

# SOMMAIRE

AVANT-PROPOS.....	14
CHAPITRE I : Les endosymbioses chimioautotrophes.....	16
<b>1. Introduction</b>	<b>16</b>
<b>2. Les milieux marins réduits contemporains</b>	<b>19</b>
<u>2.1 Les sources hydrothermales</u>	19
<u>2.2 Les suintements froids</u>	21
<u>2.3 Les zones de sédimentation</u>	21
<u>2.4 Les apports organiques imprévisibles</u>	22
<u>2.5 Points communs et divergences</u>	22
<b>3. La vie dans les milieux marins réduits</b>	<b>23</b>
<u>3.1 Producteurs primaires et chimiosynthèse</u>	23
<u>3.2 Consommateurs secondaires</u>	24
<u>3.3 Les symbioses</u>	24
<u>3.3.1 Les épisymbioses</u>	25
<u>3.3.2 Les endosymbioses</u>	26
<b>4. Les degrés d'intégration des endosymbioses chimioautotrophes</b>	<b>27</b>
<u>4.1 L'intégration cytologique</u>	27
<u>4.2 L'intégration physiologique</u>	27
<u>4.3 L'intégration génétique</u>	29
<b>5. Du symbiote à l'organite</b>	<b>30</b>
CHAPITRE II : L'organisme d'étude, <i>Riftia pachyptila</i> .....	33
<b>1. Place phylogénétique</b>	<b>33</b>
<u>1.1 Un Annélide Polychète Siboglinidae</u>	33
<u>1.2 Relations phylogénétiques au sein des Siboglinidae</u>	35
<u>1.3 Espèces symbiotiques au sein des Siboglinidae</u>	36
<b>2. Aspects du développement</b>	<b>37</b>
<u>2.1 Acquisition environnementale des symbiontes</u>	37
<u>2.2 Les étapes de l'acquisition des symbiontes</u>	37
<b>3. Anatomie de l'adulte symbiotique</b>	<b>39</b>
<u>3.1 Le panache branchial</u>	39
<u>3.2 Le vestimentum</u>	40
<u>3.3 Le tronc</u>	41
<u>3.4 L'opisthosome</u>	42
<u>3.5 Le tube</u>	42
<u>3.6 Communication entre les différents compartiments : le système circulatoire</u>	43
<b>4. Les voies métaboliques de la symbiose</b>	<b>43</b>
<u>4.1 Contribution de l'hôte : acquisition et transport des métabolites</u>	43
<u>4.1.1 Le dioxyde de carbone (CO<sub>2</sub>)</u>	44
<u>4.1.2 Le sulfure d'hydrogène (H<sub>2</sub>S) et l'oxygène (O<sub>2</sub>)</u>	46
<u>4.1.3 Composés azotés</u>	46

4.1.4 L'hôte : milieu de vie pour les bactéries	47
4.2 Contribution du symbionte	47
4.2.1 Sulfo-oxydation et assimilation du CO <sub>2</sub>	47
4.2.2 Transformation du NO <sub>3</sub> <sup>-</sup>	48
4.3 Contribution de l'hôte et du symbionte	49
4.3.1 Voie de synthèse de novo des pyrimidines	50
4.3.2 Voie de dégradation des pyrimidines	50
4.3.3 Métabolisme de l'arginine	50
4.4 Influence de la disponibilité des métabolites et limites du fonctionnement de l'association symbiotique	51
4.4.1 Autotrophie	51
4.4.2 Conditions hypoxiques	51
4.5 Questions en suspens	52
<b>6. Objectif du travail de thèse</b>	<b>53</b>

## CHAPITRE III : Matériel et Méthodes d'étude..... 57

<b>1. Echantillonnage des spécimens de <i>Riftia pachyptila</i></b>	<b>57</b>
<b>2. Extraction d'ARN total</b>	<b>57</b>
2.1 Précautions préliminaires	57
2.2 Méthode au RNABle	57
<b>3. Banques soustractives d'ADN complémentaire</b>	<b>59</b>
3.1 Purification d'ARN messager	59
3.2 Choix des échantillons sondeur et meneur	60
3.3 Préparation des ADNc de l'échantillon sondeur et de l'échantillon meneur	60
3.4 Hybridation Soustractive Suppressive (SSH)	63
3.5 Clonage	65
3.6 Séquençage	66
3.7 Analyse des séquences et recherche d'homologie	67
<b>4. Efficacité de la soustraction et vérification de l'expression différentielle : validation des banques par PCR classique</b>	<b>67</b>
4.1 Dessin d'amorces spécifiques	67
4.2 Amplification par PCR	68
<b>5. Vérification par PCR quantitative ou PCR en temps réel</b>	<b>68</b>
5.1 Description d'une amplification par PCR	68
5.2 Intérêt et principe de la méthode	69
5.3 Synthèse d'ADNc	70
5.4 Dessin d'amorces et paramètres de PCR	70
5.5 Méthode du 2 <sup>-ΔΔCt</sup> : conditions d'application	71
5.6 Méthode du 2 <sup>-ΔΔCt</sup> : détail de la méthode	73
<b>6. Autres techniques employées pour des cas plus spécifiques :</b>	<b>75</b>
6.1 Séquençage complet par Rapid Amplification of cDNA Ends (RACE) (Chapitre IV)	75
6.2 Alignements et analyses phylogénétiques (Chapitres V et VI)	76
6.3 Choix des sondes et hybridation sur membranes pour l'optimisation de l'Hybridation <i>In Situ</i> par Fluorescence (FISH) (Chapitre IV)	76
6.4 Hybridation <i>in situ</i> par Fluorescence (FISH) (Chapitre V)	79

6.4.1 Préparation des coupes	79
6.4.2 Hybridation des sondes	80
6.4.3 Détection	81
<u>6.5 Clonage d'un transcrit dans un vecteur d'expression (Chapitre VI)</u>	81
<u>6.6 Test d'activité superoxyde dismutase (SOD) (Chapitre V)</u>	83
<u>6.7 Western-blot pour détecter la Major Vault Protein (MVP) (Chapitre VII)</u>	84

## CHAPITRE IV : Premier article.....85

Application de la technique d'Hybridation Soustractive Suppressive à l'étude transcriptomique de *Riftia pachyptila*

*Identification of proteins involved in the functioning of Riftia pachyptila symbiosis by Subtractive Suppression Hybridization*

<b>1. Résumé</b>	<b>85</b>
<u>1.1 Contexte de l'étude</u>	85
<u>1.2 Principaux résultats</u>	86
<b>2. Article publié dans BMC Genomics</b>	<b>86</b>
Sanchez S., Hourdez S. and Lallier F. H., 2007. Identification of proteins involved in the functioning of <i>Riftia pachyptila</i> symbiosis by Subtractive Suppression Hybridization. <i>BMC Genomics</i> , <b>8</b> :337.	

## CHAPITRE V : Deuxième article..... 116

Identification, expression et localisation d'un nouveau transcrit d'anhydrase carbonique chez *Riftia pachyptila*

*Identification, sequencing, and localization of a new carbonic anhydrase transcript from the hydrothermal vent tubeworm Riftia pachyptila*

<b>1. Résumé</b>	<b>116</b>
<u>1.1 Contexte de l'étude</u>	116
<u>1.2 Principaux résultats</u>	116
<b>2. Article publié dans FEBS Journal</b>	<b>117</b>
Sanchez S., Andersen A. C., Hourdez S. and Lallier F. H., 2007. Identification, sequencing, and localization of a new carbonic anhydrase transcript from the hydrothermal vent tubeworm <i>Riftia pachyptila</i> . <i>FEBS Journal</i> , <b>274</b> (20):5311-5324.	

## CHAPITRE VI : Troisième article..... 146

Caractérisation, expression et étude fonctionnelle d'un transcrit de myohémérythrine de *Riftia pachyptila*

*Characterization, expression, and functional study of the myohemerythrin from the hydrothermal vent tubeworm Riftia pachyptila (Polychaeta, Siboglinidae)*

<b>1. Résumé</b>	<b>146</b>
<u>1.1 Contexte de l'étude</u>	146
<u>1.2 Principaux résultats</u>	147
<b>2. Article en préparation</b>	<b>147</b>
 CHAPITRE VII : Discussion, Conclusion et Perspectives.....	 169
<b>1. L'approche méthodologique générale</b>	<b>169</b>
<u>1.1 Efficacité de la méthode SSH dans la comparaison de tissus différents</u>	169
<u>1.2 Efficacité de la méthode SSH dans notre problématique</u>	171
1.2.1 Comparaison avec des banques d'EST classiques	171
1.2.2 Les études dites « globales »	172
1.2.3 Etude sur Anthopleura elegantissima	173
<u>1.3 Problème général lié à l'identification des séquences</u>	173
<b>2. Les anhydrases carboniques de Riftia pachyptila</b>	<b>175</b>
<u>2.1 Transcrits du panache branchial</u>	175
<u>2.2 Transcrits du trophosome</u>	176
<u>2.3 Evolution de la famille multigénique des anhydrases carboniques</u>	176
<b>3. Les protéines respiratoires</b>	<b>178</b>
<u>3.1 Transcrit de myoglobine</u>	179
<u>3.2 Transcrit de myohémérythrine</u>	180
<b>4. La Major Vault Protein (MVP)</b>	<b>180</b>
<u>4.1 Description</u>	180
<u>4.2 Indices de présence chez R. pachyptila</u>	181
<u>4.3 Possibles fonctions de la MVP ?</u>	182
<b>5. Autres pistes de protéines impliquées dans des fonctions spécifiques</b>	<b>184</b>
<u>5.1 Le Précurseur Chitinase</u>	184
<u>5.2 Autres séquences</u>	184
<b>6. Perspectives d'études</b>	<b>185</b>
 REFERENCES BIBLIOGRAPHIQUES.....	 188
 LISTE DES ABREVIATIONS.....	 11
 LISTE DES FIGURES ET TABLEAUX.....	 199
 RESUME ET ABSTRACT.....	 203
 AUTRES PUBLICATIONS DE L'AUTEUR.....	 204



## ABREVIATIONS

- 3' UTR** : 3' UnTranslated Region, région non traduite en 3'
- 5' UTR** : 5' UnTranslated Region, région non traduite en 5'
- ADN** : Acide DésoxyriboNucléique
- ADNc** : ADN complémentaire
- ADP** : Adénosine DiPhosphate
- AMP** : Adénosine MonoPhosphate
- AMV RT** : Avian Myeloblastosis Virus Reverse Transcriptase
- APS réductase** : Adénosine-5'-PhosphoSulfate réductase
- ARN** : Acide RiboNucléique
- ARN 16S** : petite sous-unité de l'ARN ribosomique (bactérien ou mitochondrial)
- ARN 18S** : petite sous-unité de l'ARN ribosomique (eucaryotes)
- ARN 28S** : grande sous-unité de l'ARN ribosomique (eucaryotes)
- ARNm** : ARN messenger
- ATP** : Adénosine TriPhosphate
- ATPF1** : complexe F1 de l'ATP synthase
- BIF** : Banded Iron Formation, gisement de fer rubané
- BR** : branchial plume, branchie
- BR-BW** : banque d'ADNc soustractive de BR (par rapport à BW)
- BRbw** : contig issu de la banque BR-BW
- BW** : body wall, paroi du corps
- BW-BR** : banque d'ADNc soustractive de BW (par rapport à BR)
- BWbr** : contig issu de la banque BW-BR
- BW-TR** : banque d'ADNc soustractive de BW (par rapport à TR)
- BWtr** : contig issu de la banque BW-TR
- CA** : anhydrase carbonique
- ccoxI** : sous-unité I de la cytochrome c oxydase étudiée chez *Riftia pachyptila* (notre étude)
- COI** : sous-unité I de la Cytochrome Oxydase
- CPSase** : carbamylphosphate synthétase
- Ct** : threshold cycle, cycle de PCR auquel l'amplification atteint un seuil donné
- CTP synthase** : Cytidine TriPhosphate
- DD** : differential display, affichage différentiel

**ddNTP** : DiDésoxyNucléotide TriPhosphate

**DEPC** : DiEthylPyroCarbonate

**DIG** : digoxygénine

**dNTP** : DésoxyNucléotide TriPhosphate

**EDTA** : acide Ethylène-Diamine-TétraAcétique

**EF-1 $\alpha$**  : Elongation Factor 1 $\alpha$ , facteur d'élongation 1 $\alpha$

**EPR** : East Pacific Rise, Dorsale de l'océan Pacifique oriental

**EST** : Expressed Sequences Tag, étiquettes de séquences transcrites

**GDHase** : Glutamate DÉHydrogénase

**GSase** : Glutamine Synthétase

**GST** : Glutathion S-Transférase

**Hb** : hémoglobine

**HBL-Hb** : hexagonal bilayer hemoglobin

**Hr** : hémérythrine

**MHr** : myohémérythrine

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

**MVP** : Major Vault Protein

**NADP** : Nicotinamide Adénine Dinucléotide Phosphate

**NADPH** : forme réduite du NADP

**NBT / BCIP** : nitro blue tetrazolium / 5-bromo-4-chloro-indolyl phosphate

**NCBI** : National Center for Biotechnology Information

**NH<sub>4</sub>OAc** : acétate d'ammonium

**PBS** : Phosphate Buffered Saline

**PCR** : Polymerase Chain Reaction, réaction de polymérisation en chaîne

**Ppi** : Pyrophosphate inorganique

**PrCh** : Précurseur Chitinase étudié chez *Riftia pachyptila* (notre étude)

**ROS** : Reactive Oxygen Species, espèces réactives de l'oxygène

**RpCAbr** : transcrit de CA spécifique de la BR chez *Riftia pachyptila*

**RpCAtr** : transcrit de CA spécifique du TR chez *Riftia pachyptila*

**RpMHr** : transcrit de MHr de *Riftia pachyptila*

**SBSA** : Specific Branchial Surface Area, surface branchiale spécifique

**SDS-PAGE** : Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis

**SSH** : Subtractive Suppression Hybridization, hybridation soustractive suppressive

**TAE** : Tris-Acetate-EDTA

**TCA** : tricarboxylic acid

**TCL** : T-cell receptor

**TR** : trophosome

**TR-BW** : banque d'ADNc soustractive de TR (par rapport à BW)

**TRbw** : contig issu de la banque TR-BW

**TSA** : Tyramide Signal Amplification

**Espèces et éléments chimiques :**

**CH<sub>4</sub>** : méthane

**CO<sub>2</sub>** : dioxyde de carbone

**HCO<sub>3</sub><sup>-</sup>** : ion bicarbonate

**H<sub>2</sub>S** : sulfure d'hydrogène

**NH<sub>3</sub>** : ammoniac

**NH<sub>4</sub><sup>+</sup>** : ammonium

**NO<sub>2</sub><sup>-</sup>** : ion nitrite

**NO<sub>3</sub><sup>-</sup>** : ion nitrate

**O<sub>2</sub>** : oxygène

**SO<sub>3</sub><sup>-</sup>** : ion sulfite

**SO<sub>4</sub><sup>2-</sup>** : ion sulphate

# AVANT-PROPOS

## AVANT-PROPOS

Cette thèse a été réalisée au sein du laboratoire ECOPHYSIOLOGIE : ADAPTATION ET EVOLUTION MOLECULAIRES dirigé par le Dr Franck Zal, appartenant à l'Unité Mixte de Recherche 7144 intitulée ADAPTATION ET DIVERSITE EN MILIEU MARIN dirigée par le Pr François Lallier. L'étude des organismes vivant en milieux réduits est le thème fédérateur de l'équipe et en particulier, l'étude des organismes hydrothermaux, tant d'un point de vue biochimique que moléculaire est l'une des préoccupations majeures de l'équipe avec un intérêt marqué pour l'étude des crustacés, des mollusques et des annélides. Le ver Siboglinidae *Riftia pachyptila* (anciennement vestimentifère) récemment admis dans le phylum des annélides et précisément parmi les polychètes, a ainsi été au centre de recherches concernant l'anhydrase carbonique et ses transporteurs impliqués (thèse de M.-C. De Cian, 2002) et l'évolution moléculaire des chaînes de l'hémoglobine extracellulaire géante caractéristique des annélides (thèse de X. Bailly, 2003).

Mon travail de thèse sur *Riftia pachyptila* a consisté en une approche plus globale, non focalisée *a priori* sur un point particulier, mais orientée sur la relation symbiotique. L'approche méthodologique envisagée, originale, et l'utilisation d'outils moléculaires performants ont rendu ce travail très intéressant. Celui-ci a par ailleurs permis d'obtenir plus de 150 nouvelles séquences de transcrits de l'hôte *R. pachyptila* qui constitueront une base de travail pour des études ultérieures.

Nous introduirons tout d'abord quelques généralités au cours du CHAPITRE I, pour replacer cette étude dans le contexte plus large de la vie dans les milieux réduits existant sur la planète. Nous décrirons en détail le fonctionnement des écosystèmes chimiosynthétiques dont les endosymbioses sont une composante majeure.

Le CHAPITRE II sera entièrement dédié à la description complète de *Riftia pachyptila* au niveau anatomique et physiologique et les interactions avec ses symbiontes bactériens seront abordées en détail. Nous introduirons la problématique qui constitue le sujet de cette thèse et la façon par laquelle nous nous proposons d'y répondre.

Dans le CHAPITRE III, nous présenterons l'approche expérimentale choisie pour le démarrage de cette thèse (la technique moléculaire d'Hybridation Soustractive Suppressive)

ainsi que les autres techniques employées pour des cas plus spécifiques (PCR quantitative, séquençage complet, Hybridation *in situ* par Fluorescence...).

Les résultats de l'identification des séquences obtenues à partir des quatre banques soustractives seront présentés dans le CHAPITRE IV. L'expression différentielle observée pour quelques transcrits sera commentée. A partir de ces travaux préliminaires, des études plus spécifiques seront développées dans les deux chapitres suivants.

Ainsi, dans le CHAPITRE V, un intérêt particulier sera porté sur l'étude de transcrits codant des isoformes d'anhydrase carbonique. Cette étude a consisté en la mise en œuvre de plusieurs méthodes complémentaires pour vérifier l'expression différentielle des transcrits dans différents tissus et a été complétée par le séquençage complet du nouveau transcrit identifié.

Puis, dans le CHAPITRE VI, l'accent sera mis sur un transcrit codant un pigment respiratoire particulier, la myohémérythrine, dont la séquence codante entière obtenue dans l'une de nos banques a pu être surexprimée en système hétérologue. Cette étude a pu être prolongée par des études fonctionnelles préliminaires.

Enfin, l'ensemble de ces résultats sera repris dans un dernier chapitre, le CHAPITRE VII, au cours duquel nous formulerons quelques hypothèses et proposerons quelques perspectives d'études dans le prolongement de ce travail.

# **CHAPITRE I**

## **Les endosymbioses chimioautotrophes**

CHAPITRE I : Les endosymbioses chimioautotrophes <sup>1</sup>
---

## 1. Introduction

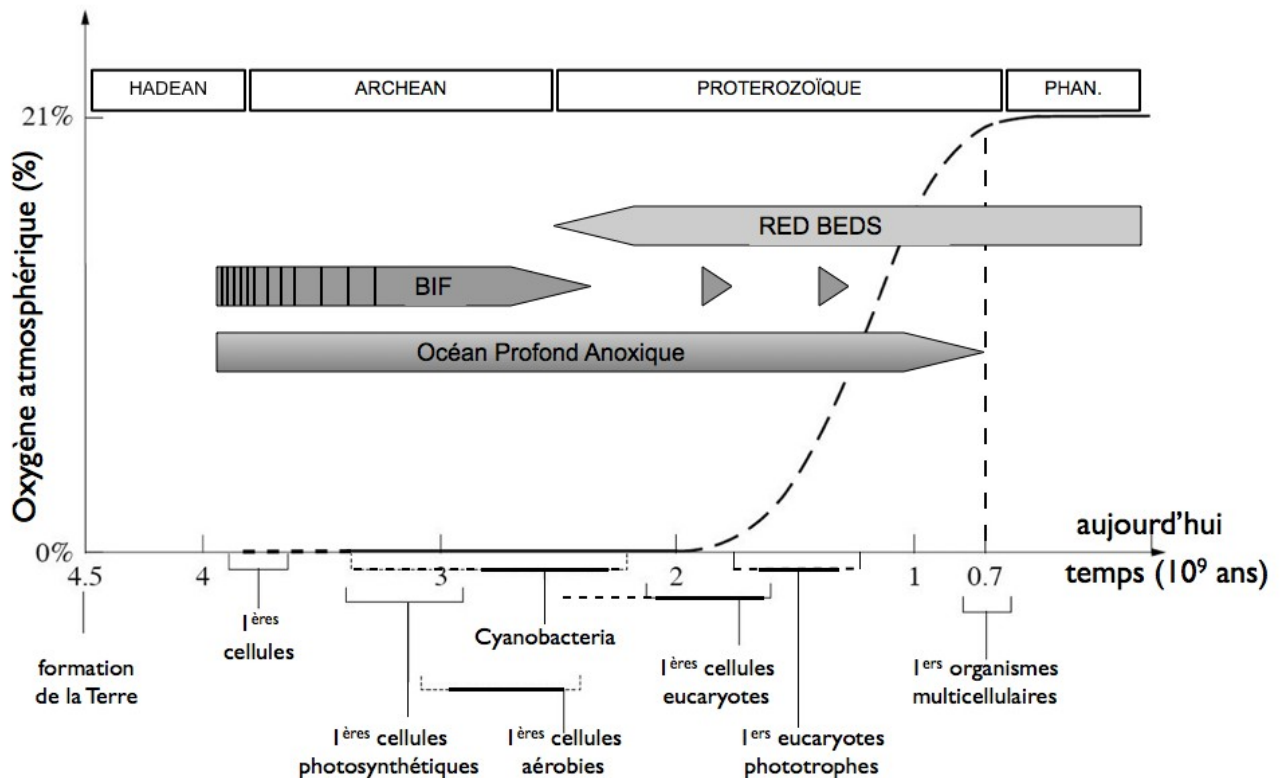
Aujourd'hui, tous les Eucaryotes sont fondamentalement des organismes aérobies, utilisant l'oxygène comme accepteur final de leur chaîne de transport des électrons. Pourtant, à l'échelle des temps géologiques, l'oxygène n'est apparu qu'assez récemment dans l'histoire de la biosphère de notre planète (Figure 1) (Canfield, 2005 ; Catling et Claire, 2005). Les premières étapes de l'évolution de la vie se sont très certainement déroulées dans un environnement anoxique et fortement réducteur et les Procaryotes ont alors développé une grande diversité de métabolismes pour exploiter l'énergie chimique, disponible sous forme de couples oxydo-réducteurs variés. Jusqu'à ce que certains procaryotes, les cyanobactéries, parviennent à utiliser très efficacement l'énergie lumineuse en couplant la réduction du dioxyde de carbone à l'oxydation de l'eau, il y a environ 2,7 milliards d'années. Une conséquence majeure de ce nouveau métabolisme est la production d'oxygène qui, à cette époque, était immédiatement absorbé par l'abondance d'éléments réduits (cela a laissé des traces dans les sédiments anciens sous la forme de "banded iron formation" ou BIF, voir la Figure 1). Ce n'est donc qu'avec un délai certain (près de 500 millions d'années) et très progressivement que l'oxygène a pu s'accumuler dans l'atmosphère et dans l'hydrosphère, et fort heureusement car il s'agit une molécule toxique pour les organismes qui ne disposent pas des mécanismes *ad hoc* pour se protéger de ses effets délétères (générés par les espèces réactives de l'oxygène, ROS). Certains Procaryotes, des  $\alpha$ -protéobactéries, ont pu alors développer un nouveau métabolisme permettant de coupler la réduction de l'oxygène, nécessaire pour lutter contre sa toxicité, avec la production d'énergie sous forme d'ATP : la respiration aérobie.

---

<sup>1</sup> Ce chapitre a été publié sous forme de revue dans le Journal de la Société de Biologie pour sa séance du 22 juin 2007.

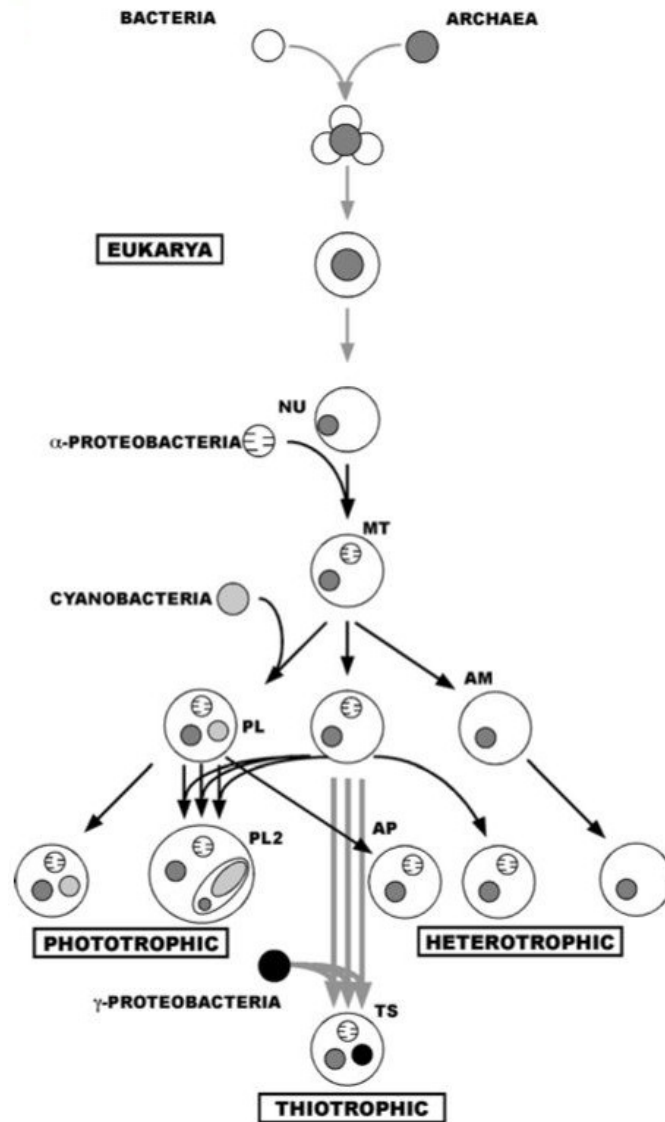
**Sanchez S.**, Hourdez S. et Lallier F. H., 2007. Les endosymbioses chimioautotrophes : des modèles contemporains de la symbiogenèse ? *Journal de la Société de Biologie*, **201**(3) : 247-257.





**Figure 1 :** Relation entre l'évolution de la teneur en oxygène dans l'atmosphère et un scénario probable de l'évolution cellulaire (adapté de Vellai et Vida, 1999). Les pointillés horizontaux indiquent les incertitudes sur la datation de ces événements anciens. La ligne verticale pointillée souligne la concordance entre l'apparition des Eucaryotes multicellulaires et un niveau d'oxygénation atmosphérique proche du niveau actuel. Les Cyanobactéries sont les premières cellules procaryotes photosynthétiques oxygéniques : avant leur apparition l'oxygène n'était produit qu'en quantité négligeable de façon abiotique par photodissociation de l'eau. BIF: "banded iron formations" ou gisements de fer rubané, témoins de la production d'oxygène sans accumulation notable dans l'atmosphère ; Red Beds : couches rouges, formations géologiques indicatrices de la présence d'oxygène atmosphérique. PHAN: Phanérozoïque.

L'origine des premières cellules eucaryotes, plus grandes et plus complexes que les Procaryotes, est encore débattue de nos jours, mais un grand nombre de spécialistes s'accordent à penser que la symbiose a pu jouer un grand rôle dans cet événement majeur de l'évolution de la vie (Figure 2) (Embley et Martin, 2006 ; Poole et Penny, 2007). Une première fois avec l'une de ces  $\alpha$ -protéobactéries capables d'utiliser efficacement l'oxygène et qui deviendront les mitochondries, ces organites caractéristiques de tous les Eucaryotes contemporains (les quelques exceptions connues correspondent à des pertes secondaires, Figure 2). Et une seconde fois avec une cyanobactérie, dotant certains Eucaryotes, les phototrophes, de la capacité d'effectuer simultanément photosynthèse oxygénique et respiration aérobie. Mais ces événements sont si anciens que les traces qu'ils ont laissées ouvrent trop souvent la voie à des interprétations divergentes quant à leur déroulement précis (Embley et Martin, 2006).



**Figure 2 :** Scénario évolutif hypothétique de l'origine et de la diversification des Eucaryotes mettant en exergue le rôle de l'endosymbiose (d'après Lallier, 2006). En haut : les premiers Eucaryotes sont peut-être issus d'une association entre Bacteria et Archaea, celle-ci étant à l'origine du noyau (NU). Au centre : les deux événements d'endosymbiose primaire qui ont donné naissance aux mitochondries (MT) et aux chloroplastes (PL) à partir d' $\alpha$ -protéobactéries et de cyanobactéries respectivement. Les événements ultérieurs comprennent des endosymbioses secondaires (PL2) et la perte des symbiotes primaires (AM, AP) pour aboutir aux différentes lignées d'Eucaryotes phototrophes et hétérotrophes. En bas : addition hypothétique d'un nouvel événement d'endosymbiose primaire avec une  $\gamma$ -protéobactérie sulfo-oxydante (TS) donnant naissance à la lignée des Eucaryotes thiotrophes.

A notre époque, dans notre monde désormais aérobie, il existe pourtant encore de nombreuses poches de milieux réduits qui ont pu au cours des âges préserver une grande diversité de Procaryotes. Parmi ceux-ci certains sont capables de réaliser la chimiosynthèse en récupérant l'énergie chimique pour réduire le dioxyde de carbone ( $\text{CO}_2$ ) en carbone organique : ce sont des chimioautotrophes. Dans certains milieux réduits particuliers, là où les autres sources de nourriture sont rares, des Eucaryotes ont su mettre à profit ces bactéries

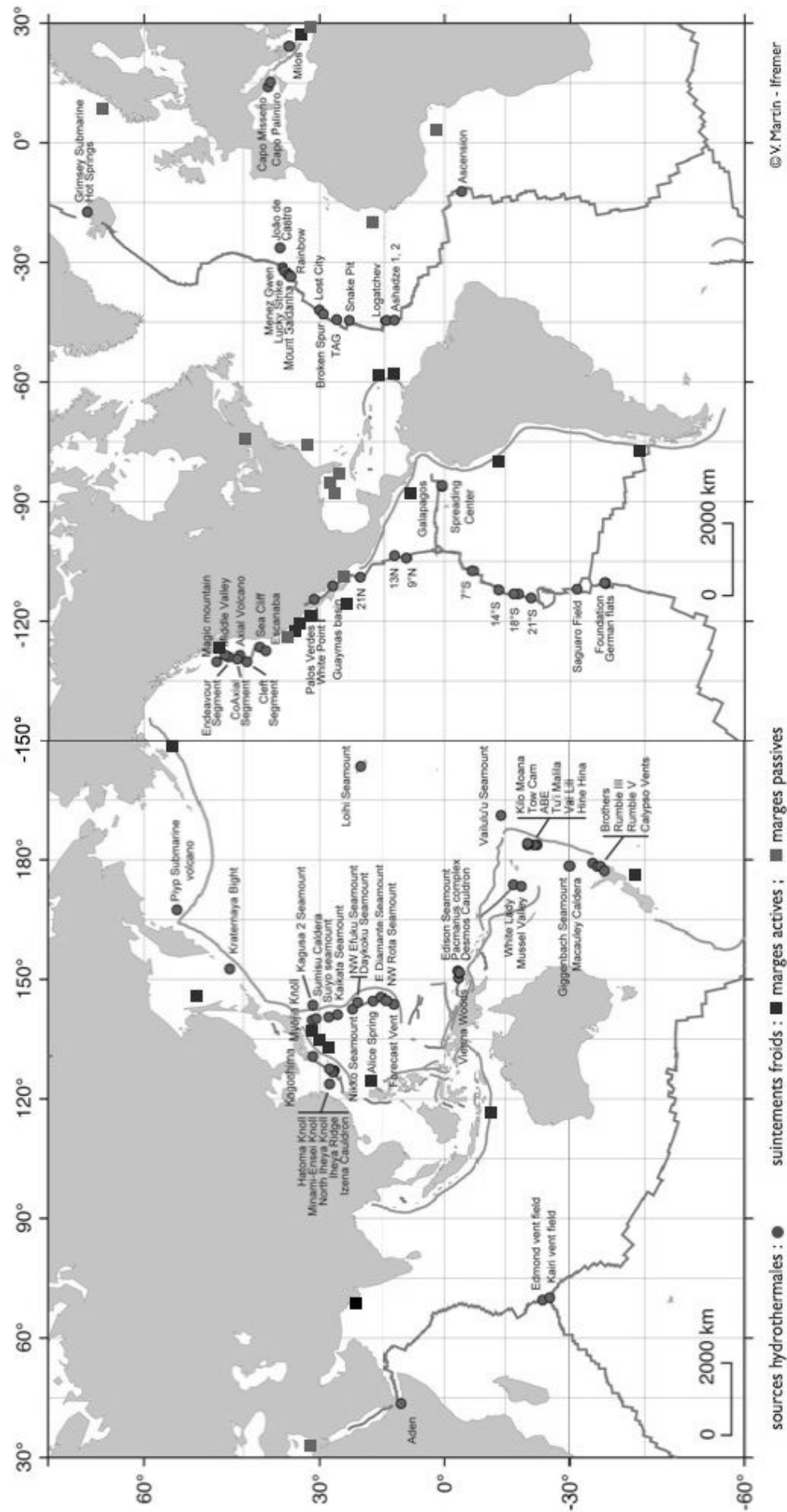
chimioautotrophes pour prospérer dans les environnements les plus inattendus : sources hydrothermales profondes, zones de suintements froids, sédiments anoxiques ou carcasses de baleines. Les animaux les plus abondants dans ces milieux vivent en fait en symbiose avec ces bactéries chimioautotrophes et présentent divers degrés d'intégration biologique: caractère facultatif ou obligatoire, ecto- ou endosymbiose, transmission horizontale ou verticale... Ces animaux symbiotiques pourraient ainsi constituer des modèles contemporains des événements supposés avoir abouti à l'intégration des  $\alpha$ -protéobactéries et des cyanobactéries, respectivement en mitochondries et chloroplastes.

Où se trouvent ces milieux réduits ? Quels types d'organismes y trouve-t-on ? Comment vivent ces symbioses chimioautotrophes ? Ce sont des questions que les scientifiques se posent depuis leur découverte, il y a maintenant une trentaine d'années.

## 2. Les milieux marins réduits contemporains

### 2.1 Les sources hydrothermales

C'est en effet à partir de 1977, après la découverte de véritables oasis de vie autour des sources hydrothermales profondes gisant par près de 2 500 mètres de fond au large des Galápagos (Corliss *et al.*, 1979), qu'un regain d'intérêt pour les écosystèmes chimiosynthétiques s'est manifesté. La circulation hydrothermale et les éruptions volcaniques qui prennent place au niveau des dorsales océaniques constituent le mécanisme majeur par lequel le globe terrestre perd quantité d'énergie sous forme de chaleur. Ce système volcanique qui s'étend sur plus de 60 000 kilomètres au fond des océans est à l'origine de l'écartement des plaques lithosphériques. Les milieux hydrothermaux profonds sont localisés précisément le long de ces rifts (voir la localisation des sites connus à ce jour sur la Figure 3). A la dorsale de l'océan Pacifique oriental (EPR, East Pacific Rise), la dorsale Médio-Atlantique, la dorsale Arctique et la dorsale de l'océan Indien, s'ajoutent les bassins arrière-arc du sud-ouest Pacifique et certains volcans intra-plaques. Pour l'ensemble de ces sites, la formation des cheminées hydrothermales suit le même processus : l'eau de mer dense et froide pénètre dans les fissures du basalte et réagissant avec la roche chaude sous-jacente, se charge en composés réduits (sulfure d'hydrogène  $H_2S$ , méthane  $CH_4$ , ammonium  $NH_4^+$ ...) et en éléments métalliques (manganèse  $Mn^{2+}$ , fer  $Fe^{2+}$ , lithium  $Li^+$ , zinc  $Zn^{2+}$ ...).



**Figure 3 :** Carte des principaux sites réduits profonds explorés depuis 1977 (Galápagos) jusqu'en 2006 (adapté de Desbruyères *et al.*, 2006) pour les sites hydrothermaux et complétée par les zones de suintements froids. A noter l'absence de sites reconnus dans l'océan Austral circum-antarctique, l'Atlantique sud et une bonne partie de l'océan Indien, certainement par défaut d'exploration à ce jour.

Le fluide hydrothermal résultant, de composition variable selon les sites mais toujours chaud, anoxique et moins dense, remonte et rejaillit à la surface sous forme de geyser ou de diffuseur. A plus de 350 °C, l'interaction du fluide chaud avec l'eau de mer froide résulte en la précipitation de minéraux métalliques générant un panache plus ou moins sombre (fumeurs noirs ou fumeurs blancs) et permettant aussi la construction progressive des cheminées hydrothermales.

### 2.2 Les suintements froids

En certains endroits des marges continentales, à des profondeurs variant de 400 à 8 000 mètres, se trouvent d'autres types de milieux réduits, les sources froides ou suintements froids ("cold seeps"). Ces milieux sont caractérisés par l'émission profonde de fluides froids riches en méthane qui peut avoir une origine biologique (décomposition de matière organique) ou thermogénique (transformation rapide de matière organique causée par de fortes températures et de fortes pressions). Dans certaines sources froides, la réduction des sulfates par des bactéries méthanotrophes produit de fortes concentrations en H<sub>2</sub>S dans les sédiments superficiels. Au sein des marges actives associées à des zones de subduction, ces milieux peuvent être constitués de volcans de boue (par exemple au Japon et en Méditerranée orientale, Figure 3). Au sein des marges passives, les écosystèmes sont généralement associés à des réservoirs de pétrole ou de gaz (notons en particulier les sites situés à la base de l'escarpement de Floride, dans le golfe du Mexique et dans l'Ouest africain) mais aussi à des volcans de boue.

### 2.3 Les zones de sédimentation

Ces milieux profonds particuliers ne sont pas les seuls environnements marins réduits : dans tous les fonds sédimentaires, l'oxygène dissous dans l'eau interstitielle est rapidement consommé par la biomasse présente dans la couche superficielle, permettant plus en profondeur le développement de métabolismes bactériens anaérobies générateurs de sulfures. Les sédiments fins des baies envasées ou les eaux calmes des mangroves sont le siège d'une forte production de matière organique génératrice de sulfures juste sous la surface du sédiment. Le phénomène peut être accentué lorsque l'eau de fond est elle-même déjà hypoxique ou anoxique, comme cela se passe dans des eaux soumises à une eutrophisation

intense et saisonnière (Kattegatt, zones d'upwelling...), ou dans des fonds isolés, coupés de la circulation profonde par un seuil (fjords, Mer Noire...).

### 2.4 Les apports organiques imprévisibles

Des apports massifs de matière organique sur les fonds abyssaux peuvent constituer des écosystèmes réduits très localisés et relativement éphémères, mais néanmoins rapidement colonisés par une microflore et une faune endémiques et opportunistes. Il s'agit par exemple des chutes de cadavres d'animaux massifs (les baleines aujourd'hui, les reptiles marins à l'ère secondaire) ou de gros débris végétaux (bois coulés). Le développement rapide de bactéries aérobies utilise localement l'oxygène ( $O_2$ ) et ne laisse que les sulfates comme accepteurs d'électrons pour d'autres bactéries anaérobies génératrices de sulfures. Ce type d'écosystèmes pourrait jouer un rôle non négligeable sur le long terme en constituant des refuges transitoires pour abriter des communautés d'organismes endémiques capables de coloniser également les écosystèmes des sources hydrothermales ou des sources froides.

### 2.5 Points communs et divergences

Bien que tous ces milieux partagent le caractère réduit qui les réunit ici, ils présentent de grandes différences dans la distribution spatiale et temporelle de ces composés réduits. Le milieu hydrothermal peut présenter des valeurs extrêmes de température, avec un fluide atteignant  $350^\circ\text{C}$  à sa sortie de la cheminée, mais offrant des températures plus basses, autour de  $10^\circ\text{C}$  dans la zone de mélange. Dans cette zone,  $O_2$ ,  $CO_2$ ,  $H_2S$  ou  $CH_4$  sont tous accessibles mais à des concentrations fortement variables à courte échelle de temps ou d'espace. En effet, par rapport aux autres milieux réduits, les sources hydrothermales sont des systèmes très dynamiques, caractérisés par des émissions chaotiques. Les suintements froids et sédiments réduits montrent au contraire des niveaux de concentrations plus stables mais très stratifiés, bien que dans la zone côtière de fortes variations saisonnières puissent s'établir.

La pérennité de ces environnements réduits est également très variable. La longévité des systèmes hydrothermaux est fortement tributaire des épisodes magmatiques et tectoniques (Delaney *et al.*, 1998 ; Shank *et al.*, 1998). Un site donné a ainsi une espérance de vie d'une dizaine d'année sur les dorsales les plus rapides (Pacifique Est) jusqu'à quelques centaines d'années sur les dorsales lentes (Atlantique), encore que nous manquions de recul dans l'étude de ces systèmes pour pouvoir l'affirmer. Les suintements froids des marges passives, sans

activité tectonique, forment au contraire des sites réduits beaucoup plus pérennes, établis sans doute sur des périodes de plusieurs dizaines de milliers d'années.

### 3. La vie dans les milieux marins réduits

Avec des concentrations importantes en  $H_2S$  et  $CH_4$  et des conditions de vie très souvent hypoxiques, voire anoxiques, ces milieux peuvent être qualifiés d'extrêmes pour la plupart des organismes eucaryotes aérobies. Pourtant, autour de la quasi-totalité des sites découverts à ce jour (Figure 3), s'établissent des communautés d'organismes caractérisées par une grande biomasse et une forte endémicité (Tunnicliffe, 1991). D'autres communautés similaires, composées d'espèces différentes mais appartenant souvent à des familles ou taxons proches ou identiques, s'établissent autour des sites de suintements froids (Sibuet et Olu, 1998 ; Tunnicliffe, 1991). Certains taxons ont une distribution restreinte : ainsi les annélides Alvinellidae, spécialistes des pôles tiède et chaud, ne sont connus que sur les sites hydrothermaux du Pacifique (Desbruyères *et al.*, 1998). Beaucoup d'autres familles endémiques ont par contre des représentants autour des sites hydrothermaux et des suintements froids, voire même sur les cadavres de baleines (Smith et Baco, 2003), comme les annélides vestimentifères (Siboglinidae) avec par exemple *Riftia pachyptila* (EPR), *Lamellibrachia luymesii* (Golfe du Mexique) ou *Escarpia southwardae* (Ouest Afrique), les bivalves Bathymodiolinae avec *Bathymodiolus thermophilus* (EPR) ou *Bathymodiolus childressii* (Golfe du Mexique), les bivalves Vesicomidae avec *Calyptogena magnifica* (EPR) ou *Calyptogena kilmeri* (Monterey bay). Certaines familles ont également des représentants côtiers et profonds comme les bivalves Lucinidae ou Thyasiridae.

#### 3.1 Producteurs primaires et chimiosynthèse

En l'absence de lumière, le réseau trophique de la plupart de ces écosystèmes ne peut pas reposer sur la photosynthèse comme dans la zone euphotique des océans ou en milieu terrestre mais repose essentiellement sur la chimiosynthèse. C'est la transformation de l'énergie chimique potentielle des composés réduits présents dans ces milieux en énergie "biologique", l'ATP, qui permet à des organismes procaryotes de synthétiser des sucres à partir du carbone inorganique issu du dioxyde de carbone ( $CO_2$ ) ou du méthane ( $CH_4$ ). Parmi ces procaryotes chimiosynthétiques, on distingue ainsi des bactéries méthanotrophes qui

utilisent le méthane à la fois comme source d'énergie et de carbone et des procaryotes chimiolithoautotrophes, dont des sulfo-oxydants qui utilisent l'énergie d'oxydation du sulfure d'hydrogène et le dioxyde de carbone comme source de carbone. Cette flore bactérienne est beaucoup plus dense en milieu hydrothermal par rapport aux eaux littorales de surface (Karl *et al.*, 1980 ; Prieur *et al.*, 1989). Sans conteste, le méthane et le sulfure jouent donc un rôle essentiel dans le maintien de ces communautés: des blooms microbiens sont par exemple observés à chaque début du cycle magmatique, formant des mattes épaisses et suggérant leur capacité à exploiter très rapidement les composés chimiques réduits dès qu'ils sont disponibles dans les émissions de fluide hydrothermal.

### 3.2 Consommateurs secondaires

Les communautés microbiennes chimiosynthétiques constituent les producteurs primaires permettant l'établissement d'un réseau trophique particulier : aux bactéries hétérotrophes, qui viennent enrichir la diversité spécifique des consortia bactériens, s'ajoutent des métazoaires, consommateurs primaires exploitant directement cette biomasse bactérienne, suivis d'une cohorte de consommateurs secondaires, prédateurs, nécrophages, brouteurs ou dépositivores. Par exemple, autour des sources hydrothermales, on trouve ainsi l'annélide polychète *Alvinella pompejana*, le gastéropode *Lepetodrilus fucensis* ou la crevette *Rimicaris exoculata* qui se nourrissent directement des procaryotes vivant à l'état libre. En milieu côtier, les communautés bactériennes benthiques des sédiments meubles forment la source trophique d'une faune psammivore, comme par exemple l'annélide polychète *Arenicola marina*, souvent présente en abondance mais peu diversifiée.

### 3.3 Les symbioses

Mais au sein des écosystèmes réduits, les espèces ayant la plus forte biomasse vivent généralement en symbiose avec des bactéries chimioautotrophes, constituant ainsi des associations métazoaires-procaryotes capables d'exploiter plus directement et plus efficacement les ressources trophiques inorganiques de ces milieux. Au sens de De Bary (De Bary, 1879), la symbiose est une association durable entre deux organismes différents vivant en étroite interaction. Les caractéristiques de telles associations sont variables : ectosymbiose, lorsque les organismes restent anatomiquement séparés, ou endosymbiose lorsque les cellules de l'un des partenaires contiennent l'autre ; caractère obligatoire ou facultatif du mode de vie



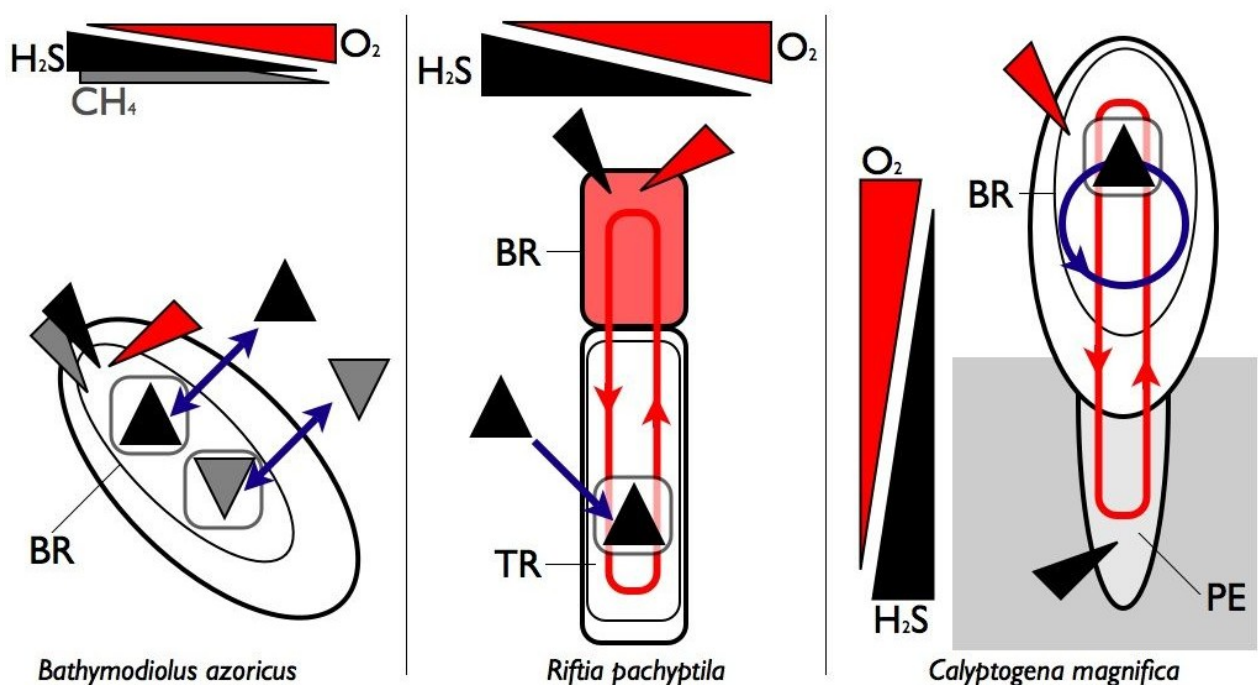
symbiotique pour l'un et/ou l'autre des partenaires ce qui sous-entend des degrés divers de dépendance physiologique ; transmission du symbiote d'une génération à l'autre de l'hôte ou acquisition *de novo* à chaque génération... Les écosystèmes chimiosynthétiques nous offrent des exemples illustrant chacune de ces caractéristiques du mode de vie symbiotique.

### 3.3.1 Les épisymbioses

Lorsque les symbiotes vivent à l'extérieur des cellules de leur hôte, on parle d'ecto- ou épisymbiose. De tels cas d'épisymbiose entre des protozoaires et des bactéries chimiosynthétiques sulfo-oxydantes ont été rapportés chez deux ciliés, le colonial *Zoothamnium niveum* et le solitaire *Vorticella* sp., communs au niveau des mangroves du récif de Belize (Bauer-Nebelsick *et al.*, 1996 ; Ott et Bright, 2004). Ces deux organismes sont complètement recouverts d'un tapis blanc bactérien, dont ils se nourrissent abondamment ce qui leur permet de maintenir un fort taux de croissance. Parmi les métazoaires, de nombreux exemples ont été décrits, notamment chez les Nématodes (Ott *et al.*, 2004). L'un des exemples les plus récemment étudiés et extrêmement démonstratif est le ver oligochète *Olavius algarvensis* qui vit dans les sédiments côtiers réduits, notamment à proximité des herbiers de Zostères. Ce ver entretient, juste sous sa cuticule, de multiples symbiotes extracellulaires (au moins 5 souches distinctes) pouvant réduire le sulfate ( $\delta$ -protéobactéries) ou oxyder l' $\text{H}_2\text{S}$  ( $\gamma$ -protéobactéries) pour générer l'énergie nécessaire (Woyke *et al.*, 2006). Les différents métabolismes réalisables en combinant les capacités de chacun des acteurs seraient mis en jeu en fonction du niveau de vie dans l'épaisseur du sédiment, permettant la vie de cette symbiose dans une large gamme de conditions physico-chimiques entre mondes anoxique et aérobie. On trouve également des ectosymbioses autour des sources hydrothermales profondes, et notamment chez les organismes colonisant le pôle chaud, à même les cheminées hydrothermales. Ainsi, *Alvinella pompejana* a le dos couvert de bactéries fixées sur des excroissances cuticulaires, certaines d'entre elles étant des chimioautotrophes sulfo-oxydantes. De manière similaire, la crevette *Rimicaris exoculata*, dont les essaims colonisent les cheminées de la dorsale médio-atlantique, "cultive" une communauté de bactéries dans ses chambres branchiales sur les épipodites de ses pièces buccales. Dans ces deux exemples, le rôle trophique de ces ectosymbioses bactériennes n'est pas complètement élucidé : culture fourragère, translocation de nutriments ou encore protection contre la toxicité du milieu, font toujours partie des hypothèses envisagées (Desbruyères *et al.*, 1998 ; Ott *et al.*, 2004).

### 3.3.2 Les endosymbioses

La biomasse la plus importante rencontrée en milieu hydrothermal et en suintement froid est néanmoins constituée de consommateurs primaires qui ont tous développé une stratégie d'association endosymbiotique. Les bactéries sont donc localisées à l'intérieur de cellules différenciées de l'hôte, les bactériocytes. Ces bactéries sont assez peu diversifiées: ce sont soit des sulfo-oxydantes (pour la plupart des  $\gamma$ -protéobactéries, mais il existe aussi des  $\epsilon$ -protéobactéries) ou des méthanotrophes ( $\gamma$ -protéobactéries) qui se répartissent dans une demi-douzaine de clades (Stewart *et al.*, 2005). Les invertébrés marins auxquels ces bactéries sont associées appartiennent aux Annélides (Polychètes et Oligochètes) ou aux Mollusques (Bivalves et Gastéropodes). Si on se limite strictement aux symbioses intracellulaires, le nombre de familles est encore plus restreint : Siboglinidae chez les Annélides, Provannidae chez les Gastéropodes, Solemyidae, Lucinidae, Thyasiridae, Vesicomidae et Mytilidae chez les Bivalves (Stewart *et al.*, 2005). Nous nous limiterons à décrire les principales caractéristiques de ces endosymbioses sur trois exemples particulièrement démonstratifs (Figure 4) : le bivalve Mytilidae *Bathymodiolus azoricus*, l'annélide Siboglinidae *Riftia pachyptila* et le bivalve Vesicomidae *Calyptogena magnifica*.



**Figure 4 :** Représentation schématique de trois symbioses hydrothermales chimioautotrophes. *Bathymodiolus azoricus* vit dans un gradient avec des concentrations variables dans le temps d'oxygène, sulfure d'hydrogène et méthane. Cette espèce abrite dans ses branchies (BR) des bactéries

Suite de la légende de la Figure 4 :

sulfo-oxydantes (triangle noir) et méthanotrophes (triangle gris) qu'elle acquiert dans l'environnement et qu'elle peut perdre (↔). *Riftia pachyptila* vit dans un gradient avec des concentrations variables dans le temps d'oxygène et de sulfure d'hydrogène qu'elle capte à travers sa branchie (BR) et transporte jusqu'au trophosome (TR) pour alimenter les bactéries endosymbiotiques sulfo-oxydantes (triangle noir) qu'elle a acquises une fois pour toutes lors de son recrutement (↘). *Calymene magnifica* vit dans un gradient spatial et capte l'oxygène dans l'eau et les sulfures dans les fissures du basalte où elle insère son pied extensible (PE) pour alimenter les bactéries endosymbiotiques sulfo-oxydantes (triangle noir) logées dans ses branchies (BR) et qu'elle transmet à sa descendance (↻).

### 4. Les degrés d'intégration des endosymbioses chimioautotrophes

#### 4.1 L'intégration cytologique

Dans l'établissement d'une endosymbiose, la première étape à laquelle on pense est l'intégration cytologique. La localisation anatomique des endosymbiontes est variable mais les bactéries endosymbiotiques finissent toujours par se retrouver dans des vacuoles intracytoplasmiques. Chez les bivalves (modioles *Bathymodiolus*, clams *Calymene*, Lucinidae *Codakia*) les bactéries sont logées dans les branchies des hôtes, dans les cellules de l'épithélium branchial. Le tissu branchial des espèces endosymbiotiques est toujours hypertrophié et d'aspect turgescents mais la distance entre bactéries et milieu extérieur reste courte, laissant supposer des échanges de métabolites milieu extérieur – hôte – bactéries relativement aisés. A noter que chez différentes espèces de bivalves Thyasiridae on peut observer une progression depuis l'état d'ectosymbiose vers celui d'endosymbiose, avec un état intermédiaire où les bactéries se logent entre les microvillosités de la surface apicale des cellules branchiales (Dufour, 2005). Chez les Siboglinidae tels que *Riftia pachyptila*, les bactériocytes sont situés dans un tissu symbiotique spécialisé, le trophosome, qui forme un organe interne dans la cavité coelomique (le tube digestif est absent chez l'adulte). Dans ce cas les bactéries se trouvent complètement isolées du milieu extérieur, et dépendent donc entièrement de leur hôte et de sa circulation sanguine pour la fourniture des composés inorganiques nécessaires à son métabolisme (Figure 4).

#### 4.2 L'intégration physiologique

Le caractère facultatif ou obligatoire de la symbiose et le degré d'interdépendance entre les deux acteurs définissent le niveau d'intégration physiologique. Chez *Bathymodiolus*

il est encore limité : les moules possèdent un tube digestif rudimentaire et sont encore capables d'ingérer des particules en suspension pour se nourrir (Page *et al.*, 1991) et l'espèce *B. azoricus* peut même perdre ses bactéries dans certaines conditions défavorables puis les acquérir de nouveau (Kadar, 2005). Cette capacité de nutrition alternée, la mixotrophie, permet aux populations de modioles de résister plus longtemps lors de l'obstruction temporaire des conduits hydrothermaux. Par contre, pour *Riftia pachyptila* ou *Calyptogena magnifica*, dépourvus de tube digestif fonctionnel, la symbiose revêt un caractère obligatoire pour l'hôte : si la source de fluide se tarit c'est l'hécatombe dans les champs de vers ou de clams, spectacle de tubes et de coquilles vides maintes fois observés sur le fond.

Dans tous les cas, le métabolisme sulfo-oxydant des bactéries nécessite des concentrations importantes de sulfure, mais surtout d'oxygène. La consommation d'oxygène de *Riftia* est multipliée par quatre lorsqu'on lui fournit des sulfures en concentration non limitante (Childress *et al.*, 1991). La zone de mélange entre fluide hydrothermal, riche en sulfure, et l'eau de fond, relativement bien oxygénée, est donc propice au développement de telles symbioses. Le bivalve *C. magnifica*, qui s'établit plus loin des sources, puise les sulfures entre les fissures du basalte en y insérant son pied extensible (Figure 4). Dans les zones de sédiments réduits, où, comme on l'a vu précédemment, les sulfures restent généralement piégés au sein de la couche anoxique du sédiment, certaines espèces procèdent de même, puisant les sulfures dans le sous-sol et l'oxygène dans l'eau sus-jacente. C'est le cas des bivalves de la famille des Thyasiridae (Dufour et Felbeck, 2003) ou des Siboglinidae qui colonisent les zones de suintement froid (Freitag *et al.*, 2001). Parfois c'est la mobilité de l'hôte, nématode ou oligochète, dans la zone d'interface entre couche oxique et anoxique, qui permet d'alimenter alternativement le métabolisme des symbiontes en oxygène et en sulfure (Bright et Giere, 2005 ; Ott *et al.*, 2004).

Enfin, le transfert de molécules organiques synthétisées par les bactéries endosymbiotiques vers leur hôte peut se faire par translocation moléculaire à travers les membranes ou par digestion intracellulaire des bactéries. La dépendance trophique de l'hôte vis-à-vis des bactéries a été démontrée pour nos trois espèces type, notamment par des mesures d'isotopes stables du carbone et de l'azote montrant une relation très nette entre la composition des bactéries et celle du tissu les renfermant (Colaco *et al.*, 2002). Par ailleurs, l'observation en microscopie électronique de coupes de tissu symbiotique a révélé l'existence de lysosomes renfermant des bactéries dégradées chez plusieurs espèces, en commençant par *Riftia* (Bosch et Grassé, 1984).

#### 4.3 L'intégration génétique

La spécificité de la relation symbiotique et les problèmes de reconnaissance interspécifiques sont très divers parmi les exemples connus d'endosymbioses. *Bathymodiolus azoricus* abrite ainsi deux souches de bactéries (Duperron *et al.*, 2006; Fiala-Médioni *et al.*, 2002), l'une thiotrophe et l'autre méthanotrophe (Figure 4). Par contre, toutes les espèces de Siboglinidae hydrothermaux de la dorsale EPR abritent une souche apparemment unique de  $\gamma$ -protéobactéries (Feldman *et al.*, 1997a). Mais, autour des sources du Pacifique Ouest, deux espèces sympatriques de gastéropodes, tous deux des Provannidae, abritent l'un des  $\epsilon$ -protéobactéries (*Alviniconcha hessleri*) et l'autre des  $\gamma$ -protéobactéries (*Ifremeria nautilei*), bien que ces deux souches soient des sulfo-oxydantes (Urakawa *et al.*, 2005). Chez les Vesicomidae on a en revanche une forte spécificité, chaque espèce de clams abritant une souche de  $\gamma$ -protéobactéries distincte, que les hôtes soient phylogénétiquement ou géographiquement proches (Hurtado *et al.*, 2003). Cette variation dans la spécificité est à mettre en relation avec le mode de transmission des bactéries endosymbiotiques d'une génération à l'autre, et plus largement avec le degré d'intégration génétique des deux partenaires. Chez les bivalves de la famille des Vesicomidae, les bactéries sont transmises à la génération suivante via les ovocytes (Cary et Giovannoni, 1993) : c'est une transmission verticale. Cela entraîne une co-évolution entre hôte et symbiote que l'on retrouve en comparant les arbres phylogénétiques des deux partenaires (Hurtado *et al.*, 2003).

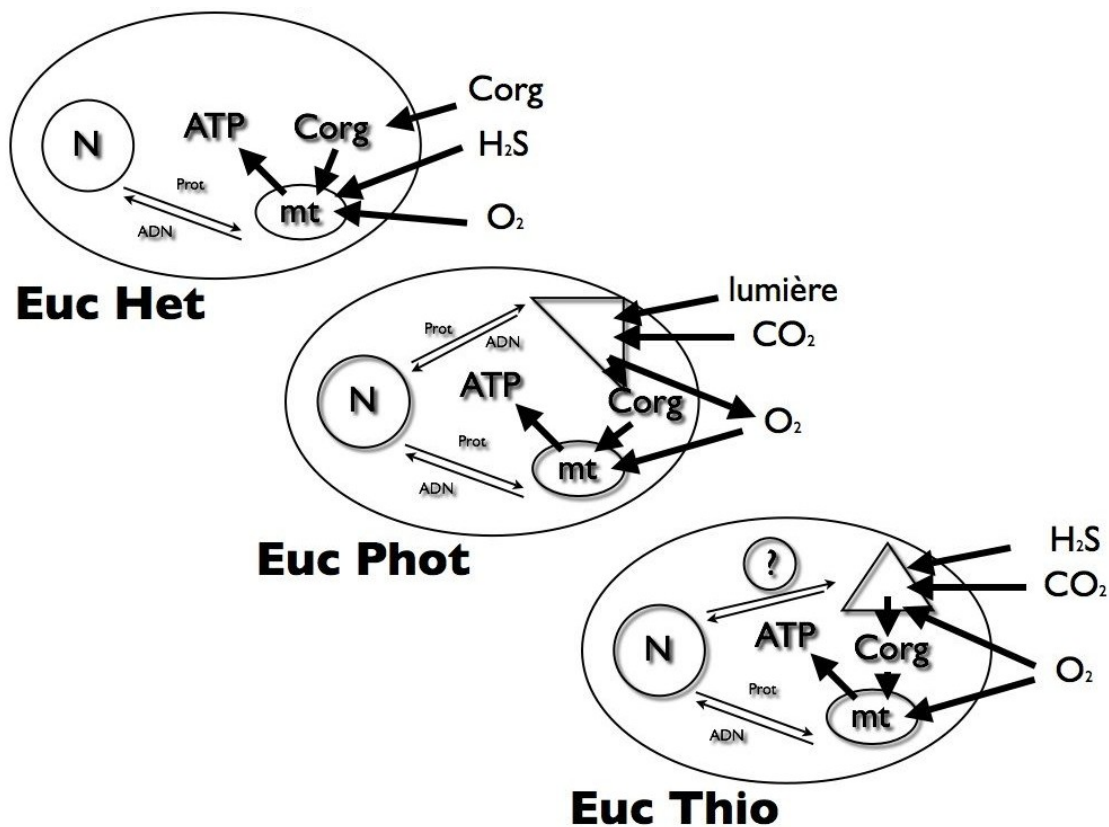
Paradoxalement, chez *Riftia*, alors que l'intégration physiologique est très poussée, la transmission des symbiotes d'une génération à l'autre se fait par transmission horizontale ou environnementale, les juvéniles de *Riftia*, dépourvus de bactéries mais pourvus d'un tube digestif, devant trouver dans leur milieu la souche bactérienne unique qui leur permettra de développer la symbiose. Plus exactement, selon les derniers travaux de l'équipe de Monika Bright (Nussbaumer *et al.*, 2006), ce serait la bactérie qui infesterait l'hôte, lequel serait alors capable de limiter et de contrôler la prolifération bactérienne au sein d'un seul type cellulaire, les bactériocytes du trophosome. Les cellules d'autres tissus infectées sur le parcours entre épiderme et trophosome montrent rapidement des figures d'apoptose (Nussbaumer *et al.*, 2006). Dans le cas des moules du genre *Bathymodiolus*, il s'agit aussi d'une transmission horizontale (Won *et al.*, 2003), sans doute plus facile étant donné la localisation branchiale des symbiotes et le caractère plus labile de la symbiose. Les modes de transmission des symbiotes utilisés par les hôtes sont fortement dépendants des contraintes biogéographiques

et écologiques qui affectent hôtes et symbiontes (Won *et al.*, 2003). Si la transmission verticale des symbiontes thiotrophes chez *Calymene* est efficace pour la colonisation de milieux riches en  $H_2S$ , le mode de transmission environnemental chez *Bathymodiolus* lui permet d'exploiter de façon flexible un régime trophique plus large, les bactéries, qu'elles soient thiotrophes ou méthanotrophes, étant déjà localement adaptées.

### 5. Du symbionte à l'organite

Comme nous l'avons vu, dans notre monde dominé par le métabolisme aérobie depuis près d'un milliard d'années (Figure 1), il subsiste encore de nombreux habitats où les sulfures, naguère prépondérants, sont encore présents (Figure 3). Les bactéries capables d'utiliser leur potentiel énergétique sont toujours là et certaines espèces d'invertébrés mettent le métabolisme de ces bactéries à profit au sein d'association endosymbiotiques très performantes (Figure 4). On peut alors se demander pourquoi de telles associations n'ont pas évolué plus avant en intégrant ces symbiontes de façon encore plus intime, les transformant de ce fait en organites. C'est ainsi que sont nées les mitochondries, à partir d' $\alpha$ -protéobactéries capables d'utiliser l'oxygène, et les chloroplastes, à partir de cyanobactéries pouvant convertir l'énergie lumineuse en énergie chimique, donnant naissance aux lignées d'Eucaryotes hétérotrophes aérobies et d'Eucaryotes photoautotrophes avec le succès évolutif que l'on sait (Figure 2). Les Eucaryotes thioautotrophes existent peut-être sans que nous les ayons encore découverts : la découverte de *Riftia* n'a que 30 ans et nous sommes loin d'avoir catalogué toutes les formes de vie de notre planète (et particulièrement pour les protistes des fonds océaniques, par exemple).

En fait, alors que les Eucaryotes hétérotrophes consomment de l'oxygène et que les Eucaryotes phototrophes en produisent également, les besoins d'une symbiose sulfo-oxydante consistent à la fois en sulfure et en oxygène à deux titres (Figure 5). L'oxygène est nécessaire aux bactéries pour fixer le carbone inorganique et à l'hôte pour oxyder le carbone organique produit par les bactéries. Si l'on prend en compte le fait que les profondeurs de l'océan sont restés anoxiques jusqu'à une période récente, peut-être moins de 600 millions d'années (Anbar et Knoll, 2002) (Figure 1), il n'est peut-être pas surprenant que les Eucaryotes thiotrophes n'aient pas encore vu le jour. D'autant que les endosymbioses primaires semblent être des phénomènes rares dans l'évolution, mitochondries et plastes étant issus d'événements uniques.



