



**HAL**  
open science

# Bases génétiques du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria* *glabrata*

Emmanuel Roger

► **To cite this version:**

Emmanuel Roger. Bases génétiques du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria glabrata*. Sciences du Vivant [q-bio]. Université de Perpignan, 2008. Français. NNT: . tel-00344751

**HAL Id: tel-00344751**

**<https://theses.hal.science/tel-00344751>**

Submitted on 5 Dec 2008

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

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

ACADEMIE DE MONTPELLIER  
UNIVERSITE VIA DOMITIA

**THESE**

pour obtenir le grade de  
DOCTEUR DE L'UNIVERSITE VIA DOMITIA

**Ecologie - Génétique - Biologie moléculaire - Biochimie**

***Spécialité : Parasitologie***

*Ecole Doctorale : Energie et Environnement*

présentée et soutenue publiquement par

**Emmanuel ROGER**

Le 23 septembre 2008

**Bases génétiques  
du polymorphisme de compatibilité  
dans l'interaction  
*Schistosoma mansoni / Biomphalaria glabrata***

devant le jury composé de :

Mr. André THERON, Directeur de recherche CNRS, Perpignan	Examineur
Mr. Raymond PIERCE, Directeur de recherche CNRS, Institut Pasteur de Lille	Examineur
Mme. Christine COUSTAU, Chargé de Recherche CNRS, Institut Pasteur de Lille	Rapporteur
Mr. Jean-Michel ESCOUBAS, Chargé de Recherche CNRS, Montpellier II	Rapporteur
Mr. Guillaume MITTA, Professeur, Perpignan	Examineur



## Remerciements

*Je tiens à remercier très sincèrement toutes les personnes qui ont contribué, de près ou de loin, à la réalisation de ce travail de doctorat.*

*Tout d'abord, je souhaite remercier très sincèrement mes deux directeurs de thèse, Monsieur André Théron et Monsieur Guillaume Mitta, de m'avoir accueilli au laboratoire de Biologie et Ecologie Tropicale et Méditerranéenne et de m'avoir permis d'effectuer ce travail de doctorat. Je les remercie également de leurs précieux conseils, de leur expérience, de leur disponibilité et de leur gentillesse.*

*Je remercie particulièrement les membres du jury qui ont accepté de consacrer de leur temps précieux afin d'évaluer ce travail : Madame Christine Coustau et Monsieur Jean-Michel Escoubas pour avoir accepté de juger ce travail en tant que rapporteurs ainsi que Monsieur Raymond Pierce.*

*Je remercie sincèrement tous les membres du laboratoire de Biologie et Ecologie Tropicale et Méditerranéenne pour leur disponibilité, leur compétence, leur sympathie et pour m'avoir permis d'évoluer dans ce laboratoire tout au long de ces quelques années passées ensemble. Merci à Richard, Jean-françois, Rémi, Anne, Bernard, Gaël, Nath, Jérôme, Betty, Olivier, Juliette, Hélène, Gabriel, Irma, Anne, Christoph, Benjamin et Céline.*

*Merci également à tous les thésards pour les bons et très bons moments passés ensemble : Rodrigue, Fred, Mathieu, Géraldine, Yves, Anne, Sophie et Nico.*

*Enfin, je tiens à remercier tout particulièrement mes grands parents, mes parents, ma soeur, mes amis qui m'ont soutenu et accompagné.*

## Table des matières

<b>INTRODUCTION</b> .....	<b>11</b>
<b>I. INTERACTION HOTE VERTEBRE / PATHOGENE</b> .....	<b>14</b>
<b>1. Hôtes vertébrés : immunité adaptative, polymorphisme et diversification</b> .....	<b>15</b>
<b>1.1. Les gnathostomes</b> .....	<b>15</b>
<b>1.1.1. Immunoglobulines et TCR</b> .....	<b>15</b>
<i>a. Organisation génomique des gènes codant les Ig/TCR</i> .....	16
<i>b. Diversification du répertoire d'Ig/TCR durant la phase indépendante de l'antigène</i> .....	16
<i>c. Diversification du répertoire d'Ig durant la phase dépendante de l'antigène</i> .....	20
<i>d. Diversité structurale et spécificité multiple des Ig</i> .....	22
<b>1.1.2. Complexe Majeur d'Histocompatibilité</b> .....	<b>24</b>
<i>a. Polygénie du CMH</i> .....	24
<i>b. Polymorphisme du CMH</i> .....	25
<b>1.1.3. Conclusion</b> .....	<b>26</b>
<b>1.2. Les agnathes</b> .....	<b>26</b>
<b>1.2.1. Un type différent de réponse adaptative</b> .....	<b>26</b>
<b>1.2.2. Les récepteurs lymphocytaires variables (VLR)</b> .....	<b>27</b>
<b>1.2.3. Conclusion</b> .....	<b>28</b>
<b>2. Pathogènes</b> .....	<b>30</b>
<b>2.1. Bactéries pathogènes et virus</b> .....	<b>30</b>
<b>2.1.1. Bactéries pathogènes</b> .....	<b>31</b>
<i>a. Les protéines Opa chez Nesseria gonorrhoeae</i> .....	32
<i>b. Les monomères de piline chez Nesseria gonorrhoeae</i> .....	34
<i>c. Les lipooligosaccharides chez Nesseria gonorrhoeae et meningitidis</i> .....	35
<i>d. Les Vlp et Vls chez Borrelia hermsii</i> .....	35
<b>2.1.2. Virus</b> .....	<b>35</b>
<i>a. L'hémagglutinine et la neuraminidase chez le virus Influenza</i> .....	35
<i>b. Les glycoprotéines gp120 et gp41 chez le virus de l'immunodéficience humaine</i> .....	36
<b>2.1.3. Conclusion</b> .....	<b>38</b>
<b>2.2. Parasites eucaryotes</b> .....	<b>38</b>
<b>2.2.1 Les VSG chez Trypanosoma brucei</b> .....	<b>39</b>
<i>a. Conditions nécessaires à la variation antigénique des VSG chez T. brucei</i> .....	41
<i>b. Régulation épigénétique de l'expression monoallélique des gènes vsg</i> .....	45
<i>c. Conclusion</i> .....	47
<b>2.2.2 Les VSP chez Giardia lamblia</b> .....	<b>48</b>
<b>2.2.3 Les PfEMP1 chez Plasmodium falciparum</b> .....	<b>48</b>
<i>a. Structure et organisation génomique des gènes var</i> .....	50

b. Transcription des gènes var.....	51
c. Régulation épigénétique de l'expression monoallélique des gènes var.....	51
d. Polymorphisme du répertoire de gènes var.....	54
e. Conclusion.....	55
<b>3. Conclusion hôte vertébré / pathogène.....</b>	<b>55</b>
<b>II. INTERACTION HOTE INVERTEBRE / PATHOGENE.....</b>	<b>57</b>
<b>1. Hôtes invertébrés : immunité innée et diversification.....</b>	<b>58</b>
1.1. Les Dscam chez les diptères.....	58
1.2. Les VCBP chez <i>Branchiostoma floridae</i> .....	61
1.3. Les FREP chez <i>Biomphalaria glabrata</i> .....	62
1.4. Conclusion.....	64
<b>2. Pathogènes.....</b>	<b>64</b>
<b>III. LE MODELE <i>S. MANSONI</i> / <i>B. GLABRATA</i>.....</b>	<b>65</b>
<b>1. Généralités.....</b>	<b>65</b>
1.1. Cycle de vie de <i>S. mansoni</i> .....	66
<b>2. Caractéristiques de l'interaction mollusque / schistosome.....</b>	<b>66</b>
2.1 Spécificité d'hôte.....	66
2.2 Multiplication clonale intense.....	68
2.3. Le polymorphisme de compatibilité.....	68
2.4. Le modèle de concordance des phénotypes et sélection des souches.....	70
<b>3. Le modèle d'étude.....</b>	<b>72</b>
<b>CHAPITRE 1.....</b>	<b>73</b>
Déterminants moléculaires du polymorphisme de compatibilité dans l'interaction <i>Biomphalaria glabrata</i> / <i>Schistosoma mansoni</i> : identification de nouveaux candidats par une approche protéomique comparative globale.	
<b>CHAPITRE 2.....</b>	<b>87</b>
Analyse de l'expression de protéines de type mucine hautement polymorphes ( <i>SmPoMuc</i> ) chez le parasite <i>Schistosoma mansoni</i> .	
<b>CHAPITRE 3.....</b>	<b>101</b>
Le polymorphisme de mucines sécrétées par un parasite métazoaire ( <i>Schistosoma mansoni</i> ) dans l'interaction avec son hôte invertébré ( <i>Biomphalaria glabrata</i> ) : recombinaison, épissage alternatif, transcription indépendante et glycosylation.	

<b>DISCUSSION</b> .....	<b>172</b>
<b>I. LES <i>SmPoMuc</i> ET LES GLYCOPROTEINES / MUCINES D'AUTRES SYSTEMES HOTE / PARASITE</b> .....	<b>174</b>
<b>1. Les <i>SmPoMuc</i> au coeur de la compatibilité</b> .....	<b>174</b>
1.1. Les <i>SmPoMuc</i> , des glycoprotéines de type mucine.....	175
1.2. Les <i>SmPoMuc</i> , polymorphisme et statut de glycosylation.....	176
<b>2. Une stratégie « écran de fumée »</b> .....	<b>177</b>
2.1. Les <i>SmPoMuc</i> chez <i>S. mansoni</i> .....	177
2.2. Les SOWgp chez <i>Coccidioides immitis</i> et <i>C. posadasii</i> .....	178
2.3. Les TcSMUG et TcMUC chez <i>Trypanosoma cruzi</i> .....	180
2.3.1. Les molécules de type mucine TcSMUG des stades en interaction avec l'insecte.....	181
2.3.2. Les molécules de type mucine TcMUC des stades en interaction avec le mammifère.....	182
2.4. Les Tc-MUC chez <i>Toxocara canis</i> .....	183
2.5. Conclusion.....	184
<b>II. BASES GENETIQUES DU POLYMORPHISME DES <i>SmPoMuc</i></b>	<b>186</b>
<b>1. Le polymorphisme des <i>SmPoMuc</i> au niveau génomique</b> .....	<b>186</b>
1.1. Evolution des gènes selon le modèle « birth and death ».....	186
1.2. Recombinaisons fréquentes à l'origine des gènes mosaïques.....	187
1.3. Recombinaisons ectopiques et diversité combinatoire.....	188
<b>2. Le polymorphisme des <i>SmPoMuc</i> au niveau des transcrits</b> .....	<b>191</b>
2.1. Epissage alternatif et aberrant au niveau de la région 3'.....	191
2.2. <i>Trans</i> -épissage et « exon repetition » au niveau de la région 5'.....	192
2.3. Polymorphisme d'expression individuel spécifique.....	193
<b>CONCLUSION ET PERSPECTIVES</b> .....	<b>195</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b> .....	<b>199</b>
<b>ANNEXE 1 :</b> .....	<b>225</b>
Travaux développés sur le modèle <i>Schistosoma mansoni</i> / <i>Biomphalaria glabrata</i>	
<b>ANNEXE 2 :</b> .....	<b>245</b>
Travaux développés sur le modèle <i>Biomphalaria glabrata</i> / <i>Echinostoma caproni</i>	

## Table des figures

<b>Figure 1</b> : Organisation germinale des locus de chaînes légères ( $\kappa$ et $\lambda$ ) et lourdes (H) des immunoglobulines chez l'homme.....	17
<b>Figure 2</b> : Réarrangement des segments géniques au niveau des chaînes lourdes et légères des immunoglobulines durant la différenciation des lymphocytes B.....	19
<b>Figure 3</b> : Hypermutations somatiques (HMS).....	21
<b>Figure 4</b> : Commutation de classe (« class switch recombinaison »).....	23
<b>Figure 5</b> : Représentation schématique de l'assemblage d'un gène <i>vlr</i> mature chez la lamproie.....	29
<b>Figure 6</b> : Régulation de l'expression des gènes <i>opa</i> chez <i>Neisseria gonorrhoeae</i> .....	33
<b>Figure 7</b> : Cycle de vie de <i>Trypanosoma brucei</i> .....	40
<b>Figure 8</b> : Diagramme schématique illustrant le processus de variation antigénique chez <i>Trypanosoma brucei</i> .....	41
<b>Figure 9</b> : Localisation génomique des gènes <i>vsg</i> chez <i>Trypanosoma brucei</i> .....	42
<b>Figure 10</b> : Différents mécanismes de remplacement des gènes <i>vsg</i> .....	43
<b>Figure 11</b> : Génération des gènes <i>vsg</i> mosaïques.....	45
<b>Figure 12</b> : Régulation épigénétique de l'expression monoallélique des gènes <i>vsg</i> chez <i>Trypanosoma brucei</i> .....	47
<b>Figure 13</b> : Cycle de vie de <i>Plasmodium falciparum</i> .....	49
<b>Figure 14</b> : Représentation de l'organisation en cluster des régions télomériques et subtélomériques chez <i>Plasmodium falciparum</i> favorable aux recombinaisons géniques.....	50
<b>Figure 15</b> : Régulation épigénétique de l'expression monoallélique des gènes <i>var</i> chez <i>Plasmodium falciparum</i> .....	52
<b>Figure 16</b> : Régulation épigénétique de l'activation des gènes <i>var</i> .....	53
<b>Figure 17</b> : Diversification des gènes <i>Dscam</i> chez les diptères par un processus d'épissage alternatif différentiel.....	60
<b>Figure 18</b> : Cycle de vie de <i>Schistosoma mansoni</i> , modèle d'étude.....	67
<b>Figure 19</b> : Modèle de concordance des phénotypes (« matching phenotype model »).....	71
<b>Figure 20</b> : Implication de séquences introniques complémentaires dans le processus d'« exon repetition ».....	194
<b>Figure 21</b> : Effet de la trichostatine A (TSA), inhibiteur des histones désacétylases (HDAC), sur le profil d'expression des <i>SmPoMuc</i> chez <i>S. mansoni</i> .....	197



## Abréviations

ADN	: Acide DésoxyriboNucléique	LRR	: Leucine-Rich Repeat
ADNc	: ADN complémentaire	NOD	: Nucleotide Oligomerization Domain
AID	: Activating Induced cytidine Deaminase	PAMP	: Pathogen-Associated Molecular Pattern
ARN	: Acide RiboNucléique	PfEMP1	: <i>P. falciparum</i> Erythrocyte Membrane Proteine 1
C	: segment génique Constant	Pro	: Proline
CDR	: Complementary-Determining Region	PRR	: Pattern Recognition Receptor
cf. d-	: conférer à une partie de la discussion	PRRs	: Proline-Rich Regions
cf. i-	: conférer à une partie de l'introduction	r1	: unité répétée de type r1
CMH	: Complexe Majeur d'Histocompatibilité	r2	: unité répétée de type r2
D	: segment génique de Diversité	RAG	: Recombination Activating Genes
DSB	: Double-Strand Break	Ser	: Sérine
Dscam	: Down syndrome cell adhesion molecule	SmPoMuc	: <i>S. mansoni</i> Polymorphic Mucin
FBG	: FiBrinoGen	TcMUC	: <i>Trypanosoma cruzi</i> MUCin
FREP	: Fibrinogen-RElated Protein	TCR	: T Cell Receptor
GPI	: Glycosyl-Phosphatidyl-Inositol	TcSMUG	: <i>T. cruzi</i> Small MUCin-like Gene
H	: chaîne lourde des immunoglobulines	TdT	: Terminal desoxynucleotidyl-Transferase
HAT	: Histone Acetyl Transferase	Thr	: Thréonine
HDAC	: Histone DeACetylase	TLR	: Toll Like Receptor
HMS	: HyperMutations Somatiques	TSA	: TrichoStatin A
Ig	: Immunoglobulines	UR	: Unités Répétées
IgA	: Immunoglobuline de type A	URg	: Unités Répétées génomiques
IgD	: Immunoglobuline de type D	V	: segment génique Variable
IgE	: Immunoglobuline de type E	VCBP	: Variable Chitin-Binding Protein
IgG	: Immunoglobuline de type G	VIH	: Virus de l'Immunodéficience Humaine
IgM	: Immunoglobuline de type M	VLR	: Variable Lymphocyte Receptor
Ig-V	: domaine protéique variable des immunoglobulines	VNTR	: Variable Number of Tandem Repeats
J	: segment génique de Jonction	VSG	: Variant Surface Glycoprotein
LOS	: LipoOligoSaccharides	κ	: chaîne légère κ des immunoglobulines
LPS	: LipoPolySaccharides	λ	: chaîne légère λ des immunoglobulines

---

## Glossaire

**Résistance** : Caractéristiques génétique, biochimique ou physiologique de l'hôte qui empêche l'établissement, la survie ou le développement du pathogène.

**Infectivité** : Caractéristiques génétique, biochimique ou physiologique du pathogène qui détermine sa capacité à infecter (envahir) l'hôte. Une infectivité élevée n'est pas forcément reliée à une virulence sévère.

**Virulence** : Réduction directe ou indirecte de la fitness de l'hôte attribuée à l'infection par le pathogène.

**Souche** : Isolat géographique partageant le même phénotype.

**Bases génétiques  
du polymorphisme de compatibilité  
dans l'interaction  
*Schistosoma mansoni* / *Biomphalaria glabrata***

## INTRODUCTION

Ce travail de thèse s'intègre dans un projet sur l'étude de la compatibilité dans les interactions hôte / parasite. Au cours de ce type d'interaction, le parasite vit aux dépens de son hôte, véritable milieu biologique, qu'il va utiliser pour se développer. L'association est obligatoire pour le parasite qui seul en tire avantage pendant l'intégralité ou une partie de son cycle de vie. Toutefois, l'hôte réagit et il s'établit alors entre les deux protagonistes un dialogue. Hôte et parasite s'influencent réciproquement et s'installe entre eux un équilibre dynamique sans que le devenir de l'un ou l'autre ne soit en règle générale menacé. Néanmoins, cet équilibre peut être rompu. C'est le cas lorsque l'hôte ne parvient plus à renflouer ses pertes ou à s'opposer aux effets toxiques du parasite ou bien encore lorsque le parasite ne parvient plus à se développer.

Les interactions hôte / parasite qui se caractérisent par cet équilibre dynamique sont qualifiées d'interactions durables (Combes, 1995) et résultent d'une co-évolution prolongée des deux partenaires. Les deux opposants, à chaque génération, vont exercer l'un sur l'autre des pressions de sélection. Ainsi, chaque fois que l'hôte s'adapte, le parasite est contraint de s'adapter, et réciproquement (Webster, Davies, 2001). C'est la théorie de la « Reine Rouge » (Van Valen, 1973) appliquée aux systèmes hôte / parasite (Combes, 2000) laquelle prédit une course aux armements avec une évolution réciproque et antagoniste des mécanismes de résistance de l'hôte (Coustau *et al.*, 2000; Trowsdale, Parham, 2004) et d'infectivité et/ou de virulence du parasite (Read, 1994). Ainsi, chez les hôtes comme chez les parasites, les molécules jouant un rôle crucial dans ces interactions devraient donc présenter de hauts niveaux de diversité et/ou de polymorphisme afin d'assurer cette dynamique co-évolutive.

Ces modèles théoriques de co-évolution ont été largement vérifiés dans des interactions impliquant des vertébrés et leurs pathogènes (cf. *i-I. Interaction hôte vertébré / pathogène*). En effet, comme nous le verrons dans cette première partie de l'introduction, hôtes vertébrés et pathogènes disposent d'une importante plasticité phénotypique qui s'exprime au niveau des molécules impliquées dans l'interaction. Dans cette course aux armements, les hôtes vertébrés (cf. *i-I.1. Hôtes vertébrés : immunité adaptative, polymorphisme et diversification*) ont développé divers mécanismes afin d'assurer cette plasticité : réarrangement somatique, diversification génique ou encore diversification structurale. Ces mécanismes sont à la base des hauts niveaux de polymorphisme et/ou de diversification des molécules de reconnaissance immunitaire telles les Ig/TCR

(Immunoglobulines/ « T Cell Receptor »), protéines du CMH (Complexe Majeur d'Histocompatibilité) ou autres VLR (« Variable Lymphocyte Receptor »). En parallèle, les pathogènes (cf. i-I.2. *Pathogènes*) ont mis en place des mécanismes intervenant au niveau génique (recombinaison, réarrangement) à l'origine de la diversification et/ou du polymorphisme de leurs antigènes de surface. Ces antigènes, tels les Opa (« Opacity protein »), les glycoprotéines gp120, les VSG (« Variant Surface Glycoprotein ») ou encore les PfEMP1 (« *P. falciparum* Erythrocyte Membrane Protein 1 ») sont impliqués dans un mécanisme de variation phénotypique permettant aux pathogènes d'échapper à la réponse immune de l'hôte.

En revanche, comme la deuxième partie de cette introduction en fera état, les choses ne sont pas aussi clairement définies dans les interactions entre hôtes invertébrés et pathogènes (cf. i-II. *Interaction hôte invertébré / pathogène*). Cet état de fait est dû, en partie, au manque de connaissances disponibles. Cependant, des travaux récents mettent en évidence que l'immunité innée des invertébrés (cf. i-II.1 *Hôtes invertébrés : immunité innée et diversification*) présente des molécules de reconnaissance diversifiées et/ou polymorphes, telles que les FREP (« Fibrinogen-RElated Protein »), les Dscam (« Down syndrome cell adhesion molecule »), ou encore les VCBP (« Variable Chitin-Binding Protein »), lesquelles « répondent » vraisemblablement au polymorphisme et/ou à la diversité moléculaire des pathogènes interagissant avec eux. Toutefois, du côté du pathogène (cf. i-II.2. *Pathogènes*), aucune molécule diversifiée n'a été caractérisée jusqu'à lors. Il n'est donc pas possible de vérifier les modèles co-évolutifs théoriques dans ces interactions invertébrés / pathogènes.

Les travaux réalisés dans le cadre de cette thèse se sont attachés à étudier cette question dans une interaction hôte invertébré / parasite. Le modèle hôte / parasite expérimental choisi dans le cadre de cette étude implique le mollusque gastéropode *Biomphalaria glabrata* et le trématode parasite *Schistosoma mansoni* (cf. i-III. *Le modèle S. mansoni / B. glabrata*). Cette interaction se caractérise, en milieu naturel, par un polymorphisme de compatibilité signifiant que la dynamique co-évolutive est accessible expérimentalement. Ceci a d'ailleurs permis la sélection de souches de parasites compatibles et incompatibles vis-à-vis de la même souche de mollusque. Ainsi, des approches moléculaires comparatives entre ces souches de parasites ont été développées afin de rechercher des gènes de *S. mansoni* impliqués dans ce polymorphisme de compatibilité. Ce travail a permis (i) de mettre en évidence des molécules parasitaires clés au cœur de la dynamique co-évolutive (cf. *CHAPITRE 1*), (ii) de caractériser ces molécules et leur niveau de

polymorphisme et/ou de diversification (cf. *CHAPITRE 2*) et (iii) d'identifier les bases génétiques à l'origine de ce polymorphisme (cf. *CHAPITRE 3*).

## I. INTERACTION HÔTE VERTEBRE / PATHOGENE

Au cours des interactions hôte vertébré / pathogène, la survie de chacune des espèces dépend de sa capacité à reconnaître et à résister à l'autre. Les hôtes vertébrés possèdent des systèmes de surveillance et un arsenal de molécules défensives. De leur côté, les agents pathogènes disposent de facteurs de virulence et de stratégies diverses leur permettant d'affaiblir ou de surmonter les défenses de l'hôte.

Les hôtes vertébrés (gnathostomes et agnathes) sont capables de se protéger des pathogènes grâce à une première ligne de défense représentée par des mécanismes rassemblés sous le terme générique d'immunité innée (non adaptative) (Medzhitov, Janeway, 1997). Le système immunitaire inné comprend une série de protéines codées par des gènes présents dans la lignée germinale. Ces gènes ne sont pas remaniés et codent des récepteurs à large spectre de reconnaissance (PRR, « Pattern Recognition Receptor »). La reconnaissance de ces récepteurs porte sur des motifs structuraux invariants communs aux pathogènes ; ils sont appelés PAMP (Pathogen-Associated Molecular Pattern). Nous verrons par la suite, que cette composante innée est également présente chez les invertébrés (cf. *i-II. Interaction hôte invertébré / pathogène*). En revanche, seuls les hôtes vertébrés disposent d'une deuxième ligne de défense beaucoup plus élaborée et appelée immunité adaptative (ou spécifique). La reconnaissance se fait ici grâce à des récepteurs extrêmement diversifiés, générés par des mécanismes somatiques durant l'ontogenèse de chaque individu. Les pathogènes, quant à eux, représentent une extrême diversité d'antigènes et ont développé des stratégies de contournement de la réponse immunitaire variées. Dans cette course aux armements, des mécanismes de diversification intervenant à différents niveaux (génomique ou au cours du processus d'expression génique) ont été sélectionnés. Ils permettent de générer des niveaux élevés de diversification et/ou de polymorphisme des molécules au coeur de l'interaction tels les antigènes majeurs de surface des pathogènes ou les molécules de reconnaissance du système immunitaire adaptatif de l'hôte vertébré.

Dans cette première partie de l'introduction nous nous focaliserons sur les systèmes impliquant hôtes vertébrés (gnathostomes et agnathes) et pathogènes (microparasites bactériens, viraux ou parasites eucaryotes). Nous ferons état des éléments permettant, à l'heure actuelle, de valider les modèles théoriques de co-évolution, tel que celui de la « Reine Rouge », appliqué à ces systèmes (Combes, 2000). Dans un premier temps (cf. *i-I.1. Hôtes vertébrés : immunité adaptative, polymorphisme et diversification*), nous traiterons les caractéristiques de certains gènes de défense des hôtes à savoir les gènes codant les immunoglobulines et les TCR ainsi que les gènes du complexe majeur d'histocompatibilité

(CMH) chez les gnathostomes, ou encore les récepteurs lymphocytaires variables (VLR) chez les agnathes. Dans un deuxième temps (cf. i-1.2. *Pathogènes*), nous nous intéresserons aux stratégies développées par les pathogènes (bactéries, virus et parasites eucaryotes) dans cette course aux armements, en portant un intérêt particulier aux mécanismes de variation antigénique.

## **1. Hôtes vertébrés : immunité adaptative, polymorphisme et diversification**

Seuls les gnathostomes (vertébrés dotés d'une « mâchoire »), en plus d'une composante immunitaire innée, possèdent une composante adaptative. Elle peut reconnaître toute attaque du non soi et mettre en place des défenses spécifiques nécessitant le réarrangement d'un nombre limité de gènes spécialisés. Ces gènes impliqués codent par exemple les immunoglobulines (Ig), les récepteurs spécifiques des lymphocytes T (TCR) (cf. i-1.1.1.1 *Immunoglobulines et TCR*) ou encore les molécules du complexe majeur d'histocompatibilité (CMH) (cf. i-1.1.1.2. *Complexe Majeur d'Histocompatibilité*). Les agnathes ne disposent pas des mêmes molécules, mais ils possèdent un système apparenté, plus simple, basé sur des molécules qui assurent une fonction similaire. Il est basé sur des gènes, présents dans la lignée germinale, qui, à la suite de réarrangements somatiques, codent des récepteurs lymphocytaires variables (VLR) dont nous reparlerons (cf. i-1.1.2.2. *Les récepteurs lymphocytaires variables (VLR)*).

### **1.1. Les gnathostomes**

#### ***1.1.1. Immunoglobulines et TCR***

Les Ig de surface des lymphocytes B et les TCR des lymphocytes T sont les récepteurs des antigènes. Ces récepteurs sont extrêmement diversifiés et permettent de se lier de manière spécifique aux antigènes présents dans leur environnement. Un mécanisme génétique complexe de diversification des gènes codant les Ig/TCR a été sélectionné au cours de l'évolution. Il permet de générer un extraordinaire répertoire de protéines hautement variables (plus de  $10^{14}$  molécules différentes possibles par individu) capable de reconnaître et « prendre en charge » tous les antigènes susceptibles de se présenter à un organisme. Ces mécanismes de diversification basés sur des recombinaisons permettent l'expression de la diversité nécessaire à partir d'une information génétique relativement

limitée. Sans ces mécanismes, cette diversité ne pourrait pas être exprimée par un seul génome. Ces mécanismes de diversification interviennent pendant la différenciation et la maturation des lymphocytes. Cette étape est divisée en deux phases. La première, indépendante des antigènes, conduit à la différenciation des cellules souches lymphoïdes en lymphocytes B et T mature. La seconde, dépendante des antigènes, qui n'intéresse que les lymphocytes B, va conduire aux cellules produisant des immunoglobulines (plasmocytes) ou aux lymphocytes B mémoires.

#### a. Organisation génomique des gènes codant les Ig/TCR

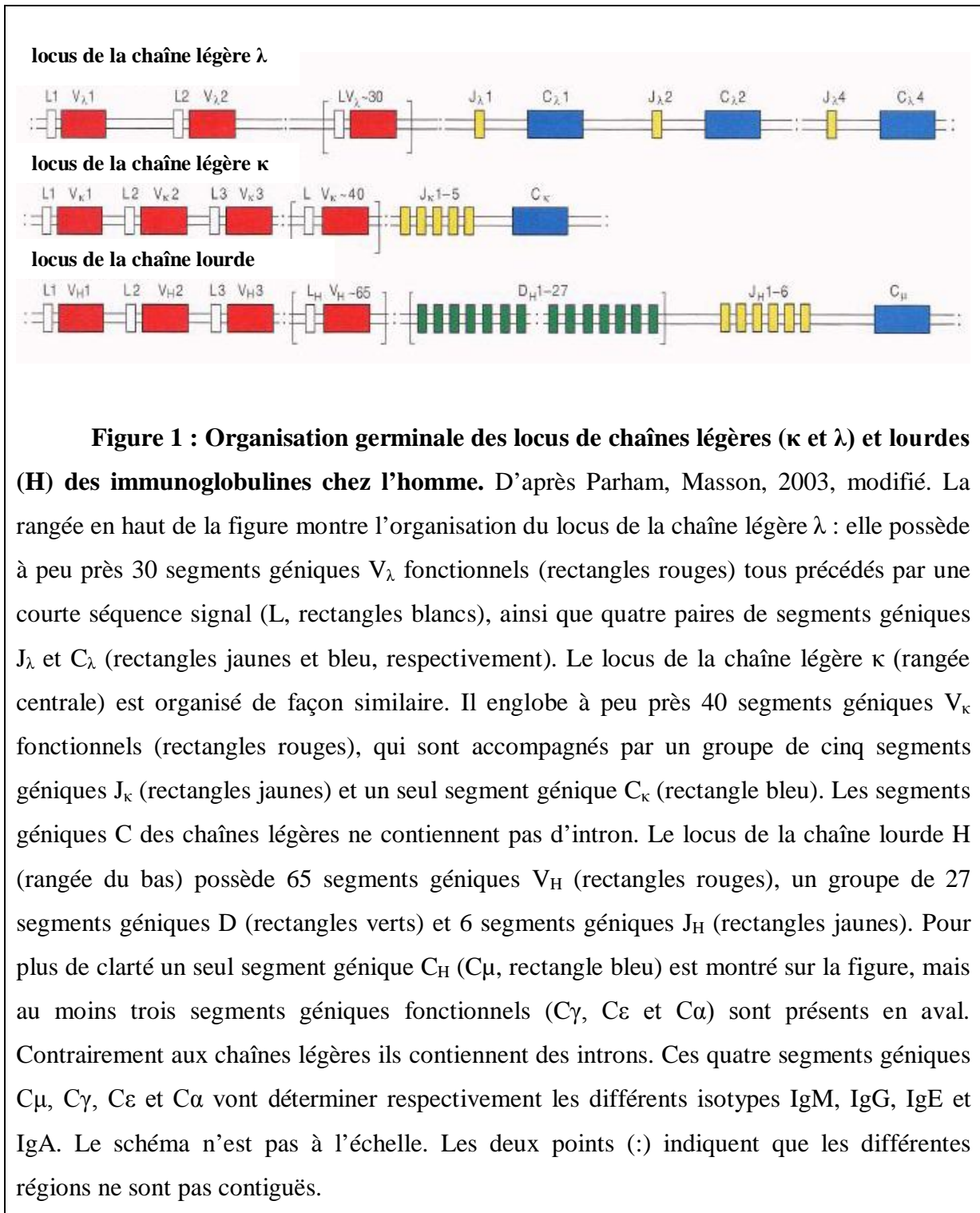
La localisation des gènes codant les Ig/TCR, ainsi que leur organisation génomique, sont à l'heure actuelle très bien caractérisées. Les gènes codant la chaîne lourde (H) des Ig sont localisés au niveau de trois *loci* (Pallares *et al.*, 1999) et ceux codant les deux chaînes légères ( $\kappa$  et  $\lambda$ ) sont localisés au niveau de deux autres *loci* (Barbie, Lefranc, 1998; Pallares *et al.*, 1998). En ce qui concerne les gènes codant les chaînes  $\alpha$ ,  $\beta$ ,  $\gamma$  et  $\delta$  des TCR, ils sont localisés au niveau de quatre *loci* (Folch, Lefranc, 2000; Lefranc, 2001; Scaviner, Lefranc, 2000). Tous ces *loci* possèdent une organisation génomique similaire (Figure 1). Ainsi, de 5' vers 3' au niveau de chaque locus, sont présents tout d'abord les segments géniques V (variable), D (diversité) et J (jonction). Ces segments sont organisés en groupes (ou clusters) et codent le domaine variable des Ig/TCR. Ensuite, viennent les segments géniques C ( $C\mu$ ,  $C\gamma$ ,  $C\epsilon$  et  $C\alpha$ ) qui codent les différents domaines constant des Ig/TCR. Seuls les *loci* des chaînes légères  $\kappa$  et  $\lambda$  et des chaînes  $\alpha$  et  $\gamma$  ne possèdent pas de segments D (Gellert, 2002). L'abréviation V(D)J sera utilisée par la suite pour désigner l'ensemble des segments.

#### b. Diversification du répertoire d'Ig/TCR durant la phase indépendante de l'antigène

L'expression des gènes codant les Ig/TCR nécessite le réarrangement préalable des segments géniques individuels en gènes fonctionnels. Ce processus s'effectue au cours de la différenciation et de la maturation des cellules de la lignée lymphoïde dans la moelle osseuse. Précisément, il s'agit d'un mécanisme de recombinaison somatique qui permet, au niveau de chaque locus, d'assembler aléatoirement un segment génique provenant de chaque cluster V, (D) et J (Figure 2). Cette recombinaison est dirigée par les séquences RSS (Recombination Signal Sequence) adjacentes à chaque segment génique (Gellert, 2002; Tonegawa, 1983). Ces RSS vont être la cible des recombinases RAG 1 et 2 (Recombination Activating Genes) qui catalysent les réarrangements V(D)J (Oettinger *et al.*, 1990). Ainsi, le



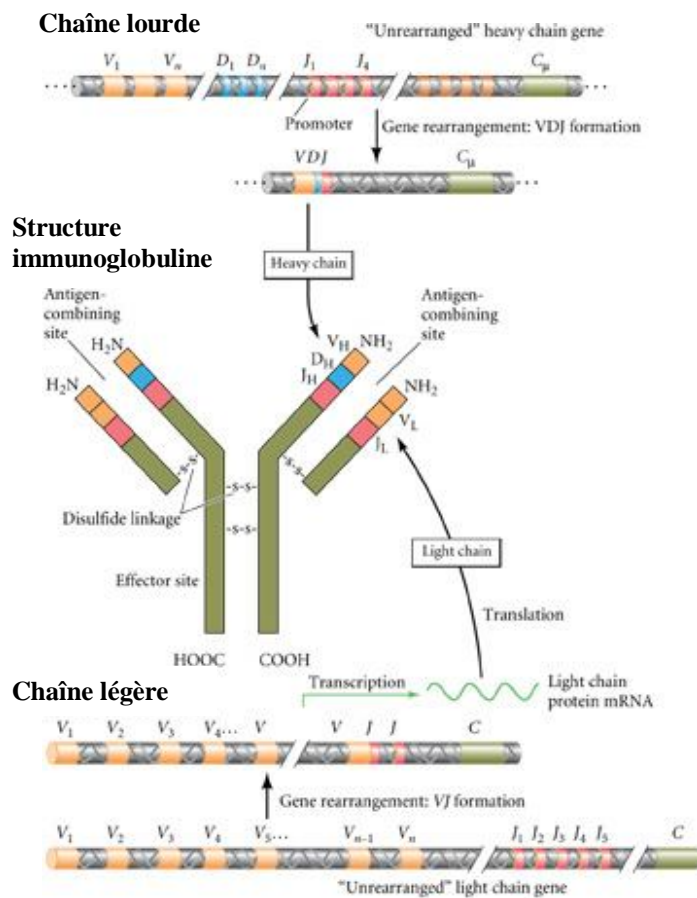
caractère aléatoire des recombinaisons et le nombre de segments géniques (Figure 1), sont à l'origine d'une partie de la diversité (diversité combinatoire) du répertoire d'Ig/TCR (Li *et al.*, 2004; Tonegawa, 1983).



Un niveau de diversité supplémentaire, la diversité de jonction, est généré au niveau de la jonction des segments géniques V, (D) et J. Il s'agit d'insertions de nucléotides qui n'étaient pas présents dans l'ADN de départ (Bogue, Roth, 1996; Tonegawa, 1983). En effet, lors de la réaction de recombinaison, RAG 1 et 2 reconnaissent et rapprochent physiquement les RSS de deux segments géniques appartenant à des clusters différents. Puis les RAG induisent une cassure double brin de l'ADN. Chaque clivage conduit à la production d'une extrémité franche au niveau des RSS et d'une structure intermédiaire en « épingle à cheveu » au niveau de la séquence codante. De manière générale, les extrémités codantes sont rapidement assemblées en jonctions codantes. Ce processus implique le clivage asymétrique des épingles à cheveux qui va favoriser l'ajout de petites séquences palindromiques (P-diversité dépendante de l'ADN) (Jackson *et al.*, 2004; Meier, Lewis, 1993). D'autre part, la désoxynucléotidyl-transférase terminale (TdT) peut ajouter, indépendamment de la matrice d'ADN, un nombre variable de nucléotides (N-diversité) à l'extrémité de la chaîne d'ADN (Gilfillan *et al.*, 1993). Le caractère aléatoire de ces insertions/délétions va cependant générer de « mauvaises » jonctions entraînant un décalage du cadre de lecture et une terminaison prématurée de la chaîne protéique. Pour remédier à cela, des recombinaisons secondaires (mécanisme de « Receptor Editing », (Casellas *et al.*, 2001)) peuvent intervenir et conduisent ainsi à un nouveau récepteur fonctionnel (Kouskoff, Nemazee, 2001). Cette variabilité de jonction présente un intérêt d'autant plus important en terme de reconnaissance antigénique puisqu'elle se situe au niveau de la séquence codant le site de combinaison antigénique. Ainsi, cette diversité de jonction amplifie considérablement la diversité combinatoire. Elle joue un rôle important dans la diversité structurale de la région hypervariable des Ig/TCR, dont le rôle est prépondérant dans la reconnaissance des antigènes.

Enfin, les chaînes des Ig/TCR, une fois synthétisées, doivent s'associer afin de former un récepteur fonctionnel. En effet, les Ig sont formées de l'association de deux chaînes lourdes identiques et de deux chaînes légères identiques (Figure 2). Les TCR résultent quant à eux de l'association soit des chaînes  $\alpha$  et  $\beta$  soit  $\gamma$  et  $\delta$  (Gellert, 2002). Ainsi, les différentes combinaisons possibles d'association vont fortement augmenter la diversité des récepteurs puisqu'ils définissent ensemble les sites de combinaison antigéniques.

**Figure 2 : Réarrangement des segments géniques au niveau des chaînes lourdes et légères des immunoglobulines durant la différenciation des lymphocytes B. D'après**



Scott, 2006, modifié. Les segments géniques V, D et J (rectangles oranges, bleu et rouges, respectivement) codant la région variable de la chaîne lourde (en haut de la figure) sont organisés en clusters. Dans un premier temps, ce sont les segments géniques D et J qui recombinent. Ensuite le segment génique V recombine avec la séquence combinée DJ, formant ainsi un exon complet codant la région variable de la chaîne lourde. Les segments géniques VDJ combinés sont transcrits en ARN pré-messager avec les segments géniques C présents en aval (rectangle vert). Les

séquences correspondant aux introns et éventuellement aux segments géniques J superflus sont ensuite excisées par épissage de façon à produire un ARN messager mature. Dans un premier temps, ce messager fonctionnel codera la chaîne lourde de l'isotype  $\mu$  (IgM) car un codon STOP est présent au niveau du premier segment  $C_\mu$ . Les segments géniques V et J (rectangles oranges et rouges, respectivement) codant la région variable de la chaîne légère (en bas de la figure) sont organisés en clusters. Un segment génique V recombine avec un segment génique J pour former un exon complet codant la région variable de la chaîne légère. Après réarrangement, le gène de la chaîne légère est composé de trois exons qui codent la séquence signal, la région variable et la région constante. Un messager fonctionnel pour le codage de la chaîne légère est produit après la transcription en ARN pré-messager et son épissage en vue d'enlever les introns. La structure classique d'une immunoglobuline est représentée au centre. Deux chaînes lourdes identiques et deux chaînes légères identiques sont reliées par des ponts disulfures. Le site de liaison à l'antigène est composé des régions variables des chaînes lourdes et légères, alors que le site effecteur est déterminé par la séquence en acides aminés de la région constante de la chaîne lourde. Le schéma n'est pas à l'échelle. Le // indique que les régions ne sont pas contiguës.

### c. Diversification du répertoire d'Ig durant la phase dépendante de l'antigène

Les processus décrits ci-dessus, dans le cas des gènes codant les Ig, se déroulent dans la moelle osseuse pendant la phase « antigène indépendante » de la lymphopoïèse B. Ils sont à l'origine, en partie, de la diversité du répertoire d'anticorps matures. En effet, des mécanismes somatiques générateurs de diversité supplémentaires vont intervenir pendant la phase que je qualifierai d'« antigène dépendante ». Les lymphocytes B matures issus de la phase « antigène indépendante » transitent par les voies sanguine et lymphatique et colonisent les follicules primaires des organes lymphoïdes secondaires. Quand ces cellules rencontrent l'antigène par l'intermédiaire de l'immunoglobuline membranaire (IgM), elles sont alors activées. Ainsi, les cellules B prolifèrent rapidement et arrivent à engendrer un centre germinatif formant un follicule secondaire. C'est au niveau de ce centre germinatif que vont avoir lieu les mécanismes à la base de ce second niveau de diversification qui va conduire à une grande affinité des sites de liaisons pour l'antigène (maturation des anticorps).

