchenko, M. S. Brown, J. L. Goldstein and J. Deisenhofer. 2002. Structure of the LDL receptor extracellular domain at endosomal pH. *Science*. **298**: 2353-2358.
- Runnegar, B. 1984. Derivation of the globins from type b cytochromes. *J Mol Evol*. **21**: 33-41.
- Saito, A., S. Pietromonaco, A. K. Loo and M. G. Farquhar. 1994. Complete cloning and sequencing of rat gp330/"megalin," a distinctive member of the low density lipoprotein receptor gene family. *Proc Natl Acad Sci U S A*. **91**: 9725-9729.
- Salvato, B., M. Santamaria, M. Beltramini, G. Alzuet and L. Casella. 1998. The enzymatic properties of *Octopus vulgaris* hemocyanin: o-diphenol oxidase activity. *Biochemistry*. **37**: 14065-14077.
- Sanger, F., G. G. Brownlee and B. G. Barrell. 1965. A two-dimensional fractionation procedure for radioactive nucleotides. *J Mol Biol*. **13**: 373-398.
- Sanna, M. T., B. Manconi, M. Castagnola, et al. 2005. Functional and structural characterization of the myoglobin from the polychaete *Ophelia bicornis*. *Biochem J*. **389**: 497-505.
- Schmidt, M., A. Giessl, T. Laufs, T. Hankeln, U. Wolfrum and T. Burmester. 2003. How does the eye breathe? Evidence for neuroglobin-mediated oxygen supply in the mammalian retina. *J Biol Chem*. **278**: 1932-1935.
- Schmidt, M., F. Gerlach, A. Avivi, et al. 2004. Cytoglobin is a respiratory protein in connective tissue and neurons, which is up-regulated by hypoxia. *J Biol Chem*. **279**: 8063-8069.
- Seamonds, B., R. E. Forster and A. J. Gottlieb. 1971. Heterogeneity of the hemoglobin from the common bloodworm *Glycera dibranchiata*. *J Biol Chem*. **246**: 1700-1705.
- Shafie, S. M., S. N. Vinogradov, L. Larson and J. J. McCormick. 1976. RNA and protein synthesis in the nucleated erythrocytes of *Glycera dibranchiata*. *Comp Biochem Physiol B*. **53**: 85-88.
- Sherman, D. R., A. P. Kloek, B. R. Krishnan, B. Guinn and D. E. Goldberg. 1992. *Ascaris* hemoglobin gene: plant-like structure reflects the ancestral globin gene. *Proc Natl Acad Sci U S A*. **89**: 11696-11700.

- Shishikura, F. 2004. The complete amino acid sequences of four globins from the land leech *Haemadipsa zeylanica* var. *japonica*. *Comp Biochem Physiol B Biochem Mol Biol.* **137**: 443-454.
- Shlom, J. M. and S. N. Vinogradov. 1973. A study of the subunit structure of the extracellular hemoglobin of *Lumbricus terrestris*. *J Biol Chem.* **248**: 7904-7912.
- Simonovic, M., K. Dolmer, W. Huang, D. K. Strickland, K. Volz and P. G. Gettins. 2001. Calcium coordination and pH dependence of the calcium affinity of ligand-binding repeat CR7 from the LRP. Comparison with related domains from the LRP and the LDL receptor. *Biochemistry.* **40**: 15127-15134.
- Standley, P. R., M. G. Mainwaring, T. Gotoh and S. N. Vinogradov. 1988. The calcium, copper and zinc content of some annelid extracellular haemoglobins. *Biochem J.* **249**: 915-916.
- Stapleton, M., J. Carlson, P. Brokstein, et al. 2002. A *Drosophila* full-length cDNA resource. *Genome Biol.* **3**: RESEARCH0080.
- Stevanin, T. M., R. K. Poole, E. A. Demoncheaux and R. C. Read. 2002. Flavohemoglobin Hmp protects *Salmonella enterica* serovar typhimurium from nitric oxide-related killing by human macrophages. *Infect Immun.* **70**: 4399-4405.
- Strand, K., J. E. Knapp, B. Bhyravhatla and W. E. Royer, Jr. 2004. Crystal structure of the hemoglobin dodecamer from *Lumbricus erythrocrutorin*: allosteric core of giant annelid respiratory complexes. *J Mol Biol.* **344**: 119-134.
- Sudhof, T. C., J. L. Goldstein, M. S. Brown and D. W. Russell. 1985a. The LDL receptor gene: a mosaic of exons shared with different proteins. *Science.* **228**: 815-822.
- Sudhof, T. C., D. W. Russell, J. L. Goldstein, M. S. Brown, R. Sanchez-Pescador and G. I. Bell. 1985b. Cassette of eight exons shared by genes for LDL receptor and EGF precursor. *Science.* **228**: 893-895.
- Sun, Y., K. Jin, X. O. Mao, Y. Zhu and D. A. Greenberg. 2001. Neuroglobin is up-regulated by and protects neurons from hypoxic-ischemic injury. *Proc Natl Acad Sci U S A.* **98**: 15306-15311.
- Sun, Y., K. Jin, A. Peel, X. O. Mao, L. Xie and D. A. Greenberg. 2003. Neuroglobin protects the brain from experimental stroke in vivo. *Proc Natl Acad Sci U S A.* **100**: 3497-3500.
- Suzuki, T., T. Takagi and T. Gotoh. 1982. Amino acid sequence of the smallest polypeptide chain containing heme of extracellular hemoglobin from the polychaete *Tylorrhynchus heterochaetus*. *Biochim Biophys Acta.* **708**: 253-258.

- Suzuki, T., H. Yasunaga, T. Furukohri, K. Nakamura and T. Gotoh. 1985. Amino acid sequence of polypeptide chain IIB of extracellular hemoglobin from the polychaete *Tylorrhynchus heterochaetus*. *J Biol Chem.* **260**: 11481-11487.
- Suzuki, T., O. H. Kapp and T. Gotoh. 1988. Novel S-S loops in the giant hemoglobin of *Tylorrhynchus heterochaetus*. *J Biol Chem.* **263**: 18524-18529.
- Suzuki, T., T. Takagi and T. Gotoh. 1990. Primary structure of two linker chains of the extracellular hemoglobin from the polychaete *Tylorrhynchus heterochaetus*. *J Biol Chem.* **265**: 12168-12177.
- Suzuki, T., T. Takagi and S. Ohta. 1990a. Primary structure of a constituent polypeptide chain (AIII) of the giant haemoglobin from the deep-sea tube worm *Lamellibrachia*. A possible H<sub>2</sub>S-binding site. *Biochem J.* **266**: 221-225.
- Suzuki, T., T. Takagi and S. Ohta. 1990b. Primary structure of a linker subunit of the tube worm 3000-kDa hemoglobin. *J Biol Chem.* **265**: 1551-1555.
- Suzuki, T. and A. F. Riggs. 1993. Linker chain L1 of earthworm hemoglobin. Structure of gene and protein: homology with low density lipoprotein receptor. *J Biol Chem.* **268**: 13548-13555.
- Suzuki, T., T. Takagi and S. Ohta. 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**: 141-146.
- Suzuki, T., T. Ohta, H. J. Yuasa and T. Takagi. 1994. The giant extracellular hemoglobin from the polychaete *Neanthes diversicolor*. The cDNA-derived amino acid sequence of linker chain L2 and the exon/intron boundary conserved in linker genes. *Biochim Biophys Acta.* **1217**: 291-296.
- 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**: 189-193.
- Suzuki, T. and K. Imai. 1998. Evolution of myoglobin. *Cell Mol Life Sci.* **54**: 979-1004.
- Suzuki, T. and S. N. Vinogradov. 2003. Globin and linker sequences of the giant extracellular hemoglobin from the leech *Macrobdella decora*. *J Protein Chem.* **22**: 231-242.
- Swank, R. T. and K. D. Munkres. 1971. Molecular weight analysis of oligopeptides by electrophoresis in polyacrylamide gel with sodium dodecyl sulfate. *Anal Biochem.* **39**: 462-477.



- Takagi, T. and J. A. Cox. 1991. Primary structure of myohemerythrin from the annelid *Nereis diversicolor*. FEBS Lett. **285**: 25-27.
- Takahashi, S., Y. Kawarabayasi, T. Nakai, J. Sakai and T. Yamamoto. 1992. Rabbit very low density lipoprotein receptor: a low density lipoprotein receptor-like protein with distinct ligand specificity. Proc Natl Acad Sci U S A. **89**: 9252-9256.
- Telfer, W. H. and J. G. Kunkel. 1991. The function and evolution of insect storage hexamers. Annu Rev Entomol. **36**: 205-228.
- Tensen, C. P., E. R. Van Kesteren, R. J. Planta, K. J. Cox, J. F. Burke, H. van Heerikhuizen and E. Vreugdenhil. 1994. A G protein-coupled receptor with low density lipoprotein-binding motifs suggests a role for lipoproteins in G-linked signal transduction. Proc Natl Acad Sci U S A. **91**: 4816-4820.
- Terwilliger, N. 1992. Molecular structure of the extracellular heme proteins. Adv Comp Environ Physiol. **13**: 193-229.
- Terwilliger, N. B. and K. O'Brien. 1992. *Cancer magister* 'Non respiratory protein' is a molt-related hemocyanin. Am Zool. **32**: 44A.
- Terwilliger, N. B. 1998. Functional adaptations of oxygen-transport proteins. J Exp Biol. **201 (Pt 8)**: 1085-1098.
- Terwilliger, R. C. 1974. Oxygen equilibria of the vascular and coelomic hemoglobins of the terebellid polychaete, *Pista pacifica*. Evidence for an oxygen transfer system. Comp Biochem Physiol A. **48**: 745-755.
- Terwilliger, R. C., R. L. Garlick and N. B. Terwilliger. 1976. Hemoglobins of *Glycera robusta*: structures of coelomic cell hemoglobin and body wall myoglobin. Comp Biochem Physiol B. **54**: 149-153.
- Terwilliger, R. C., R. L. Garlick and N. B. Terwilliger. 1980. Characterization of hemoglobins and myoglobin of *Travisia foetida*. Comp Biochem Physiol. **66B**: 261-266.
- Toulmond, A. 1970a. [Oxygen fixation by blood in the arenicola (*Arenicola marina*(L.), Annelida, Polychaeta)]. C R Acad Sci Hebd Seances Acad Sci D. **270**: 1368-1371.
- Toulmond, A. 1970b. [Fixation of carbon dioxide by the blood in the arenicola (*Arenicola marina* (L.), Annelida, Polychaeta)]. C R Acad Sci Hebd Seances Acad Sci D. **270**: 1487-1490.
- Toulmond, A. 1977. Temperature-induced variations of blood acid-base status in the lugworm, *Arenicola marina* (L.): I. In vitro study. Respir Physiol. **31**: 139-149.

- Toulmond, A. 1979. [Determination of apparent mean mass of proteins associated with heme in the hemoglobin molecule of *Arenicola marina* (L.), Annelida, Polychaeta]. C R Seances Acad Sci D. **288**: 651-654.
- Toulmond, A. 1992. Properties and functions of extracellular heme pigments. In Advance in comparative and environmental physiology (Mangum, C.P., ed.), Vol 13. 231-256.
- Trent, J. T., 3rd and M. S. Hargrove. 2002. A ubiquitously expressed human hexacoordinate hemoglobin. J Biol Chem. **277**: 19538-19545.
- Trevaskis, B., R. A. Watts, C. R. Andersson, D. J. Llewellyn, M. S. Hargrove, J. S. Olson, E. S. Dennis and W. J. Peacock. 1997. Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins. Proc Natl Acad Sci U S A. **94**: 12230-12234.
- Ullmann, B. D., H. Myers, W. Chiranand, A. L. Lazzell, Q. Zhao, L. A. Vega, J. L. Lopez-Ribot, P. R. Gardner and M. C. Gustin. 2004. Inducible defense mechanism against nitric oxide in *Candida albicans*. Eukaryot Cell. **3**: 715-723.
- van Driel, I. R., J. L. Goldstein, T. C. Sudhof and M. S. Brown. 1987. First cysteine-rich repeat in ligand-binding domain of low density lipoprotein receptor binds Ca<sup>2+</sup> and monoclonal antibodies, but not lipoproteins. J Biol Chem. **262**: 17443-17449.
- van Gelder, C. W., W. H. Flurkey and H. J. Wichers. 1997. Sequence and structural features of plant and fungal tyrosinases. Phytochemistry. **45**: 1309-1323.
- van Holde, K. E. and K. I. Miller. 1995. Hemocyanins. Adv Protein Chem. **47**: 1-81.
- van Holde, K. E., K. I. Miller and H. Decker. 2001. Hemocyanins and invertebrate evolution. J Biol Chem. **276**: 15563-15566.
- Vinogradov, S. N., C. A. Machlik and L. L. Chao. 1970. The intracellular hemoglobins of a polychaete. Some properties of the hemoglobins of *Glycera dibranchiata*. J Biol Chem. **245**: 6533-6538.
- Vinogradov, S. N. 1985. The structure of invertebrate extracellular hemoglobins (erythrocrucorins and chlorocruorins). Comp Biochem Physiol B. **82**: 1-15.
- Vinogradov, S. N., S. D. Lugo, M. G. Mainwaring, O. H. Kapp and A. V. Crewe. 1986. Bracelet protein: a quaternary structure proposed for the giant extracellular hemoglobin of *Lumbricus terrestris*. Proc Natl Acad Sci U S A. **83**: 8034-8038.
- Vinogradov, S. N., D. A. Walz, B. Pohajdak, L. Moens, O. H. Kapp, T. Suzuki and C. N. Trotman. 1993. Adventitious variability? The amino acid sequences of nonvertebrate globins. Comp Biochem Physiol B. **106**: 1-26.