**Figure 5 :** Représentation schématique des Eucaryotes hétérotrophes (Euc Het), phototrophes (Euc Phot) et thiotrophes (Euc Thio), ces derniers étant hypothétiques, représentés actuellement par des endosymbioses métazoaires –  $\gamma$ -protéobactéries (adapté de Vetter (1991) et Lallier (2004)). Les Eucaryotes hétérotrophes métabolisent le carbone organique (Corg) de manière aérobie grâce aux mitochondries (mt) pour former de l'ATP et certains d'entre eux sont aussi capables d'oxyder les sulfures (H<sub>2</sub>S) et d'en retirer un peu d'énergie (chimiohétérotrophie). Les Eucaryotes phototrophes peuvent assimiler le carbone inorganique (CO<sub>2</sub>) dans leurs chloroplastes à partir de l'énergie de la lumière, produisant du carbone organique et de l'oxygène nécessaire à leur métabolisme aérobie. Enfin, les Eucaryotes thiotrophes seraient capables d'assimiler le carbone inorganique grâce à l'énergie chimique résultant de l'oxydation des sulfures par des "thioplastes", organites hypothétiques issus d'endosymbiontes. Mais si des transferts de gènes (ADN) et de protéines (Prot) ont bien eu lieu entre mitochondries, chloroplastes et noyau, cela reste à démontrer pour les Eucaryotes thiotrophes.

En d'autres termes, on ne connaît pas d'autre endosymbiose primaire entre procaryote et eucaryote que les deux précédemment citées, à moins que leur évolution ne soit actuellement en cours à partir des différents exemples d'endosymbioses chimioautotrophes évoqués ci-dessus (Lallier, 2006). L'intégration la plus aboutie que nous connaissions aujourd'hui est sans doute *Calyptogena*. Mais qu'est-ce qui différencie un endosymbionte à transmission verticale d'un organite ? L'une des caractéristiques de ces derniers est d'avoir un génome réduit, fruit de pertes de gènes et/ou de transferts vers le génome nucléaire. Pour déceler un tel début il

faudrait connaître le génome du symbionte de *Calymmatobacter*. Ce qui est chose faite aujourd'hui (Newton *et al.*, 2007). Mais si le génome de cette bactérie est de taille modérée (1,2 Mb) il semble posséder un jeu de gènes complet. Les Eucaryotes thiotrophes ne seraient donc pas encore parmi nous.



## **CHAPITRE II**

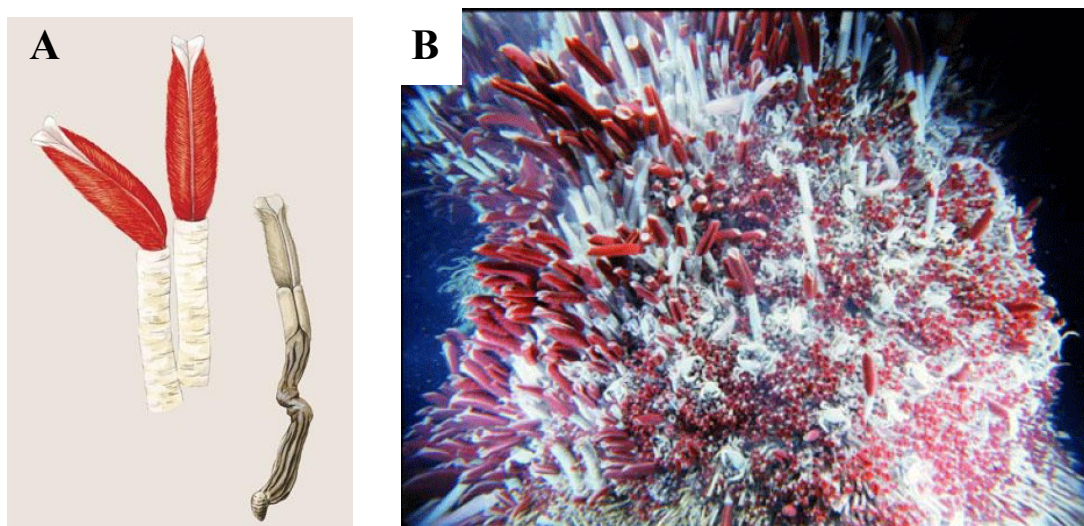
**L'organisme d'étude, *Riftia pachyptila***

## CHAPITRE II : L'organisme d'étude, *Riftia pachyptila*

### 1. Place phylogénétique

#### 1.1 Un Annélide Polychète Siboglinidae

Comme nous l'avons vu au cours du chapitre précédent, *Riftia pachyptila* est un organisme endémique des sources hydrothermales de la dorsale EPR. Cet animal tubicole (Figure 1A) forme de denses « bouquets » avec ses congénères à proximité des émissions de fluide hydrothermal (Figure 1B).



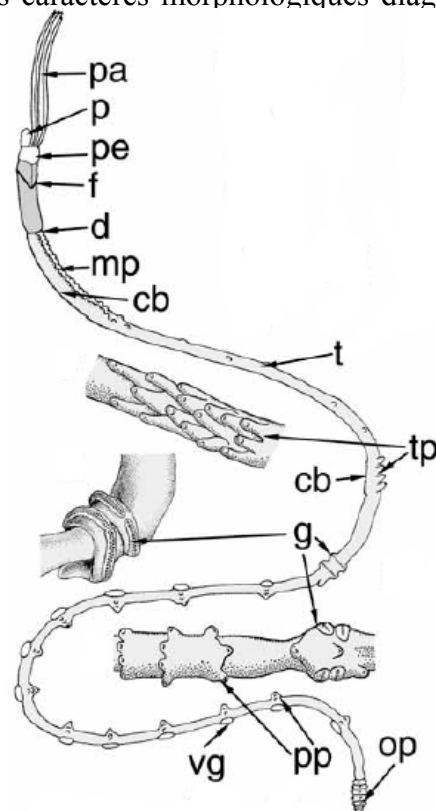
**Figure 1 :** (A) Vue dorsale de *Riftia pachyptila* dans et hors de son tube (Dessin : Michigan Science Art, USA). (B) Colonie de *Riftia pachyptila* en milieu naturel (Cliché : Ifremer).

Aujourd'hui reconnu en tant qu'Annélide, il n'est cependant pas rare de trouver les termes Pogonophore ou Vestimentifère comme seules données taxonomiques (dans GenBank par exemple). Un petit retour historique permet de comprendre la succession des appellations réservées à *Riftia pachyptila* et ses apparentés (voir Rouse (2001) pour un historique très complet). En 1914, Caullery décrit un organisme sans tube digestif apparent, *Siboglinum weberi*, qui devint le seul représentant de la famille des Siboglinidae rattachée à aucune autre famille. De son côté, Uschakov (1933) découvre un animal assez semblable, *Lamellisabella zachsi*, qu'il place aussitôt dans la famille des Sabellidae, au sein des Polychètes. Quelques

années plus tard, Johansson (1937, 1939) « range » *L. zachsi* comme seul représentant d'une nouvelle classe, les Pogonophores, plus tard érigée en tant que phylum au sein des Deutérostomiens.

C'est Ivanov (1951) qui remarque la proche parenté de *Siboglinum weberi* et *Lamellisabella zachsi*, et décrit par la suite d'autres taxons semblables qu'il regroupe tous au sein des Pogonophores. De 1967 à 1975, une controverse sur l'appartenance des Pogonophores aux Deutérostomiens ou Protostomiens est en cours. Elle ne sera tranchée qu'en 1975 avec la description d'un exemplaire complet de *Lamellibrachia* par van der Land et Nørrevang, exemplaire incluant l'opisthosome, partie postérieure segmentée et possédant des uncini, la présence de ces soies rattachant définitivement ce taxon aux Annélides. Avec la découverte des sources hydrothermales et de leur foisonnant foyer de vie en 1977, de nouveaux organismes, dont *R. pachyptila*, sont décrits et très vite rapprochés des Pogonophores, mais Jones (1985) érige un nouveau phylum pour ces espèces hydrothermales, les Vestimentifères. La monophylie de l'ensemble (Pogonophores + Vestimentifères) est enfin suggérée, puis démontrée par des données morphologiques (Southward, 1988 ; Rouse et Fauchald, 1995) et des données moléculaires (sous-unité I de la cytochrome oxydase (COI), Black *et al.*, 1997 ; 18S, Halanych *et al.*, 1998 ; EF-1 $\alpha$ , McHugh, 1997). Dans le même temps, la place des Pogonophores et Vestimentifères parmi les Annélides Polychètes initialement formulée par Uschakov (1933) se vérifie. Bientôt, de nombreux auteurs confirment la nouvelle position phylogénétique, et certains suggèrent de revenir au nom de Siboglinidae (Rouse et Fauchald, 1997 ; McHugh, 1997), première taxonomie proposée par Caullery (1914). Rouse (2001) a recensé quelques caractères morphologiques diagnostiques de la famille (Figure 2).

**Figure 2 :** Caractères morphologiques généraux pouvant être rencontrés au sein des Siboglinidae (d'après Rouse, 2001). Cb = bande ciliée, d = diaphragme, f = frenulum (chez les Pogonophores), g = ceinture d'uncini à mi-tronc, mp = papilles métamériques, op = opisthosome, lamelles enveloppantes externes, p = prostomium, pa = palpes, pe = péristomium, pp = ligne ou colonne de papilles du tronc postérieur, t = tronc, tp = zone épaisse de papilles, vg = bouclier de glandes ventrales. Note : Les Vestimentifères ont un obturaculum mais pas de frenulum, celui-ci est une caractéristique exclusive des Pogonophores.

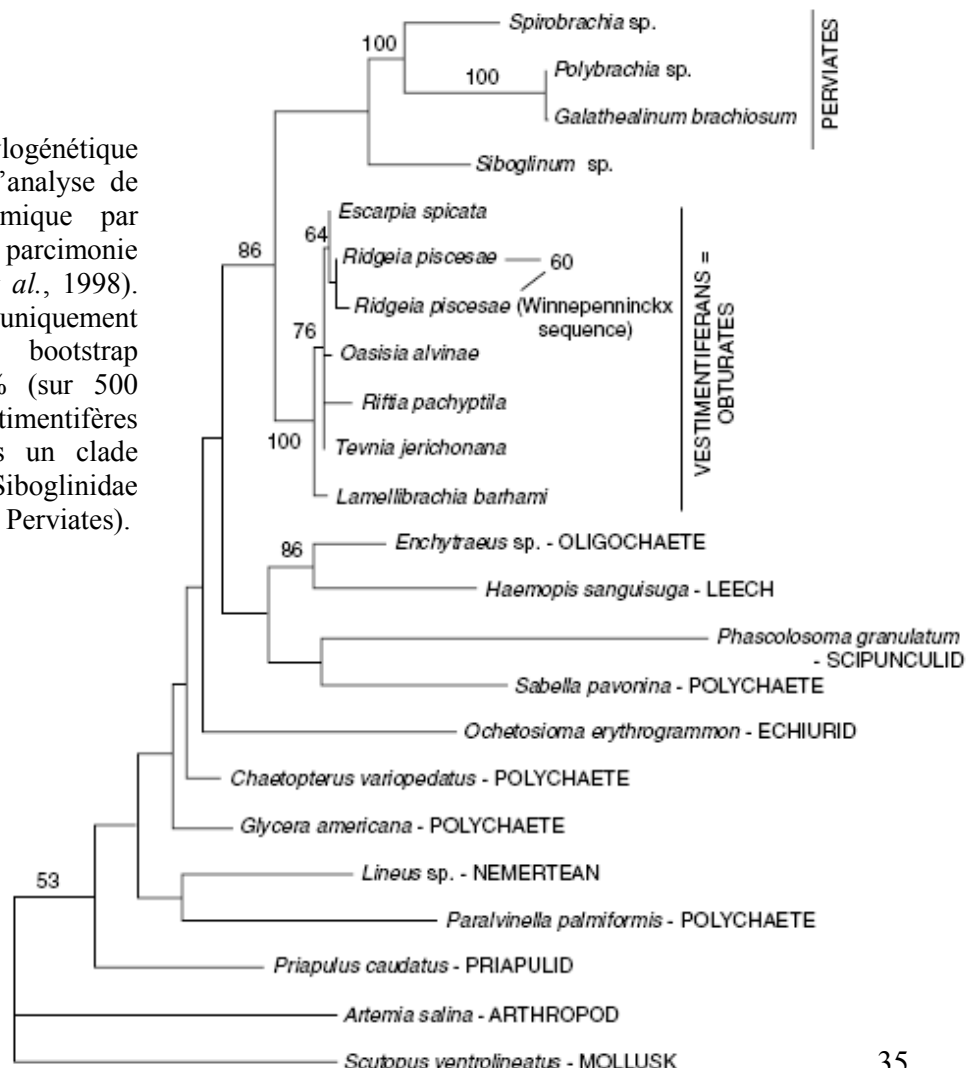


Si l'aspect vermiforme et la larve trochophore des Siboglinidae rappellent des caractères propres aux Annélides, de multiples données indépendantes montrent clairement que les Clitellates (Oligochètes et Hirudinés), Echiuridae et Siboglinidae appartiennent à la radiation des Polychètes (voir McHugh, 2000 ; Halanych, Dahlgren, et McHugh 2002 ; Shulze, 2003 ; Halanych, 2004, 2005).

### 1.2 Relations phylogénétiques au sein des Siboglinidae

Dans cette partie, nous nous focaliserons sur les relations phylogénétiques au sein des Siboglinidae hydrothermaux et des suintements froids (genres *Riftia*, *Ridgeia*, *Oasisia*, *Tevnia*, *Lamellibrachia*, *Seepiophila*, *Paraescarpia* et *Escarpia*). Celles-ci sont en effet difficiles à déterminer à partir de caractères morphologiques à cause d'une grande plasticité phénotypique. En ce qui concerne les données moléculaires, l'ARN 18S, le COI et une partie de l'ARN 28S (du côté 3') montrent de faibles niveaux de diversité nucléotidique. Cette faible diversité moléculaire se traduit notamment par de courtes branches terminales visibles sur l'arbre phylogénétique réalisé à partir de l'ARN 18S (Figure 3).

**Figure 3 :** Arbre phylogénétique obtenu à partir de l'analyse de l'ARN 18S ribosomique par maximum de parcimonie (d'après Halanych *et al.*, 1998). Sont mentionnées uniquement les valeurs de bootstrap supérieures à 50 % (sur 500 itérations). Les Vestimentifères sont regroupés dans un clade distinct des autres Siboglinidae (les Pogonophores ou Perviates).



Par ailleurs, le niveau de divergence nucléotidique des séquences de COI au sein des Siboglinidae hydrothermaux (de 0,13 à 0,20, Halanych *et al.*, 1998) pourrait être associé à des taxons d'âge très récent (moins de 50 millions d'années) remontant donc au Cénozoïque (Wray *et al.*, 1996). De même, le niveau de divergence nucléotidique des séquences de 28S (0,00171 à 0,01559, Halanych *et al.*, 1998) se situe en deçà du niveau de divergence nucléotidique des séquences de 28S entre le rat et la souris (0,02046), des taxons connus pour avoir subi une radiation entre le Paléocène et l'Eocène (entre –65 et –34 millions d'années).

D'après l'étude phylogénétique menée par Black *et al.* (1997) à partir de séquences du COI, les espèces strictement endémiques des sources hydrothermales (genres *Riftia*, *Ridgeia*, *Oasisia* et *Tevnia*) se groupent dans un même clade. Les espèces associées aux suintements froids (genres *Escarpia* et *Lamellibrachia*), paraphylétiques, émergent à la base du clade. La nature supposée opportuniste de ces vers, étayée notamment par la découverte de l'espèce *Escarpia spicata* associée à un cadavre de baleine (Feldman *et al.*, 1997b), supporte l'hypothèse de l'invasion et la colonisation du milieu hydrothermal par un ancêtre provenant des suintements froids (Black *et al.*, 1997). Ce type de phénomène est aussi mis en avant chez les bivalves Bathymodiolinae (Craddock *et al.*, 1995).

### 1.3 Espèces symbiotiques au sein des Siboglinidae

Toutes les espèces de Siboglinidae vivent en symbiose avec des bactéries. Chez les Vestimentifères, l'acquisition du symbionte est en premier lieu définie par la disponibilité de la forme libre de la bactérie dans l'environnement mais il faut également que ce symbionte potentiel soit localement adapté aux conditions environnementales auxquelles vit son hôte. Une phylogénie des symbiontes à partir du 16S a montré que ceux-ci se divisent en deux clades distincts : le premier comprenant les symbiontes associés aux vers endémiques des sources hydrothermales (*Riftia*, *Ridgeia*, *Oasisia* et *Tevnia*), le second comprenant les symbiontes associés aux vers des zones de suintement froid (*Lamellibrachia* et *Escarpia*) (Feldman *et al.*, 1997a ; McMullin *et al.*, 2003). Les symbiontes hydrothermaux apparaissent clairement spécifiques de leurs hôtes (toutefois sans spécificité de l'hôte parmi les espèces hydrothermales), tandis que différentes lignées de symbiontes caractérisent les hôtes des zones de suintement froid. Au sein du second clade, les symbiontes sont trouvés aussi bien dans l'Atlantique que dans le Pacifique, suggérant que la distribution des symbiontes au sein des différentes espèces ne paraît pas fortement affectée par la distribution géographique. Néanmoins, celle-ci serait affectée par la profondeur (McMullin *et al.*, 2003). Le fort degré

d'identité des séquences du premier clade entre elles suggère par ailleurs que les vers des sources hydrothermales s'associent avec une même lignée de symbiontes ( $\gamma$ -protéobactéries) distincte de celle des espèces associées aux suintements froids, renforçant de ce fait l'hypothèse d'une transmission horizontale ou environnementale des symbiontes.

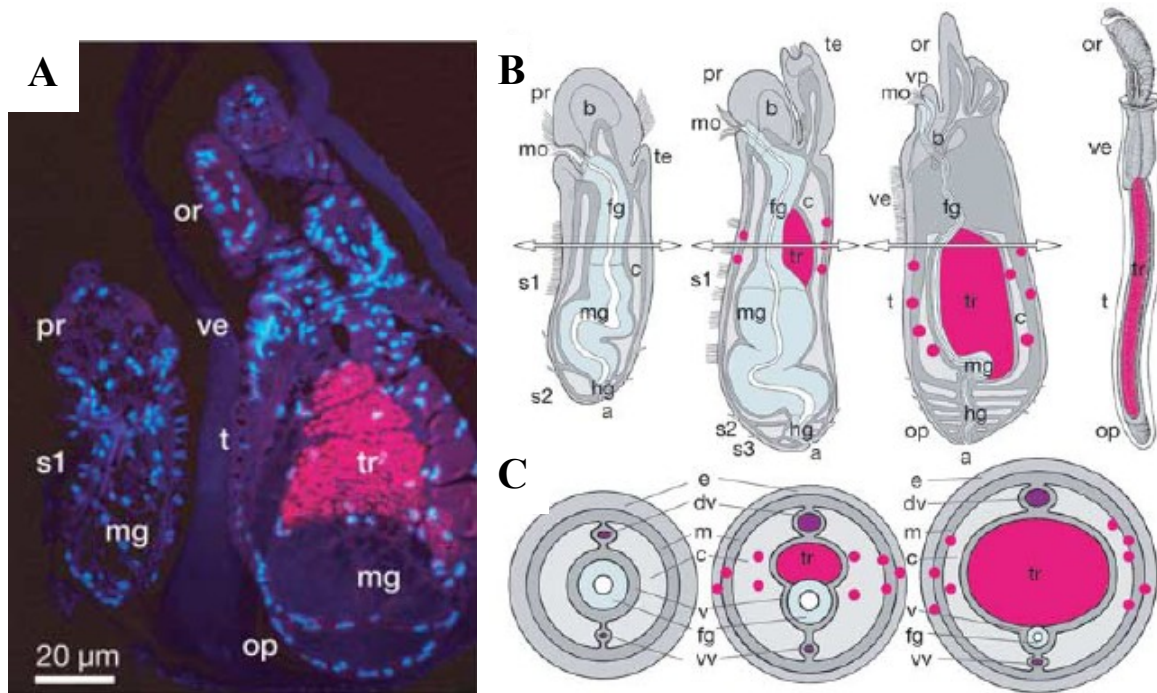
## 2. Aspects du développement

### 2.1 Acquisition environnementale des symbiontes

Si à l'état adulte, *Riftia pachyptila* vit en symbiose avec des bactéries logées dans un organe spécifique, le trophosome, la larve ne possède pas de symbionte. A l'opposé d'un mode de transmission vertical des symbiontes, par les gamètes parentaux, le ver doit trouver ses symbiontes dans l'environnement à chaque génération. La transmission horizontale par acquisition des symbiontes à partir du milieu extérieur avait déjà été suggérée chez *Ridgeia* et *Riftia* (Southward *et al.*, 1988) mais des travaux plus récents viennent préciser encore mieux ce mécanisme.

### 2.2 Les étapes de l'acquisition des symbiontes

La larve aposymbiotique peut se nourrir activement de toutes sortes de bactéries, grâce à son système digestif rudimentaire composé d'une bouche donnant sur une petite cavité buccale en position ventrale, un tube digestif et un anus (Nussbaumer *et al.*, 2006). L'hypothèse de l'établissement de la symbiose à partir de l'ingestion de bactéries au niveau du tube digestif de la larve avait été émise (Jones et Gardiner, 1988 ; Southward, 1988). Selon cette hypothèse, les symbiontes entrent par la bouche et le système digestif, et la prolifération des cellules du tube digestif (endoderme) donnent naissance au trophosome. Nussbaumer *et al.* (2006) soutiennent un autre scénario selon lequel la paroi du corps des larves est le seul site d'entrée des symbiontes, les tissus infectés étant la paroi du corps, les muscles adjacents et le tissu mésodermique non différencié (Nussbaumer *et al.*, 2006). Par hybridation *in situ* avec des sondes bactériennes spécifiques, Nussbaumer *et al.* (2006) ont démontré l'absence de symbiontes chez les stades larvaires précoces (aucun symbionte dans la bouche, le tube digestif et l'anus) et la présence de symbiontes chez des juvéniles (stade post-larvaire de 250  $\mu$ m de longueur) (Figure 4A).



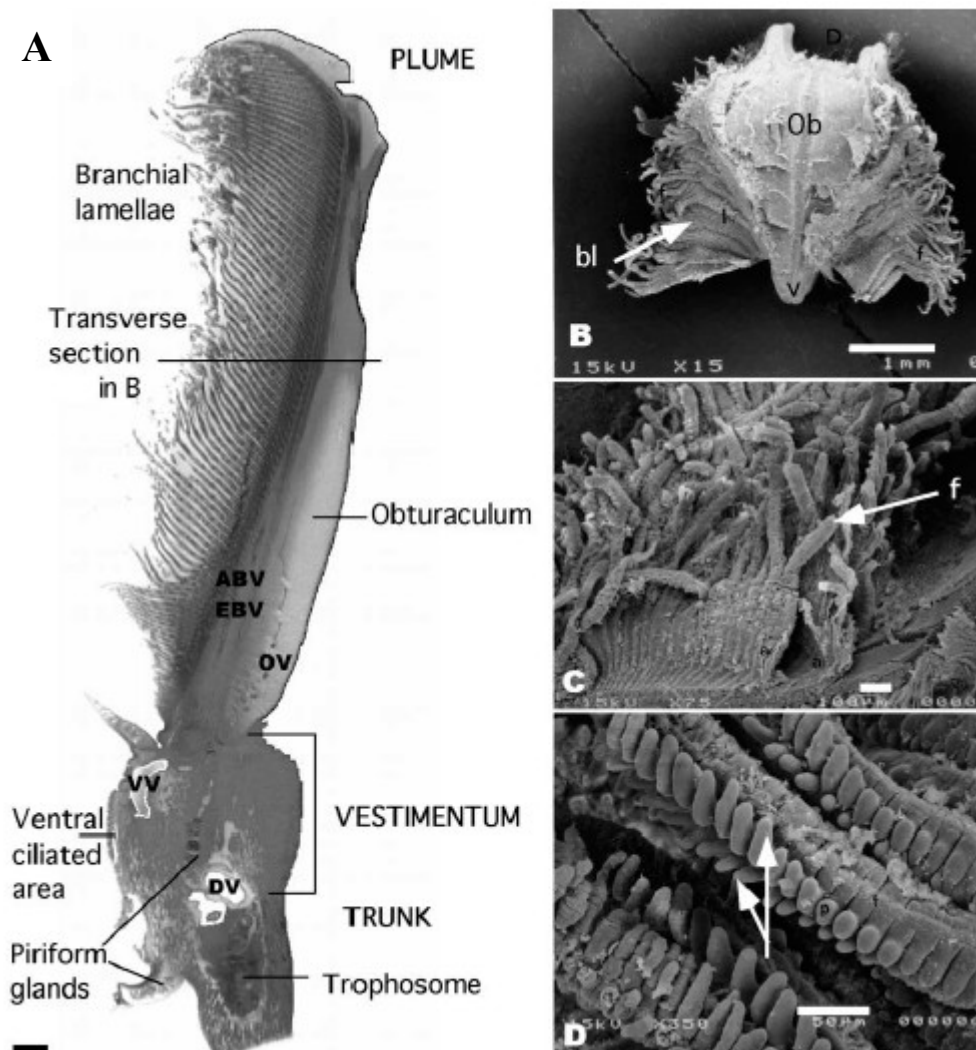
**Figure 4 :** (A) Coupes longitudinales de larve de *Riftia* aposymbiotique (gauche) et de juvénile symbiotique (droite) fixés (d'après Nussbaumer *et al.*, 2006). Le résultat de l'hybridation des sondes spécifiques des symbiontes est visible en rouge. Les noyaux sont marqués au DAPI (bleu). (B) Schémas de coupes longitudinales de larve aposymbiotique, larve infectée, juvénile et adulte (de gauche à droite). (C) Schémas de coupes transversales de larve aposymbiotique, larve infectée et juvénile (de gauche à droite). b = cerveau, pr = protostomium, s1, s2 = segments avec cavités coelomiques, fg = tube digestif antérieur, mg = tube digestif médian, mo = bouche, or = région de l'obturbacule, ve = vestimentum, t = tronc, tr = trophosome, hg = tube digestif postérieur, te = tentacules, a = anus, v = mésoderme viscéral, dv = vaisseau sanguin dorsal, vv = vaisseau sanguin ventral, c = coelome, e = épiderme, m = muscles, vp = processus ventral avec ouverture de la bouche.

Après leur entrée dans la paroi du corps des larves, les symbiontes sont vacuolisés dans les cellules mésodermiques, initiant ainsi probablement le tout début de l'établissement de la symbiose (Nussbaumer *et al.* 2006) : le trophosome d'origine mésodermique (Figure 4C) commence à se développer au niveau du mésentère dorsal à proximité de l'intestin antérieur et du coelome (Figure 4B). Une fois le trophosome formé (stades juvéniles avancés), les tissus initialement infectés (paroi du corps, muscles, tissu mésodermique non différenciés et le tube digestif larvaire) se dégradent par apoptose : c'est la fin du processus de colonisation. Désormais, l'individu, dépourvu de bouche, tube digestif et anus, vit en symbiose stricte avec ses bactéries.

### 3. Anatomie de l'adulte symbiotique

#### 3.1 Le panache branchial

Le panache branchial de *Riftia pachyptila* (Figure 5A) est le principal organe en contact direct avec le milieu environnant et constitue de ce fait le système d'acquisition principal des métabolites.



**Figure 5 :** Morphologie externe et organisation de la branchie de *Riftia pachyptila* (d'après Andersen *et al.*, 2002). (A) Section parasagittale d'un petit individu. ABV = vaisseau sanguin afférent ; DV = vaisseau dorsal ; EBV = vaisseau sanguin efférent ; OV = vaisseau obturaculaire ; VV = vaisseau ventral. Echelle représentée = 1 mm. (B-D) Photographies de la branchie en microscopie électronique. (B) Section transversale de l'obturacule (Ob) entouré de lamelles branchiales (bl). Echelle représentée = 1 mm. (C) Lamelles branchiales formées par les filaments (f). Echelle représentée = 100 µm. (D) Pinnules. Echelle représentée = 50 µm.



Des mesures de la surface branchiale totale par unité de masse fraîche sur 16 individus ont résulté en une surface branchiale spécifique (SBSA) de  $22 \text{ cm}^2.\text{g}^{-1}$  par individu, ce qui représente une surface branchiale totale équivalente à 9 fois la surface totale du reste du corps du ver (Andersen *et al.*, 2002) ! L'obturacle est une structure rigide massive constituée principalement de collagène (Andersen *et al.*, 2001). Il soutient des lamelles branchiales qui sont elles-mêmes formées par de nombreux filaments branchiaux fusionnés sur les deux tiers de leur longueur et libres à leur extrémité distale (Figures 5B, C). Des pinnules de taille croissante (Figure 5D) se rencontrent jusqu'à l'extrémité des filaments augmentant encore la surface d'échange (Andersen *et al.*, 2002). Une particularité unique caractérise le genre *Riftia* : les lamelles branchiales s'insèrent tout le long de l'obturacle (disposition type Axonobranchia, au sens de Jones, 1985) alors que chez tous les autres Vestimentifères les lamelles s'insèrent à la base de l'obturacle (disposition Basibranchia).

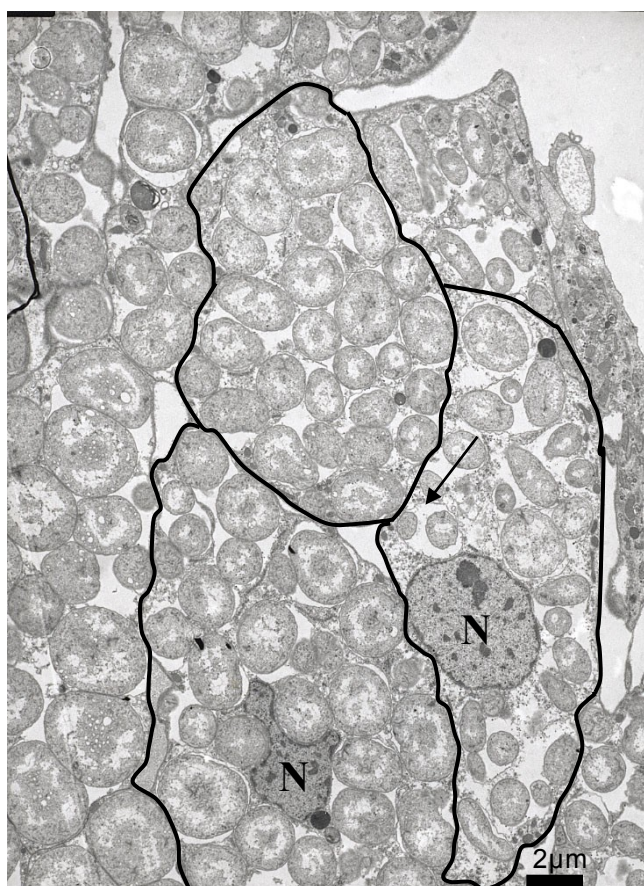
### 3.2 Le vestimentum

Situé à la base de l'obturacle (Figure 5A), le vestimentum est un organe musculeux en forme d'anneau ouvert dorsalement, formant deux ailes dorsales. Il caractérise les Vestimentifères en lieu et place du frenulum des Pogonophores. En se contractant, les muscles du vestimentum permettent au ver de se maintenir en entier dans son tube (panache branchial compris), celui-ci étant alors obturé par l'extrémité distale de l'obturacle. La contraction du vestimentum, lorsque l'animal déploie sa branchie dans le milieu, suggère également une très faible circulation d'eau dans le tube, restreignant *de facto* les échanges avec le milieu au seul panache branchial. Les individus mâles présentent sur la face dorsale, sous les ailes du vestimentum une forme en V correspondant à la jonction de deux gouttières ciliées qui aboutissent à des gonopores au niveau de la poche dorsale (Gardiner et Jones, 1993). Des glandes pyriformes réparties sur l'épiderme du vestimentum sécrètent de la chitine participant à la formation du tube du ver (Shillito *et al.*, 1993) sur laquelle nous reviendrons plus loin (voir partie 3.5). Du côté ventral du vestimentum se trouve une aire ciliée flanquée des deux branches du cordon nerveux épithélial, lequel se divise en deux uniquement au niveau du vestimentum. Cette aire ciliée sert vraisemblablement à créer un courant d'eau dans le tube à ce niveau, peut-être pour étaler la chitine nouvellement sécrétée par les glandes pyriformes, afin d'ajouter une nouvelle couche au tube.

### 3.3 Le tronc

Protégé par une épaisse paroi du corps constituée principalement d'une cuticule, d'un épiderme et de muscles, le tronc renferme le trophosome (Jones, 1981) et les gonades, étroitement imbriquées entre les lobules du trophosome, le tout baignant dans une vaste cavité coelomique remplie de liquide rouge.

Le trophosome représente environ 16 % du poids frais de l'adulte (Childress *et al.*, 1984). Cet organe est divisé en lobules ovoïdes contenant les cellules péritonéales (non symbiotiques) en périphérie, et les bactériocytes, les cellules de l'hôte qui hébergent les bactéries symbiotiques à l'intérieur de vacuoles (Figure 6). Il s'agit donc bien d'une endosymbiose (Chapitre I).



**Figure 6 :** Section transversale d'un lobule de trophosome vue en microscopie électronique (Cliché : A. Andersen). Les bactériocytes limités par leur membrane cytoplasmique (traits noirs) sont des cellules spécialisées de l'hôte renfermant des vacuoles contenant les bactéries. N = noyau. La flèche indique une vacuole qui contient deux bactéries.

Les bactéries occupent entre 15 et 35 % du volume total du trophosome avec une densité atteignant  $10^9$  cellules par gramme de tissu (Powell et Somero, 1986). Au sein de chaque lobule, différentes formes de bactéries sont reconnaissables : bactéries allongées au

centre, rondes et petites dans la zone médiane autour du vaisseau axial et rondes et grandes en zone périphérique (Bosch et Grassé, 1984a, 1984b ; Bright *et al.*, 2000 ; Bright et Sorgo, 2003). Du centre du lobule où elles sont en division, les bactéries semblent migrer jusqu'à la périphérie où elles apparaissent de plus en plus grosses et éventuellement lysées (Bosh et Grassé, 1984b). Chaque type morphologique bactérien contient des vésicules supposées participer au stockage des sulfures (Cavanaugh *et al.*, 1981 ; Hand, 1987 ; Wilmot et Vetter, 1990). Le niveau de sulfure peut atteindre 10 % du poids sec du trophosome (Fisher *et al.*, 1988 ; Vetter et Fry, 1998). La présence de sulfure influence même la couleur du trophosome qui varie de vert clair à presque noir : plus celui-ci est clair, plus il contient de sulfure (Pflugfelder *et al.*, 2005). A côté du trophosome se développent aussi les gonades, soit mâles soit femelles, reliées par des gonoductes aux gonopores situés sous les ailes du vestimentum. L'ensemble des organes du tronc baignent dans le liquide coelomique dont l'essentiel du volume se trouve au niveau du tronc.

Bien qu'en contact superficiel avec le milieu environnant, la paroi du corps n'a pas de rôle reconnu dans l'acquisition de métabolites depuis l'environnement chez *Riftia*. Il n'en est pas de même chez les vestimentifères des suintements froids qui possèdent des "racines", allongement postérieur du tronc avec un tube et une paroi du corps très minces à travers lesquels ils peuvent puiser les sulfures dans le sédiment (Freytag *et al.*, 2001).

### 3.4 L'opisthosome

A l'extrémité postérieure du tronc, l'opisthosome est une courte partie segmentée qui porte des soies en crochet permettant l'ancrage du ver au fond du tube (Rouse, 2001). Tout comme le vestimentum, l'opisthosome est riche en glandes pyriformes sécrétrices de chitine. Un modèle de croissance du tube par ses deux extrémités au cours d'une sorte de « mûre » avait été suggéré (Gaill *et al.*, 1997). Une activité chitinase au niveau de l'opisthosome (Ravaux *et al.*, 1998) appuie ainsi l'hypothèse d'une dissolution du tube pour synthétiser la base d'un nouveau tube.

### 3.5 Le tube

Le tube, de nature chitino-protéique (Gaill et Hunt, 1986), est construit par le ver lui-même grâce aux glandes sécrétrices de chitine disposées dans l'épiderme du vestimentum, du tronc et de l'opisthosome (Shillito *et al.*, 1993). Le tube de *Riftia* est quasiment iso-

diamétrique sur toute sa longueur, avec des septa vers la base, l'animal n'occupant que la partie supérieure du tube. Cette disposition caractéristique de *Riftia* (les autres genres ont un tube conique dont le diamètre s'élargit vers l'ouverture) implique un remaniement constant de l'ensemble du tube au cours de la croissance. Le tube peut en effet dépasser 1,5 mètre de hauteur pour un animal d'environ 75 cm (Jones, 1988) grâce à un taux de croissance exceptionnel, pouvant atteindre 85 cm par an (Lutz *et al.*, 1994), ce qui le place largement en tête de tous les taux rapportés chez les autres Polychètes hydrothermaux comme *Tevnia jerichonana* (plus de 30 cm par an) et plus généralement chez les organismes invertébrés marins (Lutz *et al.*, 1994). Complètement rétracté dans son tube, *Riftia pachyptila* est protégé d'éventuels prédateurs (comme le crabe *Bythograea thermydron* ou le poisson *Thermarces cerberus*) mais l'acquisition de gaz dissous est alors fortement réduite (Tunnicliffe *et al.*, 1990).

### 3.6 Communication entre les différents compartiments : le système circulatoire

Les cavités coelomiques du panache branchial, du vestimentum et du tronc ne sont pas connectées entre elles (Jones, 1985 ; Gardiner et Jones, 1993). Seul un système circulatoire clos permet une continuité d'un bout à l'autre de l'animal. Le vaisseau dorsal forme un épaississement (sorte de pompe contractile ou cœur) au niveau du vestimentum qui permet la propulsion du sang vers le panache branchial. Chaque filament branchial contient dans sa cavité coelomique un vaisseau afférent et un vaisseau efférent, qui une fois chargé de métabolites rejoint le vaisseau ventral. Au niveau de chaque lobule du trophosome, un réseau de vaisseaux afférents irrigue la périphérie puis le flux de sang artériel parvient au centre de chaque lobule au niveau d'un vaisseau axial efférent qui rejoint ensuite le vaisseau dorsal et peut retourner à la branchie (Jones, 1988 ; Felbeck et Turner, 1995 ; Bright *et al.*, 2000).

## **4. Les voies métaboliques de la symbiose**

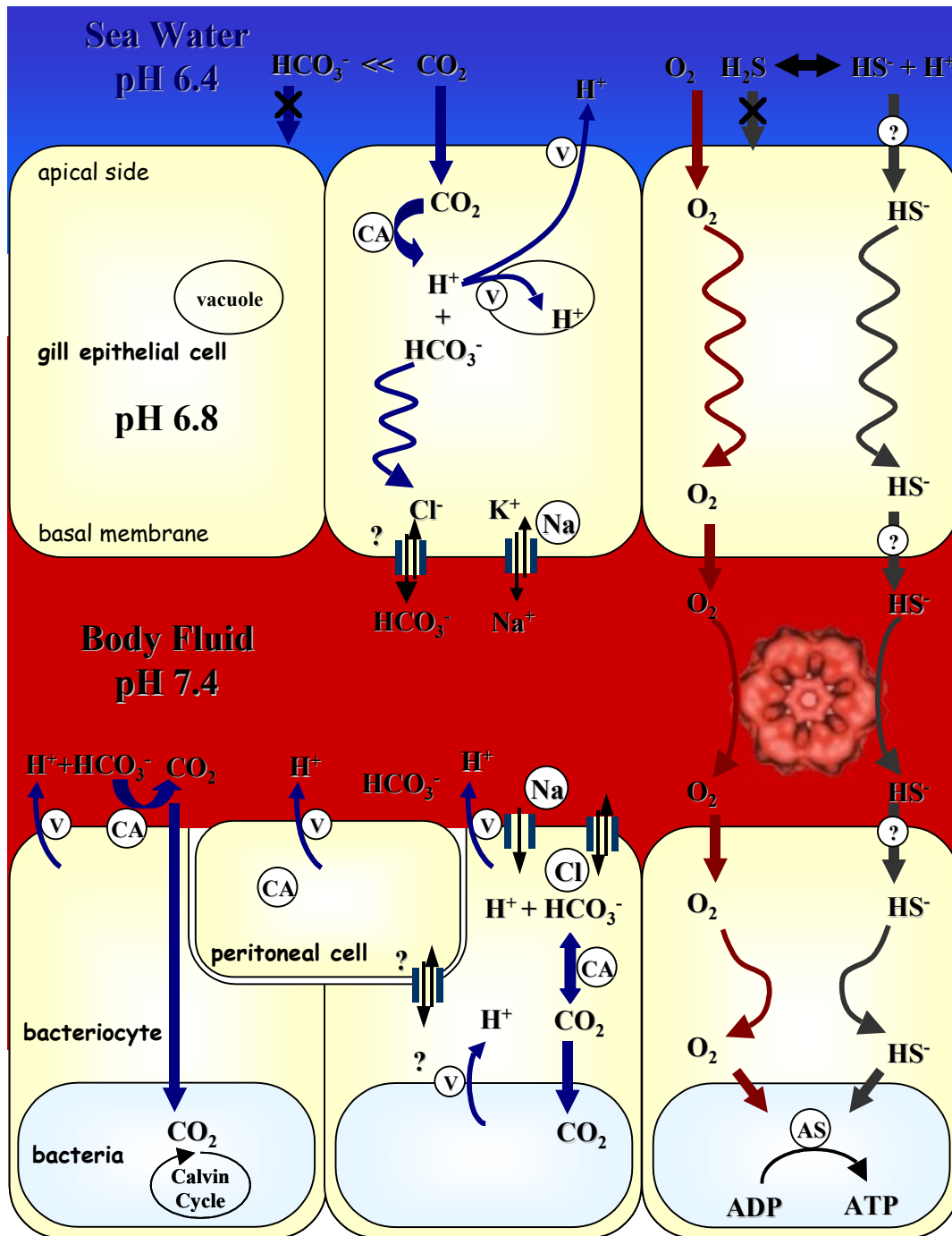
### 4.1 Contribution de l'hôte : acquisition et transport des métabolites

Comme nous l'avons introduit au cours du Chapitre I, l'hôte *Riftia pachyptila* constitue l'interface entre les bactéries et le milieu extérieur, étant donnée la position internalisée des symbiontes. Les métabolites nécessaires à son propre métabolisme et à celui

de ses bactéries sont prélevés via le panache branchial par diffusion simple ( $O_2$ ,  $CO_2$ ) ou facilitée (pour l' $H_2S$  et les ions nitrates ( $NO_3^-$ )) par des mécanismes de transport membranaires de nature encore inconnue à ce jour.

#### 4.1.1 Le dioxyde de carbone ( $CO_2$ )

Présent à des concentrations totales variant entre 3 et 5 mM dans l'environnement (Childress et Fisher, 1992 ; Childress *et al.*, 1993 ; Scott *et al.*, 1994), le  $CO_2$  est majoritairement sous forme de gaz dissous du fait du pH acide (5-6) qui règne dans la zone de mélange. Il entre par diffusion dans les cellules épithéliales branchiales (Figure 7) (Goffredi *et al.*, 1997a). Présent à une concentration de 20 à 46 mM dans le sang (Childress *et al.*, 1993), une partie du  $CO_2$  pourrait être transformée en malate puis transportée par voie sanguine vers le trophosome, une décarboxylation ayant lieu à proximité des bactériocytes (Felbeck, 1985). Mais la majorité du  $CO_2$  est sans doute convertie localement en ions bicarbonate ( $HCO_3^-$ ) par une enzyme, l'anhydrase carbonique (CA) (Figure 7). Le gradient de diffusion du  $CO_2$  est ainsi entretenu, notamment grâce à l'excrétion de protons par des ATPases qui permettent de maintenir un pH physiologique alcalin dans les liquides corporels de l'animal (Goffredi *et al.*, 1997a ; Goffredi et Childress, 2001). Les ions  $HCO_3^-$  sont ensuite acheminés par le système sanguin jusqu'aux bactériocytes à proximité desquels ils doivent être retransformés en  $CO_2$ , seule espèce chimique pouvant être incorporée dans le cycle de Calvin – Benson ou le cycle TCA (tricarboxylic acid) inverse de la bactérie. Celle-ci est en effet dépourvue de carboxysome et ne peut donc utiliser le carbone inorganique sous forme de bicarbonate (Scott *et al.*, 1999). Cette conversion fait intervenir à nouveau une anhydrase carbonique, celle-ci pouvant être cytosolique ou bien liée à la membrane des bactériocytes (De Cian *et al.*, 2003a). Des activités d'anhydrase carbonique dans la branchie et à la frontière des bactériocytes dans le trophosome de *Riftia pachyptila* ont pu être mises en évidence (Kochevar *et al.*, 1993 ; Goffredi *et al.*, 1999a ; De Cian *et al.*, 2003a) et la localisation des protéines par immunomarquage a également révélé la présence d'anhydrases carboniques dans ces deux organes (De Cian *et al.*, 2003b). Une séquence d'anhydrase carbonique a été obtenue à partir d'échantillon d'ADN complémentaire (ADNc) de trophosome par une étude moléculaire (De Cian *et al.*, 2003a).



**Figure 7 :** Schéma résumant les modes d'acquisition de l' $\text{O}_2$ ,  $\text{CO}_2$  et  $\text{H}_2\text{S}$  à travers l'épithélium branchial et leur transport jusqu'aux bactériocytes (par Y. Ferandin d'après la thèse de M.-C. De Cian, 2003).

#### 4.1.2 Le sulfure d'hydrogène ( $H_2S$ ) et l'oxygène ( $O_2$ )

Présent à des concentrations modérées dans le milieu extérieur immédiat de *Riftia* (moins de 300  $\mu M$ , Childress *et al.*, 1991 ; Goffredi *et al.*, 1997b), les sulfures y sont principalement sous forme d' $H_2S$  moléculaire. Pourtant, ce sont les ions  $HS^-$  qui sont acquis en grande majorité au niveau des cellules épithéliales branchiales (Goffredi *et al.*, 1997b ; Goffredi *et al.*, 1999b) par un transport actif de nature inconnue (Figure 7). Fortement toxiques, notamment pour les enzymes de la chaîne respiratoire mitochondriale, les ions  $HS^-$  sont neutralisés par leur fixation à des molécules d'hémoglobine (Hb) circulant dans le sang et le liquide coelomique. *Riftia pachyptila* possède trois types d'hémoglobines extracellulaires (Zal *et al.*, 1996), une coelomique et deux vasculaires, toutes trois capables de fixer réversiblement  $H_2S$  et  $O_2$  sur deux sites distincts (Arp *et al.* 1987; Zal *et al.*, 1998). L'affinité des Hb coelomique et vasculaires pour l' $O_2$  est relativement élevée (Arp et Childress, 1981) mais seules les Hb vasculaires montrent un effet Bohr (Arp *et al.*, 1985) qui permet de relarguer l' $O_2$  au niveau des bactériocytes. L'une de ces Hb est constituée de 144 chaînes de globines organisées en bicouches hexagonales (hexagonal-bilayer hemoglobin, HBL-Hb, Terwilliger *et al.*, 1980). Cette Hb géante (3,6 MDa), connue uniquement chez les Annélides (dont les Siboglinidae hydrothermaux) (Weber et Vinogradov, 2001) pourrait fixer l' $H_2S$  au moyen de résidus cystéines libres (Zal *et al.*, 1998) ou de zincs (Flores *et al.*, 2005 ; Flores and Hourdez, 2006).

#### 4.1.3 Composés azotés

L'azote organique dissous n'est probablement pas une source suffisante pour le ver étant données les faibles concentrations rapportées (moins de 0,2 nM d'acides aminés autour des communautés hydrothermales (Johnson *et al.*, 1988). *R. pachyptila* prélève plutôt les ions  $NO_3^-$  au niveau du panache branchial à un taux de 3,54  $\mu mol.g^{-1}.h^{-1}$  (Lee et Childress, 1994) ; les ions  $NO_3^-$  constituant la principale source d'azote dans le milieu environnant (concentration à environ 40  $\mu M$ , Girguis *et al.*, 2000). Le transport des nitrates vers les bactéries s'effectue via le système circulatoire où l'hémoglobine joue semble-t-il une fois encore le rôle de transporteur (Hahlbeck *et al.*, 2005).

#### 4.1.4 L'hôte : milieu de vie pour les bactéries

Si l'on considère que le milieu hydrothermal est un environnement changeant au niveau de la disponibilité des différents métabolites (Chapitre I), le développement des symbiontes dans le milieu de vie « tamponné » que constitue leur hôte leur confère un avantage certain : l'hôte est notamment capable de stocker les métabolites, leur fournissant ainsi à tout moment ce dont ils ont besoin. Ainsi, de fortes concentrations d'Hb dans le sang et le liquide coelomique permettent de stocker O<sub>2</sub> et H<sub>2</sub>S pour faire face à des conditions plus contraignantes, détaillées dans la partie 4.4. Le sang, et dans une mesure plus importante le liquide coelomique, concentrent de grandes quantités de CO<sub>2</sub> (Toulmond *et al.*, 1994). En outre, on constate la présence d'hypotaurine chez *R. pachyptila*, et en particulier des concentrations importantes de thiotaurine dans le trophosome (Brand *et al.*, 2007), composé aminé capable de stocker et éventuellement transporter les sulfures pour la consommation du symbionte. La forte affinité des Hbs pour l'oxygène et les sulfures permet d'avoir de fortes concentrations de ces composés disponibles, tout en en gardant la quantité libre minimale (Fisher *et al.*, 1988). Ceci permet de garantir un fort taux d'autotrophie des symbiontes qui sont inhibés par de trop fortes concentrations de sulfures libres.

### 4.2 Contribution du symbionte

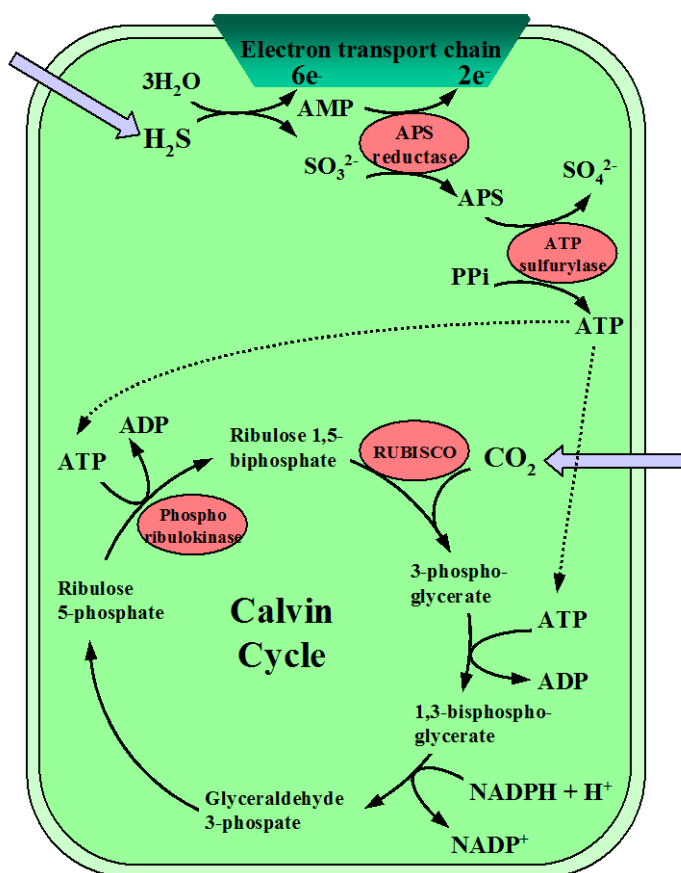
Une fois parvenus aux bactériocytes, O<sub>2</sub>, CO<sub>2</sub>, H<sub>2</sub>S et NO<sub>3</sub><sup>-</sup> (comme source d'azote, mais aussi comme accepteur final d'électrons alternatif à l'O<sub>2</sub> (Hentschel et Felbeck, 1993)) peuvent être directement utilisés par les symbiontes.

#### 4.2.1 Sulfo-oxydation et assimilation du CO<sub>2</sub>

L'H<sub>2</sub>S est oxydé en sulfite (SO<sub>3</sub><sup>2-</sup>) par un système de transport d'électron impliquant des cytochromes dans le compartiment bactérien (Figure 8) : cette oxydation permet de générer l'énergie nécessaire à l'assimilation du carbone inorganique. Les bactéries symbiotiques de *R. pachyptila* possèdent la forme II de la ribulose-1,5-biphosphate carboxylase (Rubisco) (Williams *et al.*, 1988) et acceptent uniquement l'entrée de l'espèce CO<sub>2</sub> dans le cycle de Calvin-Benson (Scott *et al.*, 1999).



**Figure 8 :** Schéma illustrant la connexion entre l'oxydation de l' $\text{H}_2\text{S}$  pour la production d'énergie et le cycle de Calvin-Benson dans une bactérie sulfo-oxydante (par Y. Ferandin d'après Minic et Hervé, 2004).



En outre, il a récemment été montré que les symbiontes de *R. pachyptila* possèdent les enzymes du cycle TCA (Markert *et al.*, 2007) qui permet aussi l'incorporation du carbone inorganique sous forme de  $\text{CO}_2$  (cycle TCA réducteur). Que ce soit par l'un ou l'autre de ces cycles, le  $\text{CO}_2$  peut alors être transformé en sucres à 3 atomes de carbone, et donner des composés organiques (succinate et glutamate). Des autoradiographies au carbone marqué ( $^{14}\text{C}$ ) ont montré qu'immédiatement après la fixation du  $\text{CO}_2$ , les symbiontes rejettent une quantité significative de carbone organique qui peut alors être assimilé par l'hôte (Bright *et al.*, 2000). Par ailleurs, comme nous l'avons indiqué au cours du Chapitre I, des stades de dégradation et des structures ressemblant à des stades de lysosomes sont observés au niveau des lobules du trophosome (Bosh et Grassé, 1984a, b ; Gardiner et Jones, 1993 ; Bright *et al.*, 2000) : cet autre mécanisme impliquant la lyse des bactéries par l'hôte en périphérie des lobules pourrait contribuer à « nourrir » directement l'hôte.

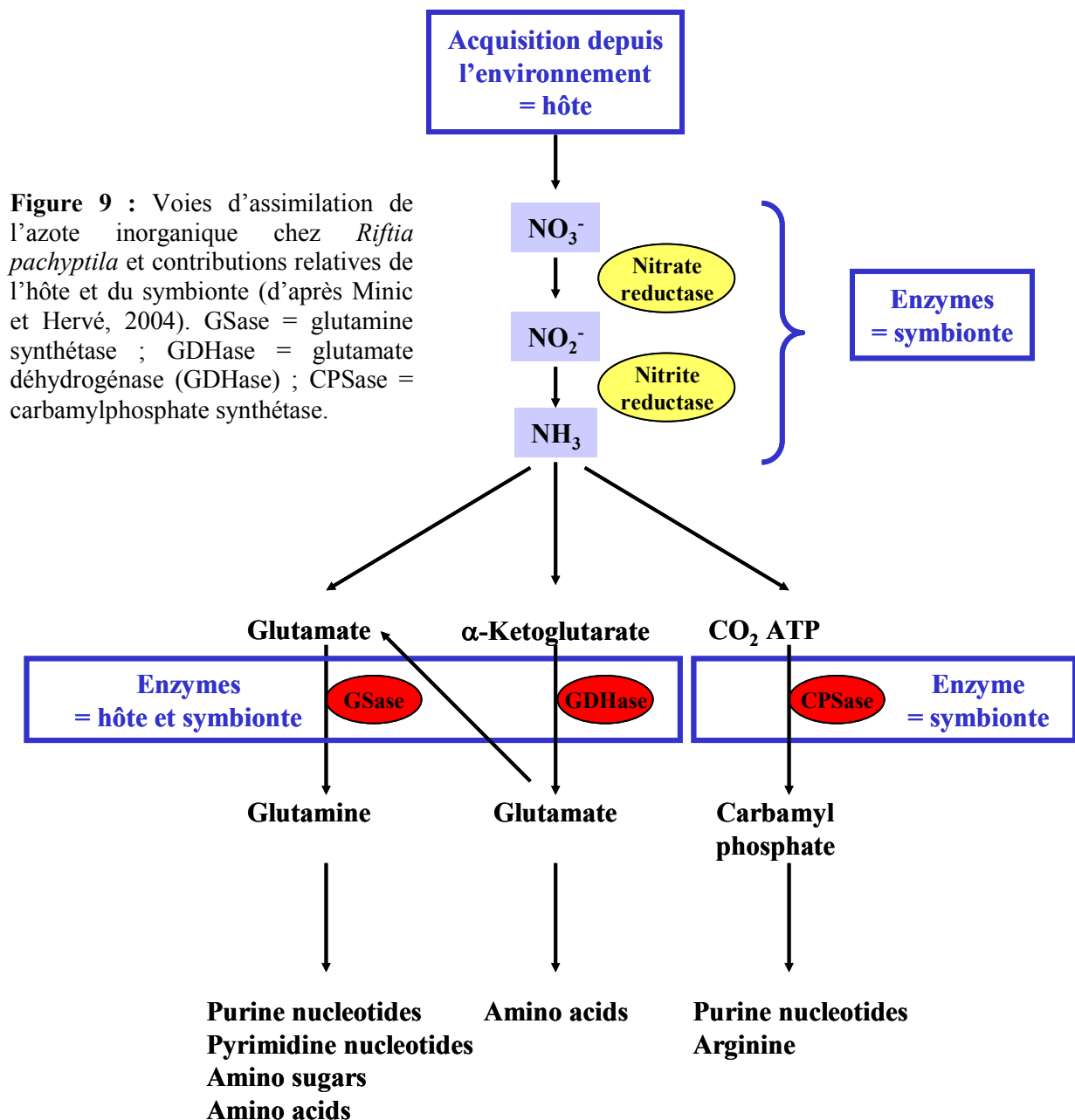
#### 4.2.2 Transformation du $\text{NO}_3^-$

La transformation des ions  $\text{NO}_3^-$  en nitrite ( $\text{NO}_2^-$ ) puis en ammoniac ( $\text{NH}_3$ ) est exclusivement réalisée par les symbiontes (Figure 9). Le  $\text{NH}_3$  est ensuite traité par différentes

enzymes (glutamine synthétase GSase, glutamate déhydrogénase GDHase et carbamylphosphate synthétase CPSase) pour produire des métabolites de base tels qu'acides aminés et nucléotides (voir Minic and Hervé, 2004 ; et ci-dessous).

#### 4.3 Contribution de l'hôte et du symbionte

Le métabolisme des pyrimidines est le résultat d'une étroite interaction des contributions de l'hôte et du symbionte (Figure 9). Ainsi, certaines réactions d'assimilation de  $\text{NH}_3$  peuvent aussi bien être réalisées par le symbionte que par *Riftia* : les enzymes GSase (une forme typique du ver et une forme typique du symbionte) et GDHase ont pu être détectées chez les deux organismes (Lee *et al.*, 1999 ; Minic *et al.*, 2001).



#### 4.3.1 Voie de synthèse de novo des pyrimidines

La voie *de novo* permet la synthèse complète des nucléotides pyrimidiques à partir de  $\text{HCO}_3^-$ , glutamine et ATP: les 3 premières enzymes de ce réseau (CPSase, aspartate transcarbamylase et dihydroorotase) sont exclusivement d'origine bactérienne. Les enzymes suivantes (dihydroorotate déhydrogénase, orotatephosphoribosyl transférase et cytidine triphosphate (CTP) synthase) sont présentes dans tous les organes de l'hôte et chez la bactérie. Si le symbionte peut donc se reposer entièrement sur ce cycle tandis que l'hôte non, la tendance s'inverse au sujet du réseau de récupération issu de la dégradation des acides nucléiques.

#### 4.3.2 Voie de dégradation des pyrimidines

Dans ce réseau impliquant les enzymes cytidine déaminase, uridine kinase et uracylphosphoribosyl transférase, des analyses biochimiques ont montré qu'à la différence de l'hôte, la bactérie isolée ne possède pas d'activité de ces enzymes (Minic *et al.*, 2002).

L'hôte peut également dégrader les acides nucléiques par un autre cycle pour « recycler » les produits de dégradation et libérer du  $\text{CO}_2$ ,  $\text{NH}_3$ , malonyl-CoA (utilisé pour la biosynthèse des acides gras) et succinyl-CoA (pouvant entrer dans le cycle de l'acide citrique). Les enzymes impliquées (5'-nucléotidase, uracil réductase et uridine phosphorylase) appartiennent exclusivement au ver (Minic *et al.*, 2002). Cette dégradation peut aussi résulter de la lyse bactérienne dans le trophosome (Bright *et al.*, 2000).

#### 4.3.3 Métabolisme de l'arginine

Le métabolisme de l'arginine est initié par une carbamylphosphate synthétase (enzyme différente de celle du métabolisme des pyrimidines) qui utilise  $\text{NH}_3$  en tant que substrat au lieu de la glutamine. Cette enzyme et les 2 enzymes suivantes (ornithine transcarbamylase et argininosuccinate synthétase) sont présentes à la fois chez l'hôte et chez la bactérie. L'arginine décarboxylase qui sert pourtant à la synthèse de polyamines, constituants jouant un rôle important dans la biosynthèse et la structure des acides nucléiques, est présente seulement chez le symbionte d'où une dépendance forte de l'hôte vis-à-vis de la bactérie. La production résultante d'agmatine et de putrescine devrait être suivie d'un transport de ces composés vers

tous les tissus de l'hôte comme cela a pu être décrit chez d'autres organismes (Kandpal et Tekwani, 1997 ; Satriano *et al.*, 2001). La dégradation des polyamines peut aussi apporter une autre source de carbone et d'azote pour le ver (Nakada *et al.*, 2001 ; Mercenier *et al.*, 1980).

### 4.4 Influence de la disponibilité des métabolites et limites du fonctionnement de l'association symbiotique

Le comportement de *Riftia pachyptila in situ* vis-à-vis de l'acquisition des métabolites est fortement influencé par les conditions environnementales (comme la position de l'individu par rapport à la source d'émission d'H<sub>2</sub>S) et par les concentrations fluctuantes de métabolites (comme la concentration en O<sub>2</sub>) requis pour son métabolisme. Par ailleurs, une destruction partielle du panache branchial par les prédateurs peut diminuer la capacité des échanges de métabolites et induire par conséquent une incapacité à fournir les éléments nécessaires aux symbiontes. S'il persiste longtemps, ce type d'événement peut finir par provoquer la mort du ver et de ses symbiontes (Tunnicliffe *et al.*, 1990).

#### 4.4.1 Autotrophie

Des expériences menées en aquarium pressurisé (120 atm) ont été réalisées par Childress *et al.*, (1991) pour observer les flux de métabolites chez le ver. En l'absence de sulfures, le ver consomme ainsi de l'O<sub>2</sub> et rejette du CO<sub>2</sub> : *R. pachyptila* a besoin d'O<sub>2</sub> pour maintenir son métabolisme aérobie (Arndt *et al.*, 1998). En présence de sulfures, une augmentation importante de la consommation d'O<sub>2</sub> est observée, correspondant aux besoins du symbionte pour oxyder l'H<sub>2</sub>S (Hentschel et Felbeck, 1993). Par ailleurs, à partir d'une consommation de HS<sup>-</sup> de 2,5 µmol. g<sup>-1</sup>. h<sup>-1</sup>, *R. pachyptila* se met à consommer plus de CO<sub>2</sub> qu'il n'en rejette : ceci caractérise l'autotrophie, avec un flux net entrant de CO<sub>2</sub> (Childress *et al.*, 1991).

#### 4.4.2 Conditions hypoxiques

Suivant la teneur en O<sub>2</sub> du milieu ou selon la durée de rétractation du ver à l'intérieur de son tube, le ver peut survivre au moins 36 heures (Childress *et al.*, 1984) et jusqu'à 60 heures (Arndt *et al.*, 1998) en condition anaérobie. Malgré des activités de lactate déshydrogénase mesurées dans la paroi du corps et le vestimentum (3 et

4  $\mu\text{mol.g}^{-1}.\text{f wt. min}^{-1}$ ), le lactate ne s'accumule pas lors de périodes hypoxiques prolongées (Arndt *et al.*, 1998). Le principal substrat utilisé par le ver dans de telles circonstances est en fait le glycogène, principalement présent dans le trophosome. A partir de la transformation du glycogène en malate, le seul produit du métabolisme anaérobie formé par *Riftia pachyptila* est le succinate qui s'accumule à de fortes concentrations dans les liquides corporels (jusqu'à 46 mM) et dans les tissus (23  $\mu\text{mol. g}^{-1} \text{ f wt}$ ) (Arndt *et al.*, 1998). Pendant l'hypoxie, les symbiontes bactériens peuvent utiliser les ions  $\text{NO}_3^-$  à la place de l' $\text{O}_2$ , ce qui leur confère un avantage pour continuer à oxyder l' $\text{H}_2\text{S}$  (Arndt *et al.*, 1998). Cependant, le symbionte pourrait également adopter un mode de vie hétérotrophe en utilisant les réserves de glycogène (par glycolyse) puis le cycle TCA oxydatif pour générer de l'énergie et des composés de biosynthèse (Markert *et al.*, 2007). En période d'hypoxie, l'acquisition du  $\text{CO}_2$  apparaît limitée (Arndt *et al.*, 1998) ce qui diminue le métabolisme autotrophe du symbionte : une baisse de l'acide aminé glutamate excrété après fixation du  $\text{CO}_2$  par les bactéries est alors observée (Felbeck et Jarchow, 1998).

#### 4.5 Questions en suspens

*Riftia pachyptila* est certainement l'un des organismes hydrothermaux les plus étudiés depuis sa découverte il y a 30 ans (tout juste !) : la banque de données Web of Science répertorie près de 400 références bibliographiques répondant au mot clé "Riftia". Néanmoins, de nombreuses questions subsistent pour atteindre une meilleure compréhension des mécanismes de la relation symbiotique entre l'hôte et les bactéries, ne serait-ce que sur le plan du métabolisme énergétique. Les flux entrants de  $\text{CO}_2$ ,  $\text{HS}^-$  ou  $\text{NO}_3^-$  nécessaires au métabolisme des bactéries sont tout à fait inhabituels pour un métazoaire. Il n'est donc pas étonnant que les mécanismes cellulaires permettant l'acquisition de ces molécules soient inconnus à ce jour. De grands flux de bicarbonate sont aussi nécessaires entre branchie et sang, et entre sang et bactériocytes, pour alimenter les bactéries en carbone inorganique, des flux dont l'intensité est comparable à celle que l'on peut mesurer pour un globule rouge de Mammifère et sans commune mesure avec les flux de  $\text{CO}_2$  observés chez des animaux aquatiques. De même, l'activité métabolique des bactéries génère un flux important de protons et de sulfates que son hôte doit pouvoir excréter. Au sein même du trophosome, quels sont les types d'échanges qui existent entre cellules péritonéales et bactériocytes, et entre bactériocytes et bactéries, notamment au niveau de la membrane vacuolaire, dernière frontière séparant l'hôte de ses bactéries ?

## 6. Objectif du travail de thèse

Confrontés aux limites des approches traditionnelles de physiologie qui visent à tester un modèle présumé, et surtout lorsque l'expérimentation est aussi difficile et parcimonieuse qu'avec un animal des sources hydrothermales profondes, nous avons voulu appréhender cette symbiose avec une approche globale au moyen d'outils moléculaires.

Le symbionte de *Riftia pachyptila* a été étudié par une étude génomique, actuellement en cours de publication, par l'équipe du Pr. H. Felbeck. Ces données ont par ailleurs été généreusement rendues accessibles au cours de ma thèse.