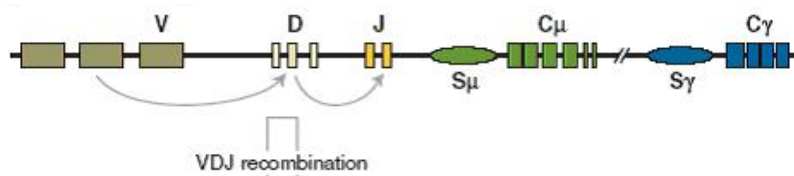
Chez l'homme et la souris, les mutations à hautes fréquences (hypermutations somatiques, HMS) constituent le mécanisme prédominant (Weill, Reynaud, 1996) (Figure 3). Ces HMS introduisent progressivement des substitutions nucléotidiques au niveau des régions codant les domaines variables et constants des chaînes lourdes et légères. Ces HMS ont une fréquence d'environ  $10^{-3}$  par paire de base par génération (Di Noia, Neuberger, 2007; Li *et al.*, 2004; Maizels, 2005). Ce processus fait intervenir la protéine AID (Activating Induced cytidine Deaminase) (Muramatsu *et al.*, 1999) qui est exclusivement exprimée dans le centre germinatif. L'AID conduit à la déamination d'une désoxycytidine en désoxyuridine générant un mésappariement U:G. Il est alors pris en charge par le système de réparation de l'ADN (« mismatch repair system ») qui va induire un gap à ce niveau là. Ces gaps vont alors être réparés par des ADN polymérases de faible fidélité (Goodman, 2002) générant ainsi des mutations ponctuelles. Le processus d'HMS est un mécanisme majeur de diversification du répertoire des lymphocytes B. Il est à la base de l'amélioration de l'affinité des anticorps lors de la réponse aux antigènes dépendants des lymphocytes T.

Chez la poule et le lapin, la conversion génique (échange non réciproque de séquences homologues ou similaires (Palmer, Brayton, 2007)) est le mécanisme prédominant (Selsing *et al.*, 1996; Weill, Reynaud, 1996). Chez la poule, le locus de la chaîne lourde (H) contient un segment de gène  $V_H$  fonctionnel, alors que chez le lapin ce locus en contient une centaine. Dans les deux cas, une centaine de gènes  $psi-V_H$  ( $\psi V_H$ ) non fonctionnels est également présente au niveau de ce locus. Ainsi, des ensembles  $V_H$ -D- $J_H$  combinés vont

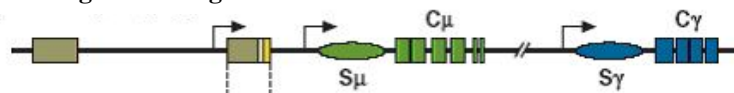
subir de nombreuses conversions géniques avec un pool de gènes  $\psi V_H$  utilisé comme donneur (Reynaud *et al.*, 1994). Les similitudes de séquences entre les gènes donneurs  $\psi V_H$  et le gène accepteur  $V_H$  vont favoriser la succession de conversions géniques. Ces processus sont souvent accompagnés de mutations ponctuelles de part et d'autre des remplacements (Weill, Reynaud, 1995). Dans ce cas encore, la conversion génique est induite à la suite d'une lésion U:G médiée par l'AID (Di Noia, Neuberger, 2007; Maizels, 2005). Bien sûr, le locus de la chaîne variable est soumis, lui aussi, à ce mécanisme de conversion génique (Reynaud *et al.*, 1987; Thompson, Neiman, 1987).

**Figure 3 : Hypermutations somatiques (HMS).** D'après Di Noia, Neuberger, 2007, modifié.

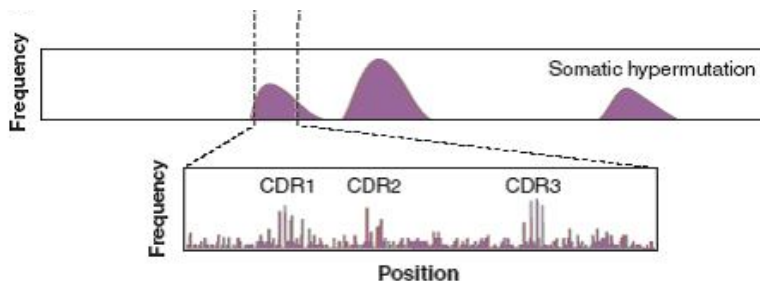
**(i) Organisation germinale du locus IgH**



**(ii) Gène IgH réarrangé**



**(iii) Distribution mutation**



Représentation schématique du processus d'HMS au niveau des gènes codant la chaîne lourde (H) des immunoglobulines (Ig). Le schéma n'est pas à l'échelle. Le double slash indique que les régions C $\mu$  et C $\gamma$  ne sont pas contiguës. Les flèches courbées représentent des événements de réarrangement génique ; les flèches droites

indiquent la position des promoteurs. **(i)** Organisation germinale d'un locus de la chaîne lourde des Ig (IgH). Recombinaison des segments géniques V(D)J. **(ii)** Gène codant une IgH réarrangée soumis aux hypermutations somatiques **(iii)** Les régions où les hypermutations somatiques s'accumulent sont indiquées par des courbes qui reflètent la fréquence des mutations. La région V(D)J est élargie sous forme d'histogramme pour illustrer la distribution non aléatoire des mutations. Les CDR (Complementary-Determining Region) constituent les domaines impliqués dans le contact avec l'antigène. Abréviations : C $\mu$ , segment génique constant  $\mu$  ; C $\gamma$ , segment génique constant  $\gamma$  ; D, segments géniques de diversité ; J, segments géniques de jonction ; S, séquence S ; V, segments géniques variables.

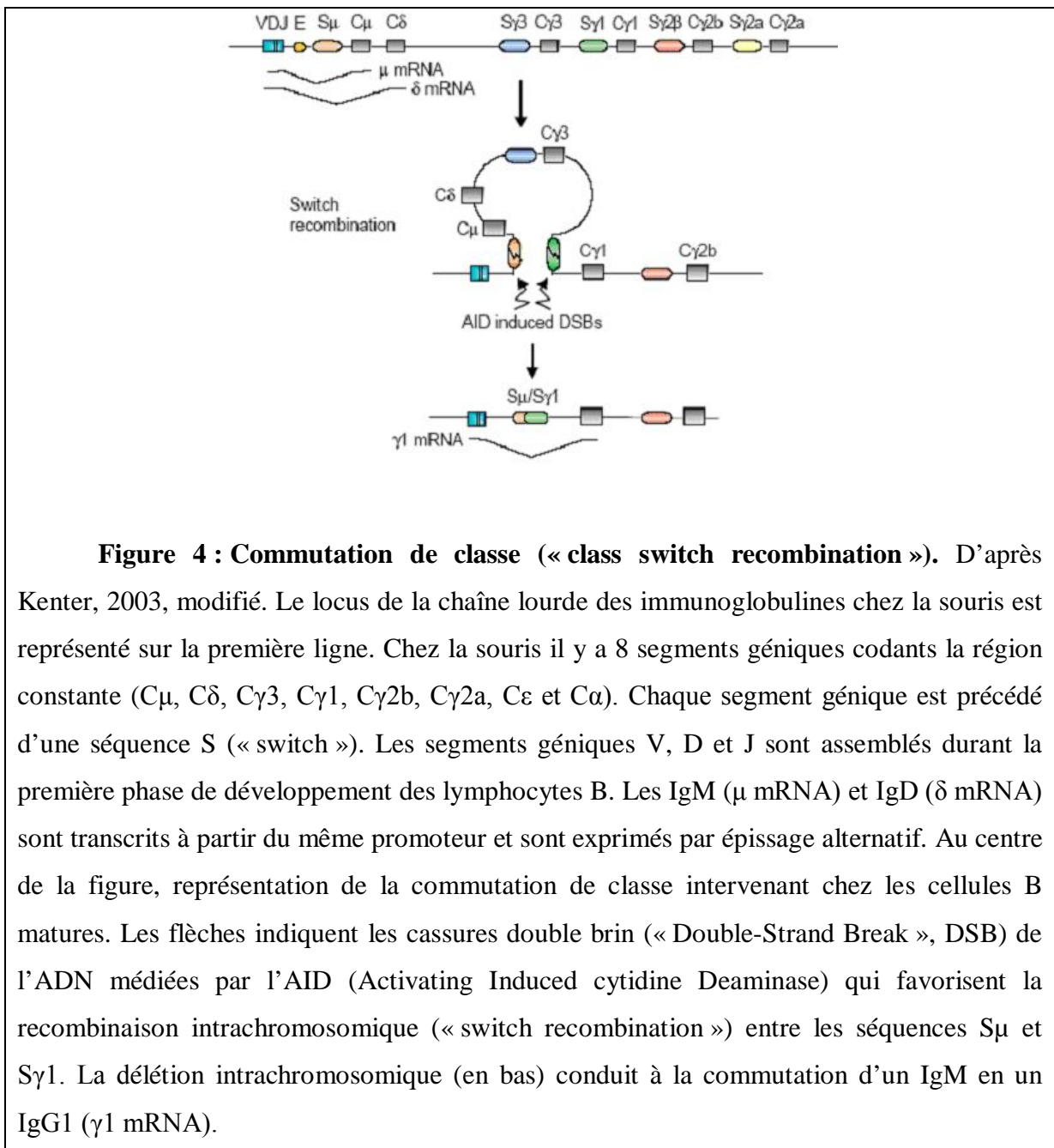
La spécificité antigénique des Ig est déterminée par les régions variables des chaînes lourdes et des chaînes légères. Les fonctions effectrices, quant à elles, dépendent des régions constantes des chaînes lourdes et varient selon les isotypes. Au cours de l'immunisation, les anticorps synthétisés par les lymphocytes B matures sont tout d'abord de type IgM, puis d'autres isotypes d'Ig (IgG ou IgE ou encore IgA) seront exprimés. Ce changement d'isotype s'appelle commutation de classe (ou « class switch recombination ») et s'opère sans changement de spécificité des clones concernés (Figure 4). Ce phénomène, qui affecte les chaînes lourdes, rapproche le segment VDJ réarrangé d'un nouveau segment génique C de classe différente ( $C\gamma$ ,  $C\epsilon$  ou  $C\alpha$ ) par un processus de recombinaison somatique. Ces commutations sont rendues possibles par la présence en 5' de chacun des segments géniques C de séquences particulières (séquences S pour « Switch »). En effet, les séquences S du segment  $C\mu$  et de l'un des 3 autres segments C vont être la cible de l'AID. Les lésions U:G vont induire des cassures double brin de l'ADN qui pourraient alors favoriser un phénomène de recombinaison intrachromosomique entre les deux séquences S. Ce processus conduit à la délétion de la région située entre les deux séquences S et au rapprochement de l'un des trois segments  $C\gamma$ ,  $C\epsilon$  ou  $C\alpha$ . (Kenter, 2003; Li *et al.*, 2004). Ainsi, ce processus de commutation de classe génère une diversité de fonction sans augmenter la diversité du répertoire d'Ig.

#### d. Diversité structurale et spécificité multiple des Ig

Nous venons de voir que la diversité du répertoire des anticorps trouve son origine au travers du réarrangement d'un nombre fini de segments de gènes, associé à la variabilité jouant au niveau de la jonction de ces différents segments, à la combinatoire chaîne lourde / chaîne légère et enfin aux hypermutations somatiques ponctuelles ou aux phénomènes de conversions géniques. Toutefois, des travaux récents (Foote, 2003; James *et al.*, 2003) établissent qu'un niveau supplémentaire de variabilité du répertoire des anticorps existe : en effet, ils démontrent qu'à une même structure primaire peut correspondre plusieurs conformations structurales, chacune étant susceptible de reconnaître un antigène spécifique. Cette démonstration repose sur l'analyse structurale de l'anticorps SPE7 (immunoglobuline anti-haptène de type E) seul ou sous la forme de complexes anticorps-antigène.

Le premier résultat révèle que l'anticorps non complexé est capable d'adopter deux conformations différentes identifiées Ab1 et Ab2. L'organisation spatiale des sites de liaisons (ou paratopes) des deux isoformes est tout à fait différente. Dans le cas de Ab1, le paratope est caractéristique d'anticorps qui lient des peptides ou protéines, alors que dans le cas de Ab2, il est caractéristique d'anticorps spécifiques d'haptène. Le second résultat

montre que lorsque l'anticorps est complexé à 3 haptènes différentes, il adopte la conformation Ab2. Cependant, une nouvelle conformation Ab3 peut être également observée. Ainsi, ces réarrangements structuraux induits permettent à l'anticorps SEP7 de s'adapter aux caractéristiques structurales des trois haptènes qu'il reconnaît. Le dernier résultat révèle que lorsque l'anticorps est complexé à une protéine, il adopte la conformation Ab1 attendue. Cependant, une nouvelle conformation Ab4 a pu être observée. Ce sont donc au total six structures différentes de l'anticorps SEP7 qui ont été caractérisées dans cette étude.



**Figure 4 : Commutation de classe (« class switch recombination »).** D'après Kenter, 2003, modifié. Le locus de la chaîne lourde des immunoglobulines chez la souris est représenté sur la première ligne. Chez la souris il y a 8 segments géniques codants la région constante (C $\mu$ , C $\delta$ , C $\gamma_3$ , C $\gamma_1$ , C $\gamma_{2b}$ , C $\gamma_{2a}$ , C $\epsilon$  et C $\alpha$ ). Chaque segment génique est précédé d'une séquence S (« switch »). Les segments géniques V, D et J sont assemblés durant la première phase de développement des lymphocytes B. Les IgM ( $\mu$  mRNA) et IgD ( $\delta$  mRNA) sont transcrits à partir du même promoteur et sont exprimés par épissage alternatif. Au centre de la figure, représentation de la commutation de classe intervenant chez les cellules B matures. Les flèches indiquent les cassures double brin (« Double-Strand Break », DSB) de l'ADN médiées par l'AID (Activating Induced cytidine Deaminase) qui favorisent la recombinaison intrachromosomique (« switch recombination ») entre les séquences S $\mu$  et S $\gamma_1$ . La délétion intrachromosomique (en bas) conduit à la commutation d'un IgM en un IgG1 ( $\gamma_1$  mRNA).

Ces travaux identifient clairement un niveau supplémentaire de diversité du répertoire des immunoglobulines fondé sur la co-existence de plusieurs structures protéiques différentes. Ainsi, un anticorps unique a la capacité d'interagir avec des antigènes totalement différents *via* des sites distincts. Il est désormais nécessaire de savoir si cela est également le cas pour l'ensemble des anticorps quelque soit leur spécificité. Le système immunitaire disposerait ainsi d'un moyen d'assurer plus efficacement la neutralisation et l'élimination des molécules reconnues. Ce qui est certain, c'est qu'il s'agit là d'une découverte importante, dont on commence à peine à entrevoir les conséquences.

### ***1.1.2. Complexe Majeur d'Histocompatibilité***

Parmi les membres de la superfamille des Immunoglobulines, il faut également compter les gènes du complexe majeur d'histocompatibilité (CMH). Ces gènes, jusqu'alors retrouvés chez tous les gnathostomes, codent des récepteurs clés qui reconnaissent et lient des peptides étrangers. Ils les transportent ensuite à la surface des cellules présentatrices des antigènes et les présentent aux cellules immunitaires spécialisées (lymphocytes T) qui induiront par la suite une réponse immune spécifique. Bien qu'il y ait une conservation des fonctions générales du CMH dans la reconnaissance immune, l'architecture génomique peut être cependant très différente entre espèces (Kelley *et al.*, 2005). Chez l'homme, plus de 260 gènes sont répertoriés et localisés sur le bras court du chromosomes 6p21.3 (Consortium, 1999; Kelley *et al.*, 2005). Ces gènes sont regroupés au niveau de trois régions codant respectivement trois types de protéines, les protéines de classe I, II et III (Dunham *et al.*, 1987; Horton *et al.*, 2004).

Deux propriétés distinctes du CMH lui permettent d'augmenter son potentiel de reconnaissance des peptides étrangers. Premièrement, le CMH est polygénique : il contient plusieurs gènes différents, organisés en familles multigéniques, de sorte qu'un individu possède une batterie de molécules du CMH avec différentes gammes de spécificité. Deuxièmement, le CMH est polymorphe ; c'est à dire qu'il y a de nombreux variants de chaque gène dans l'ensemble de la population (diversité allélique) ; les gènes du CMH sont d'ailleurs les gènes les plus polymorphes connus actuellement (Trowsdale, Parham, 2004).

#### ***a. Polygénie du CMH***

Dans cette course aux armements, un élément important, est de disposer d'un répertoire génique le plus large possible, afin de conférer plus de diversité à chaque individu.



Une des stratégies mise en place consiste à dupliquer les gènes du CMH. En effet, chaque évènement de duplication permet d'obtenir un gène paralogue au gène copié. D'après le « modèle de fonctionnalisation », le devenir d'un gène paralogue se solde par une des trois issues suivantes : (i) il dégénère sans acquérir de fonction (non fonctionnalisation ou pseudogénéisation), (ii) dans le cas de gènes à fonctions multiples, il perd réciproquement avec le gène copié certaines fonctions (sub-fonctionnalisation), (iii) il acquiert de nouvelles fonctions (néo-fonctionnalisation) (Mazet, Shimeld, 2002). Les gènes sont souvent dupliqués *in cis* et sont ainsi positionnés les uns à côté des autres, en clusters. Le fait de disposer de plusieurs gènes de fonctions similaires, pouvant être exprimés simultanément chez un individu, représente un avantage individuel mais également populationnel dans l'affrontement contre les pathogènes. Cependant, les pathogènes ont des temps de génération beaucoup plus rapide que ceux des vertébrés gnathostomes, leur permettant de développer rapidement des adaptations afin d'éviter le système immunitaire. Ainsi, les gènes du CMH évoluent selon le modèle de « birth and death » (Flajnik, Kasahara, 2001). Ils sont constamment et fréquemment renouvelés par duplications. Quant aux gènes qui n'ont plus leur utilité ils deviennent des pseudogènes non fonctionnels.

#### *b. Polymorphisme du CMH*

La polygénie n'est cependant pas suffisante à elle seule pour contrer la flexibilité des pathogènes. En effet, le nombre de gènes du CMH est fini et ne peut être augmenté à l'infini. Ainsi, disposer d'un répertoire génique le plus grand possible est une chose, mais disposer d'un polymorphisme au niveau de chaque élément du répertoire est un avantage populationnel indéniable. Cela se vérifie au niveau du CMH par un polymorphisme extraordinaire au niveau des *loci* exprimés. Il présente ainsi les gènes fonctionnels les plus variables décrits chez les vertébrés (Robinson *et al.*, 2003). Pour exemple, trois des *loci* les plus variables du CMH chez l'homme, HLA-A, HLA-B et HLA-DRB1, présentent respectivement 243, 499 et 321 allèles. La plupart des variations alléliques affectent les sillons présentoirs de l'antigène des molécules du CMH de classe I et II. Cette diversité nucléotidique est plus de deux fois supérieure à la moyenne génomique (Gaudieri *et al.*, 2000). Cela semble être une caractéristique des gènes du CMH comme le révèlent de nombreuses études réalisées dans d'autres taxons (Kelley *et al.*, 2005). Le polymorphisme allélique n'est pas seulement restreint aux mutations ponctuelles isolées (« Single Nucleotide Polymorphisms », SNP). Les échanges génétiques jouent également un rôle important dans la diversification. Par exemple, les gènes de classe I semblent échanger de courtes séquences

(10-15 nucléotides), source de nouveaux allèles (Parham *et al.*, 1995). Les pseudogènes, bien qu'ils ne soient plus exprimés, peuvent servir de réservoir de séquences qui seraient réintégrées par des phénomènes de conversions géniques et participer ainsi au polymorphisme des gènes (Doxiadis *et al.*, 2006).

### **1.1.3. Conclusion**

Le haut niveau de polymorphisme allélique, superposé à la polygénie des gènes du CMH, vont permettre à chaque individu de posséder une variété de récepteurs des antigènes. Ainsi, plusieurs combinaisons différentes par cellules seront possibles. Ces deux caractéristiques vont également permettre à la population de disposer d'une grande diversité de récepteurs, puisque chaque individu est différent. Cela permet ainsi d'augmenter la « résistance de la population » face aux nouveaux pathogènes sans augmenter à l'infinie le nombre de gènes. Ici la stratégie mise en place est donc complètement différente de celle concernant les Ig/TCR. En effet, le polymorphisme et non la diversification prédomine, favorisant une stratégie populationnelle plutôt qu'individuelle de « sauvegarde » des espèces considérées.

## **1.2. Les agnathes**

### **1.2.1. Un type différent de réponse adaptative**

Les vertébrés agnathes (vertébrés sans « mâchoire »), dont les seuls représentant aujourd'hui sont les lamproies et les myxines, ne présentent pas de système immunitaire ayant les caractéristiques du système immunitaire adaptatif. Ils possèdent la voie alternative du complément (indépendante des anticorps), mais pas la voie classique (dépendante des anticorps). Ils ne possèdent pas de tissus immunitaires majeurs comme le thymus, la rate ou le système lymphatique. Enfin, aucune des molécules du système immunitaire adaptatif (CMH, Ig, TCR, RAG, TdT) n'a été détectée chez ces organismes jusqu'à présent. Cependant, ils sont capables de développer une réponse immunitaire de type adaptative basée sur des molécules différentes. En effet, en réponse aux immunisations, les lamproies et les myxines produisent des agglutinines circulantes spécifiques qui s'apparentent aux anticorps d'un point de vue fonctionnel (Hagen *et al.*, 1985; Litman *et al.*, 1970; Raison, Hildemann, 1984; Raison *et al.*, 1994). Par ailleurs, ces espèces sont sujettes au rejet d'allogreffes (Hildemann, 1970; Perey *et al.*, 1968). Des cellules qui ressemblent

morphologiquement aux lymphocytes des vertébrés gnathostomes sont responsables de cette réponse (Gilbertson *et al.*, 1986; Mayer *et al.*, 2002; Shintani *et al.*, 2000; Uinuk-Ool *et al.*, 2002; Zapata *et al.*, 1984). Comme chez les mammifères, ces « lymphocytes-like » sont plus sensibles aux irradiations que d'autres cellules sanguines, s'agrègent et prolifèrent en réponse aux stimulations antigéniques et expriment des facteurs de transcription tels que ceux impliqués dans la différenciation des lymphocytes de mammifère (Shintani *et al.*, 2000).

### ***1.2.2. Les récepteurs lymphocytaires variables (VLR)***

Chez la lamproie, l'analyse du transcriptome des lymphocytes après immunostimulation a mis en évidence des transcrits codant des protéines hypervariables (Pancer *et al.*, 2004). La présence d'un nombre variable d'unités répétées riches en leucine différentes (« Leucine-Rich Repeat », LRR), constitue leur caractéristique principale. Cette caractéristique est également à l'origine de leur dénomination : récepteurs lymphocytaires variables (« Variable Lymphocyte Receptor », VLR). Les précurseurs des VLR possèdent tous la même structure primaire de 5' en 3' : un peptide signal conservé, un module LRR N-terminal (LRRNT), la région hypervariable composée d'un nombre variable de différents modules LRR, un peptide de connexion (CP pour « Connexion Peptide ») suivi d'un module LRR C-terminal (LRRCT) et enfin une région C-terminale conservée composée d'un domaine riche en thréonine/proline, une ancre glycosyl-phosphatidyl-inositol (GPI) et une queue hydrophobe. Un seul type de VLR est exposé, via cette ancre GPI, à la surface des lymphocytes chez des individus immunostimulés (Pancer *et al.*, 2004). Ainsi, un lymphocyte donné exprimera à sa surface un seul type de VLR. De plus, le polymorphisme des VLR s'exprime au niveau du nombre de module LRR de la région hypervariable mais également au niveau de la diversité de séquence de chaque module LRR. Toutefois, les 7 premiers acides aminés du module LRRNT et les 20 derniers acides aminés du module LRRCT sont presque toujours invariables.

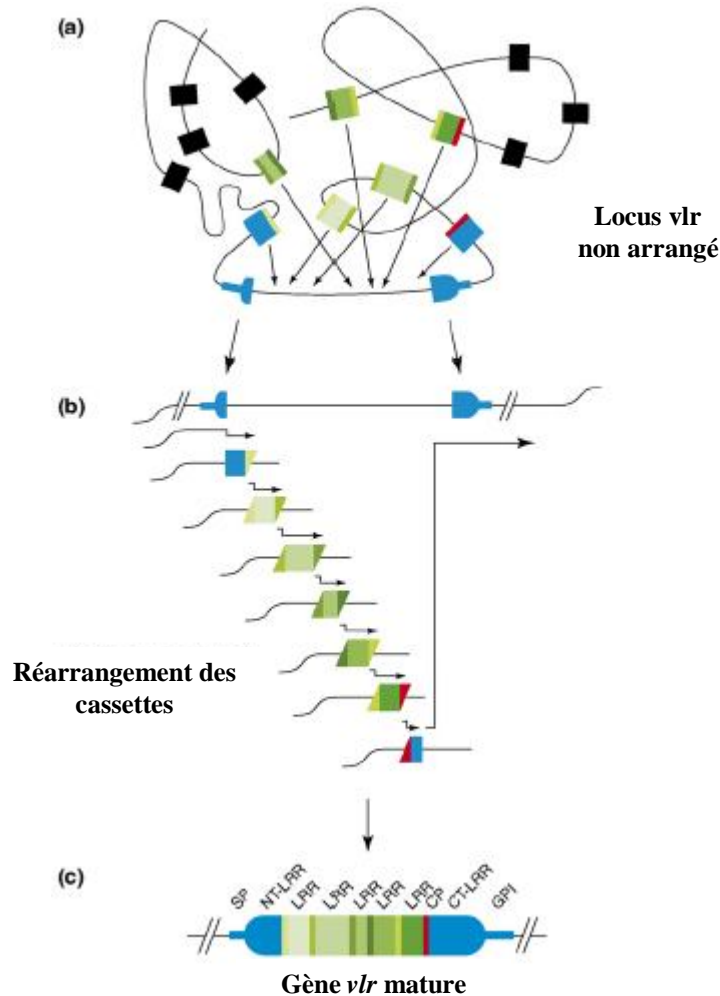
Les VLR chez les lamproies et les myxines sont codés respectivement par un et deux loci. Ces gènes sont présents dans la lignée germinale sous la forme de gènes *vlr-g* incomplets, et nécessitent l'intervention d'un mécanisme atypique afin de générer un gène *vlr* mature (Figure 5a). Les gènes *vlr-g* possèdent uniquement les exons codant les régions N-terminales et C-terminales des VLR, séparés par des régions non codantes (Pancer *et al.*, 2004). D'autre part, de nombreuses cassettes, codant des modules LRR différents, sont

organisées en clusters en amont et en aval des *vlr-g* incomplets. Durant le développement des lymphocytes, un processus complexe de réarrangement de l'ADN va permettre l'insertion de multiples cassettes au sein des *vlr-g* incomplets (Figure 5b). Ainsi, chaque lymphocyte assemble un *vlr* mature qui code un type de récepteur possédant une région hypervariable unique (Alder *et al.*, 2005) (Figure 5c).

Ces réarrangements somatiques chez les lymphocytes, à la base de la diversité des VLR, semblent faire intervenir des mécanismes de type conversion génique (Rogozin *et al.*, 2007). Ce processus serait similaire à celui identifié chez le poulet dans le cas des réarrangements des gènes codant les immunoglobulines (Reynaud *et al.*, 1987; Thompson, Neiman, 1987) (cf. *i-I.I.I.I.c. Diversification du répertoire d'Ig durant la phase dépendante de l'antigène*). En effet, différents types de séquences homologues, de 6 à 50 nucléotides, sont présents au niveau des *vlr-g* incomplets et des différentes cassettes codant les modules LRR (Figure 5b). Dans un premier temps, des séquences homologues spécifiques vont permettre la recombinaison entre les séquences codant le module LRRNT et le module LRR1. Le même scénario se déroule entre les séquences codant le module LRRCT et le peptide de connexion. Ensuite, une succession de conversion génique, guidée par les séquences homologues présentes au niveau des extrémités 5' et 3' des différentes cassettes LRR, va conduire à l'insertion de multiples cassettes (Rogozin *et al.*, 2007). De cette façon, un *vlr* mature unique est assemblé par chaque lymphocyte (Figure 5c). De plus, des gènes codant deux AID ont été identifiés chez la lamproie. Cela laisse suggérer que, comme chez le poulet, la conversion génique serait induite à la suite d'une lésion U:G médiée par l'AID au niveau des séquences homologues (Di Noia, Neuberger, 2007; Maizels, 2005).

### **1.2.3. Conclusion**

L'insertion différentielle de modules de LRR différentes dans un *vlr-g* incomplet semble donc être à l'origine d'un répertoire étendu de VLR dans la population lymphocytaire. Ce répertoire est estimé au minimum entre  $10^{14}$  et  $10^{17}$  récepteurs différents, comparable au répertoire d'anticorps des mammifères estimé à  $10^{14}$  (Alder *et al.*, 2005). Ce répertoire potentiel de VLR, bien que limité par le nombre de lymphocyte chez la lamproie ( $10^6$ ), serait toutefois suffisant pour reconnaître un large spectre d'antigène.



**Figure 5 : Représentation schématique de l'assemblage d'un gène *vlr* (« variable lymphocyte receptor ») mature chez la lamproie.** D'après Litman *et al.*, 2007, modifié. (a) Les gènes sont présents dans la lignée germinale sous la forme de gènes *vlr-g* incomplets. Ils possèdent uniquement les exons codant les régions N-terminale et C-terminale des VLR (en bleu), séparés par des régions non codantes. D'autre part, de nombreuses cassettes, codant des modules LRR (« Leucine-Rich Repeat ») différents (en dégradé de vert et en noir), sont organisées en clusters en amont et en aval des *vlr-g* incomplets. (b) Durant le développement des lymphocytes, un mécanisme atypique de réarrangement de l'ADN va permettre l'insertion de multiples cassettes (en dégradé de vert) au sein des *vlr-g* incomplets. Ce processus fait intervenir des séquences homologues (schématisées par une extrémité colorée, en diagonale), de 6 à 50 nucléotides, présentes au niveau des *vlr-g* incomplets et des différentes cassettes codant les modules LRR. (c) Ainsi, chaque lymphocyte assemble un *vlr* mature qui code un type de récepteur possédant une région hypervariable unique tant en séquence qu'en taille. NT-LRR, module LRR N-terminal (bleu); CT-LRR, module LRR C-terminal (bleu); CP, peptide de connexion (barre rouge); SP, peptide signal; GPI, ancre glycosyl-phosphatidyl-inositol.

## 2. Pathogènes

Dans cette course aux armements, face aux mécanismes de défense très sophistiqués des hôtes vertébrés (polymorphisme et/ou diversification des molécules de reconnaissance immunitaires décrit ci-dessus), les pathogènes ont développé un panel de stratégies complexes et efficaces leur permettant d'affaiblir ou de surmonter les défenses de l'hôte. Nous ferons état, dans ce chapitre, des moyens mis en oeuvre tout d'abord par les bactéries pathogènes et les virus (cf. i-I.2.1. *Bactéries pathogènes et virus*), et ensuite par les parasites eucaryotes (cf. i-I.2.2. *Parasites eucaryotes*). Parmi ce large répertoire de stratégies, nous privilégierons un mécanisme classique d'évasion, utilisé par bon nombre de parasites eucaryotes (qui vont nous intéresser dans le cadre de mon étude), mais aussi par certaines bactéries ou virus pathogènes, à savoir la variation antigénique. Nous porterons un intérêt particulier aux mécanismes génétiques permettant de générer des niveaux élevés de diversification et/ou de polymorphisme de ces antigènes de surface.

### 2.1. Bactéries pathogènes et virus

Les stratégies de virulence développées par les bactéries pathogènes et les virus sont nombreuses et très variées. Elles peuvent se matérialiser soit par une immunosuppression systémique, qui inhibe ou affaiblit les défenses immunitaires soit par une évasion immune, qui empêche la mise en place d'une réponse immunitaire efficace.

L'immunosuppression systémique passe par des mécanismes très diversifiés permettant de bloquer le développement de l'immunité cellulaire et acquise, d'inhiber le système du complément, de moduler la réponse inflammatoire ou l'apoptose, de contrer ou de tuer les cellules immunitaires (lymphocytes, macrophages), ou encore d'interférer avec les récepteurs « Toll Like Receptor (TLR) et « Nucleotide Oligomerization Domain » (NOD), d'altérer les voies de transduction des signaux intracellulaires de différentes cellules effectrices, d'empêcher l'action des cytokines et autres chemokines ou bien de neutraliser des molécules antimicrobiennes (pour revue : (Finlay, Mcfadden, 2006)).

Les stratégies passives de contournement, quant à elles, passent par des mécanismes de mimétisme ou camouflage. Par exemple, certains virus vont présenter à leur surface des molécules de l'hôte mobilisées lors du bourgeonnement. Chez les bactéries, l'altération de leurs propres motifs de surface peuvent leur permettre également d'échapper au système de reconnaissance (pour revue voir: (Finlay, Mcfadden, 2006)).

Un autre mécanisme classique d'évasion immune, utilisé aussi bien par les virus que les bactéries pathogènes, consiste à faire varier leurs molécules immunodominantes de surface. Cette variation phénotypique va garantir au parasite d'éviter le système immunitaire lors d'une infection chronique ou récurrente. Nous illustrerons cette variation antigénique chez des bactéries pathogènes du genre *Neisseria* ou encore *Borrelia* (cf. i-1.2.1.1 *Bactéries pathogènes*) ainsi que chez les virus de l'Influenza et de l'immunodéficience humaine (cf. i-1.2.1.2 *Virus*).

### **2.1.1. Bactéries pathogènes**

Chez les bactéries, une conséquence de la pression de sélection, imposée par la composante innée (TLR, NOD) et la composante adaptative (anticorps, CMH), est leur capacité à faire varier leurs composants de surface (pili, flagelles, LPS, peptidoglycanes et autres protéines de la membrane externe). Cette variation phénotypique est médiée par la combinaison des mécanismes de variation antigénique et de variation de phase.

La variation antigénique correspond à l'expression exclusive de protéines fonctionnelles, antigéniquement distinctes et conservées dans une population clonale. L'information génétique nécessaire à l'expression de la famille de variants antigéniques est disponible dans la cellule, mais seulement un variant sera exprimé à la fois.

La variation de phase, quant à elle, consiste à exprimer, ou au contraire à ne pas exprimer un composant antigénique de surface (changement réversible de l'expression de type tout ou rien). Cela va se traduire par une variation du niveau d'expression d'une ou plusieurs protéines entre des cellules d'une même population clonale (Van Der Woude, Baumberg, 2004).

De plus, ces molécules antigéniques de surface, soumises à ces mécanismes de variation antigénique et/ou de phase, présentent généralement une des trois caractéristiques suivantes : (i) elles sont codées par de multiples et différentes copies d'un gène, chacune des copies pouvant être exprimée ou non indépendamment l'une de l'autre ; (ii) elles possèdent un seul locus d'expression et de nombreuses copies silencieuses qui seront utilisées pour remplacer constamment le gène exprimé ; ou (iii) elles possèdent une région extrêmement variable pouvant changer constamment.

Les espèces pathogènes de l'homme du genre *Neisseria* (coque à gram négatif) sont responsables de la gonorrhée (*N. gonorrhoeae*) et de la méningite (*N. meningitidis*). Ces

modèles sont très intéressants dans la mesure où les trois caractéristiques citées ci-dessus sont réunies.

#### a. Les protéines Opa chez *Nesseiria gonorrhoeae*

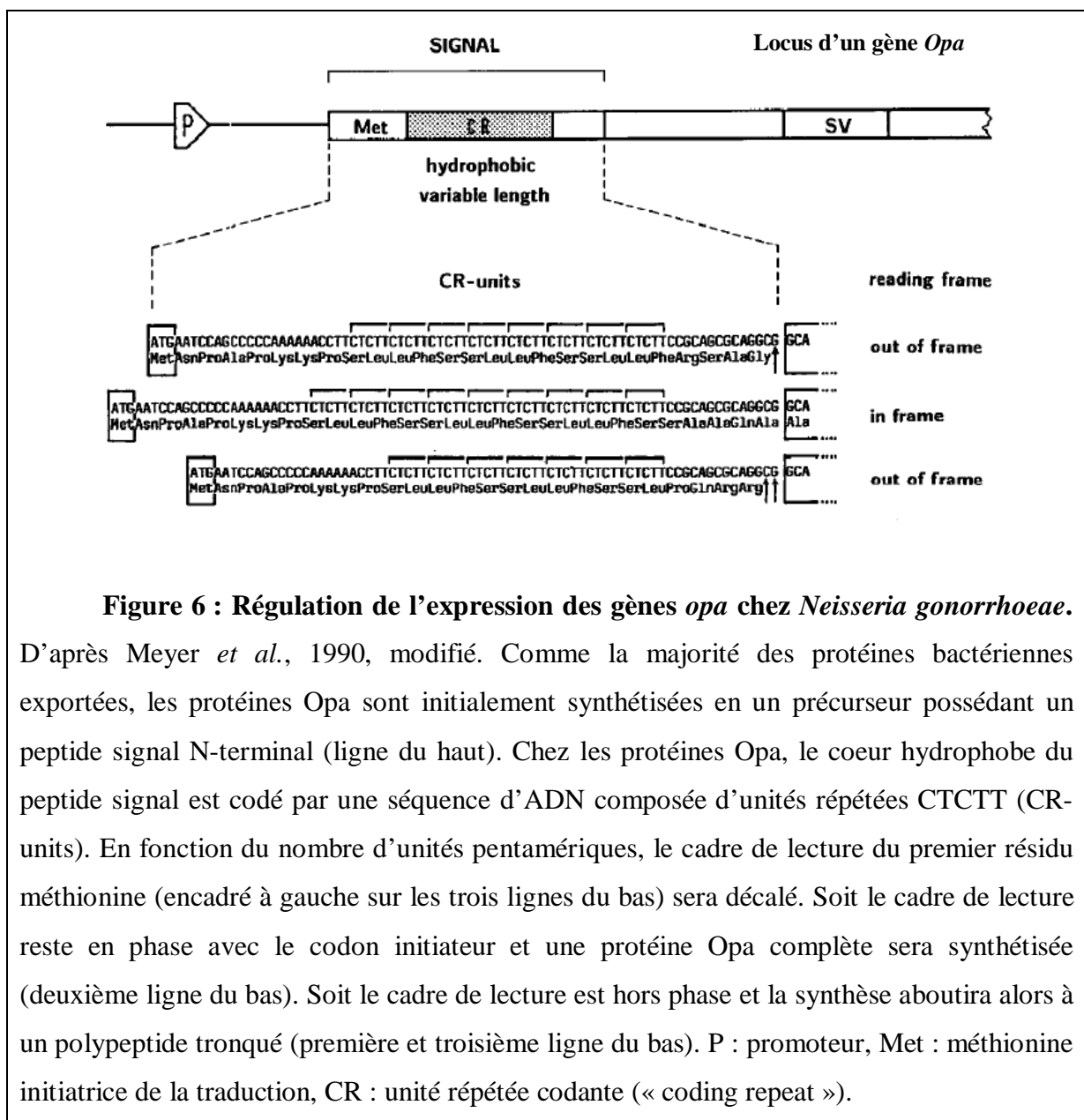
Le gonocoque *N. gonorrhoeae* possède, au niveau de sa membrane externe, une famille de protéines Opa (« opacity protein ») qui intervient dans l'adhésion et l'invasion de cellules épithéliales. Ces protéines sont codées par 11 gènes distincts qui sont transcrits constitutivement (Jerse *et al.*, 1994). Cependant, elles ne sont pas toutes présentes à la surface en même temps. Elles sont donc soumises au mécanisme de variation de phase (Van Der Woude, Baumler, 2004), et un processus de régulation de l'expression (on/off) au niveau traductionnel doit intervenir (Kline *et al.*, 2003). Ce contrôle fait intervenir la séquence du gène codant le coeur hydrophobe du peptide signal de la protéine Opa, située en aval du codon initiateur de la traduction (Figure 6). Cette séquence est composée d'un nombre variable d'unités pentamériques CTCTT (Meyer *et al.*, 1990; Stern, Meyer, 1987). Ce nombre variable d'unités répétées (de 7 à 27 répétitions) va entraîner un décalage du cadre de lecture des gènes *opa* et influencer l'issue de la traduction. En effet, si le cadre de lecture reste en phase avec le codon initiateur, alors une protéine Opa complète sera synthétisée. En revanche, si le cadre de lecture est hors phase, la synthèse aboutira alors à un polypeptide tronqué.

Cette variation de phase est donc médiée par la variation du nombre d'unités pentamériques CTCTT qui peut s'expliquer par différents mécanismes (Murphy *et al.*, 1989): par exemple des glissements (« slippage ») ou des ratés (« stuttering ») de la part des ARN polymérase lors de la transcription des gènes ; ou bien des crossing-over entre les régions codant le peptide signal de différents gènes *opa* ; ou encore un modèle impliquant un glissement d'un des deux brins d'ADN (« slipped-strand mispairing ») pendant ou en dehors de la réplication. Ce dernier phénomène est d'autant plus favorisé lorsque des unités répétées sont présentes et peut également conduire à l'insertion/délétion de répétitions (Levinson, Gutman, 1987).