- Vinogradov, S. N., D. Hoogewijs, X. Bailly, R. Arredondo-Peter, M. Guertin, J. Gough, S. Dewilde, L. Moens and J. R. Vanfleteren. 2005. Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. *Proc Natl Acad Sci U S A.* **102**: 11385-11389.
- Wakabayashi, S., H. Matsubara and D. A. Webster. 1986. Primary sequence of a dimeric bacterial haemoglobin from *Vitreoscilla*. *Nature.* **322**: 481-483.
- Watts, R. A., P. W. Hunt, A. N. Hvitved, M. S. Hargrove, W. J. Peacock and E. S. Dennis. 2001. A hemoglobin from plants homologous to truncated hemoglobins of microorganisms. *Proc Natl Acad Sci U S A.* **98**: 10119-10124.
- Weber, R. E. 1970. Relations between functional and molecular properties of Annelid Haemoglobins. Interaction between Haems in the Haemoglobin of *Arenicola Marina* L. *Comp Biochem Physiol.* **35**: 179.
- Weber, R. E. and E. Paupitit. 1972. Molecular and functional heterogeneity in myoglobin from the polychaete *Arenicola marina* L. *Arch Biochem Biophys.* **148**: 322-324.
- Weber, R. E., C. Mangum, H. Steinman, C. Bonaventura, B. Sullivan and J. Bonaventura. 1977. Hemoglobins of two terebellid polychaetes: *Enoplobranchus sanguineus* and *Amphitrite ornata*. *Comp Biochem Physiol A.* **56**: 179-187.
- Weber, R. E. 1978. Respiratory Pigments. Pp 393-446 in *Physiology of Annelids*. P.J Mill eds. Academic Press Inc, New York.
- Weber, R. E. 1981. Cationic control of O<sub>2</sub> affinity in lugworm erythrocrurin. *Nature.* **292**: 386 - 387.
- Weber, R. E. and S. N. Vinogradov. 2001. Nonvertebrate hemoglobins: functions and molecular adaptations. *Physiol Rev.* **81**: 569-628.
- Wells, R. M. and L. M. Warren. 1982. Oxygen binding of erythrocrurin and coelomic cell haemoglobin from the terebellid polychaete *Neoamphitrite figulus* related to some environmental factors. *J Mar Biol Assoc UK.* **62**: 373-384.
- Williams, E. J., C. Pal and L. D. Hurst. 2000. The molecular evolution of signal peptides. *Gene.* **253**: 313-322.
- Wittenberg, B. A., R. W. Briehl and J. B. Wittenberg. 1965. Haemoglobins of invertebrate tissues. Nerve haemoglobins of *Aphrodite*, *Aplysia* and *Halosydna*. *Biochem J.* **96**: 363-371.
- Wittenberg, J. B. 1970. Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. *Physiol Rev.* **50**: 559-636.

- Wittenberg, J. B., M. Bolognesi, B. A. Wittenberg and M. Guertin. 2002. Truncated hemoglobins: a new family of hemoglobins widely distributed in bacteria, unicellular eukaryotes, and plants. *J Biol Chem.* **277**: 871-874.
- Woese, C. R., O. Kandler and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A.* **87**: 4576-4579.
- Wohlgemuth, S. E., A. C. Taylor and M. K. Grieshaber. 2000. Ventilatory and metabolic responses to hypoxia and sulphide in the lugworm *Arenicola marina* (L.). *J Exp Biol.* **203 Pt 20**: 3177-3188.
- Wu, G., L. M. Wainwright and R. K. Poole. 2003. Microbial globins. *Adv Microb Physiol.* **47**: 255-310.
- Yochem, J. and I. Greenwald. 1993. A gene for a low density lipoprotein receptor-related protein in the nematode *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A.* **90**: 4572-4576.
- Yuasa, H. J., B. N. Green, T. Takagi, N. Suzuki, S. N. Vinogradov and T. Suzuki. 1996. Electrospray ionization mass spectrometric composition of the 400 kDa hemoglobin from the pogonophoran *Oligobranchia mashikoi* and the primary structures of three major globin chains. *Biochim Biophys Acta.* **1296**: 235-244.
- Zafar, R. S., L. H. Chow, M. S. Stern, J. S. Scully, P. R. Sharma, S. N. Vinogradov and D. A. Walz. 1990a. The cDNA sequences encoding two components of the polymeric fraction of the intracellular hemoglobin of *Glycera dibranchiata*. *J Biol Chem.* **265**: 21843-21851.
- Zafar, R. S., L. H. Chow, M. S. Stern, S. N. Vinogradov and D. A. Walz. 1990b. The heterogeneity of the polymeric intracellular hemoglobin of *Glycera dibranchiata* and the cDNA-derived amino acid sequence of one component. *Biochim Biophys Acta.* **1041**: 117-122.
- Zal, F. 1996. Structures des hémoglobines extracellulaires d'annélides et de vestimentifères colonisant des milieux extrêmes. Thèse de doctorat. Pierre et Marie Curie, Paris VI, Paris.
- Zal, F., F. H. Lallier, B. N. Green, S. N. Vinogradov and A. Toulmond. 1996a. The multi-hemoglobin system of the 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 the hydrothermal vent tube worm *Riftia pachyptila*. I. Reexamination of the number and masses of its constituents. *J Biol Chem.* **271**: 8869-8874.
- Zal, F., B. N. Green, F. H. Lallier and A. Toulmond. 1997a. Investigation by electrospray ionization mass spectrometry of the extracellular hemoglobin from the polychaete annelid *Alvinella pompejana*: an unusual hexagonal bilayer hemoglobin. *Biochemistry.* **36**: 11777-11786.
- Zal, F., B. N. Green, F. H. Lallier, S. N. Vinogradov and A. Toulmond. 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., T. Suzuki, Y. Kawasaki, J. J. Childress, F. H. Lallier and A. Toulmond. 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**: 562-574.
- 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 U S A.* **95**: 8997-9002.
- Zal, F., T. Gotoh and A. Toulmond. 1999. Fifth International Congress of Comparative Physiology and Biochemistry. *Comp Biochem Physiol.* 124A.
- Zal, F., F. Lallier and A. Toulmond. 2003. International patent submission N°10/296982.
- Zal, F., C. Chabasse, X. Bailly and M. Rousselot. 2003. Premières étapes de la synthèse in vitro d'hémoglobine extracellulaire de haut poids moléculaire, en vue d'élaborer un substitut sanguin. International patent submission N°04/02602.
- Zhang, J., X. Sun, Y. Qian and L. E. Maquat. 1998a. Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. *RNA.* **4**: 801-815.
- Zhang, W. and G. N. Phillips, Jr. 2003. Structure of the oxygen sensor in *Bacillus subtilis*: signal transduction of chemotaxis by control of symmetry. *Structure (Camb).* **11**: 1097-1110.
- Zhu, H. and A. F. Riggs. 1992. Yeast flavohemoglobin is an ancient protein related to globins and a reductase family. *Proc Natl Acad Sci U S A.* **89**: 5015-5019.

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

Zuckermandl, E. 1965. The Evolution of Hemoglobin. *Sci Am.* **212**: 110-118.

Zuckermandl, E. and L. Pauling. 1965. Molecules as documents of evolutionary history. *J Theor Biol.* **8**: 357-366.

# **Annexes**

## ANNEXES

### Annexe 1 :

#### **Premières étapes de la synthèse *in vitro* d'hémoglobine extracellulaire de haut poids moléculaire, en vue d'élaborer un substitut sanguin.**

Franck Zal, Christine Chabasse, Xavier Bailly et Morgane Rousselot.

#### **Brevet international N° 04/02602**

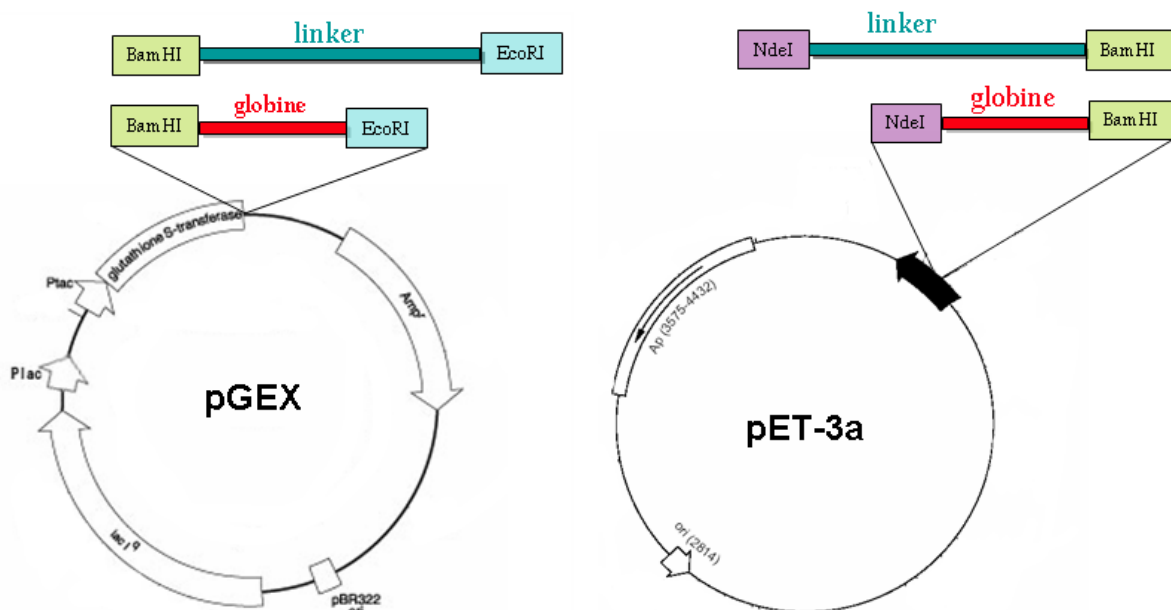
Ce brevet présente d'une part les protocoles de dissociation permettant l'obtention des huit sous-unités de globines et des deux sous-unités de chaînes de structure (linkers) de l'hémoglobine extracellulaire de *Arenicola marina*, et d'autre part les protocoles ayant permis l'obtention de cinq séquences d'ADNc de globines et d'une séquence d'ADNc de linker. Ces protocoles sont détaillés en annexe 2.



**Annexe 2 : Résultats préliminaires de la surexpression d'un monomère de globine et de linker de *Arenicola marina*.**

Ces travaux ont été entrepris par Xavier Bailly et Thomas Harnois au cours de la dernière année de cette thèse, à partir des ADNc caractérisés au cours de cette thèse.

A partir des ADNc complets du monomère A2a et du linker L2 nous avons entrepris la construction de deux types de vecteurs d'expression. L'un est issu de la lignée des plasmides **pET** et l'autre de la lignée des plasmides **pGEX** (figure 33). Des sites de restrictions différents flanquent les extrémités 5' et 3' des ADNc matrices insérés dans les vecteurs d'expression pour orienter la séquence codante dans le bon cadre de lecture, afin que l'extrémité 5' (ou NH<sub>2</sub>-terminale de la protéine correspondante) soit placée directement en aval du promoteur. Ces plasmides ont été utilisés pour transformer des bactéries BL21 DE3 directement utilisées pour des cultures bactériennes de surexpression.

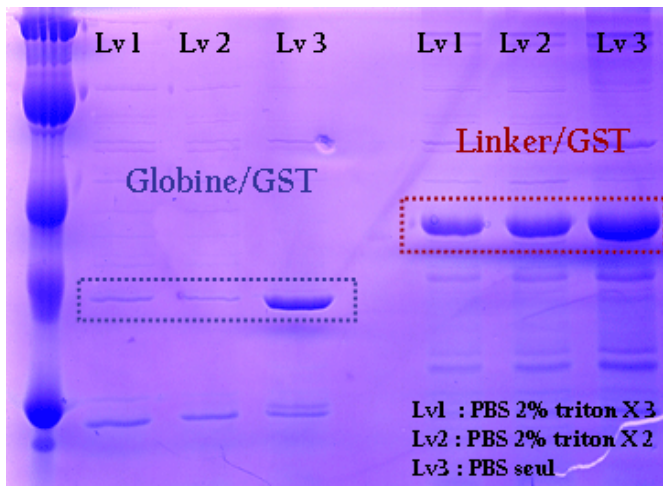


**Figure 33** : Les deux plasmides pGEX et pET ont été utilisés pour la mise au point de la surexpression d'un linker et d'une globine de *Arenicola marina*.

La nature de ces deux plasmides illustre la stratégie de mise en place de la surexpression de l'hémoglobine de *Arenicola marina*.