En ce qui concerne l'hôte *R. pachyptila*, des approches moléculaires (par séquençage d'ADN et d'ADN complémentaire (ADNc)) ou de spectrométrie de masse (avec séquençage des acides aminés) ont été précédemment réalisées mais sur des cibles particulières. Cela a notamment permis d'identifier des séquences des chaînes de globines participant à la formation des Hb coelomique ou vasculaires. Six chaînes ont ainsi été séquencées : chaîne A2, (Zal *et al.*, 1996a, 1996b, 1997), et chaînes A1, B1 et B2 (Bailly *et al.*, 2002). De même, des séquences d'ADNc d'une anhydrase carbonique (De Cian *et al.*, 2003a), de myosine (Ravaux *et al.*, 2001), de collagène fibrillaire (Sicot *et al.*, 2000) et de protéine de liaison à l'exosquelette (Chamoy *et al.*, 2001) ont été obtenues. Le génome mitochondrial de *Riftia pachyptila* est également séquencé à 80 % (Jennings et Halanych, 2005). Mais cette contribution reste encore infime à l'heure où de vastes projets de séquençage incluant notamment des générations de banques d'ADNc telles que les étiquettes de séquences transcrites (Expressed Sequences Tag, EST) permettent d'identifier des séquences par homologie avec des banques de données et peuvent être un moyen d'extrapoler des réseaux métaboliques. Nous avons décidé de mener une étude exploratoire du transcriptome de l'hôte. Celui-ci étant l'interface indispensable entre ses symbiontes et le milieu extérieur, de nombreuses clés pour comprendre l'interaction physiologique entre les deux organismes sont certainement détenues par les séquences exprimées par l'hôte lui-même. C'est pourquoi nous avons décidé de générer des banques d'ADNc chez *R. pachyptila*, en gardant à l'esprit quelques considérations...

Le phénotype d'un organisme est déterminé par l'expression des gènes. La difficulté de trouver un ARN messager (ARNm) responsable d'une fonction particulière dans un

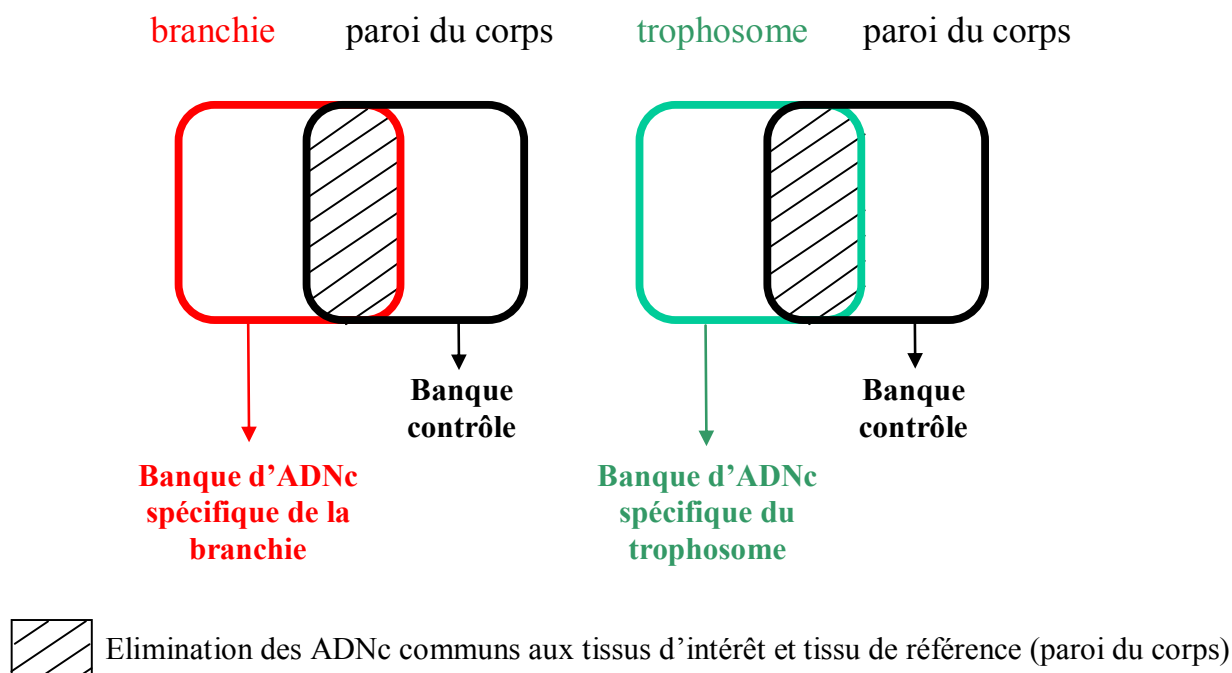
échantillon donné n'est pas étonnante si on considère que ce gène n'est peut être pas exprimé de façon abondante. En effet, il faut garder en tête les proportions relatives d'ARN messagers au sein d'une cellule donnée : sur un total de 30 000 ARNm différents, environ 99 % d'entre eux sont rares (20 000 à 70 000 fois moins abondants par rapport aux plus abondants). Ces 99 % d'ARNm ne représentent en fait que la moitié de la masse totale des ARNm dans la cellule, l'autre moitié étant représentée par les 1 % d'ARNm abondants (Wan *et al.*, 1996). Des méthodes ont alors été mises au point pour identifier les ARNm dans une population d'ARNm indépendamment de leur niveau de présence dans la population. Parmi ces méthodes, les plus utilisées ces dernières années restent l'« affichage différentiel » (Differential Display, DD, Liang et Pardee, 1992) et l'hybridation soustractive suppressive (Subtractive Suppression Hybridization, SSH, Diatchenko *et al.*, 1996 ; Gurskaya *et al.*, 1996).

La méthode SSH a été beaucoup utilisée ces dernières années pour l'étude des symbioses entre des plantes de la famille des Légumineuses et les rhizobiums (bactéries aérobies du sol du genre *Rhizobium*), de loin les symbioses les plus étudiées. La formation de nodosités, véritables organes d'échange entre les bactéries et les plantes au niveau des racines ou de la tige, est induite par les bactéries. Dans le but d'isoler des séquences liées au phénomène de nodulation chez le soja, l'utilisation de la méthode SSH entre des ADNc de nodule et des ADNc de racine non infectée a permis d'identifier des séquences spécifiquement exprimées dans le nodule (Kouchi et Hata, 1993).

Dans le cadre de notre étude, notre hypothèse de travail repose sur le principe que les ARN messagers codant les protéines impliquées dans la relation symbiotique chez *Riftia pachyptila* devraient être préférentiellement présents dans les tissus directement impliqués dans le mode de vie symbiotique. Comme nous l'avons décrit plus haut, deux organes principaux jouent un rôle prépondérant dans l'association symbiotique : la branchie, organe d'échanges avec le milieu environnant, et le trophosome qui héberge les symbiontes. L'approche idéale aurait consisté en la comparaison des séquences exprimées dans ces tissus chez des individus aposymbiotiques (en l'occurrence des larves chez *R. pachyptila*) et des séquences exprimées chez des individus strictement symbiotiques (des adultes). Ce type d'approche a notamment été mené par Rodriguez-Lanetty *et al.* (1996) sur le cnidaire *Anthopleura elegantissima* qui vit en symbiose avec une algue verte unicellulaire (le dinoflagellé *Symbodinium muscatinei*). Une telle approche s'avère impossible dans notre cas,

étant données la taille des larves et l'impossibilité de les prélever en quantité appréciable lors des missions.

Plutôt que de générer des banques d'EST classiques qui auraient certes apporté des résultats nouveaux mais demandé un effort de séquençage plus conséquent, nous avons préféré utiliser la méthode SSH dans une approche originale (Figure 10) pour accéder rapidement à des séquences spécifiques de chacun des deux tissus.



**Figure 10 :** Explication de l'approche envisagée pour la thèse. La construction de banques soustractives enrichies en séquences spécifiques de la branchie d'une part, et en séquences spécifiques du trophosome d'autre part, sera réalisée grâce à l'utilisation de la paroi du corps comme tissu de référence. La méthode SSH sera appliquée au couple branchie-paroi du corps d'une part, et au couple trophosome-paroi du corps d'autre part. La construction des banques « réciproques » sera également réalisée, ces banques serviront de contrôles destinés à vérifier l'efficacité de la méthodologie employée.

Cette technique a été utilisée pour identifier les séquences spécifiques de la branchie d'une part, et du trophosome d'autre part, lorsque chacun des tissus est comparé à un tissu de référence, *a priori* non impliqué dans les échanges de métabolites, la paroi du corps. En éliminant les séquences communes entre branchie et paroi du corps d'une part, et entre trophosome et paroi du corps d'autre part, les banques d'ADNc devraient être respectivement enrichies en séquences spécifiques de branchie et de trophosome (action de « soustraction » de la technique SSH, Chapitre III). Une liaison entre les activités de protéines connues chez *Riftia* et l'existence des transcrits de ces protéines serait dans un premier temps un résultat



intéressant. Enfin, grâce à l'action de « normalisation » de la technique SSH qui permet d'enrichir la proportion des ARNm les plus rares par rapport aux ARNm les plus abondants (Chapitre III), on peut également espérer trouver de nouvelles séquences qui avaient jusque-là échappé aux méthodes traditionnelles.

# **CHAPITRE III**

## **Matériel et Méthodes d'étude**

## CHAPITRE III : Matériel et Méthodes d'étude

### 1. Echantillonnage des spécimens de *Riftia pachyptila*

Les différents individus ont été récoltés sur les sites Rehu Marka (17°25'S – 113°12'W), Susie et Miss WormWood (17°35'S - 113°14'W) de la dorsale Est-Pacifique (EPR) à environ 2 600 mètres de profondeur, au cours de la campagne océanographique BIOSPEEDO (mars-avril 2004, chef de mission D. Jollivet ; plongées Nautila 1587, 1588 & 1590). Prélevés à l'aide des bras télémanipulés du Nautila, les individus ont été placés dans le panier isotherme du sous-marin et remontés à l'issue de la plongée. Les spécimens ont ensuite été disséqués à bord (F. H. Lallier et S. Hourdez). Pour chaque individu, un morceau de chaque type de tissu (panache branchial, trophosome et paroi du corps) a été isolé et conservé à 4°C pendant 24 heures dans du RNAlater<sup>®</sup> (Ambion), puis stocké dans de l'azote liquide jusqu'à l'utilisation en laboratoire (début de la thèse).

### 2. Extraction d'ARN total

#### 2.1 Précautions préliminaires

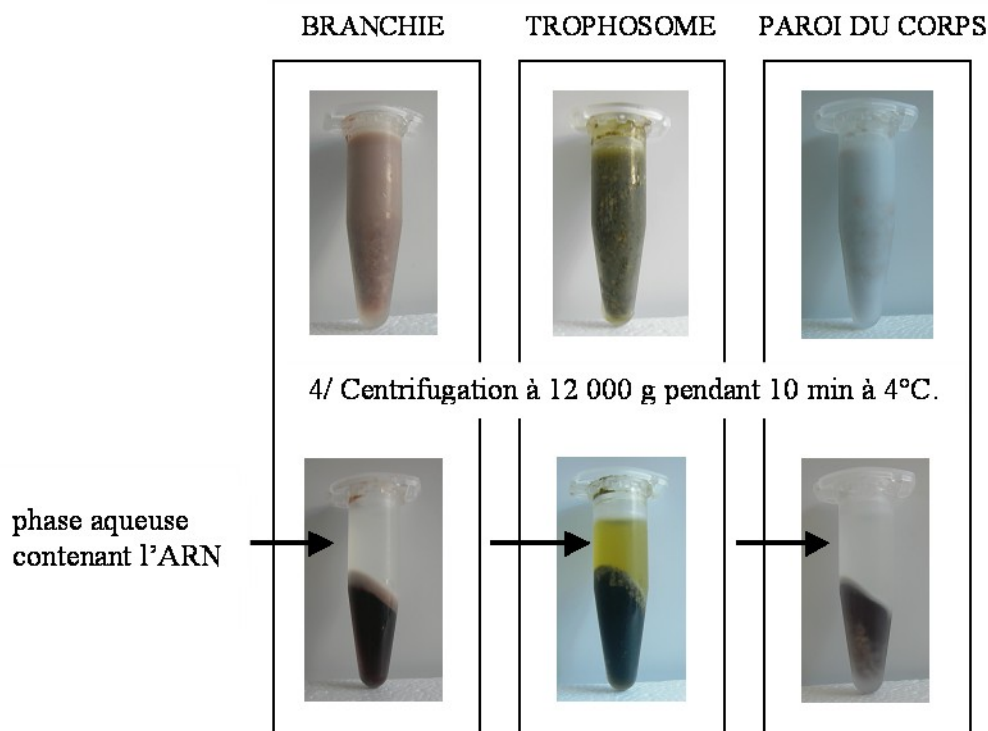
Toute expérimentation sur de l'ARN est sujette à des conditions stringentes. Le matériel utilisé (pinces, mortier) et le plan de travail doivent être exempts de toute ribonucléase par un nettoyage minutieux à l'eau stérile traitée au diéthylpyrocarbonate (DEPC), puis au RNaseZap<sup>®</sup> (Ambion) qui élimine les ribonucléases. Le matériel est ensuite « plongé » pendant quelques minutes dans l'azote liquide (-196°C) pour assurer un environnement froid lors de l'extraction (Figure 1) et éviter la dégradation des ARN. Toutes les expérimentations ultérieures sont effectuées sur glace.



1/ Morceau de tissu réduit en poudre dans de l'azote liquide.  
(Clichés : broyat de tissu branchial.)

2/ Mélanger vigoureusement 200 mg de broyat à 1 ml de RNAb.

3/ Sur glace, ajouter 100  $\mu$ l de chloroforme. Mélange vigoureux pendant 15 à 20 secondes. Repos sur glace pendant 5 min.



5/ Prélever la phase aqueuse et ajouter 2/3 d'isopropanol. Mélanger. Laisser 10 minutes à température ambiante.

6/ Centrifugation pendant 15 min à 12 000 g. L'ARN précipite sous forme de culot blanc-jaunâtre au fond du tube.

7/ Lavage du culot à l'éthanol 75% puis séchage du tube à l'envers sur la paille pendant une dizaine de minutes.

8/ Reprise du culot dans 50 à 100  $\mu$ l d'eau stérile traitée au DEPC.

**Figure 1 :** Protocole d'extraction d'ARN total utilisé pour chaque type de tissu.

## 2.2 Méthode au RNABle

L'ARN a été extrait en utilisant la solution RNABle (Eurobio). Les morceaux de tissus ont été broyés et réduits en poudre dans de l'azote liquide. Pour chaque échantillon, nous avons suivi le protocole présenté en Figure 1. Le RNABle contient un irritant (l'isothiocyanate de guanidine) et un poison (le phénol). Cette méthode d'extraction repose sur la formation de complexes entre l'ARN et l'isothiocyanate de guanidine qui se retrouvent en phase aqueuse, isolés de l'ADN et des protéines (phase phénolique inférieure). Une fois les culots d'ARN repris dans l'eau, on fait migrer 2 µl en gel d'agarose 1,5 % pour s'assurer de l'absence de dégradation d'ARN avant toute expérimentation ultérieure. Ces échantillons sont ensuite dosés au spectrophotomètre à 260 nm avant la purification d'ARN messager (partie suivante) pour constituer les banques, ou avant la synthèse d'ADN complémentaire en préambule aux expérimentations de PCR quantitative.

## **3. Banques soustractives d'ADN complémentaire**

### 3.1 Purification d'ARN messager

Nous avons utilisé les colonnes en résine Oligo (dT) (Amersham Biosciences) et suivi le protocole du fournisseur pour isoler les ARN messagers. Les ARN messagers eucaryotes possèdent en effet une longue queue poly-A (80 à 200 résidus) en 3' UTR (UnTranslated Region, région non traduite). Pour chaque individu, nos échantillons d'ARN total de branchie, trophosome et paroi du corps ont été passés sur colonne. Seuls les ARN disposant d'une queue poly-A peuvent se fixer à la colonne par hybridation avec les Oligo(dT). Certains ARN messagers bactériens possèdent aussi une queue poly-A, mais celle-ci est beaucoup plus courte (15 à 60 résidus) et n'existe que dans 2 à 60 % des transcrits de chaque gène (Sarkar, 1997). Les ARN bactériens, présents dans les échantillons d'ARN total de trophosome auront ainsi une plus faible probabilité de se fixer à la colonne par hybridation avec les Oligo(dT). On peut donc admettre que les ARN messagers de *Riftia pachyptila* pourront être isolés des ARN bactériens lors de la purification des ARN totaux du trophosome.

Une fois les ARN messagers fixés, plusieurs lavages avec des tampons fortement puis faiblement salins (0,5 M NaCl puis 0,1 M NaCl) permettent d'éliminer tous les ARN ne se liant pas ou faiblement aux Oligo(dT) : ARN ribosomiques, ARN de transfert, ou encore ARN

bactériens. Les ARN poly-A sont ensuite élués avec un tampon non salin préchauffé à 65°C. Le dosage du produit d'éluion à 260 nm permet ensuite de calculer la quantité d'ARN poly-A présent pour les prochaines étapes.

### 3.2 Choix des échantillons sondeur et meneur

Pour la réalisation de banques soustractives, l'échantillon sondeur (ou « tester ») est l'échantillon qui nous intéresse, celui dont on veut construire la banque d'ADNc, tandis que l'échantillon meneur (ou « driver ») est l'échantillon de référence. Comme énoncé précédemment (Chapitre II), nous avons choisi pour cette thèse de produire 4 banques soustractives pour lesquelles l'ADNc de chaque tissu, branchie (BR), trophosome (TR) et paroi du corps (BW) est tour à tour sondeur ou meneur :

- la banque BR-BW, où BR est le sondeur et BW est le meneur.
- sa réciproque, la banque BW-BR, où BW est le sondeur et BR est le meneur.
- la banque TR-BW, où TR est le sondeur et BW est le meneur.
- sa réciproque, la banque BW-TR, où BW est le sondeur et TR est le meneur.

Chaque banque a été réalisée à partir d'échantillons de tissus provenant d'un seul individu, les résultats généraux des banques présentés dans le Chapitre IV reflètent ainsi le transcriptome d'un seul individu.

### 3.3 Préparation des ADNc de l'échantillon sondeur et de l'échantillon meneur

La synthèse des ADNc double-brin (Tableau 1) a été effectuée avec les réactifs et selon les instructions du kit Clontech PCR-Select™ cDNA Subtraction Kit (BD Biosciences). Tous les échantillons sont ensuite digérés par l'enzyme de restriction RsaI (Tableau 2).

**Tableau 1 : Protocole de synthèse des ADNc double-brin des échantillons cibles et meneurs (d'après Clontech PCR-Select™ cDNA Subtraction Kit, BD Biosciences).** L'eau est toujours utilisée stérile et traitée au DEPC.

synthèse d'ADNc	simple- brin	2 µl d'ARN poly-A (1 µg.µl <sup>-1</sup> ), 1µl d'amorce (10 µM), 2 µl d'eau 70°C pendant 2 min, puis 4°C pendant 2 min, centrifuger.
		Ajouter 2 µl de Tampon 5X Premier-brin, 1 µl de dNTP (10mM de chaque nucléotide), 1 µl d'eau, 1 µl d'enzyme AMV Reverse Transcriptase (20 U.µl <sup>-1</sup> ). Mélanger, centrifuger, et incuber à 42°C pendant 1 heure 30.
	double- brin	Ajouter 48.4 µl d'eau, 16 µl de Tampon 5X Second-brin, 1.6 µl de dNTP (10mM de chaque nucléotide), 4 µl de cocktail d'enzyme 20X Second-brin. Mélanger, centrifuger et incuber à 16°C pendant 2 heures.
		Ajouter 6 U d'ADN Polymérase T4. Mélanger et incuber à 16°C pendant 30 min. Ajouter 4 µl d'EDTA-Glycogène 20X. Synthèse terminée.  Ajouter 100 µl de phénol : chloroforme : alcool isoamyl (25 : 24 : 1). Mélanger, centrifuger à 14 000 rpm pendant 10 min, puis prélever le surnageant. Etape renouvelée une fois. Ajouter 40 µl de NH <sub>4</sub> OAc 4M et 300 µl d'éthanol 95 %. Mélanger, centrifuger à 14 000 rpm pendant 20 min et éliminer le surnageant. Laver le culot à l'éthanol 80°C, le sécher à l'air pendant 10 min puis le dissoudre dans 50 µl d'eau.

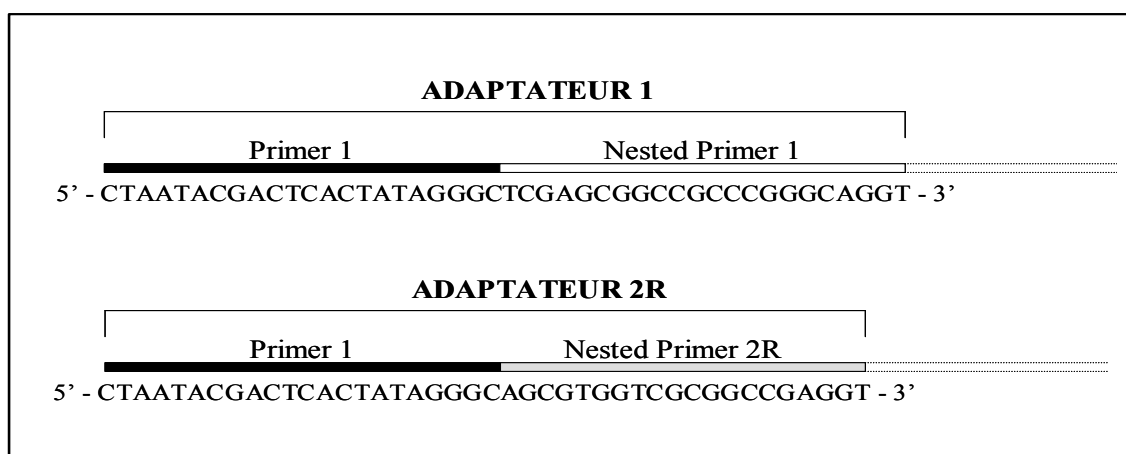
**Tableau 2 : Protocole de digestion par RsaI des ADNc double-brin des échantillons cibles et meneurs (d'après Clontech PCR-Select™ cDNA Subtraction Kit, BD Biosciences).**

Digestion par RsaI	Mélanger 43,5 µl d'ADNc double-brin, 5 µl de tampon 10X de restriction RsaI, 1,5 µl d'enzyme RsaI (10 U.µl <sup>-1</sup> ) Centrifuger brièvement et incuber à 37°C pendant 1 heure 30.
Précipitation	Ajouter 2,5 µl d'EDTA-Glycogène 20X. Réaction terminée. Ajouter 50 µl de phénol : chloroforme : alcool isoamyl (25 : 24 : 1). Mélanger, centrifuger à 14 000 rpm pendant 10 min, puis prélever le surnageant. Etape renouvelée une fois. Ajouter 25 µl de NH <sub>4</sub> OAc 4M et 187,5 µl d'éthanol 95 %. Mélanger, centrifuger à 14 000 rpm pendant 20 min à T°C ambiante et éliminer le surnageant. Laver le culot à l'éthanol 80°C, le sécher à l'air pendant 10 min puis le reprendre dans 5,5 µl d'eau. Les ADNc digérés sont conservés à -20°C.

Puis, seuls les ADNc digérés de l'échantillon sondeur sont séparés en 2 lots d'ADNc pour l'étape de ligation avec des adaptateurs différents (Tableau 3). Ces adaptateurs sont de courtes séquences oligonucléotidiques dont l'une des extrémités est une partie (GT) du site de restriction de RsaI (GT/AC). Le premier lot sera lié à l'adaptateur 1, le second à l'adaptateur 2R (Figure 2). Un échantillon constitué d'un mélange des 2 lots et dilué à la fin de la ligation à 1/1 000<sup>ème</sup> (Tableau 2) constituera par la suite l'échantillon non soustrait (dont l'utilisation sera expliquée dans ce chapitre, section 4.2).

**Tableau 3 : Protocole de ligation des adaptateurs aux ADNc digérés de l'échantillon cible (d'après Clontech PCR-Select<sup>TM</sup> cDNA Subtraction Kit, BD Biosciences).**

A partir d'ADNc digéré de l'échantillon cible dilué à 1/6ème	
Lot 1 (volume total = 10 µl)	Lot 2 (volume total = 10 µl)
2 µl ADNc digéré dilué 2 µl adaptateur 1 (10 µM) 3 µl d'eau stérile 2 µl de tampon de ligation 5X 1 µl d'ADN Ligase T4 (400 U.µl <sup>-1</sup> )	2 µl ADNc digéré dilué 2 µl adaptateur 2R (10 µM) 3 µl d'eau stérile 2 µl de tampon de ligation 5X 1 µl d'ADN Ligase T4 (400 U.µl <sup>-1</sup> )
Centrifuger brièvement et incuber à 16°C toute la nuit (tube supplémentaire : mélange de 2 µl de lot 1 et 2 µl de lot 2).  Ajout d'1 µl d'EDTA-Glycogène 20X. Ligation terminée. Incubation à 72°C pendant 5 min pour inactiver la ligase. Conservation des échantillons à -20°C.	



**Figure 2 :** Séquences nucléotidiques des adaptateurs 1 et 2R. Les adaptateurs se lient aux ADNc digérés du sondeur (en pointillés) à leur extrémité 3'. Les adaptateurs sont composés d'une séquence commune (Primer 1) et d'une séquence différente (Nested Primer 1 et Nested Primer 2R) qui permettront la fixation des amorces Primer 1, Nested Primer 1 et Nested Primer 2R au cours des deux PCR successives.



### 3.4 Hybridation Soustractive Suppressive (SSH)

La SSH (Diatchenko *et al.*, 1996 ; Gurskaya *et al.*, 1996) repose sur l'hybridation entre les ADNc communs de l'échantillon sondeur et de l'échantillon meneur, leur soustraction par l'amplification des ADNc spécifiques de l'échantillon sondeur et la régulation de l'abondance des transcrits de l'échantillon sondeur. La Figure 3 reprend de façon précise ces différentes étapes.

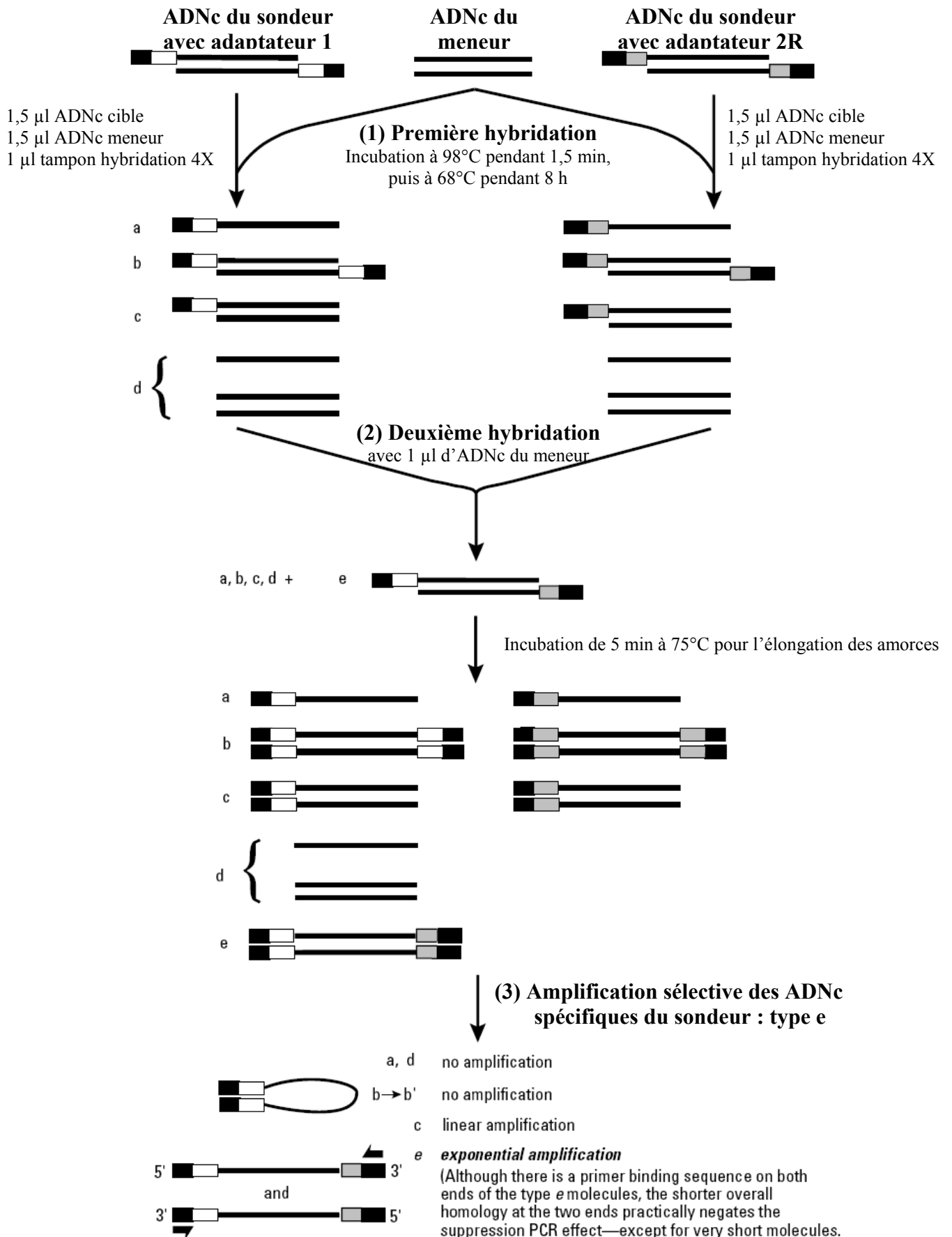
#### **(1) Première hybridation entre les ADNc communs du sondeur et du meneur (Figure 3)**

Les 3 types d'échantillons sont dénaturés puis les ADNc digérés de l'échantillon meneur sont ajoutés en excès à chacune des 2 populations d'ADNc de l'échantillon sondeur. Dès lors, les séquences complémentaires (représentant un même ADNc) vont s'hybrider entre elles. En théorie, plusieurs cas de figure sont possibles :

- les séquences spécifiques de l'échantillon sondeur ne s'hybrident pas avec les séquences de l'échantillon meneur (a, Figure 3) mais peuvent s'hybrider entre elles (b, Figure 3) si leur probabilité de rencontre est élevée (c'est-à-dire s'il s'agit de séquences abondantes dans l'échantillon) ;
- les séquences communes entre l'échantillon sondeur et l'échantillon meneur s'hybrident entre elles (c, Figure 3) mais avec plus ou moins d'efficacité selon leurs proportions relatives dans chacun des deux échantillons ;
- comme pour le cas (a) énoncé plus haut, les séquences spécifiques de l'échantillon meneur ne s'hybrident pas avec les séquences de l'échantillon sondeur mais peuvent éventuellement s'hybrider entre elles (d, Figure 3).

#### **(2) Mélange des produits d'hybridation des populations 1 et 2 du sondeur et seconde hybridation avec les ADNc du meneur (Figure 3)**

Au cours de cette étape, les séquences complémentaires entre les 2 populations d'ADNc de l'échantillon sondeur, correspondant au même ARNm, vont s'hybrider entre elles. L'ADNc double-brin résultant présentera à la fois les adaptateurs 1 et 2R à ses extrémités. L'ajout d'ADNc de l'échantillon meneur permet d'enrichir le mélange en séquences spécifiques de l'échantillon sondeur s'il subsistait encore des séquences communes simple-brin (voir (c, Figure 3) de la partie précédente).



**Figure 3 :** Etapes de la technique d'Hybridation Soustractive Suppressive (SSH) (d'après Clontech PCR-Select™ cDNA Subtraction Kit, BD Biosciences).

### **(3) Amplification sélective des ADNc spécifiques du sondeur (soustraction des ADNc identiques au meneur, et spécifiques du meneur) (Figure 3)**

Après une étape d'élongation de façon à ce que les ADNc double-brin spécifiques de l'échantillon meneur présentent des adaptateurs des deux côtés de chaque brin, deux PCR emboîtées sont conduites pour amplifier de façon sélective les ADNc spécifiques de l'échantillon sondeur (e, Figure 3). La première PCR est conduite avec des amorces spécifiques de la partie commune des adaptateurs (« Primer 1 », Figure 2). Le mélange réactionnel est constitué d'1 µl d'une dilution à 1/5<sup>ème</sup> du mélange issu de l'étape (2), 2,5 µl de tampon 10X de la Taq, 19,5 µl d'eau stérile, 0,5 µl de dNTP (10 mM), 1 µl de Primer 1 (10 µM) et 0,5 µl de 50X Advantage cDNA Polymerase Mix. Le programme PCR est constitué d'une dénaturation initiale à 94°C pendant 25 s, suivie de 27 cycles du programme : 94°C pendant 10 s, 66°C pendant 30 s, 72°C pendant 1 min 30s. Une seconde PCR avec des amorces complémentaires des parties spécifiques des adaptateurs 1 et 2R (« Nested Primer 1 » et « Nested Primer 2R », Figure 2) permet d'enrichir le mélange en séquences différentiellement exprimées. Le mélange réactionnel est constitué d'1 µl d'une dilution à 1/10<sup>ème</sup> du produit de première PCR, 2,5 µl de tampon 10X de la Taq, 18,5 µl d'eau stérile, 0,5 µl de dNTP (10 mM), 1 µl de Nested Primer 1 (10 µM), 1 µl de Nested Primer 2R (10 µM) et 0,5 µl de 50X Advantage cDNA Polymerase Mix (BD Biosciences). Le programme PCR est constitué de 12 cycles du programme : 94°C pendant 10 s, 68°C pendant 30 s, 72°C pendant 1 min 30s. Les ADNc spécifiques de l'échantillon sondeur sont alors amplifiés de façon exponentielle. Il en résulte un « smear » après migration du produit PCR sur gel d'agarose 1,5 %, correspondant à une variabilité de taille des séquences obtenues. Les ADNc de l'échantillon meneur ne présentant pas d'adaptateurs ne peuvent être amplifiés et sont ainsi « soustraits » du produit PCR.

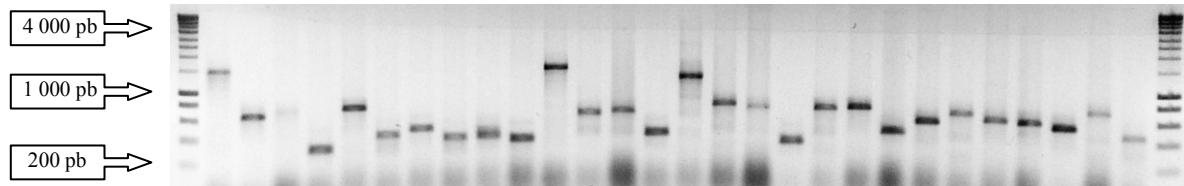
### **(4) Suppression des ADNc abondants du sondeur**

Les ADNc de type b (Figure 3) de l'échantillon sondeur présentent deux adaptateurs identiques à chacune de leurs extrémités. Ces séquences sont complémentaires sur chaque brin, si bien qu'après dénaturation lors de la PCR, ces séquences ont tendance à s'hybrider formant ainsi des boucles (b', Figure 3) empêchant toute amplification ultérieure.

## 3.5 Clonage

Nous avons décidé de séquencer à peu près 200 clones pour chaque banque générée.

Pour chaque soustraction effectuée, l'ensemble du produit PCR obtenu a pu être inséré dans le vecteur TOPO<sup>®</sup> puis cloné dans des bactéries TOP10 chimiocompétentes grâce au système TOPO TA Cloning<sup>®</sup> (Invitrogen) selon les instructions du fournisseur. Choies de façon arbitraire, les colonies sont ensuite testées par PCR pour vérifier que les plasmides ont bien intégré le vecteur TOPO<sup>®</sup> et que celui-ci contient un insert (Figure 4). Pour les 4 banques réalisées, la taille moyenne des inserts se situe autour de 400 paires de bases.



**Figure 4 :** Gamme de taille des inserts obtenus après clonage du produit de 2<sup>ème</sup> PCR de la banque BR-BW. La taille des fragments d'ADN est indiquée par le marqueur de taille Smart Ladder (Eurogentec) en paires de bases (pb), à laquelle il faut enlever environ 100 pb (correspondant aux deux parties du vecteur également amplifiées par les amorces T3 et T7). Des gammes de taille comparables ont été obtenues pour les banques BW-BR, TR-BW et BW-TR.

Le mélange réactionnel (dans lequel est plongé un cure-dent ayant été en contact avec la colonie testée) est constitué de 2,5 µl de tampon 10X de la Taq, 18,75 µl d'eau stérile, 2,5 µl de dNTP (4mM total), 0,5 µl d'amorce T3 (100 ng.µl<sup>-1</sup>), 0,5 µl d'amorce T7 (100 ng.µl<sup>-1</sup>), 0,75 µl de MgCl<sub>2</sub> et 0,2 µl de Taq DNA Polymérase (5 U.µl<sup>-1</sup>) UptiTherm (Interchim). La PCR contrôle est constituée d'une dénaturation initiale à 94°C pendant 4 min, suivie de 25 cycles du programme : 94°C pendant 45 s, 55°C pendant 45 s, 72°C pendant 1 min, puis, d'une étape finale d'élongation à 72°C pendant 5 min. L'extraction des plasmides (minipreps) des cellules bactériennes à partir de cultures de 2 ml a été réalisée à l'aide du kit Flexiprep (Amersham Biosciences) dans un premier temps. A partir de la deuxième banque, les minipreps ont été réalisés en routine sur des plaques 96 puits grâce au robot pipeteur disponible sur la plateforme de séquençage de la Station.

### 3.6 Séquençage

Le séquençage a été réalisé selon la méthode de Sanger *et al.* (1977) : l'ADN est copié à partir de l'extrémité 3'OH d'une amorce par une polymérase avec des arrêts aléatoires provoqués par l'introduction de ddNTP terminateurs. La réaction de séquence se fait avec un « premix », le kit BigDye<sup>®</sup> Terminator V3.1 Cycle Sequencing (Applied Biosystems). Le BigDye<sup>®</sup> contient une Taq Polymérase, des dNTP et des ddNTP marqués avec un

fluorochrome différent pour chaque base. Pour chaque réaction de séquence, 1 µl de miniprep (100 ng) est ajouté à 0,75 µl de tampon de BigDye<sup>®</sup>, 1 µl d'amorce T3 (ou T7) à 5 µM, 0,5 µl de BigDye<sup>®</sup> et 1,75 µl d'eau. Le programme utilisé pour les réactions de séquence est constitué d'une dénaturation initiale à 96°C pendant 5 min, suivie de 50 cycles du programme : 96°C pendant 30 s, 55°C pendant 30 s, 60°C pendant 4 min. La purification des réactions de séquence et le passage des échantillons sur un séquenceur à 16 capillaires (séquenceur Applied Biosystems 3130) ont ensuite été assurés par Morgan Perennou et Nathalie Gloaguen de la Plate-forme de Séquençage OUEST-genopole<sup>®</sup> de Roscoff.

### 3.7 Analyse des séquences et recherche d'homologie

Les chromatogrammes obtenus après séquençage ont pour la plupart été traités automatiquement par les logiciels PHRED et Seqclean pour nettoyer les séquences des vecteurs et adaptateurs. Ensuite, l'utilisation de divers logiciels (TGICL, Megablast et CAP3) permet de générer des regroupements de séquences partageant la plus grande homologie entre elles (les « clusters »). Dans chaque cluster, les séquences identiques ou montrant un recouvrement partiel ont pu être groupées dans un même « contig » (représentant un seul et même ADNc). Ces analyses ont été réalisées sur l'ensemble des séquences des 4 banques, et pour chacune des banques séparément : cette partie du travail a été réalisée par Erwan Corre du Service Informatique et Génomique de Roscoff. Au fur et à mesure du séquençage, de nouvelles séquences ont été nettoyées avec le logiciel Sequencher<sup>TM</sup> Version 4.1.4 (Gene Codes Corporation) et rajoutées manuellement aux contigs déjà formés. Les analyses d'homologie ont été réalisées par BlastX et BlastN sur le serveur NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>). Le logiciel PhyloGena (Hanekamp *et al.*, 2007) a aussi été utilisé ponctuellement pour vérifier certaines attributions d'homologie.

## **4. Efficacité de la soustraction et vérification de l'expression différentielle : validation des banques par PCR classique**

### 4.1 Dessin d'amorces spécifiques

Pour quelques séquences pour lesquelles nous montrons le plus d'intérêt, nous avons dessiné manuellement des couples d'amorces spécifiques en tenant compte de certains

critères : taille des amorces d'au minimum 18 paires de bases, contenu en bases G et C aux alentours de 50 %, amorce se terminant de préférence par 2 bases pyrimidiques (G ou C) pour une meilleure fixation à la matrice, et amorces d'un même couple présentant au maximum 4°C d'écart de  $T_m$ . La liste des couples d'amorces utilisés est disponible dans le Chapitre IV, section "Material and Methods".

#### 4.2 Amplification par PCR

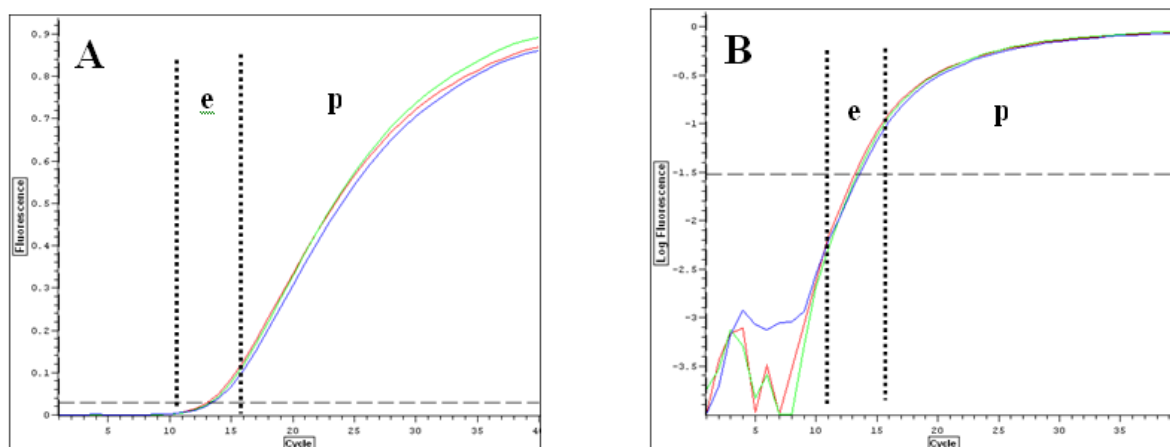
Les PCR ont été réalisées à partir d'échantillons de chaque tissu, et pour chacun des tissus dans un échantillon non soustrait (voir partie 3.3) dilué à  $1/10^{\text{ème}}$  et dans un échantillon obtenu après soustraction (produit de  $2^{\text{ème}}$  PCR issu du protocole SSH dilué à  $1/10^{\text{ème}}$  dans de l'eau stérile). Le mélange réactionnel est constitué d'1 µl de matrice, 3 µl de tampon 10X de la Taq, 22,4 µl d'eau stérile, 0,6 µl de dNTP (10 mM), 1,2 µl d'amorce Forward (10 µM), 1,2 µl d'amorce Reverse (10 µM) et 0,6 µl de 50X Advantage cDNA Polymerase Mix (BD Biosciences). Le programme PCR est constitué de 33 cycles du programme : 94°C pendant 30 s,  $T_m$  pendant 30 s, 68°C pendant 2 min. Au cours de la PCR, des aliquots sont prélevés à intervalles réguliers (à partir du 18ème cycle et tous les 5 cycles jusqu'à la fin de la PCR) puis sont mis à migrer sur gel d'agarose 1,5 % (dans du TAE 0,5 X). A nombre de cycles égal, l'amplification d'une séquence donnée a pu être comparée dans un même tissu avant et après soustraction (contrôle de l'efficacité de la soustraction) et entre un tissu sondeur et un tissu meneur (vérification de l'expression différentielle) : plusieurs types de résultats sont visibles sur les gels présentés dans le Chapitre IV, section "Results".

### **5. Vérification par PCR quantitative ou PCR en temps réel**

#### 5.1 Description d'une amplification par PCR

L'évolution de l'amplification d'un fragment d'ADN par PCR peut être représentée par une courbe dont l'allure est celle d'une sigmoïde (Figure 5A). Deux phases reconnaissables, une phase exponentielle et une phase en plateau, caractérisent la courbe (Figures 5A et B). Au cours de l'amplification exponentielle, la quantité de produit PCR obtenue est directement proportionnelle à la quantité initiale du fragment d'ADN amplifié : à chaque cycle, si l'efficacité de PCR est de 100 %, le nombre de copies double. La phase en

plateau correspond à l'épuisement des réactifs dans le mélange réactionnel. Au terme de  $n$  cycles d'une PCR classique, il est très difficile de dire si la phase plateau est atteinte, et si pour 2 amplifications à partir de 2 échantillons différents, les bandes observées sur gel d'agarose sont comparables ou non (c'est par exemple le risque encouru lors de l'utilisation de la méthode de quantification directe par analyse de l'intensité des bandes).



**Figure 5 :** Courbe d'amplification d'une réaction de PCR (en triplicat). L'axe des abscisses correspond au nombre de cycles de PCR. L'axe des ordonnées, en échelle linéaire (A) et en échelle logarithmique (B), correspond à la quantité de produit amplifié mesuré par la quantité de fluorescence émise. La phase exponentielle mesurable (e) et la phase de plateau (p) sont indiquées sur chaque courbe.

## 5.2 Intérêt et principe de la méthode

La PCR quantitative permet une mesure directe (en temps réel) de la réaction au cours de l'amplification. Au cours des cycles d'amplification, l'accumulation de produit PCR est directement corrélée à l'intensité de fluorescence (axe des ordonnées, Figure 5) émise par un fluorochrome, le SYBRGreen, lorsqu'il est incorporé dans l'ADNc double-brin formé. En se plaçant systématiquement à un même seuil d'amplification en phase exponentielle, cette méthode permet une comparaison très précise de la quantité d'un même transcrit dans plusieurs échantillons différents. En outre, chaque banque ayant été réalisée à partir d'échantillons d'un individu unique, nous pourrions valider nos résultats en analysant les tissus de plusieurs individus différents.

Pour permettre des comparaisons d'expressions entre échantillons différents, il nous faut choisir un même gène normalisateur ou gène de référence qui s'exprime de façon ubiquitaire quel que soit l'échantillon (ici le tissu) considéré. Après quelques essais préliminaires, l'ARN ribosomal 18S s'est trouvé être un bon candidat car exprimé de façon

identique dans trois tissus d'un individu : les courbes d'amplifications à partir d'une même quantité d'ARN total (1,24 µg) de branchie, trophosome et paroi du corps d'un même individu se sont avérées pratiquement identiques avec un Ct moyen de 12,3 ( $\pm$  1,3) cycles (voir la Figure 6 pour la définition du Ct).

### 5.3 Synthèse d'ADNc

A partir de chaque échantillon d'ARN total, la synthèse d'ADNc a été effectuée par transcription reverse avec l'enzyme Reverse Transcriptase M-MLV (Promega). Chaque volume réactionnel est composé de 2 µl du tampon 5X de l'enzyme, 0.5 µl de Bovine Serum Albumin (BSA) (10 mg/ml), 1 µl d'ARN total (1.24 µg/µl), 2.5 µl de dNTP (4 mM total), 1.5 µl de Random Primer 9 (Ozyme) (100 ng/µl) et 3 µl d'eau stérile traitée au DEPC. Chaque mélange est ensuite chauffé pendant 5 min à 80°C puis placé sur glace. L'enzyme Reverse Transcriptase M-MLV est ensuite ajoutée (1 µl) à chaque mélange réactionnel, puis toutes les réactions sont incubées à 42°C pendant 1 heure puis placées sur glace. Les réactions de PCR quantitative sont conduites directement après la synthèse d'ADNc préparée fraîche.

### 5.4 Dessin d'amorces et paramètres de PCR

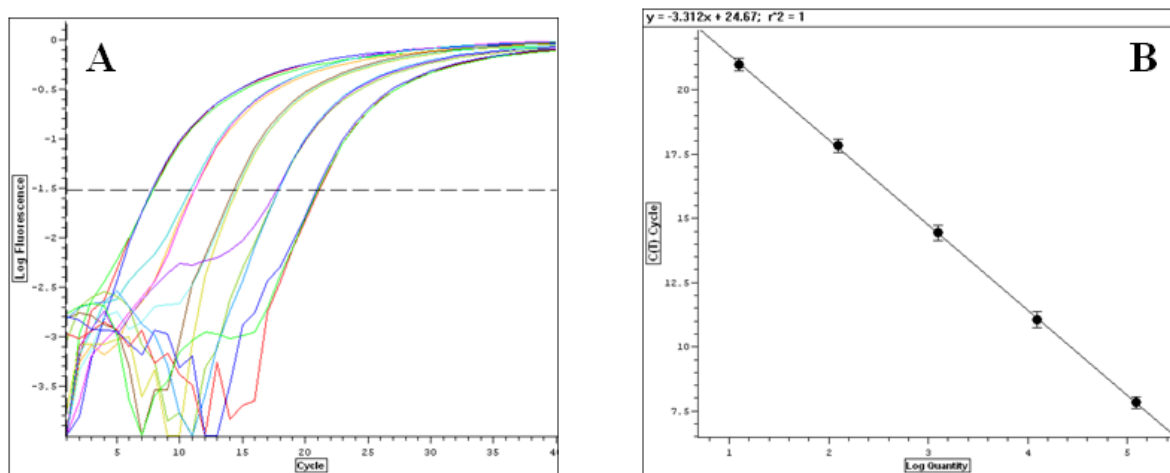
L'ARN ribosomique 18S a été amplifié avec les amorces 18L et 18h (Halanych *et al.*, 1998). Un choix judicieux des amorces de PCR est essentiel pour assurer la spécificité nécessaire : seul notre ADNc d'intérêt doit être amplifié. Pour toutes les autres amplifications menées, les couples d'amorces ont été dessinés à l'aide du logiciel Primer Express, selon des critères optimisés pour la PCR avec SYBRGreen (amplicon d'environ 50 paires de bases) et une température d'hybridation identique (60°C). Les séquences utilisées pour l'ensemble de ce travail sont disponibles dans le Chapitre IV, section "Material and Methods". Le mélange réactionnel constitué de 1 µl de matrice, 12,5 µl de Power SYBRGreen PCR master mix (PE Applied Biosystems), 0,5 µl d'amorce Forward (15 µM), 0,5 µl d'amorce Reverse (15 µM) et 9,25 µl d'eau. Le programme (95°C pendant 30s, 60°C pendant 30s, 72°C pendant 1 min) pendant 40 cycles a pu être utilisé en routine pour toutes les amplifications sur un thermocycleur Chromo4™ System CFB-3240 (BIORAD). Pour chaque couple d'amorces testé, une courbe de dissociation réalisée en fin de PCR par dénaturation de l'amplicon à haute température permet de vérifier que l'on a bien amplification d'un ADNc unique et qu'il n'y a



pas de formation de dimères d'amorces (contrôle également réalisé sur gel d'agarose 1,5 % où une seule bande à la taille de l'amplicon doit être visible).

### 5.5 Méthode du $2^{-\Delta\Delta C_t}$ : conditions d'application

Pour que cette méthode (détaillée dans la section 5.6 suivante) soit applicable, il est nécessaire que les efficacités de PCR du transcrit de référence et du transcrit cible soient approximativement égales (écart maximal aux alentours de 5 %) et les plus proches de 100 %. Aussi, pour chaque transcrit visé, une courbe standard est obtenue grâce à des PCR menées à partir de différentes quantités initiales d'ADNc (Figure 6).

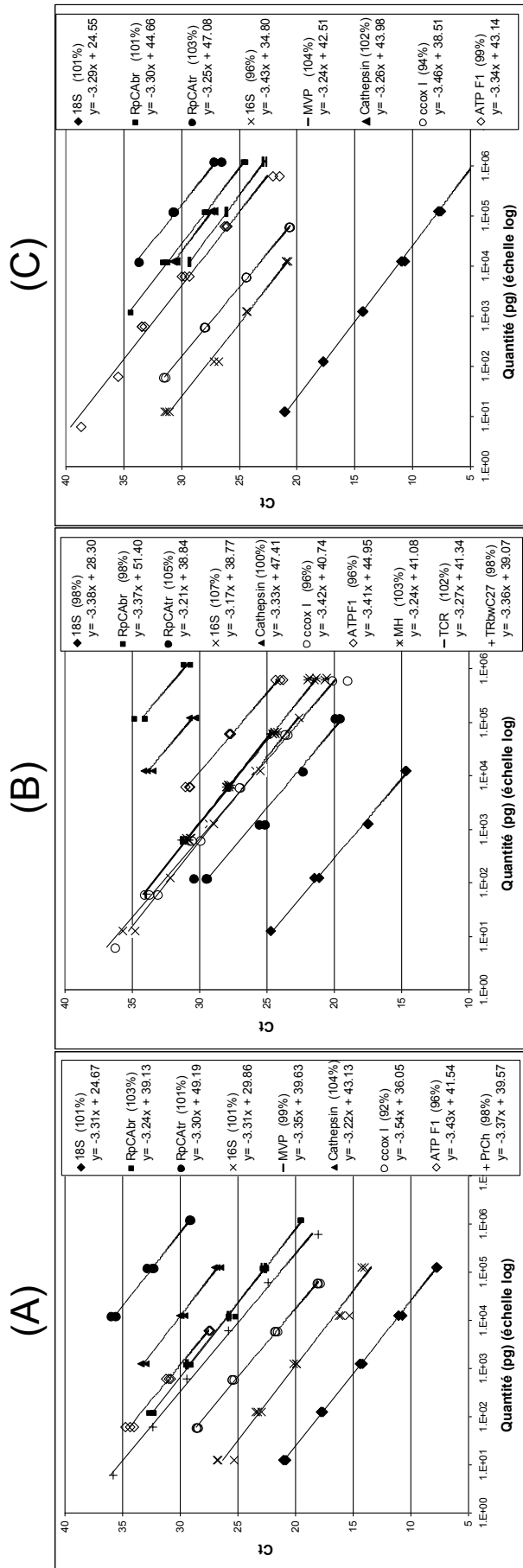


**Figure 6 :** (A) Courbes d'amplification à partir de dilutions consécutives de matrice d'ADNc. La ligne en pointillés correspond à un seuil fixé (« threshold ») de quantité de produit PCR amplifié. Pour chaque dilution, l'intersection de la courbe et du seuil donne le nombre de cycles nécessaires pour atteindre le seuil : c'est le Ct. Chaque Ct est ensuite reporté en fonction de la quantité initiale de matrice pour constituer la courbe standard (B). (Les données présentées sont celles de l'amplification du 18S à partir d'ADNc total de branchie.)

Chaque efficacité (E) est ensuite calculée à partir de la pente (S) de chaque courbe standard selon la formule :

$$E = 10^{-1/S} - 1$$

Les résultats des calculs d'efficacité de PCR pour les amplifications menées à partir d'échantillons d'ADNc de branchie, de trophosome et de paroi du corps sont présentés respective-ment Figures 7A, 7B et 7C.



**Figure 7 :** Courbes standard construites pour l'ensemble des transcrits étudiés au cours de cette thèse à partir d'ADNc total de (A) branchie, (B) trophosome et (C) paroi du corps. Pour chaque dilution, les réactions de PCR ont été réalisées en triplicats. L'efficacité de PCR (indiquée entre parenthèses) est calculée à partir de la pente de chaque droite. Pour chaque tissu, les différences d'efficacité de PCR observées entre l'amplification du 18S et celle du transcrit cible sont correctes (de 0 à 9 % dans l'échantillon de branchie ; de 0 à 9 % dans l'échantillon de trophosome ; de 0 à 7 % dans l'échantillon de paroi du corps. Notons qu'un transcrit n'a pas pu être amplifié à partir d'ADNc de paroi du corps et de trophosome mais uniquement à partir d'ADNc de branchie (PrCh). D'autres transcrits n'ont pas pu être amplifiés à partir d'ADNc de branchie et de paroi du corps mais uniquement à partir d'ADNc de trophosome (MH, TCR et TRbwc27).

### 5.6 Méthode du $2^{-\Delta\Delta Ct}$ : détail de la méthode

Cette méthode (Encadré 1) a été élaborée par Livak et Schmittgen (2001).

<p>The equation that describes the exponential amplification of PCR is</p>	<p>quence context effects on the fluorescence properties of the probe, the efficiency of probe cleavage, purity of the probe, and setting of the fluorescence threshold. Therefore, the constant <math>K</math> does not have to be equal to one. Assuming efficiencies of the target and the reference are the same,</p>
$X_n = X_0 \times (1 + E_X)^n, \quad [1]$	$E_X = E_R = E,$
<p>where <math>X_n</math> is the number of target molecules at cycle <math>n</math> of the reaction, <math>X_0</math> is the initial number of target molecules. <math>E_X</math> is the efficiency of target amplification, and <math>n</math> is the number of cycles. The threshold cycle (<math>C_T</math>) indicates the fractional cycle number at which the amount of amplified target reaches a fixed threshold. Thus,</p>	$\frac{X_0}{R_0} \times (1 + E)^{C_{TX} - C_{TR}} = K, \quad [5]$
$X_T = X_0 \times (1 + E_X)^{C_{TX}} = K_X \quad [2]$	$X_N \times (1 + E)^{\Delta C_T} = K, \quad [6]$
<p>where <math>X_T</math> is the threshold number of target molecules, <math>C_{TX}</math> is the threshold cycle for target amplification, and <math>K_X</math> is a constant. A similar equation for the endogenous reference (internal control gene) reaction is</p>	<p>where <math>X_N</math> is equal to the normalized amount of target (<math>X_0/R_0</math>) and <math>\Delta C_T</math> is equal to the difference in threshold cycles for target and reference (<math>C_{TX} - C_{TR}</math>). Rearranging gives the expression</p>
$R_T = R_0 \times (1 + E_R)^{C_{TR}} = K_R, \quad [3]$	$X_N = K \times (1 + E)^{-\Delta C_T}, \quad [7]$
<p>where <math>R_T</math> is the threshold number of reference molecules, <math>R_0</math> is the initial number of reference molecules, <math>E_R</math> is the efficiency of reference amplification, <math>C_{TR}</math> is the threshold cycle for reference amplification, and <math>K_R</math> is a constant.</p>	<p>The final step is to divide the <math>X_N</math> for any sample <math>q</math> by the <math>X_N</math> for the calibrator (cb):</p>
<p>Dividing <math>X_T</math> by <math>R_T</math> gives the expression</p>	$\frac{X_{N,q}}{X_{N,cb}} = \frac{K \times (1 + E)^{-\Delta C_{T,q}}}{K \times (1 + E)^{-\Delta C_{T,cb}}} = (1 + E)^{-\Delta\Delta C_T}, \quad [8]$
$\frac{X_T}{R_T} = \frac{X_0 \times (1 + E_X)^{C_{TX}}}{R_0 \times (1 + E_R)^{C_{TR}}} = \frac{K_X}{K_R} = K. \quad [4]$	<p>Here <math>-\Delta\Delta C_T = -(\Delta C_{T,q} - \Delta C_{T,cb})</math>. For amplicons designed to be less than 150 bp and for which the primer and <math>Mg^{2+}</math> concentrations have been properly optimized, the efficiency is close to one. Therefore, the amount of target, normalized to an endogenous reference and relative to a calibrator, is given by</p>
<p>For real-time amplification using TaqMan probes, the exact values of <math>X_T</math> and <math>R_T</math> depend on a number of factors including the reporter dye used in the probe, the se-</p>	<p>amount of target = <math>2^{-\Delta\Delta C_T}</math>, <span style="float: right;">[9]</span></p>

**Encadré 1 :** Explication de la formule du  $2^{-\Delta\Delta Ct}$  (Livak et Schmittgen, 2001).

Pour un échantillon donné, l'expression de la normalisation de l'amplification du transcrite cible par rapport à l'amplification du transcrite de référence s'écrit :

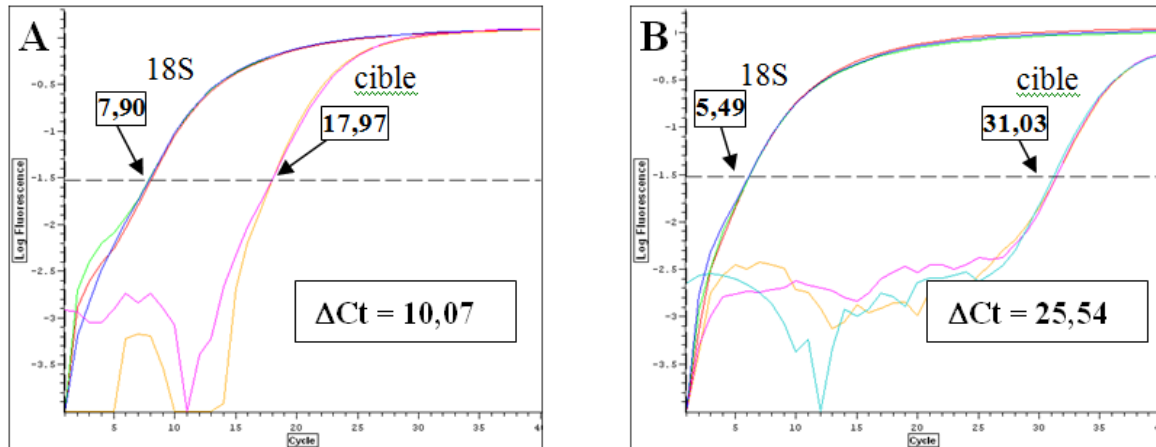
$$\Delta Ct = Ct \text{ cible} - Ct \text{ référence} = Ct \text{ cible} - Ct \text{ 18S}$$

L'étude de l'expression relative d'un gène cible consiste à comparer au moins deux échantillons (Figure 9). Ainsi, un échantillon va servir de « calibreur » et la quantification du gène cible dans les autres échantillons sera faite relativement à ce calibreur selon la formule :

$$2^{-\Delta\Delta Ct} \text{ qui peut s'écrire : } 2^{-(\Delta Ct \text{ échantillon} - \Delta Ct \text{ calibreur})}$$

Classiquement, l'échantillon calibreur est celui qui ne subit pas d'expérimentation, autrement dit l'échantillon contrôle. Dans le cas de notre étude, nous ne comparons pas des tissus identiques dans différentes conditions expérimentales, mais des tissus différents : pour

déterminer l'expression d'un gène dans un organe, le calibreur peut alors être un autre organe (Livak et Schmittgen, 2001). Nous avons choisi le calibreur comme étant le tissu où l'amplification de cible était la plus importante. Ainsi, dans notre exemple (Figure 8), le niveau d'expression observé dans l'échantillon B peut ainsi s'exprimer en fonction du niveau d'expression observé dans le calibreur A (pour lequel le niveau d'expression est fixé arbitrairement à 1).



**Figure 8 :** Calcul du  $2^{-\Delta\Delta Ct}$  appliqué aux amplifications du 18S et d'un transcrite cible à partir d'ADNc total du calibreur (A) et de l'échantillon (B).  $\Delta\Delta Ct = 15,47$  soit  $2^{-\Delta\Delta Ct} = 0.00002$  soit  $5.10^5$  fois plus de transcrite cible dans le calibreur que dans l'échantillon.

Pour plus de précision, nous avons analysé les ADNc de branchie, trophosome et paroi du corps de plusieurs individus. Pour intégrer ces répliquats « biologiques » à nos résultats, des modes différents de calcul sont envisageables :

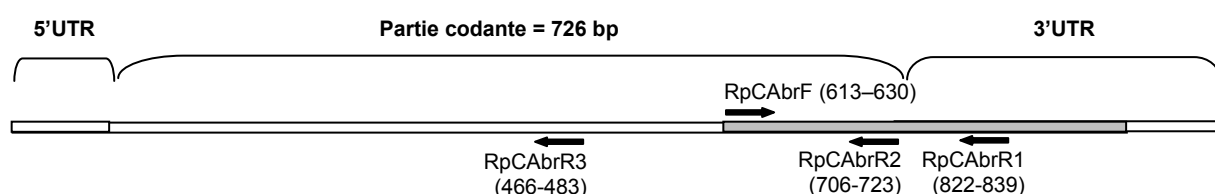
- (i) soit on effectue la moyenne des  $\Delta Ct$  de tous les échantillons et la moyenne des  $\Delta Ct$  de tous les calibreurs que l'on soustrait ensuite l'une à l'autre pour obtenir le  $\Delta\Delta Ct$ , puis le  $2^{-\Delta\Delta Ct}$ .
- (ii) soit on regarde l'expression relative par individu. Lorsque l'on dispose d'échantillons de plusieurs tissus d'un même individu, il devient possible d'apparier ces échantillons. Ainsi, pour chaque individu, on soustrait le  $\Delta Ct$  de l'échantillon au  $\Delta Ct$  du calibreur pour obtenir le  $\Delta\Delta Ct$ . Ensuite, on applique la formule  $2^{-\Delta\Delta Ct}$  pour chaque individu dont on dispose d'au moins deux tissus analysés. A terme, on obtient une moyenne d'expression par individu.

La méthode (ii) est intéressante puisqu'elle tient compte de l'unité biologique que représente chaque individu pris séparément. Elle apparaît ainsi beaucoup plus fine. Nous avons présenté nos résultats selon ce mode de calcul (Chapitre IV).

## 6. Autres techniques employées pour des cas plus spécifiques :

### 6.1 Séquençage complet par Rapid Amplification of cDNA Ends (RACE) (Chapitre IV)

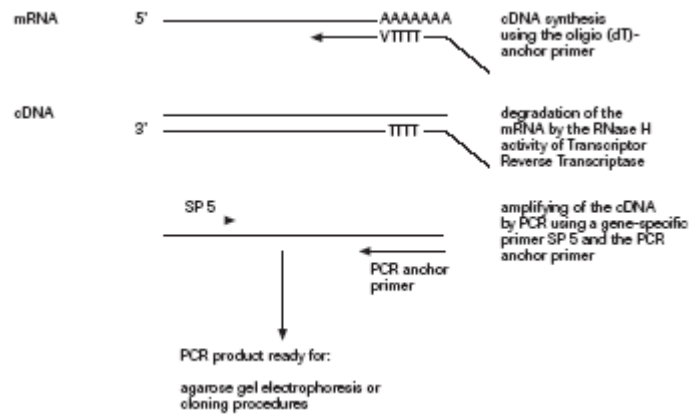
A partir de la banque BR-BW (Chapitre IV, section "Results"), nous avons obtenu une séquence partielle (partie grisée, Figure 9) d'un nouveau transcrit d'anhydrase carbonique, constituée de 170 paires de bases dans la région codante suivies d'une région 3' non traduite (3'UTR) longue mais partielle. L'obtention de la partie 3' complète par 3'RACE PCR a nécessité l'utilisation de l'amorce RpCAbrF (Figure 9), tandis que l'obtention de la partie 5' par 5'RACE PCR a nécessité l'utilisation des amorces RpCAbrR1, RpCAbrR2 et RpCAbrR3 (Figure 9), de façon à remonter progressivement vers le début de la séquence d'intérêt.



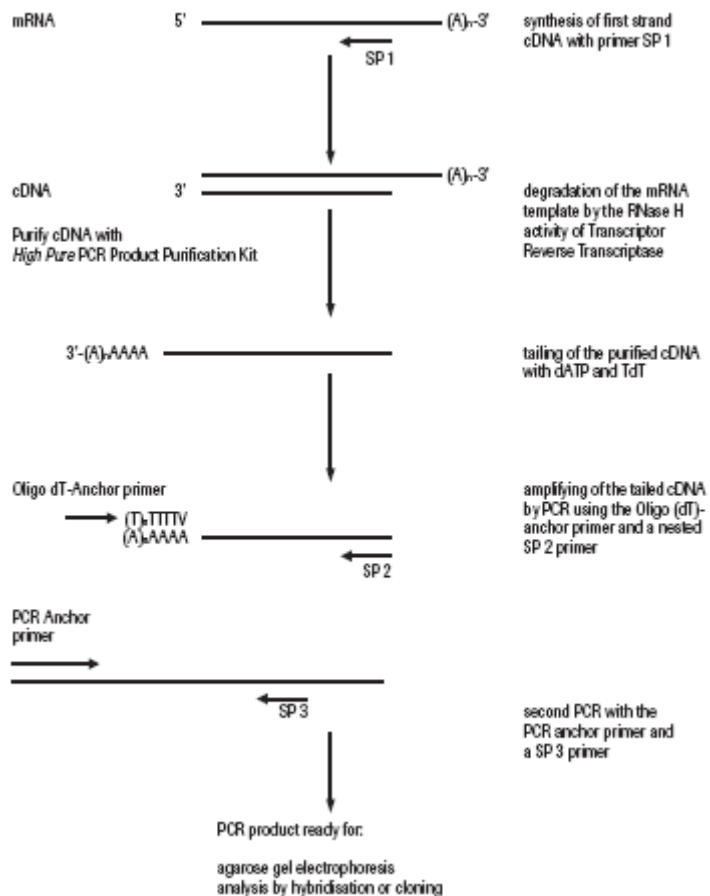
**Figure 9 :** Position des amorces utilisées pour l'obtention de la séquence complète RpCAbr par RACE. Les positions sont données à partir du début de la séquence codante. RpCAbrF : 5' TAC AAG GAT GCC ATT AGC 3' ; RpCAbrR1 : 5' CGT AGC AGT ATC AGC AGT 3' ; RpCAbrR2 : 5' AGA GCA GCA GAC CTT ACG 3' ; RpCAbrR3 : 5' GTT ACT TCC GCA GCT AGG 3'

Si le protocole de 3'RACE (Figure 11A) proposé par le fournisseur n'a pas posé de problème, celui de 5'RACE (Figure 11B) a dû être adapté : l'ajout de poly-A a été remplacé par un ajout de poly-C, et l'amorce oligodT a par conséquent été remplacée par une amorce oligodG.