Les protéines Opa sont donc soumises à un mécanisme de variation de phase conduisant à leur expression ou non expression de manière indépendante (Murphy *et al.*, 1989). Elles oscillent entre ces deux états à une fréquence d'environ  $10^{-3}$  événements par cellule par génération (Black *et al.*, 1984; Mayer, 1982). Ainsi, après une division cellulaire dans une population clonale, la majorité des cellules filles va posséder la phase d'expression de la cellule mère. Cependant, une petite partie des cellules filles aura changé de phase



d'expression et présentera donc un phénotype nouveau (Van Der Woude, Baumler, 2004). Cette variabilité phénotypique (conséquence de la variation de phase) se traduit par une population bactérienne hétérogène, contenant des cellules n'exprimant pas ou exprimant de une à plusieurs protéines Opa (Jerse *et al.*, 1994; Swanson, 1982). Les différentes combinaisons possibles des protéines exprimées sont cependant limitées par le nombre de gènes *opa* présent dans une cellule. Afin de pallier à cela, des évènements de recombinaison occasionnels interviennent entre les différents gènes *opa* diversifiant ainsi le répertoire. Ces recombinaisons vont permettre de réarranger plus particulièrement les deux régions hypervariables qui codent les domaines antigéniques variables de la protéine (Connell *et al.*, 1988).



**Figure 6 : Régulation de l'expression des gènes *opa* chez *Neisseria gonorrhoeae*.**

D'après Meyer *et al.*, 1990, modifié. Comme la majorité des protéines bactériennes exportées, les protéines Opa sont initialement synthétisées en un précurseur possédant un peptide signal N-terminal (ligne du haut). Chez les protéines Opa, le coeur hydrophobe du peptide signal est codé par une séquence d'ADN composée d'unités répétées CTCTT (CR-units). En fonction du nombre d'unités pentamériques, le cadre de lecture du premier résidu méthionine (encadré à gauche sur les trois lignes du bas) sera décalé. Soit le cadre de lecture reste en phase avec le codon initiateur et une protéine Opa complète sera synthétisée (deuxième ligne du bas). Soit le cadre de lecture est hors phase et la synthèse aboutira alors à un polypeptide tronqué (première et troisième ligne du bas). P : promoteur, Met : méthionine initiatrice de la traduction, CR : unité répétée codante (« coding repeat »).

### b. Les monomères de piline chez *Nesseiria gonorrhoeae*

Un autre exemple, toujours chez le gonocoque, concerne la piline qui constitue la sous-unité de base des pili de type IV présent à la surface de la cellule. Les pili permettent l'attachement initial du gonocoque aux cellules épithéliales des muqueuses de l'hôte (Swanson, 1973). Ils sont également importants dans l'établissement de l'infection, de telle sorte que des gonocoques qui en sont dépourvus ne sont pas capables de causer la pathologie (Kellogg *et al.*, 1963). Les monomères de piline sont codés par le gène présent au niveau de l'unique locus d'expression *pilE* (Meyer *et al.*, 1982). En plus de *pilE*, de nombreux loci transcriptionnellement silencieux (*pilS*) sont présents le long du chromosome. Chaque locus *pilS* peut contenir de 1 à 6 copies silencieuses variables de gènes codant la piline (Hamrick *et al.*, 2001). Ces copies silencieuses ne possèdent pas de séquences promotrices, ni de sites de liaison du ribosome et ne sont donc pas transcrites ni traduites (Meyer *et al.*, 1984).

Les sous-unités de piline sont soumises à une forte variation antigénique (environ  $10^{-3}$  évènements par cellule par génération) (Serkin, Seifert, 1998). Cet évènement a lieu lorsqu'une partie ou la totalité d'une copie silencieuse *pilS* recombine de façon unidirectionnelle (conversion génique) avec *pilE* (Hagblom *et al.*, 1985; Segal *et al.*, 1986). Ce mécanisme de conversion génique est médiée par de courtes séquences conservées présentes aux bornes des segments recombinés (Haas, Meyer, 1986; Meyer *et al.*, 1990). Ainsi, ce processus de recombinaison va donner lieu à de nouvelles combinaisons de séquences, conduisant ainsi à une immense variabilité des gènes exprimés (Swanson *et al.*, 1987). Le nouveau gène recombiné au locus *pilE* codera un monomère de piline antigéniquement distinct qui sera assemblé en pili fonctionnel. L'ensemble de ces processus est à l'origine de la variation antigénique de la piline. Cependant, une autre issue est également possible. Le monomère de piline peut ne pas être produit (ou assemblé inefficacement), donnant lieu à une colonie sans pili. Ce changement de phénotype (avec pili / sans pili) qui est réversible est considéré comme étant de la variation de phase. Deux raisons à cette absence totale de pili. La première est la production de piline sous une forme tronquée (piline S), qui, au lieu d'être polymérisée en pili, est sécrétée (Haas *et al.*, 1987). La deuxième est la production d'une forme beaucoup plus longue (piline L) ne pouvant s'assembler en pili. La piline L résulte d'une recombinaison inégale d'un locus *pilS* avec le locus *pilE* (Gibbs *et al.*, 1989).

### c. Les lipooligosaccharides chez *Nesseiria gonorrhoeae* et *N. meningitidis*

La variation antigénique des lipooligosaccharides (LOS) constitue un dernier exemple de variation de phase rencontré chez *N. gonorrhoeae* et *N. meningitidis*. Ces molécules sont similaires aux lipopolysaccharides (LPS) mais de taille plus petite et font aussi partie de la membrane externe des bactéries à Gram - (Plant *et al.*, 2006; Van Putten, Robertson, 1995). Elles sont exprimées en grand nombre et présentent des structures oligosaccharidiques variables (Berrington *et al.*, 2002; Schneider *et al.*, 1988). Des glycosyltransférases, responsables de l'assemblage de ces LOS, sont à l'origine de leur variation antigénique. En effet, l'expression des gènes codant ces glycosyltransférases est soumise à un mécanisme de variation de phase. Cette variation de phase est médiée, ici aussi, par des séquences répétées présentes dans la région codante (Yang, Gotschlich, 1996).

### d. Les Vlp et Vls chez *Borrelia hermsii*

Les principales lipoprotéines variables de surface Vlp (« Variable large proteins ») et Vsp (« Variable small proteins ») chez *Borrelia hermsii*, agent de la fièvre récurrente à tique, constituent un autre exemple de variation antigénique. Ces organismes présentent des plasmides linéaires qui codent des copies multiples de Vlp et Vsp. Tout comme les pili chez *Neisseria*, le transfert de ces séquences par des processus de réarrangement d'ADN (conversion génique) au niveau d'un site unique d'expression résulte dans l'expression de différentes protéines variables (Dai *et al.*, 2006).

## **2.1.2. Virus**

### a. L'hémagglutinine et la neuraminidase chez le virus Influenza

L'exemple suivant va s'intéresser aux virus, tels que les virus Influenza qui appartiennent à la famille des Orthomyxoviridae. Ce sont des virus à ARN négatif contenu dans 8 nucléocapsides. Il existe trois types de virus Influenza (types A, B et C), dont seuls les types A et B sont responsables de sévères pathologies. Ces différents types se distinguent par l'antigénicité que présente les protéines de leur capsid (Hampson, Mackenzie, 2006). L'enveloppe des virus de type A et B présente deux antigènes glycoprotéiques de surface : l'hémagglutinine (H) qui est responsable de la fixation du virus à la membrane cytoplasmique des cellules, et une neuraminidase (N), enzyme active, qui intervient dans la

maturation et le détachement du virus (Webster *et al.*, 1992). Ces deux glycoprotéines sont soumises à deux types de variations antigéniques, l'une graduelle et progressive, appelée glissement antigénique (« antigenic drift »), l'autre plus radicale, appelée saut antigénique (« antigenic shift ») (Zambon, 1999). Les antigènes de l'Influenza de type B semblent soumis uniquement aux glissements antigéniques.

Les glissements antigéniques correspondent à des modifications mineures du génome. Elles sont le résultat de l'instabilité génétique caractéristique de l'ARN et du caractère infidèle des ARN polymérase ARN dépendantes (réplicases) virales qui ne possèdent pas de mécanisme de relecture ni de correction des erreurs (Elena, Sanjuan, 2005). Ces glissements antigéniques résultent donc du changement ponctuel de quelques bases nucléiques dans les gènes codant l'hémagglutinine et la neuraminidase (potentiellement à chaque réplication) (Webster *et al.*, 1992). En ce qui concerne le gène codant l'hémagglutinine, ces changements interviennent principalement dans la région codant la partie globulaire qui possède cinq sites antigéniques hypervariables (Plotkin, Dushoff, 2003). Une conséquence de ces glissements va être l'apparition à l'intérieur du même sous-type, H3N2 par exemple, de toute une série de variants dérivant progressivement par rapport à la souche d'origine. Ces nouveaux variants seront capables d'échapper à la réponse immunitaire acquise à la suite d'une précédente infection ou d'une vaccination (Schweiger *et al.*, 2002).

Les sauts antigéniques quant à eux, correspondent à des remaniements génétiques beaucoup plus importants que des mutations ponctuelles mais beaucoup moins fréquents. Ils ont lieu lorsque deux virus différents co-infectent un hôte. Par des « réassortiments » génétiques, c'est à dire des échanges complets de gènes codant l'hémagglutinine et la neuraminidase, un nouveau virus est créé possédant des éléments des deux virus d'origine. Ces nouveaux virus contribuent ainsi à leur succès évolutif et au risque important de pandémie (Scholtissek, 1994; Webster *et al.*, 1992).

#### *b. Les glycoprotéines gp 120 et gp 41 chez le virus de l'immunodéficience humaine*

Un autre exemple de variation antigénique chez le virus de l'immunodéficience humaine (VIH). C'est un rétrovirus à ARN monocaténaire positif (présent en deux exemplaires dans la capsid) classé en type 1 (VIH-1) et type 2 (VIH-2). Du fait de son extrême variation génétique, les virus VIH-1 peuvent être classés en trois groupes, M, N et O chacun étant le résultat de différentes transmissions inter-espèces (Gao *et al.*, 1999). Le

groupe M (le plus répandu) est sous divisé en neuf sous-types et de nombreux recombinaisons (Holmes, 2004).

L'extraordinaire variation génétique du VIH-1 au sein d'un hôte est due à une combinaison de mécanismes intrinsèques. Tout d'abord, la transcriptase inverse est très infidèle et ne possède pas de mécanisme de relecture ni de correction des erreurs. Elle confère donc un fort taux de mutation, approximativement  $3,4 \times 10^{-5}$  mutations par paire de base par cycle viral. De plus, la taille du génome est d'environ  $10^4$  paires de base et le taux de base de la production virale est d'approximativement  $10^{10}$  virions par jour, ce qui conduit à la production de millions de variants viraux par jour chez une personne infectée (Perelson *et al.*, 1996; Starcich *et al.*, 1986). Des recombinaisons peuvent également conduire à un niveau supplémentaire de diversité lorsqu'une cellule est co-infectée par au moins deux souches différentes de virus. En effet, pendant le processus de transcription inverse, la transcriptase inverse peut passer d'une matrice ARN virale à l'autre. Ce processus génère ainsi de nouveaux recombinaisons viraux qui contiennent des séquences dérivées des multiples génomes parentaux (Blackard *et al.*, 2002; Robertson *et al.*, 1995; Sharp *et al.*, 1995).

Les conséquences de cette diversité génétique se répercutent principalement par des changements en acides aminés au niveau de deux glycoprotéines de l'enveloppe (Starcich *et al.*, 1986). La première est la glycoprotéine gp120 responsable de la liaison spécifique des virions aux récepteurs CD4 des cellules cibles. La seconde est la glycoprotéine gp41 responsable de la fusion des membranes virale et cellulaire lors du processus d'entrée.

La glycoprotéine gp120 possède la même organisation parmi différentes souches virales : cinq régions constantes (C1 à C5) et cinq régions variables ou boucles variables (V1 à V5) (Starcich *et al.*, 1986). Ces boucles variables sont capables d'induire efficacement des anticorps neutralisant spécifiques de souches mais qui deviendront inefficaces lorsque le virus mutera pour échapper aux pressions du système immunitaire de l'hôte. La boucle V3 est particulièrement intéressante pour plusieurs raisons : premièrement c'est un épitope de neutralisation immunodominant et deuxièmement elle participe au site de fixation des corécepteurs qui interviennent dans la liaison avec les récepteurs CD4 (Wyatt, Sodroski, 1998). Ainsi, le VIH-1 a développé un certain nombre de stratégies pour minimiser l'antigénicité et l'immunogénicité des glycoprotéines d'enveloppe, permettant au virus d'échapper à la neutralisation par les anticorps. Les boucles variables, exposées au système immunitaire, sont en évolution constante et n'induisent que des anticorps neutralisants spécifiques de souches. Elles vont également former une barrière stérique permettant de masquer les sites de liaison des récepteurs, sites conservés parmi différentes souches de virus (Kwong *et al.*, 2002).

En ce qui concerne la glycoprotéine gp41, elle est formée d'un domaine externe, d'une ancre transmembranaire et d'une queue cytoplasmique. Cette dernière présente un polymorphisme entre variants (Lambele *et al.*, 2007). Ce polymorphisme pourrait influencer la distribution intracellulaire des glycoprotéines de l'enveloppe, leur incorporation aux virions ainsi que les capacités de réplication virale.

### **2.1.3. Conclusion**

En conclusion, la variation antigénique de surface est un mécanisme clé d'évasion immunitaire mis en jeu par les bactéries pathogènes et les virus pour éviter la reconnaissance de l'hôte. Cette variation phénotypique est médiée par des mécanismes de variation antigénique et/ou de variation de phase. Ces mécanismes reposent essentiellement sur des réarrangements géniques, qui peuvent s'avérer « dangereux » car ils peuvent conduire à des erreurs. Cependant, ils représentent un avantage pour le pathogène car ils permettent à une faible proportion de la population d'être en permanence en avance sur le système de défense de l'hôte (Brunham *et al.*, 1993; Criss *et al.*, 2005; Meyer *et al.*, 1990; Zhang *et al.*, 1992).

## **2.2. Parasites eucaryotes**

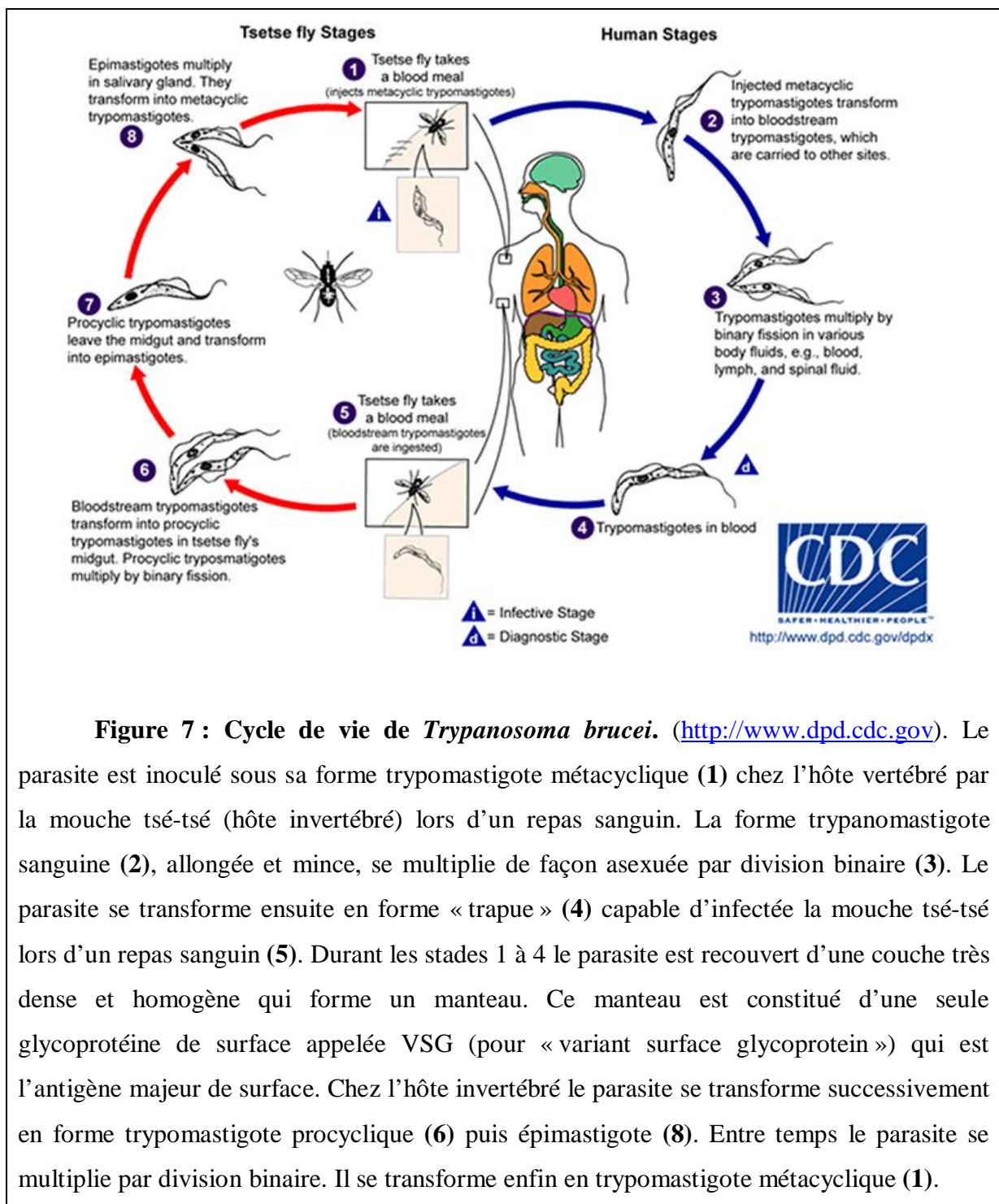
Dans cette course aux armements, en réponse au système immunitaire des hôtes vertébrés, les parasites eucaryotes ont développé diverses stratégies efficaces afin d'affaiblir ou de surmonter les défenses de l'hôte.

Prenons dans un premier temps l'exemple des helminthes parasites. Ce sont des métazoaires, responsables de l'infection de centaines de millions de personnes et d'animaux domestiques à l'échelle mondiale (Hotez *et al.*, 2008). La majorité des helminthiases est causée par des nématodes (*Ascaris*, *Trichuris*) ou des plathelminthes (*Schistosoma*). Contrairement aux bactéries et virus (cf. i-1.2.1. *Bactéries pathogènes et virus*), les helminthes (parasites pluricellulaires) ne présentent pas la même dynamique d'infection chez l'hôte vertébré. En effet, ils ne se multiplient pas de façon asexuée chez cet hôte et la charge parasitaire n'augmente que suite à des expositions et infections répétées. Ainsi, dans cette course aux armements, ils ne sont pas capables de distancer le système immunitaire en utilisant des stratégies de variation antigénique couplée à des divisions cellulaires rapides (Maizels *et al.*, 1993). En conséquence, ils ont plutôt développé des stratégies immunomodulatrices (pour revue (Maizels *et al.*, 2004)).

D'autres parasites eucaryotes, sur lesquels nous focaliserons aussi notre attention, sont les protozoaires parasites à large impact médical et vétérinaire. Afin de développer une infection chronique, ces parasites unicellulaires ont choisi différentes stratégies pour éviter leur élimination par le système de défense de l'hôte. Certains sont extracellulaires, tels que les trypanosomes africains ou les espèces de *Giardia*, et colonisent la circulation sanguine, sans se « cacher » dans les cellules hôtes. D'autres, tel que *Plasmodium sp.*, sont, au contraire, intracellulaires. Ils présentent des cycles de reproduction efficaces dans différents types de cellules de l'hôte comme les érythrocytes, les hépatocytes, les macrophages et les lymphocytes. Ils sont cependant exposés brièvement à la réponse immunitaire de l'hôte lors de leur passage de cellule à cellule au cours du processus de réinfection. Quelque soit leur mode infectieux, intra ou extracellulaires, et contrairement aux bactéries, virus ou autres helminthes, la seule stratégie d'évitement immunitaire des protozoaires parasites est la variation antigénique. Ils vont être capables de modifier certaines de leurs molécules antigéniques majeures de surface. Cette variation antigénique s'accompagne en général de l'expression d'un seul membre d'une famille multigénique codant des variants moléculaires de surface. Ce paragraphe va présenter certains de ces variants moléculaires ainsi que la régulation de leur expression chez certains parasites protozoaires. Tout d'abord ceux présents directement à la surface des parasites extracellulaires tels que les VSG chez *T. brucei* (cf. i-1.2.2.1 *Les VSG chez Trypanosoma brucei*) ou les VSP chez *G. lamblia* (cf. i-1.2.2.2. *Les VSP chez Giardia lamblia*), et ensuite ceux exposés directement à la surface de la cellule hôte, chez les parasites intracellulaires, tels que les PfEMP1 chez *P. falciparum* (cf. i-1.2.2.3. *Les PfEMP1 chez Plasmodium falciparum*).

### **2.2.1 Les VSG chez Trypanosoma brucei**

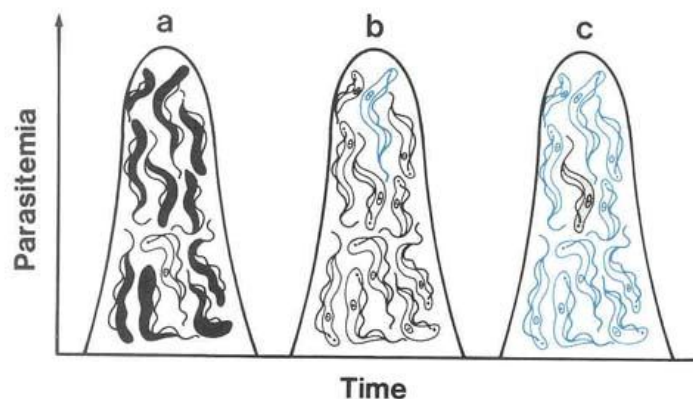
Les trypanosomes africains sont des protozoaires qui se multiplient librement dans la circulation sanguine de leurs hôtes mammifères. *Trypanosoma brucei* (Figure 7) est l'agent responsable de la maladie du sommeil chez l'homme (Barry, McCulloch, 2001). Chez l'hôte vertébré, la forme trypomastigote sanguine allongée du parasite se trouve donc directement exposé au système immunitaire. Pour s'en protéger elle est recouverte par une couche très dense et homogène qui forme un manteau. Ce manteau est constitué d'une seule glycoprotéine de surface appelée VSG (pour « Variant Surface Glycoprotein ») qui est l'antigène majeur de surface. Pour échapper à cette réponse immunitaire, le trypanosome changera de manteau régulièrement par le mécanisme de variation antigénique (Barry, McCulloch, 2001) (cf. i-1.2.1.1. *Bactéries pathogènes*).



**Figure 7 : Cycle de vie de *Trypanosoma brucei*.** (<http://www.dpd.cdc.gov>). Le parasite est inoculé sous sa forme trypanomastigote métacyclique (1) chez l'hôte vertébré par la mouche tsé-tsé (hôte invertébré) lors d'un repas sanguin. La forme trypanomastigote sanguine (2), allongée et mince, se multiplie de façon asexuée par division binaire (3). Le parasite se transforme ensuite en forme « trapue » (4) capable d'infecter la mouche tsé-tsé lors d'un repas sanguin (5). Durant les stades 1 à 4 le parasite est recouvert d'une couche très dense et homogène qui forme un manteau. Ce manteau est constitué d'une seule glycoprotéine de surface appelée VSG (pour « variant surface glycoprotein ») qui est l'antigène majeur de surface. Chez l'hôte invertébré le parasite se transforme successivement en forme trypanomastigote procyclique (6) puis épimastigote (8). Entre temps le parasite se multiplie par division binaire. Il se transforme enfin en trypanomastigote métacyclique (1).

Ainsi, le nouveau clone possédant un manteau différent sera sélectionné et pourra être initiateur, par expansion clonale, de la population suivante. Des vagues successives de variants phénotypiques VSG vont alors se mettre en place. Cette variation antigénique va permettre au trypanosome de garder un temps d'avance sur la réponse immunitaire et de causer une infection persistante (Figure 8).





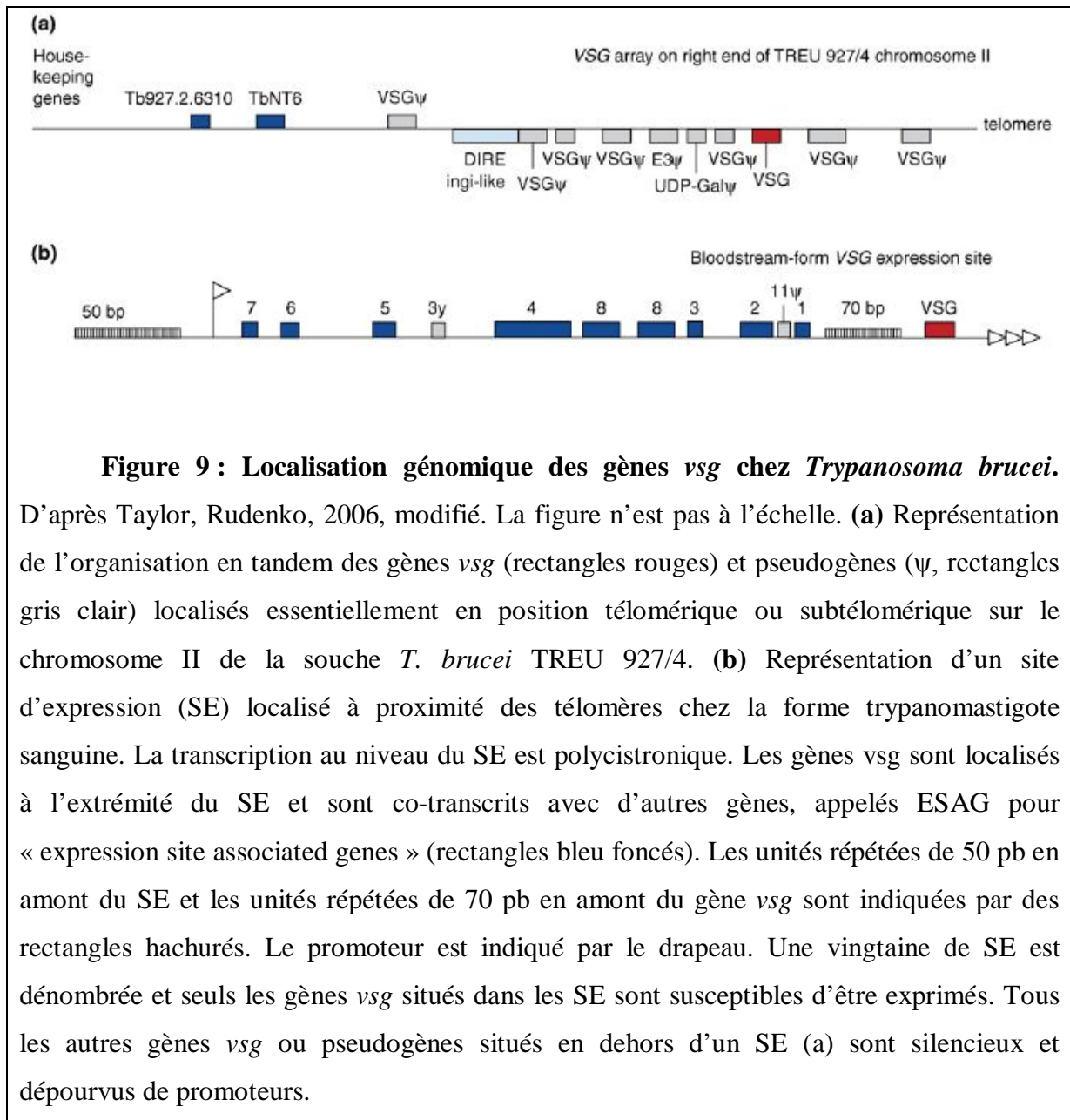
**Figure 8 : Diagramme schématique illustrant le processus de variation antigénique chez *Trypanosoma brucei*.**

(<http://www.ilri.org/InfoServ/Webpub/Fulldocs/Ilrad89/Trypano.htm>). Lorsqu'un hôte mammifère est infecté par un trypanosome, le système immunitaire de l'hôte produit des anticorps contre l'antigène majeur de surface appelé VSG (« variant surface glycoprotein ») qui forme un manteau autour du parasite. (a) Cependant, avant que tous les parasites ne soient éliminés, de nouveaux trypanosomes (forme blanche) vont apparaître dans la population clonale, possédant un manteau formé de VSG différents, qui ne seront pas reconnus par la réponse immune initiale. (b) Cette seconde vague de trypanosome se multiplie rapidement jusqu'à ce que l'hôte produise de nouveaux anticorps contre la seconde catégorie de VSG, mais certains parasites vont présenter un autre VSG (forme bleu). (c) Des vagues successives de variants phénotypiques VSG vont alors se mettre en place. Cette variation antigénique va permettre au trypanosome de garder un temps d'avance sur la réponse immunitaire.

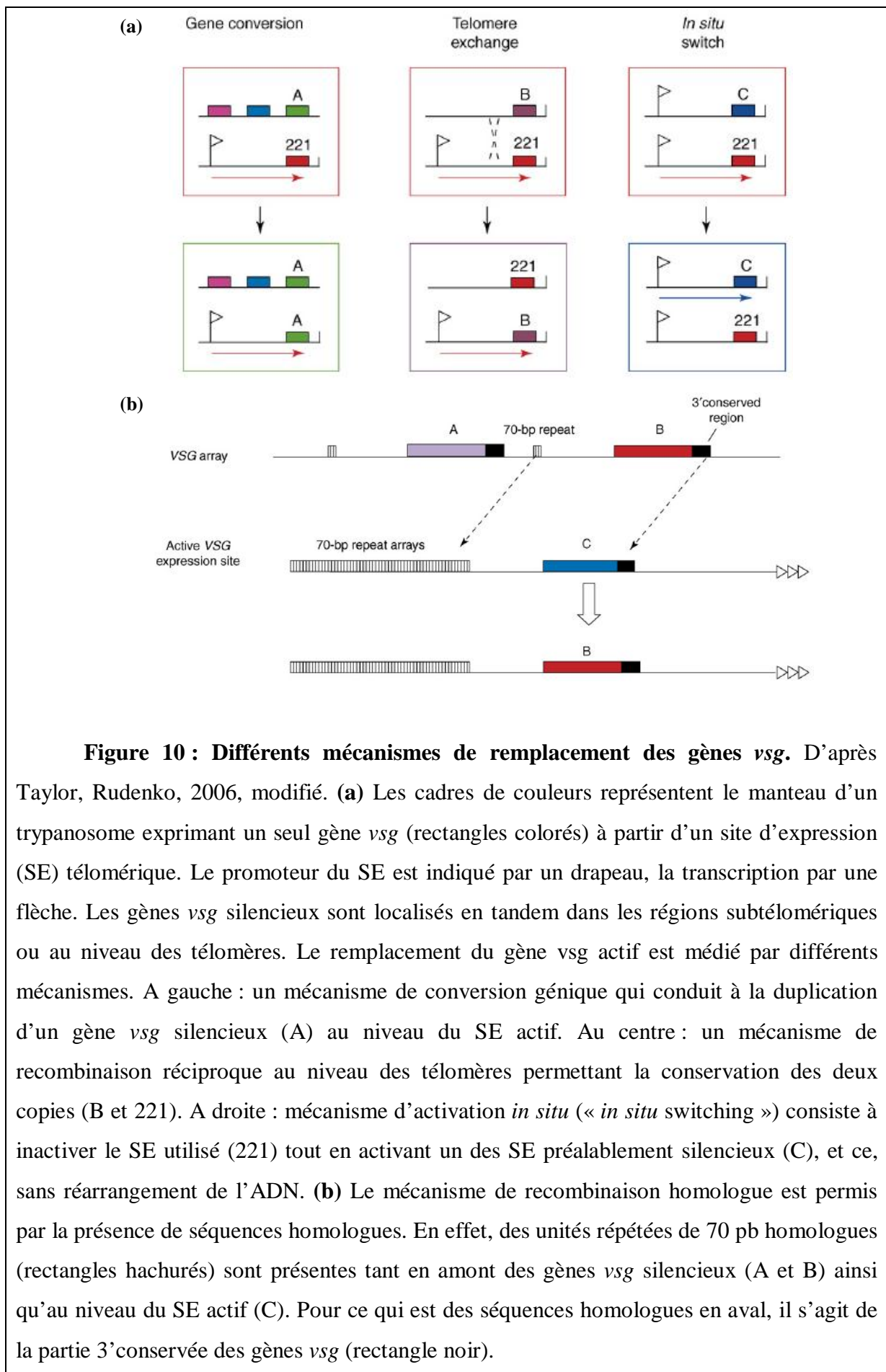
a. Conditions nécessaires à la variation antigénique des VSG chez *T. brucei*

Le séquençage récent du génome de *T. brucei* a mis en évidence la présence d'environ mille gènes (*vsg*) et pseudogènes (*pseudo-vsg*) (Berriman *et al.*, 2005). Ils sont organisés en tandem et localisés essentiellement en position télomérique ou subtélomérique (Van Der Ploeg *et al.*, 1982) (Figure 9a). Certains gènes *vsg* sont localisés à l'extrémité d'unités de transcription polycistroniques, appelées sites d'expression (SE) (Figure 9b). Une vingtaine de SE est dénombrée, toujours localisée à proximité des télomères (De Lange, Borst, 1982). Seuls les gènes *vsg* situés dans les SE sont susceptibles d'être exprimés. Pour permettre le mécanisme de variation, un seul SE doit être actif à la fois : un seul gène *vsg* est transcrit conduisant ainsi à la présence d'un seul type de VSG à la surface de la cellule. On

parle d'expression monoallélique (Donelson, 2003). Tous les autres gènes *vsg* situés en dehors d'un site d'expression sont silencieux et dépourvus de promoteurs.

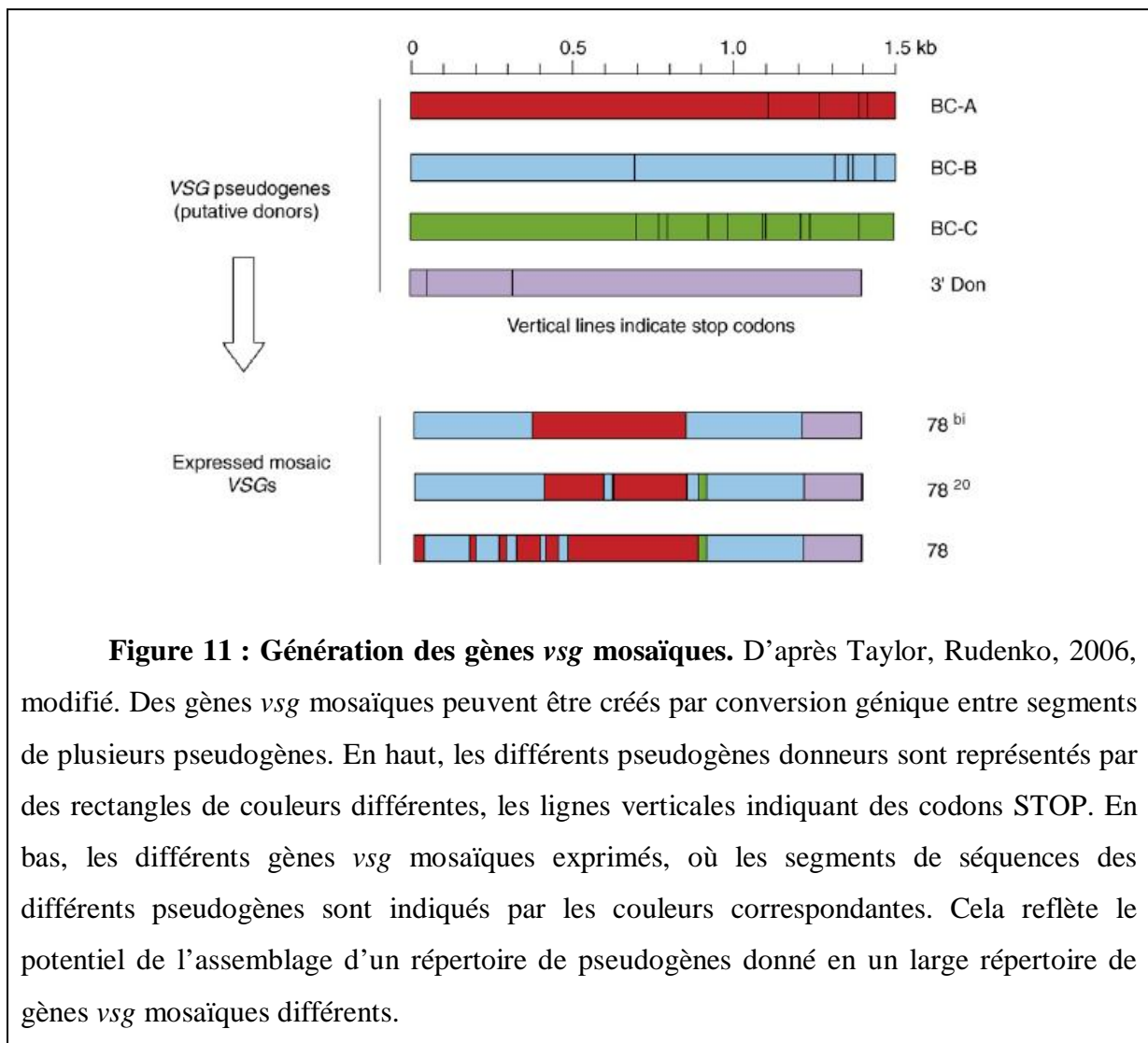


La variation antigénique nécessite également de remplacer rapidement le gène *vsg* exprimé par un autre membre de la famille (Taylor, Rudenko, 2006). Pour ce remplacement, le trypanosome utilise deux processus distincts (Figure 10a). Le premier, appelé activation *in situ* (« *in situ* ES activation » ou « *in situ* switching ») consiste à inactiver le SE utilisé tout en activant un des SE préalablement silencieux, et ce, sans réarrangement de l'ADN (Navarro *et al.*, 1999). Le second processus consiste à insérer un gène silencieux au niveau du SE actif. La recombinaison réciproque (au niveau des télomères) et la conversion génique sont les deux mécanismes permettant ce processus (Borst, Ulbert, 2001).



La recombinaison réciproque permet la conservation des deux copies du gène tandis que la conversion génique entraîne la disparition totale ou partielle du gène préalablement exprimé. Ces deux mécanismes sont dirigés par la présence de séquences homologues (Figure 9b et 10b). En effet, des unités répétées de 70 pb homologues sont présentes tant en amont des gènes *vsg* silencieux que de ceux présents au niveau du SE actif. Pour ce qui est des séquences homologues en aval, il s'agit de la partie 3' conservée des gènes *vsg* (Pays *et al.*, 1983).

La variation antigénique nécessite enfin de faire évoluer son répertoire afin que le trypanosome puisse disposer de variants VSG antigéniquement différents au cours d'une infection chronique. Les mécanismes permettant la diversification du gène *vsg* exprimé sont liés aux processus qui permettent l'intégration du gène silencieux au niveau du SE. Les gènes *vsg* forment une famille et présentent donc des homologies de séquences entre eux. La conversion génique peut ainsi s'effectuer entre deux gènes entiers (présent dans deux SE différents) mais aussi entre deux parties de gènes (Pays *et al.*, 1983). On assiste ainsi à des recombinaisons intragéniques entre des gènes fonctionnels mais également avec les pseudogènes (Taylor, Rudenko, 2006). En effet, le cas des pseudogènes est particulier car ils ne sont pas capables de coder les VSG et ne peuvent donc pas recombiner directement au niveau du SE actif. Cependant, le parasite dispose d'un mécanisme extrêmement efficace (non élucidé à l'heure actuelle) de type conversion génique (Palmer, Brayton, 2007). Ce mécanisme aboutit à la création de gènes *vsg* « mosaïques » fonctionnels constitués de fragments de différents pseudogènes (Roth *et al.*, 1989; Thon *et al.*, 1989) (Figure 11). Ces gènes mosaïques pourront alors participer à la variation antigénique en recombinant avec le SE actif et coder un nouveau variant VSG. Ce type de mécanisme augmente donc la capacité du trypanosome à disposer de variants VSG antigéniquement différents au cours d'une infection chronique (Barbet, Kamper, 1993).



### b. Régulation épigénétique de l'expression monoallélique des gènes *vsg*

Quelle que soit la manière dont la variation antigénique est réalisée, l'élément capital reste le fait qu'un seul SE soit actif à tout moment. Différents éléments pouvant être impliqués dans cette régulation, tels que le promoteur, la position télomérique, la structure de la chromatine ou encore la compartimentation nucléaire, ont fait l'objet de différentes études (Borst, Ulbert, 2001; Pays *et al.*, 2004).

Tout d'abord, la régulation de la transcription peut s'exercer au niveau du promoteur. Ce dernier se situe entre 40 et 60 kilobases (Pays *et al.*, 1989; Zomerdijk *et al.*, 1990) en amont du gène *vsg*, qui lui est localisé à la fin d'une unité de transcription polycistronique (Figure 9b). Chez *T. brucei*, les SE contenant les gènes *vsg* sont transcrits par une ARN polymérase de type I (pol I) de façon constitutive (Landeira, Navarro, 2007; Rudenko *et al.*, 1992; Scheer, Hock, 1999). Un gène rapporteur inséré en aval du promoteur d'un SE silencieux est tout de même transcrit (Navarro, Cross, 1996). De plus les gènes ESAG 6 et 7

situés au début des SE (Figure 9b) sont transcrits de manière permanente. La pol I serait donc recrutée en permanence au niveau des promoteurs, ce qui permettrait la transcription des gènes situés uniquement au début des SE. D'autre part, il semblerait qu'une séquence de 70 pb, en amont du site d'initiation de la transcription, soit suffisante pour assurer la transcription et que des éléments *cis* régulateur supplémentaires ne soient pas nécessaires (Navarro, Cross, 1998; Vanhamme *et al.*, 1995). Ainsi, alors qu'un seul gène *vsg* est exprimé à la fois, la transcription débute dans tous les sites d'expressions. La régulation de l'expression de ces gènes ne s'effectue donc pas au niveau de l'initiation de la transcription.

Ensuite, la position télomérique des SE a également fait l'objet de nombreuses d'études (Conway *et al.*, 2002; Horn, Cross, 1995; Horn, Cross, 1997). La répression de l'expression d'un gène suite à son insertion dans le voisinage d'un télomère, l'extinction télomérique, a été démontré chez la levure (Cockell, Gasser, 1999). Chez *T. brucei*, il semblerait que ce mécanisme classique ne soit pas impliqué dans l'extinction des SE inactifs (Borst, Ulbert, 2001; Pays, 2005).