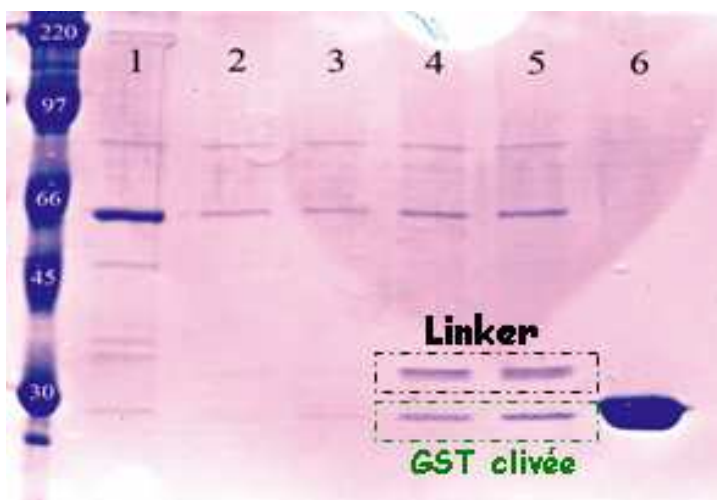
**Le plasmide pGex** permet de produire des protéines recombinantes (dites de fusion) car la protéine d'intérêt est simultanément exprimée avec la protéine GST (Glutathion S

Transférase) reliée par une courte séquence polypeptidique qui comporte un site de clivage au facteur Xa. L'affinité de la GST pour le glutathion (fixé sur des billes de sépharose) est une propriété directement utilisée pour la purification de la protéine de fusion. Cette étape permet d'extraire d'un lysat de protéines bactériennes, la protéine de fusion surexprimée. Nous avons réussi à surexprimer la globine A2 ainsi que le linker de *Arenicola marina* (Fig. 34).



**Figure 34 :** Sur ce gel d'acrylamide coloré au bleu de coomassie on observe la globine (les 3 premiers puits) et le linker de fusion (les trois derniers puits) exprimés puis purifiés dans 3 tampons de lyse différents (Lv1 : PBS 2% triton 3X, Lv2 : PBS 2% triton 2X et Lv3 : PBS seul).

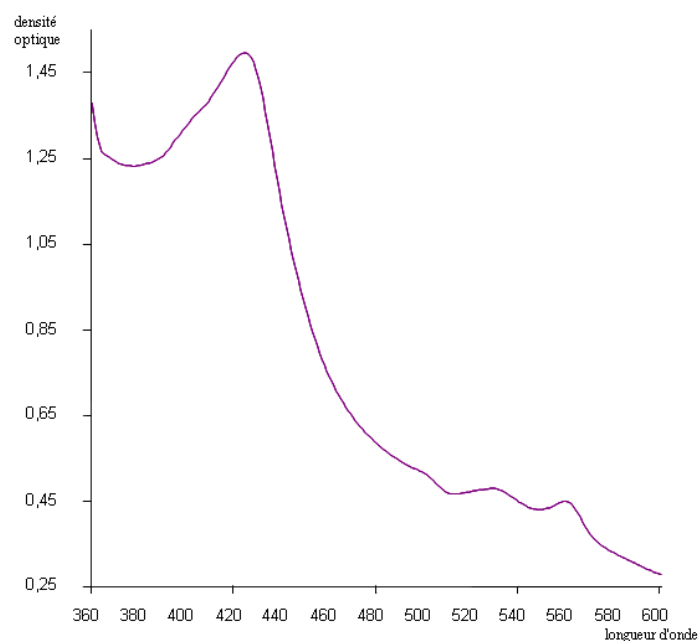
La possibilité d'utiliser le facteur Xa, une protéase qui clive spécifiquement la séquence polypeptidique reliant les deux protéines fusionnées, permet après purification du complexe GST-globine ou GST-linker de purifier avec plus de facilité la protéine recombinante, et à posteriori, de la caractériser spécifiquement sans protéines bactériennes contaminantes (Fig. 35).



**Figure 35 :** Digestion au facteur Xa à 1, 2, 3 et 4 heures (puits 2, 3, 4 et 5) à partir de la protéine fusion GST-linker fixée sur bille (puits 1). On remarquera l'augmentation de l'intensité de la bande vers 32 kD (le linker) au cours du temps.

**Le plasmide pET** permet la production de protéines recombinantes seules : le monomère de globine, de la méthionine initiatrice au codon stop. L'intérêt de cette construction est de s'affranchir des étapes de purification de la GST et de la digestion au facteur Xa. L'utilisation du plasmide pET repose donc sur la mise au point d'un protocole de purification plus direct de la protéine recombinante fonctionnelle, indispensable pour la production à l'échelle industrielle d'un substitut sanguin.

A partir de cultures surexprimant la globine non taguée (vecteur Pet3a) et en présence du groupement prosthétique indispensable pour la fonctionnalité de la protéine (l'hémine), nous avons pu caractériser la présence de la globine surexprimée dans le surnageant d'une culture bactérienne lysée. Le spectre d'absorption entre 360 et 600 nm (Fig. 36) est spécifique des hémoglobines en solution avec un pic caractéristique vers 414-420 nm.



**Figure 36** : Spectre d'absorption de la globine recombinante de *Arenicola marina*

Ces résultats préliminaires montrent que le recours aux techniques de surexpression devrait permettre d'envisager la production à plus grande échelle de larges quantités de monomères de globine de *Arenicola marina*. Le but final étant de pouvoir reconstruire une protéine multimérique et fonctionnelle à partir d'un ensemble de monomères de globines et de linkers.

### **Annexe 3 : Matériel et Méthode**

#### **A- 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 composé. 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 Tris-EDTA (10 mM Tris-HCl, 1 mM EDTA, pH 7,4).

#### **B- Purification des ARN messagers avec le kit mRNA Purification kit (Amersham).**

Le pool des ARN totaux est passé sur une colonne de chromatographie d'affinité, à travers de la cellulose ayant des extensions oligo(dT). En théorie, seuls les ARNm polyadénylés s'accrochent et sont secondairement élués, après les ARN de transfert et ribosomaux.

Après élution dans un tampon (10 mM Tris-HCl, 1 mM EDTA, pH 7,4), les ARNm sont immédiatement utilisés pour les expériences de Reverse Transcription ou de RACE PCR.

#### **C- Reverse Transcription.**

La Reverse Transcription est la synthèse d'un brin d'ADN complémentaire à partir d'une matrice ARN grâce à l'ADN polymérase ARN dépendante. Cette réaction est réalisée à partir de 0.5-2µg d'ARNm, 1 mM de dNTP, 50 pmoles d'oligo dT, 200 Unités de M-MLV-RT (Moloney Murine Leukemia Virus Reverse Transcriptase), 10mM Tris-HCl (pH 8.3), 15mM KCl, 1mM MgCl<sub>2</sub> and 2mM DTT, dans un volume final de 20 µl. Le mélange réactionnel sans la M-MLV-RT est d'abord chauffé à 72°C pendant 5 minutes pour détruire les structures secondaires des ARNm, puis incubé pendant 1 heure à 42°C avec la M-MLV-RT. Un choc thermique de 3min à 95°C stoppe la réaction en détruisant la M-MLV-RT. L'ADNc simple brin est ensuite conservé à -20°C jusqu'à son utilisation.

D- 5' et 3' RACE PCR avec le kit 5'/3' Race Kit (Roche).

**5' RACE PCR :** Cette technique permet l'amplification de la partie 5' d'une séquence d'ARN messager. Une fois le premier brin d'ADNc synthétisé par Reverse Transcription grâce à une amorce spécifique de la séquence d'intérêt (appelée SP1), une extension poly(dA) est ajoutée par une terminale transférase à l'extrémité 3' de l'ADNc. L'ADNc polyadénilé ainsi obtenu va être amplifié par PCR grâce à l'utilisation d'une autre amorce spécifique de la séquence d'intérêt (appelée SP2) et d'un oligo-dT.

**3' RACE PCR :** Cette technique utilise l'extension polyadénylée naturelle des ARN messagers. La première étape consiste à obtenir un ADNc par Reverse Transcription en utilisant un oligo(dT). Puis la séquence sera amplifiée par PCR en utilisant une amorce spécifique de la séquence d'intérêt (appelée SP5) ainsi qu'un oligo-dT.

E- Extraction d'ADN génomique.

Les tissus sont lavés une première fois avec le de tampon Protéinase K (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8, 100 mM NaCl) puis mis à incuber pendant 2h à 65°C dans 700µL de Tampon Protéinase K + 15 µl de protéinase K (10 mg/ml).

Une fois les tissus digérés, on effectue une première extraction au phénol (pH 8) puis une extraction phénol/chloroforme/ alcool isoamylique [25/24/1] et enfin une extraction au chloroforme. L'ADN est ensuite précipité par addition d'isopropanol froid [1/1] et incubation à -20°C pendant 2 heures. Le culot d'ADN est ensuite lavé à l'alcool 75% puis séché sous vide. L'ADN est ensuite dissous dans un tampon TE (10mM Tris-HCl, 0.1mM EDTA, pH 8) et placé à +4°C.

F- Marche sur le chromosome.

Cette technique permet de déterminer les séquences adjacentes d'une région connue du génome. La première étape consiste à amplifier par PCR la séquence voulue à partir d'une amorce spécifique de la séquence d'intérêt et de quatre amorces dégénérées possédant une partie commune (Ancre) en 5'. Cette étape, réalisée avec de faible température d'hybridation, permet aux amorces dégénérées de s'hybrider de façon aléatoire sur le génome. La deuxième étape de PCR utilise une amorce emboîtée spécifique de la séquence d'intérêt, ainsi que l'ancre commune aux quatre premières amorces.

**Annexe 4 : Séquences nucléotidiques et protéiques de globines et de linkers de *Arenicola marina* obtenus pendant la thèse.**

A- Séquences nucléotidiques de globines et de linkers de *Arenicola marina*.

**Globine A2a :** numéro accession AJ880690

(En rouge: partie 5'UTR. En vert: séquence codant pour le peptide signal. En noir: séquence codant pour le polypeptide mature. En bleu : partie 3'UTR)

AGCAGATCCTGCCCCACAACCTGAGGGACAGCCGCTTCAGTAACCTTAAAGGTTAGCCCTGAG  
 ATCCTCGATACACCCAGCATGAAGTCCTTGGTGGTTCTGTTTCGCCCTGGTGGCCATGGTGGC  
 TGCAGAGTGCGGCCCCATGCAGCGCCTCCTGGTCAAGACCCAGTGGAACAAGGTGTACGGCAC  
 CAGCAAGGTCAGGGACGAGGCCGGACACGTCTCTGGAAGGCTATTTTCGCCCCAGGATCCCGA  
 GACCCGGGCTCTCTTCAAGAGAGTCAACGGTGACGACATCTACTCTCCCGAGTTCATGGCTCA  
 CAGCGCCCGTGTCTTGGGTGGCCTTGACATTGCCATCTCCCTCCTCGACAACCAGGCTGACCT  
 TGACGTCGCCCTGGCTCACCTTCACGTGCAGCACGTAGAAAGGCACATCCCAACCCGCTACTT  
 CGATCTGTTCAAGAACGCCCTGATGGAGTATGCCCCAGCGCCCTGGGACGCTGCTTCGATAA  
 GACCGCCTGGAGCTCGTGCTTTGACGTCATCGCCAACGGCATCAAGGAATAGGCTGCAGTAAG  
 CCACGTGACTCATGACGTCATAACCCTCTTCAACACGCGGAATGGCGACGTGACAGCACCAACA  
 TAAACTTTTTCACGACACCTCGTCTGTCTGGCCTAGCTGACAGCTGCGATACTCATAACAGGCCT  
 CTCCTCACATCCCGTCGACCCAGAGCGTTATCTGAACTAGCTGTCCATTCCATCACCATCAAA  
 GACAGCGGGTGTATAGATCTCAGGGCAAATTATAGATCTTCAATAGTCGAAGGGAAATACCGC  
 TTATCTCTGGTTTTTGGCTACCTTTGTCCCATGAGCGCCACTCATATAGAGACATCGCCATCT  
 GGTTCCTTCAACGGTGGGTCCAAAGCTTTGACATCGGCAGTATTTCTTTTCTCGTATCCAGA  
 TGGCGTATTTCTGATTTCGACAGGGTGTCTGGATCCTGTAGAGGATTACCTCAGCAAGTTGTTT  
 ACCAGCAGAATCACCGAATAAAAAATAAAATCAAAAAGCTAAAAAAAAAAAAAGAAAAAAAAAA  
 AAAAA

**Globine A2b:** numéro accession AM109949

AAACANGCGTATCGATGTCGACTTTTTTTTTTTTTTTTCGCAGATCCTGCCCCACATCTAGAGG  
 GACAGCCGCTTCAGTAACCTTAGAGGTTAGCCCTGAGATCCTCGATACACCCAGCATGAAGT  
 TCTTGGTGGTTCTGTTTCGCCCTGGTGGCCATGGTTCGCTGCTGATTGTGGCCCCATGCAGCGCC  
 TCCTGGTCAAGGCCAGTGGAACAAGGTGTACGGCACCAGCAAGGTCAGGGACGACGCCGGAC  
 ACGTCTCTGGAAGGCTATCTTCAACCAGGATGGTGTAGACCCGCGCCCTCTTCAACAGAGTGC  
 ACGGTGACGACATCTACTCTCCCGAGTTCATGGCTCACAGCGCCCGTGTCTTGGGTGGCCTTG  
 ACATTGCCATCTCCCTCCTCGACAACCAGGCTGAGCTTGACGCTGTCCTGGCTCACCTCAAGG  
 AGCAGCACATTGAGAGGGGGATCCCAGACCGTTACTTCGACCTGTTCAAGAACGCCCTGATGG  
 AGTTTGCCCCAGCGCCTTGGGACGCTGCTTCATAAAGGACGCTTGGAGCTCATGCTTTGACG  
 TCATTGCCAACGGCATCAAGGGACAGTAAACGTCATTGCCAACGGCATCAAGGGACAGTAAAGT  
 GTCCTAACCGGCCTCCGGGTGGCAGCACCACCTAAGTTCCTTCATCACTCAACTCTTCATCGA  
 ACCAACTGTATCGCTTCAGTCAAATGAAGGGATATTCAATGCACCGACAGAAACAATAGCACC  
 AGAACGAGGCAGCGTCTGGTCTGATGCCTCGGAATGACCATTTACCATCGTGGACATTGTGGG  
 TGTAGATCTCAGGGCAAATCAGATACGAACTCTCACAGACATCGAATCGAGGGGAAAATACCG  
 CCGACCTCTGGTTTTTACACTACCGAGTTCATATGAGCGCCAATCATGGAGGAAGGACGTCCT  
 CAACATCTCGATCTAGAGCGGGTGGCTGCATAACTGTAGTTTCGGGAGTGTTTTTCTCCTCCGAT  
 GTAGACAACATAACCATCGTCTTGTTCGATTCTCTGGTAACAGCCAGCGTTGTCTTGTGTCAAC  
 CTTTGCACGTTTTTATCACTTGAATAAAGTGCGTAAATCATCAAAAAAAAAAAAAAAAAAAAAA