**Figure 10 :** Schéma résumant le protocole de la RACE-PCR en 3' (5'/3' RACE Kit, 2nd Generation, Roche Applied Science)



**Figure 11:** Schéma résumant le protocole de la RACE-PCR en 5' (5'/3' RACE Kit, 2nd Generation, Roche Applied Science). Après des essais d'amplification infructueux, l'étape de ligation d'une queue poly-A a été remplacée par la ligation d'une queue poly-C. Une première PCR avec l'amorce RpCABrR2 et l'oligodG a été réalisée à 56°C (T°C d'hybridation). A partir d'1 µl du produit obtenu, une deuxième PCR a été réalisée avec l'amorce RpCABrR2 et à nouveau l'oligodG (mêmes conditions de PCR).



## 6.2 Alignements et analyses phylogénétiques (Chapitres V et VI)

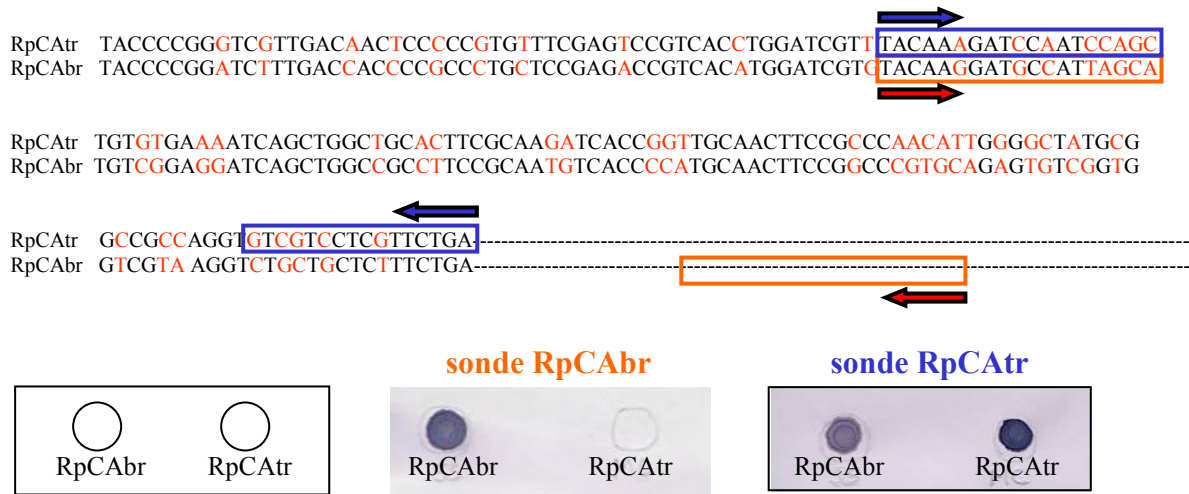
Des phylogénies ont été réalisées à partir de séquences traduites d'ADNc complets. Les alignements multiples ont été réalisés par l'algorithme ClustalW mis au point par Thompson *et al.* (1994), sous le logiciel MEGA3.1 (Kumar *et al.*, 2004) qui offre une interface très conviviale. Des méthodes de distance et de parcimonie ont été menées en parallèle de façon à identifier les clades les plus robustes comme étant retrouvés au travers des deux méthodes.

### 6.3 Choix des sondes et hybridation sur membranes pour l'optimisation de l'Hybridation *In Situ* par Fluorescence (FISH) (Chapitre IV)

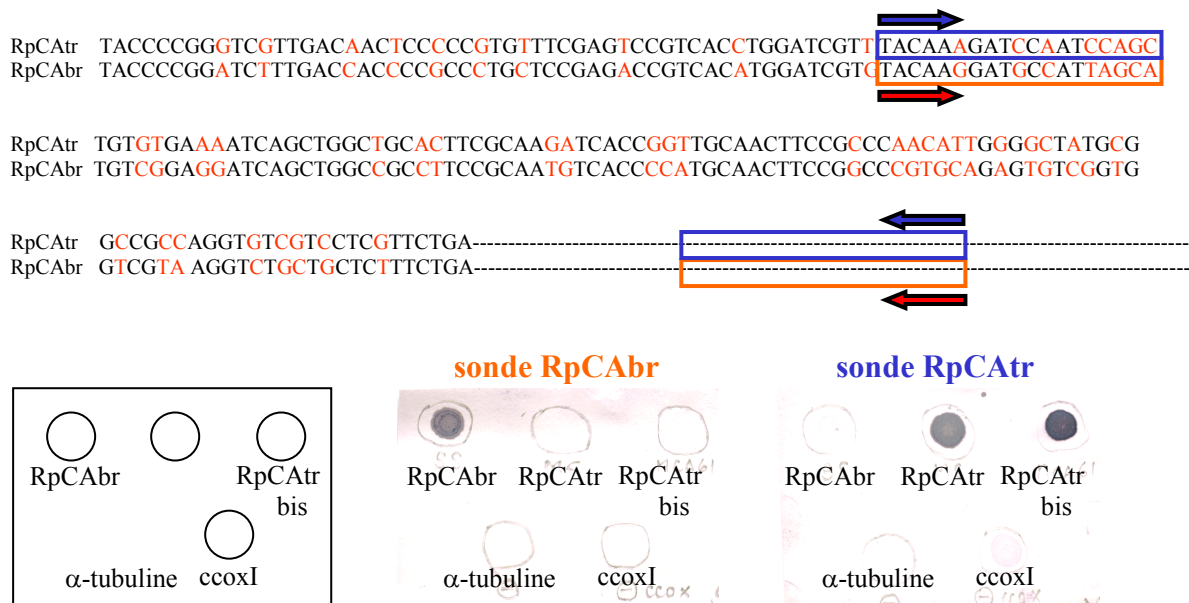
Après avoir identifié un nouveau transcrit d'anhydrase carbonique chez *Riftia pachyptila* (RpCAbr, voir le Chapitre IV) s'ajoutant à celui déjà identifié (RpCAtr) par De Cian et al. (2003), des sondes ADN spécifiques de chaque type de transcrit d'anhydrase carbonique ont été élaborées dans le but de localiser ces ARN sur des coupes de tissu branchial et de trophosome. Ces sondes anti-sens (c'est-à-dire complémentaires de l'ARN à localiser) d'environ 200 paires de bases ont été obtenues par PCR à partir de matériel biologique (ADNc de branchie pour RpCAbr et ADNc de trophosome pour RpCAtr) et marquées à la digoxygénine (DIG) par PCR selon le protocole du PCR DIG Probe Synthesis Kit (Roche Applied Science) et les instructions du fabricant.

La vérification de la spécificité des sondes (Figure 12) est une étape incontournable avant de poursuivre plus avant. Des minipreps de RpCAbr, RpCAtr et de diverses séquences constituant des témoins négatifs, tous dénaturés et dilués à  $1/20^{\text{ème}}$ , sont déposés sur une membrane de nitrocellulose. Les sondes sont diluées (2  $\mu\text{l}$  de sonde dans 4  $\mu\text{l}$  d'ADN de sperme de saumon ( $10 \text{ mg.ml}^{-1}$ ) et 36  $\mu\text{l}$  d'eau stérile) puis dénaturées à  $100^{\circ}\text{C}$  pendant 5 min. Chaque membrane est ensuite incubée à  $42^{\circ}\text{C}$  pendant 16 heures sous agitation dans un four à hybridation, soit avec la sonde RpCAbr, soit avec la sonde RpCAtr, diluées dans 2 ml de tampon d'hybridation. La composition de ce tampon et des suivants est indiquée sur le Tableau 4. Des lavages stringents sont réalisés avec une solution de lavage A (2 rinçages à température ambiante), puis une solution de lavage B (à  $55^{\circ}\text{C}$ , 2 rinçages de 30 min puis 1 rinçage de 15 min). A ce stade, une fois bien débarrassées des sondes en excès, les membranes peuvent être rincées ensemble avec une dernière solution de lavage C, puis incubées pendant 30 min dans une solution de blocage qui bloque les sites non spécifiques. Les membranes sont alors incubées pendant 30 min avec l'anticorps anti-DIG ( $150 \text{ mU.ml}^{-1}$ ) dilué à  $1/5\ 000^{\text{ème}}$  dans la solution de blocage. Suivent 2 rinçages de 15 min chacun dans la solution de lavage C puis une équilibration pendant 5 minutes dans du tampon de détection, au pH optimum de la phosphatase alcaline (pH 9,5). La dernière étape est l'incubation des membranes à l'obscurité avec la solution de substrat NBT/BCIP (nitro blue tetrazolium/5-bromo-4-chloro-indolyl phosphate) de l'enzyme : la détection colorimétrique a lieu au bout de quelques minutes. Après un premier essai, il a été nécessaire de redessiner la sonde RpCAtr qui s'hybridait de façon aspécifique avec la séquence RpCAbr (Figure 12A) de façon à ce qu'elle soit plus spécifique, en incorporant un fragment de partie 3'UTR (Figure 12B).

**A**



**B**



**Figure 12 :** Dot-blots réalisés pour vérifier la spécificité des sondes ADNc marquées au DIG. Les sites nucléotidiques différents entre les deux séquences sont indiqués en rouge. (A) La sonde RpCAtr réalisée par PCR avec les amorces exclusivement dessinées dans la partie codante n'apparaît pas spécifique (hybridation aussi bien avec la séquence RpCAbr qu'avec la séquence RpCAtr). (B) La sonde RpCAtr réalisée par PCR avec une amorce en partie codante et l'autre en partie 3'UTR apparaît bien spécifique de la séquence RpCAtr.



**Tableau 4 : Composition des tampons et solutions de lavage utilisés pour les dot-blots.**

Solution	Composition	Stockage
Tampon d'hybridation	1 M NaCl ; 50 mM Tris-HCl pH 7,4 ; 40 % formamide ; 10 % dextran sulfate ; 1 % SDS ; eau stérile	T°C ambiante
Solution de lavage A	2X SSC* ; 0,1 % SDS ; eau stérile	T°C ambiante
Solution de lavage B	0,1X SSC* ; 0,1 % SDS ; eau stérile	T°C ambiante
Solution de lavage C	0,1 M d'acide maléique** ; 0,15 M NaCl ; 0,3 % Tween 20	T°C ambiante
Solution de blocage	1/10 <sup>ème</sup> agent de blocage 10X (Roche) dans l'acide maléique**	A préparer extemporanément
Tampon de détection	0,1 M Tris-HCl ; 0,1 M NaCl ; eau stérile pH 9,5	T°C ambiante
Solution de coloration	Solution stock NBT/BCIP (Roche) diluée à 1/50 <sup>ème</sup> dans le tampon de détection	A préparer extemporanément
*Tampon SSC 20X	Dans 800 ml d'eau stérile : 175,3 g NaCl ; 88,2 g de citrate de sodium ; compléter à 1L pH 7	T°C ambiante
**Tampon d'acide maléique 0,1 M	0,1 M d'acide maléique ; 0,15 M NaCl ; eau stérile, ajuster à pH 7,5 avec NaOH solide	T°C ambiante

## 6.4 Hybridation *in situ* par Fluorescence (FISH) (Chapitre V)

### *6.4.1 Préparation des coupes*

Les coupes histologiques de 7 µm d'épaisseur ont été réalisées avec un cryomicrotome Leica CM3050S à -25°C et immédiatement déposées sur des lames de microscopie traitées au Biobond (Euromedex) à 2 % pour permettre le collage des coupes. Les coupes ont par la suite été traitées par deux types de déprotéinisation : une déprotéinisation acide et une enzymatique, selon le protocole détaillé dans le Chapitre V, section "Experimental Procedures".

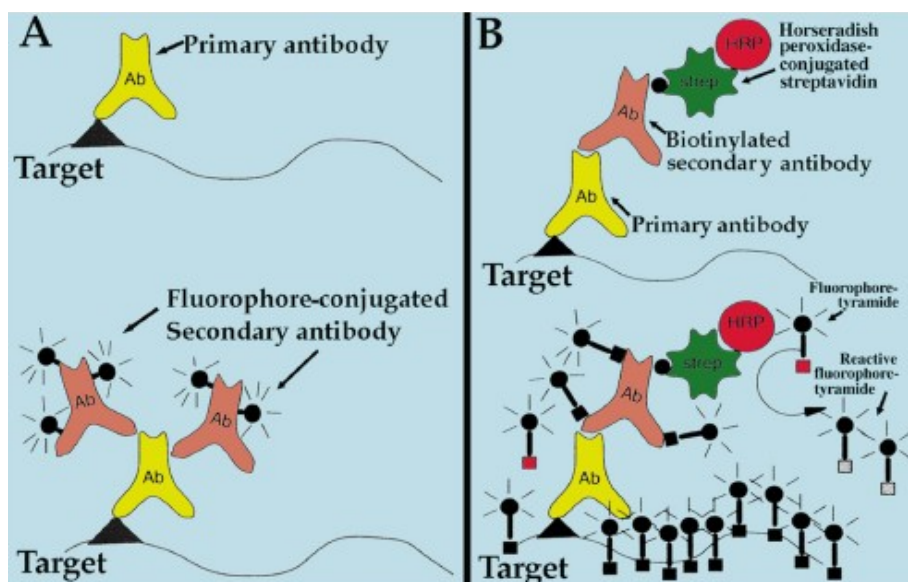
### *6.4.2 Hybridation des sondes*

Le principe de l'hybridation repose sur la complémentarité du transcrit d'intérêt avec une sonde oligonucléotidique anti-sens marquée (par exemple au DIG ou à la biotine),

formant ainsi un double-brin. Brièvement, la sonde est diluée à  $15 \text{ ng} \cdot \mu\text{l}^{-1}$  dans  $100 \mu\text{l}$  de tampon d'hybridation HB. Ce tampon contient notamment du dextran sulfate pour le rendre plus visqueux et augmenter la concentration de la sonde, du SDS pour charger négativement les protéines pour faciliter la pénétration de la sonde, du formamide déionisé dont une concentration croissante permet de diminuer la température d'hybridation. Nous avons pu réaliser l'hybridation à une température de  $44^\circ\text{C}$  pendant 20 heures dans une boîte scellée (chambre humide pour éviter toute évaporation) placée dans un four à hybridation. En parallèle, nous avons augmenté la température des tampons de rinçage de nos lames (rinçages de 20 min à 46, 52 et  $46^\circ\text{C}$ ) pour éliminer les hybridations aspécifiques.

#### 6.4.3 Détection

La sonde est reconnue par un anticorps dirigé contre elle (anticorps anti-DIG ou anticorps streptavidine) lors d'une incubation d'au moins 2 heures. Cet anticorps peut être associé à un fluorochrome (Figure 13A) pour la détection du signal. Récemment, il a été démontré que la sensibilité de la méthode pouvait être sensiblement accrue grâce à l'utilisation d'une technique d'amplification du signal, le Tyramide Amplification Signal (TSA, brevet déposé par NEN Life Science Products).



**Figure 13 :** Comparaison du système classique de détection de signal (A) et du système TSA indirect (B) (d'après Wang *et al.*, 1999) au cours duquel l'enzyme HRP réagit avec le réactif d'amplification pour générer de nombreux radicaux marqués par le fluorophore qui se déposent à proximité immédiate de l'enzyme.

Ce type de détection de signal est basé sur la présence d'une enzyme la HorseRadish Peroxydase (HRP) qui permet d'amplifier le signal, l'enzyme pouvant être ajoutée directement sur la sonde (TSA direct) ou sur un anticorps dirigé contre la sonde (TSA indirect) (Figure 13B). L'utilisation du système TSA a considérablement augmenté le ratio signal / bruit de fond au cours de nos manipulations. Le système TSA direct a pu être testé de façon concluante sur nos lames avec une sonde 18S spécifique des Eucaryotes directement marquée à l'HRP (don du Dr D. Vaultot, Equipe Plancton Océanique, SBR). Par la suite, nous avons utilisé le système TSA indirect pour la détection de nos 2 transcrits :

(1) la sonde RpCAbr a été marquée au DIG, puis reconnue avec un anticorps anti-DIG couplé à la HRP (anti-DIG-POD (Roche Diagnostics) dilué à 1/100<sup>ème</sup>). La détection du signal a lieu par l'ajout du réactif d'amplification contenant de la fluorescéine (émission dans le vert) ;

(2) la sonde RpCAtr (correspondant à la séquence testée Figure 10B) a été marquée à la biotine (biotin-16-ddUTP ajoutée par l'enzyme terminal transférase (Roche Diagnostics)), puis reconnue par un anticorps streptavidine couplé à la HRP (Perkin Elmer) et dilué à 1/100<sup>ème</sup>. La détection du signal a lieu par l'ajout du réactif d'amplification contenant de la rhodamine (émission dans le rouge).

L'ensemble du protocole et la composition des tampons utilisés sont détaillés dans le Chapitre V, section "Experimental Procedures".

### 6.5 Clonage d'un transcrit dans un vecteur d'expression (Chapitre VI)

Pour une analyse fonctionnelle, des expériences de surexpression ont été menées sur un transcrit pleine longueur codant une myohémérythrine (MHr), obtenu dans notre banque TR-BW (Chapitre IV, section "Results"). Nous avons utilisé le système Glutathion S-transférase (GST) pour l'expression, la purification et la détection de notre protéine myohémérythrine (protéine de fusion GST-MHr) produite par *Escherichia coli*.

Cette séquence a d'abord été insérée dans un vecteur d'expression, le vecteur pGEX-5X3. Pour cela, à partir du clone contenant la partie codante complète de la séquence, une amplification par PCR a été conduite avec des amorces spécifiques allongées de quelques nucléotides : l'amorce Forward (5' CGG GAT CCC GAT GTC TTT CGA GGT TCC TTC C 3') possédant en 5' un site de coupure spécifique de l'enzyme BamHI (site également présent dans le vecteur d'expression) et l'amorce Reverse (5' CGG AAT TCT TAC AGA TTT CCC TTG TAT T 3') possédant en 5' un site de coupure spécifique de l'enzyme EcoRI (site

également présent dans le vecteur d'expression). Après un clonage classique (TOPO TA Cloning<sup>®</sup>, Invitrogen) et contrôle des colonies par PCR (même protocole que dans la section 3.5 de ce chapitre), les clones positifs sont choisis pour être séquencés (même procédure que dans la partie 3.6 de ce chapitre). Cette étape de vérification permet de choisir le clone pour lequel les sites de restriction sont bien localisés aux deux extrémités de la séquence codante et que celle-ci est correcte. L'extraction du plasmide à partir d'une culture de 4 ml est ensuite réalisée (miniprep : Flexiprep kit, Amersham Biosciences). Le plasmide ainsi que le vecteur d'expression pGEX sont coupés par les enzymes de restriction BamHI et EcoRI à 37°C pendant 1 heure dans 30 µl (10 µl d'ADN, 3 µl de Tampon 10X, 2 µl EcoRI, 2 µl BamHI et 13 µl d'eau). Après migration sur gel d'agarose 1,5 %, les bandes correspondant à la séquence de myohémérythrine et au vecteur linéarisé sont purifiées avec le kit High Pure PCR Purification (Roche Diagnostics). La ligation de l'insert (séquence complète de myohémérythrine) et du vecteur pGEX-5X3 est réalisée à 16°C pendant 24 heures dans 20 µl (2 µl de Tampon 10X de l'enzyme T4 Ligase, 8 µl de vecteur, 8 µl d'insert et 2 µl d'enzyme T4 Ligase).

Le produit de ligation est ensuite cloné dans des bactéries TOP 10 afin d'augmenter le rendement de clone positif. Une seconde extraction de plasmides (miniprep) possédant notre insert s'ensuit. Un aliquot de 200 µl de bactéries BL21 (bactérie pour l'expression) compétentes est mis en présence de 2 µl du vecteur pGEX-5X3 possédant notre séquence de myohémérythrine. Après une incubation de 45 minutes à 4°C, les bactéries sont transformées par un choc thermique, 30 secondes à 42°C, puis immédiatement diluées dans 1 ml de LB et incubées à 37°C pendant 1 heure. Cette culture est étalée (généralement 100 µl) sur des boîtes de Pétri contenant du milieu LB agar supplémenté en ampicilline (100 µg/ml). Les boîtes sont mises à incuber une nuit à 37°C. Les clones sont ensuite vérifiés par PCR.

Les manipulations suivantes ont été réalisées par le Dr Thomas Harnois : le détail du protocole est disponible dans le Chapitre VI, section "Material and Methods". Brièvement, des pré-cultures ont été lancées dans de petits volumes, puis la surexpression en « batch » (volumes de 2 L) dans du milieu auto-inductible à 32°C pendant 3-4 jours a permis d'obtenir la protéine recombinante GST-MHr (Myohémérythrine fusionnée à la GST). Après centrifugation, le culot constitué des cellules bactériennes a été re-suspendu dans du tampon de lyse et les cellules lysées à la presse de French. Ensuite, le lysat est purifié par passage sur

des billes de glutathion fixatrices de GST qui retiennent ainsi la protéine de fusion GST-MHr (Figure 14).

**Figure 14 :** Fixation de la protéine de fusion GST-MHr aux billes de glutathion sépharose (Cliché : T. Harnois). Noter la couleur rose des billes, normalement blanches.



#### 6.6 Test d'activité superoxyde dismutase (SOD) (Chapitre V)

La myohémérythrine a pu être isolée des billes de glutathion en mélangeant 200 µl de billes à 1,3 ml de tampon d'élution (glutathion réduit 20 mM, Tris 50 mM) à 4°C sous agitation pendant au moins 2 heures. Après centrifugation, le surnageant a été récupéré, puis la protéine a été concentrée sur des filtres Microcon<sup>®</sup> (Millipore). En parallèle, l'hémérythrine d'un siponcle *Sipunculus nudus* récolté sur l'estran en Baie de Morgat (Finistère) a été isolée à partir du liquide coelomique de l'animal. Pour récupérer les cellules roses remplies d'hémérythrine, 1 volume de liquide coelomique est centrifugé à 1 000 g pendant 5 min et le surnageant éliminé. Les cellules sont lysées par choc osmotique après ajout de 0,5 volume d'eau distillée. Une centrifugation à 10 000 g pendant 5 min permet d'isoler la protéine soluble (dans le surnageant rose) des débris cellulaires. Celle-ci a ensuite été purifiée (D. Le Guen) sur colonne Sephadex pour séparer l'hémérythrine des protéines contaminantes. Les échantillons de myohémérythrine surexprimée de *Riftia pachyptila* et d'hémérythrine de *Sipunculus nudus* ont ensuite été dosés par la méthode de Bradford (1976). A partir d'une quantité identique de protéine (2 mg.ml<sup>-1</sup>), des tests d'activité SOD ont été menés.

Ceux-ci ont reposé sur l'action de la xanthine oxydase sur la xanthine qui produit de l'H<sub>2</sub>O<sub>2</sub>. Ce peroxyde réagit avec le NBT donnant une coloration bleue. Une protéine capable d'activité SOD va consommer l'H<sub>2</sub>O<sub>2</sub> qui ne pourra plus réagir avec le NBT : plus l'activité SOD de cette protéine sera importante, moins la coloration bleue se développera. Le protocole suivi est détaillé au Chapitre VI , section "Material and Methods".

### 6.7 Western-blot pour détecter la Major Vault Protein (MVP) (Chapitre VII)

Des western-blots sur des extraits protéiques des trois tissus ont été réalisés avec les anticorps « George » (anticorps polyclonal purifié de lapin, obtenu à partir de la purification des acides aminés 593 à 805 de la MVP de *Rattus norvegicus* surexprimée dans *E. coli* (Silva *et al.*, 2001) et « Bob » (anticorps polyclonal purifié de lapin, obtenu à partir de la purification des protéines MvpA et MvpB de *Dictyostelium discoideum* (Vasu *et al.*, 1993)), généreusement fournis par le Dr L. Rome (UCLA University, Los Angeles, USA). Des morceaux de tissus de branchie, trophosome et paroi du corps d'un même individu ont été broyés et dilués dans 500 µl d'inhibiteurs de protéases dilués dans du PBS 1X. Après centrifugation à 15 000 g pendant 30 min à 4°C, le surnageant est récupéré et dosé par la méthode de Bradford (1976). Une quantité égale (100 µg) de chaque extrait protéique est dénaturée pendant 5 min à 95°C dans du tampon de dépôt (contenant du β-mercaptoéthanol) et mise à migrer sur gel d'acrylamide SDS-PAGE à 200 V pendant 45 min. Le transfert du gel à une membrane de nitrocellulose se fait dans un montage en « sandwich » dans du tampon de transfert (0,5 volume de tampon de migration TG-SDS 10X (Amresco), 0,2 volume d'éthanol pur et 0,3 volume d'eau). Le gel est positionné du côté négatif pour un transfert vers la membrane de nitrocellulose positionnée du côté positif, dans une cuve contenant le tampon de transfert et un bloc de glace. Le transfert se déroule à 300 mA pendant 2 h en chambre froide pour éviter une surchauffe.

Le transfert terminé, la membrane est mise à saturer toute la nuit à 4°C dans un mélange de PBS 1X, Tween 0,1 % et BSA 1 %. Après 4 lavages de 10 min chacun avec du Tween 0,2 % (dans PBS 1X), la membrane est incubée avec l'anticorps primaire « George » (ou « Bob ») dilué à 1/1 000<sup>ème</sup> (ou 1/250<sup>ème</sup>) dans du Tween 0,1 % (dans PBS 1X) pendant 2 heures avec agitation. Après 4 nouveaux lavages de 10 min chacun avec du Tween 0,2 % (dans PBS 1X), la membrane est incubée avec l'anticorps secondaire anti-lapin couplé à la peroxydase dilué à 1/2 000<sup>ème</sup> dans du PBS 1X pendant 45 min avec agitation. Après 4 derniers lavages de 10 min chacun avec du Tween 0,2 % (dans PBS 1X), la membrane est mise en présence de luminol (ECL) pendant 1 min, et la révélation (révélateur et bain d'arrêt) a lieu en chambre noire.

**CHAPITRE IV**  
**Premier article**

**CHAPITRE IV : Premier article****Application de la technique d'Hybridation Soustractive Suppressive  
à l'étude transcriptomique de *Riftia pachyptila*****1. Résumé**1.1 Contexte de l'étude

Le ver tubicole géant des sources hydrothermales profondes *Riftia pachyptila* se caractérise par sa symbiose avec des bactéries chimioautotrophes sulfo-oxydantes (Chapitre I). Logés à l'intérieur du corps de l'animal dans un organe interne spécialisé, le trophosome, ces symbiontes dépendent exclusivement de l'hôte pour leur approvisionnement en O<sub>2</sub>, CO<sub>2</sub> et H<sub>2</sub>S à partir du milieu environnant (Chapitre II). Néanmoins, si l'acquisition de ces composés par le panache branchial du ver puis leur acheminement jusqu'aux symbiontes ont été intensivement étudiés, certains points restent encore mal connus.

Nous avons utilisé une approche moléculaire globale pour tenter d'identifier des protéines impliquées dans les échanges de métabolites non mises en évidence par les méthodes d'études classiques. Les transcrits codant ces protéines spécifiques devraient être préférentiellement présents dans les tissus importants dans ces échanges (filaments branchiaux et le trophosome), directement impliqués dans le mode de vie symbiotique.

Par l'utilisation de l'hybridation soustractive suppressive (SSH), nous avons identifié les séquences exprimées dans ces tissus lorsque ces derniers sont comparés à un tissu de référence, la paroi du corps, non impliqué dans les échanges de métabolites chez cette espèce. Quatre banques ont été établies : i) branchie par rapport à la paroi du corps (BR-BW), ii) sa réciproque BW-BR, iii) trophosome par rapport à la paroi du corps (TR-BW), iv) sa réciproque BW-TR.



## 1.2 Principaux résultats

Nous avons obtenu un total de plus de 700 séquences avec une moyenne d'environ 45 ADNc différents par banque. Sur la totalité des ADNc obtenus, environ la moitié ont montré des homologies avec des séquences de fonction connue dans les bases de données. Pour les transcrits qui nous ont paru les plus intéressants, nous avons mené des études de PCR classique et PCR quantitative pour vérifier leur expression différentielle : ces résultats ont corroboré les résultats des banques. La méthode SSH s'est révélée efficace pour comparer des tissus différents. Ainsi, les transcrits de Major Vault Protein, précurseur Chitinase, cathepsine et ARN 16S se sont révélés plus abondants dans le tissu branchial que dans les autres tissus étudiés (trophosome et paroi du corps). Au contraire, les transcrits de myohémérythrine, récepteur T-cell et un transcrit non identifié sont davantage présents dans le trophosome par rapport aux deux autres tissus. A partir de l'obtention de plusieurs transcrits codant pour des protéines aux fonctions étroitement liées dans la banque spécifique de la branchie, nous avons formulé quelques hypothèses sur les possibles dégradation et renouvellement de protéines particulièrement importants dans ce tissu. Si sur l'ensemble de nos banques, quelques transcrits intéressants du point de vue symbiotique (tels que des transcrits d'anhydrase carbonique) ont été obtenus, d'autres transcrits, sans lien apparent avec le mode de vie symbiotique, étaient inattendus : des discussions bibliographiques sur ces points particuliers sont également menées.

## **2. Article publié dans BMC Genomics**

**Sanchez S.**, Hourdez S. and Lallier F. H., 2007. Identification of proteins involved in the functioning of *Riftia pachyptila* symbiosis by Subtractive Suppression Hybridization. *BMC Genomics*, **8**:337.

**Identification of proteins involved in the functioning of *Riftia pachyptila*  
symbiosis by Subtractive Suppression Hybridization**

Sophie Sanchez, Stéphane Hourdez and François H. Lallier\*

Equipe Ecophysiologie: Adaptation et Evolution Moléculaires,  
UMR 7144 CNRS-UPMC

Station Biologique, Place Georges Teissier, BP 74, 29682 Roscoff Cedex, France

\* corresponding author

Tel +33 2 9829 2311

Fax +33 2 9829 2324

E-mail [lallier@sb-roscoff.fr](mailto:lallier@sb-roscoff.fr)

Short title: SSH libraries to study a chemosynthetic symbiosis

Keywords: Siboglinidae; chemoautotrophy; subtractive libraries; quantitative PCR; cDNA sequences

## ABSTRACT

**Background:** Since its discovery around deep sea hydrothermal vents of the East Pacific Rise, the chemoautotrophic symbiosis between the vestimentiferan tubeworm *Riftia pachyptila* and its symbiotic sulfide-oxidizing  $\gamma$ -proteobacteria has been extensively studied for the last twenty years. However, studies on the tubeworm host were essentially targeted, biochemical approaches. We decided to use a global molecular approach to identify new proteins involved in metabolite exchanges and assimilation by the host. We used a Subtractive Suppression Hybridization approach (SSH) in an original way, by comparing pairs of tissues from a single individual. We chose to identify the sequences preferentially expressed in the branchial plume tissue (the only organ in contact with sea water) and in the trophosome (the organ housing the symbiotic bacteria) using the body wall as a reference tissue because it is supposedly not involved in metabolite exchanges in this species.

**Results:** We produced four cDNA libraries: i) body wall-subtracted branchial plume library (BR-BW), ii) and its reverse library, branchial plume-subtracted body wall library (BW-BR), iii) body wall-subtracted trophosome library (TR-BW), iv) and its reverse library, trophosome-subtracted body wall library (BW-TR). For each library, we sequenced about 200 clones resulting in 45 different sequences on average in each library (58 and 59 cDNAs for BR-BW and TR-BW libraries respectively). Overall, half of the contigs matched records found in the databases with good E-values. After quantitative PCR analysis, it resulted that 16S, Major Vault Protein, carbonic anhydrase (RpCAbr), cathepsin and chitinase precursor transcripts were highly represented in the branchial plume tissue compared to the trophosome and the body wall tissues, whereas carbonic anhydrase (RpCAtr), myohemerythrin, a putative T-Cell receptor and one non identified transcript were highly specific of the trophosome tissue.

**Conclusions:** Quantitative PCR analyses were congruent with our libraries results thereby confirming the existence of tissue-specific transcripts identified by SSH. We focused our study on the transcripts we identified as the most interesting ones based on the BLAST results. Some of the keys to understanding metabolite exchanges may remain in the sequences we could not identify (hypothetical proteins and no similarity found). These sequences will now have to be better studied by a longer -or complete- sequencing to check their identity, and then by verifying the expression level of the transcripts in different parts of the worm.

## BACKGROUND

The vestimentiferan annelid *Riftia pachyptila* lives around hydrothermal vents on the East Pacific Rise at 2600 meters-depth. It forms an obligatory symbiosis with sulfide-oxidizing bacteria that provide all its nutrition. Adult vestimentiferans lack a mouth, gut and anus [1]. Instead, they possess a specialized tissue, called trophosome, that contains the symbiotic bacteria. Their larvae however possess a digestive tract [2], and are devoid of symbiotic bacteria which they acquire from the environment. The acquisition of bacteria occurs through the skin, and the trophosome is established from mesodermal tissue. Then, apoptosis of some cells in the host epidermis occurs at the end of the colonization process [3].

Several studies focused on the functioning of this symbiosis. Previous biochemical and enzymatic studies addressed the uptake of hydrogen sulfide [4, 5] and the transport of both oxygen and hydrogen sulfide by the giant extracellular hemoglobin [6-8]. The diffusion of carbon dioxide through the branchial plume [9] and its subsequent conversion into bicarbonate through the activity of carbonic anhydrase [10, 11] were also shown. More recently, molecular techniques were used to better understand some aspects of the exchanges mechanisms in branchial plume and trophosome, such as the existence of a carbonic anhydrase transcript [11]. The sequencing of the whole genome of the symbiont of *Riftia pachyptila* is currently under progress (Horst Felbeck, personal communication) and a proteomic approach has been carried out on the symbiont [12] revealing previously unsuspected carbon fixation pathways. However, no global genomic work is planned on the host to date.

Identification of differentially-expressed transcripts (i.e. transcripts which differ in abundance between samples being compared) has been conducted for the last ten years on symbiotic interactions between rhizobia and legumes (for review see [13]) thanks to improved molecular approaches such as Subtractive Suppression Hybridization (SSH) for example. Morel and coworkers [14] constructed cDNA libraries by a SSH procedure and performed hybridizations on arrays between two compartments of the fungus *Paxillus involutus* living in symbiosis with the plant *Betula pendula*. These methods successfully identified differentially-expressed sequences in this ectomycorrhizal symbiosis, suggesting differences in metabolism between the two studied compartments [14]. SSH seems to be a quick and efficient method to rapidly obtain many specific sequences. It is a powerful method to enrich samples for

differentially expressed transcripts by combining steps of suppression and normalization prior to differential screening and that starting from very little material.

A transcriptome analysis of a marine cnidarian-dinoflagellate symbiosis using microarrays to compare aposymbiotic and symbiotic stages of the host *Anthopleura elegantissima* revealed the existence of key genes involved in the maintenance of the symbiosis [15].

In *Riftia pachyptila*, aposymbiotic larvae/post-larvae are very small (less than 100  $\mu\text{m}$ ) and very difficult to obtain. Besides, the host cannot be kept alive without its symbionts. Therefore, comparison between aposymbiotic and symbiotic states in *R. pachyptila* cannot be considered at present. Previous studies on the host were only targeted molecular studies and no global molecular analysis has been carried out on the host *Riftia pachyptila* to date. The aim of the present study was to identify host transcripts which could be involved in metabolite exchanges in the branchial plume on the one hand, and in metabolite exchanges with the symbionts in the trophosome. We postulated that these specific protein-coding genes should be preferentially expressed in these two tissues that are directly involved in the symbiotic way of life. Instead of comparing a same tissue in two physiological states, we used SSH in an original way i.e. by comparing pairs of tissues from a single individual. The subtracted libraries obtained should therefore be enriched in specific sequences compared to a conventional library without any subtraction procedure. In theory, identical sequences between key tissue and the reference tissue (housekeeping genes sequences in particular) should be eliminated by the subtractive suppression hybridization. Only tissue specific sequences should be recovered in each library. We also maximized the chances of obtaining new sequences using the normalization procedure which increases the amount of rare transcripts during the SSH procedure.

## RESULTS

### *General results of sequencing*

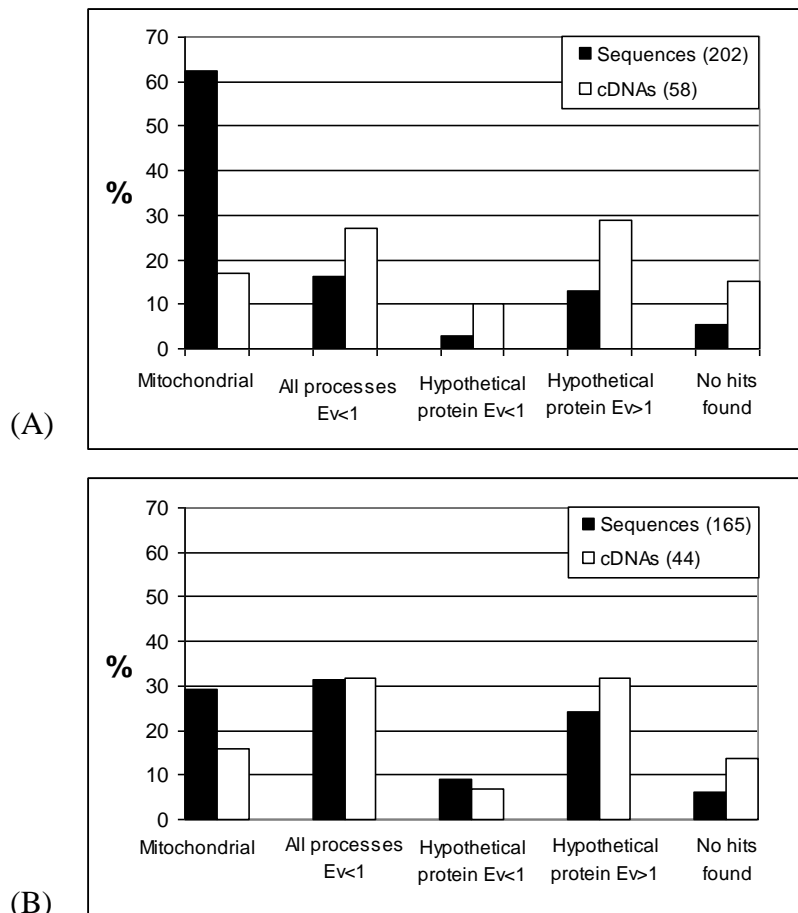
Global results including the total number of obtained sequences, contigs, singletons, and redundancy rates are given in Table 1 for all the libraries (the body wall-subtracted branchial plume library (BR-BW), the branchial plume-subtracted body wall library (BW-BR), the body wall-subtracted trophosome library (TR-BW) and the trophosome-subtracted body wall library (BW-TR)).

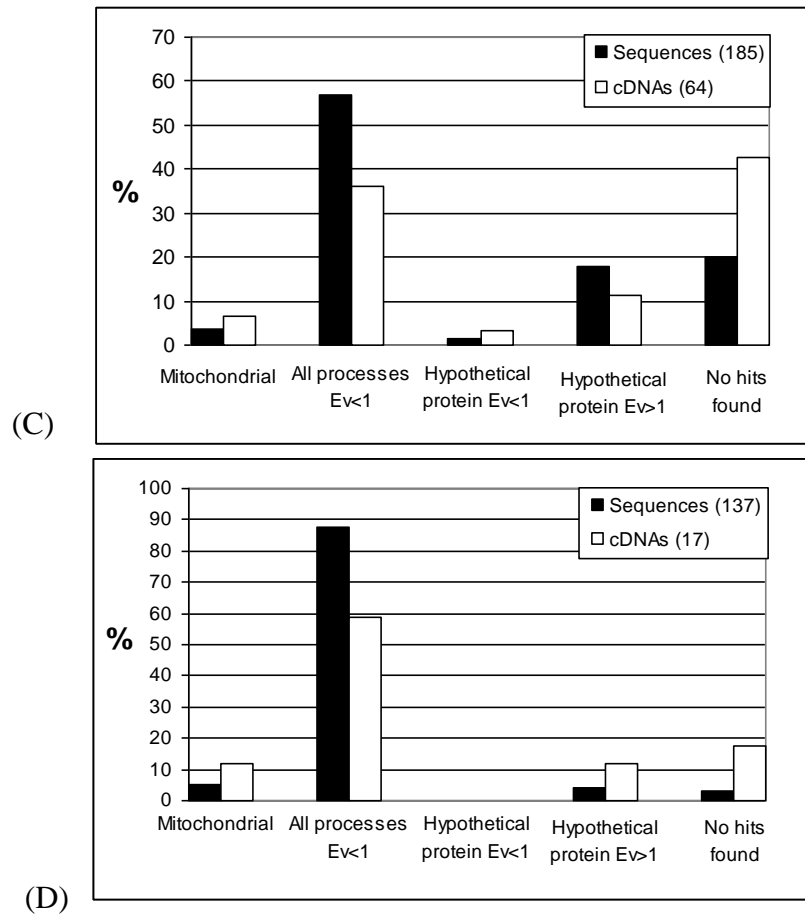
**Table 1: Overall statistics based on the analysis of each library.**

	BR-BW	BW-BR	TR-BW	BW-TR
Number of fragments sequenced	202	165	185	137
Number of cDNAs (contigs)	58	45	59	17
Number of contigs formed by one sequence (singletons)	38	17	36	6
Redundancy rate	81.2	89.7	80.5	95.6

The redundancy rates of the libraries range from 80.5 to 95.6 %. This indicates that additional sequencing should bring few or no new sequences. The sequences obtained were assembled into 58, 45, 59 and 17 different sequences (each putatively representing one cDNA) respectively for each library. Of those, 38, 17, 36 and 6 appeared as singletons, respectively.

Figure 1 shows the proportion of sequences with homology or not in the GenBank protein database for the four libraries: BR-BW (Fig. 1A), BW-BR (Fig. 1B), TR-BW (Fig. 1C) and BW-TR (Fig. 1D).





**Fig. 1:** Proportion of sequences and contigs separated into 5 main categories (mitochondrial, all processes sequences (E-value<1), hypothetical sequences (E-value<1), hypothetical sequences (E-value>1) and no similarity found). (A) Results of the BR-BW cDNA library. (B) Results of the BW-BR cDNA library. (C) Results of the TR-BW cDNA library. (D) Results of the BW-TR cDNA library.

These sequences are split into different categories: 1) mitochondrial sequences, 2) all processes sequences (with E-value < 1), 3) hypothetical sequences (with E-value < 1), 4) hypothetical sequences (with E-value > 1) and 5) no similarity found. If we consider that the two last categories cannot improve our knowledge, it remains that the proportions of cDNAs which matched with good homologies scores in GenBank database are 54.2 % in the BR-BW, 54.6 % in the BW-BR, 45.9 % in the TR-BW and 70.6 % in the BW-TR libraries. The choice of E-value=1 as the threshold to assess the degree of similarity to protein sequences in GenBank database is arbitrary, given that sequences obtained are usually short and that there is little molecular data on vestimentiferans or annelids in general.

### *Specific sequences obtained from each subtracted library*

#### *Subtracted branchial plume library (BR-BW)*

Fig. 1A shows a strikingly high proportion of mitochondrial sequences (62.4%) compared to what is observed in the three other libraries. Table 2 shows the sequences with the best E-values for the sequences obtained from the branchial plume library.

**Table 2: List of contigs with best E-values obtained for the BR-BW library sequences.**

Contig (BRbwC)	Number of sequences per contig	Putative identification on Blastx	E-value	GenBank accession number*
<b>Mitochondrial sequences</b>				
1	78	rRNA 16S large subunit 2390-2967	0	AY741662.1
2	32	rRNA 16S large subunit 2962-3499	0	AY741662.1
3	7	cytochrome c oxidase (ccox) subunit I 12-729	$4.10^{-122}$	AAU20751.1
4	2	ccox subunit I 735-1548	$2.10^{-124}$	AAU20751.1
5	1	ccox subunit Vb	0.084	<i>EF648457</i>
6	1	ccox subunit VIc	$6.10^{-13}$	<i>EF648477</i>
7	1	ccox <i>Proteus vulgaris</i>	0.19	<i>EF648458</i>
8	2	ccox subunit Va	$4.10^{-28}$	<i>EF648459</i>
9	1	ccox subunit Vb	0.11	<i>EF648460</i>
<b>All processes sequences (E-value&lt;1)</b>				
10	2	ATP F0 c subunit	$3.10^{-16}$	<i>EF648461</i>
11	1	ATP F1 $\beta$ subunit	$4.10^{-93}$	<i>EF648462</i>
12	11	carbonic anhydrase	$3.10^{-12}$	<i>EF490380</i>
13	4	Major Vault Protein	$6.10^{-20}$	<i>EF648463</i>
14	1	chitinase precursor	$3.10^{-8}$	<i>EF648464</i>
15	3	cathepsin L-like	0.004	<i>EF648465</i>
16	2	BTG1 protein	$2.10^{-21}$	<i>EF648466</i>
17	1	$\alpha$ -tubuline	$3.10^{-5}$	<i>EF648467</i>
18	1	hydroxylamine reductase	$3.10^{-20}$	<i>EF648468</i>
19	1	transcription repair coupling factor	0.67	<i>EF648469</i>
20	1	transcriptional regulator	0.66	<i>EF648470</i>
21	1	valosin containing protein	0.013	<i>EF648471</i>
22	1	Rab5 GDP/GTP exchange factor	0.65	<i>EF648472</i>
23	1	serine protease	0.56	<i>EF648473</i>
<b>Hypothetical protein (E-value&lt;1)</b>				
24	1	super cystein-rich protein	$5.10^{-8}$	<i>EF648474</i>
25	1	hypothetical protein	$2.10^{-22}$	<i>EF648475</i>
26	1	hypothetical protein	$2.10^{-5}$	<i>EF648476</i>
<b>Unknown sequences (E-value&gt;1 or No Hits Found)</b>				
27 to 58	GenBank accession numbers <i>EF648478</i> to <i>EF648509</i>			

\* new sequences are italicized

First of all, among the most redundant clones, we found a 16S ribosomal mitochondrion sequence which appeared highly redundant (contigs 1 and 2, corresponding to two fragments of mt16S, contain 78 and 32 sequences, respectively). High homologies scores were obtained for several, in particular contigs 12 (carbonic anhydrase), 13 (Major Vault Protein), 14 (chitinase precursor), 15 (cathepsin L-like), 16 (BTG1 protein), 17 ( $\alpha$ -tubuline), 18 (hydroxylamine reductase), and 24 (super cystein rich protein). The carbonic anhydrase cDNA obtained in the branchial plume (hereafter referred to as RpCAbr) was different from



the one already sequenced from a trophosome cDNA sample [11] (hereafter referred to as RpCAtr). The RpCAbr full-length sequence [GenBank:EF490380, [16]] is only 66% identical in amino acids to RpCAtr [GenBank:AJ439711, [11]].

### ***Subtracted body wall library (BW-BR)***

The sequencing of the reciprocal library (i.e. BW-BR, Table 3) revealed only 3 cDNAs in common with the BR-BW library: the two sections of the mitochondrial rRNA 16S large subunit (although they form a smaller proportion of the sequences), and the branchial carbonic anhydrase (RpCAbr) cDNAs.

**Table 3: List of contigs with best E-values obtained for the BW-BR library sequences.**

Contig (BWbrC)	Number of sequences per contig	Putative identification on Blastx	E-value	GenBank accession number*
<b><i>Mitochondrial sequences</i></b>				
1	10	rRNA 16S large subunit 2390-2967	0	AY741662.1
2	10	rRNA 16S large subunit 2962-3499	0	AY741662.1
3	1	cytochrome c oxidase (ccox) subunit I 293-471	$8.10^{-65}$	AAU20751.1
4	12	ccox subunit I <i>Littorina</i>	$2.10^{-77}$	EF648510
5	9	ccox subunit I <i>Paralvinella</i>	$3.10^{-172}$	EF648511
6	2	ccox subunit II	$6.10^{-74}$	AAU20752.1
7	4	ccox subunit IV <i>Urechis</i>	$4.10^{-50}$	EF648512
<b><i>All processes sequences (E-value&lt;1)</i></b>				
8	1	carbonic anhydrase	$3.10^{-12}$	EF490380
9	9	exoskeleton $\beta$ -chitin-specific binding	$1.10^{-43}$	AAF76890
10	9	exoskeleton $\beta$ -chitin-specific binding (2)	$1.10^{-7}$	EF648513
11	6	galaxin	$2.10^{-14}$	EF648514
12	7	myosin regulatory light chain 102-576	0	AF173680
13	4	myosin regulatory light chain <i>Eisenia</i>	$3.10^{-12}$	EF648515
14	2	extracellular hemoglobin linker	0.39	AM000033
15	2	intracellular globin	$2.10^{-9}$	EF648516
16	6	troponin C	$2.10^{-13}$	EF648517
17	2	similar to fraser syndrome 1 isoform 1	0.007	EF648518
18	1	sugar transporter <i>Arabidopsis</i>	0.79	EF648519
19	1	putative integrin	0.14	EF648520
<b><i>Hypothetical protein (E-value&lt;1)</i></b>				
20	13	hypothetical protein	0.1	EF648521
21	1	hypothetical protein	0.62	EF648522
22	1	unamed	0.13	EF648523
23	1	hypothetical protein	0.047	EF648524
24	1	hypothetical protein	0.002	EF648525
<b><i>Unknown sequences (E-value&gt;1 or No Hits Found)</i></b>				
25 to 45		GenBank accession numbers EF648526 to EF648546		

\* new sequences are italicized

Although it was not the main target, this library yielded interesting sequences involved in the formation of the tube and therefore expected to be specific of the body-wall (Table 3),

including *Riftia pachyptila* exoskeleton  $\beta$ -chitin-binding transcript (contig 9) already sequenced by Chamoy and coworkers [17]. We also obtained a different exoskeleton  $\beta$ -chitin-binding transcript (contig 10) with a best homology score with the previously sequenced one [17]. Surprisingly, contig 11 showed a high homology with galaxin, a protein involved in the formation of the calcified exoskeleton of the coral *Galaxea fascicularis*. Two transcripts coding for respiratory proteins were also found in this library: a new extracellular hemoglobin linker (contig 14) [18] which matches with *Sabella spallanzanii* linker chain sequence, and an intracellular globin (contig 15).

### Subtracted trophosome library (TR-BW)

The trophosome library yielded much less identifiable sequences (Table 4).

**Table 4: List of contigs with best E-values obtained for the TR-BW library sequences.**

Contig (TRbwC)	Number of sequences per contig	Putative identification on Blastx	E-value	GenBank accession number*
<b>Mitochondrial sequences</b>				
1	3	rRNA 16S large subunit 2542-2964	0	AY741662.1
2	2	rRNA 16S large subunit 2962-3499	0	AY741662.1
3	1	cytochrome c	$4.10^{-16}$	<i>EF648557</i>
4	1	cytochrome c oxidase subunit I <i>Arabella</i>	$3.10^{-93}$	<i>EF648558</i>
5	1	NADH dehydrogenase	$5.10^{-28}$	<i>EF648559</i>
<b>All processes sequences (E-value&lt;1)</b>				
6	5	carbonic anhydrase	$3.10^{-123}$	AJ439711
7	1	carbonic anhydrase 3'UTR	$3.10^{-136}$	AJ439711
8	19	hemoglobin (Hb) A1 chain (longer sequencing)	$9.10^{-20}$	<i>EF648449</i>
9	28	small similar part of Hb A1 chain (2)	$8.10^{-6}$	<i>EF648560</i>
10	1	small similar part of Hb A1 chain (3)	0.047	<i>EF648450</i>
11	7	small similar part of Hb A1 chain (4)	$7.10^{-3}$	<i>EF648451</i>
12	1	new Hb A2 chain (hit with <i>Arenicola</i> )	$9.10^{-16}$	<i>EF648561</i>
13	9	Hb B chain	$8.10^{-7}$	P80592
14	1	new Hb B1 chain (hit with <i>Lamellibrachia</i> )	$2.10^{-11}$	<i>EF648568</i>
15	2	Hb B2 chain	$2.10^{-12}$	AAW78354
16	1	new extracellular Hb linker (hit with <i>Sabella</i> )	$2.10^{-8}$	<i>EF648562</i>
17	5	myohemerythrin	$2.10^{-29}$	<i>EF648563</i>
18	2	T-cell receptor	0.58	<i>EF648564</i>
19	2	small nuclear ribonucleoprotein	$6.10^{-22}$	<i>EF648565</i>
20	2	oxidoreductase molybdopterin binding	0.32	<i>EF648566</i>
21	1	ubiquinol cytochrome reductase	$2.10^{-22}$	<i>EF648567</i>
22	1	putative calmodulin	0.034	<i>EF648452</i>
<b>Hypothetical protein (E-value&lt;1)</b>				
23	1	hypothetical protein	0.33	<i>EF648569</i>
24	1	hypothetical protein	0.46	<i>EF648570</i>
25	2	hypothetical protein <i>Oryza</i>	0.087	<i>EF648571</i>
26	1	hypothetical protein	0.19	<i>EF648572</i>
<b>Unknown sequences (E-value&gt;1 or No Hits Found)</b>				
27	24	unknown sequence		<i>EF648581</i>
28 to 59		GenBank accession numbers <i>EF648573</i> to <i>EF648580</i> , <i>EF648582</i> to <i>EF648601</i> , <i>EF648454</i> to <i>EF648456</i> , <i>EF648453</i>		

\* new sequences are italicized

Sequences obtained from this library were compared to the unpublished genomic sequences of the symbiont (with Horst Felbeck permission) in order to verify that they were host-specific, and not a contamination from the symbionts that are contained in this tissue. Noticeably, among the identifiable sequences, we recovered the previously sequenced carbonic anhydrase transcript (RpCAtr) [11], and transcripts coding for a large number of globin chains (contigs 8-15), for a hemoglobin linker (contig 16), and for a myohemerythrin (contig 17). Among the cDNAs with a significant blast value (all processes sequences (E-value<1)), more than 56 % are respiratory pigment proteins transcripts. These latter are partial fragments of the already known A1, A2, B1, B2 chains of the giant extracellular hemoglobin of *Riftia pachyptila*, and a probably new A2 (contig 12) and B1 (contig 14) chains. The extracellular hemoglobin linker identified here (contig 16) is different from the one identified in the BW-BR library. This brings the number of known partial linkers cDNAs up to three out of the four known types of linker chains [19]. A sequence coding for a serine-threonine rich protein was also found and matches with a T-cell receptor protein sequence (contig 18, E-value=0.58). Among the unknown sequences, the most abundant contig was composed of 24 sequences (contig 27).

### ***Subtracted body wall library (BW-TR)***

Overall, the general results found for the BW-TR library (Table 5) are similar to those found for the BW-BR library (Table 3) although with a lower number of contigs.

**Table 5: List of contigs with best E-values obtained for the BW-TR library sequences.**

Contig (BWtrC)	Number of sequences per contig	Putative identification on Blastx	E-value	GenBank accession number*
<b><i>Mitochondrial sequences</i></b>				
1	2	rRNA 16S large subunit 2433-2964	0	AY741662.1
2	5	rRNA 16S large subunit 2962-3499	0	AY741662.1
<b><i>All processes sequences (E-value&lt;1)</i></b>				
3	24	exoskeleton $\beta$ -chitin-specific binding	$5.10^{-61}$	AAF76890
4	59	exoskeleton $\beta$ -chitin-specific binding (2)	$5.10^{-34}$	<i>EF648513</i>
5	20	galaxin	$6.10^{-32}$	<i>EF648514</i>
6	7	myosin regulatory light chain	$1.10^{-132}$	AF173680
7	3	ficolin 2 precursor 108-211	$7.10^{-26}$	<i>EF648547</i>
8	2	putative integrin isoform 2	0.33	<i>EF648548</i>
9	1	actin	$6.10^{-21}$	<i>EF648549</i>
10	1	CCAAT-box DNA-binding	0.28	<i>EF648550</i>
<b><i>Unknown sequences (E-value&gt;1 or No Hits Found)</i></b>				
11 to 17	GenBank accession numbers <i>EF648546</i> , <i>EF648551</i> to <i>EF648556</i>			

\* new sequences are italicized

However, given the very large number of exoskeleton  $\beta$ -chitin-binding sequences (contig 3 and 4, comprising 24 and 59 sequences, respectively), we can suspect a less efficient normalization for these transcripts. The same new exoskeleton  $\beta$ -chitin-binding transcript, galaxin, and myosin chains were found in this library.

### ***Checking the subtraction procedure and the specificity of the libraries***

We only found 3 common cDNAs between the BR-BW and BW-BR libraries and 2 common cDNAs between the TR-BW and BW-TR ones. We were able to assess the degree of successful subtraction for several transcripts by performing regular PCR on subtracted and unsubtracted poly-A cDNA pools. Some results are shown in Fig. 2.

The profiles of abundant tissue-specific transcripts are presented in Fig. 2A (exoskeleton  $\beta$ -chitin-binding transcript) and 2B (galaxin transcript). The exoskeleton  $\beta$ -chitin-binding transcript was clearly present in the body wall tissue (Fig. 2A) in both unsubtracted cDNA sample and subtracted BW-BR cDNA sample (before and after SSH procedure respectively). The same profile was obtained for the galaxin transcript (Fig. 2B).

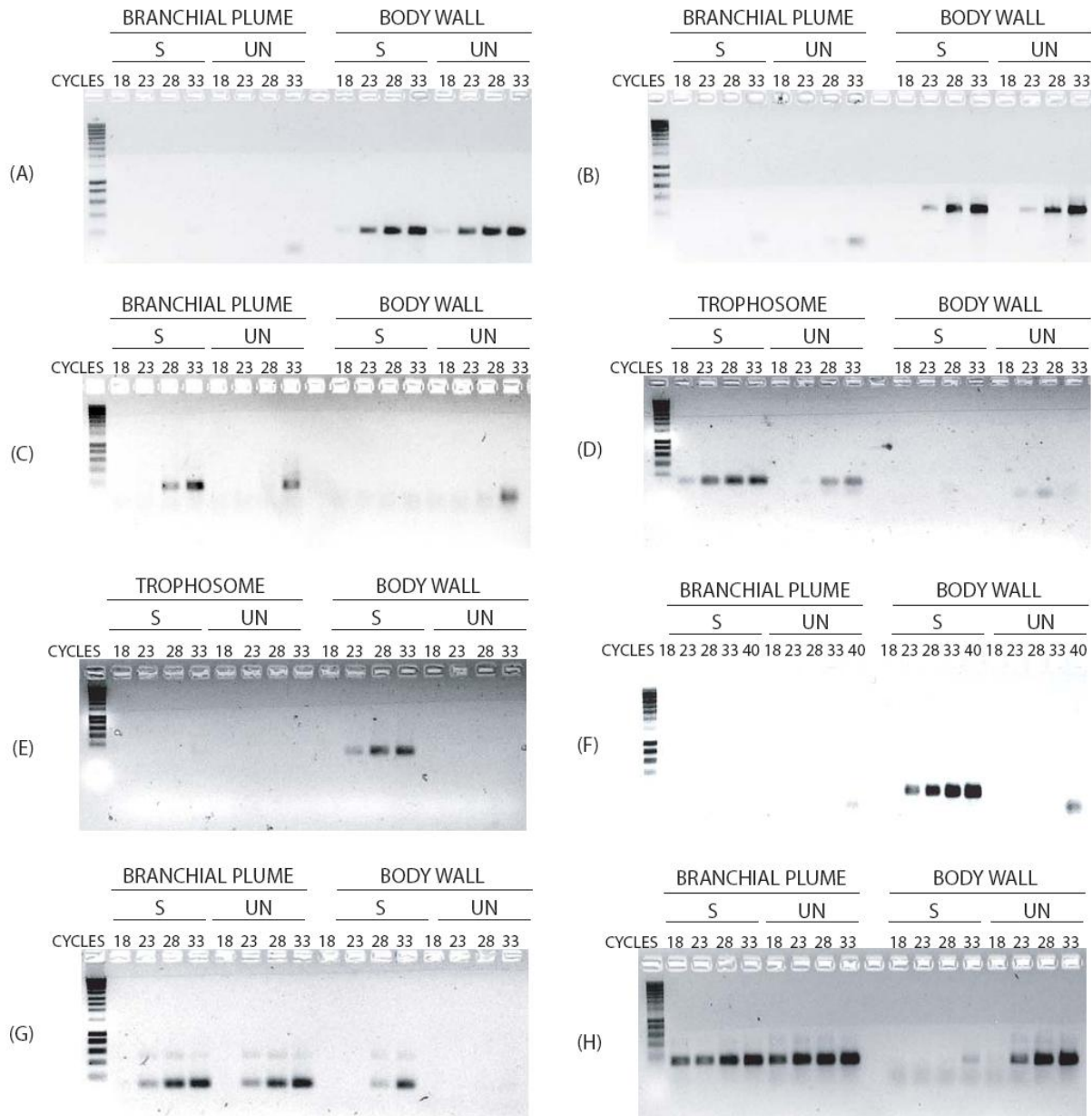
Typical profiles of enrichment of transcripts after SSH procedure are shown in Fig. 2C, D. The chitinase precursor sequence is enriched by the SSH, appearing on agarose gel after 28 cycles instead of 33 cycles on the branchial plume cDNA (Fig. 2C) and RpCAtr sequence amplification is visible after 18 cycles instead of 28 cycles from the trophosome cDNA before the SSH procedure (Fig. 2D).

Fig. 2E and 2F show profiles of amplification of rare transcripts. RpCAbr amplification can be seen on the BW-TR cDNA pool after 23 cycles and not from the body wall unsubtracted cDNA, even at 33 cycles (Fig. 2E). Amplification of the intracellular globin sequence can be seen on the BW-BR cDNA pool after 23 cycles and not from the body wall unsubtracted cDNA, even after 40 cycles (Fig. 2F).

Fig. 2G shows results of the amplification profile of MVP. This transcript appeared abundant in one tissue (the branchial plume) and rare in another (the body wall). Only the subtraction procedure allowed its detection in the body wall (not detected in the unsubtracted sample, even after 33 cycles).

Finally, Fig. 2H illustrates the difference of SSH efficiency of a same transcript in two different subtraction procedures (BR-BW and BW-BR). A successful amplification of cytochrome c oxydase I (ccox I) was obtained in both branchial plume and body wall unsubtracted cDNA pools after 18 and 23 cycles respectively. In contrast, no amplification

was observed in the BW-BR subtracted cDNA pool, in contrast with the BR-BW one. It seems that the subtraction was successful in the BW-BR library whereas this transcript could not be successfully subtracted in the BR-BW library after SSH procedure.



**Fig. 2:** Typical PCR profiles obtained after amplification of fragments of interesting cDNAs. S = subtracted sample; UN = unsubtracted sample. (A) and (B) Abundant tissue-specific transcripts: exoskeleton  $\beta$ -chitin-binding transcript (A) and galaxin transcript (B). (C) and (D) Transcripts enriched after SSH procedure: chitinase precursor transcript (C) and RpCAtr (D). (E) and (F) Rare transcripts enriched after SSH procedure: RpCAbr transcript (E) and intracellular globin transcript (F). (G) Abundant transcript in one tissue and rare in other tissue: MVP transcript. (H) Non equally subtracted transcript: cytochrome c oxidase subunit I transcript. The faint bands appearing at a smaller size than expected in some wells are interpreted as non-specific amplification (possible primer dimerization under specific conditions).

**Relative expression levels of some target genes over the three types of tissues**

We used quantitative PCR on some transcripts to further assess tissue specificity and gain data on the relative level of expression in each tissue starting from total cDNA. For all studied transcripts, PCR were performed from different initial amount of total cDNA in order to construct standard curves. The equations of the curves are reported in Table 6.

**Table 6: Equations of the standard curves obtained by amplification from total cDNA samples of branchial plume, trophosome and body wall tissues.** For each sequential dilution, PCR reactions were performed in triplicates. PCR efficiencies calculated from the slope of each curve are given.

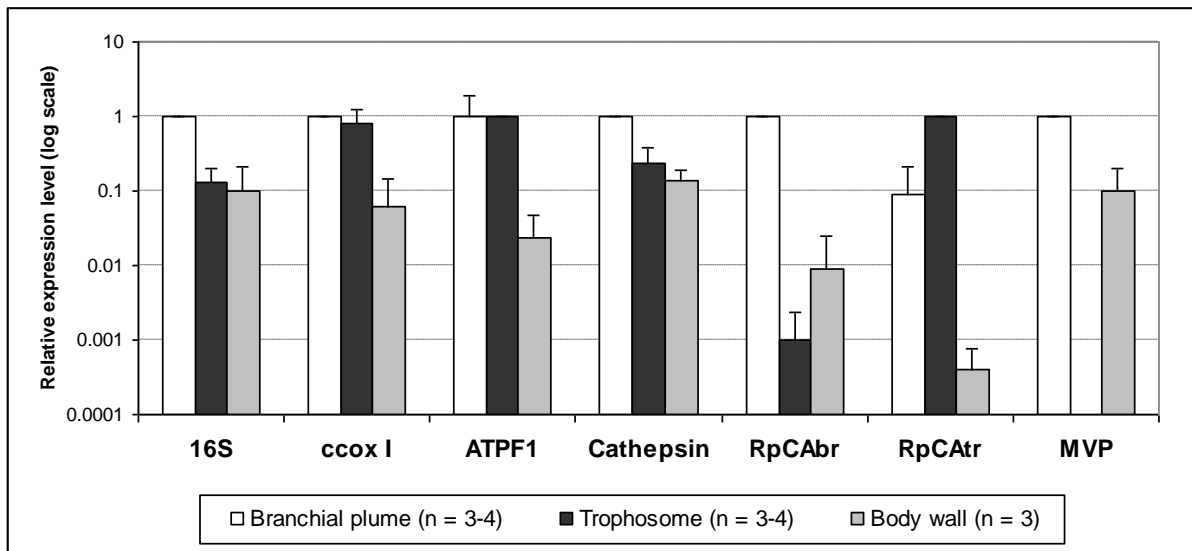
Transcript	Standard curve equation and efficiency (E) in the different tissues (calculated from one sample each time)					
	Branchial plume		Trophosome		Body wall	
18S	$y = -3.31x + 24.67$	E = 101 %	$y = -3.38x + 28.30$	E = 98 %	$y = -3.29x + 24.55$	E = 101 %
RpCAbr	$y = -3.24x + 39.13$	E = 103 %	$y = -3.37x + 51.40$	E = 98 %	$y = -3.30x + 44.66$	E = 101 %
RpCAtr	$y = -3.30x + 49.19$	E = 101 %	$y = -3.21x + 38.84$	E = 105 %	$y = -3.25x + 47.08$	E = 103 %
MVP	$y = -3.35x + 39.63$	E = 99 %	nd		$y = -3.24x + 42.51$	E = 104 %
Cathep	$y = -3.22x + 43.13$	E = 104 %	$y = -3.33x + 47.41$	E = 100 %	$y = -3.26x + 43.98$	E = 102 %
ChPr	$y = -3.37x + 39.57$	E = 98 %	nd		nd	
16S	$y = -3.31x + 29.86$	E = 101 %	$y = -3.17x + 38.77$	E = 107 %	$y = -3.43x + 34.80$	E = 96 %
ccoxI	$y = -3.54x + 36.05$	E = 92 %	$y = -3.42x + 40.74$	E = 96 %	$y = -3.46x + 38.51$	E = 94 %
ATPF1	$y = -3.43x + 41.54$	E = 96 %	$y = -3.41x + 44.95$	E = 96 %	$y = -3.34x + 43.14$	E = 99 %
MH	nd		$y = -3.24x + 41.08$	E = 103 %	nd	
TCR	nd		$y = -3.27x + 41.34$	E = 102 %	nd	
TRbwC27	nd		$y = -3.36x + 39.07$	E = 98 %	nd	

nd: Not detected

All PCR efficiencies (E in table 6), calculated based on the slopes of the curves, varied between 92 and 107%. For the dilution range we chose, we could only amplify the Major Vault Protein (MVP) transcript in the branchial plume and the body wall tissues, the chitinase precursor (ChPr) in the branchial plume tissue, and the myohemerythrin (MH), T-cell receptor (TCR) and contig 27 from TR-BW library (TRbwC27, unknown protein) transcripts in the trophosome tissue.