La structure de la chromatine a également été appréhendée dans la mesure où elle peut réduire fortement l'accès de l'ADN à de nombreuses protéines non-histoniques (éléments *trans* régulateurs). Bien que du côté du promoteur du SE aucune différence structurale n'est observée suivant l'état d'activation transcriptionnel (Navarro, Cross, 1998), des études récentes suggèrent cependant qu'une protéine de remodelage (TbISWI) de la chromatine puisse intervenir dans la régulation des SE (Hughes *et al.*, 2007). En effet, la diminution de l'expression de TbISWI par la technique d'ARN interférent, conduit à la dérégulation des promoteurs des SE silencieux. Toutefois, cela ne permet ni leur transcription totale ni une augmentation de la fréquence d'activation des SE. Ces résultats suggèrent donc que la régulation de l'activation des SE est un processus en plusieurs étapes dans lequel la chromatine a un rôle à jouer.

Enfin, un processus de compartimentation nucléaire a été suggéré (Borst, Ulbert, 2001; Navarro *et al.*, 1999; Shea *et al.*, 1987). Ce type de contrôle épigénétique, qui est héréditaire et peut modifier l'expression d'un gène sans modifier la séquence d'ADN (Bird, 2007), pourrait être à l'origine de l'expression monoallélique des gènes *vsg*. En effet, la transcription débute dans tous les SE mais il semble cependant que celle-ci soit abortive dans tous les cas sauf dans le cas du SE actif. La pol I ne serait processive qu'au niveau d'un seul SE et pourrait alors transcrire l'unité polycistronique jusqu'au bout. Ainsi, l'association sélective des machineries de maturation, d'épissage et d'export des transcrits avec un SE unique pourrait être due au recrutement par le site actif d'un ou plusieurs éléments qui ne serai(en)t présent(s) dans le noyau qu'en quantité limitante, ne permettant la transcription

complète que d'un SE à la fois (Vanhamme *et al.*, 2000). Récemment des expériences en immunofluorescence ont démontré que le SE actif et la pol I étaient localisés dans un compartiment du noyau spécifique, distinct du nucléole. Ce compartiment a été appelé « Expression Site Body » (ESB) (Navarro, Gull, 2001). L'ESB semble être impliqué dans la régulation de l'expression monoallélique des *vsg*, en recrutant spécifiquement les éléments nécessaires à l'élongation de la transcription et en empêchant ces derniers d'interagir avec les sites d'expression silencieux, non contenus dans l'ESB (Navarro *et al.*, 1999; Navarro *et al.*, 2007) (Figure 12). De plus, l'ESB est uniquement observé chez les formes sanguines où a lieu la variation antigénique et est transmis aux cellules filles lors de la mitose. Ce processus permettrait ainsi de maintenir le statut actif de transcription au cours des divisions.

**Figure 12 : Régulation épigénétique de l'expression monoallélique des gènes *vsg***



**chez *Trypanosoma brucei*.** D'après Lopez-Rubio *et al.*, 2007, modifié. Chez *T. brucei* les sites d'expression (SE) peuvent être répartis dans trois compartiments. Les SE silencieux (rond rouge) sont présents dans le nucléoplasme. La transcription du SE actif (rond vert) a lieu au niveau d'un compartiment distinct du nucléole, l'« Expression Site Body » (ESB). Les SE partiellement actifs (rond jaune) correspondant à une transcription abortive sont probablement associés avec le nucléole. A noter la présence de l'ARN polymérase de type I (pol I) au niveau de l'ESB et du nucléole. N : noyau.

### c. Conclusion

Chaque population clonale de *T. brucei* possède son propre répertoire composé d'environ mille gènes (*vsg*) et pseudogènes. Ce répertoire est issu de réarrangements géniques qui ont lieu au sein de la famille de gènes et pseudogènes codant les VSG. Par ailleurs, la variation antigénique trouve son origine dans l'alternance de l'expression d'un gène *vsg* de cette famille multigénique. Ainsi, si les cellules filles présentent en majorité un

pattern d'expression identique à celui de la cellule mère, une petite proportion de la population exprime un gène *vsg* différent. Dans ce cas, cette population ne sera pas reconnue par la réponse immune développée contre la population clonale initiale, et permettra l'expansion de nouvelles populations clonales (Meunier, 2001).

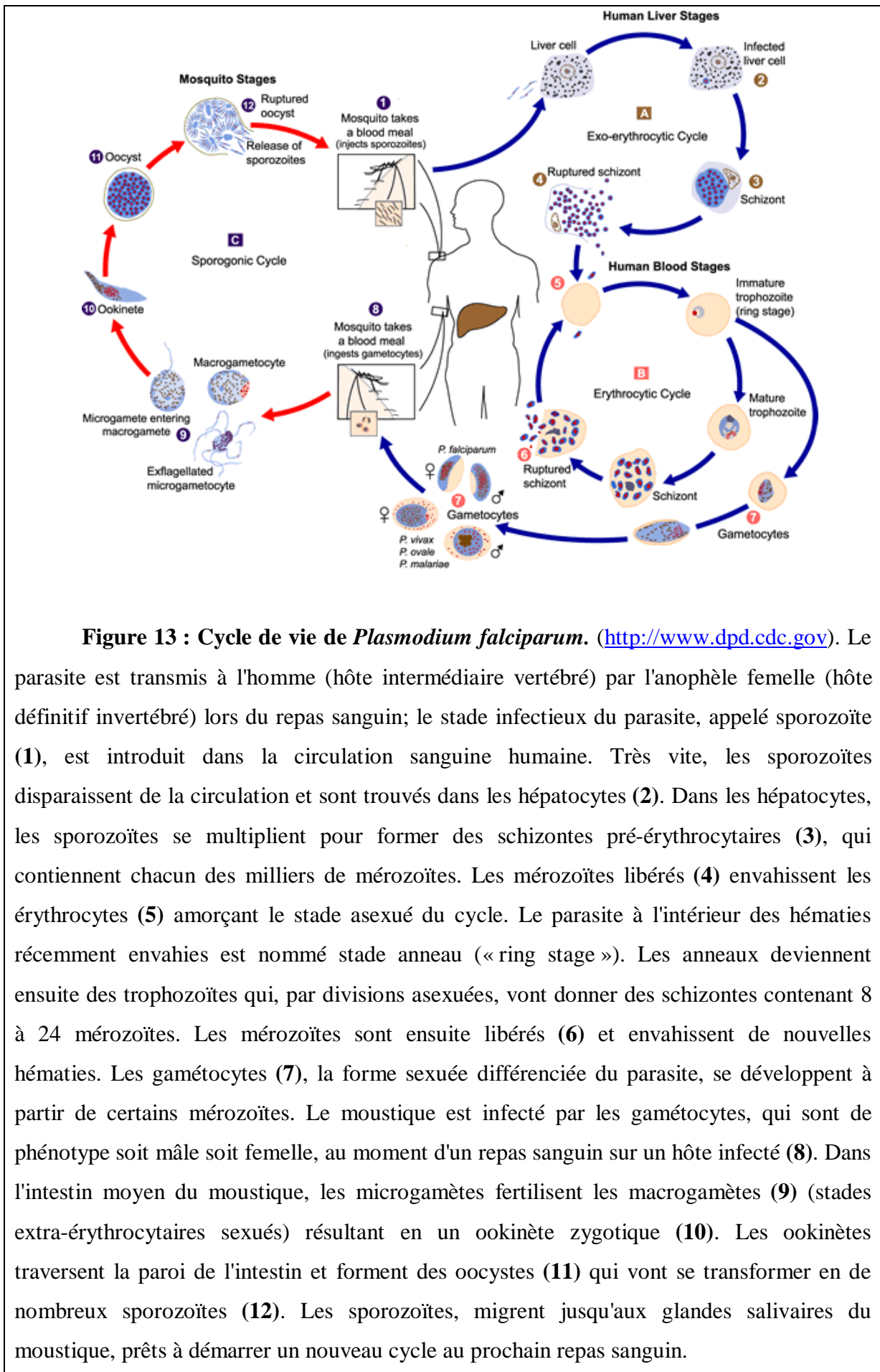
### **2.2.2 Les VSP chez *Giardia lamblia***

D'autres parasites extracellulaires ont développé des stratégies similaires, telles que celle de *Giardia lamblia*, dont le stade trophozoïte colonise le tube digestif de nombreux mammifères (Adam, 1991). Ici aussi, afin d'éviter la réponse immunitaire, les trophozoïtes présentent une variabilité phénotypique qui s'exprime au niveau des protéines de surface (« Variable Surface Protein », VSP) (Muller, Gottstein, 1998). Comme chez *T. brucei*, un seul des 150 gènes *vsp* est exprimé à la fois (exclusion allélique), excepté pendant la période de remplacement d'un variant (Nash, 2002). Les remplacements interviennent toutes les 6 à 13 générations de telle manière que les nouveaux variants générés chez l'hôte ne sont pas reconnus par les anticorps spécifiques (Nash *et al.*, 1990). Les mécanismes à l'origine de cette variation antigénique ne sont pas encore connus. Cependant des travaux récents ont mis en évidence que l'expression d'un seul gène *vsp* n'est pas associée à des réarrangements géniques mais plutôt à des acétylations des histones au niveau des séquences immédiatement en amont du gène *vsp* exprimé. Cela supporte donc l'hypothèse d'un contrôle épigénétique de la variation antigénique chez *G. lamblia* (Kulakova *et al.*, 2006).

### **2.2.3 Les PfEMP1 chez *Plasmodium falciparum***

Les protozoaires intracellulaires utilisent également la variation antigénique pour maintenir une infection persistante chez l'hôte. C'est le cas du parasite *Plasmodium falciparum* (Figure 13) responsable de la forme la plus sévère de paludisme chez l'homme, avec 300 à 500 million de cas et 1 à 2 million de morts recensés par an (Snow *et al.*, 2005).



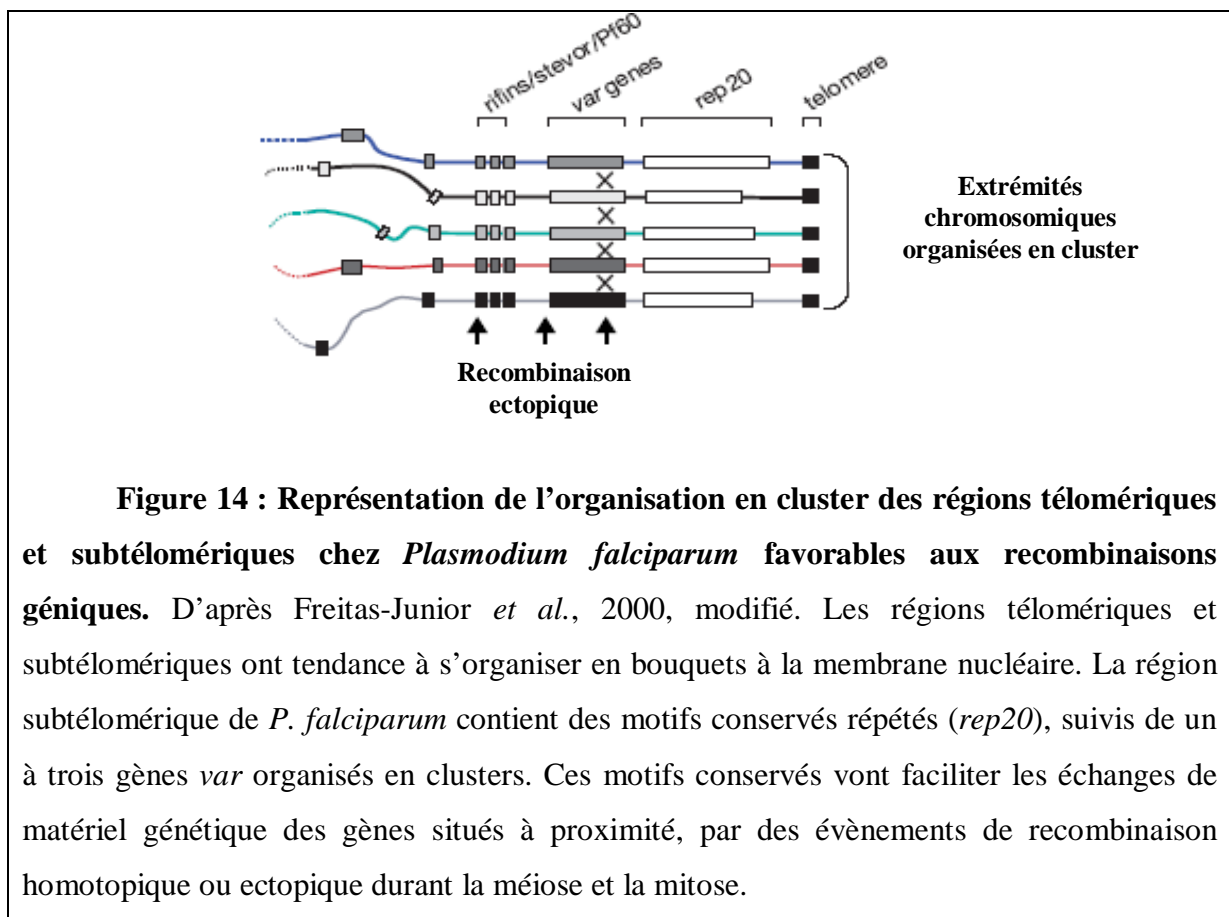


**Figure 13 : Cycle de vie de *Plasmodium falciparum*.** (<http://www.dpd.cdc.gov>). Le parasite est transmis à l'homme (hôte intermédiaire vertébré) par l'anophèle femelle (hôte définitif invertébré) lors du repas sanguin; le stade infectieux du parasite, appelé sporozoïte (1), est introduit dans la circulation sanguine humaine. Très vite, les sporozoïtes disparaissent de la circulation et sont trouvés dans les hépatocytes (2). Dans les hépatocytes, les sporozoïtes se multiplient pour former des schizontes pré-érythrocytaires (3), qui contiennent chacun des milliers de mérozoïtes. Les mérozoïtes libérés (4) envahissent les érythrocytes (5) amorçant le stade asexué du cycle. Le parasite à l'intérieur des hématies récemment envahies est nommé stade anneau (« ring stage »). Les anneaux deviennent ensuite des trophozoïtes qui, par divisions asexuées, vont donner des schizontes contenant 8 à 24 mérozoïtes. Les mérozoïtes sont ensuite libérés (6) et envahissent de nouvelles hématies. Les gamétocytes (7), la forme sexuée différenciée du parasite, se développent à partir de certains mérozoïtes. Le moustique est infecté par les gamétocytes, qui sont de phénotype soit mâle soit femelle, au moment d'un repas sanguin sur un hôte infecté (8). Dans l'intestin moyen du moustique, les microgamètes fertilisent les macrogamètes (9) (stades extra-érythrocytaires sexués) résultant en un ookinète zygotique (10). Les ookinètes traversent la paroi de l'intestin et forment des oocystes (11) qui vont se transformer en de nombreux sporozoïtes (12). Les sporozoïtes, migrent jusqu'aux glandes salivaires du moustique, prêts à démarrer un nouveau cycle au prochain repas sanguin.

### a. Structure et organisation génomique des gènes var

La gravité de la pathologie est associée à l'agrégation des érythrocytes parasités et à leur adhérence à l'endothélium (Biggs *et al.*, 1992; Miller *et al.*, 2002; Udomsangpetch *et al.*, 1989). Cette adhérence résulte de l'expression de la protéine PfEMP1 (Leech *et al.*, 1984). Cette protéine vient se localiser au niveau de la membrane des érythrocytes infectés et constitue ainsi un antigène majeur de surface de ces érythrocytes (Biggs *et al.*, 1991). La variation antigénique qui s'exprime au niveau des PfEMP1, permettra au parasite d'éviter la réponse immunitaire. Ces protéines sont codées par les gènes de la famille *var* (Gardner *et al.*, 2002), qui partagent tous la même structure à 2 exons. Le premier exon est très polymorphe et variable en taille (3,5 à 9 kb). Il code des domaines extracellulaires d'adhésion appelés « Duffy Binding-Like » (DBL) et un domaine intermédiaire riche en cystéines (CIDR). Le second exon, plus conservé et plus petit (1 à 1,5 kb), code la queue cytoplasmique (Su *et al.*, 1995).

Environ 60 gènes *var*, par génome haploïde, sont répartis sur les 14 chromosomes de *P. falciparum*. Ils ont, à plus de 80%, une localisation subtélomérique (Gardner *et al.*, 2002). En plus de ces copies subtélomériques, il existe des regroupements ou clusters de gènes *var* en position interne sur les chromosomes 4, 7, 8 et 12 (Fischer *et al.*, 1997).



**Figure 14 : Représentation de l'organisation en cluster des régions télomériques et subtélomériques chez *Plasmodium falciparum* favorables aux recombinaisons géniques.** D'après Freitas-Junior *et al.*, 2000, modifié. Les régions télomériques et subtélomériques ont tendance à s'organiser en bouquets à la membrane nucléaire. La région subtélomérique de *P. falciparum* contient des motifs conservés répétés (*rep20*), suivis de un à trois gènes *var* organisés en clusters. Ces motifs conservés vont faciliter les échanges de matériel génétique des gènes situés à proximité, par des événements de recombinaison homotopique ou ectopique durant la méiose et la mitose.

### b. Transcription des gènes var

Tout comme pour les gènes *vsg* de *T. brucei* (cf. i-2.2.1 *Les VSG chez Trypanosoma brucei*), les gènes *var* de *P. falciparum* sont exprimés de façon monoallélique permettant ainsi la présentation d'un seul variant PfEMP1 à la surface des érythrocytes infectés (Scherf *et al.*, 1998; Voss *et al.*, 2006). La variation antigénique est possible grâce à un système de commutation mutuellement exclusive entre les gènes *var*. Elle a lieu au stade précoce d'anneaux (Figure 13) et peut atteindre une fréquence de 2% par génération (Roberts *et al.*, 1992). Contrairement à *T. brucei*, elle n'est pas le résultat de réarrangements génomiques ou de mécanismes de conversion génique (Deitsch *et al.*, 1999; Horrocks *et al.*, 2002), mais a lieu *in situ* (Scherf *et al.*, 1998). En effet, la localisation chromosomique des gènes *var*, qu'elle soit subtélomérique ou interne, n'influe pas sur leur capacité à être transcrits (Fischer *et al.*, 1997; Hernandez-Rivas *et al.*, 1997).

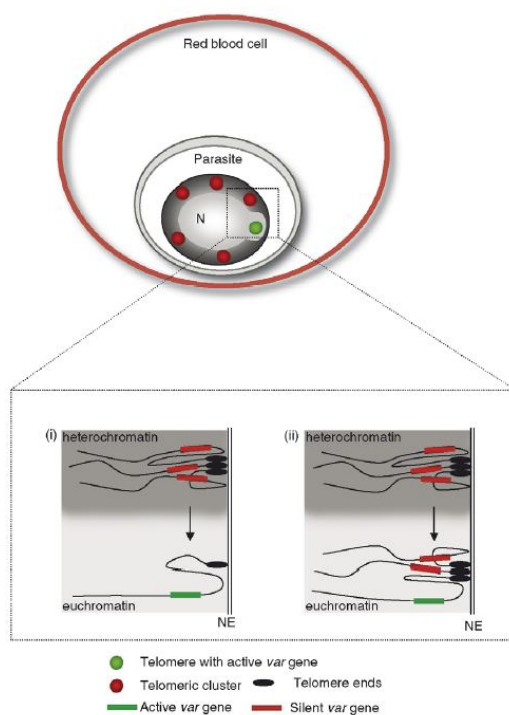
La transcription des gènes *var* commence environs 6 heures après l'invasion de l'érythrocyte par le mérozoïte. A ce stade, la transcription est non restrictive, c'est-à-dire que plusieurs gènes *var* sont transcrits (Chen *et al.*, 1998). Cependant, des expériences de « nuclear run-on » (Scherf *et al.*, 1998) et de Northern blot (Kyes *et al.*, 2003) montrent qu'un unique gène *var* est transcrit au stade trophozoïte. C'est à ce stade que la protéine PfEMP1 est traduite et transportée à la surface de l'érythrocyte. De plus, il est établi que l'activation ou l'inactivation des promoteurs des gènes *var* n'est pas sous le contrôle de facteurs *cis* ou *trans* régulateurs (Deitsch *et al.*, 1999). Il semblerait donc que des facteurs épigénétiques, tels que la topographie des gènes dans le noyau (Voss *et al.*, 2006) et la structure locale de la chromatine (Lopez-Rubio *et al.*, 2007), puissent réguler la transcription mutuellement exclusive des gènes.

### c. Régulation épigénétique de l'expression monoallélique des gènes var

L'organisation nucléaire des gènes *var* est quelque peu particulière (Figure 14, cf. i-2.2.3.d. *Polymorphisme du répertoire de gènes var*). En effet, les extrémités chromosomiques contenant la majorité des gènes *var* ont tendance à se regrouper et à former des clusters de télomères localisés à la périphérie du noyau (Freitas-Junior *et al.*, 2000). L'analyse de cette zone périphérique montre une chromatine condensée (hétérochromatine) dans laquelle quelques petites zones de chromatine non condensée (euchromatine) sont observées (Freitas-Junior *et al.*, 2005; Ralph *et al.*, 2005). Lors de son activation, la région télomérique où se trouve le gène *var* est déplacée vers une zone distincte, toujours en

périphérie. Cette observation confirme donc qu'il existe dans les zones périnucléaires des régions permissives à la transcription (Duraisingh *et al.*, 2005; Ralph *et al.*, 2005; Voss *et al.*, 2006) (Figure 15). Cependant, la relocalisation physique de la région télomérique vers la région permissive ne peut expliquer à elle seule l'expression exclusive d'un gène. En effet, cette région possède plusieurs gènes *var* proches les uns des autres et finalement un seul sera exprimé. Cette expression unique peut s'expliquer par la présence au niveau de la région permissive d'un régulateur unique tel qu'un activateur (Kyes *et al.*, 2007). Il serait nécessaire pour activer *in trans* et de manière stochastique un seul membre de la famille multigénique, comme cela a été décrit pour les récepteurs olfactifs (Lomvardas *et al.*, 2006). Il y aurait donc une compétition entre les gènes *var* pour l'interaction avec cet élément régulateur commun (Frank, Deitsch, 2006; Voss *et al.*, 2006).

**Figure 15 : Régulation épigénétique de l'expression monoallélique des gènes *var* chez *Plasmodium falciparum*.** D'après Lopez-Rubio *et al.*, 2007, modifié. Chez *P.*



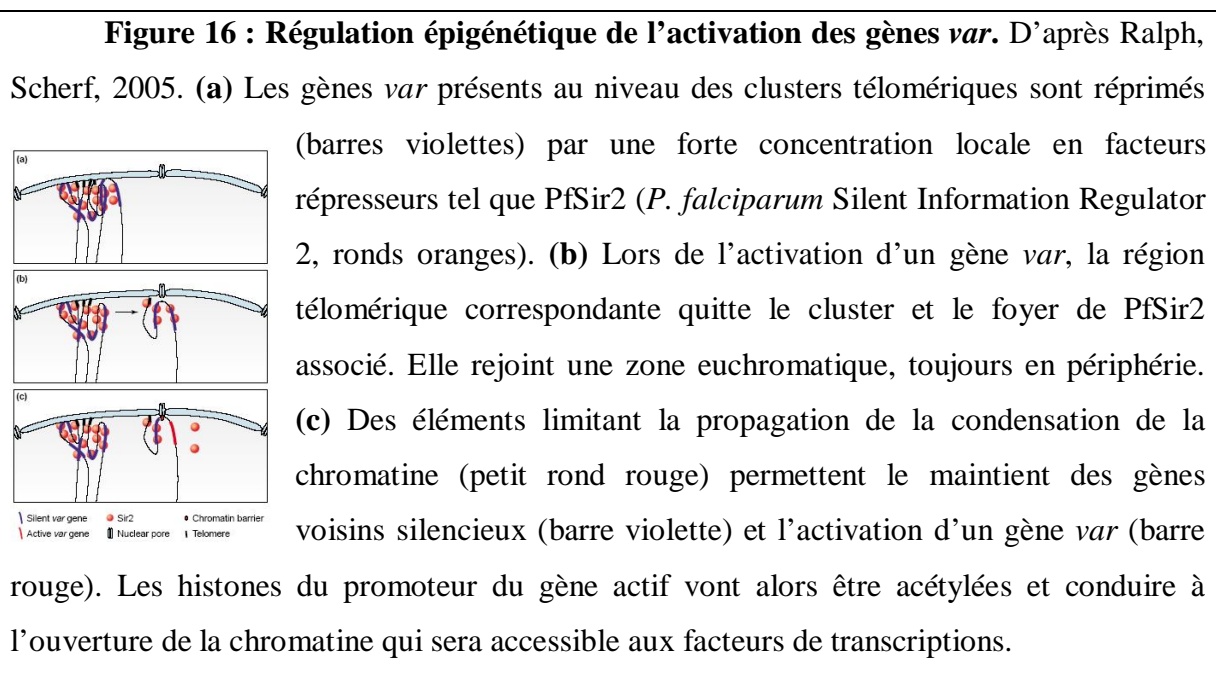
*falciparum* les télomères et les gènes *var* sont organisés en cluster, de quatre à 8 extrémités de chromosomes, à la périphérie nucléaire (rond rouge). Un seul télomère possède un gène *var* actif à la fois (rond vert). Lors de son activation, le télomère où se trouve le gène *var* est déplacé d'une région non permissive (hétérochromatine) vers une région permissive à la transcription (euchromatine). Cette relocalisation concerne soit (i) le télomère, soit (ii) le télomère au sein du cluster. N : noyau, NE : enveloppe nucléaire.

La modification structurale locale de la chromatine peut également participer à la régulation de l'expression exclusive des gènes *var*. La chromatine présente différents niveaux de condensation allant d'un état relâché (euchromatine) à un état beaucoup plus compact (hétérochromatine). Ces différents états structuraux de la chromatine facilitent ou non l'accès de l'ADN à de nombreuses protéines non-histioniques et régulent ainsi l'activation des gènes. L'état de la chromatine est conditionné par de nombreux facteurs. La méthylation de l'ADN

et les modifications post-traductionnelles des histones sont les plus courantes (Berger, 2007; Kouzarides, 2007).

La méthylation est, en règle générale, associée à la répression transcriptionnelle. Cependant, des travaux ont montré que le statut de méthylation de l'ADN étaient identique entre des gènes *var* actifs et inactifs (Scherf *et al.*, 1998). Ce facteur épigénétique ne semble donc pas intervenir dans la régulation de la transcription.

L'une des modifications post-traductionnelles des histones ayant d'importantes fonctions régulatrices de la transcription est l'acétylation. En effet, il existe une très forte corrélation entre l'acétylation des histones et l'activité transcriptionnelle de certaines zones chromatiennes (Jenuwein, Allis, 2001) : l'hyperacétylation des histones « ouvre la chromatine » et favorise l'expression des gènes alors que l'hypoacétylation « ferme la chromatine » et réduit ou annihile l'expression des gènes. Ces modifications réversibles des histones sont réalisées par des histones acétylases ou désacétylase. Chez *P. falciparum*, des expériences d'immunoprécipitation de la chromatine (ChIP) ont montré que la région en amont des gènes *var* était associée à ce type de marque épigénétique (Freitas-Junior *et al.*, 2005; Lopez-Rubio *et al.*, 2007). En effet, un gène *var* actif est associé à des histones hyperacétylées, alors que les histones associées à un gène *var* silencieux sont hypoacétylées. De plus, une histone désacétylase (PfSir2) homologue de la protéine SIR 2 (Silent Information Regulator 2) a été mise en évidence et colocalise avec les zones nucléaires les plus condensées. Ces zones hétérochromatiques seraient ainsi maintenues par la présence de PfSir2, qui désacétylerait les histones présentes au niveau des télomères, ce qui maintiendrait les gènes *var* subtélomériques voisins inactifs (Ralph, Scherf, 2005) (Figure 16).



#### d. Polymorphisme du répertoire de gènes var

La variation antigénique chez *P. falciparum* est effectuée à partir d'un répertoire de gènes *var* limité (environ 60) en comparaison au répertoire de gènes *vsg* de *T. brucei*. Il semblerait donc que le nombre de gènes de ce répertoire ne soit pas le seul élément à l'origine de la variabilité antigénique observée. Un autre élément, d'autant plus important dans l'expansion des populations clonales ou de la population au sens large, est de disposer de répertoires de gènes *var* différents. En effet, le système immunitaire doit probablement sélectionner des répertoires de gènes *var* ne se recouvrant pas. Il favoriserait ainsi la prolifération de parasite n'induisant pas l'émergence d'une réaction immunitaire secondaire, consécutive à l'exposition préalable à une ou plusieurs molécules PfEMP1. Or, un répertoire limité tel que celui-ci ne peut générer une variabilité antigénique suffisante que si il est capable de se diversifier. Des informations concernant la diversité de ce répertoire ont été apportées par le séquençage complet et l'annotation du génome de l'isolat 3D7 (Gardner *et al.*, 2002), ainsi que des projets de séquençage de différents isolats (Kraemer *et al.*, 2007). Ces études révèlent une extrême diversité des gènes *var* aussi bien au sein d'un même isolat qu'entre isolats différents. Ainsi, une stratégie populationnelle basée sur des mécanismes génétiques spécifiques à *P. falciparum*, doit exister pour générer constamment de nouveaux répertoires de gènes *var*.

L'analyse de la ségrégation des gènes *var*, à partir du croisement de deux souches parasitaires, montre que les gènes *var* de la progéniture ne sont pas toujours présents sur les mêmes chromosomes par rapport aux souches parentales (Freitas-Junior *et al.*, 2000). Deux phénomènes semblent concourir à la nature hautement recombino-génique des régions subtélomériques et à la création de variabilité génique et antigénique au sein des gènes *var* (Freitas-Junior *et al.*, 2000; Taylor *et al.*, 2000) (Figure 14). D'une part, la région subtélomérique de *P. falciparum* contient des motifs conservés répétés (*rep20*), suivis de un à trois gènes *var* organisés en clusters (Kyes *et al.*, 2007). Ces motifs conservés vont faciliter les échanges de matériel génétique des gènes situés à proximité, par des événements de recombinaison allélique ou ectopique durant la méiose. D'autre part, les régions télomériques et subtélomériques ont tendance à s'organiser en bouquets à la membrane nucléaire. Cela permet l'accolement des motifs subtélomériques conservés de deux chromosomes non homologues créant ainsi une structure favorable aux événements de recombinaison ectopique (Freitas-Junior *et al.*, 2000; Kraemer, Smith, 2003). De telles structures favorisant les recombinaisons ont été décrites chez de nombreuses espèces eucaryotes et procaryotes (Cooper *et al.*, 1998; De Lange, 1998). Par ailleurs, des structures

similaires en bouquets à la périphérie du noyau ont été mises en évidence durant le stade mérozoïte (Freitas-Junior *et al.*, 2000). Bien que cela n'ait pas été démontré, ces structures pourraient favoriser les réarrangements géniques également lors des mitoses chez l'hôte vertébré.

#### e. Conclusion

Ainsi, malgré un répertoire de gène *var* de faible envergure, *P. falciparum* exprime une variabilité antigénique élevée. Les événements de recombinaison allélique ou ectopique durant la méiose (phase sexuée) et la mitose (phase asexuée) sont à la base de cette variabilité antigénique. La diversification du répertoire de gènes *var* durant la phase érythrocytaire asexuée permet au parasite de maintenir une infection chronique chez l'hôte vertébré. Cette diversification, à l'issue de la phase sexuée chez l'hôte intermédiaire, favorise également l'expansion du parasite dans la population d'hôte vertébré en permettant à ce dernier de disposer d'un répertoire unique. Le parasite pourra alors présenter une PfEMP1 originale lors de la première infestation conditionnant ainsi le succès de l'installation du parasite chez l'hôte vertébré.

### **3. Conclusion hôte vertébré / pathogène**

L'application de la théorie de la « Reine Rouge » aux systèmes hôte / parasite (Combes, 2000) prédit une course aux armements dans laquelle les espèces devront disposer d'une plasticité phénotypique des molécules qui jouent un rôle crucial au cœur de l'interaction. Comme nous venons de l'illustrer dans cette première partie de l'introduction, ce modèle théorique se vérifie largement dans les interactions hôte vertébré / pathogène.

En effet, dans cette course aux armements, les hôtes vertébrés ont développé des mécanismes de défense très sophistiqués. Ils sont basés sur des molécules de reconnaissance immunitaires extrêmement polymorphes et/ou diversifiées (Ig/TCR, CMH et VLR). Ces niveaux de polymorphisme et/ou de diversification sont générés par des mécanismes complexes : réarrangement somatique, diversification génique ou encore diversification structurale.

En réponse, les pathogènes ont développé un panel de stratégies complexes et efficaces leur permettant d'affaiblir ou de surmonter les défenses de l'hôte. Un mécanisme classique d'évasion immune consiste à faire varier leurs molécules immunodominantes de surface. Cette variation phénotypique est médiée tout d'abord par la combinaison des

mécanismes de variation antigénique et/ou de variation de phase. Enfin, les antigènes majeurs de surface (Opa, gp120, VSG et PfEMP1), présentent des niveaux élevés de diversification et/ou de polymorphisme, générés par des mécanismes de recombinaisons et de réarrangements géniques.



## II. INTERACTION HÔTE INVERTEBRE / PATHOGENE

Qu'en est-il dans les systèmes impliquant un hôte invertébré ? Ces derniers, comme les vertébrés (cf. *i-I. Interaction hôte vertébré / pathogène*), évoluent en interaction avec de nombreuses espèces pathogènes. La survie de chacune des espèces dépend également de sa capacité à reconnaître et à résister aux différents pathogènes auxquels elles sont confrontées. Ainsi, les pressions sélectives pour les hôtes invertébrés sont comparables à celles des hôtes vertébrés (Loker *et al.*, 2004). Cependant, en réponse aux pathogènes, seuls les vertébrés ont développé une composante immunitaire adaptative (ou spécifique). L'efficacité de cette composante est basée à la fois sur un haut degré de spécificité ainsi qu'une remarquable propriété de « mémoire ». La spécificité repose sur des récepteurs extrêmement diversifiés, générés par des mécanismes somatiques durant l'ontogenèse de chaque individu (Maizels, 2005) (cf. *i-I.1.1.1. Immunoglobulines et TCR*).

Malgré l'absence d'un tel système, les invertébrés sont capables de développer des réponses immunitaires efficaces. Leur capacité à disposer de récepteurs immunitaires diversifiés, comparable à celle des vertébrés, a longtemps été écartée, car les mécanismes moléculaires présents chez les vertébrés, tels que les réarrangements somatiques à l'origine de la diversité des anticorps, n'avaient pas été identifiés chez les invertébrés (Pancer, Cooper, 2006). De plus, selon le paradigme de Medzhitov et Janeway (Medzhitov, Janeway, 1997), la réponse immunitaire des invertébrés serait exclusivement innée et reposerait uniquement sur un nombre limité de récepteurs moléculaires codés dans la lignée germinale, appelés PRR (« Pattern Recognition Receptor »). Le spectre de reconnaissance de ces récepteurs est dirigé vers des patrons moléculaires invariants. Ils sont largement répandus chez les pathogènes et absents des cellules de l'hôte, et sont appelés PAMP (« Pathogen-Associated Molecular Pattern »). Ils correspondent par exemple aux LPS des parois des bactéries à Gram négatif ou aux peptidoglycanes des bactéries à Gram positif.

Ainsi, ce type de molécules de reconnaissance invariables, contrastent fortement avec les molécules hypervariables présentes dans l'interaction hôte vertébré / pathogène (cf. *i-I. Interaction hôte vertébré / pathogène*). Elles ne reflètent pas ce que l'on est en lieu d'attendre dans les modèles théoriques de la course aux armements, où hôte invertébré et pathogène devraient maintenir et/ou générer un niveau suffisant de diversité et/ou de polymorphisme des molécules au cœur de l'interaction. Des études menées ces 10 dernières années, apportent des perspectives nouvelles quant à la compréhension du système immunitaire inné des invertébrés et soutiennent les prédictions théoriques.

Cette deuxième partie de l'introduction se focalisera donc sur des systèmes impliquant hôtes invertébrés et pathogènes. Nous traiterons, dans un premier temps, du potentiel de diversification des récepteurs immunitaires chez certains invertébrés (cf. i-II.1 *Hôtes invertébrés : immunité innée et diversification*) tels que les diptères, les céphalocordés ou encore les mollusques gastéropodes. Dans un deuxième temps, nous ferons état du manque de connaissances actuelles concernant les stratégies développées par les pathogènes d'invertébrés dans cette course aux armements (cf. i-II.2. *Pathogènes*).

## **1. Hôtes invertébrés : immunité innée et diversification**

Le paradigme de Medzhitov et Janeway (Medzhitov, Janeway, 1997) est aujourd'hui remis en cause. En effet, une série de travaux mettent en évidence l'existence d'une « mémoire » immune (Kurtz, Franz, 2003; Little *et al.*, 2003) ainsi qu'un certain degré de « spécificité » chez les invertébrés (Little *et al.*, 2005). Un exemple frappant, est la réponse immune spécifique, développée par le bourdon *Bombus terrestris*, contre deux espèces de bactéries proches appartenant au genre *Paenibacillus* (Sadd, Schmid-Hempel, 2006).

Les bases moléculaires soutenant cette immunité spécifique chez les invertébrés sont actuellement en cours d'investigation dans différents groupes de recherche. Il semblerait que la diversité génétique des gènes codant des PRR et/ou des effecteurs de la réponse immune (Schulenburg *et al.*, 2007) puisse permettre, dans la course aux armements, la reconnaissance ou l'élimination du parasite via un système de serrure / clé. Ainsi, chaque variant moléculaire de récepteur (ou d'effecteur) reconnaîtra un variant moléculaire du pathogène : c'est l'hypothèse des « matching alleles » (Agrawal, Lively, 2003). Ces gènes hôtes, pour lesquels la diversification commence à être connue, codent par exemple les Dscam (« Down syndrome cell adhesion molecule ») rencontrés chez les diptères (cf. i-II.1.1. *Les Dscam chez les diptères*), les VCBP (« Variable Chitin-Binding Protein ») chez l'amphioxus *Branchiostoma floridae* (cf. i-II.1.2. *Les VCBP chez Branchiostoma floridae*) ou encore les FREP (« Fibrinogen-Related Proteins ») chez le mollusque *Biomphalaria glabrata* (cf. i-II.1.3. *Les FREP chez Biomphalaria glabrata*).

### **1.1. Les Dscam chez les diptères**

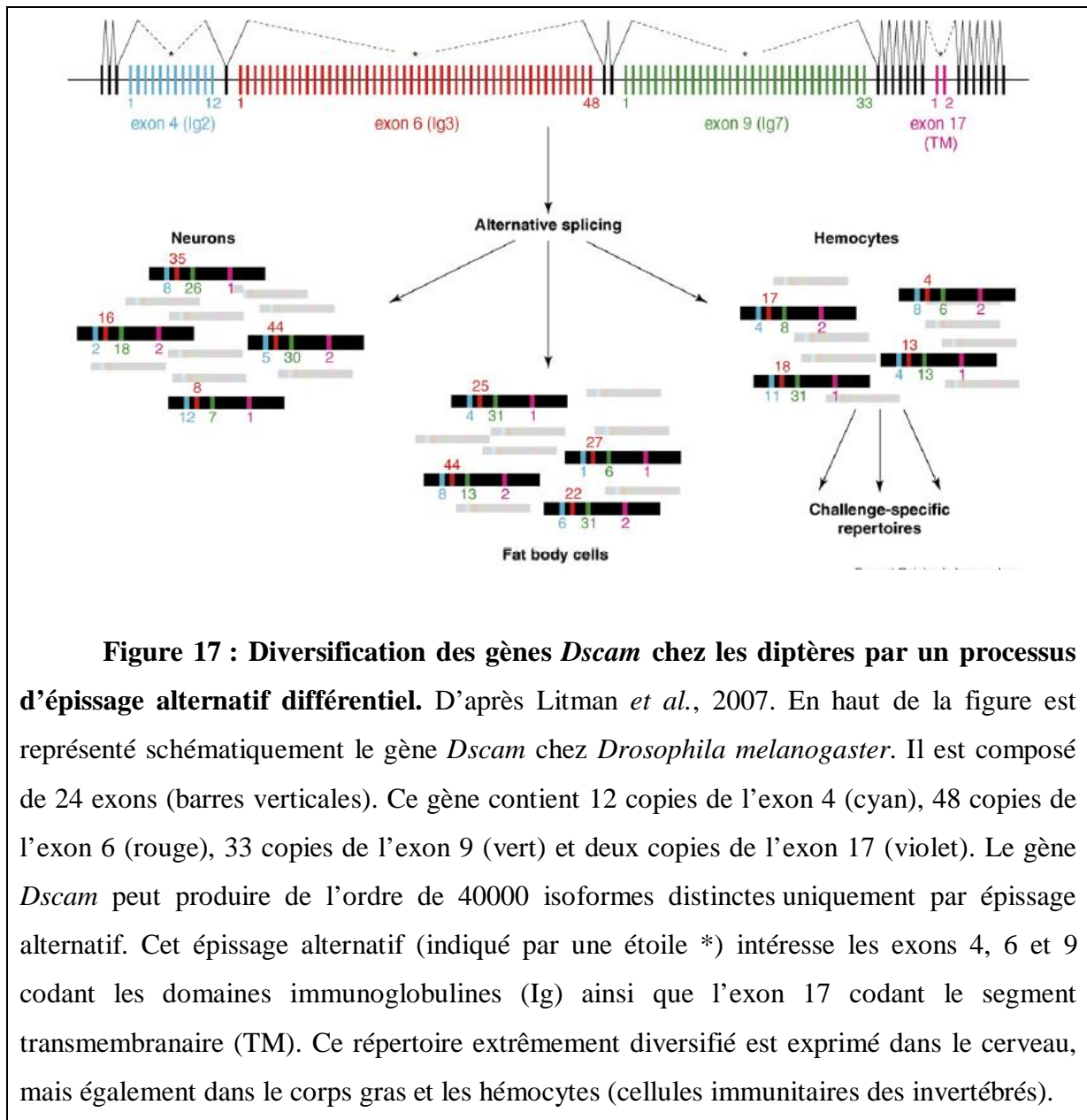
Le système immunitaire inné des insectes est activé dès qu'un patron moléculaire associé au pathogène est reconnu spécifiquement par un PRR. Ces récepteurs vont conduire à l'élimination du pathogène soit directement par des mécanismes tels que l'encapsulation ou

la phagocytose, soit indirectement en activant les voies de signalisation intracellulaire qui contrôle la transcription de gènes effecteurs (De Gregorio *et al.*, 2002; Hoffmann, 2003). Ce système immunitaire inné, malgré un nombre limité de récepteurs ou d'effecteurs (une centaine), parvient à lutter efficacement contre l'extrême diversité des pathogènes rencontrés (Christophides *et al.*, 2002).