**Globine A2c:** numéro accession AM109950

AAGCTCTCTTTAAGGCAGGGACGCTTCTGACGTGAGGTTTTGAAGCCCTGAGGTTGAATCTAG  
 CAATCATCATGAAGGTCCTGATCGTACTGATGGCTGCTTGGCCTACGTCGCCCGACTGCG

GACCTCTGCAGAGGCTGAAGGTGAAGCATCAGTGGGTGCAGGTGTACAGCGGCCATGGTTACG  
 AGCGTGAGGCGTTCCGGCAGAGAGGTCTTCCCTCGAGATGTACAACCAGGCACCCAAGGCCAAGG  
 ACCTCTTACCAGGGTCAGGGGCGAGAACGTCTTCTCCCCGAGTTCGGAGCCACATGGTCC  
 GTGTGCTCGGAGGACTCGACATGTGCATCGCTCTGCTGTCCGATGACACCGTCTCAACGCC  
 AGCTTGCTCACCTCAGCACGCAGCACAAGGACCGTGAATCCCCAACGAGTACTTCGATGTGA  
 TGAAGGTGCCCCATGAAGGTGCTCCCCGGCCACGTTTACACTTCGACTTCGATGCCTGGT  
 CTGCCTGCTATGACGTCATCGCCAACGGCATCAAGCACTAAGCGCTGAACGCCACCTGGTGAC  
 AAAACGCCACACCGCTGACTGTCTGCCATAGCTTGAACCAAACCTCTCCACCGTCAACTTCTG  
 TCTCTTCTGTCTGCAATCCGCCAAAACCCATGGTGGTTTTTACCCATTCCCAGCCGGATCG  
 CACTGGCTAGCTGAACAATACCGCCATCACCCGCGGGAGAGTACGCTGAGGTGTCACGTGCT  
 CGGAGGGTGAGAACTTACCATCTTCTCCTCACTTTTGCGGGTGCAAAAAGAATAACTGACCACC  
 ATCGAAAGTTGTTTTGTTTTCTTCCGCAATGCTGCAACCTCGTTCACCTCAGGGCTTTGAAAG  
 CGTCGCAGAACAATGGCCGACCACAATTTGTTGATGCATCCCTAATTTGTCGCGTTATCCCGA  
 GAGTTTACATCATCTGAACAACATCTGAGATGACGTCATTCAAATGGTGTGCTTGCATTGGACT  
 GGGCCTGTACATCCTCCGTCAATTATTGCTATAACCATCTGGGGCTCGGGGTTACCACCCCTCA  
 GCCAGCCCCAACTGTCAACCTAGGCCGCCCTCCCTCGCACGCCTGACCCCTCACACCTGC  
 CCGCGGTCCCATCTTCTTGCCTGTACTATGTACATCTACCCTGTTCTGTAGAACAATAA  
 AACATGAACATTAAAAAAAAAAAAAAAAAAAAAA

**Globine B1:** numéro accession AM109951

CAGCTGATCGCAGCCGTCGAACAGCAGCGGTTTAGTTGCAATAGAACTAGCAACTCAACCCCT  
 GAGTCAAGATGATGTCGGTCTGTTCTCCTCGGCCTTGTTGGCTACGCCTCCGCCTCCAGCT  
 GCTGCTCCTATGGAGACCAGCAGAAGGTCAAGGCCAGTGAACAGCCTCTGGAACACCCCTG  
 ACTCCTCCACATCCAAGATCATCTTCGGAAAGGAAGTCTTCGCACGCTTCTTCGAGGTTGACC  
 CCGAGAGCAAGAGCCTGTTCCGTTCGCTCAAGGTTGAAGACCCCGACAGCCCGAGTTCGCCG  
 GACACGTGATCCGTGTTTTGACCGGTCTGGATTTGATCATCAACTTGATGGGTGACGATGCCA  
 TGGATGCCGAGCTGGCCACCTTAACACCAGCATTGTTGGCCAGAGAGGGAATCACCGGAACCC  
 ACTTCACCGAGATGTTCAAGGTCTTGATGGATCCCTCCGCCAGGTTCTCGAGGAGTACGATT  
 CCCTGTCTTGAGGTTACTGCTTCCGTGGTCTGGGCGCCGCTCAGGGATGGTCTCCCCGCAT  
 AAATACACCGCTGCATCCGCTGAAACTGCGCGCCACGGCTCGCCAGACGTCGCCACTCACGTC  
 CTCCACCGCGCGCACGGGCCGCTCCCCACGAGCGTGGCCACGTCGCGATTGGACGGGACA  
 AGCGTGACGTCACAGTCGTGGTGACTCCTTGTAGTTTAGTCATCTCGACAAAAAACTGTTTAA  
 AAAATCCTCGCCATGCTGGACTAAGACTTTGTATATAACAAGCAATGCCGTTGACAGAGACAT  
 CCGTGAATGAAGTGAAGATAAGACCGCCATCATCGAAGGGATATCGCTGTCTAGATATTGCT  
 CTTATTTCCCAAAGCCAAAGAAGTCTTACATTGCCACCAAGCATGCGTAATAAACAGATGAAA  
 CGGTAACATCATGTCAATGGTATTTTATGGTAAATTAATTGGCGAAGGCCTTCTAACAGCGC  
 CAGAGATGTTCAATTGGGCTCCCGTTTACGACTCCGAAAACGCACCTCCATCGGCTCCCGAA  
 GTGCGTCTTCGGCTTTGCGCAGTCGCGGTGTGCACTGCATGATGGCGACAGAAGCAGACGACA  
 ATTAGCGAGCGATCATCCACAGTTCTAAAGGCACCTCTATTATAACCATCATCAACTACCATCA  
 ACGCGTAACCATGTCTCTGTTATGTCGCTGTCTGTAGGCAGTTGTTGGCCTTTCTACGGCTT  
 CTAAGAAAAAGAAAAAAATATCTGTACCGGATAAATACTGGGGCTATAGTAAATGAAAAA  
 ACATGTAAAAA

**Globine B2:** numéro accession AJ880691

AGTCTGCTGACTCAGTAAACAGAACTGAGCAGGGAACCGGCTCACATCACACGTCAACATGCT  
 TCGTTTTCGTAGCACTCTTGGCTCTGGTGGCCTGGCCGTCTGTGACGACTGTTGTACCACCGA  
 GGACCGCAAGGAGGTCCAGACGCTGTGGAGTGAGATCTGGAGTGCCAGTTCCTGTTGCGCCG  
 TGTCCAGGTTGCCAGGCTGTGTTTCGAGGACCTCTTCCGCCGCGACCCCGAGTCCAAGAACC  
 GTTCAAGCGCGTCAATGTTGACGACATGAACAGCCCCGAATTCCACGCTCACTGCATCCGTGT  
 TGTCAACGGTCTTGACACCGTGATCGGTCTCCTTGACGACCCCGACACCTGAAGTCCAGCT  
 CGAGCACTTGGCCAGCAGCACAAGGAGCGTGATGGCATCCACAAGACCCACTTCGACGAGAT  
 GTCCACGCCTTCGGCGCCGTATGCCCCAGGTCAGCAGCTGCTTCAACCCCGATGCCTGGAA  
 CCGTTGCTTCGGCTCCATCGCTACCAAGATTGCTTCCCTCCTCGAGGATTAAGCCTTCGACTT  
 AGCAAACCGCGCGGTTCAAGACTGATCTGTCTCTGCCCTTCTCCTCTCGTTGACATCTAGTGGTCT

GGTGGCCCGTGTTACTGGCTGTGCCTTGTCTCTCGGTCTCCAAGGTCGAGTAACAGGGTCCTG  
 CCATATACGCGTCCCTAGATTTAGGGGTGTAATGTAGGGATAAAAAAAGTAGAAAATAGACTA  
 TAAGTTGAAGATGCTTTGTTTGTGTGTCGTGTTTCACAACGCGTGTATAGCGGCTATGCTTC  
 TAAAGCGCCGTGAGTAGCTTGAGACTCAGGAACAGTTTTACATCTGTGCGATTTAATCCATC  
 GTTGGAGTGATGACTAAAGTCGCTCTTTGAAAGAAGAGGAACAAAGTAATGAAAGCGTCTGTG  
 ATACATTCTTGTAACCTCGTCGCCCCATGCTGTTTCAAGTGTAGCGGTTGTATGAACGGAGATCT  
 TTGGTCGAGCTTTAACGTAGCTTCATTAGCCATGCGCGCAGCTATGACGTCACTATAAAACCG  
 GCTTCACAGTGACGTCAGCGTGGCATGTACGTCACCCGTATCACAAAGCTTCCGTCCGGGACC  
 TCATTGGACAGACAACCCGATCGTTTACGAAGGCGATGAGCGCTTGTATATGTGTGGCCGT  
 TCCCTGTTTCAAGTTTTAATAAAAAATATAAAAAA

**Linker L2:** numéro accession AM000028

AGGCAGGCTCTTCTGACAACAATGAAGAGCTACGTGCTCGTGTGCTGCCTCGTGGTGGGGGCC  
 GTGGCCTACCCCCACGAGGTGATGCACCATGCCGTTGGCGCAAACCGAATGTGCAAGTGTGAT  
 GCCCCGGCAGGGAACGCCGAAACCTCCGCCGACAGAGCAGAGTCACACTCTCGATGAGTTG  
 ACCCATCAGTTGCACATGCTGCAGCAAGCCTACGACACCCGGCATGGGTGCTGTGATGACGTG  
 ATGGAGGACATGGACGACCTGTCCACAGGATCGCCGACCACGAGAAGGAACACTGTAAGAAG  
 TATAGAGAGTTCCAGTGCGGTGGTGACCATCAAAGTGCATCTCGAACCTCCTCGTCTGCGAC  
 GGTGACAACGACTGTGACAATGGAGCTGATGAGGCTCGTTGTGATGTGCTCACCGAGGCTGGT  
 AGTAGTTGGACTGGTACTGTGGTCTACGATCACTGCACCAAGCGTCGCCCAGAGACCATGAAG  
 CTCAGCATCAAGAGCGTGGATACCGTACCCTTCTTCACCACCCACCCCAAGGTCCGCGGTACC  
 GTGCTTATGGAGAAGCACACCAAGGACTACAGCGAGGTCATCAACGAGCCGGTCTCTGGCTAC  
 TGGAGCAGCGCCGATAGGAGCGCCGCTATGCCCCCGACAGCGCCGGTCACCTTGGCTTTGTC  
 TGCATCTTCCACGGCCACGACCACGACACCTGCACTGGTCTCCTCACCAAGGGCAAGGTCACA  
 GATGCCTGCGCCGAGTTCACCTTCCACAGGGATTAAATATCACATGTGATACTACACTCTGCT  
 TGTTTCGCCAACACCTGAGAGTTGCTCAAGCCTTAGCCCCGGTGTACTGCCCCCGAAAGCC  
 ACTACGTGACGCGCGCAGCCGAAAACAGGAAGTACGTCATCTAGGTTTTCCAATGGGCCGC  
 TTCTGTACGAGTATAACTCGGAGCCGTACTCGTGGCAGAGTTCATTGAAACCAATCTCCTCC  
 AAAAGTGTGCTAGTGCCTTTAGTCGCTTTCCCTGGCAGTCCATCAGACCTTGATAAGCGT  
 CCTCGTGGTTGCCTATCATTGACGTTATACTTTGCGAAGTTGATAGCTCAGGAACTGCATTT  
 GAAAGGCGCTCAGATTCTCGCGTACTTTAGAACATAACCACAGCTATTGAAATTGATGAATGA  
 AAGTTGATGCTTGCAAAAGAGGTGCCAAATGTTTTTCATTGTAAATTATTAAGTGTGGATGAG  
 TCACCAACCAACGAGCTGTTGTTCTGTACCAGAATCCGCGTCTTGTTAAAGGCCACCAATGT  
 TTGATTTGAGGGATACCCTTATCATATTATACACCGATACGTTTTCAGTTTCCATGGAGAGGG  
 GGCGGGGTTAGGGACATCTACGTCTTTTGTGTTATTATGTACGCGACCATTCTCGTTTCGAGTT  
 ACAATTGTCAGCCCTTATATATGTAGCATAAAATATGACGGAGGTAATTAATAAACATGATGC  
 AAATAACGTTTAACTCTTGGGAGAACGCGACACGATGCC~ACACGCGCAGGCAGAGATGACC  
 CTATATCTCATGTTCTCGCTCGGGTAGAGAACTCACATGTCCTCTCTCGCAGATGGGGATA  
 GAGGGCGCACAGGCACCACAAAGTAGTTCCCTCCCTCCTCCCTCCCCACCGGTATACGGT  
 ACATATCACTGGTAGTGTATGTTATGTGTGTGCTTACGATCTTACGGCTCGGTGTCCCATTA  
 AAACGCACAGACTTCCGTCCACGGCATTGTCTTTGATATAATTACAAATAAAAAGAATCACAA  
 AAAAAA