Once normalizations of the transcripts amplifications with 18S reference gene amplification were done, relative expression levels were calculated between the different tissues of a whole organism. The results from the analysis of several individuals are shown in Fig. 3. The 16S ribosomal gene has 7.5-fold and 10-fold higher expression levels in the branchial plume compared to the trophosome and the body wall, respectively. The ccox I and ATP synthase F1 transcripts were equally present in the branchial plume and trophosome tissues but comparatively less abundant in the body wall tissue (about 16-fold and 43-fold,

respectively). The new CA sequence (RpCAbr) is preferentially expressed in the branchial plume tissue compared to the trophosome (1,000-fold less expression) and the body wall (109-fold less expression) tissues. On the opposite, the RpCAtr transcript was more abundant in the trophosome tissue than in the branchial plume (12-fold less expression) and the body wall (2,500-fold less expression) tissues. Relative quantification analyses also showed that cathepsine L-like genes are also up-regulated in the branchial plume tissue compared to the trophosome (about 4-fold less expressed) and the body wall (about 7-fold less expressed) tissues of the worm. The MVP transcript was nearly 10-fold more abundant in the branchial plume than in the body wall tissue but we could not detect it in the trophosome tissue total cDNA.



**Fig. 3:** Relative expression levels of ribosomal RNA 16S, ccox I, ATPF1, Cathepsin, RpCAbr, RpCAtr, and MVP transcripts in the branchial plume, trophosome and body wall tissues. For each transcript, the calibrator tissue was chosen as the tissue with the higher expression: the branchial plume was the calibrator for ribosomal RNA 16S, ccox I, Cathepsin, RpCAbr and MVP amplifications, the trophosome was the calibrator for RpCAtr and ATPF1 amplifications. (Note: The number of tissue replicates (n) ranges from 3 to 4 and corresponds to the number of intra-individual tissue pairs we had).

Relative expression calculations for ChPr, MH, TCR and TRbwC27 sequences could not be made because we could only generate standard curves from the branchial plume total cDNA (for ChPr transcript) or from the trophosome total cDNA (for MH, TCR and TRbwC27) which indicate that these transcripts are drastically tissue-specific.

## DISCUSSION

### *Use of the SSH method for the study of symbiosis*

We used SSH on different tissues from a single individual to look for genes involved in the functioning of the symbiosis because it is not possible to obtain aposymbiotic adult *Riftia pachyptila*. The body wall was used as a reference tissue to find specific proteins expressed in the gills (main metabolite exchange organ with the milieu) and in the trophosome (organ that houses the symbiotic bacteria). We then focused our attention on some chosen transcripts for a quantitative analysis. The remaining unidentified sequences (many of which could correspond to 3'UTR portions of cDNA) could prove interesting. Their future identification will require either a RACE approach or hybridization on a full length cDNA library.

### *Efficiency of the SSH method*

All the 10 sequences that were more closely studied by quantitative PCR showed differential expression in agreement with the subtractive libraries where they were found. A transcript obtained in a given library showed the highest expression in the expected tissue, as evidenced by checking the result with rapid PCR validation (Fig. 2) and from several individuals by quantitative PCR (Fig. 4). In addition, although constitutively expressed in all cells of all tissues, only one tubulin transcript sequence was obtained from the BR-BW library, and one actin sequence from the BW-TR library. This demonstrates the adequate subtraction of these common sequences. However, some sequences are sometimes highly represented, possibly indicating a subtraction that was not as efficient. *ccox I*, for example, could not be eliminated in the BR-BW cDNA pool, but this may be due to the fact that it was more expressed in the branchial plume than in the body wall tissue. Strangely however, we recovered the *RpCAbr* transcript from the BW-BR library (Table 3) although it was 109-fold less abundant in the body wall than in the branchial plume tissues (Fig. 4).

As noticed by Ji and coworkers [20], SSH PCR favors highly differentially expressed genes. From our quantitative PCR results, some transcripts showed such high differential expression (e.g. *RpCAbr* in the branchial plume compared to the other tissues, and *RpCAtr* in the trophosome compared to the other tissues). These authors suggest that the primary application of SSH PCR should be the detection of dramatic alteration of gene expression, as it is for example the case for gene expression profiling of two different tissues. Our use of SSH for comparing pairs of tissues seems very appropriate.



### *Proteins degradation and turnover in the branchial plume tissue*

Some transcripts were preferentially expressed in the branchial plume tissue. Relative quantification of the cathepsin transcript (a degradation enzyme found in lysosomes) revealed a more important expression in the branchial plume tissue, compared to the body wall (sheltered by the tube) and the trophosome. The plume is the only organ in direct contact with sea-water and thereby strongly exposed to hydrogen sulfide and other toxic molecules such as heavy metals which are abundant in the hydrothermal vent environment. Electron-dense organelles (EDO) seem to be very common in tissues of sulfide-adapted marine annelids. Such structures have previously been observed in both the *Riftia pachyptila* epidermal body wall [21] and the branchial plume organ (Ann Andersen, personal communication). Arp and coworkers [22] hypothesized that EDO structures could actually be secondary lysosomes. This could be in agreement with our results of cathepsin expression, found to be highest in this highly exposed tissue. Julian and coworkers [23] showed that even in sulfide-tolerant organisms like the annelid *Glycera dibranchiata*, sulfide exposure poisons the mitochondria, leading to depolarization that is not reversible. Lysosomes could degrade mitochondria that have been damaged by sulfide exposure. Besides, we also found a Rab5 GDP/GTP exchange factor (BRbwC22) which is a member of the Ras superfamily of GTPases. This protein, involved in vesicle trafficking, is known to be located in early endosomes which precede the formation of lysosomes.

Interestingly, another pathway of degradation could also been involved since we obtained a Valosin-Containing Protein (VCP, BRbwC21) which is required in ubiquitin-proteasome degradation [24]. The existence of two VCP transcripts has recently been demonstrated in an annelid, the earthworm *Eisenia fetida* [25]. Our sequence shows the best homology scores with the predicted VCP protein sequence of *Strongylocentrotus purpuratus* [Expect = 0.014] and the eVCP-1 isoform from *Eisenia fetida* [Expect = 0.025] which is ubiquitously expressed in this worm [25].

If EDOs do represent autophagic degradation of organelles as suggested by Arp and coworkers [22], rapid replacement of organelles should take place [26]. The high expression level of ribosomal 16S, an essential gene for the traduction of mitochondrial messenger RNAs into proteins, and the presence of some transcripts linked to transcription (BRbwC19 and 20, Table 2) is then consistent with a high protein turnover in this tissue.

Sulfide oxidation with the concomitant production of ATP by the mitochondria of the annelid *Arenicola marina* has been shown [27]. However, no protein sequence was found here that would suggest a similar property of *R. pachyptila* mitochondria.

### ***Hydroxylamine reductase protein***

Formerly known as prismane, the hydroxylamine reductase is a member of the Hybrid-Cluster Protein (HCP) family and is thought to play a role in nitrogen metabolism. It catalyses the reduction of hydroxylamine to form ammonia using NADH. In rat liver mitochondria, this enzyme is firmly attached to the mitochondrial membrane [28] and its activity can prevent hydroxylamine to inhibit mitochondrial respiration [29]. However, blasts of our sequence did not match with the few eukaryotic sequences available but resulted in 100 bacteria sequences hits. The best ones are those of the Actinobacteria *Salinispora arenicola* [Expect = 5e-19] and the  $\alpha$ -proteobacteria *Rhodospirillum rubrum* [Expect = 5e-19]. Because this sequence did not match any sequence in the *Riftia* symbiont genomic database, it could be contamination from bacteria living close to the branchial plume of the worm.

### ***Major Vault Protein gene expression***

The Major Vault Protein (MVP) (100 kDa) is the major protein component of vaults, ribonucleic particles of 13MDa. Some studies established that vaults could be involved in nucleocytoplasmic transport of ribosomes and/or mRNA [30]. This could be coherent with our results of 16S expression obtained on the branchial plume tissue in which probable high transcription levels of this protein occur. Other studies indicate the participation of MVP in drug resistance mechanisms where it could act as a nucleocytoplasmic and vesicular transporter of drugs and/or metabolites to transport them to exocytotic vesicles or proton pumps [31, 32]. It could be evidenced that MVP gene in *Mytilus edulis* was predominantly expressed in epithelia-rich tissues such as the gills and digestive gland and could be involved in multixenobiotic resistance [33]. In our study, MVP transcript is preferentially expressed in the branchial plume tissue compared to the body wall, while no MVP transcript was detected in the trophosome samples. The presence of such a protein in the branchial plume tissue may be used to temporarily immobilize toxic molecules before they are processed.

### ***Chitinase gene expression***

Interestingly, a chitinase precursor was recovered as a branchial plume specific transcript. A previous report indicated chitinase activity in the opisthosome and branchial

plume of *R. pachyptila* [34]. Chitin is a major component of the tube of *R. pachyptila*, produced by specialized glands located in the body-wall and the vestimentum [35]. Chitinase activity was suggested to be involved in tube growth and tube shape modifications [34]. A chitinase sequence was recently discovered in the hydroid cnidarian *Hydractinia* [36] and a possible role of chitinase enzyme in pattern formation and allorecognition was suggested. Interestingly, the transcript was exclusively expressed in ectodermal tissues of the animal, and the authors also suggested a possible role in host defense against pathogens. Such a hypothesis could be interesting to explore given our quantitative PCR experiments because we only could amplify this transcript from cDNA from the branchial plume, the only organ in contact with the environmental sea water.

#### ***Tissue-specific expression of different carbonic anhydrases***

Our quantification analyses showed a higher abundance of the RpCAbr transcript in the branchial plume compared to the trophosome (present at very low levels) and the body wall tissues. In contrast, the RpCAtr transcript was very abundant in the trophosome compared to the branchial plume (medium levels) and the body wall tissues. Fluorescent *In Situ* Hybridization confirmed the co-expression of the two transcripts in the branchial plume in contrast with the trophosome where only one transcript could be detected [16]. An alignment of these translated CA cDNAs with vertebrate and non-vertebrate CA protein sequences revealed the conservation of most amino acids involved in the catalytic site, indicating that the two proteins are probably functional if the cDNAs are translated [16].

#### ***Myohemerythrin, T-cell receptor, and unidentified transcripts from the trophosome library***

A complete coding sequence obtained from the TR-BW library (contig 15) showed a very high homology score with a myohemerythrin sequence from the Sipuncula *Sipunculus nudus* [GenBank: CAG14944] (Expect = 1e-11). The complete *Riftia* sequence has an open-reading frame of 120 amino acids. Myohemerythrin is an oxygen-binding protein that participates in the storage of oxygen in muscles. Such a protein could be involved in the regulation of cadmium levels in the gut of the annelid *Nereis diversicolor* [37]. In *Hirudo medicinalis*, it would have indirect antibacterial properties by regulating free iron availability to deprive bacteria of iron essential for their growth [38].

Both TCR and TRbwC27 cDNAs showed specific expression in the trophosome tissue where they could be essential. The TCR transcript first caught our attention because it matched a sequence fragment coding for a T-cell receptor, which is a complex of integral

membrane proteins that participates in the activation of T-cells in response to the presentation of an antigen. The trophosome is mostly composed of bacteriocytes which house bacterial cells in intracellular vacuoles and cellular recognition may be very important for the functioning of this tissue. As for the TRbwC27 it is a large fragment of 273 nucleotides which is highly represented in our subtracted library but did not reveal reliable Blastx homology E-values.

## METHODS

### *Animals and sampling*

Specimens of *Riftia pachyptila* were collected at the Rehu Marka (17°25'S – 113°12'W), Susie and Miss WormWood (17°35'S - 113°14'W) sites at 2600 meters-depth along the South East Pacific Rise during the BIOSPEEDO 2004 cruise. For each individual, parts of the branchial plume, trophosome and body wall tissues were isolated on ice, placed in RNAlater (Ambion) for 24 h at 4°C and then frozen in liquid nitrogen.

### *RNA extraction*

Plume, trophosome and body wall tissue samples were ground individually in liquid nitrogen under RNase-free conditions. For each tissue, total RNA was extracted using the RNAlater buffer (Eurobio) following the manufacturer's instructions. Then, both for library constructions and for complete sequencing, messenger poly-A RNAs were purified using the oligo-dT resin column of the mRNA Purification Kit (Amersham).

### *Construction of subtractive tissue-specific cDNA libraries*

Libraries were constructed from tissues taken from a single individual, thereby representing a single organism transcriptome. A total of 4 libraries were produced: branchial plume vs. body wall subtracted library (and its reciprocal) and trophosome vs. body wall subtracted library (and its reciprocal). For all tissue pairs, cDNA synthesis as well as subtractive suppressive hybridization (SSH, [39, 40]), including steps of adaptor ligation, subtractive hybridization, and selective amplification were performed following the protocol of the Clontech PCR-Select<sup>TM</sup> cDNA Subtraction Kit (BD Biosciences).

In the BR-BW library, SSH was performed to produce a cDNA library enriched in branchial plume specific transcripts. The tester sample was the cDNA population from the

branchial filaments (BR) and the driver sample was the cDNA population from the body wall tissue (BW). In the BW/BR library the tester and driver samples were reversed, and SSH was performed to produce a cDNA library enriched in body wall specific transcripts. In the TR-BW library, the tester sample was the cDNA population from the trophosome (TR) and the driver sample was the cDNA population from the body wall tissue (BW). In the BW-TR library, the tester and driver samples were reversed.

### *Cloning and sequencing*

For each SSH procedure, the whole amplification product was cloned into the TOPO®-TA cloning vector (Invitrogen), producing a range of cDNA fragment sizes. Nearly 200 cDNA fragments were sequenced for each library. Plasmid DNA from individual colonies were purified with the FlexiPrep kit (Amersham) and used in a dye-primer cycle sequencing reaction with T3 or T7 universal primers and the Big Dye® Terminator V3.1 Cycle Sequencing kit (Applied Biosystems). Reactions were then run on a 16-capillary 3130 Applied Biosystems sequencer.

### *Sequence analysis and homology search*

Most of chromatograms obtained after sequencing were treated with PHRED [41] and Seqclean software (TGIR, the Institute for Genomic Research, Rockville, MD, USA) to remove vector and adaptors sequences. Progressively, additional sequences were treated manually. Clustering was performed with the TGICL programs, Megablast and CAP3 [42]. Clusters and contigs were formed on the whole set of sequences and also individually for each of the four libraries. Contigs were then verified manually to detect possible chimeras. BLAST analyses of the cDNA libraries sequences were performed on the NCBI server. The assembled sequences were analyzed for homology with known sequences in databases using the BlastX and BlastN programs (<http://www.ncbi.nlm.nih.gov/BLAST/>) and also treated with the PhyloGena software [43] which combines both homology searching and phylogenetic reconstruction to verify the homology attributions.

The redundancy corresponds to the probability that a newly sequenced cDNA was previously obtained. Redundancy rate was calculated with the formula:

$R = (1 - (Nu / Nt)) \times 100$ , where Nu is the number of unique sequences and Nt is the total number of sequences.

### ***Validation of differential expression by simple PCR***

Unsubtracted and subtracted PCR samples (respectively before and after SSH procedure) were obtained as recommended in the Clontech PCR-Select<sup>TM</sup> cDNA Subtraction Kit (BD Biosciences). These samples were then diluted ten-fold in sterile milliQ water. For each transcript tested, PCRs were conducted on these diluted unsubtracted and subtracted samples with specific forward and reverse primers (Table 7). Each reaction mixture was composed of 1 µl of cDNA sample; 1.2 µl of specific forward primer (10 µM), 1.2 µl of specific reverse primer (10 µM), 22.4 µl of sterile water, 3 µl of 10X PCR reaction buffer, 0.6 µl of dNTP Mix (10 mM) and 0.6 µl of 50X Advantage cDNA Polymerase Mix (BD Biosciences). The following thermal cycling program was used for 33 cycles: 94°C for 30 s, 60°C for 30 s, and 68°C for 2 min. For each cDNA pool tested, a 5µl-aliquot was removed from the reaction mixture every 5 cycles, starting at the end of cycle 18.

### ***SYBR Green quantitative PCR***

#### *Reverse transcription (RT)*

Fresh RT reaction was done with a random primer on each total RNA sample (branchial plume, trophosome and body wall). The reaction mixture was composed of 2 µl of M-MLV 5X RT buffer; 0.5 µl of BSA (10 mg/ml), 1 µl of total RNA (1.24 µg/µl), 2.5 µl of dNTP (4 mM total), 1.5 µl of Random Primer 9 (Ozyme) (100 ng/µl), 3 µl DEPC water. The reaction mixtures were then incubated at 80°C for 5 minutes and placed on ice. M-MLV RT was added (1 µl) to each reaction mixture and all reactions were incubated at 42°C for 1 hour and finally placed on ice.

#### *Amplification*

Specific pairs of primers for some target genes (Table 7) were designed using the software Primer Express. 18S rRNA transcript was chosen as a reference gene for the normalization of expression data and was amplified with the 18h and 18L primers designed by Halanych and coworkers [44]. For amplifications, the Power SYBR Green PCR master mix (PE Applied Biosystems) was used in 23 µl reaction mixtures on a Chromo4<sup>TM</sup> System CFB-3240 (BIORAD). Amplification conditions were 40 cycles with the following profile: 95°C for 30 s, 60°C for 30 s, and 72°C for 1 min.

**Table 7: Primers sequences used for the transcripts amplifications.**

Amplification of transcripts by classical PCR		
Primers	Sequence	
ExosqF	5'	TGC AGG CGA TGC GAG TGC 3'
ExosqR	5'	GCT ACA ACA GCG GTT AGG 3'
GalaxinF	5'	ATT TCG TTT GCA ACA GCC 3'
GalaxinR	5'	CTT CCT CTG CAG CAC TGG 3'
ChPrF	5'	AAT TCT GAG ACC GGT GAC C 3'
ChPrR	5'	TCC AAG ACC GTG TTG AGC 3'
RpCAtrF	5'	TAC AAA GAT CCA ATC CAG C 3'
RpCAtrR	5'	ACG AGG ACG ACA CCT GG 3'
RpCAbrF	5'	TAC AAG GAT GCC ATT AGC 3'
RpCAbrR	5'	AGA GCA GCA GAC CTT ACG 3'
IntraGlobF	5'	GGA AAG GAC GTC GAC AGC 3'
IntraGlobR	5'	TGC TGC TTG GTT AGT CCC 3'
MVPF	5'	GAG AAC AGA ATG ACA TGG 3'
MVPR	5'	TTT CAC CTG CAT CTC GGG 3'
ccoxIF	5'	ACA GGT TTA GTA GCC ACT 3'
ccoxIR	5'	GTG TTG ATA TAG GAC TGG 3'
Amplification of transcripts by quantitative PCR		
Primers*	Sequence	
RpCAbrFq	5'	TGG TTT CAC CCC GTC GAA 3'
RpCAbrRq	5'	GGT CTG GTC TTT TCT CGC CAT A 3'
RpCAtrFq	5'	GCC AGG TGT CGT CCT CGT T 3'
RpCAtrRq	5'	TCA CAA ATG TCC AGT GCC AGT T 3'
16SFq	5'	CGT AAG ACT ATA GCT GGT TTT CCA AA 3'
16SRq	5'	TTA TCA AAG ATT TTT TCT TGG TTC ACT AAT 3'
MVPFq	5'	GAT TGA GAC AAC CAA GTT CAG GAA 3'
MVPRq	5'	CTG GCG ATT GCC TGA ATT G 3'
CathepFq	5'	TAC ATG GCC CGT AAT AAG GAC A 3'
CathepRq	5'	GCT GGC TTG TGA TGC AAC AC 3'
ccoxIFq	5'	CTA ATG GGA GGC TTC GGA AAC 3'
ccoxIRq	5'	AGG TGC CCC TAG CAT TAA AGG 3'
ATPF1Fq	5'	TGC AGG ACA TCA TTG CCA TC 3'
ATPF1Rq	5'	TGT CCT CCT GGG ACA ACT CG 3'
ChPrFq	5'	GTC GTC GGA ATG GCG AGT TA 3'
ChPrRq	5'	AGC GTT GCT GGC TGT TTT G 3'
MHFq	5'	AGA GGC ACA CCA ACA ACC GT 3'
MHRq	5'	CCC GAT TGG TTC ATC ACA GC 3'
TCRFq	5'	AAT CCG ACG TGG CGA TCA T 3'
TCRFq	5'	GGT CAT TGT TGT TGC CTG GG 3'
TRbwC27Fq	5'	CGA CGG TGG TAC CCC GTA TA 3'
TRbwC27Rq	5'	CCG CAA CCT TTG AAC CTC AG 3'

\*Primers designed by Primer Express software (ABI PRISM™)

### *Standard curves*

In order to estimate the relative expression levels of each transcript by the  $2^{-\Delta\Delta C_t}$  method, we calculated PCR efficiencies of the transcripts amplifications to verify they do not highly differ from the reference transcript (rRNA 18S) amplification. PCR were performed from a dilution range of total cDNA from each tissue (from one sample). Therefore, we performed PCR starting from total cDNA amounts ranging from 6.2 pg to 620 ng. Each PCR reaction was performed in triplicate. For each initial template quantity, we looked at the threshold cycle, the  $C_t$ , corresponding to the number of cycles required to reach a set quantity of amplified cDNA during the exponential phase. The standard curves were generated by plotting the log of the initial template concentration against the  $C_t$  generated for each dilution.

### *Data analysis*

For each transcript, the efficiency (E) was calculated from the slope (S) of the standard curve using the formula:

$$E = 10^{-1/S} - 1$$

Once differences between efficiencies of reference gene and target gene amplifications were approximately equal, we calculated the relative expression level for each gene analyzed. In each tissue, amplification of the target transcript was compared to the endogenous control amplification in order to get the normalized number of cycles (NNC):

$$NNC = C_t \text{ target} - C_t \text{ 18S}$$

Then, for relative quantification measurement, we used the  $2^{-\Delta\Delta C_t}$  method [45] for individuals for which we had analyzed at least one pair of tissues. Relative quantification results were obtained by comparing levels of expression with the calibrator tissue, the latter being chosen as the tissue for which the better expression was observed with the following calculation:

$$\text{Relative expression level} = 2^{-(NNC_{\text{sample}} - NNC_{\text{calibrator}})}$$

### **Authors contribution**

SS carried out the experimental work, the sequence analysis and drafted the manuscript. SH participated in sample acquisition, sequence analysis and helped to draft the manuscript. FHL conceived the study, participated in sample acquisition and coordination, and helped to draft the manuscript. All authors read and approved the final manuscript.



**Acknowledgements**

We wish to thank the captain and crew of the NO L'Atalante, the crew of the Nautilie submersible, and the chief scientist of the BioSpeedo cruise, Didier Jollivet. SS was supported by a grant from the Région Bretagne. We also thank the personnel of the Ouest-Genopole sequencing platform in Roscoff: we thank Morgan Perennou and Nathalie Gloaguen for the sequencing work. We are also are grateful towards Erwan Corre from the Service Informatique et Genomique in Roscoff for the generation of clusters and contigs. We would like to thank Horst Felbeck and Eric E. Allen for blasting some of our trophosome sequences against the symbionts sequence library in order to verify that they were of host origin.

## REFERENCES

1. Jones ML: **On the Vestimentifera, new phylum: Six new species, and other taxa, from hydrothermal vents and elsewhere.** *Bulletin of the Biological Society of Washington* 1985, **6**:117-158.
2. Jones ML, Gardiner SL: **Evidence for a transient digestive tract in Vestimentifera.** *Proc Biol Soc Wash* 1988, **101**(2):423-433.
3. Nussbaumer AD, Fisher CR, Bright M: **Horizontal endosymbiont transmission in hydrothermal vent tubeworms.** *Nature Letters* 2006, **441**:345-348.
4. Goffredi SK, Childress JJ, Desaulniers NT, Lallier FH: **Sulfide acquisition by the vent worm *Riftia pachyptila* appears to be via uptake of HS<sup>-</sup>, rather than H<sub>2</sub>S.** *Journal of Experimental Biology* 1997, **200**(20):2609-2616.
5. Goffredi SK: **Physiological Adaptations of the Hydrothermal Vent Tubeworm *Riftia pachyptila*; Including the Uptake of Carbon Dioxide and Hydrogen Sulfide, Unprecedented pH Regulation, and a Reverse Chloride Shift.** *Dissertation*. 1999.
6. Arp AJ, Childress JJ, Vetter RD: **The sulfide-binding protein in the blood of the vestimentiferan tube worm, *Riftia pachyptila*, is the extracellular haemoglobin.** *J Exp Biol* 1987, **128**:139-158.
7. Zal F, Leize E, Lallier FH, Toulmond A, Van Dorsselaer A, Childress JJ: **S-Sulfohemoglobin and disulfide exchange: the mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins.** *Proc Natl Acad Sci U S A* 1998, **95**(15):8997-9002.
8. Flores JF, Fisher CR, Carney SL, Green BN, Freytag JK, Schaeffer SW, Royer Jr WE: **Sulfide binding is mediated by zinc ions discovered in the crystal structure of a hydrothermal vent tubeworm hemoglobin.** *PNAS* 2005, **102**(8):2713-2718.
9. Goffredi S, Childress J, Desaulniers N, Lee R, Lallier F, Hammond D: **Inorganic carbon acquisition by the hydrothermal vent tubeworm *Riftia pachyptila* depends upon high external PCO<sub>2</sub> and upon proton-equivalent ion transport by the worm.** *Journal of Experimental Biology* 1997, **200**(Pt 5):883-896.
10. Kochevar RE, Govind NS, Childress JJ: **Identification and characterization of two carbonic anhydrases from the hydrothermal vent tubeworm *Riftia pachyptila* Jones.** *Molec Mar Biol Biotechnol* 1993, **2**(1):10-19.
11. De Cian MC, Bailly X, Morales J, Strub JM, Van Dorsselaer A, Lallier FH: **Characterization of carbonic anhydrases from *Riftia pachyptila*, a symbiotic invertebrate from deep-sea hydrothermal vents.** *Proteins* 2003, **51**:327-339.

12. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, Albrecht D, Robidart J, Bench S, Feldman RA *et al*: **Physiological Proteomics of the Uncultured Endosymbiont of *Riftia pachyptila***. *Science* 2007, **315**:247-250.
13. Lievens S, Goormachtig S, Holsters M: **A critical evaluation of differential display as a tool to identify genes involved in legume nodulation: looking back and looking forward**. *Nucleic Acids Research* 2001, **29**(17):3459-3468.
14. Morel M, Jacob C, Kohler A, Johansson T, Martin F, Chalot M, Brun A: **Identification of Genes Differentially Expressed in Extraradical Mycelium and Ectomycorrhizal Roots during *Paxillus involutus*-*Betula pendula* Ectomycorrhizal Symbiosis**. *Applied and Environmental Microbiology* 2005, **71**(1):382-391.
15. Rodriguez-Lanetty M, Phillips WS, Weis VM: **Transcriptome analysis of a cnidarian-dinoflagellate mutualism reveals complex modulation of host gene expression**. *BMC Genomics* 2006, **7**(23).
16. Sanchez S, Andersen AC, Hourdez S, Lallier FH: **Identification, sequencing, and localization of new carbonic anhydrase transcripts from the hydrothermal vent tubeworm *Riftia pachyptila***. *FEBS J* 2007, **In revision**.
17. Chamoy L, Nicolai M, Ravaux J, Quennedey B, Gaill F, Delachambre J: **A Novel Chitin-binding Protein from the Vestimentiferan *Riftia pachyptila* Interacts Specifically with  $\beta$ -Chitin. Cloning, expression, and characterization**. *The Journal of Biological Chemistry* 2001, **276**(11):8051–8058.
18. Chabasse C, Bailly X, Sanchez S, Rousselot M, Zal F: **Gene Structure and Molecular Phylogeny of the Linker Chains from the Giant Annelid Hexagonal Bilayer Hemoglobins**. *J Mol Evol* 2006, **63**:365–374.
19. Zal F, Lallier FH, Wall JS, Vinogradov SN, Toulmond A: **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* 1996, **271**:8869-8874.
20. Ji W, Wright MB, Cai L, Flament A, Lindpaintner K: **Efficacy of SSH PCR in isolating differentially expressed genes**. *BMC Genomics* 2002, **3**(12).
21. Menon J, Willsie JK, Tauscher A, Arp AJ: **Epidermal ultrastructure and implications for sulfide tolerance in six species of deep-sea polychaetes**. *Invertebrate Biology* 2003, **122**(4):334-346.
22. Arp AJ, Menon JG, Julian D: **Multiple mechanisms for sulfide tolerance in *Urechis caupo***. *Amer Zool* 1995, **35**:132-144.

23. Julian D, April KL, Patel S, Stein JR, Wohlgemuth SE: **Mitochondrial depolarization following hydrogen sulfide exposure in erythrocytes from a sulfide-tolerant marine invertebrate.** *The Journal of Experimental Biology* 2005, **208**:4109-4122.
24. Dai RM, Li C-CH: **Valosin-containing protein is a multiubiquitin chain-targeting factor required in ubiquitin–proteasome degradation.** *Nature Cell Biology* 2001, **3**:740-744.
25. Suzuki T, Honda M, Matsumoto S, Stürzenbaum SR, Gamou S: **Valosine-containing proteins (VCP) in an annelid: Identification of a novel spermatogenesis related factor.** *Gene* 2005, **362**:11-18.
26. Wohlgemuth S, Arp A, Bergquist D, Julian D: **Rapid induction and disappearance of electron-dense organelles following sulfide exposure in the marine annelid *Branchioasychis americana*.** *Invertebrate Biology* 2007, **in press**.
27. Völkel S, Grieshaber MK: **Sulphide oxidation and oxidative phosphorylation in the mitochondria of the lugworm *Arenicola marina*.** *The Journal of Experimental Biology* 1997, **200**:83-92.
28. Bernheim MLC: **The Hydroxylamine Reductase of Mitochondria.** *Archives of Biochemistry and Biophysics* 1969, **134**:408-413.
29. Wikström MKF, L. SN-E: **Effect of Hydroxylamine on Respiration and Oxidative Phosphorylation.** *European J Biochem* 1969, **9**:160-166.
30. Hamill DR, Suprenant KA: **Characterization of the Sea Urchin Major Vault Protein: A Possible Role for Vault Ribonucleoprotein Particles in Nucleocytoplasmic Transport.** *Developmental Biology* 1997, **190**:117-128.
31. Scheffer GL, Wijngaard PLJ, Flens MJ, Izquierdo MA, Slovak ML, Pinedo HM, Meijer CJLM, Clevers HC, Scheper RJ: **The drug resistance-related protein LRP is the human major vault protein.** *Nature Medecine* 1995, **1**(6):578-582.
32. Mossink MH, van Zon A, Scheper RJ, Sonneveld P, Wiemer EA: **Vaults: a ribonucleoprotein particle involved in drug resistance?** *Oncogene* 2003, **22**:7458-7467.
33. Lüdeking A, Köhler A: **Identification of six mRNA sequences of genes to multixenobiotic resistance (MXR) and biotransformation in *Mytilus edulis*.** *Marine Ecology Progress Series* 2002, **238**:115-124.

34. Ravaux J, Gay L, Voss-Foucart MF, Gaill F: **Tube growth process in the deep-sea hydrothermal vent tubeworm *Riftia pachyptila* (Vestimentifera): synthesis and degradation of chitin.** *Cah Biol Mar* 1998, **39**(2):99-107.
35. Shillito B, Lubbering B, Lechaire JP, Childress JJ, Gaill F: **Chitin Localization in the Tube Secretion System of a Repressurized Deep-Sea Tube Worm.** *Biological Abstracts Vol 100, Iss 2, Ref 21637* 1995, **114**(1):67-75.
36. Mali B, Möhrlen F, Frohme M, Frank U: **A putative double role of a chitinase in a cnidarian: pattern formation and immunity.** *Developmental and Comparative Immunology* 2004, **28**:973-981.
37. Demuynck S, Bocquet-Muchembled B, Deloffre L, Grumiaux F, Leprêtre A: **Stimulation by cadmium of myohemerythrin-like cells in the gut of the annelid *Nereis diversicolor*.** *The Journal of Experimental Biology* 2004, **207**:1101-1111.
38. Vergote D, Sautière P-E, Vandenbulcke F, Vieau D, Mitta G, Macagno ER, Salzet M: **Up-regulation of Neurohemerythrin Expression in the Central Nervous System of the Medicinal Leech, *Hirudo medicinalis*, following Septic Injury.** *The Journal of Biological Chemistry* 2004, **279**(42):43828–43837.
39. Diatchenko L, Lau Y-FC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED *et al*: **Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries.** *Proc Natl Acad Sci USA* 1996, **93**:6025-6030.
40. Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL, Lukyanov SA, Sverdlov ED: **Equalizing cDNA Subtraction Based on Selective Suppression of Polymerase Chain Reaction: Cloning of Jurkat Cell Transcripts Induced by Phytohemagglutinin and Phorbol 12-Myristate 13-Acetate.** *Analytical Biochemistry* 1996, **240**:90-97.
41. Ewing B, Hillier L, Wendl M, Green P: **Base calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Research* 1998, **8**:175–185.
42. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi, B., Jennifer J *et al*: **TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST data sets.** *Bioinformatics* 2003, **19**:651-652.
43. Hanekamp K, Bohnbeck U, Beszteri B, Valentin K: **PhyloGena—a user-friendly system for automated phylogenetic annotation of unknown sequences.** *Bioinformatics* 2007, **23**(7):793-801.

44. Halanych KM, Lutz RA, Vrijenhoek RC: **Evolutionary origins and age of vestimentiferan tube-worms.** *Cahiers de Biologie Marine* 1998, **39**:355-358.
45. Livak KJ, Schmittgen TD: **Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method.** *Methods* 2001, **25**:402–408.

**CHAPITRE V**  
**Deuxième article**

## CHAPITRE V : Deuxième article

### Identification, expression et localisation d'un nouveau transcrit d'anhydrase carbonique chez *Riftia pachyptila*

#### 1. Résumé

##### 1.1 Contexte de l'étude

Une séquence partielle d'ADNc d'anhydrase carbonique (RpCAbr) a été mise en évidence chez *Riftia pachyptila* à partir de notre banque soustractive de tissu branchial (banque BR-BW), tandis qu'une autre séquence d'ADNc d'anhydrase carbonique (RpCAtr) déjà connue a été retrouvée dans notre banque soustractive de trophosome (banque TR-BW). Les deux fragments de séquence n'étant identiques qu'à seulement 66 % en acides aminés, nous nous sommes intéressés de plus près à ces deux isoformes. En effet, des études biochimiques ont révélé des activités importantes d'anhydrase carbonique au niveau de la branchie et du trophosome de *Riftia pachyptila*, mais avec des caractéristiques différentes. Nous avons utilisé les outils de PCR quantitative pour évaluer le taux d'expression de ces gènes dans chacun des tissus et vérifier leur expression différentielle suggérée par le résultat du séquençage des banques. La localisation des ARNm par FISH a également été réalisée sur des coupes histologiques des deux tissus. L'obtention de la séquence complète du nouveau transcrit par RACE-PCR a également permis d'effectuer de comparer nos séquences avec des séquences d'organismes vertébrés et « non vertébrés » disponibles dans les banques de données. Une reconstruction phylogénétique a été réalisée pour discuter de l'évolution de la famille multigénique des anhydrases carboniques.

##### 1.2 Principaux résultats

Nous avons clairement pu mettre en évidence une expression tissu-spécifique de chacune des séquences par nos études de PCR quantitative. Plus précisément, le gène RpCAbr est très fortement exprimé dans le tissu branchial par rapport à la paroi du corps et le trophosome, tandis que le transcrit RpCAtr est plus exprimé dans le trophosome que dans les



deux autres tissus. Néanmoins, ce dernier transcrit n'est qu'une dizaine de fois moins exprimé dans la branchie (résultats établis à partir de la comparaison des deux tissus au sein d'un même individu avec trois individus analysés). Ainsi, les résultats de FISH montrent que les deux types de transcrits sont présents dans le tissu branchial tandis qu'un seul est détecté dans le trophosome. Après alignement de nos deux séquences traduites avec les séquences protéiques de métazoaires, nous avons pu constater la conservation de la plupart des acides aminés impliqués dans le site catalytique des anhydrases carboniques. Dans notre reconstruction phylogénétique, les deux séquences RpCAbr et RpCAtr se regroupent ensemble et apparaissent proches des isoformes cytosoliques de vertébrés. Certains organismes « non vertébrés » ont à l'image des vertébrés plusieurs isoformes d'anhydrase carbonique apparaissant clairement polyphylétiques.

## **2. Article publié dans FEBS Journal**

**Sanchez S.**, Andersen A. C., Hourdez S. and Lallier F. H., 2007. Identification, sequencing, and localization of a new carbonic anhydrase transcript from the hydrothermal vent tubeworm *Riftia pachyptila*. *FEBS Journal*, **274**(20):5311-5324.

**Identification, sequencing, and localization of a new carbonic anhydrase transcript from the hydrothermal vent tubeworm *Riftia pachyptila***

Sophie Sanchez, Ann C. Andersen, Stéphane Hourdez and François H. Lallier\*

Equipe Ecophysiologie: Adaptation et Evolution Moléculaires,

UMR 7144 CNRS-UPMC

Station Biologique, Place Georges Teissier, BP 74, 29682 Roscoff Cedex, France

\* corresponding author

Tel +33 2 9829 2311

Fax +33 2 9829 2324

Mail: lallier@sb-roscoff.fr

Short title: Carbonic anhydrase transcripts in *Riftia*

Keywords: symbiosis; chemoautotrophy; Siboglinidae; messenger RNA; differential expression

Note: nucleotide sequence data are available in the GenBank database under the accession numbers EF490380 (RpCAbr) and EF490381 (RpCAbr2).

**ABSTRACT**

The vestimentiferan annelid *Riftia pachyptila* forms dense populations at hydrothermal vents along the East Pacific Rise at 2600 meter-depth. It harbors CO<sub>2</sub>-assimilating sulfide-oxidizing bacteria that provide all of its nutrition. In order to find specific host transcripts which could be important for the functioning of this symbiosis, we used a Subtractive Suppression Hybridization (SSH) approach to identify plume- or trophosome-specific proteins. This way we demonstrated the existence of carbonic anhydrase (CA) transcripts, a protein endowed with an essential role in generating the influx of CO<sub>2</sub> needed by the symbionts. One of the transcripts was previously known and sequenced. Our quantification analyses showed a higher expression of this transcript in the trophosome compared to the branchial plume or the body wall. A second transcript, with 69.7 % nucleotide identity with the previous one, was almost only expressed in the branchial plume. Fluorescent *In Situ* Hybridization confirmed the co-expression of the two transcripts in the branchial plume in contrast with the trophosome where only one transcript could be detected. An alignment of these translated CA cDNAs with vertebrate and non-vertebrate CA protein sequences revealed the conservation of most amino acids involved in the catalytic site. According to the phylogenetic analyses, the two *R. pachyptila* transcripts clustered together but not all non-vertebrate sequences grouped together. Complete sequencing of the new CA transcript revealed the existence of two slightly divergent isoforms probably coded by two different genes.

## INTRODUCTION

Vestimentiferan tubeworms (Polychaeta; Siboglinidae) often represent a major component of the endemic fauna at hydrothermal vents and cold seeps. These annelid worms are devoid of mouth, digestive tract, and anus [1], relying completely on their autotrophic sulfide-oxidizing symbionts to fulfill their metabolic needs [2]. These symbionts are located deep inside the body of the host, in a specialized organ called the trophosome. This location, remote from the environment that contains all the necessary nutrients for the bacteria, implies that the tubeworm host needs to transport oxygen, hydrogen sulfide and inorganic carbon compounds in large quantities for the bacteria to produce organic matter [3].

CO<sub>2</sub> is acquired from the environment by diffusion through the branchial plume [4, 5], the respiratory-exchange organ, where it is immediately converted into bicarbonate through high activities of carbonic anhydrase (CA) [6, 7]. Inorganic carbon accumulates up to very high concentrations in the body fluids (up to 30-60 mmol. l<sup>-1</sup>, [4, 5]). The pH values of these fluids remain stable and alkaline relative to the surrounding environment thus maintaining an inward CO<sub>2</sub> gradient [4, 6, 8]. Kochevar and Childress [7] also measured high CA activities in the trophosome. Indeed, once near the bacteriocytes (the cells housing the bacteria in the trophosome), a reconversion of bicarbonate into CO<sub>2</sub> is necessary because the bacterial symbionts only use molecular CO<sub>2</sub> [9] to enter the Calvin-Benson cycle or the reverse TCA cycle [10]. In this context, high activities of CA may represent an adaptation for providing the symbionts with a suitable chemical form of CO<sub>2</sub>.

CAs are zinc-containing enzymes catalyzing the reversible hydration of CO<sub>2</sub> to bicarbonate. Ubiquitous in a wide range of eukaryotic organisms, they are also widespread in the Archaea and Bacteria domains [11]. Among the broad range of physiological processes they participate to, CA can play a significant role in autotrophic organisms, serving as an inorganic carbon-concentrating component [12]. In symbiosis involving metazoa and autotrophic organisms, the host CA may help to provide a sufficient CO<sub>2</sub> flow to the symbionts, as shown for example in algal-cnidarian symbioses [13]. In the same way, measurements of CA activity in several chemosynthetic clam and vestimentiferan species indicate that CA facilitates inorganic carbon uptake, with high activities reported from clam gill, vestimentiferan plume and trophosome tissues [6, 7].

Biochemical studies on *Riftia pachyptila* [14, 15] revealed two main forms of cytosolic CA, with different kinetics and apparent molecular weight, one present in the branchial plume and the other in the trophosome. A complete cDNA was obtained by De Cian and

collaborators [15] from the trophosome tissue. Further functional and histological studies suggested the existence of several carbonic anhydrase isoforms in the trophosome tissue [16, 17], indicating the possible existence of various CA isoforms in groups other than vertebrates. Earlier studies addressed the central role of the branchial plume in oxygen, CO<sub>2</sub> and sulfide acquisition, as well as blood transport of these metabolites to the trophosome where symbionts are housed [3]. However, this review paper stressed out several points that remain to be elucidated regarding the different pathways involved in these transport processes.

In an attempt to identify yet unknown host proteins involved in branchial and trophosome functions associated with the symbiotic mode of life of *R. pachyptila*, we constructed subtractive tissue-specific cDNA libraries (Subtractive Suppression Hybridization, SSH). Among other cDNAs, we obtained a new CA transcript from the branchial tissue that is different from the one previously sequenced. Here, we show that the two CA sequences are differentially expressed in tissues of the worm. These sequences are also compared with other CA sequences from vertebrates and non-vertebrates.

## RESULTS

### *CA sequences from the SSH libraries*

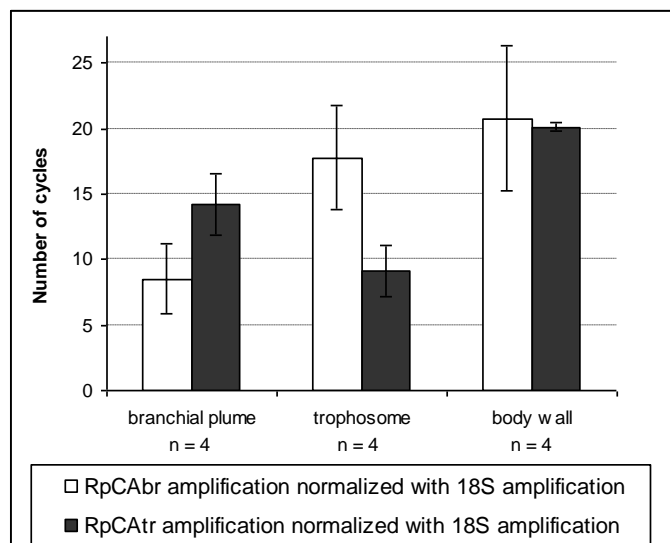
From the body wall-subtracted trophosome cDNA library, we recovered a 3' coding sequence fragment of 174 nucleotides and a partial 3' untranslated region (3'UTR) sequence of 234 nucleotides. These two fragments were strictly identical to the sequence already found by De Cian et al. ([14]; acc. # Q8MPH8), hereafter referred to as RpCA<sub>tr</sub>, standing for “*Riftia pachyptila* Carbonic Anhydrase trophosome”.

From the body wall-subtracted branchial plume cDNA library, we obtained a carbonic anhydrase transcript of 171 nucleotides, with only 66% nucleotide identity to the RpCA<sub>tr</sub> sequence, followed by a partial 3'UTR of 364 nucleotides radically different from RpCA<sub>tr</sub>. This new sequence is hereafter referred to as RpCA<sub>br</sub>, standing for “*Riftia pachyptila* Carbonic Anhydrase branchial plume”.

### *Tissue-specific expression*

The amount of each transcript that is amplified is quantitatively correlated to the fluorescence intensity emitted by the SYBRGreen fluorochrome when it was incorporated in double-stranded cDNA. The number of PCR-cycles required to amplify each CA transcript to

the same level of fluorescence, relative to the amplification of the reference transcript (18S rRNA transcript), is shown in Fig. 1.



**Fig. 1:** Normalized amplifications of RpCAbr and RpCAtr with 18S amplification. The number of cycles on the Y axis is the difference between the number of cycles required to amplify each transcript and the number of cycles required to amplify 18S. The number of tissue replicates (n) is indicated under each histogram.

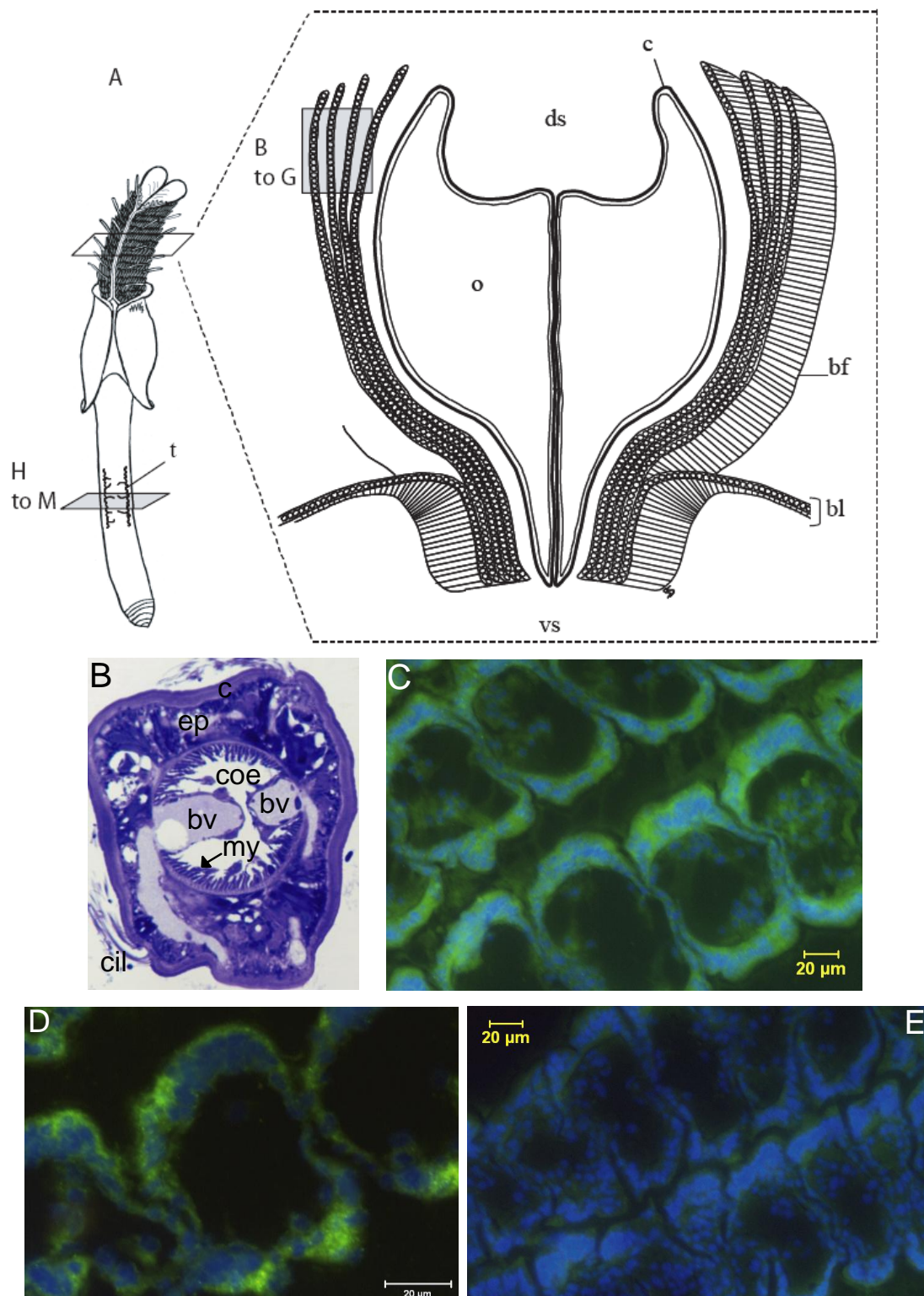
RpCAbr amplification reaches a fluorescence threshold after  $8.49 (\pm 2.68)$  cycles for branchial plume cDNA and after  $17.80 (\pm 4.02)$  cycles for trophosome cDNA (Fig. 1). Similarly, RpCAtr amplification reaches a fluorescence threshold after  $14.24 (\pm 2.33)$  cycles and  $9.11 (\pm 1.91)$  cycles for branchial plume and trophosome cDNA respectively. Nearly 10 less cycles are required to reach the threshold for the RpCAbr transcript in the branchial plume compared to the trophosome whereas about 5 less cycles are required to reach the threshold for RpCAtr in the trophosome compared to the branchial plume. Levels in the body wall are comparatively low ( $20.76 (\pm 5.55)$  cycles and  $20.14 (\pm 0.34)$  cycles are required to obtain the same quantities of RpCAbr and RpCAtr respectively).

Average values of relative expression levels resulted in 636-fold higher expression of RpCAbr in the branchial plume compared to the trophosome (tissue-pair comparisons within a single individual resulted in a 1,000-fold higher mean expression according to individuals for which we analysed the two tissues) and a 4,950-fold higher expression of RpCAbr in the branchial plume compared to the body wall (109-fold higher mean expression for paired tissues). The RpCAtr transcript showed a 184-fold higher expression in the trophosome compared to the branchial plume (12-fold higher mean expression for paired tissues) and a 2,098-fold higher expression in the trophosome compared to the body wall (2,500-fold higher

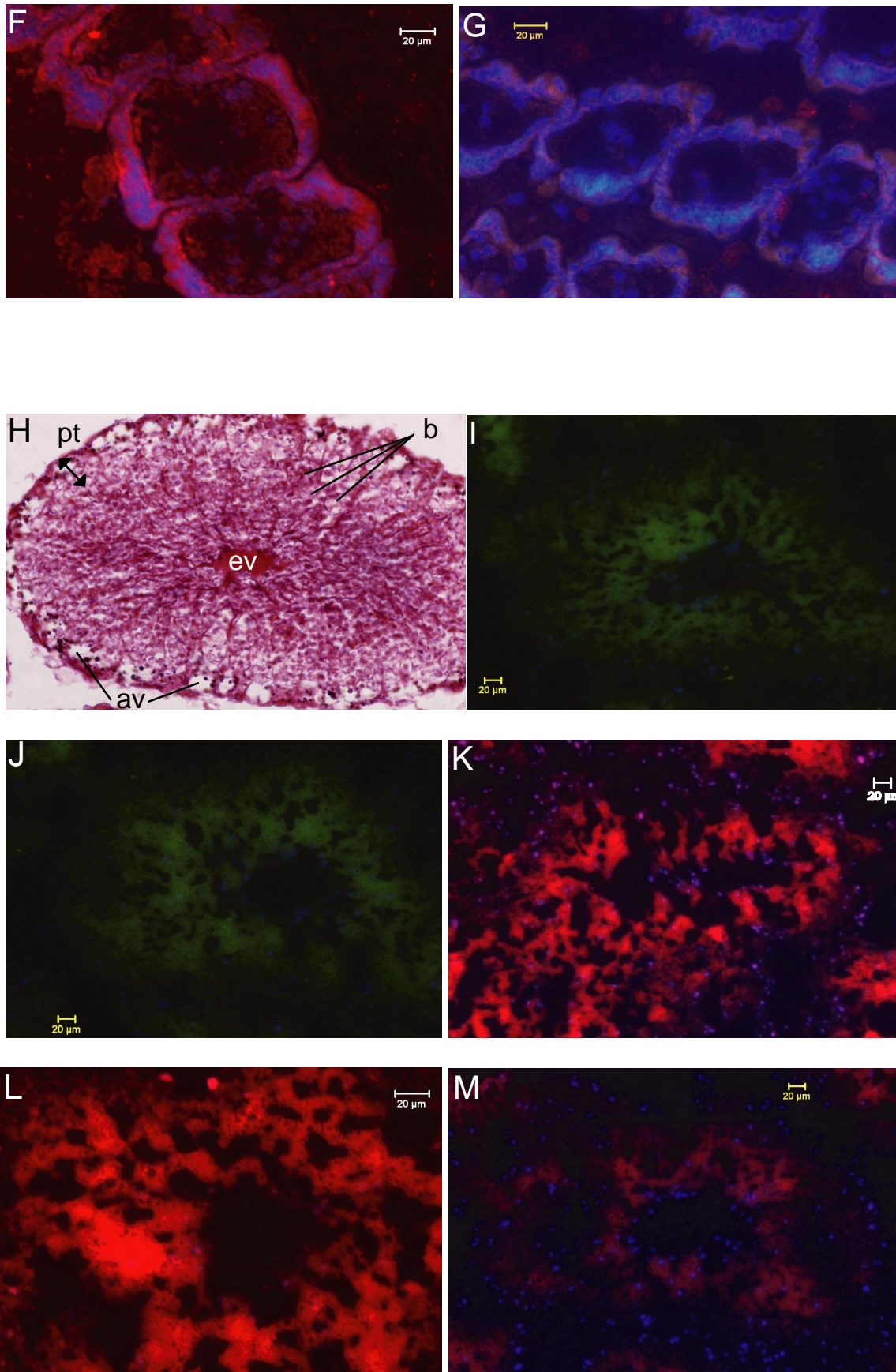
mean expression for paired tissues). Thus, the expression pattern of CAs appears tissue-specific.

### In situ hybridization

*In situ* hybridizations were performed on cross sections of the branchial plume and of the trophosome as shown in Fig. 2A.









**Fig. 2 (previous pages):** (A) Morphological representation of an adult *Riftia pachyptila* removed from its tube. Histological sections performed in this study are located at the levels indicated by shaded boxes on the drawings. t: trophosome; vs: ventral side; ds: dorsal side; o: obturaculum; c: cuticle; bf: branchial filament; bl: branchial lamellae. (B) Transverse section showing the morphological structure of a branchial filament with cuticle (c), tufts of cilia (cil), epithelial cells (ep), myoepithelium (my), blood vessels (bv) and coelome (coe). (C) to (G) FISH results on the branchial plume sections with RpCAbr probe (C to E, green FISH) and with RpCAtr probe (F and G, red FISH). Nuclei are stained in blue. (C) and (D) Positive staining with the anti-sense RpCAbr probe. (E) Negative control with the sense RpCAbr probe. (F) Positive staining with the anti-sense RpCAtr probe. (G) Negative control with the sense RpCAtr probe. (H) Transversal section of a trophosome lobule showing peritoneal cells (pt), bacteriocytes (b), afferent blood vessel (av) and efferent blood vessel (ev). (I) to (M) FISH results on the trophosome with the RpCAbr probe (I and J, green FISH) and with the RpCAtr probe (K to M, red FISH). Nuclei are stained in blue. (I) Positive staining with the anti-sense RpCAbr probe. (J) Negative control with the sense RpCAbr probe. (K) and (L) Positive staining with the anti-sense RpCAtr probe. (L) Higher magnification of the lobule showing the intensity of the labeling throughout the bacteriocytes. (M) Negative control with the sense RpCAtr probe.

The branchial plume is composed of a central obturaculum, mainly made of extracellular matrix, supporting at its periphery many branchial filaments. The branchial filaments are composed of a single layer of epidermal cells, on top of a myoepithelium that surrounds a central coelomic cavity and the two blood vessels it contains (Fig. 2B). The cytoplasm of the branchial epithelial cells is clearly stained with the RpCAbr cDNA probe (Fig. 2C). The staining is cytoplasmic as it generally corresponds to the rough reticulum area around the nucleus and is maximal in the cytoplasmic apex of the branchial epidermis. In contrast, the staining is very weak basally along the myoepithelium that lines the internal coelomic cavity. Although nuclei appear clustered on one side of each filament (Fig. 2D), a homogenous fluorescence was observed in the cytoplasm of the cells. The staining appears to be specific of the probe sequence as the staining is clear with the complementary sequence to RpCAbr, but not with the sense probe (negative control, Fig. 2E). The same hybridization procedure with the anti-sense RpCAtr cDNA probe on gill filaments sections resulted in similar staining and localization than the RpCAbr probe (Fig. 2F). The sense probe to the RpCAtr transcript did not give any signal above background level (Fig. 2G).

The trophosome tissue is composed of bacteriocytes grouped in lobules surrounding a central efferent vessel, and lined by peritoneal cells that are supplied with many small afferent blood capillaries (Fig. 2H). The bacteriocytes house the bacterial symbionts inside vacuoles of their cytoplasm. RpCAbr anti-sense probe did not stain the trophosome lobule more than its negative control (Fig. 2I, J). With the tissue specific RpCAtr, an intense staining is observed

in the cytoplasm of all the bacteriocytes (Fig. 2K, L) compared to its negative control (Fig. 2M).

### ***Full-length sequencing***

The complete RpCAbr sequence (acc. # EF490380) was obtained from the branchial plume cDNA with an open reading frame of 726 nucleotides and 5' and 3' UTR sequences of 171 and 442 nucleotides, respectively. Positions of the primers on the complete cDNA are given in Table 1. A poly-A tail signal (AAUAAA) occurred 405 nucleotides downstream from the in-frame stop codon and 19 nucleotides upstream from the poly-A tail. Search of motifs with the PROSITE server (ScanProsite: [18]) showed the presence of an  $\alpha$ -CA signature from amino acids 96 to 112: S-E-[HN]-x-[LIVM]-x(4)-[FYH]-x(2)-E-[LIVMGA]-H-[LIVMFA](2). The new RpCAbr sequence is 69.7% identical in nucleotides (and 66.8% in amino acids) to the previously known RpCAtr sequence (acc. # Q8MPH8). The best results of Blastx on NCBI server are shown in Table S1. In addition to RpCAtr, five out of fifteen most closely related protein sequences that matched with our sequence belonged to non-vertebrates (Table S1). The blast analysis shows that RpCAbr appears close both to CAI and CAII *Mus musculus* isoforms sequences.

### ***Alignment***

Full-length RpCAbr and RpCAtr were aligned with other metazoan sequences (Fig. 3). A noteworthy difference between RpCAbr and RpCAtr is the deletion of one amino acid (proline) in the RpCAbr sequence at position 85, while a majority of the aligned sequences exhibit a proline. The three histidine residues (named H94, H96 and H119 in reference to positions 94, 96 and 119 in CAII from *Homo sapiens*) which are directly involved in binding the zinc cofactor, are conserved in the two *R. pachyptila* sequences (positions labeled “Z” in Fig. 3). These residues are hydrogen bond donors to Q92 (position 129, shared by all organisms of Fig. 3 with the exception of *Riftia* and *Caenorhabditis* sequences where it is replaced by a serine residue), N244 (position 297, conserved) and E117 (position 156, conserved) respectively. Other amino acids involved in the hydrogen bond network surrounding the active site are also conserved (positions labeled “\*” in Fig. 3) with few exceptions. For example, at position 98, the two *Riftia* sequences exhibit a hydrophobic amino acid (leucine) instead of the histidine that is shared by almost all other sequences. The same amino acid replacement occurs in the two isoforms CAa and CAb of *Drosophila melanogaster*.

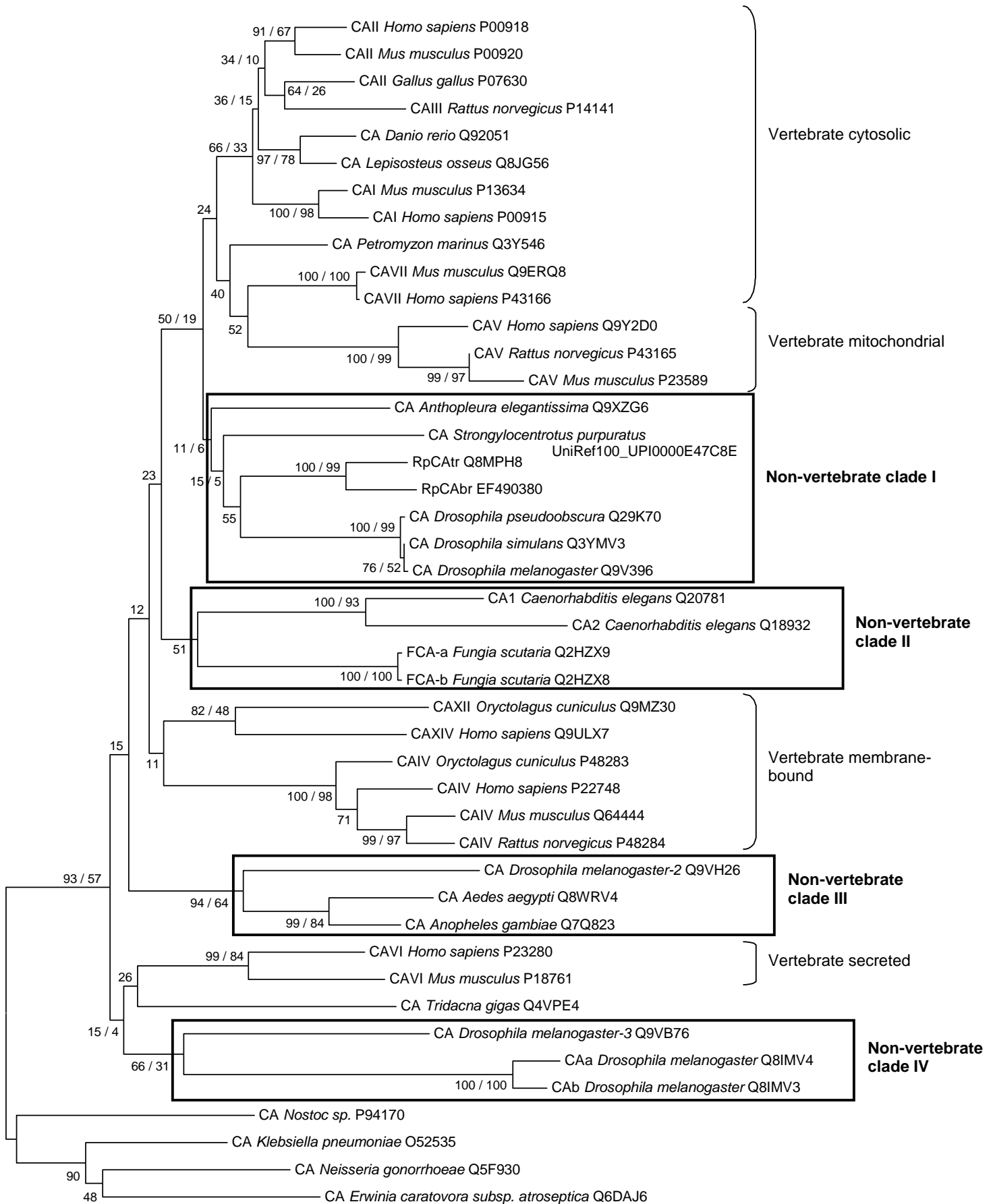


**Fig. 3 (previous page):** Alignment of complete RpCAbr and RpCAtr amino acids sequences with some representative metazoan CA protein sequences. Identical and similar amino acids shared by at least 50% of the isoforms are shown in black and grey, respectively. Histidine residues involved in zinc binding in the catalytic site are indicated by a “Z”; important amino acids involved in the hydrogen bond network are indicated by a “\*”; framed amino acids are commented in the section “Results” and positions indicated above the frame refer to the reference positions in CA II *Homo sapiens* sequence. The last few amino acids of the alignment have been omitted.

### *Phylogenetic analyses*

Neighbor-Joining (NJ) and Maximum Parsimony (MP) trees produced similar topologies. Only the NJ tree is presented in Fig. 4 but bootstrap values (BP) for both NJ and MP analyses are shown near the recurrent nodes found in both distance and parsimony methods. Given the high number of taxa used in these reconstructions, BP values are generally low, and lower in MP tree than in the NJ one.

Non-vertebrate CA sequences are clearly polyphyletic. Some non-vertebrate CA sequences form a single clade (clade I in Fig. 4) comprising cnidarian, protostomian and deuterostomian sequences. Although supported by very low bootstrap values ( $BP_{NJ}=15$  and  $BP_{MP}=5$ ), this clade is found in both NJ and MP analyses. In this clade, RpCAbr is most closely related to the previously sequenced RpCAtr ( $BP_{NJ}=100$  and  $BP_{MP}=99$ ). *Fungia scutaria* (FCA-a and FCA-b) and *Caenorhabditis elegans* (CA1 and CA2) sequences fall outside of clade I and are more closely related to each other ( $BP_{NJ}=51$ ) (clade II in Fig. 4). Although not supported by high bootstrap values, we believe the isolation of clade I from the rest of non-vertebrate sequences is well supported as the group consisting of clade I, vertebrate cytosolic, and vertebrate mitochondrial sequences is found in both NJ and MP analyses ( $BP_{NJ}=50$  and  $BP_{MP}=19$ ). We can notice that *Drosophila* spp. sequences form three distinct groups: the first one (CA *D. melanogaster* + CA *D. pseudoobscura* + CA *D. simulans*) belongs to clade I; the second one (CA *D. melanogaster*-2) belongs to clade III and the third one (CAa *D. melanogaster* + CAb *D. melanogaster* + CA *D. melanogaster*-3) forms clade IV. This latter clade is most closely related to the non-vertebrate clam *Tridacna gigas* and the CAVI vertebrate sequences in both NJ and MP analyses but with very low support ( $BP_{NJ}=15$  and  $BP_{MP}=4$ ).



**Fig. 4 (previous page):** Neighbor-joining (NJ) tree obtained after a multiple alignment of 40 complete metazoan CA amino acids sequences. Four bacterial  $\alpha$ -CA sequences from *Nostoc* sp., *Klebsiella pneumoniae*, *Erwinia carotovora* subsp. *atroseptica* and *Neisseria gonorrhoeae* are used as outgroups. Some nodes were also recovered from Maximum Parsimony (MP) analysis. Numbers are bootstrap values (BP) calculated from 1000 replicates from NJ (BP<sub>NJ</sub>) and MP (BP<sub>MP</sub>) analyses and are represented as (BP<sub>NJ</sub> / BP<sub>MP</sub>). Nodes with only one number (BP<sub>NJ</sub>) are only found from NJ analysis.