Contrairement aux vertébrés, les insectes ne produisent pas d'anticorps. Toutefois, les analyses des génomes de *Drosophila melanogaster* et *Anopheles gambiae* (vecteur du paludisme) ont révélé la présence d'environ 140 à 150 gènes codant des protéines possédant des domaines de type Ig. Certains de ces gènes, le gène *Dscam* (« Down syndrome cell adhesion molecule ») chez *D. melanogaster* et son orthologue *AgDscam* chez *A. gambiae*, ont fait l'objet de nombreuses études (Chen *et al.*, 2006; Schmucker *et al.*, 2000; Vogel *et al.*, 2003; Watson *et al.*, 2005). Ils sont présents en copie unique et codent une molécule transmembranaire qui contient des domaines extracellulaires de type Ig et fibronectine ainsi qu'un domaine cytoplasmique (Schmucker *et al.*, 2000).

Le gène *Dscam*, et de manière similaire son orthologue chez *A. gambiae*, est composé de 24 exons (Figure 17). Quatre exons (4, 6, 9 et 17) ont été dupliqués et sont présents en plusieurs copies (12, 48, 33 et 2, respectivement) organisées en cassette. Les exons 4, 6 et 9 codent les domaines de type Ig de la protéine. Le fait marquant est que ce gène *Dscam* peut produire de l'ordre de 40000 isoformes distinctes par épissage alternatif : ces isoformes contiennent des combinaisons variables de domaines de type Ig (Dong *et al.*, 2006; Watson *et al.*, 2005). Ce répertoire extrêmement diversifié est exprimé non seulement dans le cerveau, mais également dans le corps gras et les hémocytes (cellules immunitaires des invertébrés). Ces études ont mis en évidence une implication des *Dscam* dans l'immunité. En effet, la modification de l'expression du gène codant les *Dscam* chez *D. melanogaster*, par la technique d'ARN interférence, altère la capacité des hémocytes à phagocyter des bactéries. De même, chez *A. gambiae*, à la suite d'une infection par *Plasmodium*, l'extinction du gène *AgDscam* augmente de 65% le nombre de parasites qui traversent avec succès l'intestin moyen (Dong *et al.*, 2006). Les *Dscam* de diptères semblent donc intervenir dans la réponse immunitaire innée, mais de manière spécifique. En effet, chez *A. gambiae*, la réponse à une infection ne se traduit pas par une augmentation globale des transcrits du gène *AgDscam*. Il en résulte plutôt une modification de la représentativité de certains exons variables, au sein du répertoire, par épissage alternatif. De plus, cette modification est spécifique car des infections par différents pathogènes vont induire la production de répertoires différents. Ainsi, suite à une infection, le répertoire va s'enrichir en

récepteurs possédant certains domaines de type Ig ayant une affinité prononcée vis-à-vis des « antigènes » d'un pathogène donné (Dong *et al.*, 2006).



**Figure 17 : Diversification des gènes *Dscam* chez les diptères par un processus d'épissage alternatif différentiel.** D'après Litman *et al.*, 2007. En haut de la figure est représenté schématiquement le gène *Dscam* chez *Drosophila melanogaster*. Il est composé de 24 exons (barres verticales). Ce gène contient 12 copies de l'exon 4 (cyan), 48 copies de l'exon 6 (rouge), 33 copies de l'exon 9 (vert) et deux copies de l'exon 17 (violet). Le gène *Dscam* peut produire de l'ordre de 40000 isoformes distinctes uniquement par épissage alternatif. Cet épissage alternatif (indiqué par une étoile \*) intéresse les exons 4, 6 et 9 codant les domaines immunoglobulines (Ig) ainsi que l'exon 17 codant le segment transmembranaire (TM). Ce répertoire extrêmement diversifié est exprimé dans le cerveau, mais également dans le corps gras et les hémocytes (cellules immunitaires des invertébrés).

Les *Dscam* sont donc des PRR hypervariables essentiels du système immunitaire des diptères. Ils possèdent une plasticité phénotypique leur permettant de s'adapter spécifiquement à différents spectres de pathogènes (Crayton *et al.*, 2006; Graveley *et al.*, 2004). De plus, la diversification de ces récepteurs du système immunitaire des invertébrés est en grande partie générée au niveau somatique. Cette observation est en désaccord avec le paradigme de Medzhitov et Janeway selon lequel la diversité des récepteurs de l'immunité innée serait limitée et codée dans la lignée germinale (Medzhitov, Janeway, 1997). Comme chez les vertébrés, les invertébrés utilisent (i) des récepteurs codés dans la lignée germinale

spécifiques des PAMP et (ii) des récepteurs diversifiés générés par des mécanismes somatiques capables de réagir rapidement au changement antigénique du pathogène (Kurtz, Franz, 2003).

## 1.2. Les VCBP chez *Branchiostoma floridae*

Une famille multigénique codant des protéines contenant des domaines Ig a également été mise en évidence chez l'amphioxus *Branchiostoma floridae*, un céphalocordé ne possédant pas de système immunitaire adaptatif. Ces protéines, nommées VCBP (« Variable Chitin-Binding Protein »), possèdent une région N-terminale variable composée de deux domaines de type Ig-V (domaine protéique variable des immunoglobulines), organisés en tandem. Un domaine de liaison à la chitine est également présent au niveau de la région C-terminale (Cannon *et al.*, 2002). Les VCBP sont solubles et sont exprimées spécifiquement dans l'intestin. Ces protéines se répartissent en cinq familles relativement variables (VCBP 1-5), qui partagent de 27 à 41% de similarités de séquence en acides aminés. Toutefois, les domaines de type Ig-V présents chez les protéines appartenant aux familles VCBP2 et VCBP5 partagent 78% de similarités de séquence en acides aminés (Cannon *et al.*, 2004). Les résultats obtenus jusqu'à présent, ont mis en évidence, au niveau de la région N-terminale variable des protéines VCBP2, une petite région hypervariable de 10 à 13 résidus. La variabilité de cette région s'exprime aussi bien au niveau de sa taille que de sa séquence en acides aminés.

Les cinq familles de VCBP sont codées par cinq familles de gènes (*vcbp* 1-5). L'analyse de la région hypervariable des VCBP2, au niveau génomique et au niveau des transcrits, a été réalisée à partir de différents amphioxus provenant de la même région géographique. Cette analyse révèle un polymorphisme de cette région au niveau intra- et inter-individuel (Cannon *et al.*, 2002; Cannon *et al.*, 2004). Ainsi, chaque individu dispose d'un répertoire VCBP2 unique codé dans la lignée germinale. De plus, la comparaison de ces séquences génomiques aux séquences d'ADN complémentaires, issues d'un même individu, montre qu'elles sont identiques. Il semblerait donc que l'hyperdiversité des VCBP2 ne soit pas dirigée par des mécanismes somatiques. Le mécanisme à la base du polymorphisme génétique des VCBP2 observé reste à élucider.

Les VCBP semblent être des molécules bifonctionnelles. Tout d'abord la région N-terminale présente les caractéristiques des récepteurs de l'immunité adaptative ainsi qu'un domaine hyperdiversifié. Ensuite, le domaine de liaison à la chitine pourrait jouer un rôle

dans l'immunité innée chez les invertébrés (Destoumieux *et al.*, 2000). Ce sont donc des candidates de choix en tant que médiateur alternatif de la reconnaissance immune chez *B. floridae*. De plus, la capacité à maintenir un niveau de diversification suffisant de ces récepteurs représente un avantage face aux stratégies d'évitement des pathogènes. Ainsi, chez les espèces caractérisées par de grandes populations, tel que *B. floridae*, un haut niveau de polymorphisme des récepteurs de l'immunité innée pourrait compenser l'absence de réarrangement somatique.

### 1.3. Les FREP chez *Biomphalaria glabrata*

Les bases moléculaires de la reconnaissance immune ont également été appréhendées au niveau d'un embranchement important des invertébrés, celui des mollusques. Ces études ont été réalisées sur le mollusque gastéropode *Biomphalaria glabrata*, qui est l'hôte intermédiaire obligatoire du trématode parasite *Schistosoma mansoni*. Ce trématode est l'agent responsable de la bilharziose, seconde endémie parasitaire mondiale après le paludisme (Chitsulo *et al.*, 2000) (cf. i-III. *Le modèle S. mansoni / B. glabrata*).

Les travaux menés dans le groupe de Sam Loker ont permis de mettre en évidence des molécules pouvant jouer un rôle dans la défense interne du mollusque (Adema *et al.*, 1997). Il s'agit de protéines solubles et exprimées par les hémocytes. Leur expression est induite spécifiquement en réponse à une infection par un trématode (Adema *et al.*, 1997). En effet, ces protéines ne sont pas présentes dans l'hémolymphe de mollusques sains ou infestés par des bactéries. Elles possèdent également des propriétés de type lectine, leur permettant de précipiter des antigènes solubles issus de trématodes (Adema, Loker, 1997; Hertel *et al.*, 1994). Le précurseur de ces protéines possède un peptide signal, une région N-terminale composée d'un ou deux domaines de type Ig (Ig-1 et Ig-2), et une région C-terminale composée d'un domaine fibrinogène (FBG). En conséquence, ces protéines ont été appelées FREP (« Fibrinogen-Related Proteins ») et présentent une juxtaposition originale de domaines de type Ig et fibrinogène (Adema, Loker, 1997; Leonard *et al.*, 2001). Les domaines de type Ig sont impliqués dans les phénomènes de reconnaissance, de liaison et d'adhésion des cellules. Les plus connus sont ceux du système immunitaire des vertébrés (Du Pasquier, 2000). Plusieurs membres possédant de tels domaines ont été décrits chez d'autres invertébrés (Mendoza, Faye, 1999) : l'hémoline chez *Hyalophora cecropia* (Sun *et al.*, 1990) ou la MDM pour « Molluscan Defense Molecule » chez *Lymnaea stagnalis* (Hoek *et al.*, 1996). Quant aux molécules contenant des domaines fibrinogènes, elles sont présentes chez les vertébrés et invertébrés et peuvent être impliquées dans la réponse immunitaire

innée (Adema *et al.*, 1997; Baker *et al.*, 1990; Dimopoulos *et al.*, 2000; Gokudan *et al.*, 1999; Kurachi *et al.*, 1998; Xu, Doolittle, 1990).

Actuellement, 13 différentes sous-familles de gènes *FREP* ont été identifiés chez *B. glabrata* (Zhang, Loker, 2003; Zhang, Loker, 2004). Les séquences complètes de quatre gènes (*FREP* 2, 3, 4 et 7) sont connues. Les gènes *FREP* 2 et 4 possèdent quatre exons et code des FREP possédant un seul domaine Ig. Le premier exon code un peptide signal, le second et le troisième codent le domaine Ig et le quatrième code le domaine fibrinogène (Leonard *et al.*, 2001). Les gènes *FREP* 3 et 7 présentent une organisation similaire, à ceci près qu'ils possèdent 6 exons, et codent des FREP possédant 2 domaines Ig (Zhang *et al.*, 2001). Chaque sous-famille est répartie au niveau de 1 à 8 loci (2 à 16 allèles) (Zhang, Loker, 2004).

Une étude, ciblée sur les régions codant les domaines de types Ig et FBG de la sous-famille *FREP* 3, montre une grande diversité. Ainsi, 45 séquences différentes ont été isolées à partir d'un seul individu (Zhang *et al.*, 2004). Ce nombre de séquences est en désaccord avec le nombre de loci identifié. Les auteurs suggèrent alors que ce haut niveau de diversification, peut être généré à partir d'un nombre limité de gènes sources (correspondant au nombre de loci) présents dans la lignée germinale. Ils seraient soumis, durant la différenciation des cellules somatiques, à deux types de mécanismes non exclusifs. Tout d'abord, les mutations ponctuelles, qui permettent de générer de nouveaux variants suite à la modification de quelques nucléotides du gène source. Ensuite, des recombinaisons somatiques entre gènes sources, qui conduiraient à de nouveaux gènes mosaïques. Ces processus seraient ainsi capables de générer un niveau de diversité des gènes codant les FREP. A l'heure actuelle, bien que ces mécanismes somatiques soient présents chez les vertébrés (cf. i-1.1. *Hôtes vertébrés : immunité adaptative, polymorphisme et diversification*), ils n'ont jamais été décrits chez les invertébrés.

Chez *B. glabrata*, les FREP apportent une composante « spécifique » au système immunitaire inné. Les mécanismes à la base de la diversification des gènes codant les FREP vont apporter un avantage certain au niveau individuel. En effet, chaque individu est capable d'exprimer une certaine diversité de récepteurs à partir d'un nombre limité de gènes présents dans la lignée germinale.

## 1.4. Conclusion

Ces résultats récents apportent une nouvelle perception de l'immunité innée des invertébrés, qui présente désormais une composante diversifiée analogue au système immunitaire adaptatif des vertébrés. En effet, comme ces derniers, les invertébrés utilisent des récepteurs diversifiés générés par des mécanismes somatiques capables de réagir rapidement au changement antigénique du pathogène (Kurtz, Franz, 2003).

## 2. Pathogènes

Certains parasites eucaryotes (*P. falciparum*, *T. brucei*, *S. mansoni*) responsables de maladies humaines, sont dépendant d'hôtes invertébrés (diptères, mollusques) au cours de leur cycle de vie. Cependant, bon nombre des recherches se sont focalisées jusqu'alors sur l'interaction entre ces pathogènes et leur hôte vertébré. Ceci explique le décalage important des connaissances acquises concernant leur interaction avec l'hôte vertébré et invertébré. Comme nous l'avons mis en évidence dans la première partie de cette introduction (cf. i-I. *Interaction hôte vertébré / pathogène*), les interactions hôte vertébré / pathogène sont très bien référencées. Les modèles protozoaires parasites, tel que *P. falciparum* ou *T. brucei* par exemple, ont permis de mettre en évidence les stratégies de variation antigénique (gènes *var* et *vsg*, respectivement) en réponse à la pression de sélection imposée par le système adaptatif de l'hôte vertébré. Les molécules polymorphes et/ou diversifiées au coeur de cette interaction sont désormais bien connues et répondent aux modèles théoriques de co-évolution (cf. i-I.3. *Conclusion hôte vertébré / pathogène*).

En revanche, les données concernant l'interaction avec les hôtes invertébrés restent fragmentaires et insuffisantes pour permettre de vérifier les modèles évolutifs. En effet, des molécules polymorphes chez certains invertébrés ont récemment été mises en évidence (cf. i-II.1 *Hôtes invertébrés : immunité innée et diversification*). Toutefois, aucune molécule polymorphe et/ou diversifiée n'a été caractérisée du côté « pathogène ».

Il est donc nécessaire à l'heure actuelle d'appréhender expérimentalement la réponse du pathogène face au potentiel de diversification des récepteurs immunitaires de certains invertébrés. En effet, est-ce que le pathogène est capable de générer, lui aussi, une plasticité phénotypique de ces antigènes ? Possède-t-il des antigènes polymorphes et/ou diversifiés ? Les réponses à ces questions permettraient ainsi de valider ou d'invalider les modèles théoriques, tel que celui de la « Reine Rouge », dans les systèmes hôte invertébré / pathogène.



### III. LE MODELE *S. MANSONI* / *B. GLABRATA*

Afin de vérifier le modèle théorique de la « Reine Rouge », appliqué aux systèmes hôte invertébré / parasite, il est nécessaire de disposer d'un système où la dynamique co-évolutive est accessible expérimentalement. C'est le cas de certains modèles hôte invertébré / parasite dans lesquels les mécanismes de défense de l'hôte et d'infektivité du parasite sont à l'équilibre en population naturelle. Il en résulte un polymorphisme de compatibilité qui se traduit par la présence de certains phénotypes hôte et parasite compatible, d'autres étant incompatibles. C'est le cas dans l'interaction entre le trématode parasite *Schistosoma mansoni* et son hôte intermédiaire mollusque *Biomphalaria glabrata* (cf. i-III.1. *Généralités*). Le polymorphisme de compatibilité, caractéristique de cette interaction (cf. i-III.2. *Caractéristiques de l'interaction mollusque / schistosome*), a été exploité pour sélectionner dans le laboratoire des souches de parasites compatible et incompatible (cf. i-III.3. *Le modèle d'étude*). Elles représentent des outils expérimentaux de choix pour mener à bien nos investigations par des approches moléculaires comparatives. De plus, dans ce modèle, comme nous l'avons mis en évidence par ailleurs (cf. i-II.1.3. *Les FREP chez Biomphalaria glabrata*), l'hôte *B. glabrata* présente une variabilité phénotypique de certains récepteurs immunitaires (FREP). Ainsi, dans l'hypothèse d'une course aux armements, les mécanismes de diversification présents chez l'hôte intermédiaire, devraient se retrouver chez le parasite.

#### 1. Généralités

Appartenant à la famille des Schistosomatidae (embranchement des Plathelminthes, classe des Trématodes) les espèces du genre *Schistosoma* sont des vers parasites à large impact médical et vétérinaire. Parmi les 19 espèces du genre *Schistosoma*, trois espèces (*Schistosoma mansoni*, *S. haematobium* et *S. japonicum*) sont responsables d'une pathologie chronique humaine, la bilharziose (ou schistosomose). Il s'agit de la seconde endémie parasitaire mondiale après le paludisme. Plus de 600 millions de personnes sont exposées, 200 millions sont touchées et 5% des cas présentent des complications cliniques sévères, dans 74 pays des régions tropicales et subtropicales du globe (Chitsulo *et al.*, 2000). *S. mansoni* (agent de la bilharziose intestinale) est présent principalement en Afrique et Amérique du Sud ; *S. haematobium* (agent de la bilharziose urinaire) en Afrique et au moyen Orient ; et *S. japonicum* (agent de la bilharziose hépto-intestinale) est endémique de l'Asie.

## 1.1. Cycle de vie de *S. mansoni*

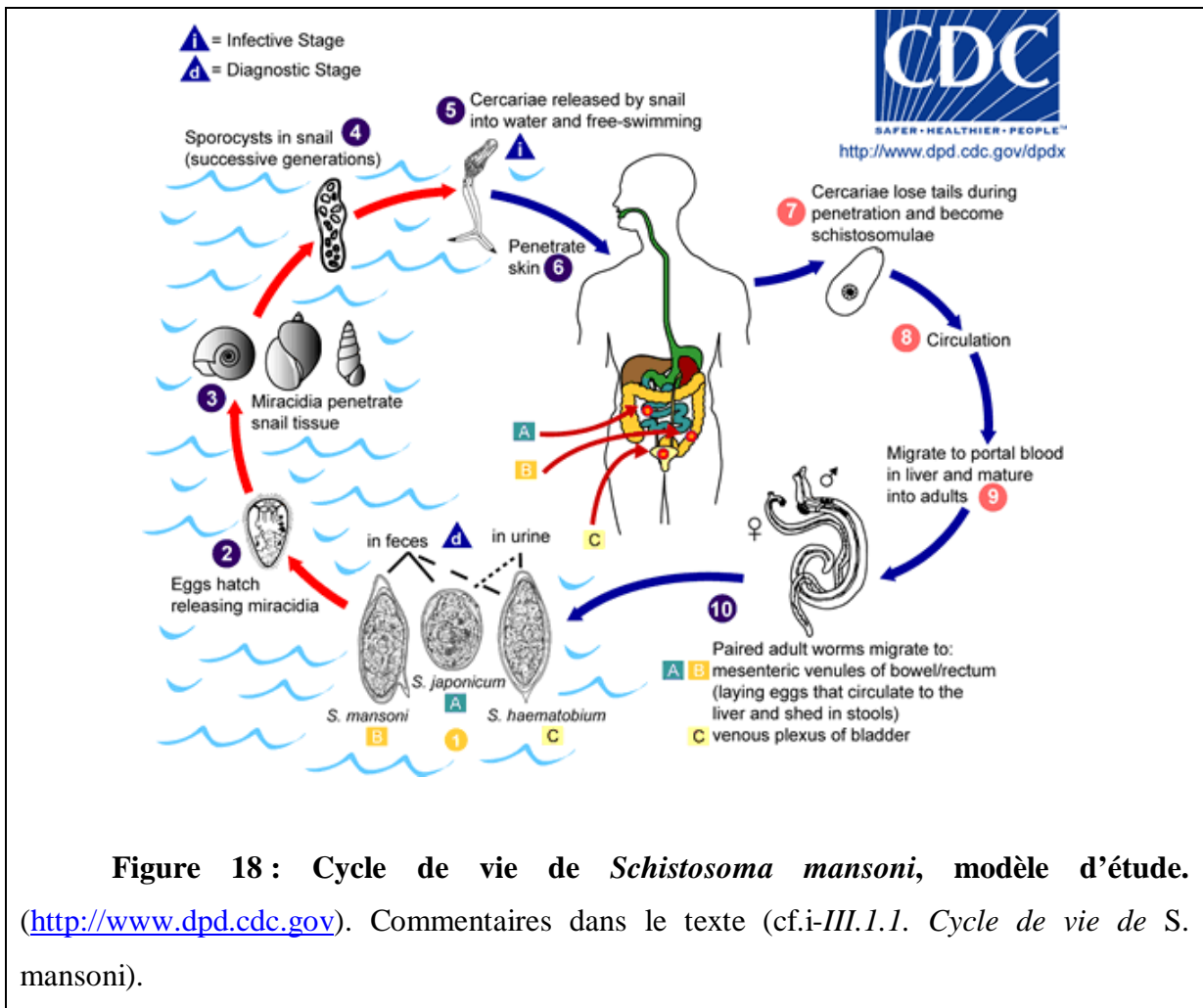
Les parasites du genre *Schistosoma* ont un cycle de vie complexe (Figure 18) (Ross *et al.*, 2002). Ils utilisent un mollusque (3) d'eau douce comme hôte intermédiaire (*Biomphalaria* spp. vecteur de *S. mansoni* (B); *Bulinus* spp. vecteur de *S. haematobium* (C) et *Oncomelania* spp. vecteur de *S. japonicum* (A)) chez lequel ils se multiplient intensément de manière asexuée. Cette association est spécifique et indispensable au développement du parasite. Les larves nageantes (cercaires, 5) émises par les mollusques sont directement infestantes pour l'hôte définitif vertébré (6) chez lequel elles se développent en vers adultes (7, 8, 9 et 10) dans le système veineux porte mésentérique (A, B) ou vésical (C). Caractérisés par un fort dimorphisme sexuel, les schistosomes adultes s'accouplent. Les femelles pondent alors des œufs, principaux responsables des complications pathologiques. Une partie des œufs pondus par les femelles est piégée dans les tissus de l'hôte, l'autre partie est évacuée vers le milieu extérieur par les fèces ou l'urine (1). Les œufs atteignent le milieu aquatique et libèrent une larve libre et nageante (miracidium, 2). Cette larve infeste un nouveau mollusque hôte intermédiaire (3) et s'y transforme en sporocyste (stade intra-molluscal, 4). Chez *S. mansoni* il y a deux générations de sporocystes : les sporocystes primaires (Sp1 ou sporocystes mères) généralement présent au niveau des tissus périphériques du pied de l'hôte et les sporocystes secondaires (Sp2 ou sporocystes filles) généralement présent au niveau de la glande digestive. Les Sp2, produit par les Sp1, sont capables de produire non seulement des cercaires mais également des générations supplémentaires de sporocystes. La pénétration et le développement d'un miracidium dans le mollusque conduit à la production de plusieurs milliers de cercaires (Theron, 1986).

## 2. Caractéristiques de l'interaction mollusque / schistosome

### 2.1. Spécificité d'hôte

Les interactions mollusque / schistosome, comme les interactions mollusque / trématode en général, se caractérisent par une forte spécificité d'hôte. En effet, chaque espèce de parasite n'est capable de se développer que dans une seule famille (voir même une seule espèce au sein de cette famille) d'hôte mollusque (Adema, Loker, 1997) : plusieurs espèces du genre *Biomphalaria* (*B. glabrata*, *B. alexandrina*, *B. pfefferi*, *B. tenagophila*, etc ...) peuvent servir d'hôte intermédiaire à *Schistosoma mansoni* (Campbell *et al.*, 2000). Un certain nombre d'explications sont proposées pour interpréter cette caractéristique des

associations mollusque / trématode. Tout d'abord, le comportement du miracidium, qui ne peut se fixer et pénétrer spécifiquement que chez certains hôtes (Haas *et al.*, 1991). Ensuite, un deuxième filtre intervient lorsque le miracidium a pénétré et qu'il se trouve dans l'environnement interne du mollusque. Les conditions spécifiques de ce milieu (osmolarité, pH, température, oxygène, nutriments) peuvent ne pas être favorable au développement intramolluscal du parasite (« unsuitability », (Sapp, Loker, 2000)). Enfin, un dernier filtre intervient lorsque le parasite s'installe chez le mollusque. Le devenir du parasite dépend de ces capacités à surmonter le système de défense interne de l'hôte. De ce fait, l'incapacité de générer une réponse adaptée peut réduire le spectre d'hôtes du parasite (Adamson, Cairn, 1994). Cette spécificité est très importante puisqu'elle va permettre de limiter les espèces trématodes d'intérêt médical et vétérinaire aux aires géographiques occupées par l'hôte intermédiaire compatible (Van Der Knaap, Loker, 1990).



**Figure 18 : Cycle de vie de *Schistosoma mansoni*, modèle d'étude.** (<http://www.dpd.cdc.gov>). Commentaires dans le texte (cf.i-III.1.1. Cycle de vie de *S. mansoni*).

## 2.2 Multiplication clonale intense

Chez le schistosome, si la durée de vie maximale des stades infestant dans le milieu extérieur peut atteindre 20 heures, l'infectivité des *miracidia* ou des cercaires n'est que de quelques heures. Cette courte infectivité limite les possibilités de dispersion du parasite. Cette dispersion est également limitée par les capacités du mollusque. Seulement une faible proportion de la population est parasitée (1%) et la charge parasitaire par mollusque est faible (1,1 génotype de parasite larvaire *S. mansoni*) (Theron *et al.*, 2004). En revanche, si l'interaction avec le mollusque limite les capacités de dispersion du parasite, la phase de multiplication asexuée intense chez cet hôte a toute son importance d'un point de vue démographique. Elle permet, en effet, de compenser les pertes énormes subies par le parasite lors de chaque étape de changement d'hôte. En effet, en moyenne seulement six cercaires produites par un mollusque parasité parviennent à atteindre un hôte définitif alors que des milliers sont produites quotidiennement (Theron, 2005).

Cette phase d'intense multiplication clonale ajoutée à cette spécificité d'hôte font du mollusque un facteur clé de la transmission du parasite. Ainsi, cet hôte intermédiaire est un acteur essentiel de la distribution et de l'expansion de la parasitose en milieu naturel.

## 2.3. Le polymorphisme de compatibilité

Cette interaction spécifique entre *S. mansoni* et *B. glabrata* est le résultat d'un « dialogue subtil » entre les mécanismes de défense de l'hôte et d'infectivité et/ou de virulence du parasite. Dans ce modèle ces mécanismes sont à l'équilibre conduisant ainsi, en population naturelle, à un polymorphisme de compatibilité. Ce polymorphisme se traduit par la présence de certain phénotype hôte et parasite compatible. A l'heure actuelle, nous ne disposons sur le plan fonctionnel que de peu de connaissance pour expliquer le maintien et l'évolution de ce polymorphisme en population naturelle. Plusieurs hypothèses sont souvent avancées, telle que la théorie de la « Reine Rouge » (Van Valen, 1973) ou encore la théorie des « trade-offs » coût/bénéfice (Webster, Davies, 2001). Ces théories restent largement du domaine conceptuel en particulier dans les modèles impliquant des métazoaires parasites comme les modèles schistosome / mollusque.

Cependant, quelques études récentes sur le modèle *S. mansoni* / *B. glabrata* tentent de valider de manière empirique ces différents modèles théoriques. En effet, Webster et collaborateur (Webster *et al.*, 2004; Webster *et al.*, 2007), mettent en évidence que architecture et variabilité génétique sont présentes chez l'hôte et le parasite. De plus, ils sont

soumis aux contraintes sélectives qu'ils exercent l'un sur l'autre. Ces éléments sont essentiels pour permettre l'évolution de certains traits, tels que la résistance / susceptibilité, la virulence ou encore l'infectivité. Le polymorphisme génétique de la compatibilité serait alors maintenu par des compromis (« trade-offs ») coût-bénéfice (Webster, Davies, 2001) entre résistance / susceptibilité, virulence et infectivité. Tout d'abord, un coût de la résistance a été suggéré chez le mollusque (Coustau *et al.*, 2000). Il se traduit par une fitness du génotype résistant inférieure à celle du génotype sensible, en absence de parasite, permettant ainsi un maintien du polymorphisme de compatibilité. Ensuite, chez le parasite, l'infectivité de l'hôte mollusque et la production cercarienne sont corrélées négativement à l'infectivité de l'hôte définitif et la production miracidiale. Deux conditions expliqueraient le « trade-off » du succès reproducteur du parasite entre l'hôte intermédiaire et définitif. Premièrement, un coût de l'infectivité au niveau de l'hôte intermédiaire. Il empêcherait la fixation des allèles responsables de ce trait (Frank, 1993) et déterminerait également dans certaines situations le niveau d'infection optimal (Davies *et al.*, 2001). Deuxièmement, un bénéfice de la virulence (multiplication clonale), qui est corrélée positivement au succès de la transmission à l'hôte définitif. Une sélection fluctuante, par laquelle différents génotypes ou différentes allèles pourraient être favorisés dans les différents hôtes obligatoires, maintiendrait ainsi un polymorphisme génétique et phénotypique de l'infectivité et de la virulence dans ce système. Un élément supplémentaire à prendre en compte est la compétition intra-hôte (entre différents parasites au sein d'un même hôte). En effet, cette dernière va favoriser les génotypes avirulents, contrairement au succès de la transmission à l'hôte définitif qui favorise les génotypes virulents, concourant ainsi au maintien du polymorphisme en population naturelle (Gower, Webster, 2005).

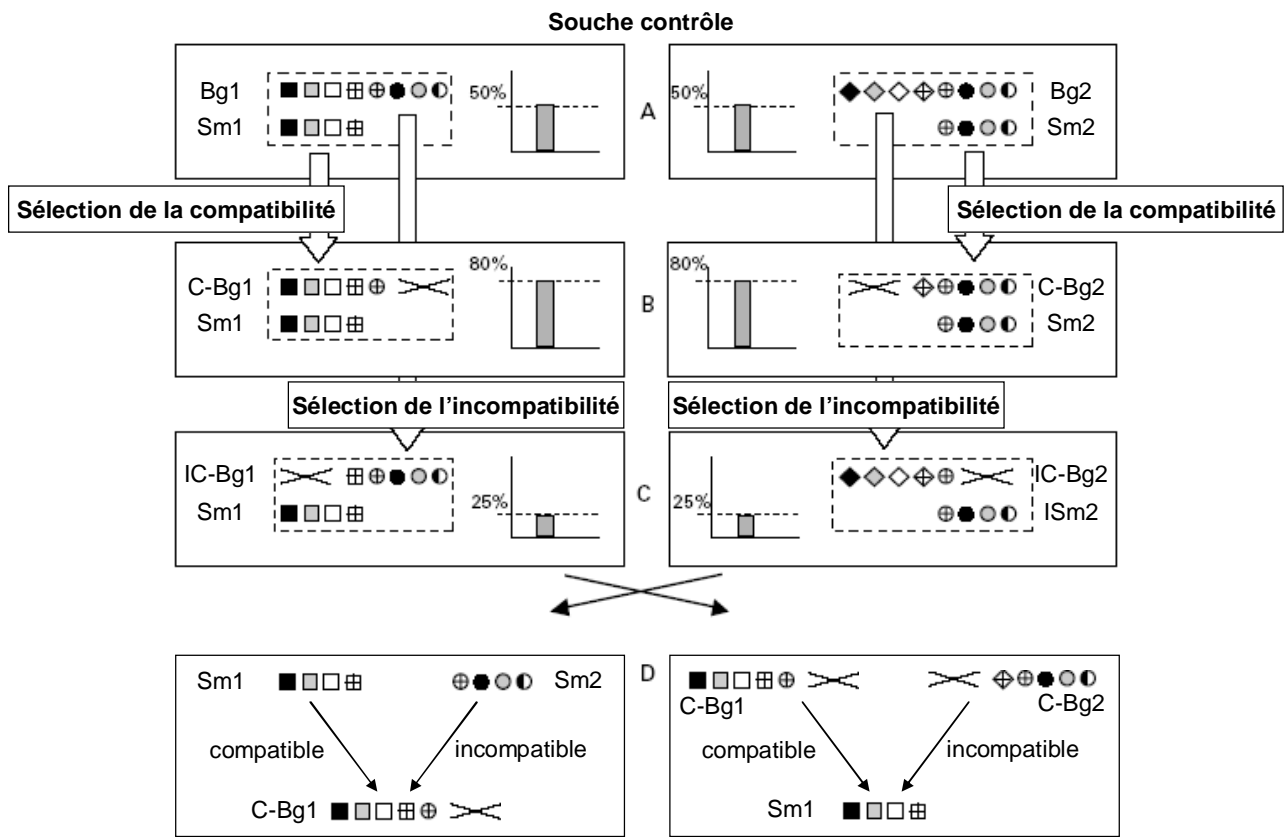
Les résultats concernant les « trade-offs » coût-bénéfice représentent des pistes intéressantes quant à la compréhension des processus qui sous-tendent le polymorphisme de compatibilité. Cependant d'autres hypothèses avancent que le polymorphisme de compatibilité n'est que le reflet de la dynamique co-évolutive de certains systèmes hôte / parasite à l'équilibre où au sein d'une population seuls certains individus hôtes ou parasites sont compatibles. Il s'agit de l'hypothèse de concordance des phénotypes (« matching phenotype ») proposée dans l'interaction *S. mansoni* / *B. glabrata*. La compatibilité serait testée indépendamment pour chaque individu parasite et hôte, et le succès ou l'échec de l'infection dépendrait de la concordance (« matched ») ou de la non concordance (« mismatched ») des phénotypes de l'hôte et du parasite (Theron, Coustau, 2005).

## 2.4. Le modèle de concordance des phénotypes et sélection des souches

Afin de tester la compatibilité dans le modèle *S. mansoni* / *B. glabrata*, des travaux ont été menés dans notre laboratoire (Theron, Coustau, 2005). Notamment, des coupes histologiques ont été réalisées 48 heures après l'infestation d'un mollusque par plusieurs *miracidia* de *S. mansoni*. Les observations révèlent la présence de sporocystes mère développés mais également encapsulés (par les cellules hémocytaires) au sein du même mollusque. Ces résultats, en accord avec Basch (Basch, 1975), suggèrent que la compatibilité serait testée de façon indépendante pour chaque miracidium lors de l'interaction avec un même hôte. Dans ce cas, le succès ou l'échec de l'infection du mollusque, qui se traduit respectivement par la survie ou l'encapsulation du parasite, résulterait principalement de la capacité ou incapacité d'un miracidium donné à éviter le système immunitaire de l'hôte qu'il pénètre. Ainsi, la réaction de défense de l'hôte dirigée contre le miracidium non infestant ne serait plus envisagée comme un mécanisme de résistance spécifique contre *S. mansoni* mais plutôt comme une réponse de défense générale contre un organisme vivant, reconnu comme non-soi.

A l'heure actuelle, la nature des gènes et des produits impliqués dans l'interaction *B. glabrata* / *S. mansoni* restent inconnus. Cependant, l'hypothèse selon laquelle les sporocystes échappent à la réponse hémocytaire en exprimant des épitopes (probablement des glycoprotéines) qui miment les molécules de l'hôte est largement acceptée (Adema, Loker, 1997). La compatibilité dans l'interaction *S. mansoni* / *B. glabrata* peut être alors définie comme une concordance entre les molécules de l'hôte et du parasite impliquées respectivement dans les mécanismes de reconnaissance et de mimétisme. Un miracidium peut être infestant ou non infestant si son phénotype est respectivement concordant ou non concordant avec le phénotype de l'hôte. En d'autres termes, le succès ou l'échec de l'infestation dépendrait de la concordance (« matched ») ou de la non concordance (« mismatched ») des phénotypes de l'hôte et du parasite. Un tel mécanisme est compatible avec un modèle de concordance des phénotypes (« matching phenotype model», Figure 19) ainsi qu'avec l'existence d'un polymorphisme de compatibilité dans le modèle *S. mansoni* / *B. glabrata* (Theron, Coustau, 2005).

Une représentation du modèle de concordance des phénotypes (Figure 19) illustre la sélection de couples d'hôte / parasite compatibles et incompatibles, utilisés dans nos investigations. Les souches de *B. glabrata* de laboratoire sont, en moyenne, compatibles à 50% vis-à-vis de leurs souches sympatriques de *S. mansoni* (Webster, Woolhouse, 1998).



**Figure 19 : Modèle de concordance des phénotypes (« matching phenotype model »).** D'après Theron, Coustau, 2005, modifié. Représentation schématique de la sélection de couples compatibles et incompatibles à partir de deux souches d'isolats naturels de *Biomphalaria glabrata* (Bg1 et Bg2) et *Schistosoma mansoni* (Sm1 et Sm2) sous l'hypothèse de « matching phenotype ». Les symboles représentent la diversité phénotypique représentative des mollusques et des parasites. **(A)** Souches contrôles de mollusques (Bg1 et Bg2) dont 50% des phénotypes concordent respectivement avec les phénotypes des souches de parasites Sm1 et Sm2. **(B)** Sélection de couples compatibles. **(C)** Sélection de couples incompatibles. **(D)** Obtention de modèles adaptés à nos investigations : à gauche, deux souches de parasite, l'une compatible (Sm1) et l'autre incompatible (Sm2) vis-à-vis de la même souche de mollusque (C-Bg1) ; à droite deux souches de mollusque l'une compatible (C-Bg1) et l'autre incompatible (C-Bg2) vis-à-vis de la même souche de parasite (Sm1).

Ainsi, selon le modèle de « matching phenotype », cela revient à dire que la souche de parasite possède une variété de phénotype qui concorde seulement avec la moitié des phénotypes présents chez la souche d'hôtes (Figure 19A). La sélection de couples compatibles consiste en l'élimination des phénotypes de la souche de mollusque non concordant avec les phénotypes de la souche de parasite (Figure 19B). En revanche, la sélection de couples incompatibles consiste en l'élimination des phénotypes de la souche de mollusque concordant avec les phénotypes de la souche de parasite (Figure 19C). Il est alors possible, à partir de ces couples d'hôte / parasite compatibles ou incompatibles, de disposer de modèles d'études tout à fait intéressant, comme par exemple, des souches de parasites compatible et incompatible vis-à-vis de la même souche de mollusque (Figure 19D).

### **3. Le modèle d'étude**

Le polymorphisme de compatibilité, qui caractérise l'interaction *S. mansoni* / *B. glabrata*, a permis de sélectionner au laboratoire deux souches de *S. mansoni*, une compatible (souche C) et une incompatible (souche IC), vis-à-vis de la même souche de mollusques *B. glabrata*. Ces souches de *S. mansoni* représentent des outils expérimentaux de choix pour mener à bien nos recherches. En effet, il est possible de développer des approches moléculaires comparatives entre sporocystes primaires des souches compatible et incompatible. Ces approches comparatives ont pour objectif d'approcher les bases moléculaires du polymorphisme de compatibilité dans l'interaction *S. mansoni* / *B. glabrata*. Les travaux réalisés dans le cadre de cette thèse vont être maintenant exposés dans les trois chapitres suivants. Ils se sont attachés (i) à mettre en évidence des molécules parasitaires clés au cœur cette interaction (cf. *CHAPITRE 1*), (ii) à caractériser ces molécules et leur niveau de polymorphisme et/ou de diversification (cf. *CHAPITRE 2*) et (iii) à identifier les bases génétiques à l'origine de ce polymorphisme (cf. *CHAPITRE 3*).



# CHAPITRE 1

## Déterminants moléculaires du polymorphisme de compatibilité dans l'interaction *Biomphalaria glabrata* / *Schistosoma mansoni* : identification de nouveaux candidats par une approche protéomique comparative globale

### Publication n°1

Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: new candidates identified by a global comparative proteomics approach. **Molecular & Biochemical Parasitology** (2008);157:205-16.

Les travaux réalisés dans le cadre de cette thèse se sont attachés à étudier l'interaction hôte invertébré / parasite, avec comme modèle d'étude l'interaction entre le mollusque gastéropode *Biomphalaria glabrata* et le trématode parasite *Schistosoma mansoni*. Cette interaction se caractérise par un polymorphisme de compatibilité, en milieu naturel, qui a permis la sélection de souches de *S. mansoni* compatible (C) et incompatible (IC) vis-à-vis de la même souche de *B. glabrata*.

Afin de caractériser cette interaction *in vivo*, nous avons réalisé une étude histologique, chez le mollusque, du processus d'encapsulation du stade infestant, compatible et incompatible. Cette étude révèle que dans le cas d'une combinaison compatible, le parasite pénètre et se développe normalement chez le mollusque. En revanche, dans le cas d'une combinaison incompatible, le parasite pénètre mais est reconnu comme non soi (une heure post-infestation). Il est alors encapsulé et détruit par le système de défense interne du mollusque 8 à 10 heures post-infestation.