**B- Séquences protéiques de globines et de linkers de *Arenicola marina*.**

**Globine A2a :** numéro accession AJ880690

(En rouge: peptide signal. En noir: le polypeptide mature)

MKSLVVLFA~~L~~VAM~~V~~AAECGPMQRLLVK~~T~~QWNKVYGT~~S~~KVRDEAGHVLWKAIFAQDPETRALFK  
 RVNGDDIYSPEFMAHSARVLGGLDIAISLLDNQADLDVALAHLHVQHVERHIPTRYFDL~~F~~KNA  
 LMEYAP~~S~~ALGR~~C~~FDKTAWSSCFDVIANGIKE



**Globine A2b:** numéro accession AM109949

MKFLVVLFALVAMVAA DCGPMQRLLVKAQWNKVYGT SKVRDDAGHVLWKAIFNQDGETRALFN  
RVHGDDIYSPEFMAHSARVLGGLDIAISLLDNQAELDAVLAHLKEQHIERGIPDRYFDLKFNA  
LMEFAPSALGRCFIKDAWSSCFDVIANGIKGQ

**Globine A2c:** numéro accession AM109950

MKVLIVLMACLAYVAA DCGPLQRLKVKHQWVQVYSGHGYEREA FGREVFLEMYNQAPKAKDLF  
TRVRGENVFSPEFGAHMVRVLGGLDMC IALLSDDTVLNAQLAHLSTQHKDRGIPNEYFDVMKV  
ALMKVVPGHVSHFDFDAWSACYDVIANGIKH

**Globine B1:** numéro accession AM109951

MMSVVFLGLVAYASA SSCCSYGDQQKVKQWNSLWNT PDSSTSKIIFGKEVFARFFEVD PES  
KSLFGRVKVEDPDSPEFAGHVIRVLTGLDLI INLMGDDAMDAELAHLNTQHLAREGITGTHFT  
EMFKVLDGSLRQVLEEYDSL SWRYCFRGLGAALRDGLPA

**Globine B2:** numéro accession AJ880691

MLRFVALLALVGLAVC DDCCTTEDRKEVQTLWSEIWSAQFTGRRVQVAQAVFEDLFRRD PESK  
NLFKRVNVDDMNSPEFHAHCIRVVNGLDTVIGLLDDPDTLKSQLEHLAQQHKERDGIHKTHFD  
EMSHAFGAVMPQVSSCFNPDAWNRCFGSIATKIASLLED

**Linker L2:** numéro accession AM000028

MKS YVLVCCLVVGAVAYPHEVMHHA VGANRMCKCDAPAGNAETSADREQSHTLDELTHQLHML  
QQAYDTGMGRVDDVMEDMDDL SHRIADHEKEKHCKKYREFQCGGDHPKCI SNLLVCDGDND CDN  
GADEARCDVLTEAGSSWTGTVVYDHCTKRRPETMKLSIKSVDTVPF FTTHPKVRGTVLMEKHT  
KDYSEVINEPVSGYWSSADRSAAMP PDSAGHLGFVCI FGHGDHDTCTGLLTKGKVTDACAEFT  
FHRD

**Annexe 5 : Séquences nucléotidiques et protéiques des globines intracellulaires non-circulantes et circulantes et des linkers d'annélides obtenues pendant la thèse.**

**Séquences nucléotidiques et protéiques partielles de la myoglobine MbIIa de *Arenicola marina* : numéro accession AJ880692**

GCGGACCAGATGGCAGCTGTAAAGGCTAACATCGCCGCCGTGAAGGCTGGTGACGTCACCAAG  
ACCGCCGGCGATTTCTTCGTCTTCCTCTTCAAGAAGTTCCCCGCCCTCCAGGACAAGTTCCCC  
AACTACAAGGGCAAGTCTGTGGACTCTCTGAGCTCTGTTGCCACCTTCGCCCCCACACCACC  
AAGGTCGTGGCCCGCGTCTTGACCTTGTGGCGAAGGCCGGAGACGCCGGGGTCTGGCTGGC  
GCCGCCAAGCAGGTCGTGCTGATCACGTGTCCCGCGGAGTCGTCAGCGGAGCCGAATACT  
GACCTGTTTCGCGGCCCTCGTCCCCTTCCTGGCCGCCGCCCTGGGTGGCGCCTGCGACCAGGCC  
GCCTGGACCAGCAACGGGA

ADQMAAVKANIAAVKAGDVTKTAGDFVFLFKKFPALQDKFPPNYKGSVDSLSSVATFAPHTT  
KVVAAVLDLVAKAGDAGVLAGAAKQVVDHVSRRGVVSGAEYTDLFAALVPPFLAAALGGACDQA  
AWTAATG

**Séquences nucléotidiques et protéiques partielles de la globine intracellulaire de *Alvinella pompejana* : numéro accession AJ880693**

GCGGACAACATAGCGGCGGTTAGGGGTGATGTCTCCACCCATGCCATGAACATCTTCGTGGAG  
TACTTCAAGAAGTTCCCACAGCACCAGAATGCTTTCGCTGACTACAAGGGCAAGGATCCCGAG  
AGCCTGAAGTCTCTGCCAAGTTCAGACGCACACCACCAAGGTCGTCAGCAAGCTGCTCGAC  
ATCGTCGAGAAGGCCAGCGACTCCGGAGCTCTCCAGAGCAACTGCACCACCCTGGCCAAGATG  
CCTCAGCACAAGGTCTCAACCAGCAACAGTTCGCTGATCTTGGCGCTGTCTTGGTTCCATAC  
CTGCAGAAGGCGCTCGGCGGCGCCTGCGACTCGGCTGCATGGGAACAAGCATAACAAC

ADNIAAVRGDVSTHAMNIFVEYFKKFPQHQNFAFYKGDPELKLKSLPKFKTHTTKVVSLLD  
IVEKASDSGALQSNCTTLAKMPQHKGLNQQQFADLGAVLVPYLQKALGGACDSAAWEQAYN

**Séquences nucléotidiques et protéiques partielles du linker LY de *Riftia pachyptila* :**

numéro accession AM000033

TCGAGCGGCCCGCCGGGCAGGTATCATGGCAGGCCTGGTGGCACTCGCCATTATCCTACTGGC  
GGCCGTCCAGGTCGAATCAGATGCGAGTAGAAAAGATGTAAACGACACGCGACTCAATGATTT  
GCAAGACAAGGTTGATGAACTTCAGGGCATCTGGAGGAAAGGGGAGCCATACGCGACCAGAG  
GCTGAAGGAGTCTCTGAACTCCAATCACGCGTCGCCAAACTCAAAGGCTCTCGTTGTGGTGT  
CAGGGAATTCCAGTGCACAAACAGCGCCATCTTCTGTATTACGACATACTCGTGTGTGACGG  
TGCCAAGGACTGTCCCGATGGTTCGGATGAAGAATTTTGCACCAATCCCGCCAAAGCCGGAAG  
CACTTTCACTGGTGTACGAACAACATTAAGTGCTCAACCGGATAACAAGGGGGATGTGCTGTG  
TTTTGACATCATTGGCGAGAAACGATAACAAGACTTCCCAGCCATTGTCTTCCCTAGAGGTTCA  
GACCAGCATCGACGGCCAAACGTTGCCATTTAAATTCGACGGTGTGTACCTCGGCCGCGACCA  
CGCT

MAGLVALAIILLAAVQVESDASRKDVNDTRLNDLQDKVDELQGILEERGAIRDQRLKEFSELO  
SRVAKLKGSRGCVREFQCTNSAIFCIHDILVCDGAKDCPDGSDEEFCTNPAKAGSTFTGVTNN  
IKCSTGYKGDVVCFDIIGEKRYNDFPAIVFLEVQTSIDGQTLFPKFDGVYLRDHA

**Séquences nucléotidiques et protéiques du linker L1 de *Alvinella pompejana* :** numéro accession AM000029.

(En rouge: peptide signal. En noir: le polypeptide mature)

ATGGCAGGCCTGGTGGCACTCGCCATTATCCTACTGGCGGCCGTCCAGGTCGAATCACACGAG  
CGCAGAGATGTCAATGACATGCGACTGAATGATCTTCAGGGAAAGATTGAGGAACTCCAACAG  
ACGCTGGACGAACGGGCCAAGACCCGGGATCAGCGACTGAGAGAGTTTCGAGAGCTAACAGCT  
CGAGTACACAACTGAAAGAGAGCCACTGTGGTCCACGTGAGTTTGAGTGTACTGAGAGCGCC  
AATCACTGTATCCATGACATTCTCGTCTGTGATGGTGCCAATGACTGTCCGGACGGATCCGAC  
GAGAAGAACTGCGGTAACCCCGCCACGCCGGTGCCACGTTCAAGGGCGTGATCCTGAACAGC  
CAGTGCCAGACCGAAAACGTGGCCAAGAACATGCAGATCGACATCGTTCGGCGAGAAGAGATAC  
TCCGACTTCCCGACGATCAGCGTCTCGAATTACTGGTCACTGGACGACCACCAGGACTTG  
TACAACGGTATCTACAGCTACGGCCGAAAGGCCCTGGTGTCTTTGGCAAGGGAGCGGGCGGT  
CTGGGCATGGTTTGTATTTTACACCGACGACGGCAAGTTCTGCAAAGCCGAGTTCCTAAGC  
ATCGTCTCCAAGGAAGTCTGTGGCACAGCCATATTGACCAGTGATTAA

MAGLVALAI ILLAAVQVESHERRDVNDMRLNDLQGKIEELQOTLDERAKTRDQRLREFAEFTA  
RVHKLKESHCGPREFECTESANHCIHDILVCDGANDCPDGSDEKNCGNPAHAGATFKGVILNS  
QCQTENVAKNMQIDIVGEKRYSDFP TISVLELLVTLDDHQDLYNGIYSYGRKALVSFGKGGGG  
LGMVCYFD TDDGKFKAEFLSIVSKEVCGTAILTSD

**Séquences nucléotidiques et protéiques partielles du linker L2 de *Alvinella pompejana* :** numéro accession AM000030.

ATGAGAATCCTTATCGTCTATGTCTAGCCGCCCTGGTACGGGCCGGAGCCCGCTTCGACGAG  
ATCGGAGAGGAGATCGCCGAGCTGACCGCCGAAGTGGTCGCTTTTCGAGAACATGATCGAGGAC  
AGAATCGAACAGGCCGAGGCTAGACACGCGCAGTGGCATGTCTTCTCCTCCCGCATTGACGCC  
GTCGAAGGATCCGGATGTGACGACGCTATCCAGATGCAGTGTGGTGGTGATACTCCAGACTGT  
GTATCCCGTCTGCTCATATGTGATGGAGAGAATGACTGTCTGAATGGAGCCGACGAATCACAA  
TGCCGTGTTTTACACCTGAGGGCAGCAAATGGAAAGCAGAATTCACATGGGACACATGCACA  
AAGCGCAAGCCAAAGGAAGTGTCCGGTACATTCACACACTACGAAGTGCCAGAGGAATTACCG  
TCCTTCCCTGAGGTAAAAGCTGAAGCATAATGAGCTCTGAAAACGATGATTTCTCATACTCT  
GCCAACCTTATTCTATCGGGTTACATGGACCTTAAGCCAGGCCAGAACAACTTTACCTTCGC  
CCGCCAGAGTCCGATGGCTTGGGTCTCCAGTGTATATTTGACGGCGTCAATGACGACCACTGC  
CCCGGT

MRILIVLCLAALVRAGARFDEIGEEIAELTAEVVAFENMIEDRIEQAEARHAQWHVFSRIDA  
VEGSGCDDAIQMOCGGDTPDCVSRLLICDGENDCLNGADESQCRVFTPEGSKWKAFTWDTCT  
KRKPKELSVTFTHYEVPEELPSFPEVKAEAYMSENDDFSYSANLILSGYMDLKPGQNKLYLR  
PPESDGLGLQCIFDGVNDDHCPG

**Séquences nucléotidiques et protéiques partielles du linker L3 de *Alvinella pompejana* :** numéro accession AM000031.

GACGTGATAAGGAGATGGAAGTGTGGAAAGAAATAATCAGACAGGCCGAACCCAGCGGCTGT  
AATGACCGATTGCAATTCCAGTGCGGCGGCCATCACCCGAAGTGCATCTCGCGTCTGCTCGTC  
TGCGACAGCTACGCGGACTGCGGAGAACGCCCGGACGAGAACGAACAATGCAGGGTCTACACG  
CCGGCCGGCAGCACCTGGGAGGGACGAGTTGAGTCCGACACGTGCACCAAGCGCCGACCAAAG  
GAGGTCAGGCTGACGATCGACAGCTACAAGATCTTCGACTACCTGCGCTCCTTCCACGCGTG  
GACGGCCCACTGGAGATCGATCCGCACTCGCCGGACTTCGACCCGACGAACAACGTCAAGTT  
GGAGGGATCCATGGATCTGATGCTTGGCGACAACATCCTGTACTTCGGTGTCTCCGGAGCAGGA  
CGGCCTTGCCCTCCAGTGCCTCTTCGACGGCATTGA