### ***RpCAbr isoforms***

In addition to RpCAbr, amplification with RpCAbrR3 primer (Table 1) gave another partial cDNA with an open reading frame of 483 nucleotides and a 175 nucleotide-long 5'UTR. RpCAbr and this other transcript's (RpCAbr2, acc. # EF490381) partial coding regions are very similar to each other and exhibited only 3 non-synonymous substitutions (99.38 % nucleotides identity and 98.14 % amino acids identity). However, the two transcripts strongly differ in their 5'UTR sequence from nucleotides 18 to 140, although a fragment of thirty-five nucleotides is very well conserved at the end of both 5'UTR sequences. This latter fragment may have important properties since investigations on 5'UTR regions by the search engine UTRscan [19] revealed the presence of an internal ribosome entry site (IRES) for both 5'UTR of RpCAbr (nucleotides 83 to 171) and RpCAbr2 (nucleotides 82 to 175) transcripts.

A phylogenetic analysis with this partial sequence (data not shown) revealed that, as expected, RpCAbr and RpCAbr2 grouped together and were a sister-group of RpCAtr. Other analyses (data not shown) showed that the adult *Fungia scutaria* CA sequence (only partial and therefore not used in our phylogenetic construction) was most closely related to CA *Anthopleura elegantissima*.

## **DISCUSSION**

### ***Differential expression***

We demonstrated that the RpCAbr gene is highly, and preferentially, expressed in the branchial plume tissue whereas the RpCAtr gene is preferentially expressed in the trophosome but significantly expressed in the branchial plume tissue as well. Fluorescent *in situ* hybridization on histological sections corroborated these findings with the detection of RpCAtr mRNA in both the epidermal cytoplasm of the branchial filaments and in the cytoplasm of the trophosomal bacteriocytes. We could only detect RpCAbr mRNA in the

epidermal cytoplasm of the branchial filaments (we could not detect this transcript in the trophosome probably because of high signal background noise).

This is the first report of tissue-specific expression of cytosolic CAs in a non-vertebrate species. Such a protein is essential for the symbiotic association of the worms with their bacteria. Studies on *Anthopleura elegantissima*, a cnidarian with symbiotic dinoflagellate, already showed that CA expression is enhanced in the presence of symbionts [20]. We could not reproduce such an approach on *Riftia* since the aposymbiotic stage is limited to the larval phase of its life cycle [21]. Thus it is first difficult to obtain these stages in the hydrothermal vent environment, and second the aposymbiotic-specific expression condition could be masked by the developmental condition.

### ***Comparison with western-blot and CA activities studies***

Previous studies by western-blot and SDS-PAGE on cytosolic fractions [14, 15] concluded that there were two CA proteins: one of 27 kDa in the branchial plume, and another of 28 kDa in the trophosome. From the differential expression results we obtained, RpCAbr could correspond to the 27 kDa protein and RpCAtr to the 28 kDa one. However, from our translated sequences, we calculated the total molecular weight of each translated transcripts and found 26 973 Da for RpCAbr and 27 084 Da for RpCAtr. The difference of almost 1 kDa obtained for the trophosome CA protein (observed on gel) could be attributed to a differential migration behavior of the protein in the SDS-page gel or to post-translational modifications such as phosphorylations. For instance, 3 glycosylation, 3 phosphorylation and 6 myristyl sites were found in the translated RpCAbr transcripts using Motif Scan ([22], MyHits Swiss Institute of Bioinformatics). In the RpCAtr protein sequence, 10 more phosphorylation sites were found (1 glycosylation, 13 phosphorylation and 3 myristyl sites).

Different CA activities were previously measured in *Riftia pachyptila* [6, 14, 15]. In these studies, high affinities and activities of CA had been found in the plume and in the trophosome. CA from the branchial plume tissue had an affinity of 13.9 mmol.L<sup>-1</sup> and an activity of 253.7 μmol CO<sub>2</sub>. min<sup>-1</sup>. g<sup>-1</sup> wet wt. CA from the trophosome tissue had an affinity of 7.2 mmol.L<sup>-1</sup> and an activity of 109.4 μmol CO<sub>2</sub>. min<sup>-1</sup>. g<sup>-1</sup> wet wt. Given our results of differential expression, RpCAbr and RpCAtr could be the transcripts coding for the two different CAs identified by Kochevar and coworkers [14] based on a biochemical study. However, in the protein extracts analyzed by these authors [14] in the branchial plume, only one CA form had been identified. Therefore, we may wonder whether Kochevar and coworkers did not detect the second CA form (corresponding to RpCAtr transcript) because

its protein concentration was below the detection threshold. However, we do not know exactly in what proportions the two different CA proteins are present since we only have indications about the expression level of their genes which may not reflect protein levels.

### ***Branchial plume CA isoforms***

From the full-length sequences, it appears that two isoforms (RpCAbr and partial RpCAbr2) could correspond to two different genes expressed in the branchial plume. It is unlikely that the two sequences correspond to different alleles of the same gene as the divergence of the 5'UTRs is high. No eukaryotic specific splicing consensus sequences could be found in either RpCAbr or RpCAbr2 5'UTR sequences. These two transcripts have different 3'UTRs (data not shown), a strong support toward the existence of two distinct genes. These transcripts are then likely the result of the transcription of two different genes that evolved independently after a duplication event. This possible duplication event could illustrate a strategy to increase the number of transcripts instead of having a strong transcription promoter. The fact that several genes can be the source of several isoforms in the branchial plume could increase global carbonic anhydrase activity.

The two isoforms possess a relatively well conserved region in their 5'UTRs. This conserved region contains IRES motifs. This IRES sequence is an alternative mode of 40S recruitment to the mRNA instead of 5'capping recruitment (for review, see [23]). The occurrence of such a mechanism could enhance the regulation capacity for CA translation and may be correlated to an inhibition of cap-dependant translation in the branchial plume tissue. Indeed, some IRES are only active in specific tissues [24]. But we cannot conclude to any IRES activity in our case, given that only one IRES prediction from just looking at the 5'UTR should be checked by further studies of the structural elements (as enzymes and translation factors) that drive this mechanism. Interestingly however, RpCAtr did not exhibit any IRES in its 5'UTR.

### ***A membrane-bound CA in *Riftia pachyptila*?***

Two models exist for CO<sub>2</sub>-concentrating mechanisms in autotrophic organisms [12]. Bicarbonate ions may enter the cells through specific anionic exchangers and then be converted to CO<sub>2</sub> intracellularly with the help of cytosolic CA; alternatively, membrane-bound CA can catalyze bicarbonate conversion to CO<sub>2</sub> extracellularly in the boundary layer and thereby locally increases CO<sub>2</sub> gas diffusion into the cells. The existence of a membrane-bound CA has been postulated in *Riftia* bacteriocytes on the basis of inhibitor experiments



performed on isolated cells [17]. The two *Riftia* sequences presented in this study (RpCAbr and RpCAtr) do not seem to be membrane-bound isoforms. The RpCAbr and the RpCAtr transcripts are phylogenetically related and both distant from the vertebrate membrane-bound (CAIV) isoforms, and from the larval *Fungia scutaria* sequences (FCA-a and FCA-b) which may be membrane-bound isoforms [25]. Moreover, as shown in the alignment, *R. pachyptila* CAs do not share any specific feature with CAIV isoforms when FCA-a and FCA-b do [25]. The *Riftia* sequences are also phylogenetically distant from the *Aedes aegypti* and *Anopheles gambiae* CA sequences, and do not contain any GPI-anchored site (tested with the PSORT II server, <http://psort.hgc.jp/>) binding the protein to the membrane, whereas the mosquitoes sequences do [26]. In addition, no evidence of signal peptide in 5' coding regions of *Riftia pachyptila* CA sequences could be found.

### ***Catalytic mechanism***

The zinc catalytic active site works in two main steps. During the first step, the zinc-bound hydroxide reacts with CO<sub>2</sub> forming a zinc-bound bicarbonate, which is then replaced by water. During the second step of catalytic activity, a proton is transferred from the zinc-bound water to the external buffer via a shuttle group, H64 (using amino acids positions in CAII *Homo sapiens* sequence as a reference from here on, see Fig. 3). This proton transfer is necessary to regenerate the zinc-bound hydroxide which is the catalytically active species [27, 28]. This H64 (position 98 in the alignment, Fig. 3) combined to a histidine cluster consisting of residues H3, H4, H10, H15 and H17 explains the general high efficiency of CAII isoforms as a catalyst [27, 29] because it could constitute a very appropriate channel to efficiently transfer protons from the active site to the reaction medium [30]. H64 can be replaced by less efficient proton shuttle groups such as K64 (in CAIII *Rattus norvegicus* for example) or Y64 (in CAV *Mus musculus*, CA *Anthopleura elegantissima*, CA *Drosophila melanogaster* and CA *D. pseudoobscura*).

Among non-vertebrates sequences, *Strongylocentrotus purpureus* and *Fungia scutaria* larvae sequences have a H64 also shared by *Anopheles gambiae*, *Aedes aegypti*, *Tridacna gigas*, *Drosophila melanogaster*-2 and *D. melanogaster*-3 sequences (data not shown). In contrast, *Riftia pachyptila* amino acid sequences do not have any of these CA II features. Indeed, they have neither H64 nor any specific histidine cluster. Besides, the two *Riftia pachyptila* sequences exhibit a hydrophobic amino acid (leucine) instead of H64. That point is problematic since this amino acid cannot receive any proton. *D. melanogaster* CAa and CAb sequences also share this peculiar trait. To our knowledge, there has been no study on specific

CA activity in this latter species. CA activity is however present in *Riftia pachyptila*, and if these transcripts encode for functional proteins, a possibility of replacement of H64 could be the involvement of another group, E106, which, although a less likely candidate, has been suggested to be able to transfer protons [31]. However, without an overexpression approach of RpCAbr and RpCAtr, we cannot know the functional effect of changes of some key amino-acids.

### ***Origin and number of non-vertebrate CAs***

Although the bootstrap values of the deep branches are low, we can draw some tentative conclusions from the phylogeny. Our study cannot exclude that the clades I and II could have a common origin with cytosolic CAI, CAII, CAIII, CAVII and mitochondrial CAV vertebrate isoforms as was previously suggested by [15]. The two clades could have a common ancestor being either a CAII-like [32] or a CAVII-like [33] protein. But, in contrast to De Cian and coworkers [15] phylogenetic analysis where only three non-vertebrate sequences were included, the extended set of invertebrate sequences now available for our study did not strictly group together. Our phylogenetic reconstruction shows on one hand, a close relationship of *Riftia pachyptila* CA sequences with one of the CA *Drosophila melanogaster* sequences, and on the other hand, the other *D. melanogaster* sequences more closely related to the CAVI vertebrate isoforms (CAa, CAb and *D. melanogaster-3*) or to the mosquitoes sequences (*D. melanogaster-2*). Del Pilar Corena and coworkers [34] suggested that several CA isoforms also exist in *Aedes aegypti*. The cnidarian *Fungia scutaria* also possesses multiple CA transcripts [25]. The adult *Fungia scutaria* sequence is more closely related to *Riftia pachyptila* and *Anthopleura elegantissima* CA transcripts (data not shown). On the contrary, the two larval *Fungia* CA transcripts included in our phylogenetic reconstruction seem evolutionarily distant from clade I as already reported [25]. Vertebrate cytoplasmic CAs could have evolved through duplication events over the course of 600 million years [33]. In De Cian and coworkers study [15], the three non vertebrate sequences analyzed (RpCAtr, CA *A. elegantissima* and CA *D. melanogaster*) formed a distinct cluster apart from the secreted (CAVI) and membrane-bound (CAIV) isoforms. The present study could support the existence of a more ancient  $\alpha$ -CA-like ancestor for both vertebrate and non-vertebrate CAs.

## EXPERIMENTAL PROCEDURES

### *Animals and sampling*

Specimens of *Riftia pachyptila* were collected at the Rehu Marka (17°25'S – 113°12'W), Susie and Miss WormWood (17°35'S - 113°14'W) sites at 2600 meters-depth along the South East Pacific Rise during the BIOSPEEDO 2004 cruise. For each individual, parts of the branchial plume, trophosome and body wall tissues were isolated on ice, placed in RNAlater (Ambion) for 24-h at 4°C and frozen in liquid nitrogen.

### *RNA extraction*

Plume, trophosome and body wall tissue samples were pulverized individually in liquid nitrogen in RNase-free conditions. For each tissue, total RNA was extracted using the RNAlater solution (Eurobio) following the manufacturer's instructions. Then, both for libraries constructions and complete sequencing, messenger poly-A RNAs were purified using the oligo-dT resin column of the mRNA Purification Kit (Amersham).

### *Construction of subtractive tissue-specific cDNA libraries*

Libraries were constructed from tissues belonging to one individual therefore representing one organism transcriptome. A total of 4 libraries were obtained: branchial plume vs body wall subtracted library (and its reciprocal) and trophosome vs body wall subtracted library (and its reciprocal).

For all tissue pairs, cDNA synthesis as well as subtractive suppressive hybridization (SSH, [35, 36]) comprising steps of adaptor ligation, subtractive hybridization and selective amplification were performed following the protocol of the Clontech PCR-Select™ cDNA Subtraction Kit (BD Biosciences). For each SSH procedure, the whole amplification product was cloned in TOPO®-TA cloning vector (Invitrogen), giving a range of cDNA fragment sizes. Nearly 200 cDNA fragments were sequenced for each library.

### *SYBR Green quantitative PCR*

#### *Reverse transcription (RT)*

From each total RNA sample (branchial plume, trophosome and body wall) fresh RT was conducted with a random primer. Each reaction mixture was composed of: 2 µl of M-MLV RT buffer; 0.5 µl of BSA (10 mg/ml), 1 µl of total RNA (1.24 µg/µl), 2.5 µl of dNTP

(4mM total), 1.5 µl of Random Primer 9 (Ozyme) (100 ng/µl), 3 µl DEPC water. Then, reaction mixtures were incubated at 80°C for 5 minutes and placed on ice. M-MLV RT was added (1 µl) to each reaction mixture and all reactions were incubated at 42°C for 1 hour and finally placed on ice.

### Amplification

Specific pairs of CA primers (Table 1) located in the 3'untranslated region of each transcript were designed using the software Primer Express.

**Table 1: Primers sequences for *Riftia pachyptila* carbonic anhydrase transcripts: RpCAbr and RpCAtr. Positions are given using the initiation codon as a reference.**

Amplification of RpCAbr and RpCAtr by quantitative PCR		
Primers	Sequence	Position
RpCAbrFq*	5' TGG TTT CAC CCC GTC GAA 3'	932 - 949
RpCAbrRq*	5' GGT CTG GTC TTT TCT CGC CAT A 3'	966 - 987
RpCAtrFq*	5' GCC AGG TGT CGT CCT CGT T 3'	710 - 728
RpCAtrRq*	5' TCA CAA ATG TCC AGT GCC AGT T 3'	757 - 778
Full length sequencing of RpCAbr		
Primers	Sequence	Position
RpCAbrF	5' TAC AAG GAT GCC ATT AGC 3'	613 - 630
RpCAbrR1	5' CGT AGC AGT ATC AGC AGT 3'	822 - 839
RpCAbrR2	5' AGA GCA GCA GAC CTT ACG 3'	706 - 723
RpCAbrR3	5' GTT ACT TCC GCA GCT AGG 3'	466 - 483
Probes amplification for FISH		
Primers	Sequence	Position
RpCAbrF	5' TAC AAG GAT GCC ATT AGC 3'	613 - 630
RpCAbrR1	5' CGT AGC AGT ATC AGC AGT 3'	822 - 839
RpCAtrFprobe	5' TAC AAA GAT CCA ATC CAG C 3'	616 - 634
RpCAtrRprobe	5' TAA GAT TAC CAG AAT TGC 3'	844 - 861

\*Primers designed by Primer Express software (ABI PRISM™)

18S rRNA transcript was chosen as a reference gene for the normalization of expression data and was amplified with the 18h and 18L primers [37]. For amplifications, the Power SYBR Green PCR master mix (PE Applied Biosystems) was used with 23 µl reaction mixtures in a Chromo4™ System CFB-3240 (BIORAD). PCR reactions were performed in triplicates. Amplification conditions were 40 cycles with the following profile: 95°C for 30 s, 60°C for 30 s, and 72°C for 1 min. For each kind of tissue, standard curves were generated for

18S and the CA transcripts over a large range of template cDNA quantity in order to calculate PCR efficiencies which are critical for correct quantification.

### *Data analysis*

For each transcript, the efficiency (E) was calculated from the slope (S) of the standard curve using the formula:

$$E = 10^{-1/S} - 1$$

Once differences between efficiencies of reference gene and target gene amplifications were approximately equal (i.e. did not exceed 5% difference in each tissue), we first looked at the normalizations of the CA transcripts amplifications compared to the endogenous control amplification for each tissue in order to get the normalized number of cycles (NNC):

$NNC = Ct_{\text{target}} - Ct_{18S}$ , where Ct is the Threshold Cycle i. e. the number of cycles required to reach a same quantity of amplified cDNA during the exponential phase.

Then, for relative quantification measurement, we used the  $2^{-\Delta\Delta Ct}$  method [38]. For each transcript, relative quantification result was obtained by comparing the level of expression in each tissue with the level of expression in the calibrator tissue, the latter being chosen as the tissue for which the better expression was observed, with the following calculation:

$$\text{Relative expression level} = 2^{-(NNC_{\text{sample}} - NNC_{\text{calibrator}})}$$

We also calculated the relative expression level of the transcripts in the tissues of a whole individual and performed the calculation over several individuals (this could only be done for individuals for which we had at least two tissues to be compared to each other).

### ***Rapid Amplification of cDNA Ends (RACE)***

Full-length cDNA was obtained by RACE-PCR from a branchial plume poly-A RNA sample. 3' amplification was conducted according to the manufacturer's instructions (Roche Diagnostics). For the 5' amplification, the protocol was modified as follows: poly(A) tailing of first-strand cDNA was replaced by poly(C) tailing. As a consequence, for the next PCR amplification, the oligo-dT anchor primer was replaced by an oligo-dG primer. Specific internal primers used for the 5' and 3' amplifications and their positions are shown in Table 1.

### ***Sequencing***

Plasmid DNA from individual colonies were purified with a FlexiPrep kit (Amersham) and used in a dye-primer cycle sequencing reaction with universal primer T3 or T7 and the

Big Dye® Terminator V3.1 Cycle Sequencing kit (Applied Biosystems). Reactions were then run on a 16-capillary 3130 Applied Biosystems sequencer.

### ***Preparation of histological sections***

Pieces of branchial plume- and trophosome-tissues were cryo-embedded in Tissue-teck (Sakura Finetek) and conserved in liquid nitrogen. 5-7  $\mu\text{m}$  sections were made at  $-25^{\circ}\text{C}$  on a Leica CM3050S cryo-microtome. The sections were collected on glass slides coated with 2% Biobond (Electron Microscopy Sciences EMS, Euromedex). They were deproteinized for 10 min with  $0.2 \text{ mol l}^{-1}$  HCl at room temperature, treated for 10 min in 2x SSC at  $70^{\circ}\text{C}$ , and rinsed twice in PBS ( $0.1\text{M}$  pH 7.5  $1100 \text{ mOsm}$ ) for 5 min. The slides were then incubated with  $50 \mu\text{g ml}^{-1}$  Proteinase K (Sigma) for 15 min at  $37^{\circ}\text{C}$  in a moist chamber and rinsed twice in cold PBS for 5 min.

### ***Production of the probes for in situ hybridization***

Specific oligonucleotides primers (Table 1) were designed on the basis of the alignment of the two sequences RpCAbr and RpCAtr. These primers were chosen to amplify a region overlapping the coding and 3'UTR regions for best specificity (hybridization specificity was checked by dot-blot, results not shown).

Once fragments of RpCAbr and RpCAtr transcripts were amplified from branchial plume and trophosome cDNAs respectively, migration of PCR products on agarose gel allowed us to check the size of the amplicons. Then these PCR products were purified with the High Pure PCR Purification Kit (Roche Diagnostics) following the manufacturer's instructions.

Different fluorochrome labeling of the probes was chosen to detect RpCAbr mRNAs and RpCAtr mRNAs by green and red fluorescence respectively. The anti-sense RpCAbr probe was synthesized by incorporation of DIG-conjugated UTP from the purified PCR product by a linear amplification with reverse primer (RpCAbrR1) with the PCR DIG Probe Synthesis Kit (Roche Applied Science). The sense probe (negative control) was produced with the RpCAbrF primer by the same procedure. The anti-sense RpCAtr probe was synthesized in two steps. First, linear amplification of the purified PCR product was performed with the reverse primer RpCAtrR probe only in order to enrich the PCR product for anti-sense RpCAtr fragments. Then, addition of biotin-16-ddUTP to the 3'OH ends of 100 pmol of this cDNA amplification was performed with the Terminal Transferase Recombinant

(Roche Diagnostics). The sense probe (negative control) was produced with the RpCAtrF probe following the same procedure.

### ***In Situ Hybridization***

Sections were prehybridized at 44°C for half an hour in hybridization buffer (HB: 0.9M NaCl, 20mM Tris-HCL pH 7.5, 0.01% SDS, 10% dextran sulfate, 2% Blocking Reagent (BR, Roche Diagnostics), 40% deionized formamide) in a moist chamber. Then, the probe (15 ng.µl<sup>-1</sup> in pre-warmed HB) was added to each slide and the hybridization was conducted for 20 hours at 44°C. After three stringent washes with HHB buffer (20 mM Tris-HCl pH 7.5, 28mM NaCl, 0.01% SDS, 5mM EDTA) at 46°C (20 min), 52°C (20 min) and 46°C (20 min), hybridizations were blocked with 100 µl of TNB Buffer (0.1M Tris-HCl pH 7.5, 0.15M NaCl, 0.5% BR) in a moist chamber for 30 min at room temperature.

### ***Fluorescence detection***

Tyramide Amplification Signal (TSA) system (Perkin Elmer) with the use of Horse-Radish Peroxidase (HRP) gave the best signal/background ratio results. To detect the DIG-labeled RpCAbr probe, anti-DIG-POD (Roche Diagnostics) was added at a 1:100 concentration in TNB Buffer. To detect the biotin-labeled RpCAtr probe, streptavidin-HRP was added at a 1:100 concentration in TNB Buffer. Slides were incubated for 2 hours in a moist chamber at room temperature. Then, they were washed twice for 10 min each in fresh TNT Buffer (0.1M Tris-HCl pH 7.5, 0.15M NaCl, Tween 20) at room temperature and for 15 min in TNT buffer at room temperature in the dark. RpCAbr was detected by a TSA reaction that took place for 30 min at room temperature in a 1:50 dilution of Fluorescein Tyramide (Green FISH) in equal volumes of 1X Amplification buffer (Perkin Elmer) and 40% dextran sulfate. RpCAtr was similarly detected in a 1:50 dilution of Tetra Methyl Rhodamine (Red FISH) in equal volumes of 1X Amplification buffer (Perkin Elmer) and 40% dextran sulfate for 30 min at room temperature. Slides were then washed twice for 20 min each at 55°C in new pre-warmed TNT Buffer to stop the enzymatic reaction and to remove dextran sulfate [39]. The nuclei were counter-stained with a 2 µg.ml<sup>-1</sup> DAPI solution for 10 min. Sections were mounted in Citifluor antifading reagent (Electron Microscopy Sciences), covered with coverslips and sealed with nail varnish.

### ***Homologies search, alignment, and phylogenetic analyses***

BLAST analyses (blastx and blastn) of the cDNA libraries sequences were conducted on the NCBI server (<http://www.ncbi.nlm.nih.gov/BLAST/>). Accession numbers (NCBI Entrez Proteins) of the sequences used in the phylogenetic reconstruction are given on the tree presented in Fig. 4. All metazoan protein sequences used in [15] for phylogenetic reconstruction were also used in our study. In order to test the hypothesis of the “non-vertebrate” clade previously observed by De Cian and coworkers [15], we added CA protein sequences from our newly identified RpCAbr translated sequence as well as sequences from the sea urchin *Strongylocentrotus purpuratus*, the nematode *Caenorhabditis elegans*, the fruitflies *Drosophila melanogaster* and *D. pseudoobscura*, the mosquitoes *Aedes aegypti* and *Anopheles gambiae*, the clam *Tridacna gigas* and larval sequences from the cnidarian *Fungia scutaria*. Finally, we chose an outgroup comprising  $\alpha$ -CAs from a cyanobacteria (*Nostoc* sp.) and three proteobacteria (*Klebsiella pneumoniae*, *Erwinia carotovora* subsp. *atroseptica* and *Neisseria gonorrhoeae*).

All the 44 complete sequences were first automatically aligned with ClustalW [40] in MEGA 3.1 [41] and the alignment was then adjusted visually. The Neighbor-Joining tree was constructed under the Dayhoff matrix model (PAM matrix) and the Maximum Parsimony tree was constructed with the close-neighbor-interchange search option. For each method, bootstrap tests were conducted over 1000 replicates.

## ACKNOWLEDGEMENTS

We wish to thank Dr. D. Vaultot’s research group for the use of their microscope and their useful advice on FISH-TSA experiments (UMR 7144 CNRS-UPMC Roscoff, France). We also are grateful to Dr. Didier Jollivet, chief scientist of the BIOSPEEDO cruise (2004) and to the crews of the N. O. L’Atalante and the submersible Nautilie for providing the samples used in this study. We also thank the three referees whose remarks have considerably improved this paper. Funding for this project was provided by the Région Bretagne.



## REFERENCES

1. Jones ML (1985) On the Vestimentifera, new phylum: Six new species, and other taxa, from hydrothermal vents and elsewhere. *Biol Soc Wash Bull* **6**, 117-158.
2. Childress JJ & Fischer CR (1992) The biology of hydrothermal vent animals: physiology, biochemistry and autotrophic symbioses. *Oceanogr Mar Biol Annu Rev* **30**, 337-441.
3. Goffredi SK, Childress JJ, Lallier FH & Desaulniers NT (1998) How to be the perfect host: CO<sub>2</sub> and HS<sup>-</sup> accumulation and H<sup>+</sup> elimination in the hydrothermal vent tubeworm *R. pachyptila*. *Cah Biol Mar* **39**, 297-300.
4. Childress JJ, Lee RW, Sanders NK, Felbeck H, Oros DR, Toulmond A, Desbruyères D, Kennicut II MC & Brooks J (1993) Inorganic carbon uptake in hydrothermal vent tubeworms facilitated by high environmental pCO<sub>2</sub>. *Nature* **362**, 147-149.
5. Goffredi S, Childress J, Desaulniers N, Lee R, Lallier F & Hammond D (1997) Inorganic carbon acquisition by the hydrothermal vent tubeworm *Riftia pachyptila* depends upon high external PCO<sub>2</sub> and upon proton-equivalent ion transport by the worm. *J Exp Biol* **200**, 883-896.
6. Goffredi SK, Girguis PR, Childress JJ & Desaulniers NT (1999) Physiological Functioning of Carbonic Anhydrase in the Hydrothermal Vent Tubeworm *Riftia pachyptila*. *Biol Bull* **196**, 257-264.
7. Kochevar RE & Childress JJ (1996) Carbonic anhydrase in deepsea chemoautotrophic symbioses. *Mar Biol* **125**, 375-383.
8. Toulmond A, Lallier FH, De Frescheville J, Childress JJ, Lee R, Sanders NK & Desbruyères D (1994) Unusual carbon dioxide-combining properties of body fluids in the hydrothermal vent tubeworm *Riftia pachyptila*. *Deep-Sea Res* **10**, 1447-1456.
9. Scott KM, Bright M, Macko SA & Fisher CR (1999) Carbon dioxide use by chemoautotrophic endosymbionts of hydrothermal vent vestimentiferans: Affinities for carbon dioxide, absence of carboxysomes, and delta super(13)C values. *Mar Biol* **135**, 25-34.
10. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, Albrecht D, Robidart J, Bench S, Feldman RA, Hecker M & Schweder T (2007) Physiological Proteomics of the Uncultured Endosymbiont of *Riftia pachyptila*. *Science* **315**, 247-250.
11. Smith KS, Jakubzick C, Whittam TS & Ferry JG (1999) Carbonic anhydrase is an ancient enzyme widespread in prokaryotes. *Proc Natl Acad Sci* **96**, 15184–15189.

12. Raven JA (1991) Implications of inorganic carbon utilization: ecology, evolution, and geochemistry. *Can J Bot* **69**, 908-924.
13. Weis VM, Smith GJ & Muscatine L (1989) A "CO<sub>2</sub> supply" mechanism in zooxanthellate cnidarians: role of carbonic anhydrase. *Mar Biol* **100**, 195-202.
14. Kochevar RE, Govind NS & Childress JJ (1993) Identification and characterization of two carbonic anhydrases from the hydrothermal vent tubeworm *Riftia pachyptila* Jones. *Molec Mar Biol Biotechnol* **2**, 10-19.
15. De Cian MC, Bailly X, Morales J, Strub JM, Van Dorsselaer A & Lallier FH (2003) Characterization of carbonic anhydrases from *Riftia pachyptila*, a symbiotic invertebrate from deep-sea hydrothermal vents. *Proteins* **51**, 327-339.
16. De Cian MC, Andersen AC, Bailly X & Lallier FH (2003) Expression and localization of carbonic anhydrase and ATPases in the symbiotic tubeworm *Riftia pachyptila*. *J Exp Biol* **206**, 399-409.
17. De Cian MC, Andersen AC, Toullec JY, Biegala I, Caprais JC, Shillito B & Lallier FH (2003) Isolated bacteriocyte cell suspensions from the hydrothermal-vent tubeworm *Riftia pachyptila*, a potent tool for cellular physiology in a chemoautotrophic symbiosis. *Mar Biol* **142**, 141-151.
18. Gattiker A, Gasteiger E & Bairoch A (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl Bioinf* **1**, 107-108.
19. Pesole G & Liuni S (1999) Internet resources for the functional analysis of 5' and 3' end untranslated regions of eukaryotic mRNA. *TIG* **15**, 378.
20. Weis VM & Reynolds WS (1999) Carbonic Anhydrase Expression and Synthesis in the Sea Anemone *Anthopleura elegantissima* Are Enhanced by the Presence of Dinoflagellate Symbionts. *Physiol Bioch Zool* **72**, 307-316.
21. Nussbaumer AD, Fisher CR & Bright M (2006) Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**, 345-348.
22. Falquet L, Pagni M, Bucher P, Hulo N, Sigrist CJ, Hofman K & Bairoch A (2002) The PROSITE database, its status in 2002. *Nucleic Acids Res* **30**, 235-238.
23. Vagner S, Galy B & Pyronnet S (2001) Irresistible IRES. *EMBO reports* **2**, 893-898.
24. Créancier L, Morello D, Mercier P & Prats AC (2000) Fibroblast Growth Factor 2 Internal Ribosome Entry Site (IRES) Activity Ex Vivo and in Transgenic Mice Reveals a Stringent Tissue-specific Regulation. *J Cell Biol* **150**, 275-281.

25. deBoer ML, Krupp DA & Weis VM (2006) Two Atypical Carbonic Anhydrase Homologs From the Planula Larva of the Scleractinian Coral *Fungia Scutaria*. *Biol Bull* **211**, 18-30.
26. Seron TJ, Hill J & Linser PJ (2004) A GPI-linked carbonic anhydrase expressed in the larval mosquito midgut. *J Exp Biol* **207**, 4559-4572.
27. Silverman DN & Lindskog S (1988) The catalytic mechanism of carbonic anhydrase: implications of a rate-limiting protolysis of water. *Acc Chem Res* **21**, 30-36.
28. Tu C, Silverman DN, Forsman C, Jonsson BH & Lindskog S (1989) Role of histidine 64 in the catalytic mechanism of human carbonic anhydrase II studied with a site-specific mutant. *Biochemistry* **28**, 7913-7918.
29. Liang J-Y & Lipscomb WN (1987) Hydration of carbon dioxide by carbonic anhydrase: internal proton transfer of  $\text{Zn}^{2+}$ -bound  $\text{HCO}_3^-$ . *Biochemistry* **26**, 5293-5301.
30. Briganti F, Mangani S, Orioli P, Scozzafava A, Vernaglione G & Supuran CT (1997) Carbonic Anhydrase Activators: X-ray Crystallographic and Spectroscopic Investigations for the Interaction of Isozymes I and II with Histamine. *Biochemistry* **36**, 10384-10392.
31. Silverman DN (1991) The catalytic mechanism of carbonic anhydrase. *Can J Bot* **69**, 1070-1078.
32. Hewett-Emmett D, Hopkins PJ, Tashian RE & Czelusniak J (1984) Origins and molecular evolution of the carbonic anhydrase isozymes. *Ann N Y Acad Sci* **429**, 388-358.
33. Hewett-Emmett D & Tashian RE (1996) Functional diversity, conservation, and convergence in the evolution of the alpha-, beta-, and gamma-carbonic anhydrase gene families. *Mol Phylogenet Evol* **5**, 50-77.
34. del Pilar Corena M, Seron TJ, Lehman HK, Ochrietor JD, Kohn A, Tu C & Linser PJ (2002) Carbonic anhydrase in the midgut of larval *Aedes aegypti*: cloning, localization and inhibition. *J Exp Biol* **205**, 591-602.
35. Diatchenko L, Lau Y-FC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED & Siebert PD (1996) Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* **93**, 6025-6030.
36. Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL, Lukyanov SA & Sverdlov ED (1996) Equalizing cDNA Subtraction Based on Selective Suppression of Polymerase Chain Reaction: Cloning of Jurkat Cell Transcripts Induced by Phytohemagglutinin and Phorbol 12-Myristate 13-Acetate. *Anal Biochem* **240**, 90-97.

37. Halanych KM, Lutz RA & Vrijenhoek RC (1998) Evolutionary origins and age of vestimentiferan tube-worms. *Cah Biol Mar* **39**, 355-358.
38. Livak KJ & Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods* **25**, 402–408.
39. Not F, Simon N, Biegala IC & Vaulot D (2002) Application of fluorescent in situ hybridization coupled with tyramide signal amplification (FISH-TSA) to assess eukaryotic picoplankton composition. *Aquat Microb Ecol* **28**, 157–166.
40. Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673-4680.
41. Kumar S, Tamura K & Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinf* **5**, 150-163.

SUPPLEMENTARY MATERIAL

**Table S1: Best Blastx Hits obtained for the identification of RpCAbr transcript.**

Putative identification on Blastx	GenBank acc. numb.	E-value	Identities (%)	Positives (%)
<i>Riftia pachyptila</i> CA	CAD29128.1	$3.10^{-95}$	67	81
<i>Anthopleura elegantissima</i> CA	AF140537	$2.10^{-52}$	43	59
<i>Strongylocentrotus purpuratus</i> CA	XP_782997.2	$5.10^{-51}$	41	57
<i>Petromyzon marinus</i> CA	AAZ83742.1	$8.10^{-51}$	44	59
<i>Drosophila pseudoobscura</i> CA	XP_001356245.1	$6.10^{-49}$	42	57
<i>Lepisosteus osseus</i> CA	AAM94169.1	$1.10^{-48}$	42	57
<i>Drosophila simulans</i> CA	AAY56646.1	$1.10^{-48}$	42	57
<i>Drosophila melanogaster</i> CA	NP_523561.1	$4.10^{-48}$	42	57
<i>Pseudopleuronectes americanus</i> CA	AAV97962.1	$4.10^{-47}$	42	58
putative <i>Monodelphis domestica</i> CA	XP_001366749.1	$2.10^{-46}$	42	55
<i>Danio rerio</i> CA	NP_571185.1	$3.10^{-46}$	42	57
<i>Mus musculus</i> CAI	NP_033929.2	$6.10^{-46}$	41	56
<i>Mus musculus</i> CAI	P13634	$6.10^{-46}$	41	56
<i>Macaca mulatta</i> CAI	XP_001096360.1	$8.10^{-46}$	42	56
<i>Oncorhynchus mykiss</i> CA1	BAD36835.2	$1.10^{-45}$	41	58
<i>Mus musculus</i> CAII	AAH55291.1	$1.10^{-45}$	40	57

**CHAPITRE VI**  
**Troisième article**

## CHAPITRE VI : Troisième article

### Caractérisation, expression et étude fonctionnelle d'un transcrit de myohémérythrine de *Riftia pachyptila*

#### 1. Résumé

##### 1.1 Contexte de l'étude

Obtenu à partir de notre banque soustractive de trophosome (banque TR-BW) et spécifiquement exprimé dans ce tissu, le transcrit de myohémérythrine (MHr) nous a paru important sur plusieurs points. Tout d'abord, il code une protéine respiratoire autre qu'une hémoglobine (Hb). Dans notre banque TR-BW élaborée à partir d'un seul individu, des séquences codant des chaînes d'Hb extracellulaires et des séquences codant la MHr avaient été obtenues simultanément : si les séquences sont traduites, on peut supposer la présence de deux types de pigments respiratoires chez *Riftia pachyptila*. Chez les Siponcles, le transport d'O<sub>2</sub> est réalisé exclusivement par des hémérythrines (Hr) et le stockage dans les tissus par la MHr. Cependant, relativement peu d'annélides possèdent ce pigment dans l'état actuel de nos connaissances. Sur le plan évolutif il semblait donc nécessaire de resituer cette séquence de MHr dans un contexte phylogénétique. Par ailleurs, sur le plan fonctionnel, la liaison à l'O<sub>2</sub> n'est peut être pas la seule fonction de cette protéine : des études réalisées sur des annélides signalent en effet des propriétés de liaison au cadmium (chez *Nereis diversicolor*) et des propriétés anti-bactériennes (chez *Hirudo medicinalis*). Dans le cadre de notre problématique, ces éléments sont intéressants du fait (i) de l'exposition du ver à des métaux lourds présents dans l'environnement tels que le cadmium, et surtout (ii) la présence exclusive des bactéries symbiotiques dans le trophosome de *R. pachyptila*. La séquence codante complète a pu être surexprimée pour réaliser des tests d'activité superoxide dismutase.

## 1.2 Principaux résultats

Le gène de la MHr de *R. pachyptila* (RpMHr) est exclusivement exprimé au niveau du trophosome (pas de détection du transcrit dans des échantillons de branchie et paroi du corps). Les analyses de PCR quantitative montrent également un niveau d'expression supérieur dans le trophosome des individus des sites Rehu Marka, Susie et Miss WormWood (Sud-EPR) par rapport au trophosome des individus du site Genesis (Nord-EPR). Ce résultat pourrait peut-être être expliqué par des différences de concentrations en métaux entre les deux sites. L'alignement de la séquence traduite de RpMHr avec d'autres séquences protéiques d'Hr and MHr de Sipunculides et Annélides révèle que les acides aminés participant à la fixation de l'O<sub>2</sub> sur la protéine sont bien conservés. La séquence primaire de RpMHr apparaît phylogénétiquement proche des séquences d'Achètes. Alors qu'une couleur rose caractéristique de l'état oxygéné de la protéine Hr a pu être observée lors du protocole de surexpression de RpMHr, le spectre d'absorbance n'a pas révélé l'existence de pics traduisant la fixation de ligands au site actif, au contraire de l'Hr de Siponcle utilisée comme contrôle. Pourtant, une activité superoxide dismutase semblerait bien présente.

## **2. Article en préparation**



**Characterization, expression, and functional study of the myohemerythrin  
from the hydrothermal vent tubeworm *Riftia pachyptila* (Polychaeta,  
Siboglinidae)**

Sophie Sanchez, Thomas Harnois, Stéphane Hourdez and François H. Lallier\*

Equipe Ecophysiologie: Adaptation et Evolution Moléculaires,

UMR 7144 CNRS-UPMC

Station Biologique, Place Georges Teissier, BP 74, 29682 Roscoff Cedex, France

\* corresponding author

Tel +33 2 9829 2311

Fax +33 2 9829 2324

E-mail lallier@sb-roscoff.fr

Short title: Myohemerythrin in *Riftia*

Keywords: symbiosis; chemoautotrophy; Siboglinidae; messenger RNA; differential expression

Note: nucleotide sequence data available in the GenBank database under the accession number EF648563.

**ABSTRACT**

The annelid *Riftia pachyptila* lives close to hydrothermal vents along the East Pacific Rise at 2600 meter-depth, and harbors CO<sub>2</sub>-assimilating sulfide-oxidizing bacterial symbionts in an internal organ, the trophosome, that provide all of its nutrition. A Subtractive Suppression Hybridization (SSH) approach was used to construct subtractive libraries in order to find specific host transcripts which could be important for the functioning of this symbiosis. From the trophosome specific library, we obtained a myohemerythrin (MHr) transcript which appeared exclusively present in this tissue compared to other tissues of the worm (branchial plume and body wall). Our quantification analyses also showed differential expression between individuals from two different vent sites that may be related to the physiological condition of the trophosome. Protein sequences alignment of hemerythrin (Hr) and MHr of several annelids and sipunculids revealed the conservation of amino acids involved in the active site for iron-binding and O<sub>2</sub>-binding. Although a pink color was observed during the overexpression protocol (typical of a Hr or MHr with bound O<sub>2</sub>), the purified RpMHr did not seem to be able to reversibly bind O<sub>2</sub>. We were however able to measure a superoxide dismutase activity.

## INTRODUCTION

The Siboglinidae *Riftia pachyptila* (Annelida; Polychaeta) is a tubeworm endemic of hydrothermal vents from the East Pacific Rise, living at depths of ~2,600 meters. This worm has no mouth, digestive tract, and anus [1], and rely completely on its autotrophic sulfide-oxidizing symbionts to live [2]. These symbionts are located deeply inside the body of the host, in the cells (bacteriocytes) of a specialized organ called the trophosome. The metabolites CO<sub>2</sub>, H<sub>2</sub>S, O<sub>2</sub> and NO<sub>3</sub><sup>-</sup> are acquired from the environment through the branchial plume [3, 4], the only organ in contact with the surrounding water. Once in the body fluids of the worm, O<sub>2</sub>, H<sub>2</sub>S and possibly NO<sub>3</sub><sup>-</sup> are bound to extracellular hemoglobins (Hb) whereas most of the CO<sub>2</sub> is transported as dissolved HCO<sub>3</sub><sup>-</sup>. Two vascular Hbs (V1 and V2) and one coelomic Hb are known in *Riftia pachyptila*: they are all able to reversibly and simultaneously bind O<sub>2</sub> and H<sub>2</sub>S on two different sites of the protein [5]. Hbs can transport the two molecules to the bacteriocytes where they are released to be used by the symbionts. The highly vascularized trophosome lobules must be the site of intense cellular exchanges, between body fluids and bacteriocytes, and between bacteriocytes and vacuolized bacteria, to provide the bacteria with the necessary molecules for their chemoautotrophic metabolism and to recover organic compounds from the bacteria.

In an attempt to identify yet unknown host proteins involved in branchial and trophosome functions associated with the symbiotic mode of life of *R. pachyptila*, we constructed subtractive tissue-specific cDNA libraries (Subtractive Suppression Hybridization, SSH) (Sanchez et al., submitted). Among other cDNAs, we obtained a complete transcript coding for a myohemerythrin (MHr) in the trophosome subtractive library.

Oxygen-binding proteins provide oxygen transport to and oxygen storage in the tissues of metazoan organisms. The binding site can contain either iron or copper. Iron can be found in intra- or extracellular circulating hemoglobin (Hb), chlorocruorin, myoglobin (Mb), intracellular circulating hemerythrin (Hr), and myohemerythrin (MHr). Copper is found in the molluscan and arthropod hemocyanins (Hc) (for a review, see [6]). Although globins (Hb and Mb) and hemerythrins (Hr and MHr) use iron in their active site, there differ on some points: Hbs and Mbs use a heme group in which one atom of iron is the metal, whereas in Hrs and MHrs, there are two iron atoms, no heme group, and the iron atoms are directly bound the protein through covalent bonds. Hr is a homopolymer consisting of polypeptides, 113 to 117

amino acid-long, and is contained in circulating cells, the hemerythrocytes, in the vascular system [7]. MHr is a monomeric protein of 120 amino acids classically recovered in muscles [8, 9] and is involved in intracellular oxygen transfer and storage. Hemerythrins are common in Sipuncula, Brachipoda and Priapulida [6, 10, 11]. Although the great majority of annelids possess globins (Hb and Mb), species of the polychaete family Magellonidae possess Hr as a respiratory pigment [8]. In addition, a few species of annelids in the polychaete family Nereididae and the achaete family Glossiphoniidae are now known to express a form of Hr (referred to Hr or MHr, according to the studies) [12-14]. It is unlikely that these are used either as oxygen storage or as an oxygen transporter (these species possess Hbs and Mbs in much greater amounts). In *Nereis diversicolor*, the expression of the myohemerythrin has been shown to increase after exposure to cadmium and to a bacterial infection [11, 25-27]. This antibacterial function has also been shown in some leech [15]. In the bacterium *Desulfovibrio vulgaris*, a hemerythrin-like protein has a superoxide dismutase activity, probably protecting the cells from oxidative damage [16].

Here, we determined the relative expression level of this myohemerythrin gene from different tissues of the worm and particularly from trophosome cDNA samples of worms collected from two vent sites. Alignment of protein sequences from several annelid and sipunculid myohemerythrins was performed to investigate the conservation of important amino acids for oxygen binding and to perform phylogenetic reconstruction. Knowing the broad range of biological functions have been reported in MHr other than O<sub>2</sub> binding, we also overexpressed this sequence to test some functional properties of the protein.

## RESULTS

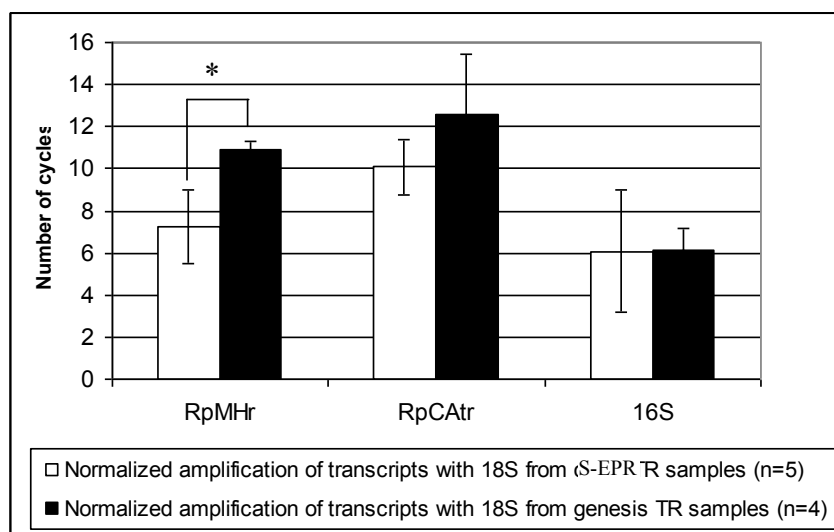
### *Complete coding myohemerythrin transcript and partial 5' and 3'UTRs*

We recovered a contig of 5 sequences from the body wall-subtracted trophosome cDNA library that comprises of a small 5'UTR of 45 nucleotides (probably a partial sequence), an open reading frame of 363 nucleotides thus coding for a 120 amino acids-protein, followed by a very long 3' UTR of 732 nucleotides. A poly-A tail signal (AAUAAA) occurred 336 nucleotides downstream from the in-frame stop codon, but no poly-A tail could be sequenced (we probably have a partial 3'UTR). Search of motifs with the PROSITE server (ScanProsite: [17]) showed the presence of a hemerythrin signature from amino acids 56 to 79: H-F-x(2)-[EQ]-[ENQ]-x(2)-[LMF]-x(4,7)-[FY]-x(5,6)-H-x(3)-[HR]. Results of homology by BlastX analysis on the NCBI server resulted in a high homology score with the

myohemerythrin cDNA sequences from the Sipunculidae *Phascolopsis gouldii* (acc. # P27686) and *Sipunculus nudus* (acc. # CAG14944). The *Riftia pachyptila* MHR sequence is hereafter referred to as RpMHR (acc. # EF648563), standing for “*Riftia pachyptila* Myohemerythrin”.

### ***Tissue-specific expression of myohemerythrin***

We compared the expression levels between individuals coming from the South sites (S-EPR) and individuals coming from the Genesis site (N-EPR). No amplification of RpMHR could be detected from branchial plume cDNA samples nor from body wall cDNA samples. In contrast, amplification was successfully for all trophosome cDNA samples studied. RpMHR amplification reaches a fluorescence threshold after 7.23 ( $\pm 1.77$ ) cycles for trophosome cDNA of individuals coming from the S-EPR sites, and after 10.88 ( $\pm 0.46$ ) cycles for trophosome cDNA of individuals coming from the Genesis site (Fig. 1).

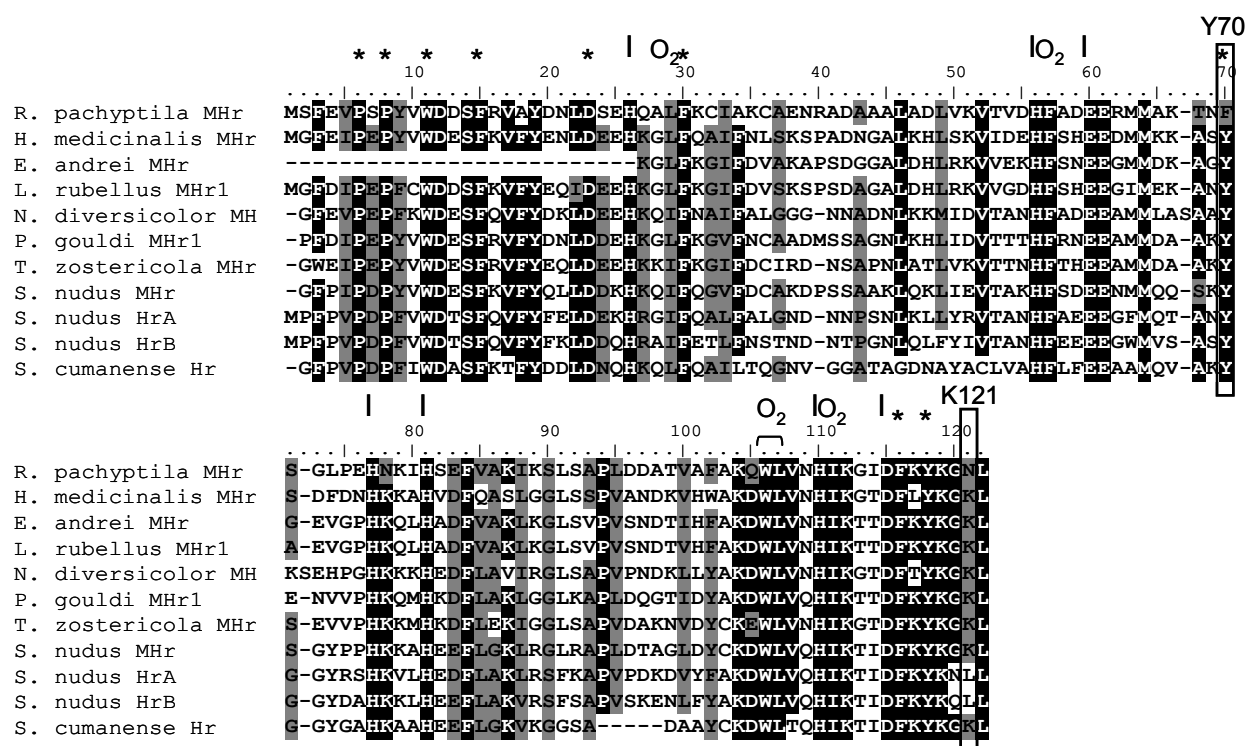


**Fig. 1:** Normalized amplifications of RpMHR, RpCAtr and 16S with 18S amplification from trophosome (TR) samples belonging to individuals from the S-EPR (n=5) and Genesis (n=4) sites. The number of cycles on the Y axis is the difference between the number of cycles required to amplify each transcript and the number of cycles required to amplify 18S. Asterisks indicate significant differences (Student's t-test,  $p < 0.01$ ).

This difference of 2.65 cycles is significant (Student's t test,  $p < 0.01$ ) and corresponds to a 12-fold higher expression of RpMHR in the trophosome of S-EPR individuals compared to the trophosome of Genesis individuals. On average, the carbonic anhydrase (RpCAtr, [18]) and the 16S [19] do not show a significant expression level difference between worms of the two vent sites.

### Alignment and phylogenetic reconstruction

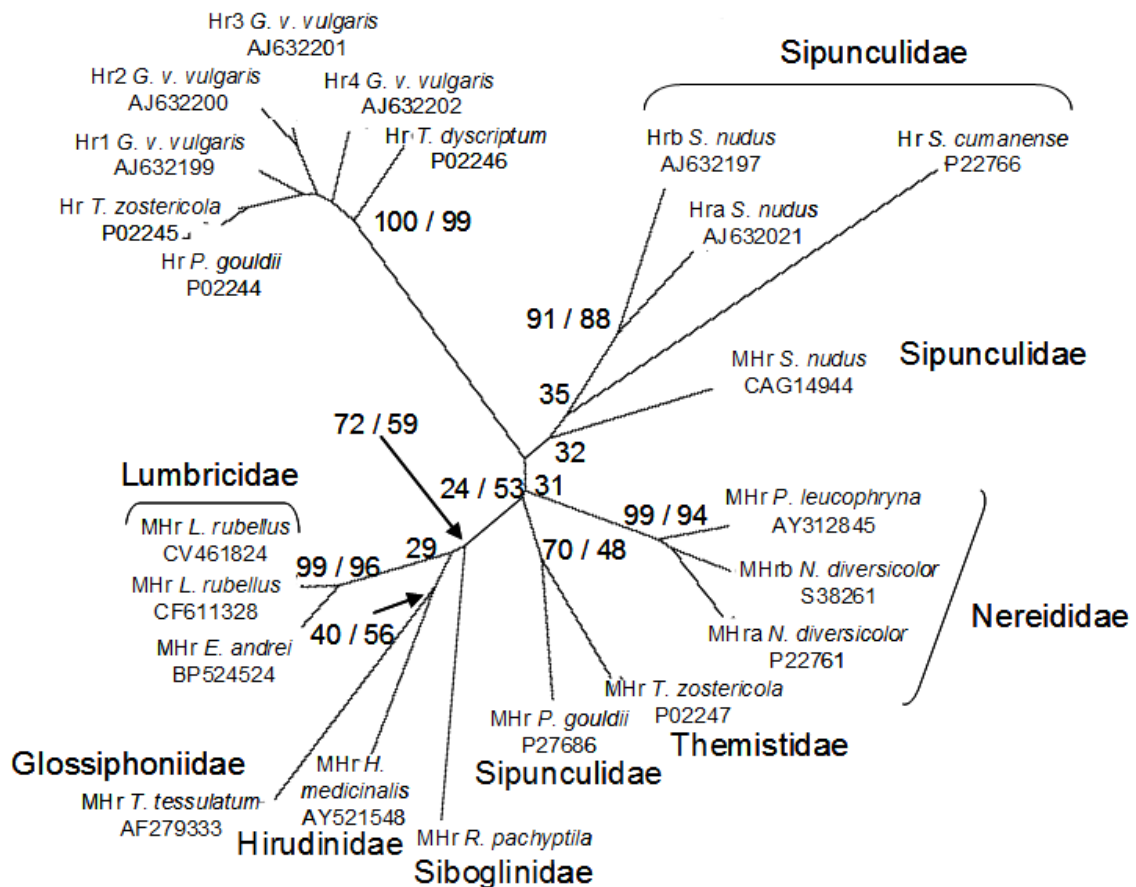
For alignment and phylogenetic reconstruction, Annelida MHR sequences were represented by the Nereididae (*Nereis diversicolor*, *Periserrula leucophryna*), the Hirudinidae (*Hirudo medicinalis*), and the Glossiphoniidae (*Theromyzon tessulatum*) families. Sipunculida MHR sequences were represented by the Themistidae (*Themiste zostericola*) and Sipunculidae (*Sipunculus nudus*, *Phascolopsis gouldii*, and *Siphonosome cumanense*) families. We also used Hr sequences from Annelida and Sipunculida coding for monomeric, trimeric and octameric proteins. Lastly, we added some Annelida sequences (from *Lumbricus rubellus* and *Eisenia andrei*) recently available in the databases from EST libraries. All these sequences were unambiguously aligned: an alignment with a reduced data set is presented in Fig. 2.



**Fig. 2:** Alignment of complete RpMHR amino acids sequence with some representative protostomian MHR and Hr protein sequences. Identical and similar amino acids shared by at least 80% of the isoforms are shown in black and grey, respectively. Residues involved in iron binding in the catalytic site are indicated by a "I"; important amino acids involved in the structure are indicated by a "\*"; residues involved in the O<sub>2</sub>-binding pocket are indicated by a "O<sub>2</sub>"; framed amino acids are commented in the section "Results" and positions indicated above the frame refer to the reference positions in RpMHR sequence.

Over all the sequences that were aligned (not all represented in Fig. 2 but in the phylogenetic tree in Fig. 3), 26 sites out of 123 are conserved (21.1 % identity). Over all MHR sequences

(all represented in Fig. 2 and Fig. 3), 38 sites are conserved (30.9 % identity). The *Lumbricus rubellus* MHR (acc. # CV461824) and RpMHR sequences on the one hand, and the *Themiste zostericola* MHR and RpMHR sequences on the other hand, share nearly the same number of amino acids (62 and 63 respectively) giving 50.4 % and 51.2 % amino acids identity respectively. All MHR sequences share a five-residue segment which is usually missing in Hr sequences (see the *S. cumanense* sequence), except in *S. nudus* HrA and HrB sequences. All the residues involved in iron binding (H26, H56, D/E60, H77, H81, H110, D115) [8] (using positions in RpMHR sequence as a reference from here on) are strictly conserved in the RpMHR amino acids sequence. Residues that have been indicated to be important for the structure (P6, P8, W11, F15, D23, F30, F116 and Y118) were also found in RpMHR sequence except the Y70 which is replaced by a Phenylalanine (F) amino acid. The K121 is found in all sequences excepted *S. nudus* HrA and HrB (L) and RpMHR where it is replaced by a similar amino acid (N). Finally, all residues involved in the O<sub>2</sub>-binding pocket (L29, F57, W106, L107, I111) [14, 20] are conserved in the RpMHR sequence.

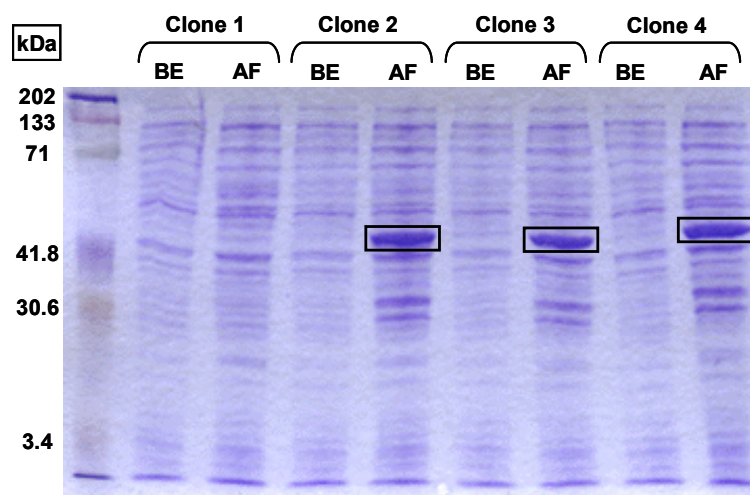


**Fig. 3 (previous page):** Neighbor-joining (NJ) tree obtained after a multiple alignment of protostomian MHR and Hr amino acids sequences. Some nodes were also found in the Maximum Parsimony (MP) analysis. Numbers are bootstrap values (BP) calculated from 1000 replicates from NJ ( $BP_{NJ}$ ) and MP ( $BP_{MP}$ ) analyses and are represented as ( $BP_{NJ} / BP_{MP}$ ). Nodes with only one number ( $BP_{NJ}$ ) are only found in the NJ tree.

A phylogeny of the Hr and MHR was constructed by NJ and MP (Fig. 3). The octameric Hr from the Sipuncula families Golfingiidae (*Golfingia*), Sipunculidae (*Phascolopsis*), and Themistidae (*Themiste*) form a very well supported group. RpMHR forms a monophyletic group with sequences from Lumbricidae (oligochaetes) and achaetes ( $BP_{NJ} = 72$  and  $BP_{MP} = 59$ ) from which the Nereididae (polychaetes) are excluded. Annelida MHR sequences form a paraphyletic group in this phylogeny, with some (but not all) sipunculid MHR mixed. A monophyly of the Annelida could only be obtained by removing the oligochaete (Lumbricidae) sequences, although the bootstrap support was weak in the MP tree (data not shown,  $BP_{NJ}=69$  and  $BP_{MP}=12$ ).

### Functional analysis

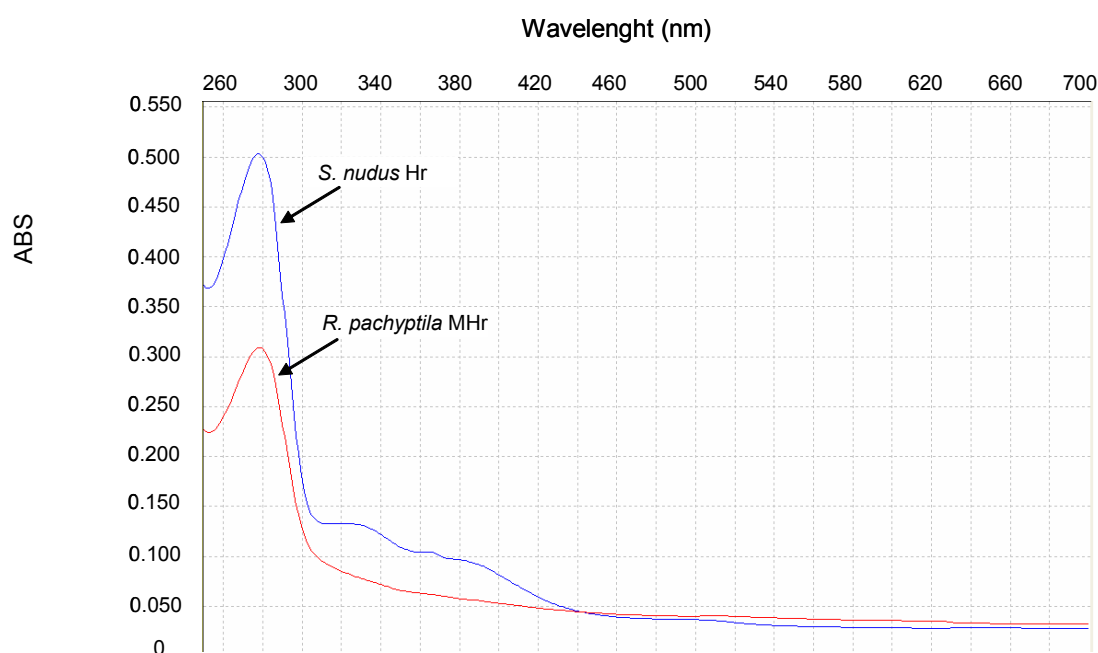
In order to study the protein function into greater detail, the myohemerythrin cDNA sequence was cloned and overexpressed in *Escherichia coli* (BL21 StarTM (DE3)) as a GST fusion protein. Migrations on a polyacrylamide gel of the proteins from crude extracts of bacteria containing the GST-RpMHR overexpression plasmid prior to the induction (BE), and after 2-3 days of growth in an autoinducible medium (AF) are presented in Fig. 4.



**Fig. 4:** Migrations of bacterial extracts before (BE) and after (AF) induction for four different clones. Clone 1 did not show any difference before and after induction whereas the three others (Clones 2, 3 and 4) showed the presence of GST-RpMHR overexpressed fusion protein (at about 45 kDa) only after induction. The other bands at lower molecular weights that also differ between before and after induction could be the result of degradation.



After induction, we can observe a strong band at 45 kDa compared to the migration profile before induction. This is close to the expected molecular mass for the fusion protein 40.51 predicted from the RpMHr amino acid sequence (13.51 kDa) and the molecular mass of the GST protein (27 kDa). This mass is typical of purified MHrs from the Nereididae *N. diversicolor* [9] and the sipunculan *Dendrostomum pyroides* [21]. During the purification process on glutathion-Sepharose 4B beads, we observed a pink coloration of the beads following the fixation of the GST-RpMHr fusion protein. This color frequently disappeared during washings but the protein was not eluted. This would indicate an oxygenation/deoxygenation during the washings (data not shown). We measured the absorption spectrum of the overexpressed GST-RpMHr fusion protein and compared it to that of *S. nudus* after purification (Fig. 5).



**Fig. 5:** Absorption spectra obtained for *Riftia pachyptila* GST-RpMHr and *Sipunculus nudus* Hr. No other peak than the one corresponding to proteins (278 nm) could be detected for RpMHr whereas three peaks at 323, 365 and 502 nm were detected for *S. nudus* Hr.

For the two samples analyzed, the peak at 278 nm corresponds to the presence of protein in the solution. Peaks at 323 nm, 365 nm and 502 nm are detected on the spectrum of the *S. nudus* Hr. No such peaks could be detected during the analysis of GST-RpMHr. This could indicate that the protein is not associated to oxygen anymore.

These samples and an *Arenicola marina* hemoglobin (Hb) sample were analyzed to detect a possible SOD activity following Oberley and coworkers' [22] protocol. The activity measured in the GST-RpMHR fusion protein was similar to that found in *Arenicola marina* HBL-Hb but only 1/3 of that found in *Sipunculus nudus* Hr (Table 1).

**Table 1: SOD activities**

	<i>Arenicola marina</i>	<i>Sipunculus nudus</i>	<i>Riftia pachyptila</i>
	HBLHb	Hr	MHR*
<b>SOD activity (U.g<sup>-1</sup>)</b>	1846	6133	2174

\*Calculations were made by removing the GST contribution to the total mass of the fused protein that was used.

## DISCUSSION

### *Expression of the RpMHR gene*

The RpMHR transcript clearly appears tissue-specific in our study as it could only be amplified from trophosome cDNA samples and not from branchial plume and body wall cDNA samples. The trophosome tissue of Genesis site individuals had a dark aspect (almost black) in comparison to the trophosome tissue of the S-EPR site individuals (data not shown). This difference of color has been shown to correspond to a different content of elemental sulfur [23, 24], thought to be storage for the symbionts' metabolic needs. Interestingly, the carbon fixation rate of lighter colored trophosome was often higher than that of the dark-colored trophosomes, although the reason is not clear (KT Scott, pers. com.). Individuals coming from the S-EPR site (light-colored trophosome) had a higher RpMHR expression levels compared to individuals from the Genesis site. As a control, no difference of expression level was observed from the analysis of the 16S transcript. On average, the RpCAtr transcript was also less expressed in the Genesis worms (although the difference is not significant due a large variability among individuals), a lesser expression that could be coherent with the lower carbon fixation rate in the dark trophosomes. Additional physiology and gene expression level studies are needed to better understand the functions of the proteins specifically expressed in the trophosome.

### *Phylogenetic reconstruction*

Our phylogenetic reconstruction showed that the *Hirudo medicinalis* sequence (mentioned in the databases as a Hr) could be a MHR isoform, as previously suggested by Vanin and coworkers [25]. The RpMHR and Lumbricidae sequences grouped more closely with known MHR. However, we recovered the monophyly of MHR Annelida sequences only when Oligochaeta sequences were not included in the phylogenetic reconstruction. The position of RpMHR sequence in a monophyletic group comprising Achaeta MHR sequences is well supported in both NJ and MP analyses performed. However, it did not group closely to the other Polychaeta MHR sequences (Nereididae). The sipunculan MHR sequences did not form a monophyletic group as previously shown [25]. Additional *L. rubellus* sequences identified as Hr exist (acc. # CO046773 and DR076331) but they were finally not included in our analysis because of their very long branches which could have induced long branch attraction. However, they seemed to emerge close to the monomeric and trimeric Hr sequences (represented by *S. nudus* and *S. cumanense* sequences, data not shown) instead of being close to the other *L. rubellus* sequences included in our analysis (Fig. 3). The addition of new Hr and MHR annelid sequences may improve the resolution of the relationships among monomeric and trimeric Hr sequences in further phylogenetic reconstructions. The polychaete family Magelonidae is known for using Hr as the sole oxygen carrying protein in its body but no sequence is currently available. This sequence is needed to address the orthologous status of the different sequences studied here.