Ainsi, afin d'identifier les déterminants moléculaires potentiellement impliqués dans cette compatibilité différentielle, nous avons développé une approche protéomique comparative globale entre sporocystes primaires (stade intramolluscal) C et IC. L'analyse qualitative et quantitative des profils protéiques, a permis de mettre en évidence des protéines différentiellement représentées entre sporocystes des deux souches. Ces protéines peuvent être réparties en trois grands groupes : (i) des enzymes anti-oxydantes, (ii) des protéines du métabolisme primaire et (iii) des protéines polymorphes de type mucine

possédant un domaine composé d'un nombre variable d'unités répétées en tandem (VNTR « Variable Number Tandem Repeat »). Ce dernier groupe revêt un intérêt particulier car les mucines sont impliquées dans les mécanismes d'évasion immune, d'invasion de l'hôte ou encore dans la protection du parasite. Cela suggère que ces nouvelles protéines de type mucine de *S. mansoni* pourraient être un facteur déterminant de la compatibilité dans l'interaction *S. mansoni* / *B. glabrata*.

## Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: New candidates identified by a global comparative proteomics approach<sup>☆</sup>

Emmanuel Roger, Guillaume Mitta, Yves Moné, Anne Bouchut, Anne Rognon, Christoph Grunau, Jérôme Boissier, André Théron, Benjamin E.F. Gourbal<sup>\*</sup>

*Parasitologie Fonctionnelle et Evolutive, UMR 5244, CNRS, EPHE, UPVD, Biologie & Ecologie Tropicale et Méditerranéenne, Université de Perpignan, 52 Ave Paul Alduy, 66860 Perpignan Cedex, France*

Received 20 July 2007; received in revised form 23 October 2007; accepted 2 November 2007

Available online 17 November 2007

### Abstract

The co-evolutionary dynamics that exist in host–parasite interactions sometimes lead to compatibility polymorphisms, the molecular bases of which are rarely investigated. To identify key molecules that are involved in this phenomenon in the *Schistosoma mansoni*/*Biomphalaria glabrata* model, we developed a comparative proteomics approach using the larval stages that interact with the invertebrate host. We used qualitative and quantitative analyses to compare the total proteomes of primary sporocysts from compatible and incompatible parasite strains. The differentially expressed proteins thus detected belong to three main functional groups: (i) scavengers of reactive oxygen species, (ii) components of primary metabolism, and (iii) mucin-like proteins. We discuss the putative roles played by these protein families as determinants of compatibility polymorphism. Since mucins are known to play key roles in the host–parasite interplay, we consider the newly discovered *S. mansoni* mucin-like proteins (*SmMucin*-like) as the most promising candidates for influencing the fate of host–parasite interactions. An analysis of their expression is presented in a paper published in the same journal issue.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** *Biomphalaria glabrata*; *Schistosoma mansoni*; Compatibility polymorphism; Proteomics; Mucin-like

### 1. Introduction

Schistosomiasis or bilharzia is a tropical disease caused by worms of the genus *Schistosoma*. The main disease-causing species are *S. haematobium*, *S. japonicum*, and *S. mansoni*. An estimated 200 million people in 74 countries suffer from schistosomiasis [1,2]. The World Health Organization expert committee (WHO Technical Report Series 912: prevention and control of schistosomiasis and soil transmitted helminthiasis (WHO, Geneva, 2002)) concluded that yearly deaths could be as high as 200,000, making schistosomiasis the most severe tropical disease after malaria in terms of morbidity [1].

The life cycle of the parasite requires contamination of surface water by excrement, specific freshwater snails as intermediate hosts, and human to water contact. The snail hosts have attracted significant attention in the past, due to their medical and epidemiological importance. In addition, the interaction between snail and schistosome has been extensively studied as a model in evolutionary biology. This host–parasite relationship is a model of choice for the study of co-evolution or reciprocal adaptation [3,4]. Host parasite co-evolution is driven by the reciprocal evolution of host resistance and parasite infectivity and or virulence [4]. To study and understand this phenomenon, it is necessary to work with a host–parasite model in which such co-evolutionary dynamics exist. This is the case in our model that involves *Schistosoma mansoni*, the main schistosome species infecting humans in Africa, the Arabian Peninsula, and South America, and *Biomphalaria glabrata*, its intermediate snail host in the New World. In the *S. mansoni*/*B. glabrata* interaction, a compatibility polymorphism was observed [5]: in natural populations some snail/schistosome combinations are compatible

<sup>☆</sup> Note: Nucleotide sequence data reported in this paper are available in the GenBank™ database under the accession numbers: EU042595, EU042596, EU042597, and EU042598.

<sup>\*</sup> Corresponding author. Tel.: +33 4 68 66 20 50; fax: +33 4 68 66 22 81.  
E-mail address: [benjamin.gourbal@univ-perp.fr](mailto:benjamin.gourbal@univ-perp.fr) (B.E.F. Gourbal).

cleavages per peptide were allowed and some variable modifications were taken into account in the search such as carbamidomethylation of cysteine, oxidation of methionine, histidine or tryptophan, and phosphorylation of serine, threonine or tyrosine. Searches were performed without constraining protein molecular weight or isoelectric point. A protein was considered to be correctly identified if at least two peptides were confidently matched with database sequences. An overall MASCOT score of greater than 50 was considered significant ( $p < 0.05$ ) [19].

#### 2.4. RNA extraction, reverse transcription and quantitative real-time PCR

Total RNA extractions from Sp1 were performed using the Trizol Reagent (Invitrogen) and the manufacturer's protocol. Reverse transcription was performed according to previously described procedures [20]. Real-time PCR analysis of gene expression was done on cDNAs coding for the proteins identified in the proteomic approach as described before [20]. Primers for real-time quantitative PCR were designed with the Light Cycler Probe Design™ version 1.0. Primers for ornithine amino-transferase (OAT), superoxide dismutase Cu/Zn (SOD), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), myosin light chain (MyoLC), proteasome alpha subunit ( $\alpha$ -SubProt), ribosomal protein P2 (Ribo-P2) and for the 28 s ribosomal gene (Ribo s28) used as a control, are shown in Table 1.

#### 2.5. Gene characterization and protein primary structure analysis

ESTs (GenBank dbEST of *S. mansoni* 162,839 sequences) identified from MASCOT analysis were assembled with Sequencher™ (Gene Codes Corporation). Complete CDSs were found for all proteins identified in the comparative analysis except the mucin-like proteins [21].

To characterize potential polymorphisms and difference between strains in the ornithine amino-transferase (OAT) and creatine kinase (CK) genes, complete CDSs in Sp1 of the C and IC strains were amplified using the following primers (OATCDSF1 AAATCACCGCACAGATTGTTCAAG, OATCDSRI GCATAGTTGTATGTTTTTTCCTGTG, CKCDSF1 GTGCACAATCAGAATGCAGGTTG, and CKCDSRI ATATGTACCATCAGGGGATTCTG). PCR products were ligated into the pCR2.1-TOPO™ vector according to the manufacturer's instructions (Invitrogen). Plasmids were sequenced with M13 forward and reverse primers using a dideoxy-dye-terminator method (CEQ™ DTCS-Quick Start kit, Beckman Coulter) and a CEQ™ 8000 capillary sequencer (Beckman Coulter).

Based on complete CDSs, primary structure analyses were performed using ProtFun 2.2 (<http://www.cbs.dtu.dk/services/ProtFun/>) and SignalP 3.0 (<http://www.cbs.dtu.dk/services/SignalP/>). Specific active domain searches were done using ScanProsite (<http://www.expasy.ch/tools/scanprosite/>) and MotifScan ([http://myhits.isb-sib.ch/cgi-bin/motif\\_scan](http://myhits.isb-sib.ch/cgi-bin/motif_scan)). Expected molecular weights (MW) and isoelectric points (pI) were calculated using Scansite (<http://scansite.mit.edu/>

[calc.mw.pi.html](http://calc.mw.pi.html)) (see Table 2). When necessary multiple alignments were realized using ClustalW 1.7 (<http://www.ebi.ac.uk/clustalw/>).

### 3. Results

#### 3.1. Histological study of the encapsulation process

We observed that miracidia of compatible (C) and incompatible (IC) strains of *S. mansoni* penetrate the snail epithelium in a similar manner (same number of larvae, same speed of entry). Nevertheless, drastic differences were evident between C and IC strains in the first hour after penetration. Parasites from the IC strain were immediately recognized by haemocytes that were in contact with the surface of the parasites 1 h post-infection (data not shown) and sporocysts were entirely encapsulated 4–8 h post-infection. At this stage, sporocysts were clearly degraded (Fig. 1B). Most sporocysts lacked tegumental cytoplasm. Germinal cells and other internal structures showed extensive pathological changes (increase of intercellular spaces and pycnotic nuclei). In contrast, miracidia of the C strain remained unaffected and did not undergo encapsulation, no haemocytes were observed close to the sporocysts. Normal developing Sp1 structures were observed (Fig. 1A). In addition, miracidia of the C strain attained deeper regions in the organs of *B. glabrata* than did the IC strain. It is apparent that fast encapsulation of IC strain miracidia after penetration prevents larvae from moving further into the snail tissues. C strain parasites then pursued normal growth and development (data not shown). Our results suggest constitutive antigenic differences between strains that allow immediate recognition, encapsulation and killing of parasites of the IC strain. These antigenic differences were investigated in the global comparative proteomics approach developed here.

#### 3.2. Global comparative proteomics approach

Proteomic studies have been used to study molecules involved in the host–parasite interplay in *Schistosoma*. Most of these previous studies were done on adult worms interacting with the vertebrate host [22–24]. Here we compared Sp1 proteomes from the *S. mansoni* C and IC strains. Five independent samples for each strain were separated on 2D gels and representative gels are shown in Fig. 2. Samples were compared using the 2D gel analysis software package PD Quest and spots with more than twofold differences in intensity were examined individually. Spots of interest were excised from the gels and identified by mass spectrometry. The results are listed in Table 2; for both strains, names of proteins identified in databases, species and Genbank accession numbers are shown. The number and sequences of identified peptides, the MASCOT score and the coverage of the protein are also reported. Where possible, theoretical values for molecular weight (MW) and isoelectric point (pI) were calculated using the Scansite program and these values were compared to the position on the gel.

Our qualitative and quantitative analyses showed that patterns and/or intensities of 23 spots were different between Sp1 C and

Table 1  
Primers used for quantitative real-time PCR

Gene name	Primer forward	Primer reverse
Ornithine amino-transferase (OAT)	TGCCCGGATTTGTAC	ACAGCGGATACTGGAAG
Superoxide dismutase Cu/Zn (SOD)	AGTGGACTCAAGGCTG	CCACGGCCTAAATCAT
Glyceraldehydes-3-phosphate dehydrogenase (GAPDH)	GCGAGGTTTCGACTGA	AACAACGAACATGGGTG
Myosin light chain (MyoLC)	ATGGTCGAGATGGTATGA	AGCATAGGTGCCAGTT
Proteasome alpha subunit ( $\alpha$ -SubProt)	GCTGTGAGTAACTTGGC	GCACCTTCACTACCAGA
Ribosomal protein P2 (Ribo-P2)	CGTGATGCGTTACTTGG	TGACATCTGCTTGGCG
28 s ribosomal (Ribo s28)	GCTGTAGTGGATCTGTGC	CTACGTCATGGGACGG

IC proteomes (Fig. 2B). Among these 23 spots, 11 were present solely in the C strain (7719, 7720, 7721, 7723, 7322, 0729, 0719, 0723, 0539, 0505, and 0508), and 6 in the IC strain (7704, 7705, 7706, 7727, 0726, and 0534) (Fig. 2B). Concerning the quantitative differences, 5 spots of the C strain (0011, 1123, 5019, 6219, 0017) showed a greater than twofold increase in intensity compared to the IC strain and one IC strain spot (1749) was present in at least double the amount seen in the C strain (Fig. 2B and Table 2).

### 3.3. Characterization of proteins resulting from the comparative proteomic analysis

Three of the spots respectively specific for the C strain (7719, 7720, and 7721) or for the IC strain (7704, 7705, and 7706) could not be identified.

#### 3.3.1. Proteins involved in metabolism

**3.3.1.1. Creatine kinase.** Spots 7723 to 7727 were identified as *S. mansoni* creatine kinase (*SmCK*) (Fig. 2). ScanProsite revealed the presence of a specific active domain of the enzyme family: the ATP: guano phosphotransferases (Prosite #: PS00112). Analysis of the *SmCK* precursor by ProtFun 2.2 predicts a cytosolic or mitochondrial localization. This is in agreement with previous studies in other species [25]. In each strain a train of 4 spots was observed on the 2D gel. Their position is in agreement with the theoretical molecular weight calculated

from the deduced amino acid sequence. Nevertheless, a shift to a more acidic *pI* in the IC strain enzymes distinguishes them from the C strain *SmCK*s (Fig. 2B and Table 2). The observed *pI* shift could be due to different post-translational modifications of these *SmCK* in the strains.

This hypothesis is strengthened by previous data showing that mammalian CK is extensively autophosphorylated [26] and this could also be the case for *SmCK*. The addition of phosphate groups on proteins does not change the MW (at least not sufficiently to be visible on 2D gels), but drastically changes the isoelectric point. The result is a stretch of spots commonly called a “phosphorylation train”, the number of spots depending on the number of phosphates that the proteins carry [27]. These trains were observed in our gels and differential levels of phosphorylation could explain the observed shift.

In addition, *SmCK* cDNA sequencing in the C and IC strains revealed differences in the deduced amino acid sequence: N15S; K252R; G493D for C versus IC strain, respectively (GenBank accession numbers: EU042595 and EU042596). These differences do not change the theoretical MW (76.7 kDa for the 2 strains), but the calculated *pI* is more acidic for *SmCK* from the IC strain (C: 8.08; IC: 7.87). In addition, the difference at N15 (S in the IC strain) provides an additional putative phosphorylation site (NetPhosK 1.0 phosphorylation prediction server (<http://www.cbs.dtu.dk/services/NetPhosK/>)). Taken together, these data provided a possible basis to explain the shift in *pI* values observed for *SmCK*.

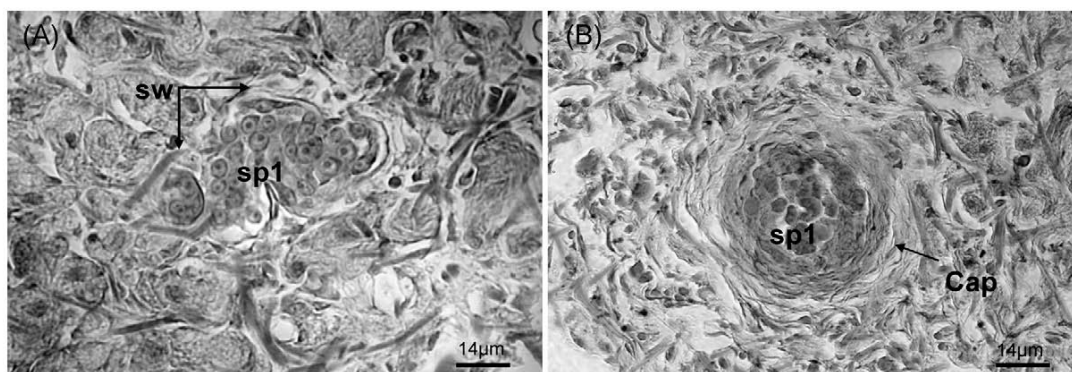


Fig. 1. Histological sections of *S. mansoni* Sporocysts (Sp1) in a compatible (C strain) host–parasite combination (A) or in an incompatible (IC strain) host–parasite combination (B) located in head-foot tissues of *B. glabrata* snails (Brazil strain) 6 h post-infection. (A) Normal developing sporocyst (Sp1) structures, sporocyst wall (sw) is intact; no haemocytes present in the vicinity of the sporocyst. (B) The sporocyst is encapsulated by hemocytes (cap = capsule), the sporocyst wall is already clearly degraded and most of sporocyst cells are destroyed.

Table 2  
Protein identification in a comparative screen between Sp1 C strain and Sp1 IC strain of *Schistosoma mansoni*

Function	Spot #	Protein name	Strain	Fold change	Species	Accession # <sup>a</sup>	MW (kDa) (exp/obs) <sup>b</sup>	pI (exp/obs) <sup>b</sup>	Coverage (%)	Peptides #	Peptides Sequence data	Mascot score
Metabolism	7723	Creatine kinase	C		<i>S. mansoni</i>	CD092123	73/75	8.7/8.7	30	7	LIEGINAIGK QLVEDHFLFK DAGYRDWVPVGR IISMQQGNLAAYK IISMQQGNLAAYK <sup>c</sup> YGYTCCPSNLGTSMR <sup>d</sup> FAHSDKYGYTCCPSNLGTSMR <sup>d</sup>	285
	7727	Creatine kinase	IC		<i>S. mansoni</i>	CD085199	73/75	8.7/8.1	3	2	LPFGDLDPTGK TYVEGFGFPTLTK TQSETLTLTSR VIFAEGNFWGR	121
	7522	Ornithine aminotransferase	C		<i>S. mansoni</i>	CD064137	47/48	8.5/8.4	41	7	VLPMNSGVEAGETSIIK VLPMNSGVEAGETSIIK <sup>c</sup> YGAHNYHPLPVLSK AFYNVCLGFEFEEMAAK <sup>d</sup> AFYNVCLGFEFEEMAAK <sup>d,c</sup> AGISLNNNFVK VVDLITHMHK	287
	6219	Glyceraldehyde-3-phosphate dehydrogenase	C/IC +2.7		<i>S. mansoni</i>	AI067530	36/37	8.1/7.8	42	7	GAMQNIIPASTGAAK GAMQNIIPASTGAAK <sup>c</sup> LVSWYDNEFGYSCR <sup>d</sup> GILEYTEDEVVSSDFYGSTSSIFDAK GILEYTEDEVVSSDFYGSTSSIFDAK	276
Antioxidant	5019	Superoxide dismutase Cu/Zn	C/IC +2.2		<i>S. mansoni</i>	AM043786	16/16	6.1/6.3	35	7	LACGVVGLAAE <sup>d</sup> AVCVMTGTAGVK <sup>d</sup> AVCVMTGTAGVK <sup>d,c</sup> TMVIHENEDDLGR TMVIHENEDDLGR <sup>c</sup> FTQETDNGPVHVHAEFSGLK HVGDLGNVVAAGADGNVYNAIDK	283
Host-parasite interaction	0729	<i>S. m</i> Mucin-like	C		<i>S. mansoni</i>	CD161554	nd/110	nd/3.7	23	6	RDWSGLLR WAVHYLSK IEAQLNIVK EQYYEAYSTLAK SLKEQYYEAYSTLAK QLTNTENKIEAQLNIVK WAVHYLSK IEAQLNIVK EQYYEAYSTLAK SLKEQYYEAYSTLAK WAVHYLSK FLSLEPYTR APQYHBECLR <sup>d</sup> WAVHYLSK IEAQLNIVK EQYYEAYSTLAK SLKEQYYEAYSTLAK WAVHYLSK FLSLEPYTR	246
	0719	<i>S. m</i> Mucin-like	C		<i>S. mansoni</i>	CD161554	nd/100	nd/3.9	17	4	WAVHYLSK IEAQLNIVK EQYYEAYSTLAK SLKEQYYEAYSTLAK QLTNTENKIEAQLNIVK WAVHYLSK IEAQLNIVK EQYYEAYSTLAK SLKEQYYEAYSTLAK WAVHYLSK FLSLEPYTR	146
	0723	<i>S. m</i> Mucin-like	C		<i>S. mansoni</i>	CD119636	nd/90	nd/4.1	17	3	WAVHYLSK FLSLEPYTR APQYHBECLR <sup>d</sup> WAVHYLSK IEAQLNIVK EQYYEAYSTLAK SLKEQYYEAYSTLAK WAVHYLSK FLSLEPYTR	100
	1749	<i>S. m</i> Mucin-like	C/IC -6.7		<i>S. mansoni</i>	CD161554	nd/80	nd/4.3	23	5	WAVHYLSK FLSLEPYTR APQYHBECLR <sup>d</sup> WAVHYLSK IEAQLNIVK EQYYEAYSTLAK SLKEQYYEAYSTLAK WAVHYLSK FLSLEPYTR	159

Table 2 (Continued)

Function	Spot #	Protein name	Strain	Fold change	Species	Accession # <sup>a</sup>	MW (kDa) (exp/obs) <sup>b</sup>	pI (exp/obs) <sup>b</sup>	Coverage (%)	Peptides #	Peptides Sequence data	Mascot score
	0539	<i>S. m</i> Mucin-like	C		<i>S. mansoni</i>	CD119646	nd/57	nd/3.3	28	5	WAVHYLSK FLSLEPYTR APQYHEECLR <sup>d</sup> EILDVDESAYIAEPR APQYHEECLRLEK <sup>d</sup>	212
	0505	<i>S. m</i> Mucin-like	C		<i>S. mansoni</i>	CD119646	nd/58	nd/3.4	21	3	WAVHYLSK APQYHEECLR EILDVDESAYIAEPR	170
	0508	<i>S. m</i> Mucin-like	C		<i>S. mansoni</i>	CD119624	nd/53	nd/3.4	26	4	WAVHYLSK FLSLEPYTR APQYHEECLR <sup>d</sup> EILDVDESAYIAEPR	184
	0726	<i>S. m</i> Mucin-like	IC		<i>S. mansoni</i>	AA140578	nd/120	nd/3.5	33	3	ILLITLTPK MIQAMDKSTNCS <sup>d</sup> ILLITLTPKDEQIK	120
	0534	<i>S. m</i> Mucin-like	IC		<i>S. mansoni</i>	CD119646	nd/55	nd/3.4	31	6	LYFTK WAVHYLSK FLSLEPYTR APQYHEECLR <sup>d</sup> EILDVDESAYIAEPR EREILDVDESAYIAEPR	221
Other functions	0017	Myosin light chain	C/IC	+2.1	<i>S. mansoni</i>	AW017362	18/18	4.7/4	54	11	FIDLR LEDPEVDAILK EGQGFISAAEMR EGQGFISAAEMR <sup>c</sup> CSGINPTIALTVK <sup>d</sup> ETGYADYMEAFK ETGYADYMEAFK <sup>c</sup> FDEFPCYEALF <sup>d</sup> EDLDGNIKYEELIK EDLDGNIKYEELIK CSGINPTIALTVKKGATIK <sup>d,e</sup> GVNTFSPEGR VNSTLIPSSIEK LFQVEYAEATK RVNSTLIPSSIEK DIPQLIAEGSKK TVLNSYGHSDSER TVLNSYGHSDSERLEK	386
	1123	Proteasome alpha subunit	C/IC	+2.2	<i>S. mansoni</i>	T14568	27/25	5.2/5.2	26	4		196
	0011	Ribosomal protein P2	C/IC	+9.7	<i>S. mansoni</i>	A1067566	12/16	4.7/3.5	23	3		152

nd: not determined.

<sup>a</sup> Genbank accession number.<sup>b</sup> Expected and observed molecular weight (MW) and isoelectric point (pI) for each identification.<sup>c</sup> Methionine oxidation.<sup>d</sup> Cysteine carboxymethylation.<sup>e</sup> Serine/threonine phosphorylation.

**3.3.1.2. Ornithine aminotransferase.** Spot 7322 was present solely in the C strain (Fig. 2) and was identified as an ornithine aminotransferase (*SmOAT*; GeneDB #: Smp\_000660). The calculated theoretical MW and *pI* based on protein primary structure are in good agreement with the position in the gel (Table 2). ScanProsite revealed the presence of the co-enzyme (pyridoxal phosphate) fixation site (Prosite #: PS00600). The alignment of *SmOAT* with OAT of other species revealed conservation of the amino acids involved in the active site and tertiary conformation [28] (data not shown). No signal peptide was predicted by SignalP 2.0, in agreement with the mitochondrial localization of the enzyme (ProtFun 2.2) [28]. Ornithine aminotransferase plays a central role in ornithine biosynthesis [29]. The absence of *SmOAT* in the IC strain is therefore surprising. To verify differences between C and IC strains, we investigated *SmOAT* transcription levels (Fig. 3). Both strains expressed OAT, with a slightly, but not significantly, higher level in the C strain. In order to characterize this gene further, we sequenced C and IC *SmOAT* cDNA obtained by PCR using primers flanking the *SmOAT* CDS. Sequencing revealed differences in the deduced amino acid sequence: N30S; E122K, for C versus IC strain, respectively (GenBank accession numbers: EU042597 and EU042598). These differences do not change the theoretical MW (48.5 kDa for the 2 strains) but change the calculated *pI* value, for the C strain *pI*=8.83 and *pI*=8.97 for the IC strain. This shift in *pI* value could explain why *SmOAT* was identified solely in C strain (Fig. 2).

**3.3.1.3. Glyceraldehyde-3-phosphate dehydrogenase.** Glyceraldehyde-3-phosphate dehydrogenase (*SmGAPDH*; GeneDB #: Smp\_056970.1; spot 6219) is over-expressed in strain C (Table 2). The enzyme active site was identified by ScanProsite (Prosite #: PS00071). Alignment of the protein CDS with GAPDHs from various other organisms confirmed the presence of conserved amino acids involved in the catalytic domain, cysteine 153 of the active site, and the domains associated with the binding of NAD<sup>+</sup> [30] (data not shown). GAPDH is involved in the Embden–Meyerhof–Parnas pathway. In this pathway, the glyceraldehyde 3-phosphate generates a high-potential phosphorylated compound, the 1,3-bisphosphoglycerate. This compound is formed by the action of the GAPDH producing concomitantly NADH.

### 3.3.2. Antioxidant enzymes

**3.3.2.1. Superoxide dismutase.** Spot (5019), was over-expressed in the C strain (Fig. 2B, Table 2) and was identified as Cu/Zn superoxide dismutase (*SmSOD*; GeneDB #: Smp\_176200.2). Two characteristic active sites of the enzyme were revealed using ScanProsite: two histidines (H45 and H47) are involved in copper binding (Prosite #: PS00087), and cysteines (C57 and C145) form disulfide bonds that are involved in the tertiary structure of the protein (Prosite #: PS00332) [31]. The calculated MW of 16 kDa and the expected *pI* of 6.3 are in good agreement with the position of the protein in the gel (Fig. 2B, Table 2). SOD is involved in oxidative

stress scavenging and its specific actions will be described in the Section 4 of the paper.

### 3.3.3. Molecules involved in host–parasite interaction

**3.3.3.1. Mucins.** Spots 0719, 0729, 0723, 0539, 0505, and 0508 are solely present in the C strain and spots 0726 and 0534 are exclusively present in the IC strain. Spot 1749 is over-expressed in IC (Fig. 2B, Table 2). These nine proteins represent the main difference between the Sp1 C and IC proteomes. ESTs identified from MASCOT analysis were assembled using Sequencher™, and partial CDSs were deduced. Comparison of this nucleotide sequence and the deduced amino acid sequence with databases (Basic BLAST; <http://www.ncbi.nlm.nih.gov/BLAST/>) revealed no similarities with known sequences.

A series of complete cDNAs were obtained and we provide a comprehensive description of these proteins in another article [21]. Briefly, the protein consists of an N-terminal signal peptide (22 amino acids) and a conserved C-terminal region. The N-terminal structure of the deduced protein shows internal tandem repeats of nine amino acids enriched in serine, threonine, and proline residues that potentially serve as a scaffold for the extensive addition of O-glycans. This particular feature relates them to the group of mucin-like molecules in mammals, and other parasites [32,33]. These proteins were thus named *S. mansoni* mucin-like proteins (*SmMucin*-like).

### 3.3.4. Proteins with other functions

**3.3.4.1. Myosin light chain.** Protein spot 0017 was over-expressed in the C strain and was identified as a myosin light chain (*SmMyoLC*; GeneDB #: Smp\_045220.1) (Fig. 2B, Table 2). ScanProsite revealed the presence of two EF-hand calcium-binding domains (prosite #: PS50222). The position of the protein in the gel corresponds to the calculated MW and *pI*.

**3.3.4.2. Proteasome alpha subunit.** Spot 1123 was over-expressed in the C strain and corresponds to the Proteasome alpha subunit (*Sm $\alpha$ -SubProt*; GeneDB #: Smp\_032580.2) (Fig. 2B, Table 2). ScanProsite identified the specific signature of a proteasome subunit (Prosite #: PS00388). The theoretical MW and *pI* based on the protein primary structure are in good agreement with the position in the gel (Table 2). No signal peptide was detected by SignalP 2.0 which is in agreement with the predicted cytosolic or nuclear localization of the protein (ProtFun 2.2) [34].

**3.3.4.3. Ribosomal protein P2.** This protein (spot 0011), over-expressed in C (Table 2), was identified as ribosomal protein P2 (*SmRibo-P2*; GeneDB #: Smp\_029820). The MotifScan tool revealed the presence of specific signature of 60s acidic ribosomal protein in the Pfam motif database (Pfam #: PF00428.9). The Ribosomal P2 protein belongs to this ribosomal protein family. A signal peptide was predicted by SignalP 2.0 HMM with a cleavage site between residues 18 and 19. Calculated MW and *pI* based on protein primary structure do not correspond to the position in the gel for either strain (Table 2). This could be related to



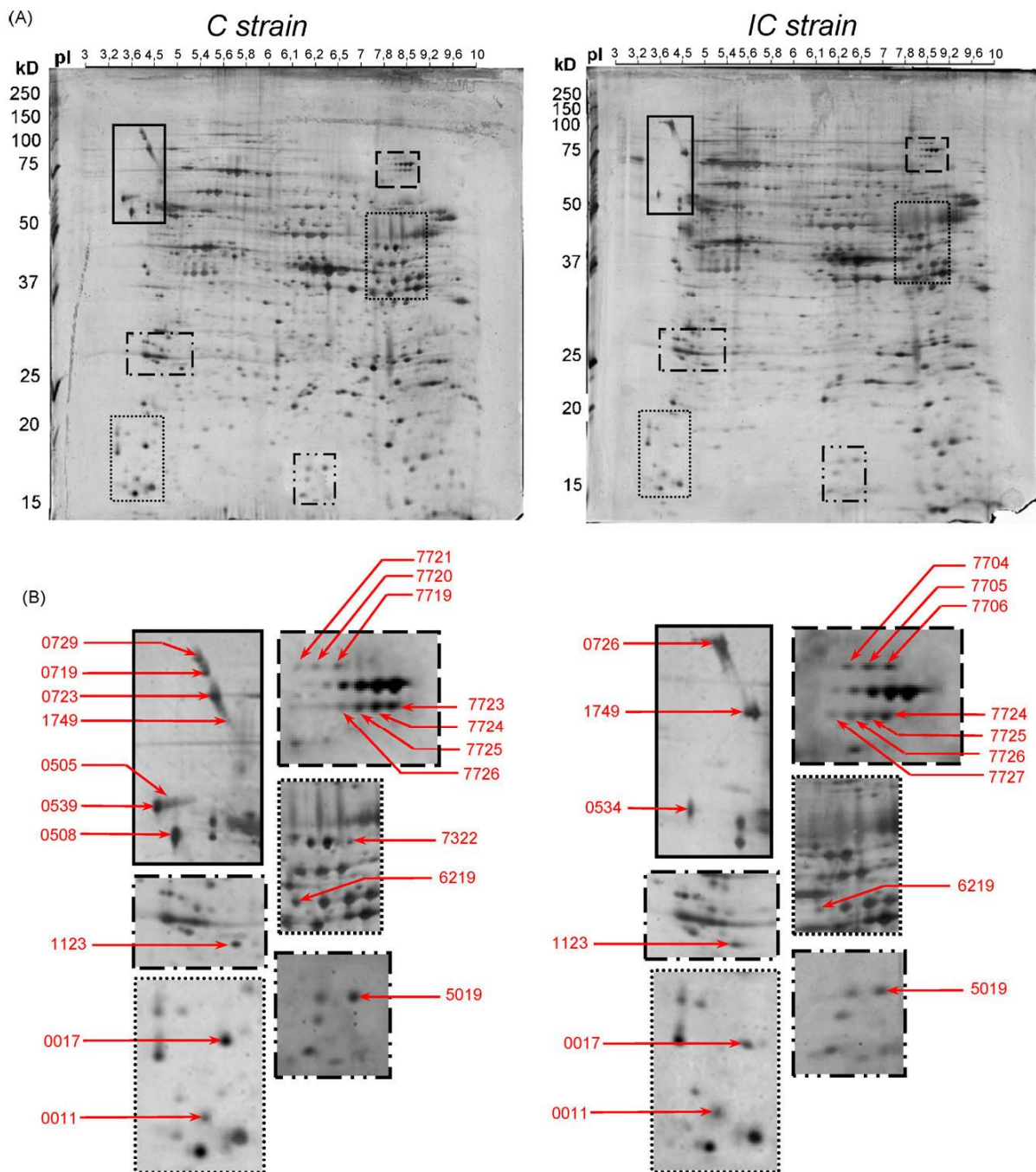


Fig. 2. (A) Proteomic profiles of sporocyst total proteome from *S. mansoni* C and IC strains. (B) High magnification of gel areas showing the qualitative and quantitative differences underlined using PDQuest image analysis software (see Table 2). X-axis is the isoelectric point scale (pI). Y-axis is the molecular weight scale in kDa. Proteins were focused using a 3–10 non-linear pH range as the first dimension and a 12% polyacrylamide gel electrophoresis (SDS-PAGE) as the second dimension. Gels were silver-stained.

post-translational modifications resulting in an increase of MW and acidification.

### 3.4. Quantitative real-time PCR

To confirm differences shown by the comparative proteomics approach for MyoLC, Ribo-P2,  $\alpha$ -SubProt, GAPDH, and SOD,

we investigated their transcript levels in sporocysts from the two strains, using quantitative RT-PCR on total RNA from 10 000 sporocysts. None of the 5 transcripts showed significant differences in transcript levels (Fig. 3). However, post-transcriptional or post-translational modifications cannot be excluded, and could explain the observed differences in protein expression. In the light of the small differences observed for these five pro-

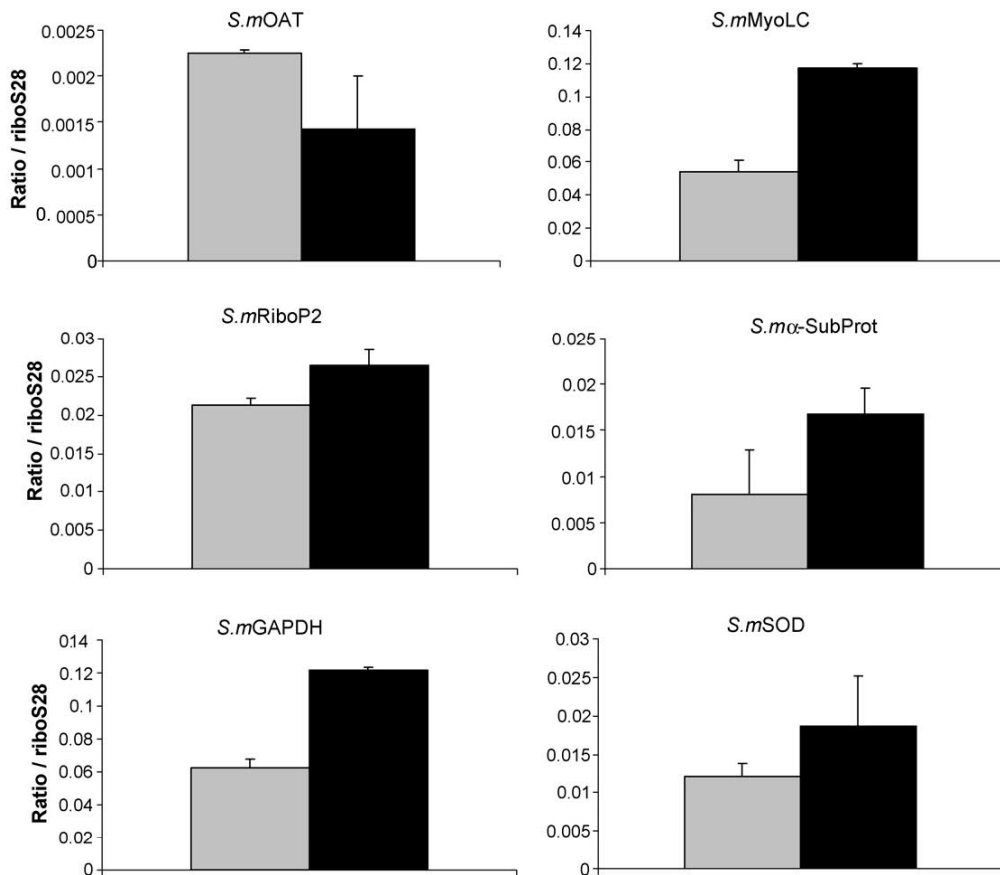


Fig. 3. Ratios of transcripts corresponding to *S.mOAT*, *S.mMyoLC*, *S.mRiboP2*, *S.mα-SubProt*, *S.mGAPDH*, and *S.mSOD*. Transcript levels in sporocysts (Sp1) from *S. mansoni* C strain (grey histograms) and *S. mansoni* IC strain (dark histograms) were measured using real-time quantitative PCR and are shown relatively to S28 expression levels (ratio/riboS28). Each histogram is the average value of duplicates  $\pm$  S.D.

teins, the significance of which for host-compatibility remains to be determined, we decided to focus our further work on the principal difference between the C and IC strain: the *SmMucin*-like gene family.

#### 4. Discussion

During the life cycle of *S. mansoni*, Sp1 larval stages develop in the mollusk intermediate hosts. Parasites need to penetrate into this host, develop, multiply asexually and finally leave the host to continue their life cycle [35,36]. Parasites therefore face many challenges such as gaining enough energy to grow and to evade the defense system of the host [37–39]. In parallel, hosts have to co-evolve with their parasites to avoid being infected. This evolutionary battle leads to an arms race between host and parasites. In our model, this struggle results in compatibility polymorphism [5]. In natural populations some snail/schistosome combinations are compatible and others are not.

In compatible interactions, the parasite penetrates and develops normally within the snail, giving rise to the next parasite stage, the cercariae. Alternatively, in incompatible interactions, the larval trematode penetrates but is immediately recognized as

non-self, encapsulated and destroyed by the mollusk's internal defense system. The success or failure of an infection does not depend on the snail susceptibility/resistance status, but on the "matched" or "mismatched" status of the host and parasite phenotypes ("matching phenotype" hypothesis, [5]). Indeed, many fundamental studies [4] have been conducted on compatibility polymorphism, but the molecular basis of this polymorphism has rarely been studied and needs to be investigated [6].

In the *S. mansoni*/*B. glabrata* interaction, incompatible miracidia are recognized as non-self and encapsulated in the first few hours following penetration. Compatible strains remain unaffected. These observations suggest constitutive antigenic differences between strains that are responsible for the compatibility polymorphism. To identify and characterize the underlying molecular determinants, we conducted a global comparative proteomic approach using primary sporocysts (Sp1) from C and IC strains. Our findings reveal several qualitative and quantitative differences, and we discuss here their putative involvement in compatibility polymorphism.

We found several proteins that are involved in oxidative stress scavenging processes. All these proteins (*SmSOD*, *SmGAPDH*, and *Smα-SubProt*) are over-represented in strain C. Host haemocytes generate oxidative stress by producing reactive oxygen

species (ROS) such as hydrogen peroxide ( $H_2O_2$ ) [40]. *SmSOD* catalyzes the dismutation of superoxide anion ( $O_2^{\bullet-}$ ) into hydrogen peroxide ( $H_2O_2$ ) and  $O_2$ .  $H_2O_2$  is the most toxic oxygen species for *S. mansoni* sporocysts [40]. But, SOD also has a peroxidative activity that uses its own dismutation product,  $H_2O_2$ , as a substrate to produce the hydroxyl radical ( $\bullet OH$ ) [41,42]. Using its peroxidative function, SOD could inactivate  $H_2O_2$  and produce  $\bullet OH$  that is less toxic for the sporocysts [40]. Another protein that could be involved in detoxification of ROS is *SmGAPDH*. *SmGAPDH* is involved in the Embden–Meyerhof–Parnas pathway. In this pathway, glyceraldehyde 3-phosphate generates a high-potential phosphorylated compound, 1,3-bisphosphoglycerate by the reduction of  $NAD^+$  to  $NADH$ . Recently, we have put forward the hypothesis that this synthesis of  $NADH$  could be involved in the cell redox cycle [9,43]. In many other parasites, glycolysis is involved in the defence against oxidative stress [44,45]. Finally, *Sm $\alpha$ -SubProt* is part of a proteolytic complex, the proteasomal system, which is the major proteolytic system responsible for the removal of oxidized cytosolic proteins. One of the hallmarks of oxidative stress is the accumulation of oxidized proteins, which tend to form high-molecular weight aggregates. The major proteolytic system responsible for the removal of oxidized cytosolic and nuclear proteins is the proteasome [46]. Over-expression of *Sm $\alpha$ -SubProt* in the C strain could be a sign of a higher proteolytic activity of the proteasomal system acting as secondary antioxidant system that maintains the integrity of cell homeostasis. The over-expression of these proteins in the C strain is in favor of a model in which a more active oxidative stress scavenging system provides a better cellular antioxidant defense for the parasite. This would allow sporocysts to combat the first line of the host's internal defense system, and to pursue their development in the snail host.

The qualitative difference in *SmOAT* expression observed in this study (Fig. 2) appears to be related to a shift in *pI* value. RT-PCR shows that both strains express *SmOAT* at similar levels, and cDNA sequencing revealed amino acid changes in the deduced sequence of the precursor protein. The apparent absence of *SmOAT* in IC is therefore certainly due to a *pI* modification that changes the position in the gel. Frameshift and point mutations in the ornithine aminotransferase gene were reported in the literature and possible effects of these events on the structural stability or the catalytic competence of the enzyme were discussed [28]. Ornithine aminotransferase plays a central role in ornithine biosynthesis [29] catalyzing the transfer of the delta-amino group of L-ornithine to 2-oxoglutarate, producing L-glutamate-gamma-semialdehyde, that in turn spontaneously cyclizes to pyrroline-5-carboxylate, and L-glutamate. This is the first step of metabolic processes in Krebs cycle and neoglucogenesis [47]. Differences in amino acid sequence could result in *SmOAT* activity differences between strains, but this remains to be investigated.