DVDKEMEVLEE IIRQAEPSPGCNDRFEFQCGGHPKCI SRLLVCD SYADCENAADENEQCRVYT  
PAGSTWEGRVESDTCTKRRPKVRLTIDSYKIFDYLRSPRVDGHTGDRSALAGLRPDEQRQV  
GGIHGSDAWRQHPVLRCSGAGRPCPPVRLRRH

**Annexe 6 : Séquence des introns des globines extracellulaires d'annélides.****Globine A2a *Arenicola marina* intron 1**

GTAAGTAACAACAACCTCAGCACCATGCGTGGGTCCAAGATCCAGATTATATATATTATTTTAA  
 ACTTTATAATATTATTTTAACTTTAGAAATATTATTTTACTTCAGAATATTATTTTAA  
 CTTCAGAATATTATGTTTAACTTCAGAATATTATTTGTAATTTTCAAGAATTTAACAGAATATT  
 ATTTTAACTTCAGAATATTATTTGTAACCTTCAGAATATTATTTGTAACCTTCAGAATATTATT  
 TGTAACCTTTAGAAATTTATGTTCAATTGTATATACAATTATATTAATTGTATTTACAATTATA  
 TTAATTTTAGAGAATTTCAATGGCATTTACACGTGTTTCGATCTGTTTCCTGTAG

**Globine A2a *Arenicola marina* intron 2**

GTAAGTATTAATTGTAATATTATTATTAATTGCTAATTGTATATACAGTTTTATATAGCTTTG  
 TAGTTTAGTCCTTCTTATTACAAAGGTACAATGGACAAGTTAAAGCTGCTCATACAGCAATT  
 GTCCTCCAACATTATGAGATTCCAAATCTTTTGGAGTATGTCCTGGACACAATATCGGACAGT  
 GTGAATGTTAATTTGGAATGGAGTGGAAATATTATGTGACAAATTTATAGCCACCAGTAATGC  
 TCGAATTATGATGACGTAGTACACAAGAAAGGCATACCTTCCCCACACAAGATGCAGCAGTTT  
 GTTTCAAGAAGTGGCTTAACAACATGCCCTGTTATTGATTATGTGCAATAAGTGTGTTCTTAT  
 ACCCTAAAACCTTAAATATATATTTTCAATTTACATTCGAGAGATTTAGAACATTATTTTCAGG  
 AAATAAAATATTAATGAGTAACTTTAAGTATATACATATATTCATAGATATGCTTACATGTCT  
 ACTGTAAACCGTCGGTAATCTCGTTATACAAGTAAAGTTATCCAAGCAAACCTGTCCGGTGC  
 GAAATAACAAAAGTGAATTTTAAATAGAGACGTCAAATACTTGAACTATATCGCTGAAAACAG  
 GAATTACGTACTAAAGGGCTGAAATGAATTTATTTTCCCGTCCGTTTTAGCGCGAGATGTTTC  
 TTGGCGCGACTTAGCCCTTGTTTGCACAAAACCTGCCTAAACCTCCGATTCATTCATTCCTGT  
 ATATTTTATCTCTATAG

**Globine B2 *Arenicola marina* intron 1**

GTAAGTCCAGATGGCTGTCTATATTTATCTCATTTCATGTGCGCCCAGTTTTATATGTATACGCG  
 CACAAAAGTGGCCGTAAGCGTTTCGTGTTTCTTAATCATGCAATGCATGATGAATAAGATATG  
 CCATTAAGGAGTTGGATAGGATATTTGTGAGTAGTATGCATATATATCTATATATATATGT  
 ATATATATATATTACATGCGTCATTACGTCAATTTGTGGTTTCGGAAATATCAATTATTAATTA  
 AGATTTAATAGGTTGTGTTTCTTGTGCTATAATCAACAGGCGAGTAATGGTGTGTTGGCGTA  
 ACGCGCATAAGTTTTACAAAACCGTTTACACATCTGGGCGCCAAAGAAATCTCCCTCAGGGAA  
 GATCTTTTTAACCTATATCTTCGCACGTTAATACTACTACTAACAACCGCTTCATCCGGAAT  
 ACTGTCTCATGTGTTCTTCTTCGTTACCCCTCCGCACTTTACAG

**Globine B2 *Arenicola marina* intron 2**

GTAAGTCTATTCAGCCCGTTAACTAAGCCGGAAGCGGAAGAAATGCCATAACACTAGACTAAC  
 TAAACACCGAGCTTCAAACGATAACCGATAATGATTTTGAAATCCATTTTGTACAGATCTT  
 TTAACCTTTCCTAAATAGGTATTGTAGGTATTGATGATAATGTCAACTGGTAATATGCGAAAT  
 ATAAAAATTAGGAGGCAAAGATATTTATTTCAAATAACAGGAAATGTGTATTTACAGAAGGAT  
 TTATAACGATAACCGCTAATGGTTTTGAAAGCTATTTTGTGCTAGATCCTTTAACCTCACCAA  
 AATAGGTATTGTAGGCAGCGATGATGATGTCACCTGGTTGTACGCCAAGCATAGGAATTAGTT  
 GGCACAGATATTTATTTAAATTTACCGGAAATGTGTATTTACAGAAGGCGCAATATAAATATAA  
 CTGTTACACCATTATGCCATGTGGTCTTCATTGTTTGCCCTCTGACCCCTGCAG

**Globine A1 *Riftia pachyptila* intron 1**

GTAAGTTGAATATACATCGTGTGTTGCACACGCACTATTGGTGGCAACATTTCTCCATTTCTGA  
 TCCTACCCTCTATCTCTCGGTCCGTGCTTCTTGAGAATATCTCAAACAGTTTCGTGTAACCTCT  
 TGGTTATTTGTGCTCTTAAGTACTGATTTTCAATTTGCGATTAGCAAGTCATTTCAATATATT  
 AAATATTTCCATCTTTCGTAACACCATCGCAACTGTCTCAATCACGAACAGTCATCATTATT



**Globine B1b *Riftia pachyptila* intron 1**

GTAAGTGCATTCTCCAATATTTTCGACATGACATGTGAACAGAAGAAGCGCTGAGCGTACTATT  
TATAGCCCGGTTTGGAGTATCGTGGTTGTCATTGTATGCATTTCGCTACACAGTGTAAAGTGGC  
AAGACTTTGCACATGAAGATCATTATACATGTTACCTTATTGAAACTATGAAAGCTGCACATT  
TTACGGACGATAATAAATCGTTTTCTTCAATGTATTTTTTACGGGCTTCTAGTCACGTTACTGC  
TGCCGTTGGACGTATTTTTATTGTAATGATCGCAATATGTTAGGCAGCACTCACTAAATACGA  
TGATTAGAGAGGTTACTATTAGATGGCGAATAAAACAGGACGGTGGGTGGCACTCTTGTGGCA  
ATCCCTGGAACACTTATACCACACACACACACACGCACACAAGCGCACGCACGCACGCGCGCA  
CGCACGCGCGCGCACAAATCAAATATGCACACACGCACGCATTCTATTACGCGTGCACGCG  
CGCACACAAACGCACGCATATATTTTTCTCCATTTGAATGCCCTGACATCGACTGTTTAACCG  
TCATTCCGTCTAACTCTGAGTATTCATTTACAG

**Globine B1b *Riftia pachyptila* intron 2**

GTAGAGTCACCCAAACATCAACGCAGTGTTTGGACATTTAATCGCATCTAATTGACTGATACG  
GAAATCGATACTGTACCAGCCAGATATGAATGTTGGACAAATTTAGAAATACGTAATTTTAAA  
TCGGTGTTCCTAAATAAATATTTGTTAACACCGAAATGCGCAATGTTATTTTTATGGTGTG  
ACACGATTAGTTTCCGCGGCCGTCGTTACAAATGAGCGTGCCACATCGTTATTTTGGCGCGC  
TCATTTGTAAGACGGCCGCGGAACTATCCGATATGTCTAAAAATTAATAATAACCGATAA  
ATAAGTTATTATGTTAATGATATTTGATGACGTTGTCCGTAATCTACACGCAATGGGGTTA  
GAAACAACAGGACAACCTATGTATATTACATCATTCTCTTTTATTACATCACTCAATTTATTA  
CACAACTCTATTCTTTACACAAAACACTGAGACATTCTTTGCTATAACACTCTCAGTTTATTAC  
GATTTTTTTGTTTTCTATCGCACGCATGAACACACAAACACGCACACGCCCGCGCGCGCGC  
GCGCGCACACACACACGCACTGTTTAATGATTGCAATACTTTCTGTTTATTTTCGACCTGGTCT  
GTTTATCTCTATTTATTCTTTGCAG

**Globine B1c *Riftia pachyptila* intron 1**

GTTTGTGCCATTCTCTCGTACAGACAAGAAACAAAATCGGAATTGATGTTAGCTAGCGCATT  
GAACGTGCGCGTGAAAGACGAATATGAGAATAGCCTTGAAGACGTCTAGCATGACATATCGGT  
TGAGGGGCGGATTCCGCGTTCGGTAAACACTTAGTAAACTCCGTTGTAGTTGGTTGGAAGCT  
GATGTGCATTCATGACTAAACGATAAAAATAACAAGTTGAAGGGAATGAATAAAAAATACATTT  
ATTTCACTGTCTGTTGTTTGCACAGCGTCTATCTGCATCACTACTTTTCATGTAGTATAAAG  
GTTTATATAAGGAATAATGCAGCGTTTTTAACGAATTTCCATTTGAATGGCATGTGCAAAATCA  
TATCTGCGATGGAAAACAAGATAATTTAACAAGGACGCTTTTAACTTGGCTTTAGCAGGAA  
GCCTGTGTAGTTGTGTTTGTGCTTAAATACCTTAGCTAATTGCGTCTAGTATAAACCATGTAG  
CTCCCCGGGAAAACATATGATGAATATTTTCGCTATTAGAATTTCTCCATTTTCAAAAA  
CGTACTTTTCGGCCGTTTTTTGAATTCACATGTTAATCACAGAATAGTGGTCAATGTCAAATA  
AACTTGTGCTTGACAGGCGCTTGGCACAATCTGCTATTGATTACGGAGCGTGGAAAGCACGGA  
AATAGGATTGTGCTTGATCATCACAAGATATTGTTAATTGGGCCCTCTCCGGGATTAATACCA  
ATTAGGATTGTTTTGAAACTGCATGTTTTGTTTATCATTACAATTGCTGCGACGTGCAACAC  
TAAACGATTCACCAACAAAACACTGGTTAGTTAAATTACAATTCGGCAATAACACGCTGTCAGTA  
AGTTACAGCAGATGCAACTTCCAGTGTAGGCATCATCTTATAATGAAAACGAAGATGTTGTG  
TCATTTTAATGTCGGCACGCACATGTACATAATGACAACACACTAGAGAGCGGTATCACTGAC  
AACATGTCCTTTTGTGTTGCAG

**Globine B1c *Riftia pachyptila* intron 2**

GTAAAGATCGTCTGTAACAGATTCGTCATGTCCTTAGTGTTACCGAAACTGGAATGGAACAGAA  
ATGCATCATTTAAATAGTTTTATGATGACTACTTTTTTCATTTAAATATAAACGCCTCCTAATA  
TGCATAAATCATTTTGCATAAGGACAAGCCGTTAACATGGTAAAACAGTTGGAAGAAACCCAG  
TACAATGATCGAAACCCTGCTACAGGAACCTCAGAGAGACAACAACAAGAATCGCTATTGTAT  
TGGTTATTGTAATATTTACTCAATTGTTATCGTCAAAAGATAGTCCGAGACGTCTCATTGTCT

CTATCGTGTGGATCCCATAGCCCAAAAATAGTCCTCGTTTTTAGATTTTCTTTTTTATTCTTG  
TTTCTTGCCCTGCAGTCTCCAGAAAATATCCCACCCCCTGTCTAACTCCTCCAACCCTCTCC  
GAACACCTTCATACCCTGGTCATGACACTCGACTCACGCTGATTGCAG



**Annexe 7 : Séquence des introns des linkers d'annélides.****Linker L2 *Arenicola marina* intron 1**

GTAAGTGGTGGGAAATAACGCCCCATAATGCCTCGCCCGCTTGTCTATCTGATGGGCTCGCTA  
 ATACGTCCGCGCGCAAAGCTCTAGTGCTCGTGTCTTTTCGTCAGCATTAGAGTAGCAGATTAT  
 ACAGATTTATATATTTTTCCCTTGGGTATTATGGCGCGGAAGGCTCATTATATAGATGTTAT  
 AGTCATTGTTTTGTACTACGATTGGATGGTTCTGGGAGCATTGAGAGTGCTATGAAAACGTTT  
 CTTTCAAGAATACTTTGAATGTTCTGCATGATTCCTCTTAGGGGTGGGGCAGCTTCTGCCGAC  
 TTGCCGCGAAAACATTGCCTGTGTGTAGCTCACGTAATTCTCACGCCAATCACTCAGTACAAT  
 TACAATCTATTTCTGTAACAGCTAACATCTGAACATGTTTCTTTGTTTTTCTAG