### ***O<sub>2</sub> binding in S. nudus Hr and GST-RpMHR***

Between the deoxyHr and oxyHr states, other states are possible. One of them, the metHr, can occur: it corresponds to the chemical oxidation of deoxyHr in the absence of O<sub>2</sub> (for a review, see [26]). Resonance Raman studies of oxyHr established that the oxidation level of the bound O<sub>2</sub> to iron center is the peroxide (O<sub>2</sub><sup>2-</sup>) ion. Once oxygenation takes place, the diferrous center in deoxyHr is oxidized to the diferric form in oxyHr: the bound O<sub>2</sub> is reduced to O<sub>2</sub><sup>2-</sup>, giving rise to absorbance bands at 330, 360 and 500 nm. Ultraviolet-visible spectra of oxidized Hr classically display four peaks: at 320-340 nm, 360-380 nm, 440-510 nm and 590-750 nm [27]. The UV-visible spectrum of *S. nudus* Hr corresponds well with this expectation, although no peak was detected after 590 nm. In the sipunculan *Dendrostomum pyroides*, similar spectra (slightly different wavelengths) were obtained for which the authors concluded they were spectra characteristic of metazidehemerythrin metN<sub>3</sub>- [21]. The spectra were realized for the Hr, MHR and coelomic Hr that co-exist in the animal. For all these three proteins, the absorption spectra were nearly identical: all three proteins (MHR included)

showed the characteristic optical absorption and circular dichroism properties of hemerythrin [21]. We expected similar results for RpMHR and *S. nudus* Hr, but our overexpressed RpMHR did not exhibit any peak typical of oxyMHR. Additional experiments under oxic and anoxic conditions showed the capacity of *S. nudus* Hr to reversibly bind O<sub>2</sub> (change of the UV-visible absorption spectrum) whereas GST-RpMHR fusion protein did not reveal significant binding or release of O<sub>2</sub> (data not shown). We can then wonder if the GST-RpMHR fusion protein is in a spatial conformation which prevents the accessibility of O<sub>2</sub> to the iron-binding sites or strongly decreases its capacity to keep bound oxygen. However, the color observed during the purification of the GST-RpMHR fusion protein (at very high concentration) is typical of the violet-pink oxyhemerythrin [27]. The protein is colorless when the iron is in the Fe(II) state (deoxygenated state) [10, 28] while the violet-pink protein is typical of the oxygenated state [27]. The conformation of the protein could have changed in the elution buffer and a following buffer change to Riftia saline buffer did not return the protein to its oxygenated state.

### ***MHR protein function***

MHR is typically found in muscles of the sipunculans that also harbor hemerythrocytes (circulating cells that contain Hr). However, MHRs are not restricted to these organisms and can also be found in Annelida that use hemoglobin as a respiratory pigment such as *Riftia pachyptila*. The two iron atoms in MHR are coordinated by five histidine residues which appeared all conserved in the RpMHR sequence, therefore suggesting the probable functionality of the protein if translated. Several studies showed that MHRs did not have oxygen binding capacity as sole function, but could act as a detoxification agent (through a cadmium binding capacity) in the polychaete annelid *N. diversicolor* [29]. In addition, one of the *Lumbricus rubellus* sequences (acc. # CF611328) used in our phylogenetic reconstruction and which emerged close to the other Annelida MHR sequences was obtained from the cadmium-exposed earthworm cDNA library (J. Chaseley and coworkers, unpublished results). Anti-bacterial properties of MHRs were demonstrated in the leech *Hirudo medicinalis* [14, 30] and more recently in the sand worm *N. diversicolor* [31]. We did not test directly any anti-bacterial capacity of the overexpressed RpMHR. Other functions than O<sub>2</sub> binding have been reported for Hr and Hr-like proteins. For example, a non-heme iron protein (rubrerythrin) with hemerythrin-like binuclear iron cluster had been isolated from the anaerobic sulfate-reducing bacterium *Desulfovibrio vulgaris*. SOD activity assays demonstrated that the protein can play a role as a scavenger of oxygen radicals [16]. Our results on RpMHR protein seemed to

corroborate this tendency. It really seems that these proteins are able to play multiple biological functions in protecting the cells from the oxygen radicals (following changes of anaerobic to aerobic environment). It fits well with *R. pachyptila* biology. Indeed, the trophosome, housing the symbiotic bacteria, is particularly exposed to the compounds released by the sulfur-oxidizing bacteria, and to sulfide, known for producing free radicals in the presence of metals.

## CONCLUSION

*R. pachyptila* MHR is specifically expressed in the trophosome. It showed high homology with other annelid MHR sequences and emerged close to sequences from achaetes in our phylogenetic reconstruction. Although RpMHR showed SOD activity, we do not know if this activity is a primary function of the protein. The role of MHR in *Riftia* is not clear but its sole localization in the trophosome, as well as its antibacterial function in other species raises the possibility of a direct link with the bacterial activity in this tissue.

## EXPERIMENTAL PROCEDURES

### *Animal collection and sampling*

Specimens of *Riftia pachyptila* coming from the South-EPR sites were collected at the Rehu Marka (17°25'S – 113°12'W), Susie and Miss WormWood (17°35'S - 113°14'W) sites at 2600 meters-depth along the South East Pacific Rise during the BIOSPEEDO 2004 cruise. Additional specimens of *R. pachyptila* collected at the Genesis site (12°48'N, 103°56'W, 2500 m depth) during the EPR2003 cruise were used for quantitative PCR experiments. For each individual, parts of the branchial plume, trophosome and body wall tissues were isolated on ice, placed in RNAlater (Ambion) for 24-h at 4°C and frozen in liquid nitrogen.

### *RNA extraction*

Plume, trophosome and body wall tissue samples were pulverized individually in liquid nitrogen under RNase-free conditions. For each tissue, total RNA was extracted using the RNAlater buffer (Eurobio) following the manufacturer's instructions. Then, both for libraries constructions and complete sequencing, messenger poly-A RNAs were purified using the oligo-dT resin column of the mRNA Purification Kit (Amersham).

### ***Construction of subtractive tissue-specific cDNA libraries***

Libraries were constructed from tissues belonging to a single individual therefore representing one organism transcriptome. A total of 4 libraries were obtained: branchial plume vs body wall subtracted library (and its reciprocal) and trophosome vs body wall subtracted library (and its reciprocal). For all tissues pairs, cDNA synthesis as well as subtractive suppressive hybridization (SSH [32, 33]) comprising steps of adaptor ligation, subtractive hybridization and selective amplification were performed following the protocol of the Clontech PCR-Select™ cDNA Subtraction Kit (BD Biosciences). For each SSH procedure, the whole amplification product was cloned into TOPO®-TA cloning vector (Invitrogen), giving a range of cDNA fragment sizes.

### ***SYBR Green quantitative PCR***

#### *Reverse transcription (RT)*

From each total RNA sample (branchial plume, trophosome and body wall) fresh RT was conducted with a random primer. Each reaction mixture was composed of: 2 µl of M-MLV RT buffer; 0.5 µl of BSA (10 mg/ml), 1 µl of total RNA (1.24 µg/µl), 2.5 µl of dNTP (4mM total), 1.5 µl of Random Primer 9 (Ozyme) (100 ng/µl), 3 µl DEPC water. Then, reaction mixtures were incubated at 80°C for 5 minutes and placed on ice. M-MLV RT was added (1 µl) to each reaction mixture and all reactions were incubated at 42°C for 1 hour and finally placed on ice.

#### *Amplification*

Specific pairs of primers (Table 2) were designed using the software Primer Express to amplify the RpMHR, RpCAtr and 16S transcripts. 18S rRNA was chosen as a reference gene for the normalization of expression data and was amplified with the 18h and 18L primers previously used by Halanych and co-workers [34]. The Power SYBR Green PCR master mix (PE Applied Biosystems) was used for amplifications, with 23 µl reaction mixtures in a Chromo4™ System CFB-3240 (BIORAD). PCR reactions were performed in triplicates. Amplification conditions were 40 cycles with the following profile: 95°C for 30 s, 60°C for 30 s, and 72°C for 1 min. Standard curves were generated for 18S and the transcripts of interest over a large range of template cDNA quantity in order to calculate PCR efficiencies.

**Table 2: Primers sequences used for amplifications in this study**

<b>Amplification of transcripts by quantitative PCR</b>		
Primers*	Sequence	
MHFq	5'	AGA GGC ACA CCA ACA ACC GT 3'
MHRq	5'	CCC GAT TGG TTC ATC ACA GC 3'
RpCAtrFq	5'	GCC AGG TGT CGT CCT CGT T 3'
RpCAtrRq	5'	TCA CAA ATG TCC AGT GCC AGT T 3'
16SFq	5'	CGT AAG ACT ATA GCT GGT TTT CCA AA 3'
16SRq	5'	TTA TCA AAG ATT TTT TCT TGG TTC ACT AAT 3'
<b>Amplification of RpMHR for overexpression</b>		
Primers	Sequence	
BamHIMHF	5'	CGG GAT CCC GAT GTC TTT CGA GGT TCC TTC C 3'
EcoRIMHR	5'	CGG AAT TCT TAC AGA TTT CCC TTG TAT T 3'

\*Primers designed by Primer Express software (ABI PRISM™)

### *Data analysis*

For 18S and RpMHR, RpCAtr and 16S transcripts, the efficiency (E) was calculated from the slope (S) of the standard curve using the formula:

$$E = 10^{-1/S} - 1$$

Once differences between efficiencies of reference gene and target gene amplifications were approximately equal, we first looked at the normalizations of the myohemerythrin transcript amplifications compared to the endogenous control amplification for each trophosome sample in order to get the normalized number of cycles (NNC):

NNC = Ct target – Ct 18S, where Ct is the Threshold Cycle i. e. the number of cycles required to reach a same quantity of amplified cDNA during the exponential phase.

Then, for relative quantification measurement of myohemerythrin transcript in the trophosome cDNA samples between the two vent sites, we used the  $2^{-\Delta\Delta Ct}$  method [35]. Relative quantification results were obtained by comparing levels of expression with the calibrator tissue, the latter being chosen as the tissue for which the better expression was observed with the following calculation:

$$\text{Relative expression level} = 2^{-(NNC_{\text{sample}} - NNC_{\text{calibrator}})}$$

### *Sequencing*

Plasmid DNA from individual colonies were purified with a FlexiPrep kit (Amersham) and used in a dye-primer cycle sequencing reaction with universal primer T3 or T7 and the Big Dye® Terminator V3.1 Cycle Sequencing kit (Applied Biosystems). Reactions were subsequently run on a 16-capillary 3130 Applied Biosystems sequencer.

### ***Homologies search, alignment, and phylogenetic analyses***

BLAST analyses (blastx and blastn) of the cDNA libraries sequences were conducted on the NCBI server (<http://www.ncbi.nlm.nih.gov/BLAST/>).

The sequences were first automatically aligned with ClustalW [36] in MEGA 3.1 [37] and the alignment was then adjusted manually. The Neighbor-Joining tree was constructed under the Dayhoff matrix model (PAM matrix) and the Maximum Parsimony tree was constructed with the close-neighbor-interchange search option. For each method, bootstrap analyses were conducted for 1,000 re-sampling replicates.

### ***Cloning of the complete RpMHR cDNA into a pGEX expression vector***

We used a Glutathione S-transferase (GST) fusion vector for the expression of RpMHR in *Escherichia coli* to allow an easy purification. Total myohemerythrin coding cDNA was amplified by specific PCR primers (Table 2) containing BamHI restriction site (in the forward primer) and EcoRI restriction site (in the reverse primer). After cloning the PCR product with the TOPO TA Cloning® kit (Invitrogen), and extracting the plasmid with the Flexiprep kit (Amersham Biosciences), the sequence was checked by sequencing and digested by BamHI and EcoRI enzymes. Then, the sequence was cloned into the pGEX-5X3 vector using the BamHI and EcoRI restriction sites, thereby fusing the myohemerythrin sequence to the GST gene separated by a linker region possessing a cleavage site. The sequence of positive clones checked in order to choose the correct pGEX recombinants for fusion protein expression.

### ***Expression and extraction of GST-MHR***

BL21 DE3 *E. coli* cells, containing the pGEX-5X3-MHR, were grown overnight at 37°C in 4 ml Luria-Bertani medium supplemented with 100µg/ml ampicillin. The ZYP-50-52 auto-induction medium [38], used for the expression of recombinant GST-MHR, contained 1 X ZY (10 g tryptone, 5 g yeast extract, 925 ml water); 20X NPS (1 M Na<sub>2</sub>HPO<sub>4</sub>, 1 M KH<sub>2</sub>PO<sub>4</sub>, 500 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>); 50 X 5052 (25% glycerol, 2.5% glucose, 10% α-lactose monohydrate) and 2 mM MgSO<sub>4</sub>. This culture was supplemented with ampicillin (100 µg.ml<sup>-1</sup>). The cells were grown 3-4 days in 2 liter-flasks containing 500 ml ZYP-50-52 at 32°C until



the optical density at 600 nm reached 10. Expression of recombinant GST-MHr was induced by the  $\alpha$ -lactose present in media after all other carbohydrates were consumed (2-3 days growth). The bacterial cells were harvested by centrifugation. The cell pellet was resuspended in lysis buffer (sucrose 20 %; glycerol 10 %; Tris-HCL (pH 8.0) 50 mM;  $\text{Na}_2\text{S}_2\text{O}_5$  0.2 mM;  $\text{MgCl}_2$  2 mM; anti-protease cocktail and DNase I), lysed by French press (Constant Cell Disruption Systems) at 1.6 Kbar. The cell debris were removed by centrifugation at 4°C for 40 min at 10 000 rpm and the supernatants were saved.

### ***Purification of the fusion protein on glutathion-Sepharose 4B beads***

The supernatant containing the recombinant GST-MHr was mixed with glutathion-Sepharose 4B beads (Amersham Biosciences). First, the beads were washed by 10 (bead) volumes of PBS. The overexpression lysate was then applied on the column and left to elute by gravity. The beads were then washed by 3 (bead) volumes of NaCl 200 mM to remove non-specific binding on the bead and finally, the beads were washed by 10 (bead) volumes of PBS. The pink beads were stored frozen at -20°C or eluted by the addition of eight bead volumes of elution buffer (50 mM Tris-HCl pH 8.0; reduced glutathione 20 mM). The GST fusion proteins in the eluted solution were concentrated by centrifugation on a Microcon<sup>®</sup> Centrifugal Filter Devices 10kDa (Millipore).

### ***Protein analysis***

Different aliquots were collected during expression, lysis, and purification of recombinant proteins. The samples were analysed by SDS/PAGE (Ready Gels Tris-HCl Glycine 4-20% BioRad). Gels were stained with Coomassie brilliant blue R-250.

### ***Superoxide Dismutase (SOD) activity***

Hemerythrin and hemerythrin-like proteins can have a SOD function in the cells, as shown for a non-heme iron protein with hemerythrin-like binuclear iron cluster isolated from the anaerobic sulfate-reducing bacterium *Desulfovibrio vulgaris* [16]. To measure this activity, Oberley and Spitz 's procedure [39] was followed. A mix for 20 assays comprised 15.7 ml of DETAPAC (Sigma) buffer 2mM in potassium phosphate buffer 50 mM pH 7.8, 100  $\mu\text{l}$  of catalase (Sigma) 0.2  $\text{U} \cdot \mu\text{l}^{-1}$ , 100  $\mu\text{l}$  of NBT 28 mM and 100  $\mu\text{l}$  of xanthine 30 mM. For each assay, 100  $\mu\text{l}$  of sample is added to 800  $\mu\text{l}$  of mix and 100  $\mu\text{l}$  of xanthine oxydase (50  $\text{U} \cdot \mu\text{l}^{-1}$ ). For each assay, the absorbance was read at 560 nm for 10 minutes. The proper

concentration of xanthine oxydase is determined so that the DO differences along the kinetics results in a  $\Delta DO$  in the blank comprised between 0.015 and 0.025  $\text{min}^{-1}$ .

### Acknowledgements

We are very grateful to the chief scientists (J. Voight and D. Jollivet) for the two cruises on which the samples were collected. All this work would not have been possible without the work of the crews of the ships (Atlantis and L'Atalante) and of the submarines (Alvin and Nautile). We wish to thank Dr Delphine Dutheil from the Ecophysiology lab for her technical assistance and her useful advice concerning the SOD assays. We thank D. Le Guen for her help in purifying *S. nudus* Hr on DEAE columns and the technical personnel from the Station Biologique de Roscoff for providing the *Sipunculus nudus* sample used to extract the hemerythrin. The research was supported by a Region Bretagne PhD fellowship to SS and NSF DEB-0072695 grant to J. Voight (The Field Museum, Chicago).

## REFERENCES

1. Jones ML: **On the Vestimentifera, new phylum: Six new species, and other taxa, from hydrothermal vents and elsewhere.** *Bulletin of the Biological Society of Washington* 1985, **6**:117-158.
2. Childress JJ, Fischer CR: **The biology of hydrothermal vent animals: physiology, biochemistry and autotrophic symbioses.** *Oceanogr Mar Biol Annu Rev* 1992, **30**:337-441.
3. Goffredi SK, Childress JJ, Lallier FH, Desaulniers NT: **How to be the perfect host: CO<sub>2</sub> and HS<sup>-</sup> accumulation and H<sup>+</sup> elimination in the hydrothermal vent tubeworm *R. pachyptila*.** *Cah Biol Mar* 1998, **39**(3-4):297-300.
4. Childress JJ, Lee RW, Sanders NK, Felbeck H, Oros DR, Toulmond A, Desbruyères D, Kennicut II MC, Brooks J: **Inorganic carbon uptake in hydrothermal vent tubeworms facilitated by high environmental pCO<sub>2</sub>.** *Nature* 1993, **362**:147-149.
5. Zal F, Lallier FH, Wall JS, Vinogradov SN, Toulmond A: **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* 1996, **271**:8869-8874.
6. Terwilliger NB: **Functional adaptations of oxygen-transport proteins.** *J Exp Biol* 1998, **201**:1085-1098.
7. Kurtz DM, JR **Molecular structure/function relationships of hemerythrins.** In *Advances in Comparative and Environmental Physiology Blood and Tissue Oxygen Carriers* (ed C P Mangum) Heidelberg: Springer Verlag 1992, **13**:151-171.
8. Hendrickson WA, Klippenstein GL, Ward KB: **Tertiary structure of myohemerythrin at low resolution.** *Proc Natl Acad Sci USA* 1975, **72**:2160-2164.
9. Takagi T, Cox JA: **Primary structure of myohemerythrin from the annelid *Nereis diversicolor*.** *FEBS Lett* 1991, **285**:25-27.
10. Klotz IM, A. KT: **Interaction between oxygen and oxygen-carrying proteins.** *Science* 1955 Sep 23;122(3169):559 1955, **122**(3169):559.
11. Demuyne S, Sautiere P, van Beeumen J, Dhainaut-Courtois N: **Homologies between hemerythrins of sipunculids and cadmium-binding metalloprotein (MP II) from a polychaete annelid, *Nereis diversicolor*.** *CR Acad Sci III* 1991, **312**:317-322.
12. Jushi JG, Sullivan B: **Isolation and preliminary characterization of hemerythrin from *Lingula unguis*.** *Comp Biochem Physiol* 1973, **44B**:857-867.
13. Manwell C, Baker CMA: ***Magellona papillicornis* haemerythrin: tissue specificity, molecular weights and oxygen equilibria.** *Comp Biochem Physiol* 1988, **89B**: 453-463.
14. Coutte L, Slomianny M-C, Malecha J, Baert J-L: **Cloning and expression analysis of a cDNA that encodes a leech hemerythrin.** *Biochim Biophys Acta* 2001, **1518**:282-286.
15. Salzet M: **Neuropeptide-Derived Antimicrobial Peptides from Invertebrates for Biomedical Applications.** *Curr Medic Chem* 2005, **12**:2663-2681.
16. Lehmann Y, Meile L, Teuber M: **Ruberrythrin from *Clostridium perfringens*: Cloning of the Gene, Purification of the Protein, and Characterization of Its Superoxide Dismutase Function.** *J Bacteriol* 1996, **178**(24):7152-7158.
17. Gattiker A, Gasteiger E, Bairoch A: **ScanProsite: a reference implementation of a PROSITE scanning tool.** *Appl Bioinf* 2002, **1**:107-108.

18. Sanchez S, Andersen AC, Hourdez S, Lallier FH: **Identification, sequencing, and localization of new carbonic anhydrase transcripts from the hydrothermal vent tubeworm *Riftia pachyptila*. *FEBS J* 2007, In revision.**
19. Sanchez S, Hourdez S, Lallier FH: **Identification of proteins involved in the functioning of *Riftia pachyptila* symbiosis by Subtractive Suppression Hybridization. *BMC Genomics* 2007, Submitted.**
20. Farmer CS, Kurtz DMJ, Liu ZJ, Wang BC, Rose J, Ai J, Sanders-Loehr J: **The crystal structures of *Phascolopsis gouldii* wild type and L98Y methemerythrins: structural and functional alterations of the O<sub>2</sub> binding pocket. *J Biol Inorg Chem* 2001, **6**:418-429.**
21. Klippenstein GL, Van Riper DE, Oosterom EA: **A Comparative Study of the Oxygen Transport Proteins of *Dendrostomum pyroides*. Isolation and characterization of hemerythrins from muscles, the vascular system and the coelome. *J Biol Chem* 1972, **247**(18):5959-5963.**
22. Oberley LW, Douglas RS: **Assay of Superoxide Dismutase Activity in Tumor Tissue. *Meth Enzym* 1984, **105**:457-464.**
23. Fisher CR, Childress JJ, Arp AJ, Brooks JM, Distel D, Favuzzi JA, Macko SA, Newton A, Powell MA, Somero GN *et al*: **Physiology, morphology, and biochemical composition of *Riftia pachyptila* at Rose Garden in 1985. *Deep-Sea Res* 1988, **35**:1745-1758.**
24. Pflugfelder B, Fisher CR, Bright M: **The color of the trophosome: elemental sulfur distribution in the endosymbionts of *Riftia pachyptila* (Vestimentifera; Siboglinidae). *Mar Biol* 2005, **146**:895-901.**
25. Vanin S, Negrisolo E, Bailly X, Bubacco L, Beltramini M, Salvato B: **Molecular Evolution and Phylogeny of Sipunculan Hemerythrins. *J Mol Evol* 2006, **62**:32-41.**
26. Vincent JB, Olivier-Lilley G, Averill BA: **Proteins Containing Oxo-Bridged Dinuclear Iron Centers: A Bioinorganic Perspective. *ChemRev* 1990, **90**:1447-1467.**
27. Garbett K, Darnall DW, Klotz IM, Williams RJP: **Spectroscopy and Structure of Hemerythrin. *Arch Biochem Biophys* 1969, **103**:419-434.**
28. Klotz IM, Klotz TA, Fiess HA: **The nature of the active site of hemerythrin. *Arch Biochem Biophys* 1957, **68**(2):284-299.**
29. Nejmeddine A, Dhainaut-Courtois N, Baert J-L, Sautikre P, Fournet B, Boulenguer P: **Purification and characterization of a cadmium-binding protein from *Nereis diversicolor* (Annelida Polychaeta). *Comp Biochem Physiol* 1998, **89C**:321-326.**
30. Dhainaut A, Raveillon B, M'Berri M, Porchet-Hennere E, Demyunck S: **Purification of an antibacterial protein in the coelomic fluid of *Nereis diversicolor* (Annelida, Polychaeta). Similitude with a cadmium-binding protein. *Comp Biochem Physiol* 1989, **C94**:555-560.**
31. Deloffre L, Salzet B, Vieau D, Andries J-C, Salzet M: **Antibacterial properties of hemerythrin of the sand worm *Nereis diversicolor*. *Neuroendocrinol Lett* 2003, **24**:40-46.**
32. Diatchenko L, Lau Y-FC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED *et al*: **Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 1996, **93**:6025-6030.**
33. Gurskaya NG, Diatchenko L, Chenchik A, Siebert PD, Khaspekov GL, Lukyanov SA, Sverdlov ED: **Equalizing cDNA Subtraction Based on Selective Suppression of Polymerase Chain Reaction: Cloning of Jurkat Cell Transcripts Induced by**

- Phytohemagglutinin and Phorbol 12-Myristate 13-Acetate.** *Analytical Biochemistry* 1996, **240**:90-97.
34. Halanych KM, Lutz RA, Vrijenhoek RC: **Evolutionary origins and age of vestimentiferan tube-worms.** *Cahiers de Biologie Marine* 1998, **39**:355-358.
  35. Livak KJ, Schmittgen TD: **Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method.** *Methods* 2001, **25**:402–408.
  36. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
  37. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinf* 2004, **5**:150-163.
  38. Studier FW: **Protein Production by Auto-Induction in High-Density Shaking Cultures.** *Protein Expr Purif* 2005, **41**:207-234.
  39. Oberley LW, Spitz DR: **Assay of Superoxide Dismutase Activity in Tumor Tissue.** *Meth Enzymol* 1984, **105**:457-464.

# **CHAPITRE VII**

## **Discussion, Conclusion et Perspectives**

## CHAPITRE VII : Discussion, Conclusion et Perspectives

### 1. L'approche méthodologique générale

*La liste complète des séquences obtenues au cours de ce travail de thèse est disponible Chapitre IV. Chaque séquence est classée dans sa banque d'origine (Tableaux 2, 3, 4 et 5) avec son numéro d'accès dans la base de données GenBank (serveur NCBI).*

#### 1.1 Efficacité de la méthode SSH dans la comparaison de tissus différents

L'application de la méthode SSH dans le but de détecter des gènes différentiellement exprimés dans les tissus de *Riftia pachyptila* s'est avérée particulièrement efficace. La vérification du niveau d'expression de plusieurs gènes dans les trois tissus étudiés (branchie, paroi du corps et trophosome) a confirmé l'expression différentielle suggérée par les banques. Lorsqu'un transcrit a été étudié par au moins deux des méthodes utilisées au cours de ce travail (SSH, PCR classique, PCR quantitative et FISH), les résultats obtenus ont été largement congruents (Tableau 1). Comme on peut le constater, les séquences testées par PCR quantitative présentent un taux d'expression qui peut varier de 4 à  $2,5 \cdot 10^3$  fois supérieur dans le tissu dans lequel elles sont le plus exprimées, selon les séquences considérées. Des modèles théoriques et des études expérimentales ont démontré que pour un enrichissement efficace en séquences différentiellement exprimées, il faudrait idéalement que (i) la concentration du transcrit se situe au-dessus de 0,01 % de la totalité des transcrits présents dans l'échantillon sondeur (invérifiable dans notre cas étant donné qu'en l'absence de génome complet de *R. pachyptila* on ne connaît pas *a priori* le nombre total de transcrits possibles) et (ii) le rapport entre la concentration du transcrit dans l'échantillon sondeur et celle dans l'échantillon meneur soit au moins égal à 5 (Ji *et al.*, 2002). C'est effectivement le cas lorsque l'on regarde les taux d'expression des transcrits analysés par PCR quantitative au cours de notre étude. Les mêmes auteurs préconisent en effet l'utilisation de la méthode SSH dans le cas de comparaison inter-tissus. Nos résultats sont cohérents avec cette analyse.

**Tableau 1 : Récapitulatif des différences d'expression constatées après l'étude de plusieurs transcrits au cours de la thèse.** av. SSH = avant le protocole SSH ; ap. SSH = après le protocole SSH ; + = présence ; - = absence de détection ; nt = non testé. La couleur rouge symbolise la présence préférentielle d'un transcrit dans le tissu branchial, le jaune dans la paroi du corps, le vert dans le trophosome.

Banque	Transcrit étudié	PCR classique										PCR quantitative				FISH						
		BR		BW				BR		BW		TR		BR		TR						
		av. SSH		ap. SSH		av. SSH		ap. SSH		BR		BW		TR		BR		TR				
		18	23	28	33	18	23	28	33	18	23	28	33	18	23	28	33	18	23	28	33	
BR-BW	cytochrome c oxydase I	+	+	+	+	+	+	+	+	-	-	-	-	+	+	+	+	+	+	+	+	
	ATP synthase (F0 c)	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	
	ATP synthase (F1 b)	-	+	+	+	+	+	+	+	-	-	-	-	+	+	+	+	+	+	+	+	
	Major Vault Protein	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	Précurseur chitinase	-	-	+	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	
	Cathepsine	nt										+	+	+	+	+	+	+	+	+	+	+
BR-BW	Glycogen synthase	-	+	+	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	
	Sérine protéase	-	-	+	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	
	RpCAbr	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	Protéine liée à la chitine	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	
	Galaxine	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	
	Linker d'Hb extracellulaire	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	
BW-BR	Globine intracellulaire	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	
	Glutathione S-transferase	-	+	+	+	-	+	+	+	-	-	-	-	+	+	+	+	+	+	+	+	
	Troponine C	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	
		-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	
		nt										+	+	+	+	+	+	+	+	+	+	+
		nt										+	+	+	+	+	+	+	+	+	+	+
TR-BW	RpCAtr	-	-	+	+	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	
	Myohémérythrine	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
	récepteur T cellulaire	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
	Non identifié (TRbwC27)	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
	16S	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	
		nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	nt	



Néanmoins, ces résultats doivent être nuancés. En effet, si nous avons choisi de considérer un seul individu pour la construction de nos banques afin de rendre compte d'un transcriptome unique, il n'en reste pas moins que le choix de l'individu est aléatoire et les gènes exprimés dans ses tissus sont propres à l'individu. De ce point de vue, les études de PCR quantitative, qui ont été menées sur plusieurs individus, ont permis de renforcer les résultats de différences d'expression suggérées par les banques. Toutefois, nos calculs ont révélé que le taux d'expression d'un gène cible entre deux tissus différents peut fortement varier d'un individu à l'autre (sur les 4 individus testés). Des variations individuelles au niveau physiologique pourraient se traduire par des transcriptomes légèrement différents, et ainsi expliquer ce résultat. Notons que le choix de l'échantillonnage réalisé en mission a reposé sur l'aspect clair du trophosome (caractéristique de bactéries plus riche en vésicules contenant des sulfures par rapport à un trophosome plus sombre (Pflugfelder *et al.* 2005)) garantissant *a priori* un état physiologique comparable entre les individus considérés. Ceci dit, il s'agit d'un critère visuel plutôt subjectif, et de plus, entre le prélèvement des animaux sur le site et leur dissection à bord, de légères variations de métabolisme ne sont pas à exclure (décompression, absence de source d'H<sub>2</sub>S, exposition brutale à l'O<sub>2</sub> de l'eau de fond). Enfin, le nombre d'individus analysés pourrait être augmenté pour affiner ces premiers résultats.

### 1.2 Efficacité de la méthode SSH dans notre problématique

#### *1.2.1 Comparaison avec des banques d'EST classiques*

Une vérification rapide de l'utilité de l'approche soustractive pour identifier par exemple le transcrit RpCAbr consisterait à regarder la fréquence de ce transcrit dans une banque classique (non soustractive) de branchie de *R. pachyptila*. Nous ne sommes pas encore en mesure de réaliser cette comparaison. Chez le Polychète Alvinellidae *Alvinella pompejana*, des banques EST de tissu branchial, paroi du corps et corps complet sont sur le point d'être rendues publiques dans le cadre du consortium *Alvinella* (coordinateur biologique : F. Zal). Pour information, le nombre de transcrits d'anhydrase carbonique obtenus dans la banque non soustractive d'ADNc de branchie s'élève à 44 (contre 11 dans notre banque soustractive BR-BW de *R. pachyptila*) sur 10 355 EST séquencées (contre 202 dans notre banque soustractive BR-BW de *R. pachyptila*). Soit une proportion de 0.4 % de séquences d'anhydrase carbonique dans le transcriptome de branchie d'*A. pompejana* contre 5.4 % dans celle de *R. pachyptila*.

Toutefois, ces résultats sont difficiles à interpréter du point de vue biologique, d'autant plus que le mode de vie des deux vers est extrêmement différent.

### 1.2.2 Les études dites « globales »

Trouver des séquences exprimées en réponse à des interactions hôte-symbionte est l'une des préoccupations majeures des scientifiques travaillant sur les modèles symbiotiques. Comme énoncé à la fin du Chapitre II, les études moléculaires à haut-débit sur les interactions hôte-symbionte chez les légumineuses sont particulièrement bien avancées dans ce domaine, et par exemple chez *Medicago truncatula*, de très nombreuses banques EST ont pu être réalisées ces dernières années à partir d'ADNc de racine (sans symbionte) de la plante, de nodule (contenant des bactéries symbiotiques) ou de mycorrhizes (contenant des champignons symbiotiques), et cela sous différentes conditions physiologiques (sans apport d'azote, etc...) et à différents stades d'acquisition des symbiontes. De nombreux gènes potentiellement impliqués dans le « programme symbiotique » ont été identifiés sur la base de la différence de distribution des séquences dans les différentes banques EST. Ainsi, certains gènes codant des protéines impliquées dans la défense cellulaire, ou encore un gène codant un transporteur de phosphates exprimé par l'hôte, apparaissent plus exprimés dans la banque de mycorrhizes que dans celle de racine (Journet *et al.*, 2002). Dans notre cas, la principale contrainte imposée par l'organisme *R. pachyptila* est que le tissu dans lequel sont localisés les symbiontes n'est pas le lieu d'acquisition des métabolites qui sont nécessaires à ces derniers. Au cours de notre travail, de nouvelles séquences ont pu être identifiées comme différentiellement exprimées, et pour certaines d'entre elles l'hypothèse d'un lien avec la prise en charge des métabolites nécessaires au fonctionnement de la symbiose est peut-être une piste (nous détaillerons le cas de la MVP un peu plus loin).

D'autres études globales se focalisent sur les symbiontes. De ce point de vue, l'étude des endosymbioses entre les insectes aphides et les  $\gamma$ -protéobactéries du genre *Buchnera* ont permis de grandes avancées. Le séquençage du génome complet de *Buchnera aphidicola* est ainsi apparu 80 % plus réduit que celui de bactéries libres apparentées telles qu'*Escherichia coli*. Cette perte de gènes est cohérente avec l'étroite intégration des fonctions métaboliques de l'hôte et du symbionte (Wernegreen, 2002). En effet, le symbionte fournit à son hôte des acides aminés essentiels que celui-ci ne possède pas, et en retour, le symbionte acquiert la plupart de ce dont il a besoin de la part de son hôte (comme les sucres) ce qui lui permet de survivre même avec un si petit génome. Au contraire, malgré sa taille réduite (50 % de la

taille du génome de la bactérie apparentée *Thiomicrospira crunogena*), le génome de la bactérie sulfo-oxydante *Candidatus Ruthia magnifica* qui vit en symbiose avec *Calymene magnifica*, possède une grande partie des gènes codant pour des protéines impliquées dans le métabolisme chimiosynthétique (biosynthèse des vitamines, cofacteurs et les 20 acides aminés nécessaires au clam) (Newton *et al.*, 2007).

### 1.2.3 Etude sur *Anthopleura elegantissima*

L'analyse comparée des transcriptomes (sur l'organisme entier) du cnidaire *Anthopleura elegantissima* avec et sans symbionte révèle que plutôt que des gènes « spécifiques » de la symbiose, ce sont des gènes régulant un éventail très large de processus biologiques qui ont montré une expression différentielle (28 gènes), parmi lesquels des gènes codant des protéines impliquées dans le métabolisme des lipides, de la prolifération cellulaire, l'apoptose et le stress oxydatif (Rodriguez-Lanetty *et al.*, 2006). Nous n'avons pas trouvé de telles séquences dans nos banques de branchie et trophosome. Par ailleurs, le transcrite d'anhydrase carbonique d'*A. elegantissima* ne figure pas parmi ces 28 gènes. Néanmoins, les auteurs préconisent à présent de vérifier l'expression différentielle de ces gènes uniquement au niveau du tissu gastrodermique (qui héberge les symbiontes) avant et après infection. En effet, les gènes identifiés comme différentiellement exprimés montrent des taux d'expression très faibles (entre 1,2 et 4,2 fois) et il est possible que les plus forts taux aient été masqués par l'expression de ces mêmes gènes dans les tissus non symbiotiques. C'est peut-être la raison pour laquelle le transcrite d'anhydrase carbonique n'est pas identifié dans leur étude.

### 1.3 Problème général lié à l'identification des séquences

Si les transcrits codant l'anhydrase carbonique dans la branchie ont été obtenus en proportion non négligeable dans la banque BR-BW, nous n'avons curieusement pas obtenu dans cette banque de transcrits codant pour des pompes à protons ATPase de type vacuolaire ( $H^+V$ -ATPase) que nous nous attendions à découvrir étant données la grande quantité de protons engendrés par l'activité de l'anhydrase carbonique dans la branchie, la nécessité physiologique pour *Riftia* de maintenir un pH interne alcalin (Goffredi *et al.*, 1997a ; Girguis et Childress, 2006), et l'activité démontrée de ces transporteurs (Goffredi et Childress, 2001). De plus, ce type de pompe à protons a pu être mis en évidence au niveau des cellules épithéliales branchiales de *R. pachyptila* par immunolocalisation (De Cian *et al.*, 2003b). Un

anticorps dirigé contre la sous-unité  $\beta$  de la  $H^+V$ -ATPase de la levure a fonctionné pour la localisation de la protéine chez *R. pachyptila* suggérant que cette protéine est très bien conservée, ce qui n'aurait alors pas posé de difficulté d'annotation s'il on avait obtenu un tel transcrit dans nos banques, après confrontation avec les banques de données. L'efficacité de PCR (au stade de la construction des banques) est peut-être en cause. L'amplification du pool d'ADNc au cours du protocole SSH (Chapitre III) s'est faite aux mêmes paramètres de PCR. Même si des amorces identiques ont été utilisées pour amplifier l'ensemble des séquences différentiellement exprimées, on ne peut exclure la sous-représentation de certaines séquences à l'issue de la PCR.

La SSH apporte de nombreux avantages mais le problème majeur reste la taille des séquences obtenues. Celle-ci est complètement inhérente à la méthode à cause de la digestion des ADNc par l'enzyme de restriction pour lier les adaptateurs dès les premières étapes du protocole. Dépendante de la fréquence de rencontre du site de coupure de l'enzyme *RsaI*, la taille des différents ADNc digérés peut alors varier, et les plus petits fragments auront alors peu de chance de pouvoir être identifiés de façon certaine. Le taux d'ADNc non identifiés s'échelonne en effet de 29,4 à 54,1 % selon les banques réalisées au cours de cette étude (Chapitre IV) et est maximal pour la banque TR-BW. Dans cette banque en effet, plus de la moitié des contigs n'a pu être annotée faute de résultat significatif d'homologie. Ces séquences ne correspondent pas au génome des symbiontes non plus, comme l'a montré leur Blast sur la banque de données, aimablement rendue accessible par Horst Felbeck et ses collaborateurs, avant sa sortie publique. Il faut savoir que la limite générale à cette approche globale est de toute évidence la teneur des bases de données en séquences de taxons proches de notre organisme d'étude (nombre très réduit dans notre cas, nous y reviendrons plus loin). De plus, une majorité des séquences qui ont pu être identifiées par BlastX possèdent une courte partie codante se terminant par un codon stop suivie d'une longue région non traduite (partie 3' de l'ADNc). On peut dès lors supposer qu'une majorité des séquences non identifiées qui résultent de notre étude sont constituées de parties 3' non traduites (3'UTR) qui, sans pression de sélection, évoluent plus rapidement et sont en général très spécifiques de chaque taxon.

## 2. Les anhydrases carboniques de *Riftia pachyptila*

Chez les cnidaires vivant en symbiose avec des zooxanthelles, l'anhydrase carbonique permet d'augmenter la quantité de CO<sub>2</sub> (par la conversion des ions HCO<sub>3</sub><sup>-</sup> en CO<sub>2</sub>) disponible pour les symbiontes, au plus près du site de la photosynthèse. Aussi, il a été suggéré qu'à l'image des anhydrases carboniques des chlorophytes *Chlamydomonas reinhardtii* et *Chlorella vulgaris*, l'activité des enzymes des cnidaires symbiotiques pourrait avoir un rôle de concentration du CO<sub>2</sub> (Weis *et al.*, 1989). Chez *Riftia pachyptila*, un système encore plus complexe consiste en (i) la transformation immédiate du CO<sub>2</sub>, une fois dans les cellules épithéliales branchiales, en HCO<sub>3</sub><sup>-</sup> entraînant un flux entrant continu de CO<sub>2</sub> vers l'intérieur des cellules et (ii) la reconversion des ions HCO<sub>3</sub><sup>-</sup> en CO<sub>2</sub> au plus près des bactériocytes, répondant ainsi à l'exigence métabolique de la symbiose. Notre travail met pour la première fois en évidence chez un organisme protostomien l'existence de deux transcrits d'anhydrase carbonique tissu-spécifiques à l'état adulte.

### 2.1 Transcrits du panache branchial

Comme le transcrit précédemment décrit (RpCAtr), le nouvel ARNm d'anhydrase carbonique mis en évidence par notre étude (RpCAbr) est localisé au niveau du cytoplasme des cellules épithéliales branchiales. Les expériences d'immunolocalisation précédemment réalisées (De Cian *et al.*, 2003b) avaient sans aucun doute permis de détecter les deux protéines simultanément, l'anticorps étant dirigé contre l'anhydrase carbonique cytosolique de poulet. La localisation des protéines et des ARNm matures au niveau du cytoplasme des cellules épithéliales branchiales suggère qu'après traduction des ARNm, les protéines ne sont pas exportées hors des cellules. D'ailleurs, aucun peptide signal d'adressage n'a pu être détecté dans la partie codante en 5' de l'un ou l'autre de ces deux transcrits. Les deux gènes coderaient en fait des protéines très semblables, dont le champ d'action semblerait réduit au cytosol des cellules. Le transcrit RpCAbr est quasiment absent de l'autre tissu à forte activité d'anhydrase carbonique, le trophosome. La présence des deux types de transcrits dans le tissu branchial serait-elle donc une stratégie pour augmenter la production d'anhydrases carboniques dans ce tissu ? Signalons au passage qu'une séquence partielle très proche de RpCAbr (appelée RpCAbr2, Chapitre V) a également été mise en évidence à partir d'échantillons d'ADNc de branchie, suggérant que le nombre d'isoformes au sein de ce tissu est très certainement sous-estimé. En plus d'augmenter la production de protéines, l'existence

de plusieurs transcrits codant le même type de protéine pourrait être expliquée par des propriétés enzymatiques différentes entre les deux protéines (toutefois peu probables vue la forte identité observée entre la séquence RpCAbr et la séquence partielle RpCAbr2) ou bien par une localisation différente dans les cellules épithéliales branchiales qui montrent en effet une différence de polarité entre le niveau apical et le niveau basal.

### 2.2 Transcrits du trophosome

Pourtant, c'est au sein du trophosome, que nous nous attendions à découvrir un autre type d'anhydrase carbonique, une isoforme membranaire. La nécessité de reconvertir les ions  $\text{HCO}_3^-$  en  $\text{CO}_2$  dans ce tissu pourrait faire intervenir en plus de la protéine cytosolique déjà mise en évidence par immunolocalisation et western-blots (De Cian *et al.*, 2003b), une protéine liée à la membrane des bactériocytes. L'existence d'une anhydrase carbonique membranaire avait notamment été suggérée à la suite d'expériences sur des bactériocytes isolés soumis à l'action de différents inhibiteurs d'anhydrase carbonique (De Cian *et al.*, 2003c). Malgré cela, un seul type d'ADNc d'anhydrase carbonique a été obtenu après séquençage de notre banque spécifique du trophosome, dont la séquence est identique à celle déjà connue, RpCAtr (De Cian *et al.*, 2003a). Nos résultats de PCR quantitative et d'hybridation *in situ* ont cependant apporté des éclairages supplémentaires : ce transcrit est présent dans le trophosome en plus forte proportion par rapport à la branchie.

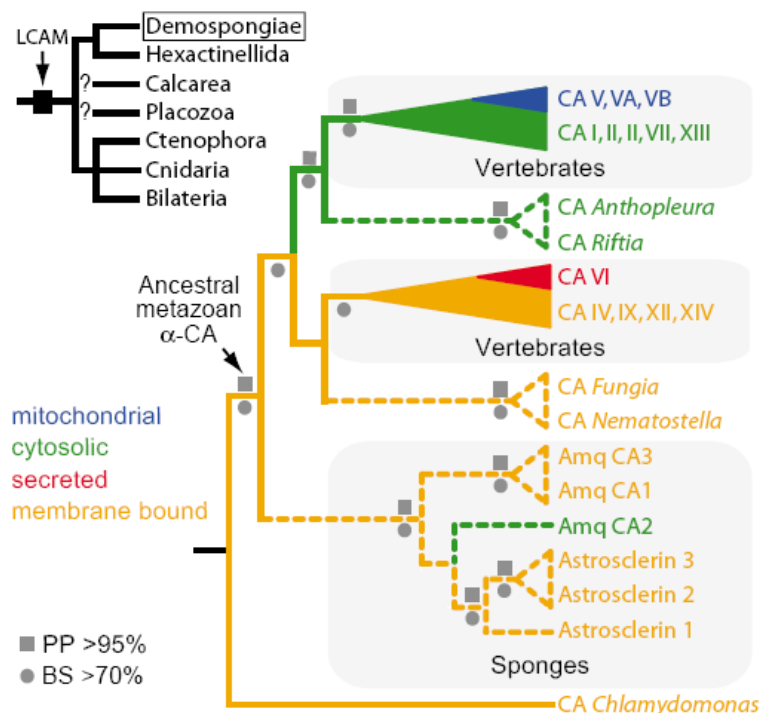
L'absence de preuve ne nous permet pas d'écarter l'existence d'anhydrase carbonique membranaire dans ce tissu. Un moyen d'essayer d'obtenir la séquence d'un tel transcrit serait de dessiner un couple d'amorces dégénérées du côté 5' de la séquence codante. En effet, c'est précisément au niveau de cette région que les séquences d'anhydrases carboniques membranaires partagent quelques acides aminés en commun. Si l'on arrive à amplifier un petit morceau de cette séquence, on pourra alors essayer de le séquencer dans le but de dessiner une amorce plus spécifique qui pourra alors être utilisée pour le protocole RACE-PCR en 3'.

### 2.3 Evolution de la famille multigénique des anhydrases carboniques

D'un point de vue purement évolutif, l'obtention de séquences d'anhydrases carboniques d'organismes « non vertébrés » est un réel besoin pour affiner la compréhension de l'évolution de cette famille multigénique. Nos résultats d'analyse phylogénétique (Chapitre

V, section Results), loin d'être robustes, se sont voulus les plus exhaustifs possible (44 séquences analysées) : les mesures statistiques aux nœuds de l'arbre sont par conséquent faibles. Toutefois, il apparaît que les séquences de *Drosophila melanogaster* sont clairement polyphylétiques, et plus généralement l'ensemble des séquences de protostomiens ne se regroupent pas préférentiellement dans un clade unique. Des données indépendantes montrent également que le cnidaire *Fungia scutaria* possède une isoforme cytosolique dont la séquence est proche de celle de *Riftia pachyptila* et *Anthopleura elegantissima* en plus de deux autres séquences proches des séquences membranaires de vertébrés. D'une part, ces tendances montrent clairement l'existence de duplications de gènes chez les protostomiens au même titre que celles connues depuis de nombreuses années chez les vertébrés, et d'autre part, la possibilité que les gènes d'anhydrase carbonique des eumétazoaires soient issus d'une duplication à partir d'une séquence ancestrale de type cytosolique ou membranaire. Très récemment publiée, l'étude effectuée par Jackson *et al.* (2007) vient préciser cette hypothèse, en révélant l'existence de trois transcrits d'anhydrase carbonique chez l'éponge corallienne *Astrosclera willeyana*. Ces trois transcrits présentent une très forte homologie avec les  $\alpha$ -anhydrases carboniques des métazoaires. La présence de peptides signaux indique la forte probabilité qu'ils codent des protéines extracellulaires ou membranaires. Par recherche bioinformatique, les auteurs ont trouvé des gènes homologues chez une autre démosponge (*Amphimedon queenslandica*) et l'ensemble des séquences de démosponges émerge en tant que groupe-frère des autres  $\alpha$ -CA de métazoaires (Figure 1).

**Figure 1** : Phylogénie simplifiée résultant de l'analyse de séquences d'anhydrases carboniques de type  $\alpha$  par les méthodes bayésienne et de maximum de vraisemblance (d'après Jackson *et al.*, 2007). Les séquences des éponges *A. willeyana* (Astrosclerin 1, 2 et 3) et *A. queenslandica* (Amq CA1, 2 et 3) forment un groupe monophylétique à la base des séquences d'eumétazoaires. Les localisations cellulaires sont mentionnées par un code couleur et sont distinguées les localisations expérimentalement prouvées (trait plein) et prédites (pointillés). LCMA = Last Common Ancestor of the Metazoa; PP = probabilité postérieure en méthode bayésienne; BS = bootstrap en maximum de vraisemblance.



Les auteurs suggèrent par ailleurs que ce type de peptide signal devait être présent dans la séquence ancestrale chez le dernier ancêtre commun aux métazoaires (LCAM), indiquant que la protéine ancestrale devait être sécrétée ou bien membranaire, ce qui apparaît compatible avec le rôle de telles enzymes chez les éponges, à savoir la biocalcification.

### **3. Les protéines respiratoires**

Les Annélides forment un taxon au sein duquel une grande diversité de pigments respiratoires est répertoriée (Tableau 2). Les hémoglobines extracellulaires de type HBL sont sans doute les plus étudiées et les travaux les plus récents sur la phylogénie des globines montrent bien leur particularité par rapport aux hémoglobines intracellulaires (Bailly *et al.* 2007) ainsi que l'évolution des chaînes de structures (Chabasse *et al.*, 2006) qui permettent l'édification de ces complexes macro-moléculaires. Comme énoncé au cours du Chapitre II, *Riftia pachyptila* possède deux hémoglobines extracellulaires vasculaires et une hémoglobine extracellulaire coelomique. Six séquences partielles de chaînes de globines avaient précédemment été obtenues chez *R. pachyptila* (une séquence de type A1, une séquence de type A2, trois séquences de type B1 et une séquence de type B2) (Bailly *et al.*, 2002). Au cours de cette thèse, nous avons pu retrouver trois de ces séquences (les séquences A1, B2 et une des séquences B1) dans notre banque soustractive du trophosome (banque TR-BW). De plus, de nouvelles séquences de chaînes de globines (probablement des chaînes de type A2 et B1) et de linkers (deux nouvelles séquences, une dans la banque BW-BR et une dans la banque TR-BW, Chapitre IV section « Results ») ont pu être mises en évidence. Le tissu hématopoïétique, qui sécrète ces protéines circulantes, est anatomiquement proche du trophosome et a donc du très certainement être disséqué en même temps que celui-ci, ce qui expliquerait la forte proportion de transcrits codant les chaînes d'hémoglobine dans la banque TR-BW (Chapitre IV, section « Results »).



**Tableau 2 : Distribution des transporteurs d'oxygène chez les animaux (adapté de Terwilliger, 1998).** La contribution de notre étude dans l'obtention de nouveaux transcrits codant des protéines respiratoires chez *R. pachyptila* est symbolisée par l'encadré.

	Myoglobine	Hémoglobine intracellulaire	Hémoglobine extracellulaire	Hémocyanine type mollusque	Hémocyanine type arthropode	Hémérythrine	Myohémérythrine
Protista	x						
Platyhelminthes	x						
Nematoda	x		x				
Phoronida		x					
Nemertea	x	x	x				
Priapulida						x	
Brachiopoda						x	
Sipunculida						x	x
Annelida	x	x	x			x	x
<b>Siboglinidae</b>	<b>x</b>		<b>x</b>				<b>x</b>
Echiura	x	x					
Mollusca	x	x	x	x			
Arthropoda			x		x		
Echinodermata		x					
Chordata	x	x					

En plus de ces nouvelles séquences, nos résultats montrent pour la première fois l'existence d'un transcrit de globine intracellulaire (ou de globine tissulaire) obtenu à partir de la banque BW-BR et d'un transcrit de myohémérythrine obtenu à partir de la banque TR-BW (Tableau 2). Si ces transcrits sont traduits en protéines fonctionnelles chez *R. pachyptila*, cela amènerait à trois le nombre de types de pigments respiratoires chez cet organisme (Tableau 2).

### 3.1 Transcrit de myoglobine

Comme leur nom l'indique, ces protéines montrent une localisation préférentielle dans les muscles. Parmi les Annélides, certains Polychètes comme le Terebellidae *Pista pacifica* utilisent plusieurs types de protéines respiratoires dans différents compartiments corporels (Terwilliger, 1974). Ainsi, cet animal tubicole comme *R. pachyptila* prélève l'O<sub>2</sub> dans l'eau de mer environnante par ses branchies bien irriguées. *P. pacifica* possède des Hb extracellulaires géantes avec une faible affinité pour l'O<sub>2</sub> dans son système vasculaire, capables de transférer l'O<sub>2</sub> à des Hb intracellulaires circulantes dans le liquide coelomique à affinité modérée, qui à leur tour peuvent transférer l'O<sub>2</sub> à des myoglobines à forte affinité pour l'O<sub>2</sub> présentes dans la paroi du corps de l'animal. Les expériences préliminaires que nous avons menées sur le transcrit de *R. pachyptila* codant la myoglobine (uniquement par PCR classique et trophosome non testé) indiquent effectivement une présence spécifique dans la paroi du corps par rapport à la branchie (Chapitre IV, section « Results », Fig. 2F). La

myoglobine pourrait servir de protéine de transfert d'O<sub>2</sub> intracellulaire et de stockage d'O<sub>2</sub> dans les muscles de la paroi du corps : la présence d'une telle protéine et sa fonction précise restent à démontrer.

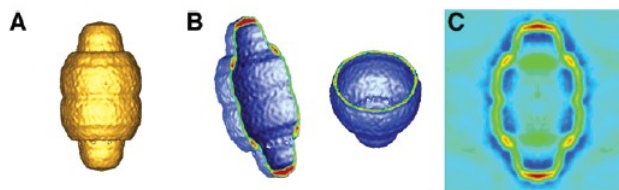
### 3.2 Transcrit de myohémérythrine

Le transcrit de myohémérythrine mis en évidence dans notre banque TR-BW apparaît spécifique du trophosome. Ce type de transcrit code classiquement une protéine monomérique non héminique dont le site de fixation à l'O<sub>2</sub> consiste en deux atomes de fer reliés aux résidus histidine du site actif. Ces protéines montrent en général une forte affinité pour l'O<sub>2</sub> (comme c'est le cas des myoglobines). Toutefois, nous ne sommes pas parvenus à démontrer la possibilité de fixation de l'O<sub>2</sub> après avoir surexprimé la protéine. Des expériences supplémentaires seront nécessaires pour vérifier s'il ne s'agirait que d'un biais dû à la fusion de la myohémérythrine à la protéine GST. Il s'agira de cliver la protéine de fusion avec le facteur Xa pour reconduire à nouveau des analyses fonctionnelles. Par contre, une activité SOD a pu être mise en évidence et correspond au piégeage de radicaux libres, ces derniers pouvant notamment être relargués par les symbiontes bactériens dans ce tissu. Egalement dans l'hypothèse d'un lien avec les symbiontes bactériens, la fonction anti-bactérienne des MHR de *H. medicinalis* et *N. diversicolor* pourrait être une orientation possible pour d'autres types d'études.

## **4. La Major Vault Protein (MVP)**

### 4.1 Description

La protéine Vault est le plus grand complexe ribonucléoprotéique connu ( $12,9 \pm 1$  MDa). Découvertes en 1986 par Nancy Kedersha and Leonard Rome, ces protéines ont été appelées Vault à cause de leur structure rappelant les voûtes des cathédrales (Figure 2).



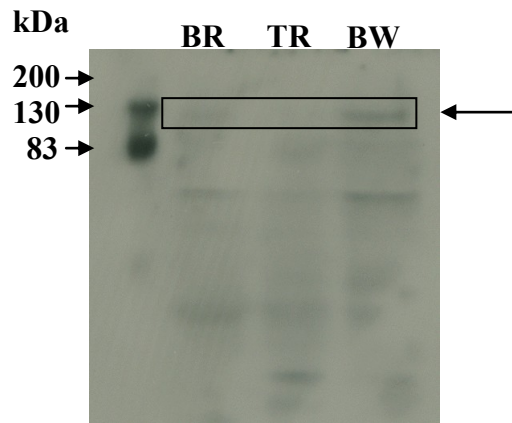
**Figure 2 :** Reconstructions tridimensionnelles par cryo-microscopie électronique de la protéine Vault de rat traitée à la RNase (d'après Kong *et al.*, 1999). (A) Vue extérieure. (B) Vues sectionnées de la Vault de rat intacte montrant la large cavité interne. Les plans coupés sont exposés avec la plus forte densité en rouge et la plus faible en vert. (C) Section de densité à partir de la Vault de rat intacte montrant la faible densité interne en vert, où les capsules terminales sont attachées à la partie en tonneau de la particule.

La Major Vault Protein (MVP, 100 kDa) est le composant majeur du complexe protéique et forme la partie en tonneau de la protéine Vault. Il a été démontré que la seule expression du gène codant la MVP du rat suffit à former une structure similaire à celle de la protéine Vault et aux mêmes caractéristiques biochimiques (Stephen *et al.*, 2001). Cette structure unique est fortement conservée chez les eucaryotes, indiquant que sa fonction, encore inconnue, doit être essentielle.

#### 4.2 Indices de présence chez *R. pachyptila*

Jusqu'à présent, la protéine MVP a été caractérisée chez relativement peu d'organismes, notamment le microeucaryote *Dictyostelium discoideum*, le mollusque *Mytilus edulis*, le poisson-chat *Ictalurus punctatus*, la raie électrique *Torpedo marmorata* ainsi que chez les mammifères (homme, rat et souris). La courte partie codante dont nous disposons montre 63 % d'identité en acides aminés avec la séquence codant la MVP d'un échinoderme (*Strongylocentrotus purpuratus*), témoignant d'une structure primaire effectivement très conservée. Les analyses par PCR classique et PCR quantitative ont révélé une expression préférentielle dans la branchie par rapport aux deux autres tissus testés (Chapitre IV, transcrit 10 fois moins abondant dans la paroi du corps et non détecté dans le trophosome par PCR quantitative). Pour aller un peu plus loin, nous avons testé la présence de la protéine à partir d'échantillons protéiques de branchie, trophosome et paroi du corps d'un même individu par western-blot (Figure 3).

**Figure 3 :** Western-blot pour tester la présence de la protéine MVP à partir de 100 µg d'extraits protéiques de branchie (BR), trophosome (TR) et paroi du corps (BW), avec l'anticorps « Bob » (don de L. Rome).

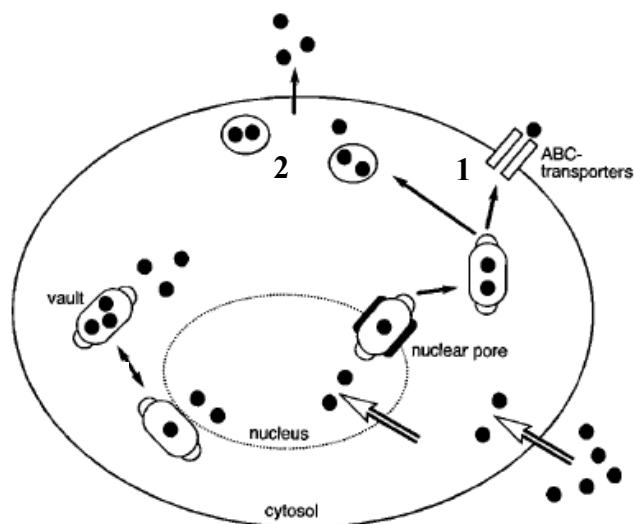


A la taille escomptée (environ 100 kDa), on peut observer une bande plus intense dans le puits contenant l'échantillon de paroi du corps que dans celui contenant l'échantillon de tissu branchial. De façon congruente avec l'incapacité à détecter le transcrit dans ce tissu, il n'y a pas de détection de la protéine dans l'échantillon de trophosome.

#### 4.3 Possibles fonctions de la MVP ?

Des études ont démontré la présence de la protéine Vault dans le cytoplasme associé au cytosquelette, mais également à la périphérie du noyau et au niveau des pores nucléaires ce qui mène à penser que la protéine Vault pourrait se charger du transport nucléocytoplasmique des ribosomes et/ou de l'ARNm (Hamill and Surprenant, 1997). L'abondance du transcrit 16S observé dans la branchie et le renouvellement supposé important de protéines dans le tissu branchial (Chapitre IV) cadreraient bien avec cette fonction chez *R. pachyptila* en permettant l'augmentation du transport des sous-unités ribosomales et des ARNm dans le cytoplasme.

Une autre hypothèse a attiré notre attention. Chez les mammifères, les Vaults sont présentes à des milliers d'exemplaires dans chaque cellule, mais ce sont les cellules tumorales, résistantes à différents traitements, qui en contiennent le plus. L'éventualité du transport nucléocytoplasmique de drogues et/ou métabolites par les protéines Vault selon le schéma présenté Figure 4 a été formulée par de nombreuses études.



**Figure 4 :** Schéma représentant le rôle hypothétique de la protéine Vault dans le transport nucléocytoplasmique de drogues et/ou métabolites (d'après Mossink *et al.*, 2003). 1 = transporteur actif ; 2 = vésicules d'exocytose. Les flèches blanches représentent l'entrée des métabolites (points noirs) par diffusion.

En séquestrant les composés toxiques, les protéines Vault, telles des gondoles, pourraient intervenir dans leur transport loin de leur cible intracellulaire vers des transporteurs actifs ou des vésicules d'exocytose (Scheffer *et al.*, 1995 ; Mossink *et al.*, 2003). Chez *Mytilus edulis*, le transcrit de la MVP est préférentiellement exprimé dans les tissus très riches en épithélium que sont les branchies et les glandes digestives (Lüdeking et Köhler, 2002) et a été interprété comme un composant du mécanisme de résistance aux xénobiotiques chez cet organisme marin. La fonction de la protéine Vault chez *R. pachyptila* pourrait également s'apparenter à un mécanisme de résistance aux métaux lourds et autres composants toxiques émanant du fluide hydrothermal (toutefois, de façon surprenante, la confrontation de notre séquence avec la banque d'EST d'*Alvinella pompejana* n'a pas montré d'homologie significative). La présence de la protéine dans la branchie et la paroi du corps de *R. pachyptila* est assez logique, les deux tissus étant à des degrés divers au contact de l'eau de mer. Dans ces tissus, la séquestration temporaire des sulfures sous forme  $H_2S$  ou  $HS^-$  serait envisageable : bien que plus exposée que la paroi du corps, la branchie possède aussi les hémoglobines HBL qui prennent en charge les ions  $HS^-$ , cela expliquerait peut-être la différence de quantité de protéine observée entre les deux tissus.

## 5. Autres pistes de protéines impliquées dans des fonctions spécifiques

### 5.1 Le Précurseur Chitinase

Des activités chitinase avaient été mesurées au niveau de l'opisthosome de *Riftia pachyptila* et de façon inattendue au niveau du panache branchial (Ravaux *et al.*, 1998), alors qu'il s'agit d'un tissu qui ne sécrète pas de chitine (Shillito *et al.*, 1993). Les auteurs exprimaient l'hypothèse que cette activité soit due à une chitinase ou à un lysozyme, hydrolase acide connue pour dégrader la chitine (Jollès et Jollès, 1984). Une activité de lysozyme avait effectivement été observée dans tous les tissus de *Riftia pachyptila* (sauf l'opisthosome, non testé) (Boetius et Felbeck, 1995). Toutefois, notre étude met en évidence l'existence d'un transcrit codant un précurseur chitinase à partir de notre banque soustractive de branchie (banque BR-BW). Sur les trois tissus que nous avons ensuite testés par PCR quantitative (branchie, paroi du corps et trophosome), la séquence mise en évidence par notre étude apparaît spécifique du tissu branchial. Les tissus de paroi du corps et de trophosome n'ont à notre connaissance pas été testés pour une possible activité chitinase. Néanmoins, le transcrit de précurseur chitinase obtenu n'est retrouvé dans aucun des 2 tissus (et sur 4 individus testés).

Une poursuite du travail pourrait être envisagée pour tester la présence éventuelle de ce transcrit dans l'opisthosome (que nous n'avons pu tester en l'absence d'échantillon). L'utilisation de nos amorces dessinées en partie codante pourra être envisagée pour tester la présence de ce transcrit et mesurer son niveau d'expression relativement à la branchie. Bien entendu, l'existence de la protéine chitinase (obtenue après clivage du précurseur chitinase) reste encore à démontrer (par western blot par exemple).

### 5.2 Autres séquences

Des clichés pris en microscopie électronique montrent la présence de lysosomes au niveau de la branchie, de la paroi du corps et du trophosome chez *R. pachyptila*. Nous avons montré la présence d'une séquence codant pour la cathepsine, enzyme lysosomale, dans la branchie, le trophosome et la paroi du corps (Chapitre IV). Sa présence dans le tissu branchial cadre bien avec la présence des organites denses aux électrons (appelées EDO, Chapitre IV) qui pourraient correspondre au produit de fusion des lysosomes et de mitochondries endommagées pour dégrader ces dernières. Dans le trophosome, une telle enzyme pourrait

participer à la dégradation de composés peptidiques, ou peut-être même à la digestion des bactéries.

Enfin, sur l'ensemble de nos banques, nous n'avons pas trouvé de séquence qui puisse être identifiée avec certitude comme une métallothionéine. Pourtant, une séquence assez proche, montrant une forte homologie avec une intégrine, a été trouvée dans la banque de paroi du corps. Curieusement, pour finir, une séquence présentant des motifs de 30 acides aminés répétés en tandem, identifiée comme étant une « galaxine » (du corail *Galaxea fascicularis*) semble exprimée préférentiellement dans la paroi du corps de *R. pachyptila*. Une drôle de coïncidence que de constater que dans l'étude de la symbiose entre le calmar *Euprymna scolopes* et la bactérie *Vibrio fischeri*, une séquence à forte homologie avec la « galaxine » montre une expression différentielle au cours du cycle circadien de l'adulte et une étude comparative réalisée sur des individus au stade symbiotique et aposymbiotique montre que son expression semblerait induite par la présence des symbiontes (M. McFall-Ngai, communication ISS Congress, Vienne, 2006).

### 6. Perspectives d'études

Le travail réalisé au cours de cette thèse aura permis de cibler quelques pistes d'études intéressantes pour la suite de l'étude du Siboglinidae *Riftia pachyptila*. Nous avons pu mettre en évidence des transcrits d'intérêt par leurs taux d'expression importants dans les tissus impliqués dans le mode de vie symbiotique. Loin d'avoir donné des réponses aux nombreuses interrogations suscitées par ce modèle symbiotique exceptionnel, les banques réalisées au cours de cette thèse ont l'avantage de présenter un panel de séquences qui sont loin d'avoir été complètement exploitées. Il reste à tester l'expression différentielle de quelques autres transcrits que cette étude n'a pu traiter, tels que la séquence « oxidoreductase molybdopterin binding protein » (TRbwC20, Chapitre IV) et l'« ubiquinol cytochrome reductase » (TRbwC21, Chapitre IV) toutes deux isolées à partir du trophosome et qui pourraient avoir un lien avec le métabolisme des symbiontes.