We have identified another protein that is known to be involved in a metabolic process, *Sm* Creatine kinase (*SmCK*). CK participates in temporal ATP buffering, in regulating inorganic phosphate (Pi) levels, which play a key role in glycogenolysis and proton buffering, and CK has a role in intracellular energy

transport. CK is an enzyme that catalyses the reversible transfer of the high-energy phosphoryl group of phosphocreatine to ADP and regenerates ATP [25,48]. CK is able to autophosphorylate *in vitro* [26] and the addition of a variable number of phosphate groups leads to a spot train like that observed in this study (Fig. 2). CK phosphorylation affects enzyme activity by modulating the reversibility of the CK reaction [26]. After autophosphorylation, the forward reaction is far more inhibited than the reverse reaction towards ATP synthesis. For *SmCK* a *pI* shift of the train was observed between C and IC strain. We hypothesize that this shift is associated with differential levels of phosphorylation between strains. Given the impact of phosphorylation on CK activity, we propose different energetic and metabolic capacities for the C and IC strains.

Finally, our comparative proteomics approach unearthed an unusual family of proteins that are promising candidates as key elements in the host compatibility mechanism. Blast searches found no similarities with proteins of known function. However, amino acid sequences deduced from EST assemblies revealed structural features characteristic of mucins from other parasites or from mammals [32,33]. The N-terminal part of the precursor showed internal tandem repeats of nine amino acids containing serine, threonine, and proline residues that could serve as a scaffold for the addition of O-glycans. This particular feature relates them to the group of mucin-like molecules.

Mucins are known to play a key role in host–parasite interplay and are involved in immune evasion, host invasion and parasite protection [33,49–51]. The surface of the protozoan parasite *Trypanosoma cruzi* is covered by GPI anchored mucins, which might provide protection against the vector and/or vertebrate host defence mechanisms [51]. Cleavage of the GPI anchor mucins by phosphatidylinositol-phospholipase C results in the shedding of mucins into the parasite environment. Immunogenicity of mucin carbohydrate epitopes might thus be used as an immunological decoy to form immunocomplexes with antibodies and block the opsonisation process [51]. This phenomenon was also described for other parasite species. Secreted mucins from the parasitic nematode *Toxocara canis* create a “smoke screen” that prevents deleterious antigen recognition [50,52].

The *SmMucin*-like proteins are therefore good candidates as key determinants of compatibility polymorphism in the *S. mansoni*/*B. glabrata* interaction. *SmMucin*-like are the proteins (9 spots out of 23) that clearly distinguish compatible and incompatible parasite strains. This pronounced qualitative and quantitative polymorphism (Fig. 2, Table 2) and the fact that mucins are known to interact with host molecules [33] leads us to hypothesize that *SmMucin*-like could be recognized as “self” molecules by the host (molecular mimicry hypothesis), or alternatively could avoid recognition of the parasite by the host (smoke screen hypothesis).

The proteomic approach conducted here has led to the identification of proteins putatively involved in the compatibility polymorphism. Several proteins involved in oxidative stress scavenging and metabolic processes were discovered but the most promising candidates are proteins belonging to the mucin family. These *SmMucin*-like constitute candidates of choice as molecular determinants of compatibility poly-

morphism and could represent the molecular basis of the “matching-phenotype” hypothesis developed by Theron and Coustau [5]. The polymorphism of these *SmMucin*-like that distinguishes compatible and incompatible parasite strains led us to investigate in more detail the role played by this protein family. Studies on *SmMucin*-like precursor structure, analysis of mucin-like expression and polymorphism are comprehensively described in a paper presented in the same issue [21].

## Acknowledgments

This work was supported by the CNRS. The mass spectrometry facility used in this study was funded by the European Community (FEDER), the Region Nord-Pas de Calais (France), the CNRS, the Genopole of Lille and the University of Sciences and Technologies of Lille. We are grateful to Adeline Page (Lille University), for performing mass spectrometry analysis. We thank Bernard Dejean, Rémi Emans and Pierre Tisseyre, for technical assistance and Dr. R.J. Pierce, for his critical reading of the manuscript.

## References

- [1] Chitsulo L, Loverde P, Engels D. Schistosomiasis. *Nat Rev Microbiol* 2004;2:12–3.
- [2] Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. *Lancet* 2006;368:1106–18.
- [3] El-Ansary A, Al-Daihan S. Important aspects of *Biomphalaria* snail–schistosome interactions as targets for antischistosome drug. *Med Sci Monit* 2006;12:RA282–92.
- [4] Webster JP, Davies CM. Coevolution and compatibility in the snail–schistosome system. *Parasitology* 2001;123(Suppl.):S41–56.
- [5] Theron A, Coustau C. Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *J Helminthol* 2005;79:187–91.
- [6] Lewis FA, Patterson CN, Knight M, Richards CS. The relationship between *Schistosoma mansoni* and *Biomphalaria glabrata*: genetic and molecular approaches. *Parasitology* 2001;123(Suppl.):S169–79.
- [7] Theron A, Pages JR, Rognon A. *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp Parasitol* 1997;85:1–9.
- [8] Mattos AC, Kusel JR, Pimenta PF, Coelho PM. Activity of praziquantel on in vitro transformed *Schistosoma mansoni* sporocysts. *Mem Inst Oswaldo Cruz* 2006;101(Suppl. 1):283–7.
- [9] Guillou F, Roger E, Mone Y, et al. Excretory–secretory proteome of larval *Schistosoma mansoni* and *Echinostoma caproni*, two parasites of *Biomphalaria glabrata*. *Mol Biochem Parasitol* 2007;155:45–56.
- [10] Chemin E. Observations on hearts explanted in vitro from the snail *Australorbis glabratus*. *J Parasitol* 1963;49:353–64.
- [11] Coustau C, Mitta G, Dissous C, et al. *Schistosoma mansoni* and *Echinostoma caproni* excretory–secretory products differentially affect gene expression in *Biomphalaria glabrata* embryonic cells. *Parasitology* 2003;127:533–42.
- [12] Vergote D, Bouchut A, Sautiere PE, et al. Characterisation of proteins differentially present in the plasma of *Biomphalaria glabrata* susceptible or resistant to *Echinostoma caproni*. *Int J Parasitol* 2005;35:215–24.
- [13] Patton WF. Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci* 2002;771:3–31.
- [14] Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 1996;68:850–8.
- [15] Bouchut A, Sautiere PE, Coustau C, Mitta G. Compatibility in the *Biomphalaria glabrata*/*Echinostoma caproni* model: potential involvement of proteins from hemocytes revealed by a proteomic approach. *Acta Trop* 2006;98:234–46.
- [16] Vergnes B, Gourbal B, Girard I, Sundar S, Drummel-Smith J, Ouellette M. A proteomics screen implicates HSP83 and a small kinetoplastid calpain-related protein in drug resistance in *Leishmania donovani* clinical field isolates by modulating drug-induced programmed cell death. *Mol Cell Proteomics* 2007;6:88–101.
- [17] Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 2001;70:437–73.
- [18] Ashton PD, Curwen RS, Wilson RA. Linking proteome and genome: how to identify parasite proteins. *Trends Parasitol* 2001;17:198–202.
- [19] Kayser JP, Vallet JL, Cerny RL. Defining parameters for homology-tolerant database searching. *J Biomol Tech* 2004;15:285–95.
- [20] Guillou F, Mitta G, Dissous C, Pierce R, Coustau C. Use of individual polymorphism to validate potential functional markers: case of a candidate lectin (BgSel) differentially expressed in susceptible and resistant strains of *Biomphalaria glabrata*. *Comp Biochem Physiol B Biochem Mol Biol* 2004;138:175–81.
- [21] Roger E, Gourbal BEF, Grunau C, Pierce RJ, Galinier R, Mitta G. Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. *Mol Biochem Parasitol* 2007, doi:10.1016/j.molbiopara.2007.11.015, in press.
- [22] Liu F, Lu J, Hu W, et al. New perspectives on host–parasite interplay by comparative transcriptomic and proteomic analyses of *Schistosoma japonicum*. *PLoS Pathog* 2006;2:e29.
- [23] Liu F, Hu W, Cui SJ, et al. Insight into the host–parasite interplay by proteomic study of host proteins copurified with the human parasite *Schistosoma japonicum*. *Proteomics* 2007;7:450–62.
- [24] Perez-Sanchez R, Ramajo-Hernandez A, Ramajo-Martin V, Oleaga A. Proteomic analysis of the tegument and excretory–secretory products of adult *Schistosoma bovis* worms. *Proteomics* 2006;6(Suppl. 1):S226–36.
- [25] Ellington WR. Evolution and physiological roles of phosphagen systems. *Annu Rev Physiol* 2001;63:289–325.
- [26] Stolz M, Hornemann T, Schlattner U, Wallimann T. Mutation of conserved active-site threonine residues in creatine kinase affects autophosphorylation and enzyme kinetics. *Biochem J* 2002;363:785–92.
- [27] Carter MD, Southwick K, Lukov G, Willardson BM, Thulin CD. Identification of phosphorylation sites on phosphocin-like protein by QTOF mass spectrometry. *J Biomol Tech* 2004;15:257–64.
- [28] Shen BW, Hennig M, Hohenester E, Jansonius JN, Schirmer T. Crystal structure of human recombinant ornithine aminotransferase. *J Mol Biol* 1998;277:81–102.
- [29] Gafan C, Wilson J, Berger LC, Berger BJ. Characterization of the ornithine aminotransferase from *Plasmodium falciparum*. *Mol Biochem Parasitol* 2001;118:1–10.
- [30] Jenkins JL, Tanner JJ. High-resolution structure of human D-glyceraldehyde-3-phosphate dehydrogenase. *Acta Crystallogr D Biol Crystallogr* 2006;62:290–301.
- [31] Cardoso RM, Silva CH, Ulian De Araujo AP, Tanaka T, Tanaka M, Garratt RC. Structure of the cytosolic Cu, Zn superoxide dismutase from *Schistosoma mansoni*. *Acta Crystallogr D Biol Crystallogr* 2004;60:1569–78.
- [32] Van Klinken BJ, Dekker J, Buller HA, Einerhand AW. Mucin gene structure and expression: protection versus adhesion. *Am J Physiol* 1995;269:G613–27.
- [33] Theodoropoulos G, Hicks SJ, Corfield AP, Miller BG, Carrington SD. The role of mucins in host–parasite interactions. Part II. Helminth parasites. *Trends Parasitol* 2001;17:130–5.
- [34] Castro-Borges W, Cartwright J, Ashton PD, et al. The 20S proteasome of *Schistosoma mansoni*: a proteomic analysis. *Proteomics* 2007;7:1065–75.
- [35] Davies SJ, Mckerrow JH. Developmental plasticity in schistosomes and other helminths. *Int J Parasitol* 2003;33:1277–84.
- [36] Parker GA, Chubb JC, Ball MA, Roberts GN. Evolution of complex life cycles in helminth parasites. *Nature* 2003;425:480–4.
- [37] De Jong-Brink M, Bergamin-Sassen M, Solis Soto M. Multiple strategies of schistosomes to meet their requirements in the intermediate snail host. *Parasitology* 2001;123(Suppl.):S129–41.
- [38] Meuleman EA, Bayne CJ, Van Der Knaap WP. Immunological aspects of snail–trematode interactions. *Prog Clin Biol Res* 1987;233:113–27.
- [39] Van Der Knaap WP, Loker ES. Immune mechanisms in trematode–snail interactions. *Parasitol Today* 1990;6:175–82.

- [40] Hahn UK, Bender RC, Bayne CJ. Involvement of nitric oxide in killing of *Schistosoma mansoni* sporocysts by hemocytes from resistant *Biomphalaria glabrata*. *J Parasitol* 2001;87:778–85.
- [41] Kim SM, Kang JH. Peroxidative activity of human Cu, Zn-superoxide dismutase. *Mol Cells* 1997;7:120–4.
- [42] Yim MB, Chock PB, Stadtman ER. Enzyme function of copper, zinc superoxide dismutase as a free radical generator. *J Biol Chem* 1993;268:4099–105.
- [43] Kum-Tatt L, Tan IK, Seet AM. A new colorimetric method for the determination of NADH/NADPH-dependent glutathione reductase in erythrocytes and in plasma. *Clin Chim Acta* 1975;58:101–8.
- [44] Cordeiro AT, Michels PA, Delboni LF, Thiemann OH. The crystal structure of glucose-6-phosphate isomerase from *Leishmania mexicana* reveals novel active site features. *Eur J Biochem* 2004;271:2765–72.
- [45] Roth Jr E. *Plasmodium falciparum* carbohydrate metabolism: a connection between host cell and parasite. *Blood Cells* 1990;16:453–60 [discussion 461–6].
- [46] Bader N, Grune T. Protein oxidation and proteolysis. *Biol Chem* 2006;387:1351–5.
- [47] Haslett MR, Pink D, Walters B, Brosnan ME. Assay and subcellular localization of pyrroline-5-carboxylate dehydrogenase in rat liver. *Biochim Biophys Acta* 2004;1675:81–6.
- [48] Stein LD, Ham DA, David JR. A cloned ATP:guanidino kinase in the trematode *Schistosoma mansoni* has a novel duplicated structure. *J Biol Chem* 1990;265:6582–8.
- [49] Hicks SJ, Theodoropoulos G, Carrington SD, Corfield AP. The role of mucins in host–parasite interactions. Part I. Protozoan parasites. *Parasitol Today* 2000;16:476–81.
- [50] Rathore D, Nagarkatti R, Jani D, et al. An immunologically cryptic epitope of *Plasmodium falciparum* circumsporozoite protein facilitates liver cell recognition and induces protective antibodies that block liver cell invasion. *J Biol Chem* 2005;280:20524–9.
- [51] Buscaglia CA, Campo VA, Frasch AC, Di Noia JM. Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nat Rev Microbiol* 2006;4:229–36.
- [52] Loukas A, Hintz M, Linder D, et al. A family of secreted mucins from the parasitic nematode *Toxocara canis* bears diverse mucin domains but shares similar flanking six-cysteine repeat motifs. *J Biol Chem* 2000;275:39600–7.



## CHAPITRE 2

### Analyse de l'expression de protéines de type mucine hautement polymorphes (*SmPoMuc*) chez le parasite *Schistosoma mansoni*

#### Publication n°2

Expression analysis of highly polymorphic mucin proteins (*SmPoMuc*) from the parasite *Schistosoma mansoni*. **Molecular & Biochemical Parasitology** (2008);157:217-27.

Afin de mettre en évidence les déterminants moléculaires du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria glabrata*, une approche protéomique comparative globale a précédemment été développée (Roger *et al.*, 2008a). La comparaison des profils protéiques des sporocystes primaires (stade intramolluscal) compatibles et incompatibles a conduit à l'identification de plusieurs candidats, dont des protéines polymorphes de types mucines. De nombreuses études montrent que les mucines jouent un rôle majeur dans les interaction hôte / parasite. Ainsi, ces protéines polymorphes de type mucine de *S. mansoni* pourraient être des acteurs essentiels de la compatibilité dans l'interaction avec *B. glabrata*.

Afin de mieux cerner l'implication de ces protéines dans le polymorphisme de compatibilité qui caractérise notre modèle, nous avons étudié leur structure, leur expression et leur polymorphisme au cours d'une deuxième étude, qui est présentée dans ce chapitre.

Nous avons tout d'abord analysé les transcrits correspondant à ces protéines de type mucine, dans les souches compatible et incompatible. Il ressort de cette analyse que la région VNTR (« Variable Number Tandem Repeat ») est extrêmement polymorphe que ce soit au niveau intra-souche ou inter-souche. Nous avons également montré que cette région VNTR est fortement glycosylée. Ces différentes caractéristiques nous ont conduit à nommer ces protéines *SmPoMuc* pour « *Schistosoma mansoni* Polymorphic Mucins ».

Nous montrons également dans cette 2<sup>ème</sup> étude que les *SmPoMuc* sont exprimées spécifiquement dans les stades miracidium et sporocyste primaire, stades interagissant avec l'hôte mollusque. Les *SmPoMuc* sont localisées dans la glande apicale, qui est un organe présent chez ces deux stades parasitaires. De plus, ces protéines sont présentes dans les

produits d'excrétion/sécrétion (ES) du stade sporocyste primaire. Cette étude nous a permis d'avancer une hypothèse quant au rôle que pourrait jouer les *SmPoMuc* dans l'interaction : elles formeraient, dans l'environnement immédiat du parasite, un « écran de fumée » et bloqueraient ainsi les PRR (« Pattern Recognition Receptor ») empêchant la reconnaissance et l'élimination par le système immunitaire du mollusque.





## Expression analysis of highly polymorphic mucin proteins (*Sm* PoMuc) from the parasite *Schistosoma mansoni*<sup>☆</sup>

Emmanuel Roger<sup>a</sup>, Benjamin Gourbal<sup>a</sup>, Christoph Grunau<sup>a</sup>, Raymond J. Pierce<sup>b</sup>,  
Richard Galinier<sup>a</sup>, Guillaume Mitta<sup>a,\*</sup>

<sup>a</sup> Parasitologie Fonctionnelle et Evolutive, UMR 5244, CNRS Université de Perpignan, 52 Ave Paul Alduy, 66860 Perpignan Cedex, France

<sup>b</sup> Inserm, U 547, Université Lille 2, Institut Pasteur de Lille, IFR 142, 1 rue du Professeur A. Calmette, 59019 Lille, France

Received 20 July 2007; received in revised form 23 October 2007; accepted 27 November 2007

Available online 3 December 2007

### Abstract

The co-evolutionary dynamics that exist in many host-parasite interactions sometimes lead to a compatibility polymorphism, of which the molecular bases are unknown. To identify key molecules involved in this phenomenon in the *S. mansoni*/*B. glabrata* model, we developed a comparative proteomics approach for the larval stages that interact with the invertebrate host. The comparison of the proteomes of compatible and incompatible parasite strains led to the identification of a new family of schistosome antigens that share molecular characteristics with the molecules of the mucin family. In particular, they possess a domain containing a variable number of tandem repeats (VNTR). The pronounced polymorphism of these proteins, that distinguishes compatible and incompatible parasite strains, led us to further investigate the role that this protein family plays in the compatibility polymorphism in our model. In the present study, we examine precursor structure, report analysis of mucin-like expression and describe their polymorphism. Our data show that these proteins share structural characteristics with highly glycosylated secreted mucins. The proteins are (i) only expressed in larval stages that interact with the mollusc, (ii) located in the apical gland of miracidia and sporocysts and (iii) secreted and released in excretion–secretion products. Finally, we show that these mucins display a high degree of polymorphism and that extensive differences are observed between *S. mansoni* strains. These different characteristics led us to name this novel gene family “*S. mansoni* polymorphic mucins” (*Sm* PoMuc).

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** *Biomphalaria glabrata*; *Schistosoma mansoni*; Host-parasite interactions; Compatibility polymorphism; *Sm* polymorphic mucins

### 1. Introduction

The comprehensive understanding of host-parasite interactions represents a major challenge in evolutionary biology. Because parasites are responsible for substantial deleterious effects, they represent a major driving force for the evolution of their hosts. In parallel, parasites have to co-evolve with their host to avoid elimination. This adaptation of the Red Queen hypothesis to host-parasite systems [1] predicts that an arms race will lead to the evolution of mechanisms that generate diversity and polymorphism of molecules that play a key role in host-parasite

interplay. For vertebrate hosts, the most striking example is the exceptional diversity of antigen-specific receptors of the adaptive immune system of jawed vertebrates, which depends on somatic gene rearrangement and hypermutation [2,3]. For the pathogen counterpart, a variety of mechanisms permitting evasion of the host immune response exist in pathogenic bacteria and viruses [4]. Antigenic variation is often essential or even the sole strategy for most of the eukaryotic parasites [5]. Concerning the interactions between invertebrate hosts and their parasites, the picture should be completely different since it is believed that invertebrates have no acquired adaptive immunity and that their immune system is non-specific. Nevertheless, recent studies have shaken this paradigm by providing evidence of novel and diverse immune receptor sequences in jawless vertebrates (lamprey [6]), protochordates (*Amphioxus* [7]), in echinoderms (sea urchin [8]), insects (*Drosophila melanogaster* [9]) and molluscs (*Biomphalaria glabrata* [10]). These results suggest

<sup>☆</sup> Note: Nucleotide sequence data reported in this paper are available in the GenBank™ database under the accession numbers: EU042599 to EU042636.

\* Corresponding author. Tel.: +33 4 68 66 21 88; fax: +33 4 68 66 22 81.

E-mail address: mitta@univ-perp.fr (G. Mitta).

the existence of individual or population-based polymorphism permitting the survival of species confronted with parasites. These novel observations raise the questions of whether diversity and polymorphism exist in antigens of parasites or of specific parasite stages (for multi-host parasites) that interact with invertebrate hosts, whether these antigens are subject to variation, and whether molecular polymorphism is at the core of interaction with the host immune system.

To address these questions, it is necessary to work with a host-parasite model where the coevolutionary dynamics are accessible. This is the case in certain host-parasite models in which only some particular host and parasite phenotypes are compatible. The model we have chosen to study is the interaction between *Schistosoma mansoni*, the agent of human intestinal schistosomiasis [11] and its invertebrate intermediate host, the gastropod mollusc *Biomphalaria glabrata*. In this interaction, a compatibility polymorphism occurs [12]: in natural populations some snail/schistosome combinations are compatible and others are not. This phenomenon is probably due to diversification mechanisms acting on key molecules like pattern recognition receptors of the host and antigens of the parasite. In *B. glabrata*, immunoglobulin superfamily (IgSF) genes have been characterized recently [13] and their polymorphism and diversification was demonstrated [10]. These genes encode lectin-like hemolymph polypeptides that can precipitate soluble antigens derived from trematodes [13]. They consist of one or two amino-terminal IgSF domains and a carboxyl-terminal fibrinogen domain. These molecules undergo mutations and recombinatorial processes that lead to diversification [10]. According to the arms race hypothesis, polymorphic antigenic variants expressed by schistosome larvae in intermediate hosts could explain the observed compatibility polymorphism. However, no data exist as yet concerning the mechanisms that generate such polymorphism in the parasite.

A recent study, presented in a companion paper in this issue, sought to identify molecules that play a central role in this compatibility polymorphism by using a comparative proteomics approach in the larval stages that interact with the invertebrate host. We compared the proteomes of parasite strains that are compatible (C) and incompatible (IC) with a specific *B. glabrata* strain. This led us to identify a family of schistosome antigens that share some characteristics with the molecules of the mucin family. In particular, they display a domain containing a variable number of tandem repeats (VNTR). Since mucins are known to play key roles in host-parasite interplay and are involved in immune evasion, host invasion and immune protection [14–17], we decided to further investigate the role of this new protein family could play in the compatibility polymorphism in our system.

In the present work, we examined precursor structure and report the analysis of expression and polymorphism of this mucin-like family. The data show that these proteins share structural characteristics with highly glycosylated secreted mucins. We show that these proteins are (i) only expressed in larval stages interacting with the mollusc, (ii) located in the apical gland of miracidia and sporocysts and (iii) secreted and released

in excretion–secretion products. Finally, we show that these mucins, that we have named *Sm* PoMuc, display a high degree of polymorphism and that important differences exist between C and IC *S. mansoni* strains.

## 2. Materials and methods

### 2.1. Parasite strains and different development stages

Two strains of *Schistosoma mansoni* were used in this study: a strain that originally came from Brazil and a Guadeloupean strain. Each strain was maintained in its sympatric *Biomphalaria glabrata* strain, and in hamsters (*Mesocricetus auratus*) as described previously [18,19]. In the past, compatibility had been tested: one hundred snails of the Bg.Bra strain were individually exposed to 10 miracidia of either *S. mansoni* strain. Infection success was examined 4 weeks after exposure [19]. Of these snails, 100% and 4% respectively exposed to the Brazilian and Guadeloupean *S. mansoni* strains were infected. The compatible Brazilian and incompatible Guadeloupean strains are called C and IC strains in the present work.

Eggs from *S. mansoni* strains were axenically recovered from 50-day infected hamster livers and miracidia were hatched from eggs as previously described [19]. Eggs and miracidia were concentrated by sedimentation on ice for 1 h and directly submitted to RNA extraction. Miracidia were used for protein extraction or immunocytochemical procedures (see below). Primary (mother) sporocysts (Sp1) were obtained *in vitro*. After three washes in sterile pond water containing an antibiotic/antimycotic mixture (penicillin 100 U/ml, streptomycin 0.1 mg/ml, amphotericin 0.025 µg/ml; Sigma), miracidia were placed in 50 ml tissue-culture plates containing sterile Chermín's balanced salt solution (CBSS) [20] and the antibiotic/antimycotic mixture at 26 °C under atmospheric conditions. Full transformation of miracidia to Sp1 occurred within the first 24 h in culture [21]. After 24 h, the transformed sporocysts were centrifuged ( $600 \times g$  for 5 min) in a 15 ml tube. Pellets were directly submitted to RNA extraction, protein extraction or immunocytochemical procedures (see below). The sporocyst excretory–secretory products (ESP) in the culture medium were recovered and filtered through a 0.22 µm membrane, lyophilized and analyzed by Western blot (see below).

Cercariae were recovered from infected snails (4 weeks post-infection) and harvested on ice by pipetting. They were then pelleted by centrifugation ( $1000 \times g$  for 10 min). Pellets were directly submitted to RNA extraction (see below).

Eight-week adult worms were recovered by portal perfusion of mice with RPMI 1640 containing 10 mM HEPES (Sigma) and 4 U/ml heparin (Sigma). RNA was extracted directly from adult worms (see Section 2.3).

### 2.2. Protein separation, staining, glycodetection and identification

Total proteome extraction was realized using 10,000 sporocysts (from each strain) incubated during 2 h at room temperature in 2D lysis buffer (8 M urea, 40 mM Tris, 4% CHAPS,

60 mM DTT). The lysate was centrifuged at  $15,000 \times g$  to pellet remaining cells and nuclear material. The protein concentration in the supernatant was assayed using the 2D Quant Kit (Amersham), 100  $\mu$ g protein were loaded on the first dimension, on 17 cm Ready Strips IPG Strips, pH 3–10 non-linear gradient (Bio-Rad). Isoelectrofocusing (IEF) was performed as previously described [22] with the IEF voltage increasing gradually to 8000 V and running for 90,000 Vh at 20 °C.

Proteins were separated in the second dimension by SDS–PAGE [22], and proteins or glycoproteins were visualized. Proteins were detected by silver staining according to a method compatible with mass spectrometry analysis [23]. 2D gels were scanned using a densitometer (GS-800 Calibrated Densitometer, Bio-Rad). The carbohydrate moieties were detected using the GlycoProfile™ III fluorescent detection kit supplied by Sigma. Briefly, after undergoing SDS–PAGE, proteins were fixed in the gel with an acetic acid (3%)–methanol (50%) solution. The carbohydrates present in the proteins were oxidized to aldehydes with periodic acid. A hydrazide dye reacting with the aldehydes was then used to form a stable fluorescent conjugate. After extensive washes with water, the gel was viewed with a standard fluorescent UV-transilluminator. The glycosylated proteins were visible in the form of bright fluorescent spots.

Gel plugs containing the proteins of interest were excised, processed and characterized by MALDI-ToF MS and nanoscale capillary liquid chromatography–tandem mass spectrometry (nanoLC–MS/MS) as described previously [22,24]. Endo Lysine-C (Roche) was used with the protocol provided by the manufacturer. MALDI-ToF MS analysis of enzymatic digests was performed on a Voyager DE Pro™ (Applied Biosystems). Peptide masses were determined after peak smoothing and internal calibration using the two autolysis trypsin fragments  $[M + H]^+ = 2211.1$  Da and 842.51 Da. NanoLC-MS/MS analysis of the digested proteins was performed using a capillary LC system coupled to an LCQ Deca XP<sup>+</sup> Thermo-electron tandem mass spectrometer. The mass data recorded during nanoLC-MS/MS analysis were processed and converted into “Merge” .DTA peak lists format prior to searching with the search engine MASCOT (<http://www.matrixscience.com/>) against protein or nucleotide sequence databases, available through the National Center for Biotechnology Information website (<http://www.ncbi.nlm.nih.gov>). One or two missed cleavages per peptide were allowed and some variable modifications were taken into account in the search such as carbamidomethylation of cysteine, oxidation of methionine, histidine or tryptophan, and phosphorylation of serine, threonine or tyrosine. Searches were performed without constraining protein molecular weight or isoelectric point.

### 2.3. RNA extraction and reverse transcription

Total RNA extractions from eggs, miracidia, Sp1, cercariae and adult worms were performed using the Trizol Reagent (Life Technologies) and the manufacturer’s protocol. Reverse transcription was performed according to previously described procedures [25].

### 2.4. Gene characterization

ESTs identified from MASCOT analysis were imported in a fasta format and assembled using Sequencher™ software (Gene Codes Corporation). Complete CDSs were obtained using the GeneRacer™ Kit (Invitrogen). Briefly, first-strand cDNAs were prepared from sporocyst total RNAs with oligo d(T)<sub>17</sub> as the primer using the Superscript II™ (Invitrogen). The 3′ and 5′ ends of the cDNAs were amplified by PCR using specific primers (S123R1 CTGTGGTGCTCTTGTGGAAGCTCG for 5′ RACE-PCR and S123F1 TTGGAATCGAAGGCGATACAAGC for 3′ RACE-PCR) designed from contigs obtained by EST assembly. PCR was performed using Advantage® 2 PCR Enzyme System (Clontech). Primers were used at a concentration of 0.2  $\mu$ M. Cycling conditions were: 40 cycles consisting of 30 s at 95 °C, 30 s at 65 °C and 90 s at 68 °C.

The PCR products were ligated into the pCR2.1-TOPO™ vector according to the manufacturer’s instructions (Invitrogen). Dideoxy sequencing reactions of the recombinant plasmids were sequenced using a dideoxy-dye-terminator method (CEQ™ DTCS-Quick Start kit, Beckman Coulter) and a CEQ™ 8000 apparatus (Beckman Coulter). M13 forward and reverse primers were used for sequencing. Sequences were obtained using the CEQ™ 8000 sequence analysis software.

For polymorphism and strain difference analysis, the CDS was amplified using the following forward (Ex1-F12 GGAA-GAATGAACAAGAAAATTATTCTC) and reverse (Ex1-R TGACACAGAAAAGTGTAAACGATCC) primers which encompass the complete CDS. PCR was performed using Advantage® 2 PCR Enzyme System with 40 cycles consisting of 30 s at 95 °C, 30 s at 64 °C and 3 min at 68 °C. Cloning and sequencing were performed as described above.

### 2.5. Real-time PCR

Real-Time PCR analysis of gene expression was performed on cDNAs. PCR and relative quantification were performed according to previously described procedures [25] with a Light cycler™ apparatus (Roche Molecular Biochemicals, Germany). Specific primers for real-time quantitative PCR were designed with the Light Cycler Probe Design™ Software version 1.0. The following primers were used for *S. mansoni* mucins (MSPPtotF2/AATTGGAATCGAAGGCG and MSPPtotR2/TGTTTCATCTTTGGGTAGTGT) and for 28s RNA (ARN28sF1/GCTGTAGTGGATCTGTGC and ARN28sR1/CTACGTCATGGGACGG).

### 2.6. Anti-Sm PoMuc antibody

A mixture of 4 BSA-coupled peptides (supplied by Genpep, France) was used to immunize New Zealand Rabbits (Proteogenix, France) as described [26]. Peptide sequences used were CEHLKTRIKHYKDAY, MIIQAMDKSTNCS, RASSTRAPQIYHEEC and CRILLTTLPKDEQI. Sera of immunized rabbits were collected and tested for the presence of specific anti-mucin Igs 3 months after the initial injection using ELISA [27] with uncoupled peptide adsorbed onto Maxisorp

plates (Nunc). For Western blotting the purified IgG fraction was used [28].

### 2.7. Western blot for *Sm* PoMuc in *Sp1*, miracidia and excretory–secretory proteins

*Sp1*, miracidia or lyophilised ESP were resuspended and disrupted in an equal volume of 2× lysis buffer (125 mM Tris–HCl, pH 6.8; 4% SDS; 3% β-mercaptoethanol; 17.5% glycerol; 5 mM Na<sub>3</sub>VO<sub>4</sub>; 20 μg/ml leupeptin (Roche Molecular Biochemicals); 20 μg/ml aprotinin (Roche Molecular Biochemicals); 0.0025% bromophenol blue). The samples were boiled for 10 min and vortexed frequently to ensure that parasites were properly disrupted. The protein content of the supernatant obtained after centrifugation at 20,000 × *g* for 30 min was measured using the 2D Quant Kit (Amersham). For protein blot analysis, 5 μg of protein were separated on a 12% SDS–polyacrylamide gel and electrophoretically transferred onto a nitrocellulose membrane. The protein blot was incubated for 1.5 h in 1× PBS, 5% non-fat milk, 0.05% Tween-20 with a 1:200 dilution of the purified IgG anti-MSPP antibody (Proteogenix, France). The blot was washed three times with 1× PBS containing 0.05% Tween-20 and incubated with a 1:5000 dilution of horseradish peroxidase-conjugated goat anti-rabbit IgG (Pierce) in 1× PBS, 5% non-fat milk, 0.05% Tween-20 for 1 h, washed again with PBS-Tween, and then subjected to autoradiography in the presence of an enhanced chemiluminescent substrate (Pierce).

### 2.8. Immunocytochemical procedures

Miracidia and sporocyst pellets (each containing 1000 individuals) were fixed in PBS pH 7.4 containing 4% paraformaldehyde during 1 h at 4 °C. After fixation, parasites were centrifuged (1 min, 800 × *g*) on poly-L-lysine coated slides (Cyto-Tek centrifuge) and immersed in a permeabilizing PBS solution containing 0.5% Triton X-100 during 15 min. A saturation step was performed in PBS buffer containing 1% gelatine hydrolysate (Bellon, France), 1% normal goat serum (NGS, Sigma) and 0.1% NaN<sub>3</sub> (Sigma) during 1 h at room temperature. Parasites were then successively incubated with the anti-*Sm* PoMuc antibody (10 μg/ml) for 2 h at room temperature and with an Alexa Fluor 488 anti-rabbit IgG (BioSys, France) diluted 1:200 for 1 h at room temperature. Slides were mounted in glycerine buffer and examined by epifluorescence and light microscopy using a Zeiss axioscope 2 microscope (Carl Zeiss AG) and a Leica DC350FX camera (Leica).

Controls with anti-*Sm* PoMuc antibody preadsorbed overnight at 4 °C with pure peptides (10 μg of peptides per μg of purified IgG) were performed on western blots and centrifuged parasites.

## 3. Results

### 3.1. *Sm* PoMuc sequences have a mucin-like structure and are highly polymorphic

After assembly of ESTs obtained from databases, partial cDNA sequences corresponding to *S. mansoni* mucin-like pro-

teins were assembled (described in the companion paper). To obtain the complete precursor sequences, their complete CDS were characterized using 5' and 3' RACE-PCR experiments on the two strains. Five clones derived from each strain and RACE experiment were sequenced. Primers were then designed to amplify the complete CDS of *S. mansoni* mucin-like genes by RT-PCR. Visualization of PCR products after size separation in agarose gels revealed a large number of different amplicons ranging approximately from 900 to 2500 bp (data not shown). These different molecular variants were cloned and 76 clones were sequenced for each strain (insert sizes ranging from 877 to 2350 bp). The sequences are very diverse but all the precursors share the structural characteristics shown in Fig. 1A. The deduced amino acid sequence begins with a 22-residue signal peptide with a typical hydrophobic core. The cleavage site for signal peptidase is most likely located after the alanine residue preceding the serine in position 1 of the mature protein as predicted by Signal P VI.1 software [29]. The sequence following the cleavage site corresponded to a first domain composed of tandem repeats of 9 aa. The number *n* of repeats is variable ( $n=1$  to  $n \approx 55$ ). Three different types of repeats were identified: r1, r1' and r2 (Fig. 1B). Two of them (r1 and r1') are very similar and differ only by 1 residue. All the repeats contain S, T and P residues. Such repetitive structures with similar amino acid compositions were described in different mucins, and O-glycosylation of T residues in these repeats (a typical feature of mucins) is predicted using the NetOGlyc 3.1 server (<http://www.cbs.dtu.dk/services/NetOGlyc/>, data not shown). After this tandem repeat domain, the ORF codes for a C-terminal region of 234 residues. Similarity searches (<http://www.ncbi.nlm.nih.gov/BLAST>) carried out using (i) blastn revealed no significant similarities, (ii) tblastn revealed a significant score ( $e$ -value =  $9e-48$ ) for hypothetical proteins from *S. japonicum* (AAX26027) and (iii) blastp revealed similarities ( $e$ -value =  $1e-4$ ) for the repetitive region of circumsporozoite proteins from *Plasmodium vivax* (Swissprot accession numbers: P08677 and Q03110). In addition, neither motifs nor domains were found using motif prediction software like RPS Blast, ScanProsite or MotifScan. Since some proteins of the mucin family are glycosylphosphatidylinositol (GPI)-anchored [30], we investigated this possibility. No GPI modification sites were found using different prediction software (big-PI Predictor, [http://mendel.imp.ac.at/sat/gpi/gpi\\_server.html/GPI-SOM](http://mendel.imp.ac.at/sat/gpi/gpi_server.html/GPI-SOM), <http://gpi.unibc.ch/>). In addition, no transmembrane domain was detected using standard software (HMM-TOP, <http://www.enzim.hu/hmmtop/> or TMPred, [http://www.ch.embnet.org/software/TMPRED\\_form.html](http://www.ch.embnet.org/software/TMPRED_form.html)).

The different variants were classified in three main groups named 1, 2 and 3. A fourth group of variants (approximately one third of the variants obtained for each strain) displayed another level of polymorphism based on alternative splicing or sequence insertion events. This fourth group of variants could correspond to truncated proteins which were not detected by western blotting (see Fig. 5). The classification of the first three groups takes into account the sequence similarities between the coding sequences corresponding to the 234-residue C-terminal region

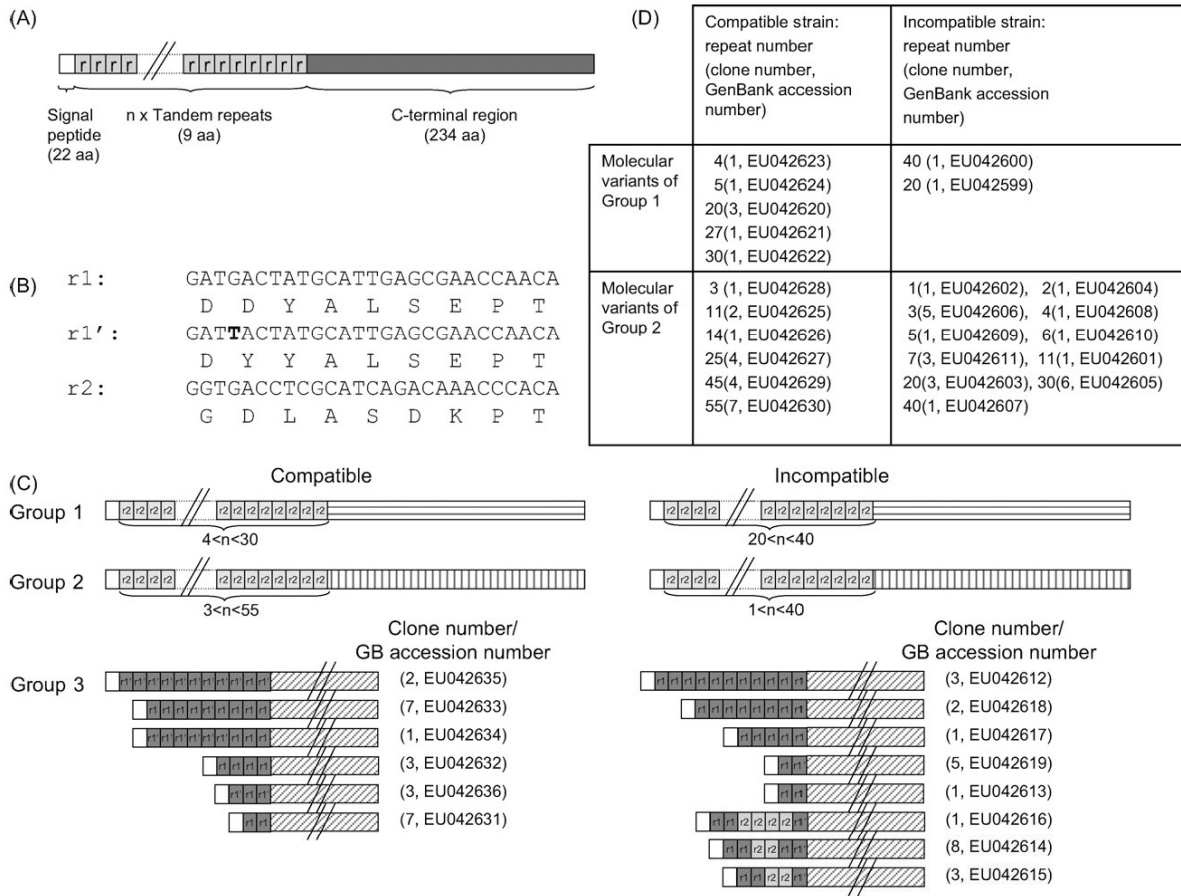


Fig. 1. *Sm* PoMuc precursor: structure and polymorphism. (A) Schematic representation of the precursor structure corresponding to *Sm* mucins; r=tandem repeats. (B) Three kinds of repeats were identified in *Sm* mucins. Their nucleic and deduced amino acid sequences are presented. Two of them, r1 and r1', display only 1 nucleotide (bold) and 1 amino acid difference. The third repeat r2 is highly divergent. (C) Schematic representation of the different groups of molecular variants obtained in the two strains. For the third group, the number of clones is given. (D) The table presents the molecular variants (discriminated by repeat number) obtained for the two first groups and for each strain. The number of clones is also indicated.

of the molecular variants. The C-terminal amino acid sequences deduced from the consensus cDNA sequences obtained for each group are shown and aligned in Fig. 3. These three groups of variants are expressed in the two *S. mansoni* strains. For both strains, groups 1 and 2 share common characteristics: they are always associated with r2 tandem repeats, and a wide variability between variants of this group is observed in the number of repeats (Fig. 1C). In groups 1 and 2, the repeat number varies from 4 to 40, and 1 to 55, respectively. The number of repeats associated with the different variants and the number of molecular variants obtained for each strain are described in Fig. 1D. The number of repeats in variants containing more than 23 repeats was estimated using agarose gel electrophoresis. The nucleotide sequence identity of repetitions makes assembly of the large region containing more than 23 repeats impossible. Concerning the third group of molecular variants, major differences between strains emerge. This latter group of molecular variants is preferentially associated with r1 and r1' repeats. The same variability in repeat number as the first two groups was observed (Fig. 1C) for the C strain.