**Linker L2 *Arenicola marina* intron 2**

GTCAGTGCTTAGTCTTTATTTACTACATTTGTTATGAGTTCCTTTTTTACCGTTTCGTCATTT  
 CTACTTTTCTCTATATCTTTTCATTCCCTTTGTTTTAATCATTTAATATTTATTTCCCTTTAATC  
 ATTTGTTATCATTATTTTCATTTCTAATGTTTTTGTATTACCTGTTATTGCGAATCTGGAGTC  
 ATTTATTATGAATATTATTGCATGAATTCATATTTCGTACCATTACTAATATTTAACTCGTC  
 GTTGTATATATACGTGAGACATGTTTTGACCGTGTTCGGTCTACACTGTACCCCTCCCTCTC  
 TCTATTTGCTTTCTCTCCCTCTCTGGTCACAG

**Linker L2 *Arenicola marina* intron 3**

GTAAGACAAGGCAATGCCCGCCTAGATACATGTCACGTGTTTGTCTTATTGCCATCGGCCCC  
 TAAGGTAAAAACCACACCATGCGCCCTACATTTTCATTGGATGATTTTATTACGTTTTACTTAT  
 GTAGGCAGTTACCTTGCTCGAAGCACCTCTTTCTCTAATTAACAACATGCACTTAACTT  
 CACTATTCTAATTTTCTGCAAGTTTACCTCCTCTGCCCTCTCCTCTCCCTACTCTACTTAC  
 CCAGCTACCTTACCCTACCCTACCCTGTCAAACATACGTATCCTTTCTCCCTCCCCCCCCCTC  
 CCAATTCCCTCCATACTGTGCCCTCCTATCCCCTACCTCCCACCCACTCCCCCTCCTCTATC  
 AACCCTGTCCACCGACCCAAAGCCCACCGCCACCTCAACCACCCAACAACACCCCCCCCAC  
 CCCCCACCGCCACCGCTAACTCCACCTTCCCACACAACCCCTCCAACACACACGCCCGATAT  
 CGTTGCAACGGCTGCAATGAATAACACGCGACGCATGTTGTCTGTATCCGTCCACCAG

**Linker L2 *Arenicola marina* intron 4**

GTAGGTTGTGTTTGTGGCGTCCACCGATAGACACCCTAAGTATACCTGCCAGTCCGTCCAC  
 CGATAGACACCCCTAGTATATCTGCCAGTCCGTCCCCCGATAGACACCCCTAGTATACCTGCC  
 AGTCCGTCCACCGATAGACACCCCTAGTATATCTGCCAGTCCGTCCACCGATAGACACCCCTA  
 GTATATCTGCCAGTCCGTCCCCCGATAGACACCCCTAGTATATCTGCCAGTTCGTCCCC  
 CGATAGACACCCCTAGTATATCTGCCAGTTCGTCCCCCGATAGACACCCCTAGTATATCTGCC  
 AGTCCGTCAACCGATAGACACCCCTAGTATATCTGCCAGTCCGTCAACCAATAGACACCCCTA  
 GTATACCTGCCAGTCCGTCCACCGATAGACACCCCTAGTATATCTGCCAGTTCGTCCCCCGAT  
 AGACACCCCTAGTATATCTGCCAGTCCATCCACCGATAGAACCACCTGGTGTCTCTATCACAA  
 TAACGGTTCGCGTATCCCCCTTTCATAGGGGTAGGTGTTGGCCTGCCTGTAAGTCCACTAGGT  
 GGATGATCTAAGTCCCTTACCTCTAATCGCTCGCTCACTTTCCACAATGGTAACAGCAGTTCAC  
 GACCCAGTACCTCTACACTGTTTCTATAATAGCCATAGAAATTGGTGTATGTATCATTATG  
 ATATTTGCCTCTATCTTTTTTCTGTGGTATAAAATATGTATTTATGTATTTGTGTCTATTTCT  
 TGCAG

**Linker L1 *Lumbricus terrestris* intron 2**

GTGAGTAGAAGGACAAGAAGCATCTTCCTCTGTCTTTTCATTTATTTCCCTTCCCTCTCTAAATA  
 ATAATAAATAAATAAAAAATAACATAAAAAATCATCATCATGTACGAAATAGAGTATCCC  
 CCGTATTAATTGTATCCCCAAAATTTAAAAAAGTCGAAAAAACCGCTTCCGCGTGTATTTTTT  
 TAAATTAATTTTTATCTGCTTTGTTTTGAACAATGACACTAACAGGTAGCCTATGTTTATTTCC

TACTAGATAATTTGTTTTACCTGTTTTAGAGCACTTAAAACATTTAAAACACTTAAGAACA  
 CTCAGAGTCACTAAATAAATTCGAGTTGCCTTTATCTGAGCATTGCGTAGAACACTTAAAACA  
 ATTAGAGTAACCTTTAAATTTTGATTGCAGCATAACATTTTTTTATTATGCGCATTTTTCAA  
 AGGAGTCAGAATGGCTCATTCCCCCGGATAGTCTAGACCGGTTAATTCCATCCGGGGGAT  
 AGATTAATTCATCCCCCGGGTAGCTTTCAGGGGGGATACTCCGTTTCGTGTCGCAGGCTCTC  
 TTTGGTATATGTGTATGTGCATTATTATGTACGAACGAATGTGTGTATGTATGTATGTATGTA  
 TGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT  
 ACATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGTATGT  
 CTAAAACCTGCATATGTCTACATGGATGAATGACTGCATATCTAGTAATCTGGGTGCATGCATT  
 GTACAATGTACAATGCATGATACATTATCGCCAACGATATAGTACTATCGAGAGACTAGAGG  
 CATTTACCGTCGCTACATTATTATTTACCGACATAATAGTTATCGTTGTTTTTTCTTTACGCT  
 TCGCAG

### **Linker L2 *Neanthes diversicolor* intron 2**

GTGAGTGCATCTCATAAGTCATAAACTAAAGTTTATTATATTAGCTAATAATCTAATCACACT  
 CTTACAACGGATTGTTGTTTTAACTATGTATTGTTGAAAGTGACCAACCAATAATATAGATA  
 GAATTCATCCATTACATGTTTTTAAATTATCCAATACATTAGTTATTTCAACCAATAAAATGG  
 TTAATTGGCCAGATTCCAATCGTAAATCTGTTGAATAAACCAATCCGTTTGAAGGGCGCAGC  
 CTCCAGGAATGGATATATACAGCATAAGCCAGTGATTCATAATTGTGCATGTTCTATTGGTTC  
 TCTTGGTCTATTGGTCTCTTGGTGTTCAG

### **Linker L2 *Neanthes diversicolor* intron 3**

GTAGGTGTCCTGCTCATTTTTAACGTGCAAATGAGTAATTCACTTCAAATATCTAAACTATGGC  
 TTTTAGTCGTACATGGAAGCTGATAGTCAAGATTCACAAAACGTGATGAAAACCAACACAATG  
 AGCCTAAATATCTATGTATACCATTATCAGTTCAGCAAGTAATGTAATAATTTCTATCTATCT  
 AGAGTGACAGTCAGCTTATTGATGAAAACAGTAGAAACATGTCCAAAAAATCAATTTTACGGT  
 AAATTTGGTTTTGGAACCACTTCCAAAATTTACATTGAAACAATTATGTCTGTAAATAAATGG  
 TACTACTTAGCTGTACTTGTAAATTGACTGATCATGTGCTGACACTAGTTCTAACCGTCGTCT  
 TGCTTGTTCAG

### **Linker L2 *Neanthes diversicolor* intron 4**

GTAAGTGTTCATAAGTTTGATTCTTCAAATTTATAACCAATTGAAGAGTAGACTCGGCTCAA  
 GTTTGACAAGACAAAGTTTGACAAGAAGCCCTGAAGTTGATTAATCTTAAGCATAGATAGACT  
 CCTATCATTAGCACACACAATTCCTTAGACCATGCTTAGACGTAATGATGAGATGACATATCAT  
 AGTAAATAACGCGAGTTTTAAATAACCGATTGCAGATTGACTCTTAAAATTAATAAAACTTG  
 AAATCAATTAACGAAACCCTATTTAAATATTCTACAATGTATGGTGGCCTATCAAATGGTA  
 TCAC'TTTAGCAAATGAGTTTTCTCATCGAATTTCTTGATTACAG

### **Linker LX *Riftia pachytila* intron 3**

GTAAGCACTCTAGCCTTTTGCAACTGATGTGCCCTATTTAAAAAGTTAGTTGCCCATCTGT  
 TTTGCATAAGGTCAGCATAGCGAAGCTTGAATGCACTGTGTGCTCAAATCTATCCGCAAGAA  
 AGTTCCACCATAACCTACAGCATATTAGTGATTTTATGACATTTCAATATGTAATGACTGGCATTA  
 CGATTTAACGATAAATCATCAAGCGTAGAAGTGGCGCATACCTGTTTCACTGGATGATTTATT  
 AATCTATTTAGTGTCGGTTTACCGTAATGTCTGACAAAACAATCCAGCATCTGCCATAGGAAA  
 GATAAAACATATATAACATATAACAATGCGAACACATGATATGTGAGTAGCTTAAAACGCACA  
 AAATAACAGTATTTGGTATTGGGTTTCCTAAGGAAGCCAATCAATTAGATGCATGCCATTTAC  
 CCAGAGTAAGTCAGGTCATTCCACACACGATAAGAGAAGACAACAACAATAGAATGTCCAA  
 GTTAATAGTTTCTATTACTAGTAGATGTATATCAAAATCATGGATTACCTCTAATAAGTGTTA  
 AATCAATGTCACTGCTCGAAAAACCGCAAAATATGTTATTGTTTTTGGCTATGTTACGAGTAGT  
 ATTCATGAAGTTGTAGACTTGACTTTGATTCTGTCTTTTCAG

**Linker LX *Riftia pachyptila* intron 4**

GTGAGCGCAGACGACAATGTTCTCCCCCTTGTTGCAGAAGTCATGCACGTTCTCTATTGGTCC  
TCGGTGTTTTGCTTGCCCTTCTGCCAGTGTCCGTCTTGTTTTGTTGCCTTCCCCGCCATGTGAC  
CCACTCCCCATTCTTCTCTTCCCTTCCCACTCCATTCCCTTATTACTCCCATTATACTACAC  
CCCAAAGGCGCTCACTCGTTTTAGTAACCGTCCCTTCCGGTCACGCCGCGTGACTGCAGCAA  
ACAAACATGCTGCCCTTTAGCGAACTAGCCATGAGTCTGGACAAATGTATCCACATTTACACG  
ATTCCAGTAAAGCATTATGTAGACACTCAAAGGTCAAGGGTCATTACGATTACGTTTCATT  
TCGTTTTTCAAATTATAGATTATAAAAAGTTATCGTGGAATCGTTTTGAAAATAAAAATAAA  
ATTAATCGTGCTACTGTCTTGTCTTCCACAATTAACGCTTTGCATGTGTGGGCGCACGCCAA  
ATAGCGACCAACATTAATAACCACGTTATTTTTGCAATCGAGTATATATTCTTGCTCTTACCC  
CTTAATAATGCTATTTACGACAATTCCCGTCAGTACAGTTAACCGTATGTTAAGACTTGTGTC  
ATTACTAACCGCAGTTTTCTTTCCACAG

**Annexe 8: Séquences des introns des globines intracellulaires circulantes et non circulantes d'annélides.**

**Myoglobine MbIIa *Arenicola marina* intron 1**

GTAAGTTACTATAACAACCTGCCGCAACTATTGCTTAATGTCTTAATTAATGATAATATACAGA  
ATAAGGATGATGCAATTATTTATAGGTAAACCAGATCATACAATAACCTAATAATATATACAT  
GAGATTACATATCATGAGATAATGCCATGAGAATCAATATTAATAAACGTACGGCTTTTCCTT  
ATCCGTCAGATTACGATAGATAATTAATTTCCAGTGCCTGATTAAATCAATCAAGGATATTCG  
AAGCTTTTTAACCCCTCAATTAACGCCTGGTCTTTCTCTCTCCACAG

**Myoglobine MbIIa *Arenicola marina* intron 2**

GTAAGTATCCTATCGGAATTGACGGGTCACCGGAAATGATACCCTCATTGTATAAAGCCGGTG  
CCCTTGCGTTTTTCGGGATTTAGGATAAAAAATAAGAGAGGAATGAGGCGGGGTGGGGGGGGG  
GGGAGATGACTACGATTGCATATGCAGTTTCTGAGACATTGACATTTTCATCCAATTGTTTC  
GTTAAGAGAAATCGCGGTTTGTCCATTAATGTCAGTTTGGGAACACAATCCGTAAAATGGACC  
TAGATAAAATTGGTTTGCATGCGTGAGAGGGACCTAGTTGGTTTTCTGGTCAAAGGAACATA  
AAGGACTCGAGTCGCGTCTTCAAACCTTGTGGCCATGGAAATTAATTTTAAAATGAAACAAT  
CGATGAATGTGTAATTGTTGGAGACCGATACGGCACGACCATTATTCTATCTTGTATGAATAT  
TGCTTAGCTATGCAGTACAAATACACTGACATGTAATGTCGTTTTCGTTTTGTTCCGCCAG

**Hémoglobine intracellulaire de *Alvinella pompejana* intron 1**

GTAAGTACATTTTTTCCCACAAAGTTATCTACGAATAGGGTCGCCTTTTCATACGGTTAGCTT  
GTCCACAAATCGGCCGCCTTTTGCTGGAGAAGGCATTAGCGACTAACACCTAATCCGTGTCTT  
CCCACCCAG