En plus des expériences suggérées sur des points particuliers tout au long de cette discussion, une étude fonctionnelle des deux anhydrases carboniques codées par chacun des transcrits RpCAbr et RpCAtr serait grandement souhaitable. En effet, étant données les fortes différences d'expression constatées entre les différents tissus pour chacun des deux transcrits,

on peut évidemment suspecter que le transcrit RpCAbr coderait pour la protéine mise en évidence dans la branchie et le transcrit RpCAtr pour celle mise en évidence dans le trophosome. La surexpression de ces transcrits en système hétérologue (comme cela a été réalisé pour l'étude de la myohémérythrine, Chapitre VI) puis la migration des protéines purifiées sur gel SDS-PAGE à deux dimensions permettra de comparer les propriétés intrinsèques de chaque protéine (poids moléculaire, point isoélectrique...). Ces données pourront être confrontées aux propriétés observées lors des études des protéines purifiées à partir du tissu branchial et du trophosome : la synthèse *in vitro* de ces protéines par génie génétique permettra à terme d'effectuer des tests d'activité.

Pour certains transcrits, susceptibles d'avoir une fonction importante à l'image du transcrit TRbwC27 spécifiquement présent dans le trophosome (Chapitre IV), l'identification reste malheureusement indéterminée. De nombreuses séquences, si un projet d'EST sur *R. pachyptila* se réalise, pourront être comparées sur une plus grande longueur et peut-être annotées. Par ailleurs, tant que le séquençage du génome complet d'un organisme protostomien « modèle » n'aura pas été réalisé, l'annotation des gènes des espèces d'Annélides, entre autres, restera difficile. Pour l'instant, aucun génome complet de Lophotrochozoa n'est encore disponible. Il est paradoxal que de tels modèles soient intensivement utilisés pour des questions aussi diverses que le développement, l'évolution moléculaire et cellulaire, la neurobiologie, la toxicologie, l'aquaculture et la médecine sans qu'un effort supplémentaire ne soit mis du côté du séquençage haut-débit. Plusieurs projets de séquençage complet de génomes sont à l'étude (la sangsue *Helobdella robusta*, les polychètes *Platynereis dumerilii* et *Alvinella pompejana*, la moule *Mytilus edulis*, l'huître *Crassostrea gigas*...) mais aucun n'a encore abouti.

Si les symbiotes impliqués dans les symbioses chimiosynthétiques sont maintenant bien étudiés au niveau génomique (génomes complets du symbiote du Vesicomylidae *Calymene magnifica* (Newton *et al.*, 2007) et du symbiote de *R. pachyptila* (bientôt disponible, H. Felbeck, comm. pers.)), les connaissances moléculaires chez les hôtes sont très limitées. Pourtant, la comparaison entre la taille du génome de l'hôte et celle de son symbiote pourrait apporter à terme des réponses sur la complexité de l'interaction symbiotique, avec l'éventualité de la réduction de taille du génome de la bactérie par rapport à la forme libre. En attendant, la mise en œuvre originale de la technique SSH dans notre étude pourrait aussi être appliquée à d'autres modèles symbiotiques chimioautotrophes. En effet,



l'approche SSH pourrait se révéler tout à fait pertinente pour identifier des gènes exprimés au stade symbiotique chez *Bathymodiolus azoricus* mais aussi chez les espèces littorales du genre *Codakia* (Lucinidae). L'avantage offert par ces organismes est que dans ces cas précis il est possible de leur faire perdre leurs symbiontes en les privant des substrats nécessaires. La soustraction entre un tissu branchial rendu aposymbiotique et le même tissu après l'établissement de la symbiose pourra être effectuée.

L'étude transcriptomique menée au cours de cette thèse avait pour but de mettre en évidence des gènes différentiellement exprimés dans les tissus de *Riftia pachyptila* pour tenter d'inférer des hypothèses sur la présence de telle ou telle protéine dans un tissu donné, et sur sa fonction biologique. Les variations des taux d'ARN messagers sont-elles le reflet des variations du niveau des protéines ? Comme l'a laissé présager le résultat du western-blot de la MVP, un niveau de transcription 10 fois plus important dans la branchie ne conduit pas forcément à une quantité plus importante de protéine dans ce même tissu. Ainsi, selon des estimations récentes, si 73 % des protéines sont régulées au niveau transcriptionnel chez la levure, au moins 25 % des protéines suivent un autre type de régulation et ne seraient pas détectées avec une seule méthode transcriptomique (Newman *et al.*, 2006). L'étude protéomique s'avère indispensable pour déceler les gènes régulés au niveau post-traductionnel. Effectivement, une approche protéomique pourrait être envisageable pour l'étude du tissu branchial afin de déterminer les protéines exprimées dans ce tissu et les comparer par la suite avec nos résultats d'expression différentielle. Néanmoins, en ce qui concerne le trophosome, l'approche est un peu plus délicate pour discriminer les protéines spécifiques de l'hôte et du symbionte.

## **REFERENCES BIBLIOGRAPHIQUES**

## REFERENCES BIBLIOGRAPHIQUES

- Anbar, A. D., and A. H. Knoll.** 2002. Proterozoic Ocean Chemistry and Evolution: A Bioinorganic Bridge? *Science* **297**:1137-1142.
- Andersen, A. C., L. Hamraoui, and D. Zaoui.** 2001. The oburaculum of *Riftia pachyptila* (Annelida, Vestimentifera): ultrastructure and function of the obturacular muscles and extracellular matrix. *Cah. Biol. Mar.* **42**:219-237.
- Andersen, A. C., S. Jolivet, S. Claudinot, and F. H. Lallier.** 2002. Biometry of the branchial plume in the hydrothermal vent tubeworm *Riftia pachyptila* (Vestimentifera; Annelida). *Can. J. Zool.* **80**:320-332.
- Arndt, C., D. Schiedek, and H. Felbeck.** 1998. Anaerobiosis in the hydrothermal vent tube-worm *Riftia pachyptila*. *Cah. Biol. Mar.* **39**:271-273.
- Arp, A. J., and J. J. Childress.** 1981. Blood function in the hydrothermal vent vestimentiferan tube worm. *Science* **213**:342-344.
- Arp, A. J., J. J. Childress, and C. R. Fisher.** 1985. Blood gas transport in *Riftia pachyptila*. *Biol. Soc. Wash. Bul.* **6**:289-300.
- Arp, A. J., J. J. Childress, and R. D. Vetter.** 1987. The sulfide-binding protein in the blood of the vestimentiferan tube worm, *Riftia pachyptila*, is the extracellular haemoglobin. *J. Exp. Biol.* **128**:139-158.
- 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. EPHE-Université Pierre et Marie Curie, Paris.
- Bailly, X., C. Chabasse, S. Hourdez, S. Dewilde, S. Martial, L. Moens, and F. Zal.** 2007. Globin gene family evolution and functional diversification in annelids. *FEBS J.* **274**:2641-2652.
- 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.
- Bauer-Nebelsick, M., C. F. Bardele, and J. A. Ott.** 1996. Redescription of *Zoothamnium niveum* (Hemprich & Ehrenberg, 1831) Ehrenberg, 1838 (Oligohymenophora, Peritrichida), a ciliate with ectosymbiotic, chemoautotrophic bacteria. *Eur. J. Protistol.* **32**:18-30.
- Black, M. B., K. M. Halanych, P. A. Y. Maas, W. R. Hoeh, J. Hashimoto, D. Desbruyeres, R. A. Lutz, and R. C. Vrijenhock.** 1997. Molecular systematics of vestimentiferan tubeworms from hydrothermal vents and cold-water seeps. *Mar. Biol.* **130**:141-149.
- Boetius, A., and H. Felbeck.** 1995. Digestive enzymes in marine invertebrates from hydrothermal vents and other reducing environments. *Marine Biology* **122**:105-113.
- Bosch, C., and P. P. Grassé.** 1984a. Cycle partiel des bactéries chimioautotrophes symbiotiques et leurs rapports avec les bactériocytes chez *Riftia pachyptila* Jones (Pogonophore Vestimentifère). I. Le trophosome et les bactériocytes. *C. R. Acad. Sci. Paris, Ser III* **299**:371-376.
- Bosch, C., and P. P. Grassé.** 1984b. Cycle partiel des bactéries chimioautotrophes symbiotiques et leurs rapports avec les bactériocytes chez *Riftia pachyptila* Jones (Pogonophore Vestimentifère). II. L'évolution des bactéries symbiotiques et des bactériocytes. *C. R. Acad. Sci. Paris, Ser III* **299**:413-419.

- Bradford, M. M.** 1976. A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal. Biochem.* **72**:248-254.
- Brand, G. L., R. V. Horak, N. Le Bris, S. K. Goffredi, S. L. Carney, B. Govenar, and P. H. Yancey.** 2007. Hypotaurine and thiotaurine as indicators of sulfide exposure in bivalves and vestimentiferans from hydrothermal vents and cold seeps. *Mar. Ecol. An Evolutionary Perspective* **28**:208-218.
- Bright, M., and O. Giere.** 2005. Microbial symbiosis in Annelida. *Symbiosis* **38 (1)**:1-45.
- Bright, M., H. Keckeis, and C. R. Fisher.** 2000. An autoradiographic examination of carbon fixation, transfer and utilization in the *Riftia pachyptila* symbiosis. *Mar. Biol.* **136**:621-632.
- Canfield, D. E.** 2005. The early history of atmospheric oxygen: Homage to Robert M. Garrels. *Ann. Rev. Earth Planet. Sci.* **33**:1-36.
- Cary, S. C., and S. J. Giovannoni.** 1993. Transovarial inheritance of endosymbiotic bacteria in clams inhabiting deep-sea hydrothermal vents and cold seeps. *Proc. Natl. Acad. Sci. USA* **90**:5695-5699.
- Catling, D. C., and M. W. Claire.** 2005. How Earth's atmosphere evolved to an oxic state: A status report. *Earth Planet. Sci. Lett.* **237**:1-20.
- Caullery, M.** 1914. Sur les Siboglinidae, type nouveau d'invertébrés recueillis par l'expédition du Siboga. *C. R. Acad. Sci., Série III* **158**:2014-2017.
- Cavanaugh, C. M., S. L. Gardiner, M. L. Jones, H. W. Jannasch, and J. B. Waterbury.** 1981. Prokaryotic cells in the hydrothermal vent tube-worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. *Science* **213**:340-342.
- Chabasse, C., X. Bailly, S. Sanchez, M. Rousselot, and F. Zal.** 2006. Gene Structure and Molecular Phylogeny of the Linker Chains from the Giant Annelid Hexagonal Bilayer Hemoglobins. *J Mol Evol* **63**:365-374.
- Chamoy, L., M. Nicolaï, J. Ravaux, B. Quennedey, F. Gaill, and J. Delachambre.** 2001. A Novel Chitin-binding Protein from the Vestimentiferan *Riftia pachyptila* Interacts Specifically with  $\beta$ -Chitin. Cloning, expression, and characterization. *J. Biol. Chem.* **276**:8051-8058.
- Childress, J. J., A. J. Arp, and C. R. Fischer.** 1984. Metabolic and blood characteristics of the hydrothermal vent tube-worm *Riftia pachyptila*. *Mar. Biol.* **83**:109-124.
- Childress, J. J., and C. R. Fischer.** 1992. The biology of hydrothermal vent animals: physiology, biochemistry and autotrophic symbioses. *Oceanogr. Mar. Biol. Annu. Rev.* **30**:337-441.
- Childress, J. J., C. R. Fischer, J. A. Favuzzi, R. Kochevar, N. K. Sanders, and A. M. Alayse.** 1991. Sulfide-driven autotrophic balance in the bacterial symbiont-containing hydrothermal vent tubeworm, *Riftia pachyptila* Jones. *Biol. Bull.* **180**:135-153.
- Childress, J. J., R. W. Lee, N. K. Sanders, H. Felbeck, D. R. Oros, A. Toulmond, D. Desbruyères, M. C. Kennicut II, and J. Brooks.** 1993. Inorganic carbon uptake in hydrothermal vent tubeworms facilitated by high environmental  $PCO_2$ . *Nature* **362**:147-149.
- Colaco, A., F. Dehairs, and D. Desbruyeres.** 2002. Nutritional relations of deep-sea hydrothermal fields at the Mid-Atlantic Ridge: a stable isotope approach. *Deep-Sea Res. Part I-Oceanographic Research Papers* **49**:395-412.
- Corliss, J. B., J. Dymond, L. I. Gordon, J. M. Edmond, R. P. V. Herzen, R. D. Ballard, K. Green, D. Williams, A. Bainbridge, K. Crane, and T. H. van Andel.** 1979. Submarine thermal springs on the Galapagos rift. *Science* **203**:1073-1083.
- Craddock, C., W. R. Hoeh, R. G. Gustafson, R. A. Lutz, J. Hashimoto, and R. C. Vrijenhoek.** 1995. Evolutionary relationships among deep-sea mytilids (Bivalvia:

- Mytilidae) from hydrothermal vents and cold-water methane/sulfide seeps. *Mar. Biol.* **121**:477-485.
- De Bary, H. A.** 1879. Die Entstehung der Symbiose, Strassburg.
- De Cian, M. C.** 2002. Transport du CO<sub>2</sub> et échanges ioniques chez un animal autotrophe, le vestimentifère *Riftia pachyptila*. Etude de l'anhydrase carbonique et des transporteurs impliqués. Université Pierre et Marie Curie, Paris.
- De Cian, M. C., A. C. Andersen, X. Bailly, and F. H. Lallier.** 2003b. Expression and localization of carbonic anhydrase and ATPases in the symbiotic tubeworm *Riftia pachyptila*. *J. Exp. Biol.* **206**:399-409.
- De Cian, M. C., A. C. Andersen, J. Y. Toullec, I. Biegala, J. C. Caprais, B. Shillito, and F. H. Lallier.** 2003c. Isolated bacteriocyte cell suspensions from the hydrothermal-vent tubeworm *Riftia pachyptila*, a potent tool for cellular physiology in a chemoautotrophic symbiosis. *Mar. Biol.* **142**:141-151.
- De Cian, M. C., X. Bailly, J. Morales, J. M. Strub, A. Van Dorsselaer, and F. H. Lallier.** 2003a. Characterization of carbonic anhydrases from *Riftia pachyptila*, a symbiotic invertebrate from deep-sea hydrothermal vents. *Proteins* **51**:327-339.
- Delaney, J. R., D. S. Kelley, M. D. Lilley, P. A. Butterfield, J. A. Baross, W. S. D. Wilcock, R. W. Embley, and M. Summit.** 1998. The Quantum Event of Oceanic Crustal Accretion: Impacts of Diking at Mid-Ocean Ridges. *Science* **281**:222-230.
- Desbruyères, D., P. Chevaldonné, A.-M. Alayse, D. Jollivet, F. H. Lallier, C. Jouin-Toulmond, F. Zal, P.-M. Sarradin, R. Cosson, J.-C. Caprais, C. Arndt, J. O'Brien, J. Guezennec, S. Hourdez, R. Riso, F. Gaill, L. Laubier, and A. Toulmond.** 1998. Biology and ecology of the Pompei worm (*Alvinella pompejana* Desbruyères and Laubier), a normal dweller on an extreme deep-sea environment: a synthesis of current knowledge and recent developments. *Deep-Sea Res. Part II* **45**:383-422.
- Desbruyères, D., M. Segonzac, and M. Bright.** 2006. Handbook of deep-sea hydrothermal vent fauna. Linz-Dornach, Austria.
- Diatchenko, L., Y.-F. C. Lau, A. P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov, and P. D. Siebert.** 1996. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* **93**:6025-6030.
- Dufour, S. C.** 2005. Gill Anatomy and the Evolution of Symbiosis in the Bivalve Family Thyasiridae. *Biol Bull* **208**:200-212.
- Dufour, S. C., and H. Felbeck.** 2003. Sulfide mining by the superextensible foot of symbiotic thyasirid bivalves. *Nature* **426**:65-67.
- Duperron, S., C. Bergin, F. Zielinski, A. Blazejak, A. Pernthaler, Z. P. McKiness, E. DeChaine, C. M. Cavanaugh, and N. Dubilier.** 2006. A dual symbiosis shared by two mussel species, *Bathymodiolus azoricus* and *Bathymodiolus puteoserpentis* (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environ. Microbiol.* **8**:1441-1447.
- Embley, T. M., and W. Martin.** 2006. Eukaryotic evolution, changes and challenges. *Nature* **440**:623-630.
- Felbeck, H.** 1985. CO<sub>2</sub> fixation in the hydrothermal vent tube worm *Riftia pachyptila* (Jones). *Physiol. Zool.* **58**:272-281.
- Felbeck, H., and J. Jarchow.** 1998. Carbon release from purified chemoautotrophic bacterial symbionts of the hydrothermal vent tubeworm *Riftia pachyptila*. *Physiol Zool* **71**:294-302.

- Felbeck, H., and P. J. Turner.** 1995. CO<sub>2</sub> transport in catheterized hydrothermal vent tubeworms, *Riftia pachyptila* (Vestimentifera). *Journal of Experimental Zoology* **272**:95-102.
- Feldman, R. A., M. B. Black, C. S. Cary, R. A. Lutz, and R. C. Vrijenhoek.** 1997a. Molecular phylogenetics of bacterial endosymbionts and their vestimentiferan hosts. *Molecular Marine Biology and Biotechnology* **6**:268-277.
- Feldman, R. A., T. M. Shank, M. B. Black, A. Baco, C. R. Smith, and R. C. Vrijenhoek.** 1997b. Vestimentiferan on a whale fall. *Biol. Bull. mar. biol. Lab., Woods Hole* **6**:268-277.
- Fiala-Médioni, A., Z. P. McKiness, P. Dando, J. Boulegue, A. Mariotti, A. M. Alayse-Danet, J. J. Robinson, and C. M. Cavanaugh.** 2002. Ultrastructural, biochemical, and immunological characterization of two populations of the mytilid mussel *Bathymodiolus azoricus* from the Mid-Atlantic Ridge: evidence for a dual symbiosis. *Mar. Biol.* **141**:1035-1043.
- Fisher, C. R., J. J. Childress, A. J. Arp, J. M. Brooks, D. Distel, J. A. Favuzzi, S. A. Macko, A. Newton, M. A. Powell, G. N. Somero, and T. Soto.** 1988. Physiology, morphology, and biochemical composition of *Riftia pachyptila* at Rose Garden in 1985. *Deep-Sea Res.* **35**:1745-1758.
- Fisher, C. R., J. J. Childress, and N. K. Sanders.** 1988. The role of vestimentiferan hemoglobin in providing an environment suitable for chemoautotrophic sulfide-oxidizing endosymbionts. *Symbiosis* **5**:229-246.
- 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. *PNAS* **102**:2713-2718.
- Flores, J. F., and S. M. Hourdez.** 2006. The zinc-mediated sulfide-binding mechanism of hydrothermal vent tubeworm 400-kDa hemoglobin. *Cah. Biol. Mar.* **47**:371-377.
- Freytag, J. K., P. R. Girguis, D. C. Bergquist, J. P. Andras, J. J. Childress, and C. R. Fisher.** 2001. A paradox resolved: Sulfide acquisition by roots of seep tubeworms sustains net chemoautotrophy. *Proceedings of the National Academy of Sciences of the United States of America* **98**:13408-13413.
- Gaill, F., and S. Hunt.** 1986. Tube of deep-sea hydrothermal vent worms *Riftia pachyptila* (Vestimentifera) and *Alvinella pompejana* (annelida). *Mar. Ecol. Prog. Ser.* **34**:267-274.
- Gaill, F., B. Shillito, F. Menard, G. Goffinet, and J. J. Childress.** 1997. Rate and process of tube production by the deepsea hydrothermal vent tubeworm *Riftia pachyptila*. *Marine Ecology Progress Series* **148**:135-143.
- Gardiner, S. L., and M. L. Jones.** 1993. Vestimentifera, p. 371-460. *In* S. L. Gardiner (ed.), *Microscopic anatomy of invertebrates*. Volume 12: Onychophora, Chilopoda and lesser Protostomata. Wiley-Lyss, New York.
- Girguis, P. R., and J. J. Childress.** 2006. Metabolite uptake, stoichiometry and chemoautotrophic function of the hydrothermal vent tubeworm *Riftia pachyptila*: responses to environmental variations in substrate concentrations and temperature. *Journal of Experimental Biology* **209**:3516-3528.
- Girguis, P. R., R. W. Lee, N. Desaulniers, J. J. Childress, M. Pospesel, H. Felbeck, and F. Zal.** 2000. Fate of nitrate acquired by the tubeworm *Riftia pachyptila*. *Applied and Environmental Microbiology* **66**:2783-2790.
- Goffredi, S., J. Childress, N. Desaulniers, R. Lee, F. Lallier, and D. Hammond.** 1997a. Inorganic carbon acquisition by the hydrothermal vent tubeworm *Riftia pachyptila*

- depends upon high external  $PCO_2$  and upon proton-equivalent ion transport by the worm. *J Exp Biol* **200**:883-96.
- Goffredi, S. K., and J. J. Childress.** 2001. Activity and inhibitor sensitivity of ATPases in the hydrothermal vent tubeworm *Riftia pachyptila* : a comparative approach. *Marine Biology* **138**:259-265.
- Goffredi, S. K., J. J. Childress, N. T. Desaulniers, and F. H. Lallier.** 1997b. Sulfide acquisition by the vent worm *Riftia pachyptila* appears to be via uptake of  $HS^-$ , rather than  $H_2S$ . *J Exp Biol* **200**:2609-2616.
- Goffredi, S. K., J. J. Childress, F. H. Lallier, and N. T. Desaulniers.** 1999b. The ionic composition of the hydrothermal vent tube worm *Riftia pachyptila*: evidence for the elimination of  $SO_4^{2-}$  and  $H^+$  and for a  $Cl^-/HCO_3^-$  shift. *Physiol Biochem Zool* **72**:296-306.
- Goffredi, S. K., P. R. Girguis, J. J. Childress, and N. T. Desaulniers.** 1999a. Physiological Functioning of Carbonic Anhydrase in the Hydrothermal Vent Tubeworm *Riftia pachyptila*. *Biol Bull* **196**:257-264.
- Gurskaya, N. G., L. Diatchenko, A. Chenchik, P. D. Siebert, G. L. Khaspekoy, S. A. Lukyanov, and E. D. Sverdlov.** 1996. Equalizing cDNA Substraction Based on Selective Suppression of Polymerase Chain Reaction: Cloning of Jurkat Cell Transcripts Induced by Phytohemagglutinin and Phorbol 12-Myristate 13-Acetate. *Analyt. Biochem.* **240**:90-97.
- Hahlbeck, E., M. A. Pospesel, F. Zal, J. J. Childress, and H. Felbeck.** 2005. Proposed nitrate binding by hemoglobin in *Riftia pachyptila* blood. *Deep-Sea Research Part I-Oceanographic Research Papers* **52**:1885-1895.
- Halanych, K. M.** 2005. Molecular phylogeny of siboglinid annelids (a.k.a. pogonophorans): a review. *Hydrobiol.* **535**:297-307.
- Halanych, K. M.** 2004. The new view of animal phylogeny. *An. Rev. Ecol. Evol. Syst.* **35**:229-256.
- Halanych, K. M., T. G. Dahlgren, and D. McHugh.** 2002. Unsegmented annelids? Possible origins of four lophotrochozoan worm taxa. *Integ. Comp. Biol.* **42**:678-684.
- Halanych, K. M., R. A. Lutz, and R. C. Vrijenhoek.** 1998. Evolutionary origins and age of vestimentiferan tube-worms. *Cah. Biol. Mar.* **39**:355-358.
- Hamill, D. R., and K. A. Suprenant.** 1997. Characterization of the Sea Urchin Major Vault Protein - a Possible Role For Vault Ribonucleoprotein Particles in Nucleocytoplasmic Transport. *Dev. Biol.* **190**:117-128.
- Hand, S. C.** 1987. Trophosome ultrastructure and the characterization of isolated bacteriocytes from invertebrate-sulfur bacteria symbioses. *Biol. Bull* **173**:260-276.
- Hanekamp, K., U. Bohnbeck, B. Beszteri, and K. Valentin.** 2007. PhyloGena—a user-friendly system for automated phylogenetic annotation of unknown sequences. *Bioinformatics* **23**:793-801.
- Hentschel, U., and H. Felbeck.** 1993. Nitrate respiration in the hydrothermal vent tubeworm *Riftia pachyptila*. *Nature* **366**:338-340.
- Hurtado, L. A., M. Mateos, R. A. Lutz, and R. C. Vrijenhoek.** 2003. Coupling of bacterial endosymbiont and host mitochondrial genomes in the hydrothermal vent clam *Calymene magnifica*. *Appl. Environ. Microbiol.* **69**:2058-2064.
- Ivanov, A. V.** 1951. On including the genus *Siboglinum* Caullery in the class Pogonophora. *Doklady Akadami Nauk SSSR* **76**:739-742.
- Jackson, D. J., L. Macis, J. Reitner, B. M. Degnan, and G. Wörheide.** 2007. Sponge Paleogenomics Reveals an Ancient Role for Carbonic Anhydrase in Skeletogenesis. *Science* **316**:1893-1895.

- Jennings, R. M., and K. M. Halanych.** 2005. Mitochondrial Genomes of *Clymenella torquata* (Maldanidae) and *Riftia pachyptila* (Siboglinidae): Evidence for Conserved Gene Order in Annelida. *Mol. Biol. Evol.* **22**:210-222.
- Ji, W., M. B. Wright, L. Cai, A. Flament, and K. Lindpaintner.** 2002. Efficacy of SSH PCR in isolating differentially expressed genes. *BMC Genomics* **3**.
- Johansson, K. E.** 1939. *Lamellisabella zachsi* Uschakow, ein Vertreter eine neuen Tierklasse Pogonophora. *Zoologiska Bidrag från Uppsala* **18**:253-268.
- Johansson, K. E.** 1937. Über *Lamellisabella zachsi* und ihre systematische Stellung. *Zoologischer Anzeiger* **117**:23-26.
- Johnson, K. S., J. J. Childress, R. R. Hessler, C. M. Sakamoto-Arnold, and C. L. Beehler.** 1988. Chemical and biological interactions in the Rose Garden hydrothermal vent field. *Deep-Sea Res.* **35**:1723-1744.
- Jollès, P., and J. Jollès.** 1984. What's new in lysozyme research ? *Mol. Cell. Biochem.* **63**:165-189.
- Jones, M. L.** 1985. On the Vestimentifera, new phylum: Six new species, and other taxa, from hydrothermal vents and elsewhere. *Biol. Soc. Wash. Bull.* **6**:117-158.
- Jones, M. L.** 1981. *Riftia pachyptila* Jones: observations on the Vestimentiferan worm from the Galapagos rift. *Science* **213**:333-336.
- Jones, M. L.** 1988. The Vestimentifera, their biology, systematic and evolutionary patterns. *Oceanol. Acta* **8**:69-82.
- Jones, M. L., and S. L. Gardiner.** 1988. Evidence for a transient digestive tract in Vestimentifera. *Proc. Biol. Soc. Wash.* **101**:423-433.
- Journet, E. P., D. van Tuinen, J. Gouzy, H. Crespeau, V. Carreau, M. J. Farmer, A. Niebel, T. Schiex, O. Jaillon, O. Chatagnier, L. Godiard, F. Micheli, D. Kahn, V. Gianinazzi-Pearson, and P. Gamas.** 2002. Exploring root symbiotic programs in the model legume *Medicago truncatula* using EST analysis. *Nucleic Acids Res.* **30**:5579-5592.
- Kádár, E., R. Bettencourt, V. Costa, R. S. Santos, A. Lobo-da-Cunha, and P. Dando.** 2005. Experimentally induced endosymbiont loss and re-acquirement in the hydrothermal vent bivalve *Bathymodiolus azoricus*. *J. Exp. Mar. Biol. Ecol.* **318**:99-110.
- Kandpal, M., and B. L. Tekwani.** 1997. Polyamine transport systems of *Leishmania donovani* promastigotes. *Life Sci.* **60**:1793-1801.
- Karl, D. M., C. O. Wirsen, and H. W. Jannasch.** 1980. Deep-sea primary productivity at the Galapagos hydrothermal vents. *Science* **207**:1345-1347.
- Kedersha, N. L., D. F. Hill, K. E. Kronquist, and L. H. Rome.** 1986. Subpopulations of Liver Coated Vesicles Resolved by Preparative Agarose Gel Electrophoresis. *J. Cell Biol.* **103**:287-297.
- Kochevar, R. E., N. S. Govind, and J. J. Childress.** 1993. Identification and characterization of two carbonic anhydrases from the hydrothermal vent tubeworm *Riftia pachyptila* Jones. *Molec. Mar. Biol. Biotechnol.* **2**:10-19.
- Kong, L. B., A. C. Siva, L. H. Rome, and P. L. Stewart.** 1999. Structure of the vault, a ubiquitous celular component. *Struct. Folding Des.* **7**:371-379.
- Kouchi, H., and S. Hata.** 1993. Isolation and characterization of novel nodulin cDNAs representing genes expressed at early stages of soybean nodule development. *Mol. Gen. Genet.* **238**:106-119.
- Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinf.* **5**:150-163.
- Lallier, F. H.** 2004. Oxygène et sulfure d'hydrogène : deux frères ennemis ? *Océanis* **30**:293-310.



- Lallier, F. H.** 2006. Thioautotrophic symbiosis: towards a new step in eukaryote evolution? *Cah. Biol. Mar.* **47**:391-396.
- Lee, R. W., and J. J. Childress.** 1994. Assimilation of inorganic nitrogen by marine invertebrates and their chemoautotrophic and methanotrophic symbionts. *Appl. Environm. Microbiol.* **60**:1852-1858.
- Lee, R. W., J. J. Robinson, and C. M. Cavanaugh.** 1999. Pathways of inorganic nitrogen assimilation in chemoautotrophic bacteria-marine invertebrate symbioses: Expression of host and symbiont glutamine synthetase. *J Exp Biol* **202**:289-300.
- Liang, P., and A. B. Pardee.** 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**:967-971.
- Livak, K. J., and T. D. Schmittgen.** 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Methods* **25**:402-408.
- Lüdeking, A., and A. Köhler.** 2002. Identification of six mRNA sequences of genes to multixenobiotic resistance (MXR) and biotransformation in *Mytilus edulis*. *Marine Ecology Progress Series* **238**:115-124.
- Lutz, R. A., T. M. Shank, D. J. Fornari, R. M. Haymon, M. D. Lilley, K. L. Von Damm, and D. Desbruyeres.** 1994. Rapid growth at deep-sea vents. *Nature* **371**:663-664.
- Markert, S., C. Arndt, H. Felbeck, D. Becher, S. M. Sievert, M. Hügler, D. Albrecht, J. Robidart, S. Bench, R. A. Feldman, M. Hecker, and T. Schweder.** 2007. Physiological Proteomics of the Uncultured Endosymbiont of *Riftia pachyptila*. *Science* **315**:247-250.
- McFall-Ngai, M.** 2006. The molecular mechanisms of host signaling in the establishment of the squid-*Vibrio* symbiosis. 5th International Symbiosis Society congress, Vienna.
- McHugh, D.** 1997. Molecular evidence that echiurans and pogonophorans are derived annelids. *Proc. Natl. Acad. Sci. USA* **94**:8006-8009.
- McHugh, D.** 2000. Molecular phylogeny of Annelida. *Can. J. Zool.* **78**:1873-1884.
- McMullin, E. R., S. Hourdez, S. W. Schaeffer, and C. R. Fisher.** 2003. Phylogeny and biogeography of deep sea vestimentiferan tubeworms and their bacterial symbionts. *Symbiosis* **34**:1-41.
- Mercenier, A., J. P. Simon, D. Haas, and V. Stalon.** 1980. Catabolism of L-arginine by *Pseudomonas aeruginosa*. *J. Gen. Microbiol.* **116**:381-389.
- Minic, Z., and G. Hervé.** 2004. Biochemical and enzymological aspects of the symbiosis between the deep-sea tubeworm *Riftia pachyptila* and its bacterial endosymbiont. *Eur. J. Biochem.* **271**:3093-3102.
- Minic, Z., S. Pastra Landis, F. Gaill, and G. Hervé.** 2002. Catabolism of pyrimidine nucleotides in the deep-sea tube worm *Riftia pachyptila*. *J. Biol. Chem.* **277**:127-134.
- Minic, Z., V. Simon, B. Penverne, F. Gaill, and G. Hervé.** 2001. Contribution of the Bacterial Endosymbiont to the Biosynthesis of Pyrimidine Nucleotides in the Deep-sea Tube Worm *Riftia pachyptila*. *J. Biol. Chem.* **276**:23777-23784.
- Mossink, M. H., A. van Zon, R. J. Scheper, P. Sonneveld, and E. A. Wiemer.** 2003. Vaults: a ribonucleoprotein particle involved in drug resistance? *Oncogene* **22**:7458-7467.
- Nakada, Y., Y. Jiang, T. Nishijyo, Y. Itoh, and C. D. Lu.** 2001. Molecular characterization and regulation of the aguBA operon, responsible for agmatine utilization in *Pseudomonas aeruginosa* PAO1. *J. Bacteriol.* **183**:6517-6524.
- Newman, J. R., S. Ghaemmamghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman.** 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**:840-6.
- Newton, I. L. G., T. Woyke, T. A. Auchtung, G. F. Dilly, R. J. Dutton, M. C. Fisher, K. M. Fontanez, E. Lau, F. J. Stewart, P. M. Richardson, K. W. Barry, E. Saunders,**

- J. C. Detter, D. Wu, J. A. Eisen, and C. M. Cavanaugh.** 2007. The *Calymene* *magnifica* chemoautotrophic symbiont genome. *Science* **315**:998-1000.
- Nussbaumer, A. D., C. R. Fisher, and M. Bright.** 2006. Horizontal endosymbiont transmission in hydrothermal vent tubeworms. *Nature* **441**:345-348.
- Ott, J., and M. Bright.** 2004. Sessile ciliates with bacterial ectosymbionts from twin cays, Belize. *Atoll Res. Bull.* **516**.
- Ott, J., M. Bright, and S. Bulgheresi.** 2004. Marine Microbial Thiotrophic Ectosymbioses. *Oceanog. Mar. Biol.: An Annual Review* **42**:95-118.
- Page, H. M., A. Fiala-Medioni, C. R. Fisher, and J. J. Childress.** 1991. Experimental evidence for filter-feeding by the hydrothermal vent mussel, *Bathymodiolus thermophilus*. *Deep-Sea Res. I* **38**:1455-1461.
- Pflugfelder, B., C. R. Fisher, and M. Bright.** 2005. The color of the trophosome: elemental sulfur distribution in the endosymbionts of *Riftia pachyptila* (Vestimentifera; Siboglinidae). *Mar. Biol.* **146**:895-901.
- Poole, A. M., and D. Penny.** 2007. Evaluating hypotheses for the origin of eukaryotes. *BioEssays* **29**:74-84.
- Powell, M. A., and G. N. Somero.** 1986. Adaptations to sulfide by hydrothermal vent animals: sites and mechanisms of detoxification and metabolism. *Biol. Bull.* **171**:274-290.
- Prieur, D., N. Benbouzid-Rollet, S. Chamroux, P. Durand, G. Erauso, E. Jacq, C. Jeanthon, and G. Mevel.** 1989. Distribution de divers types métaboliques bactériens sur un site hydrothermal profond (dorsale du pacifique oriental à 13°N). *Cah. Biol. Mar.* **30**:515-530.
- Ravaux, J., L. Gay, M. F. Voss-Foucart, and F. Gaill.** 1998. Tube growth process in the deep-sea hydrothermal vent tube-worm *Riftia pachyptila* (Vestimentifera): synthesis and degradation of chitin. *Cah. Biol. Mar.* **39**:99-107.
- Ravaux, J., A. Hassanin, J. Deutsch, F. Gaill, and U. Markmann-Mulisch.** 2001. Sequence analysis of the myosin regulatory light chain gene of the vestimentiferan *Riftia pachyptila*. *Gene* **263**:141-9.
- Rodriguez-Lanetty, M., W. S. Phillips, and V. M. Weis.** 2006. Transcriptome analysis of a cnidarian-dinoflagellate mutualism reveals complex modulation of host gene expression. *BMC Genomics* **7**.
- Rouse, G. W.** 2001. A cladistic analysis of Siboglinidae Caullery, 1914 (Polychaeta, Annelida): formerly the phyla Pogonophora and Vestimentifera. *Zool. J. Linn. Soc.* **132**:55-80.
- Rouse, G. W., and K. Fauchald.** 1995. The articulation of annelids. *Zool. Scr.* **24**:269-301.
- Rouse, G. W., and K. Fauchald.** 1997. Cladistics and polychaetes. *Zool. Scr.* **26**:139-204.
- Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**:5463 - 5467.
- Sarkar, N.** 1997. Polyadenylation of mRNA in Prokaryotes *Annu. Rev. Biochem.* **66**:173-197.
- Satriano, J., M. Isome, R. A. J. Casero, S. C. Thomson, and R. C. Blantz.** 2001. Polyamine transport system mediates agmatine transport in mammalian cells. *Am. J. Physiol. Cell. Physiol.* **281**:C329-C334.
- Scheffer, G. L., P. L. J. Wijngaard, M. J. Flens, M. A. Izquierdo, M. L. Slovak, H. M. Pinedo, C. J. L. M. Meijer, H. C. Clevers, and R. J. Scheper.** 1995. The drug resistance-related protein LRP is the human major vault protein. *Nat. Med.* **1**:578-582.
- Schulze, A.** 2003. Phylogeny of Vestimentifera (Siboglinidae, Annelida) inferred from morphology. *Zool. Scr.* **32**:321-342.

- Scott, K. M., M. Bright, S. A. Macko, and C. R. Fisher.** 1999. Carbon dioxide use by chemoautotrophic endosymbionts of hydrothermal vent vestimentiferans: Affinities for carbon dioxide, absence of carboxysomes, and delta <sup>13</sup>C values. *Mar. Biol.* **135**:25-34.
- Scott, K. M., C. R. Fisher, J. S. Vodenichar, E. R. Nix, and E. Minnich.** 1994. Inorganic carbon and temperature requirements for autotrophic carbon fixation by the chemoautotrophic symbionts of the giant hydrothermal vent tube worm, *Riftia pachyptila*. *Physiol. Zool.* **67**:617-638.
- Shank, T. M., D. J. Fornari, K. L. Von Damm, M. D. Lilley, R. M. Haymon, and R. A. Lutz.** 1998. Temporal and spatial patterns of biological community development at nascent deep-sea hydrothermal vents (9°50'N, East Pacific Rise). *Deep-Sea Res. II.* **45**:465-515.
- Shillito, B., J.-P. Lechère, and F. Gaill.** 1993. Microvilli-like structures secreting chitin crystallites. *J. Struct. Biol.* **111**:59-67.
- Sibuet, M., and K. Olu.** 1998. Biogeography, biodiversity and fluid dependence of deep-sea cold-seep communities at active and passive margins. *Deep Sea Res. Part II: Topical Studies in Oceanography* **45**:517-567.
- Sicot, F. X., M. Mesnage, M. Masselot, J. Y. Exposito, R. Garrone, J. Deutsch, and F. Gaill.** 2000. Molecular adaptation to an extreme environment: origin of the thermal stability of the pompeii worm collagen. *J. Mol. Biol.* **302**:811-820.
- Siva, A. C., S. Raval-Fernandes, A. G. Stephen, M. J. La Femina, R. J. Scheper, V. A. Kickhoefer, and L. H. Rome.** 2001. Up-regulation of Vaults may be necessary but not sufficient for multidrug resistance. *Int. J. Cancer* **92**:195-202.
- Smith, C. R., and A. R. Baco.** 2003. Ecology of whale falls at the deep-sea floor. *Oceanog. Mar. Biol.: an Annual Review* **41**:311-354.
- Southward, E. C.** 1988. Development of the gut and segmentation of newly settled stages of *Ridgeia* (vestimentifera): implications for relationship between vestimentifera and pogonophora. *J. Mar. Biol. Assoc. UK* **68**:465-487.
- Stephen, A. G., S. Raval-Fernandes, T. Huynh, M. Torres, V. A. Kickhoefer, and L. H. Rome.** 2001. Assembly of vault-like particles in insect cells expressing only the major vault protein. *J. Biol. Chem.* **276**:23217-23220.
- Stewart, F. J., I. L. G. Newton, and C. M. Cavanaugh.** 2005. Chemosynthetic endosymbioses: adaptations to oxic-anoxic interfaces. *Trends Microbiol.* **13**:439-448.
- Studier, F. W.** 2005. Protein Production by Auto-Induction in High-Density Shaking Cultures. *Protein Expr. Purif.* **41**:207-234.
- Terwilliger, N. B.** 1998. Functional adaptations of oxygen-transport proteins. *J. Exp. Biol.* **201**:1085-1098.
- Terwilliger, R. C., N. B. Terwilliger, and E. Schabtach.** 1980. The structure of hemoglobin from an unusual deep sea worm (Vestimentifera). *Comp. Biochem. Physiol.* **65B**:531-535.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673-4680.
- Toulmond, A., F. H. Lallier, J. De Frescheville, J. J. Childress, R. Lee, N. K. Sanders, and D. Desbruyères.** 1994. Unusual carbon dioxide-combining properties of body fluids in the hydrothermal vent tubeworm *Riftia pachyptila*. *Deep-Sea Res.* **10**:1447-1456.
- Tunnicliffe, V.** 1991. The Biology of Hydrothermal Vents - Ecology and Evolution. *Oceanog. Mar. Biol.* **29**:319-407.

- Tunnicliffe, V., J. F. Garrett, and H. P. Johnson.** 1990. Physical and biological factors affecting the behaviour and mortality of hydrothermal vent tube worms (vestimentiferans). *Deep-Sea Res.* **37**:103-125.
- Urakawa, H., N. Dubilier, Y. Fujiwara, D. E. Cunningham, S. Kojima, and D. A. Stahl.** 2005. Hydrothermal vent gastropods from the same family (Provannidae) harbour epsilon- and gamma-proteobacterial endosymbionts. *Environ. Microbiol.* **7**:750-754.
- Uschakov, P. V.** 1933. Eine neue Form aus der Familie Sabellidae (Polychaeta). *Zoologischer Anzeiger* **104**:205-208.
- van Der Land, J., and A. Z. Nørrevang.** 1975. The systematic position of *Lamellibrachia* (Annelida, Vestimentifera). *Zoologische Systematik und Evolutionsforschung Sonderheft* **1**:86-101.
- Vasus, S. K., N. L. Kedersha, and L. H. Rome.** 1993. cDNA Cloning and Disruption the Major Vault Protein  $\alpha$  Gene (mupA) in *Dictyostelium discoideum*. *J. Biol. Chem.* **268**:15356-15360.
- Vellai, T., and G. Vida.** 1999. The origin of eukaryotes: the difference between prokaryotic and eukaryotic cells. *Proc. Royal Soc. B: Biological Sciences* **266**:1571-1577.
- Verne, J.** 1869. Vingt mille lieues sous les mers p. 242. Atlas 2006, Evreux.
- Vetter, R. D.** 1991. Symbiosis and the evolution of novel trophic strategies: thiotrophic organisms at hydrothermal vents, p. 219-245. *In* L. Margulis, Fester, R. (ed.), *Symbiosis as a source of evolutionary innovation: speciation and morphogenesis*. MIT Press Cambridge, MA.
- Vetter, R. D., Fry, B.** 1998. Sulfur contents and sulfur-isotope compositions of thiotrophic symbioses in bivalve molluscs and vestimentiferan worms. *Mar. Biol.* **132**:453-460.
- Wan, J. S., S. J. Sharp, G. M.-C. Poirier, P. C. Wagaman, J. Chambers, J. Pyati, Y.-L. Horn, J. E. Galindo, A. Huvar, P. A. Peterson, M. R. Jackson, and M. G. Erlander.** 1996. Cloning differentially expressed mRNAs. *Nat. Biotechnol* **14**:1685-1691.
- Wang, G., C. L. Achim, R. L. Hamilton, C. A. Wiley, and V. Soontornniyomkij.** 1999. Tyramide Signal Amplification Method in Multiple-Label Immunofluorescence Confocal Microscopy. *Methods* **18**:459-464.
- Weber, R. E., and S. N. Vinogradov.** 2001. Nonvertebrate Hemoglobins: Functions and Molecular Adaptations. *Physiol. Rev.* **81**:569-627.
- Weis, V. M., G. J. Smith, and L. Muscatine.** 1989. A "CO<sub>2</sub> supply" mechanism in zooxanthellate cnidarians: role of carbonic anhydrase. *Mar. Biol.* **100**:195-202.
- Wernegreen, J. J.** 2002. Genome evolution in bacterial endosymbionts of insects. *Nat Rev Genet* **3**:850.
- Williams, C. A., D. C. Nelson, B. A. Farah, H. W. Jannasch, and J. M. Shively.** 1988. Ribulose biphosphate carboxylase of the procaryotic symbiont of a hydrothermal vent tube worm: Kinetics, activity and gene hybridization. *Fed. Eur. Microbiol. Soc. Microbiol. Lett.* **50**:107-112.
- Wilmot, D. R. J., and R. D. Vetter.** 1990. The bacterial symbiont from the hydrothermal vent tubeworm *Riftia pachyptila* is a sulfide specialist. *Mar. Biol.* **106**:273-283.
- Won, Y.-J., S. J. Hallam, G. D. O'Mullan, I. L. Pan, K. R. Buck, and R. C. Vrijenhoek.** 2003. Environmental Acquisition of Thiotrophic Endosymbionts by Deep-Sea Mussels of the Genus *Bathymodiolus*. *Appl. Envir. Microbiol.* **69**:6785-6792.
- Woyke, T., H. Teeling, N. N. Ivanova, M. Huntemann, M. Richter, F. O. Gloeckner, D. Boffelli, I. J. Anderson, K. W. Barry, H. J. Shapiro, E. Szeto, N. C. Kyrpides, M. Musmann, R. Amann, C. Bergin, C. Ruehland, E. M. Rubin, and N. Dubilier.** 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**:950-955.

- Wray, G. A., J. S. Levinton, and L. Shapiro. 1996. Molecular evidence for deep pre-Cambrian divergences among the metazoan phyla. *Science* **274**:568-573.
- Zal, F., B. N. Green, F. H. Lallier, and A. Toulmond. 1997. 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., F. H. Lallier, B. N. Green, S. N. Vinogradov, and A. Toulmond. 1996b. 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. 1996a. 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., E. Leize, F. H. Lallier, A. Toulmond, A. Van Dorsselaer, and J. J. Childress. 1998. S-Sulfohemoglobin and disulfide exchange: the mechanisms of sulfide binding by *Riftia pachyptila* hemoglobins. *Proc Natl Acad Sci USA* **95**:8997-9002.

# **LISTE DES FIGURES ET TABLEAUX**

# LISTE DES FIGURES ET TABLEAUX

## Chapitre I - Les endosymbioses chimioautotrophes

- Figure 1 :** Relation entre l'évolution de la teneur en oxygène dans l'atmosphère et un scénario probable de l'évolution cellulaire (d'après Vellai et Vida, 1999). 17
- Figure 2 :** Scénario évolutif hypothétique de l'origine et de la diversification des Eucaryotes mettant en exergue le rôle de l'endosymbiose (d'après Lallier, 2006). 18
- Figure 3 :** Carte des principaux sites réduits profonds explorés depuis 1977 (Galápagos) jusqu'en 2006 (d'après Desbruyères *et al.*, 2006) pour les sites hydrothermaux et complétée par les zones de suintements froids. 20
- Figure 4 :** Représentation schématique de trois symbioses hydrothermales chimioautotrophes. 26
- Figure 5 :** Représentation schématique des Eucaryotes hétérotrophes, phototrophes et thiotrophes, ces derniers étant hypothétiques, représentés actuellement par des endosymbioses métazoaires –  $\gamma$ -protéobactéries (d'après Vetter (1991) et Lallier (2004)). 31

## Chapitre II - L'organisme d'étude, *Riftia pachyptila*

- Figure 1 :** Vue dorsale de *Riftia pachyptila* dans et hors de son tube (Dessin : Michigan Science Art, USA) et colonie de *Riftia pachyptila* en milieu naturel (Cliché : Ifremer). 33
- Figure 2 :** Caractères morphologiques généraux pouvant être rencontrés au sein des Siboglinidae (d'après Rouse, 2001). 34
- Figure 3 :** Arbre phylogénétique obtenu à partir de l'analyse de l'ARN 18S ribosomique par maximum de parcimonie (d'après Halanych *et al.*, 1998). 35
- Figure 4 :** Coupes longitudinales de larve de *Riftia pachyptila* aposymbiotique et de juvénile symbiotique fixés (d'après Nussbaumer *et al.*, 2006). 38
- Figure 5 :** Morphologie externe et organisation de la branchie de *Riftia pachyptila* (d'après Andersen *et al.*, 2002). 39
- Figure 6 :** Section transversale d'un lobule de trophosome vue en microscopie électronique (Cliché : A. Andersen). 41
- Figure 7 :** Schéma résumant les modes d'acquisition de l'O<sub>2</sub>, CO<sub>2</sub> et H<sub>2</sub>S à travers l'épithélium branchial et leur transport jusqu'aux bactériocytes (par Y. Ferandin d'après la thèse de M.-C. De Cian, 2003). 45
- Figure 8 :** Schéma illustrant la connexion entre l'oxydation de l'H<sub>2</sub>S pour la production d'énergie et le cycle de Calvin-Benson dans une bactérie sulfo-oxydante (par Y. Ferandin d'après Minic et Hervé, 2004). 48

**Figure 9 :** Voies d'assimilation de l'azote inorganique chez *Riftia pachyptila* et contributions relatives de l'hôte et du symbionte (d'après Minic et Hervé, 2004). 49

**Figure 10 :** Explication de l'approche envisagée pour la thèse. 55

### Chapitre III - Matériel et Méthodes d'étude

**Figure 1 :** Protocole d'extraction d'ARN total utilisé pour chaque type de tissu. 58

**Tableau 1 :** Protocole de synthèse des ADNc double-brin des échantillons cibles et meneurs (d'après Clontech PCR-Select™ cDNA Subtraction Kit, BD Biosciences). 61

**Tableau 2 :** Protocole de digestion par *RsaI* des ADNc double-brin des échantillons cibles et meneurs (d'après Clontech PCR-Select™ cDNA Subtraction Kit, BD Biosciences). 61

**Tableau 3 :** Protocole de ligation des adaptateurs aux ADNc digérés de l'échantillon cible (d'après Clontech PCR-Select™ cDNA Subtraction Kit, BD Biosciences). 62

**Figure 2 :** Séquences nucléotidiques des adaptateurs 1 et 2R. 62

**Figure 3 :** Etapes de la technique d'Hybridation Soustractive Suppressiv (SSH) (d'après Clontech PCR-Select™ cDNA Subtraction Kit, BD Biosciences). 64

**Figure 4 :** Gamme de taille des inserts obtenus après clonage du produit de 2ème PCR de la banque BR-BW. 66

**Figure 5 :** Courbe d'amplification d'une réaction de PCR (en triplicat). 69

**Figure 6 :** Courbes d'amplification à partir de dilutions consécutives de matrice d'ADNc et courbe standard. 71

**Figure 7 :** Courbes standard construites pour l'ensemble des transcrits étudiés au cours de cette thèse à partir d'ADNc total de branchie, trophosome et paroi du corps. 72

*Encadré 1 :* Explication de la formule du  $2^{-\Delta\Delta Ct}$  (Livak et Schmittgen, 2001). 73

**Figure 8 :** Calcul du  $2^{-\Delta\Delta Ct}$  appliqué aux amplifications du 18S et d'un transcrit cible à partir d'ADNc total du calibre et de l'échantillon. 74

**Figure 9 :** Position des amorces utilisées pour l'obtention de la séquence complète RpCAbr par RACE. 75

**Figure 10 :** Schéma résumant le protocole de la RACE-PCR en 3' (5'/3' RACE Kit, 2<sup>nd</sup> Generation, Roche Applied Science) 75

**Figure 11:** Schéma résumant le protocole de la RACE-PCR en 5' (5'/3' RACE Kit, 2<sup>nd</sup> Generation, Roche Applied Science). 76

**Figure 12 :** Dot-blots réalisés pour vérifier la spécificité des sondes ADNc marquées au DIG. 78

**Tableau 4 :** Composition des tampons et solutions de lavage utilisés pour les dot-blots. 79



**Figure 13 :** Comparaison du système classique de détection de signal et du système TSA indirect (d'après Wang et al., 1999). 80

**Figure 14 :** Fixation de la protéine de fusion GST-MHr aux billes de glutathion sépharose (Cliché : T. Harnois). 83

Chapitre IV - Premier article : Application de la technique d'Hybridation Soustractive Suppressive à l'étude transcriptomique de *Riftia pachyptila*  
*Identification of proteins involved in the functioning of Riftia pachyptila symbiosis by Subtractive Suppression Hybridization*

**Table 1: Overall statistics based on the analysis of each library.** 91

**Fig. 1:** Proportion of sequences and contigs separated into 5 main categories (mitochondrial, all processes sequences (E-value<1), hypothetical sequences (E-value<1), hypothetical sequences (E-value>1) and no similarity found). 91

**Table 2: List of contigs with best E-values obtained for the BR-BW library sequences.** 93

**Table 3: List of contigs with best E-values obtained for the BW-BR library sequences.** 94

**Table 4: List of contigs with best E-values obtained for the TR-BW library sequences.** 95

**Table 5: List of contigs with best E-values obtained for the BW-TR library sequences.** 96

**Fig. 2:** Typical PCR profiles obtained after amplification of fragments of interesting cDNAs. 98

**Table 6: Equations of the standard curves obtained by amplification from total cDNA samples of branchial plume, trophosome and body wall tissues.** 99

**Fig. 3:** Relative expression levels of ribosomal RNA 16S, ccox I, ATPF1, Cathepsin, RpCAbr, RpCAtr, and MVP transcripts in the branchial plume, trophosome and body wall tissues. 100

**Table 7: Primers sequences used for the transcripts amplifications.** 108

Chapitre V - Deuxième article : Identification, expression et localisation d'un nouveau transcrit d'anhydrase carbonique chez *Riftia pachyptila*  
*Identification, sequencing, and localization of new carbonic anhydrase transcript from the hydrothermal vent tubeworm Riftia pachyptila*

**Fig. 1:** Normalized amplifications of RpCAbr and RpCAtr with 18S amplification. 122

**Fig. 2:** Localization of RpCAbr and RpCAtr mRNAs by Fluorescent *In Situ* Hybridization. 123

**Fig. 3:** Alignment of complete RpCAbr and RpCAtr amino acids sequences with some representative metazoan CA protein sequences. 127

**Fig. 4:** Neighbor-joining (NJ) tree obtained after a multiple alignment of 40 complete metazoan and 4 bacterial CA amino acids sequences. 129

<b>Table 1: Primers sequences for <i>Riftia pachyptila</i> carbonic anhydrase transcripts: RpCAbr and RpCAtr.</b>	<b>136</b>
---	------------

<b>Table S1: Best Blastx Hits obtained for the identification of RpCAbr transcript.</b>	<b>145</b>
---	------------

Chapitre VI - Troisième article : Caractérisation, expression et étude fonctionnelle d'un transcrit de myohémérythrine de *Riftia pachyptila*  
*Characterization, expression, and functional study of the myohemerythrin from the hydrothermal vent tubeworm Riftia pachyptila (Polychaeta, Siboglinidae)*

<b>Fig. 1:</b> Normalized amplifications of RpMHr, RpCAtr and 16S with 18S amplification from trophosome samples belonging to individuals from the S-EPR (n=5) and Genesis (n=4) sites.	<b>152</b>
---	------------

<b>Fig. 2:</b> Alignment of complete RpMHr amino acids sequence with some representative protostomian MHr and Hr protein sequences.	<b>153</b>
---	------------

<b>Fig. 3:</b> Neighbor-joining (NJ) tree obtained after a multiple alignment of protostomian MHr and Hr amino acids sequences.	<b>154</b>
---	------------

<b>Fig. 4:</b> Migrations of bacterial extracts before and after induction for four different clones.	<b>155</b>
---	------------

<b>Fig. 5:</b> Absorption spectra obtained for <i>Riftia pachyptila</i> GST-RpMHr and <i>Sipunculus nudus</i> Hr.	<b>156</b>
---	------------

<b>Table 1: SOD activities</b>	<b>157</b>
--------------------------------	------------

<b>Table 2: Primers sequences used for the transcripts amplifications by quantitative PCR.</b>	<b>162</b>
--	------------

Chapitre VII – Discussion, Conclusion et Perspectives

<b>Tableau 1 : Récapitulatif des différences d'expression constatées après l'étude de plusieurs transcrits au cours de la thèse.</b>	<b>170</b>
--	------------

<b>Figure 1 :</b> Phylogénie simplifiée résultant de l'analyse de séquences d'anhydrases carboniques de type $\alpha$ par les méthodes bayésienne et de maximum de vraisemblance (d'après Jackson <i>et al.</i> , 2007).	<b>177</b>
--	------------

<b>Tableau 2 : Distribution des transporteurs d'oxygène chez les animaux (d'après Terwilliger, 1998).</b>	<b>179</b>
---	------------

<b>Figure 2 :</b> Reconstitutions tridimensionnelles par cryo-microscopie électronique de la protéine Vault de rat traitée à la RNase (d'après Kong <i>et al.</i> , 1999).	<b>181</b>
--	------------

<b>Figure 3 :</b> Western-blot pour tester la présence de la protéine MVP à partir de 100 $\mu$ g d'extraits protéiques de branchie, trophosome et paroi du corps, avec l'anticorps « Bob » (don de L. Rome).	<b>182</b>
---	------------

<b>Figure 4 :</b> Schéma représentant le rôle hypothétique de la protéine Vault dans le transport nucléocytoplasmique de drogues et/ou métabolites (d'après Mossink <i>et al.</i> , 2003).	<b>183</b>
--	------------

## **RESUME ET ABSTRACT**

## RESUME - ABSTRACT

### **Approche transcriptomique pour identifier des protéines impliquées dans le fonctionnement de la symbiose *Riftia pachyptila*.**

Le ver tubicole géant des sources hydrothermales profondes *Riftia pachyptila* (Annélide Polychète Siboglinidae) est probablement l'un des organismes symbiotiques vivant avec des bactéries chimioautotrophes sulfo-oxydantes les plus étudiés. Logés dans un organe interne spécialisé, le trophosome (TR), ces symbiontes dépendent de l'hôte pour leur approvisionnement en O<sub>2</sub>, CO<sub>2</sub> et H<sub>2</sub>S, acquis du milieu au niveau du panache branchial (BR) de l'animal. Une approche moléculaire globale par Hybridation Soustractive Suppressive (SSH) a été utilisée pour tenter d'identifier des protéines impliquées dans les échanges de métabolites dans ces tissus. Quatre banques d'ADNc soustractives ont été constituées, en prenant la paroi du corps (BW) comme tissu de référence (BR-BW, BW-BR, TR-BW, BW-TR). Plus de 700 séquences ont été obtenues avec une moyenne de 45 ADNc différents (dont la moitié a pu être identifiée) par banque. L'expression différentielle des transcrits les plus intéressants a été validée par PCR quantitative. Un nouveau transcrit d'anhydrase carbonique (RpCAbr) spécifiquement exprimé dans la branchie a été obtenu et s'ajoute à celui déjà connu (RpCAtr) davantage exprimé dans le trophosome (confirmation par hybridation fluorescente *in situ*). Un transcrit de myohémérythrine, obtenu dans la banque TR-BW, apparaît clairement spécifique du trophosome. Pigment peu fréquent chez les Annélides, ce transcrit a été surexprimé en système hétérologue pour explorer un aspect plus fonctionnel. Au niveau du panache branchial, la protéine Major Vault pourrait constituer une piste d'étude intéressante dans le transport de métabolites.

**Mots-clés :** Siboglinidae ; chimioautotrophie ; banques d'ADNc soustractives ; PCR quantitative ; anhydrase carbonique ; myohémérythrine.

---

### **Transcriptomics approach to identify proteins involved in the functioning of *Riftia pachyptila* symbiosis**

The giant hydrothermal vent tubeworm *Riftia pachyptila* (Siboglinid Polychaete Annelid) is probably one of the most extensively studied organisms that are symbiotic with sulfide-oxidizing bacteria. Located in a specialized internal organ, the trophosome (TR), these symbionts depend on the host for their supply in O<sub>2</sub>, CO<sub>2</sub> and H<sub>2</sub>S that are acquired from the environment by the branchial plume (BR) of the worm. We used a global molecular approach by Subtractive Suppression Hybridization (SSH) to try to identify proteins involved in the metabolites exchanges in these tissues. Four subtractive cDNA libraries were constituted, with the body wall (BW) as a reference tissue (BR-BW, BW-BR, TR-BW, BW-TR). More than 700 sequences were obtained with an average of 45 different cDNAs per library, of which half could be identified. The differential expression of the most interesting transcripts was confirmed by quantitative PCR. A new carbonic anhydrase transcript (RpCAbr) specifically expressed in the gills was obtained in addition to the one already known (RpCAtr) that is mostly expressed in the trophosome (checking by fluorescent *in situ* hybridization). A myohemerythrin transcript, obtained in the TR-BW library, appears clearly specific of the trophosome tissue. Given that it is a particularly rare pigment in Annelids, this transcript was overexpressed in a heterologous system to explore a more functional aspect. In the branchial plume tissue, the Major Vault Protein could constitute an interesting clue for further studies about the transport of metabolites.

**Key words :** Siboglinidae; chemoautotrophy; subtractive cDNA libraries; quantitative PCR; carbonic anhydrase; myohemerythrin.

## **AUTRES PUBLICATIONS**

## AUTRES PUBLICATIONS DE L'AUTEUR

Chabasse C., Bailly X., **Sanchez S.**, Rousselot M., and Zal F., 2006. Gene Structure and Molecular Phylogeny of the Linker Chains from the Giant Annelid Hexagonal Bilayer Hemoglobins. *Journal of Molecular Evolution*, 63: 365-374.

Equipe Ecophysiologie: Adaptation et Evolution Moléculaires, UPMC—CNRS UMR 7144, Station Biologique, BP 74, 29682 Roscoff cedex, France

Giant extracellular hexagonal bilayer hemoglobin (HBL-Hb), found only in annelids, is an ~3500-kDa heteropolymeric structure involved in oxygen 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 13 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 four-intron and five-exon structure and gave rise to the two families subsequent to a duplication event.

**Sanchez S.**<sup>1</sup>, Dettai A.<sup>1</sup>, Bonillo C.<sup>2</sup>, Ozouf-Costaz C.<sup>2</sup>, Dietrich B.<sup>3</sup> and Lecointre G.<sup>1</sup>, 2007. Molecular and morphological phylogenies of the Nototheniidae, with on taxonomic focus on the Trematominae. *Polar Biology*, 30: 155-166.

<sup>1</sup>Département Systématique et Evolution, UMR 7138 CNRS Systématique, Adaptation, Evolution, Muséum National d'Histoire Naturelle, CP26, 43 rue Cuvier, 75231 Paris Cedex 05, France

<sup>2</sup>IFR 101 CNRS, Service de Systématique Moléculaire, Muséum National d'Histoire Naturelle, CP26, 43 rue Cuvier, 75231 Paris Cedex 05, France

<sup>3</sup>Northeastern University, Boston, MA 02115, USA

Four independent molecular data sets were sequenced in order to solve longstanding phylogenetic problems among Antarctic teleosts of the family Nototheniidae. The anatomical data of Balushkin (2000) were also coded into a matrix of 106 characters in order to test the parsimony of his taxonomic conclusions. Molecular results confirm Balushkin's Pleuragrammatinae but not his Nototheniinae. Different genes used here found the "clade A" establishing the paraphyly of the Nototheniinae *sensu lato*; i.e. *Lepidonotothen* and *Patagonotothen* are more closely related to the Trematominae than to *Notothenia*. The genus *Notothenia* is *paraphyletic* and *Paranotothenia* should become *Notothenia*. Previously no molecular data set could assign a reliable position for the genus *Gobionotothen*. For the first time robust results are obtained for the phylogeny among the Trematominae. *Trematomus scotti* is the sister-group of all others, then *Trematomus newnesi* emerges, then *Trematomus eulepidotus*. Among the crown group, three clades emerge: 1: *Trematomus hansonii* + *Trematomus bernacchii* + *Trematomus vicarius*; 2: *Trematomus pennellii* + *Trematomus lepidorhinus* + *Trematomus loennbergii*; 3: *Trematomus (Pagothenia) borchgrevinkii* + *Trematomus nicolai*. *Pagothenia* should become *Trematomus* to make the genus *Trematomus* monophyletic. The *Trematomus* tree found here did not match the topology obtained with Balushkin's morphological matrix. The tree shows that the tendencies shown by some trematomines to secondarily colonize the water column are not gained through common ancestry.

Galewski T.<sup>1</sup>, Tilak M.-k.<sup>1</sup>, **Sanchez S.**<sup>2</sup>, Chevret P.<sup>1</sup>, Paradis E.<sup>1,3</sup> and Douzery E. J. P.<sup>1</sup>, 2006. The evolutionary radiation of Arvicolinae rodents (voles and lemmings): relative contribution of nuclear and mitochondrial DNA phylogenies. *BMC Evolutionary Biology*, 6(1):80.

<sup>1</sup>Laboratoire de Paléontologie, Phylogénie et Paléobiologie – CC064, Institut des Sciences de l'Evolution UMR 5554/CNRS, Université Montpellier II; Place E. Bataillon, 34 095 Montpellier Cedex 05 –, France

<sup>2</sup>Ecophysiologie : adaptation et évolution moléculaires, Station Biologique, Place Georges Teissier – BP 7429 680 Roscoff –, France

<sup>3</sup>Institut de Recherche pour le Développement, UR175 CAVIAR, GAMET – BP 5095, 361 rue Jean François Breton, 34196 Montpellier Cedex 5 –, France

## Background

Mitochondrial and nuclear genes have generally been employed for different purposes in molecular systematics, the former to resolve relationships within recently evolved groups and the latter to investigate phylogenies at a deeper level. In the case of rapid and recent evolutionary radiations, mitochondrial genes like cytochrome b (CYB) are often inefficient for resolving phylogenetic relationships. One of the best examples is illustrated by Arvicolinae rodents (Rodentia; Muridae), the most impressive mammalian radiation of the Northern Hemisphere which produced voles, lemmings and muskrats. Here, we compare the relative contribution of a nuclear marker – the exon 10 of the growth hormone receptor (GHR) gene – to the one of the mitochondrial CYB for inferring phylogenetic relationships among the major lineages of arvicoline rodents.

## Results

The analysis of GHR sequences improves the overall resolution of the Arvicolinae phylogeny. Our results show that the Caucasian long-clawed vole (*Prometheomys schaposnikowi*) is one of the basalmost arvicolines, and confirm that true lemmings (*Lemmus*) and collared lemmings (*Dicrostonyx*) are not closely related as suggested by morphology. Red-backed voles (*Myodini*) are found as the sister-group of a clade encompassing water vole (*Arvicola*), snow vole (*Chionomys*), and meadow voles (*Microtus* and allies). Within the latter, no support is recovered for the generic recognition of *Blanfordimys*, *Lasiopodomys*, *Neodon*, and *Phaiomys* as suggested by morphology. Comparisons of parameter estimates for branch lengths, base composition, among sites rate heterogeneity, and GTR relative substitution rates indicate that CYB sequences consistently exhibit more heterogeneity among codon positions than GHR. By analyzing the contribution of each codon position to node resolution, we show that the apparent higher efficiency of GHR is due to their third positions. Although we focus on speciation events spanning the last 10 million years (Myr), CYB sequences display highly saturated codon positions contrary to the nuclear exon. Lastly, variable length bootstrap predicts a significant increase in resolution of arvicoline phylogeny through the sequencing of nuclear data in an order of magnitude three to five times greater than the size of GHR exon 10.

## Conclusion

Our survey provides a first resolved gene tree for Arvicolinae. The comparison of CYB and GHR phylogenetic efficiency supports recent assertions that nuclear genes are useful for resolving relationships of recently evolved animals. The superiority of nuclear exons may reside both in (i) less heterogeneity among sites, and (ii) the presence of highly informative sites in third codon positions, that evolve rapidly enough to accumulate synapomorphies, but slow enough to avoid substitutional saturation.