**Hémoglobine intracellulaire de *Alvinella pompejana* intron 2**

GTAAGTGTACACGCATTTTATTGTTTTGTTTTGTTAGTACAGAAAATCTGAATCAATTATTCG  
ACCATTATTTATGAATTAACAACGTTTTCGATTATCAACCAGTTTGTCTGTTCGTTCCCTATCTT  
ATGTTAACC CGCTGTGTATCGTACCATTTCACAAGCATGCAGCCTGTCTGCTGCTTCGGAA  
TCTAACGACCTGACAGACATTATTACATGTGGCGTGCCACTTCGTCTTTCAG

**Hémoglobine intracellulaire mIV de *Glycera dibranchiata* intron 1**

GTAAGTATGATAAATTTTCAGCAAGCTGTACTATAGTGAGAATACATGATGATGGATAGCTGCA  
ATATTTCACTGAACCTATTGCAGCTAACCATCACGGTCATGTGTTTTTTTTTGCATTCATCCTA  
TGATGATCAATGTATTGAAAGGCAAACACAGATTTCAAATTTGGTGCTGAGAAGCGTGATTGG  
TTCAGTACCTTGTAACATTCTCACAGTATTTACCAAGCGAATATCACCTCCCAGGCCCTC  
TGCATAGGATATGTCCAAGTGACTTTCTCAGCTGGTAACAAGAACTTTCTATTATTGCAAATT  
CCACACAAGTCTTTGTAAAATGCAATGTGTTTAAAGGTGCACTCTTTGGTCAAATGGATACTAC  
TTCACATTGTCAGCACCATTGCAACTGAGATAGTCTGTTTCGCCTATTTCGACAAGTAACCAATC  
AGCCACATGCACTACATTCGGTTTTACCTTTTTGTAGCACCCACTTTTCTTCAAGGAATGCTC  
GCTTAATACATCTTTGAGTAATACAAAAAACCAAAAACGTATTCGCATTTCTCGCACGTTGTT  
CCTTCAAATAAACCCACTTCCAGGTTTTCTAGCATGTTTCTAGCCTTTTTCTAGATGACGATGA  
TGATGATGATGGTGATGATGATGATGGTGATGATGATGATGATGATGGTGATGATGATGATGA  
TGATGATGATGATGATGATGATGATGATGATGATAACCACCCATATCCGTCCATCAATCCATC  
CATCCATCTCCATTCAATTCATCGATCCATCGATCCACCCCTCCATTCATATATCCGACTCTC  
GAAGTTACGACTTTCCATTCTAGATTGTTTCGCATTATGTCAAATTGTCAGACATTCCCCTATG  
GGTATGCTCGAGTCGATTTGATGAAAACATAGGCATGAACTGCAGTATAGAGAATTCCCAATT  
CCTTTTGACTACAGGTCGATGTGTGTGTTTTTCAG

**Hémoglobine intracellulaire mIV de *Glycera dibranchiata* intron 2**

GTAGGTGGTCAACTGTCACTACTGCCATGTCTCTTGAATCCCATAGCACTTCCCTTCTCACC  
TCAATTCCCATCCATGTAGGTACTTTTCCAATCCTTGCTAGGGGACCGGGATCTTTGGCATGC  
CAAGGGAGCAGAATGGTCATATGTTCCACCCTGTCAATGATGACTAATGGGAGGTTCAAT  
ATAGCCCATGTGACGCTAAACAGGAAAAGTAACCCCTAACCACCAATCAATCAATCACCAATTT  
GCAGCTCACCTTTGCAAGTGGGGGACTGGTGAATATCGATATGGCGAGGTCCTAGTGCCTTC  
CCTGCAAACGCATTACGTCGCTTCGACGAAAATAGGGGTCAATGCTACAGCACTCGCCTTGCAC  
TATCGTGGGACTTCCCTTTGTTGTATAGTTTCATATAATGTCACCCAGTTATCCAGGTTTCAGAGA  
AGTATCCCAGTCAGTCATTTTAGTATACGATTGCCTACCCTGGCGGTGGAAACATGAATGTTT  
TCCGTCAATCCTTGTCTATTGCATAGACATCGCGAAACAGCACATCAGGACCTGACGAGAGGTC  
CGGAATTTCTTTCAAGGATCAATTCATCATCAACTCCATTCCATTTGAGGGACATTTAAGCT  
AGTCTCATTCAGCGAACACATGACGTCGTCACGACATCGCCACAACGTCAAGTGACCACGCTA  
TCATCATCGTCATCGTCATCGTCATCGTCGTCATCGCCGTCTTCGTCAATCGTCATCGTC  
ATCGTACCACCACCACCACCACCACAACCACCACCACCATCATGATTCAATCGTCAAATC  
ATGATCAATTTATCAATCAACATTTGTATGTTTTGCAAGTAGAGCGAAGACCTTGAAGTTTTG  
CATAACCATGTTTGAATTTGTCCATTTTCAG

**DHP A de *Amphitrite ornata* intron 1**

GTAAGCAAATAAAAACGCATGTTTCCAAGAAAAAAAATGTGTATTATTGCGTTCATAGGATC  
AGTGATATTTCAAATCGTCAAGTGACTCCTCGTTTCGTTACCTTGCAG

**DHP A de *Amphitrite ornata* intron 2**

GTGAATAGAAATAATTATGATCTCAAATACTTGTAAAAAAAATGTGATATCAACGGGCACCCC  
ATGAAATTTTGCATTTTGTAACTTAGCATAACGACGAGGTGTGTGAGATGATTTTAAAATATT  
GACCTGGGCTTCCGCCTATGTATGTTACTGCCTGTTGAGGCAAATAAATTTGAAACGAAAA  
GGAAATAATGATGCCAAGATTAAGCACAGATAACGTTTTTAAACGTTTGATAATGTGGGTCC  
TCTACAGTTCACCTAATAAACAACAAGCAATGACAAATAGAACGTCGTGCAC TTCAGGAGAAC  
ATTTGATAATAGATTGAATTTTAGACATCCTTCTGATTATATCAAAC TATATATCTCAACACT  
GGAAAGGACATAAATTATCTGAAGTGCAACTATGAATCCTATTTATTGGTCTAGTCGAAAGGA  
TCCTTAGATGATTAGTGGCAGAGTTATAATTAATAATGCGTTTTTGCCGGTTTTCTGTGACGTTG  
ACAACAGAAAATGAACCCCTCAGTTTTGGAGCAGACCAGTGGTGATGGTCTTCAAATTTGG  
TGATGACGCATTGAAAACGAGAAAGCCAGAGTAGTTTTCAAATAGGTAATAAAGAATAACTTT  
AAATAGTTGACATAGAAATCGTTTTATGTGCTACAATATCATTTTTTCATGTATCTGATTCATT  
GTCACCATGTTAACGTTATATGTCAATTAACACTATGTTTCGCATTATAATGAGTACTACCTCGAA  
ATTGTCATGGAAATACATTTTCTAAGTGGTTAGTACTAGTTACCATGGCTATATTTCTCTCTG  
TACACAAGAGCCAACACTCAACTATCACACAAAAGTAAAATATGGTTTCTGACTTCTGAAAAT  
TAACATAATTCTGATATTTATGGCGCATAACTTATGTGCACGCCCTAATATCAGGTTTATCCT  
ATTTAAATCCGACCAAAATGTATTTTACGTTGTAATGACCAATCAACTTTTTGTTTCTTGGGC  
TTCTATAACATCAATTTTATTGGGTCGACAATGCCGAAAATCGATTTATTTTCTTATAAGAAG  
AAGAAGATTTTGTAGCTATGCGAGTTCTATCCCCGTGCCAAAATGTGCGTAGAACTGGTTTAA  
AACTCACGCAGAGTTTTGTCATGAAATGGCTTTGTATTTGGAAATACCATAAAACTCGTGGGTC  
GAGTTCTATCGTAAATATACAGTATCGAAATAGTTATAATGGCAACACATGTCAATCCGGTTT  
CACTGAAATTTGATTTCCACAAATACAAATTAACACTCACGTAATATCACATAAAATAAAAATTAG  
AACTTTTGGGACTACTTTTGGAAATTGACTTTATTTATGACGTCATTCTGACGTTCTGTTGGGC  
AAAATTTGTTTGTGCAAGGGAGCTTTCTTCAAATCAGTCC TAGTCTGCGTAATTTGTATCG  
GATCTCATAACCGTTCAATTTGCTACGACATTTTGTGGCTAAATTGAACGCTTTGTCTATACG  
TATTTTGTTTTTCAGAAAACGTATTTTTTCAACCATTTGGCACAAATCAGACCACCATGTTAGT  
ATAAGACAAACAACCTCAGACAAGAGAAGCACACTCCAAAGTTTGGTTCGAGTTTGTATTTGGC  
TTCAATATAAATCTTGATGATTTCATGCGTGAAATTAGAGAGAATCGTAATTAATACTGCATGC  
GTTTAGATATAAGTCATATTCCTATCAGGACACCAGTCCCATCGCGTCCCCGGATCAACATGG  
ATCAACACGGTTGTGATTTTCCGGGGGGGGGGGGTGGGGGGCGTGTGAATCGGTATTGGACAAA  
TTGGTCTCTATCGAGCCCTGCTGATCCTGGATGCCAAATACGGCAAAGCTGATATGTCAATCT  
CTGCCATTTTGAATCAAGTTTCGTATGTGTGCTACCGTGAGGGGGCCGAGGGAATGCCGTGTT  
GGAACAGTGATCATCGGGGGTTACTTTATTAGCATGCGATACCCGGATGAACACGTCGTGAC  
ACGGTGGCAACGGTTGACCATCGAGCAAATAACATCCCCGGATCACCCCGGATGAACGCTGTTT

TGCTTCCGCATTGATTCATGGTTATCCGGGAATACCGGTCTACGGTCGTAGTATGGTGGGACT  
GGGACAAAATGAAAGTGATGTAGCCGAAAACATTGGAATGTCAGCGAACTGGTGTTCACATAA  
TCGGGACCATATCGAATATAAGATCCAACAAAGGCTATATGCATATTCAGATAAAACACTCCT  
TGCAATTTAAAAAAAAAACATTCATAAGATATTGTGCGGTGGTATTTAATGCCAACGGATTTG  
TAGGAAAGGGTTAACACAACCTATAAGCAGTGGTTATGACTGTAAGATTGAAATATGACGAATA  
ATGTTTGAAATATTGATACCTTCTCTTCAG

**Résumé :** *Caractérisation des familles multigéniques des globines et des linkers codant l'hémoglobine extracellulaire de Arenicola marina, dans le cadre de la mise au point d'un substitut sanguin humain.*

Les annélides possèdent plusieurs types de globines : des globines intracellulaires non circulantes (de type myoglobine), des globines intracellulaires circulantes et des globines extracellulaires. Les hémoglobines extracellulaires en bicouche hexagonale (HBL-Hb) sont composées de sous-unités de globines, qui ont la propriété de lier et transporter l'oxygène, et des chaînes de structure, appelées Linkers, impliquées entre autre dans le maintien de la structure quaternaire. L'HBL-Hb de l'annélide polychaete *Arenicola marina* présente des propriétés fonctionnelles qui font de cette protéine un substitut sanguin humain potentiel. Pour permettre la production de cette hémoglobine par génie génétique nous avons cherché à caractériser les différents gènes des familles multigéniques qui codent pour cette hémoglobine extracellulaire. Nous avons identifié cinq gènes de globines et un gène de Linker de *Arenicola marina*.

Grâce à la détermination et à l'analyse de la position des introns des gènes de globine de *Arenicola marina* et du séquençage de globines et positions d'introns d'autres annélides, nous avons pu proposer un scénario évolutif montrant les relations phylogénétiques existant entre les trois types de globines au sein des annélides. L'analyse des séquences de Linkers et la découverte d'introns supplémentaires nous ont permis d'apporter de nouveaux éléments permettant d'affiner l'évolution moléculaire des Linkers, protéines spécifiques des annélides. Enfin, nous avons pu entreprendre les premières expériences de surexpression d'un monomère de globine et de linker.

**Mots-clés:** *Arenicola marina*, Annélide, Globine, HBL-Hb, Linker, Myoglobine, Evolution, Globine intracellulaire, Globine extracellulaire

**Abstract:** *Characterisation of the globin and linker multigenic families encoding for Arenicola marina extracellular hemoglobin in order to develop a human blood substitute.*

Annelids exhibit three different types of globin : intracellular non-circulating globin (or myoglobin), intracellular circulating globin and extracellular globin. Extracellular hexagonal bilayer hemoglobin (HBL-Hb) are composed of globin subunits, which reversibly bind and transport oxygen, and structural chains called Linkers, required for the assembly of the HBL quaternary structure.

*Arenicola marina* (annelida, polychaeta) HBL-Hb exhibits functional properties which have led the laboratory to consider this protein as a potential human blood substitute. In order to produce this hemoglobin by *in vitro* synthesis, we have envisaged characterizing the genes belonging to the multigenic families and encoding for the extracellular Hb. We have formally identified five globin genes and one linker gene from *Arenicola marina*.

Sequencing and analyses regarding the intron positions of *Arenicola marina* globin gene, as well as others annelids have led us to propose an evolutionary scenario regarding the phylogenetic relationships between the three types of globin in annelids. Discovery of additional intron position in linkers' genes, in addition to analyses of the linkers' primary sequences allowed us to bring new elements in order to refine the molecular evolution of Linkers. We also managed to conduct the preliminary experiments of over-expression of one globin subunit and one linker subunit.

**Key-words:** *Arenicola marina*, Annelids, Globin, HBL-Hb, Linker, Myoglobin, Evolution, Intracellular globin, Extracellular globin.