This is also true for the IC strain, but in contrast to C an additional subgroup exists in this strain: about half of the variants contain combinations of the two types of repeats r1 (or r1') and r2.

The cDNA data allows for estimation of *pI* and molecular weight of our mucin-like proteins. The calculated *pI* varies between 4.3 and 5 and these values are in good agreement with those determined by IEF. The calculated molecular weights range from 29 to 72 kDa considering the variation in length of the repeat domain but this theoretical molecular weight window is not in agreement with measured values (Fig. 2A) which range between 55 and 130 kDa. Proteins of the mucin family are often highly glycosylated and carbohydrate side chains contribute to up to 60% of the dry weight of mucin glycoproteins [30]. To verify the glycosylation of our protein of interest, we performed an in-gel glycodetection after 2D separation. The results are shown in Fig. 2B and demonstrate that our mucin-like proteins are indeed glycosylated. Consequently, we propose that the observed molecular weight shift is due to a high degree of glycosylation.

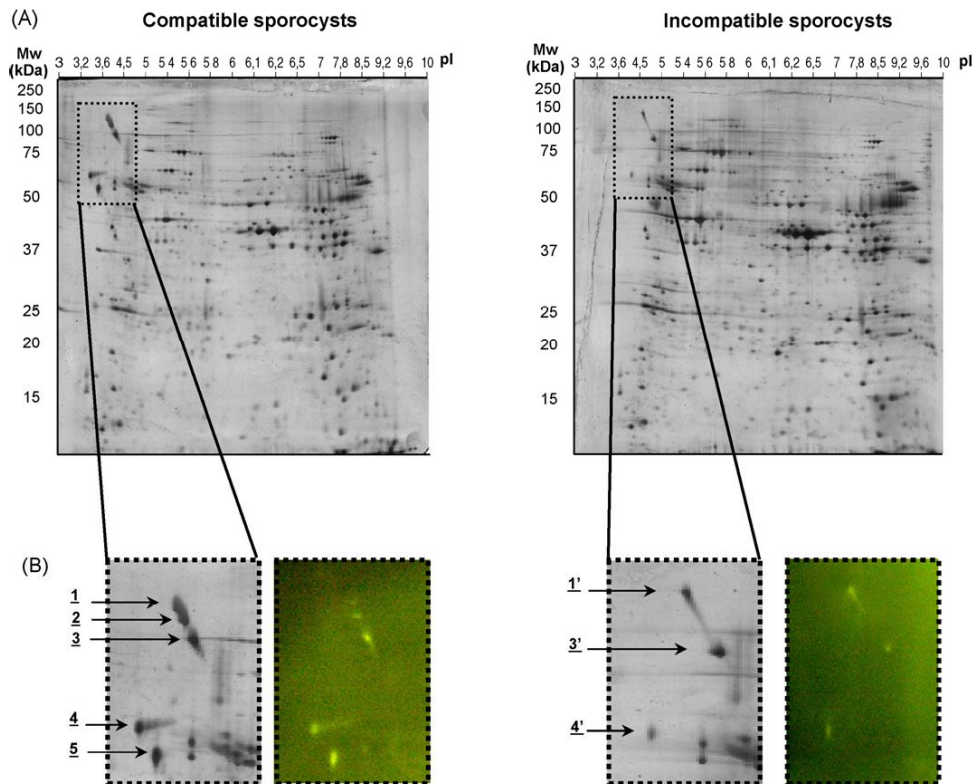


Fig. 2. *Sm PoMuc* in compatible and incompatible strains: position in 2D-gels and glycodetection. (A) Silver-stained two-dimensional gels of protein patterns from Sp1 of compatible and incompatible strains. Molecular mass (M) calibration is indicated on the left of the gel and pI calibration at the top of the gel. The pI calibration was done with standards run in the same conditions on a separate gel. Areas containing *Sm PoMuc* are boxed. (B) Enlarged views of the area of interest revealed by silver staining or glycodetection (green background). Positions of spots corresponding to *Sm PoMuc* are indicated by arrows.

To identify the different mucin-like proteins (Fig. 2B) and their corresponding transcripts, the deduced amino acid sequences of the three groups of mucin-like transcripts were compared with peptide sequences and peptide masses obtained after enzymatic digestion (trypsin and endo Lysine-C) by nanoLC-MS/MS and MALDI-ToF, respectively. No peptides corresponding to the repeat containing domains were identified after this analysis. This is probably due to post-translational modifications like the glycosylation described above. Consequently, these identifications were based on tags and fingerprints identified in the C-terminal domain of our protein of interest (see Fig. 3 for identifications). Spots 1, 2 and 3' (Fig. 2B) correspond to group 2 with coverages of 92%, 60% and 64%, respectively. Spots 3 and 1' correspond to group 1 with coverages of 56% and 64%. Finally, spots 4, 5 and 4' correspond to the third group with coverages of 32%, 19% and 27%, respectively.

The data obtained from the biochemical and precursor characterizations strongly suggests that our proteins of interest are mucins. Therefore, we decided to call them *S. mansoni* polymorphic mucins (*Sm PoMuc*). A rapid analysis of the genome assembly from *S. mansoni* reveals that *Sm PoMuc* are encoded by a multigene family, the three groups of *Sm PoMuc* being encoded by different genes. The analysis of gene structure and the genomic organization of this multigene family will be presented in a future publication.

### 3.2. *Sm PoMuc* are expressed only by larval stages interacting with molluscs

To identify the *S. mansoni* life-cycle stages that express *Sm PoMuc*, transcript levels were measured by quantitative RT-PCR on total RNA extracted from eggs, miracidia, sporocysts (Sp1), and adult worms (male and female) from each strain. The primers were chosen to amplify simultaneously the transcripts corresponding to the three groups of *Sm PoMuc*. As shown in Fig. 4, transcript expression was detected in eggs, miracidia and sporocysts with a clear maximum in the larval stage infective for molluscs, the miracidia. Since RNA was extracted from mature eggs that contain mature miracidia, the latter probably express the transcripts detected in eggs.

Western blotting reveals the same protein expression profile for miracidia and Sp1 for the two strains (Fig. 5). As in the 2D gel profile (Fig. 2A), differences in the molecular weights of *Sm PoMuc* glycoproteins between strains were detected (Fig. 5).

### 3.3. *Sm PoMuc* are located in the apical gland and secreted

Since our data indicates that the N-terminal VNTR region could be extensively glycosylated, the peptides designed for antibody production were chosen in the C-terminal conserved region of the protein. Consequently, we are unable to discrimi-

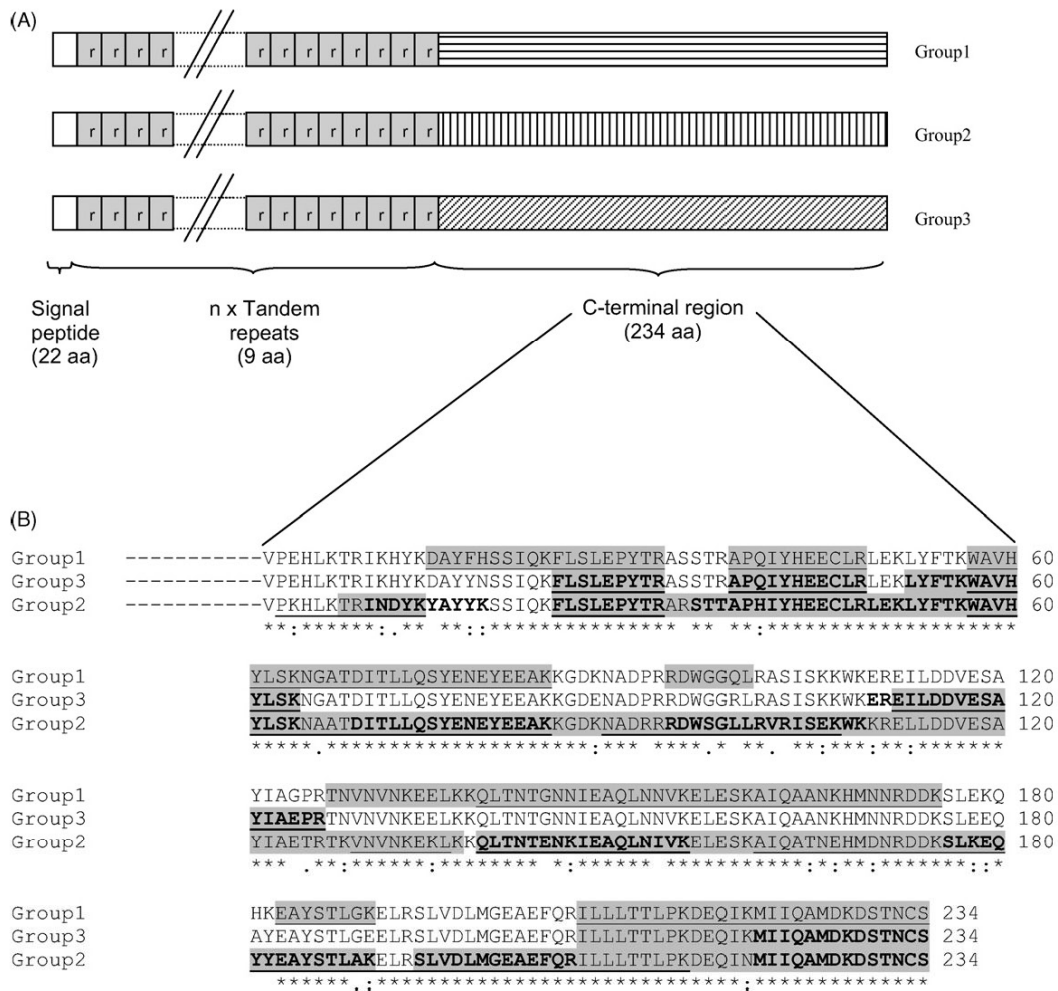


Fig. 3. (A) Schematic representation of the precursor structure corresponding to the three groups of *Sm* PoMuc; r= tandem repeats. (B) Deduced amino acid sequences and alignment of the C-terminal domains of the precursors corresponding to the three groups of *Sm* PoMuc cDNA. Conserved positions are indicated by an asterisk. Tags (obtained by LC-MSMS) and theoretical fragments matching the peptide mass fingerprint (obtained by Maldi-ToF after digestion with trypsin or Endo Lysine-C) obtained from the different spots are shown. Spots 1, 2 and 3' (see Fig. 2B) correspond to group 2 and identified peptides are highlighted in grey, in bold and underlined, respectively. Spots 3 and 1' correspond to group 1 and identified peptides are underlined and highlighted in grey, respectively. Finally, spots 4, 5 and 4' correspond to the third group and identified peptides are highlighted in grey, underlined and in bold, respectively.

nate between *Sm* PoMuc 1, 2 and 3 and consider positive signals as representing all forms of *Sm* PoMuc. Furthermore, because no information is available on the three-dimensional structure of this C-terminal region, a mixture of 4 peptides distributed along this 234-residuc tract was used for immunization. IgGs were purified by immunochemical procedures and we confirmed by ELISA that the antibody recognized all 4 peptides.

Immunofluorescence experiments showed that both *Sm* PoMuc expressing stages (miracidia and Sp1) are labelled (Fig. 6). Pre-adsorption of anti-*Sm* PoMuc IgG with synthetic peptides abolished the positive staining, demonstrating its specificity (data not shown). The overlay of light microscopy and the corresponding fluorescent images shows the localization of *Sm* PoMuc in the apical gland (Fig. 6).

Since the apical gland is known to contain secreted proteins, and because the precursors of *Sm* PoMuc possess a predicted signal peptide (see above), we investigated the presence of *Sm*

PoMuc antigens in ESP obtained by *in vitro* culture of Sp1. *Sm* PoMuc are detected by Western blot in ESP (Fig. 5B). The revealed bands display molecular weights different from those revealed in Sp1 or miracidial extracts. We conclude that these secreted *Sm* PoMuc are subjected to post-translational processing. In addition, it is possible that the paucity of material in ESP leads to fewer and less intense bands compared to those revealed in Sp1 and miracidia (Fig. 5A).

#### 4. Discussion

In order to identify molecular determinants involved in the compatibility polymorphism in *B. glabrata*/*S. mansoni* interactions [12], we developed a proteomics approach comparing two strains of *S. mansoni*, one compatible (strain C), the other incompatible (strain IC) towards the same mollusc host strain. This study, published in the same issue of this journal,

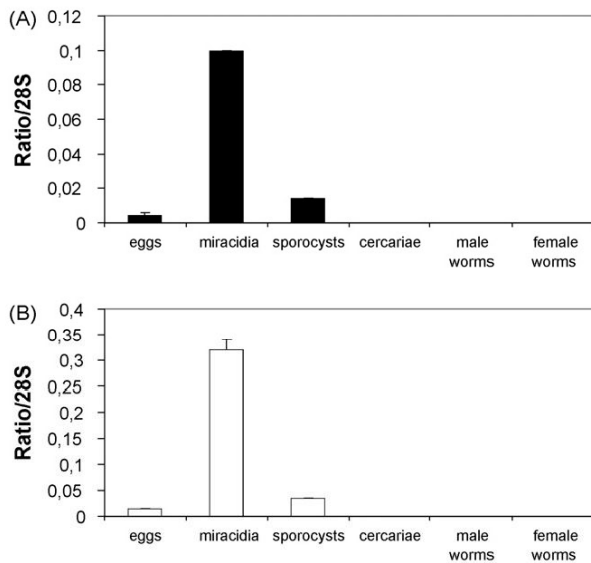


Fig. 4. Ratios of transcript levels corresponding to *Sm* PoMuc (compatible strain, A and incompatible strain, B) in different developmental stages of *S. mansoni*. Ratios were determined using real-time quantitative PCR and are expressed relative to rRNA 28s expression levels. Each histogram represents the average value of triplicates  $\pm$  S.D.

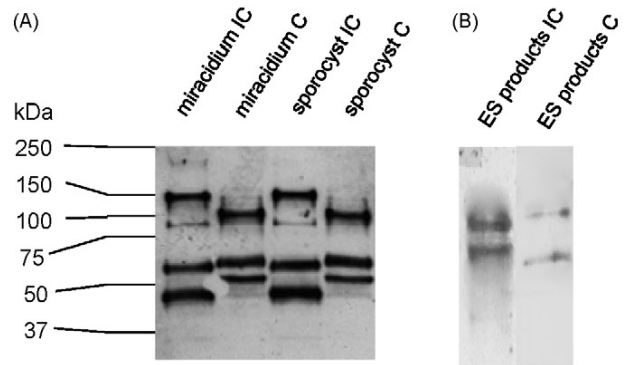


Fig. 5. Western blot for *Sm* PoMuc in extracts of miracidia and Sp1 (A) and in Excretion–Secretion Products (B) from the compatible, C, and incompatible, IC, strains.

revealed several candidate proteins, among which mucin-like proteins were identified. These proteins, that we called *Sm* PoMuc, display a domain containing a variable number of tandem repeats (VNTR). The key roles of mucins in host-parasite interactions have been extensively reported [14,15,17]. We hypothesized therefore that *Sm* PoMuc could contribute to the compatibility polymorphism of the *S. mansoni*/*B. glabrata* interaction. To further investigate the possible role played by *Sm* PoMuc, we decided to analyze their expression patterns in detail.

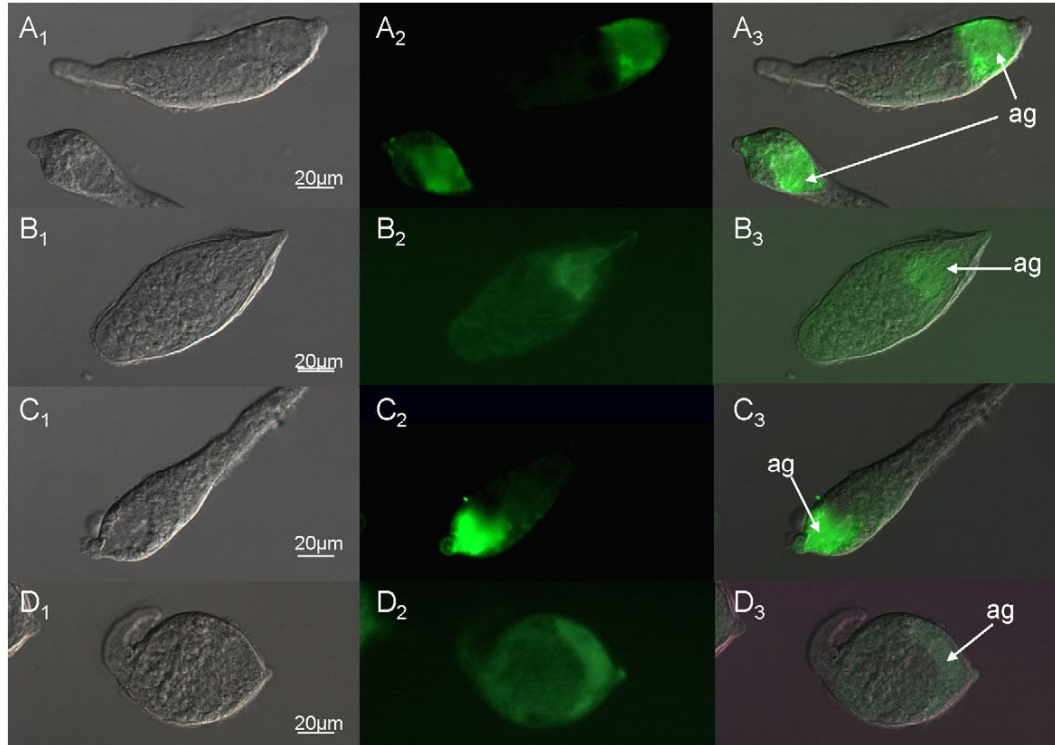


Fig. 6. Immunodetection of *Sm* PoMuc in *S. mansoni* sporocyst (A and C) and miracidium (B and D) of the compatible (A and B) and incompatible (C and D) strains. Light microscopy image (X<sub>1</sub>) and the corresponding fluorescent image (X<sub>2</sub>) documenting the immunostaining by epifluorescence using anti-*Sm* PoMuc IgG detected by secondary antibody labelled with Alexa Fluor 488. Overlay (X<sub>3</sub>) shows the localization of immunostaining in the apical gland, ag.



EST assembly and 5'/3'-RACE-PCR experiments allowed us to generate PCR primers that amplified the complete CDS of the proteins of interest. PCR experiments revealed a high degree of polymorphism, and comprehensive sequence analysis of the PCR products revealed four groups of cDNA sequences in both strains. The first three groups correspond to the different proteins identified in the proteomic approach and are discussed here. The fourth group corresponds to truncated or chimeric transcripts (data not shown). The corresponding precursors are composed of three distinct domains: a signal peptide of 22 amino-acids followed by a domain containing a variable number of 9-residue tandem repeats and a C-terminal domain of 234 amino-acids. Three types of repeats were identified and named r1, r1' and r2. Each group contains cDNAs that are identical in the deduced amino-acid sequence of the 234-residue C-terminal domain. Extensive analysis revealed a high degree of polymorphism within and between the IC and C strains. The polymorphism revealed within the two first groups consists in a variable number of r2 repeats (1–55). The major difference between strains is observed for the third group of *Sm* PoMuc. This latter displays a VNTR domain containing only r1 and/or r1' in strain C. In the strain IC, this VNTR combination was observed, but interestingly, a strain-specific subgroup of molecular variants containing both repeats r2 and r1/r1' was also identified.

This VNTR region of *Sm* PoMuc is of particular interest since it is rich in S, T and P and *in silico* prediction identifies this region as a potential target of glycosylation. Moreover, we demonstrated that *Sm* PoMuc proteins are indeed highly glycosylated. The VNTR region is therefore an excellent candidate for polymorphic glycosylation. Differences in the repeat type and number in the molecular variants could be responsible for differences in glycosylation status. Although there is no sequence similarity between *Sm* PoMuc and other vertebrate or parasite mucins, our findings clearly show that *Sm* PoMuc share the structural characteristics of the mucin family of proteins.

The polymorphism we described in the present study occurs at different levels. The first level is due to differences in repeat number within the VNTR region. This phenomenon was observed previously for numerous mucins from protozoa and helminths [14,15], but also for other parasite molecules displaying a VNTR region. Examples are the SOWgp cell surface proteins from the pathogenic fungi *Coccidioides immitis* [31] and *C. posadasii* [32] and different proteins from *Plasmodium*, the circumsporozoite proteins (CSP [33]) or the merozoite surface proteins (Msp-1 and Msp-2 [34,35]). It was proposed that these latter molecules could serve as a “smoke-screen” defence that renders the immune response of the host ineffective [32,36]. In the case of *Plasmodium* proteins, the repetitive regions are not glycosylated but they are, as in the case of *Sm* PoMuc, highly polymorphic [37,38]. Polymorphism generated by VNTR is well documented and new alleles can arise by duplication/deletion of the repeated DNA sequences through recombination involving unequal crossing-over [37]. Our data suggest that a similar process operates to generate the variability of *Sm* PoMuc. The second level of polymorphism is unique to *Sm* PoMuc. We demonstrated a combinatorial polymorphism specific for the IC strain in which we have identified molecular

variants containing a combination of repeats that does not exist in strain C. Both VNTR and tandem repeat combination polymorphism could rapidly generate a considerable diversity from a limited number of genes.

Expression analysis of *Sm* PoMuc strengthens the potential of these molecules as key determinants of the compatibility polymorphism in *S. mansoni*/*B. glabrata* interaction. It could be argued that so far no direct evidence exists for such a function for *Sm* PoMuc. However, we provide ample indirect evidence. *Sm* PoMuc genes are only transcribed in stages that interact with molluscs with the peak of expression in the mollusc infective stage, the miracidium. Furthermore, peptide signal prediction indicates a possible secretion of these molecules. This hypothesis was strengthened by (i) immunolocalization of *Sm* PoMuc in the apical gland and (ii) the detection of *Sm* PoMuc in the ESP by Western blotting. Our data show that these molecules are released in the immediate environment of the miracidia and Sp1. The apical gland is an organ involved in producing secretions necessary for the successful penetration of miracidia into snail tissue [39,40] and is known to persist for a number of days after miracidium to sporocyst transformation *in vitro* [41,42] and *in vivo* [43,44]. The gland cells could therefore play a role in infestation and post-infestation processes in the first hours following the contact between the host and its parasite. The rapid recognition, encapsulation and killing of the IC strain would easily be explained by the differences evidenced by our two studies. Several lines of evidence point therefore towards a model in which these differences are responsible for recognition or non-recognition in the case of the IC and C strains, respectively.

Parasite-derived mucins or mucin-like molecules have been extensively described in different protozoan and helminth parasites, but their function has not yet been completely elucidated. They could have roles in host recognition, penetration, adhesion and invasion of host cells, immunoprotection, immunomodulation and in the avoidance of host immune processes [14,15]. Among these molecules described in parasites, the mucin-like proteins of *Trypanosoma cruzi* share numerous characteristics with *Sm* PoMuc (see Ref. [17] for review), such as their structure, a high level of glycosylation and polymorphism. A major difference is that these molecules, encoded by the TeMUC I and II genes, like other surface components of parasitic protozoa, are typically anchored to the outer phospholipid layer of the plasma membrane by GPI. Nevertheless, secreted mucins have been characterized in helminths such as the infective larvae of the parasitic nematode *Toxocara canis* [45]. These latter secreted mucins might simply create an immunological smoke-screen by generating antigen-antibody complexes away from the parasite [46]. An attractive hypothesis could be that *Sm* PoMuc could also create an immunological smoke-screen able to block pattern recognition receptors, thus avoiding recognition and the host immune reaction.

However, compatibility polymorphism evidenced in our system could alternatively be based on a lock and key mechanism involving *Sm* PoMuc and host recognition partner molecules. In this case, positive recognition of the compatible variable antigen would determine success of the infection. Consequently, to further investigate this hypothesis, the potential partners of

*Sm* PoMucs will be investigated using co-immunoprecipitation experiments. Diversified host molecules like FREPS [10,13] could be ideal candidates.

The data described in the two papers published in this issue describe novel mucin-like proteins from *S. mansoni*. These *Sm* PoMucs are attractive candidates to explain the compatibility polymorphism occurring in our system. Studies on gene structure, genomic organization and functional characterization are underway and will shed light on the exact role of these proteins in the infection of invertebrate hosts.

## Acknowledgements

This work was supported by the CNRS. The Proteomics facility used in this study was funded by the European Community (FEDER) the Région Nord-Pas de Calais (France), the CNRS, the G enopole de Lille and the Universit e des Sciences et technologies de Lille. We are grateful to Adeline Page (Lille University) for performing mass spectrometry analysis. We thank Dr Val erie Delorme (Perpignan University) for assistance and free access to her epifluorescence microscope. We thank Jean-Fran ois Allienne, Anne Rognon, R emi Emans, Bernard Dejean and Pierre Tisseyre for technical assistance.

## References

- [1] Combes C. Selective pressure in host-parasite systems. *J Soc Biol* 2000;194:19–23.
- [2] Du Pasquier L, Zucchetti I, De Santis R. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol Rev* 2004;198:233–48.
- [3] Azumi K, De Santis R, De Tomaso A, et al. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: “waiting for Godot”. *Immunogenetics* 2003;55:570–81.
- [4] Finlay BB, Mcfadden G. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* 2006;124:767–82.
- [5] Scherf A. A greedy promoter controls malarial variations. *Cell* 2006;124:251–3.
- [6] Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. *Nature* 2004;430:174–80.
- [7] Cannon JP, Haire RN, Schmitker N, Mueller MG, Litman GW. Individual protochordates have unique immune-type receptor repertoires. *Curr Biol* 2004;14:R465–6.
- [8] Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. *Proc Natl Acad Sci USA* 2000;97:13156–61.
- [9] Watson FL, Puttmann-Holgado R, Thomas F, et al. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 2005;309:1874–8.
- [10] Zhang SM, Adema CM, Kepler TB, Loker ES. Diversification of Ig superfamily genes in an invertebrate. *Science* 2004;305:251–4.
- [11] Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. *Lancet* 2006;368:1106–18.
- [12] Theron A, Coustau C. Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *J Helminthol* 2005;79:187–91.
- [13] Adema CM, Hertel LA, Miller RD, Loker ES. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc Natl Acad Sci USA* 1997;94:8691–6.
- [14] Hicks SJ, Theodoropoulos G, Carrington SD, Corfield AP. The role of mucins in host-parasite interactions. Part I-protozoan parasites. *Parasitol Today* 2000;16:476–81.
- [15] Theodoropoulos G, Hicks SJ, Corfield AP, Miller BG, Carrington SD. The role of mucins in host-parasite interactions. Part II. Helminth parasites. *Trends Parasitol* 2001;17:130–5.
- [16] Rathore D, Nagarkatti R, Jani D, et al. An immunologically cryptic epitope of *Plasmodium falciparum* circumsporozoite protein facilitates liver cell recognition and induces protective antibodies that block liver cell invasion. *J Biol Chem* 2005;280:20524–9.
- [17] Buscaglia CA, Campo VA, Frasch a C, Di Noia JM. Trypanosoma cruzi surface mucins: host-dependent coat diversity. *Nat Rev Microbiol* 2006;4:229–36.
- [18] Dissous C, Capron A. Isolation and characterization of surface antigens from *Schistosoma mansoni* schistosomula. *Mol Biochem Parasitol* 1981;3:215–25.
- [19] Theron A, Pages JR, Rognon A. *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp Parasitol* 1997;85:1–9.
- [20] Chemin E. Observations on hearts explanted in vitro from the snail *Australorbis glabratus*. *J Parasitol* 1963;49:353–64.
- [21] Coustau C, Mitta G, Dissous C, et al. *Schistosoma mansoni* and *Echinostoma caproni* excretory-secretory products differentially affect gene expression in *Biomphalaria glabrata* embryonic cells. *Parasitology* 2003;127:533–42.
- [22] Vergote D, Bouchut A, Sautiere PE, et al. Characterisation of proteins differentially present in the plasma of *Biomphalaria glabrata* susceptible or resistant to *Echinostoma caproni*. *Int J Parasitol* 2005;35:215–24.
- [23] Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 1996;68:850–8.
- [24] Bouchut A, Sautiere PE, Coustau C, Mitta G. Compatibility in the *Biomphalaria glabrata/Echinostoma caproni* model: potential involvement of proteins from hemocytes revealed by a proteomic approach. *Acta Trop* 2006;98:234–46.
- [25] Guillou F, Mitta G, Dissous C, Pierce R, Coustau C. Use of individual polymorphism to validate potential functional markers: case of a candidate lectin (BgSel) differentially expressed in susceptible and resistant strains of *Biomphalaria glabrata*. *Comp Biochem Physiol B* 2004;138:175–81.
- [26] Vaitukaitis J, Robbins JB, Nieschlag E, Ross GT. A method for producing specific antisera with small doses of immunogen. *J Clin Endocrinol Metab* 1971;33:988–91.
- [27] Hancock CD, Evans GI. Production and characterization of antibodies against synthetic peptides. *Methods in Molecular Biology: Immunochemical Protocols*. Humana Press; 1992.
- [28] Porath J, Olin B. Immobilized metal ion affinity adsorption and immobilized metal ion affinity chromatography of biomaterials. Serum protein affinities for gel-immobilized iron and nickel ions. *Biochemistry* 1983;22:1621–30.
- [29] Nielsen H, Engelbrecht J, Brunak S, Von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997;10:1–6.
- [30] Schenkman S, Ferguson MA, Heise N, De Almeida ML, Mortara RA, Yoshida N. Mucin-like glycoproteins linked to the membrane by glycosylphosphatidylinositol anchor are the major acceptors of sialic acid in a reaction catalyzed by trans-sialidase in metacyclic forms of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 1993;59:293–303.
- [31] Hung CY, Ampel NM, Christian L, Seshan KR, Cole GT. A major cell surface antigen of *Coccidioides immitis* which elicits both humoral and cellular immune responses. *Infect Immun* 2000;68:584–93.
- [32] Johannesson H, Townsend JP, Hung CY, Cole GT, Taylor JW. Concerted evolution in the repeats of an immunomodulating cell surface protein, SOWgp, of the human pathogenic fungi *Coccidioides immitis* and *C. posadasii*. *Genetics* 2005;171:109–17.
- [33] Nussenzweig V, Nussenzweig RS. Circumsporozoite proteins of malaria parasites. *Cell* 1985;42:401–3.
- [34] Holder AA. The precursor to major merozoite surface antigens: structure and role in immunity. *Prog Allergy* 1988;41:72–97.
- [35] Smythe JA, Coppel RL, Brown GV, Ramasamy R, Kemp DJ, Anders RF. Identification of two integral membrane proteins of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 1988;85:5195–9.

- [36] Nardin EH, Nussenzweig RS. T cell responses to pre-erythrocytic stages of malaria: role in protection and vaccine development against pre-erythrocytic stages. *Annu Rev Immunol* 1993;11:687–727.
- [37] Rich SM, Hudson RR, Ayala FJ. *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. *Proc Natl Acad Sci USA* 1997;94:13040–5.
- [38] Rich SM, Ayala FJ. Population structure and recent evolution of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 2000;97:6994–7001.
- [39] Yoshino TP, Lodes MJ, Rege AA, Chappell CL. Proteinase activity in miracidia, transformation excretory-secretory products, and primary sporocysts of *Schistosoma mansoni*. *J Parasitol* 1993;79:23–31.
- [40] Dresden MH, Sung CK, Deelder AM. A monoclonal antibody from infected mice to a *Schistosoma mansoni* egg proteinase. *J Immunol* 1983;130:1–3.
- [41] Basch PF, Diconza JJ. The miracidium-sporocyst transition in *Schistosoma mansoni*: surface changes in vitro with ultrastructural correlation. *J Parasitol* 1974;60:935–41.
- [42] Wipperfsteg V, Kapp K, Kunz W, Grevelding CG. Characterisation of the cysteine protease ER60 in transgenic *Schistosoma mansoni* larvae. *Int J Parasitol* 2002;32:1219–24.
- [43] Pan SC. The fine structure of the miracidium of *Schistosoma mansoni*. *J Invertebr Pathol* 1980;36:307–72.
- [44] Pan SC. *Schistosoma mansoni*: the ultrastructure of larval morphogenesis in *Biomphalaria glabrata* and of associated host-parasite interactions. *Jpn J Med Sci Biol* 1996;49:129–49.
- [45] Loukas A, Hintz M, Linder D, et al. A family of secreted mucins from the parasitic nematode *Toxocara canis* bears diverse mucin domains but shares similar flanking six-cysteine repeat motifs. *J Biol Chem* 2000;275:39600–7.
- [46] Marin MS, Prieto M, Martin JM, Casais R, Boga JA, Parra F. Identification and expression of a *Fasciola hepatica* gene encoding a gut antigen protein bearing repetitive sequences. *Mol Biochem Parasitol* 1992;55:155–65.



## CHAPITRE 3

### **Le polymorphisme de mucines sécrétées par un parasite métazoaire (*Schistosoma mansoni*) dans l'interaction avec son hôte invertébré (*Biomphalaria glabrata*) : recombinaison, épissage alternatif, transcription indépendante et glycosylation**

#### **Publication n°3**

Polymorphism of Mucins secreted by a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*): involvement of recombination, alternative splicing, locus-specific transcription and glycosylation. **PLoS Neglected Tropical Diseases** (2008); *Accepted*.

Les résultats obtenus précédemment par l'approche protéomique comparative (Roger *et al.*, 2008a) révèlent que la principale différence entre des souches de *S. mansoni* compatible et incompatible est la présence de variants protéiques de type mucine. La structure du précurseur, le profil d'expression, la localisation des *SmPoMuc* ainsi que le polymorphisme qui s'exprime au niveau intra- et inter-souche ont également été décrits (Roger *et al.*, 2008b) et nous ont amené à nommer ces protéines *SmPoMuc*.

Nous nous sommes focalisés, dans ce 3<sup>ème</sup> travail, sur la recherche des mécanismes à la base du polymorphisme des *SmPoMuc*. Ces protéines sont codées par une famille multigénique composée d'un faible nombre de gènes et de pseudogènes. Le haut degré de polymorphisme des *SmPoMuc* semble trouver son origine à différents niveaux. Tout d'abord au niveau génomique, la structure des gènes et leur organisation, comportent des caractéristiques favorables aux événements de recombinaison. Ces événements, sont vraisemblablement à la base de l'évolution de cette famille de gènes et en partie du polymorphisme observé au niveau intra- et inter-souche. Ensuite, au niveau des transcrits, il semble que des mécanismes de *trans*-épissages, épissages alternatifs ou encore épissages aberrants soient également à l'origine du polymorphisme des *SmPoMuc*. Nous avons montré que ce polymorphisme se répercutait au niveau protéique et plus particulièrement sur le statut de glycosylation. Le rôle important que jouent les carbohydrates dans les interactions hôte / parasite, suggère qu'un statut de glycosylation différentiel pourrait être à l'origine

d'une reconnaissance différentielle par le système immunitaire de l'hôte. Enfin, nous avons mis en évidence un polymorphisme d'expression des *SmPoMuc* au niveau individuel. Il se traduit par un profil d'expression spécifique pour chaque individu. En effet, chaque larve de *S. mansoni* est capable d'exprimer une combinaison de variant(s) *SmPoMuc* unique. Des résultats préliminaires suggèrent que le polymorphisme d'expression des *SmPoMuc* pourrait être régulé épigénétiquement.

Nous montrons ici pour la première fois la présence de variants moléculaires polymorphes chez un parasite métazoaire (*S. mansoni*) interagissant avec un hôte invertébré (*B. glabrata*). De plus, cette cascade de mécanismes complexes – un « chaos contrôlé » – permet de générer un haut niveau de polymorphisme et ceci à partir d'un nombre réduit de gènes. Cette caractéristique est unique pour des variants moléculaires de parasites eucaryotes et contraste en effet avec ceux des parasites protozoaires générés à partir d'un grand répertoire de gènes.

**Polymorphism of Mucins secreted by a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*) : involvement of recombination, alternative splicing, locus-specific transcription and glycosylation.**

Research Article

<sup>a</sup>Emmanuel Roger, <sup>a</sup>Christoph Grunau, <sup>b</sup>Raymond J Pierce, <sup>c</sup>Hirohisa Hirai, <sup>a</sup>Benjamin Gourbal, <sup>a</sup>Richard Galinier, <sup>a</sup>Rémi Emans, <sup>d</sup>Italo M Cesari, <sup>a</sup>Céline Cosseau, <sup>a,\*</sup>Guillaume Mitta.

<sup>a</sup>*Parasitologie Fonctionnelle et Evolutive,*

*UMR 5244, CNRS Université de Perpignan,*

*52 Ave Paul Alduy, 66860 Perpignan Cedex, France*

<sup>b</sup>*Inserm, U 547, Université Lille 2, Institut Pasteur de Lille, IFR 142, 1 rue du Professeur A. Calmette, 59019 – Lille, France*

<sup>c</sup>*Primate Research Institute, Kyoto University, Inuyama, Aichi 484-8506, Japan*

<sup>d</sup>*Laboratoire de Parasitologie, Faculté de Médecine, U.L.B CP 616, 808 route de Lennik- B-1070 Bruxelles, Belgique*

**\*Corresponding author: Guillaume Mitta**

*UMR 5244, CNRS Université de Perpignan,*

*52 Ave Paul Alduy, 66860 Perpignan Cedex, France*

Phone: 33 (0)4 68 66 21 88

Fax: 33 (0)4 68 66 22 81

e-mail: [mitta@univ-perp.fr](mailto:mitta@univ-perp.fr)

Keywords: host-parasite interaction, polymorphism, multi-gene family, *Schistosoma mansoni*, *SmPoMuc*, molecular variants.

Running head: *SmPoMuc* Polymorphism: different levels of control

### **Abbreviations**

C strain: Compatible strain

IC strain: Incompatible strain

FREP: Fibrinogen Related Protein

IgSF: Immunoglobulin Superfamily

PAMP: Pathogen-Associated Molecular Pattern

PRR: Pattern Recognition Receptor

*SmPoMuc*: *Schistosoma Mansoni* Polymorphic Mucins

TR: Tandem Repeats



## **Abstract**

Recent studies have shown that invertebrate immunity relies on diverse immune receptors. This suggests the existence of individual or population-based polymorphism permitting the survival of individuals or species confronted with parasites and raises the question of whether antigenic diversity and polymorphism also exist in parasites of invertebrates, as an adaptive response expected in arms races. We investigated this diversity/polymorphism counterpart of parasite molecules (*Schistosoma mansoni* Polymorphic Mucins, *SmPoMucs*) that were recently identified in our laboratory by comparing two schistosome strains that are compatible and incompatible towards the same mollusc host strain of *Biomphalaria glabrata*. In the present work, we elucidate the complex cascade of mechanisms that generates polymorphism in *SmPoMucs*. Our data support the view that *S. mansoni* has evolved a complex hierarchical system that efficiently generates high polymorphism based on a relatively low number of genes.

## **Introduction**

The comprehension of host-parasite interactions represents a major challenge in evolutionary biology. Parasites are responsible for substantial deleterious effects on their hosts, and therefore represent a major driving force for their evolution. In parallel, parasites have to cope with the evolving host-defence mechanisms, i.e. they must co-evolve with their host to avoid elimination. This adaptation of the Red Queen hypothesis (Van Valen 1974) to host-parasite systems predicts that an arms race takes place in which both host and parasite develop mechanisms that generate diversity and polymorphism of molecules that play key roles in the host-parasite interplay (Ebert 2008).

In vertebrate hosts, the most striking example is the exceptional diversity of antigen-specific receptors of the adaptive immune system of jawed vertebrates. This system depends on somatic gene rearrangement and hypermutation (Tonegawa 1983; Reynaud et al. 1994; Di Noia and Neuberger 2007). For the pathogen counterparts, a variety of mechanisms permitting evasion of the host's immune response exist in pathogenic bacteria and viruses (Finlay and McFadden 2006). In contrast, antigenic variation is the predominant, or even the sole strategy for most of the eukaryotic parasites (Lopez-Rubio, Riviere and Scherf 2007). In the case of invertebrate hosts and their parasites, the picture is believed to be completely different since the prevailing view is that invertebrates have no acquired adaptive immunity, and that their immune system is innate and "non-specific". The detection of parasites by invertebrates was thought to rely exclusively on invariable germline-encoded Pattern Recognition Receptors (PRRs) that recognize pathogen-associated molecular patterns (PAMPs) (Medzhitov and Janeway 1997). Nevertheless, recent studies have shaken this paradigm by providing evidence for novel and diverse immune receptor sequences in protochordates (*Amphioxus*; (Cannon et al. 2004)), in echinoderms (sea urchin; (Pancer 2000)), insects (*Drosophila melanogaster* and *Anopheles gambiae*; (Watson et al. 2005;

Dong, Taylor and Dimopoulos 2006)) and molluscs (*Biomphalaria glabrata*; (Zhang et al. 2004)). These results suggest the existence of individual or population-based polymorphism permitting the survival of individuals or species confronted with parasites. These recent observations raise the question of whether diversity and polymorphism also exist in key compatibility molecules expressed by parasites (or parasite stages in the case of multi-host parasites) interacting with invertebrate hosts, whether these molecules are subject to variation and whether molecular polymorphism is at the core of interaction with the invertebrate host immune system.

To address these questions, we focused our study on a host-parasite model where the co-evolutionary dynamics is accessible: a model in which only some particular host and parasite phenotypes are compatible. We analyzed the interaction between *Schistosoma mansoni*, the agent of human intestinal schistosomiasis (Gryseels et al. 2006) and its invertebrate intermediate host, the gastropod mollusc *B. glabrata*. In this interaction, compatibility polymorphism occurs (Theron and Coustau 2005), i.e. in natural populations some snail/schistosome combinations are compatible and others are not. We hypothesized that this compatibility polymorphism is dependent on diversification mechanisms that act on key molecules such as the PRRs of the immunoglobulin superfamily (IgSF) characterized recently in *B. glabrata* (FREPs, (Adema et al. 1997)) and parasite antigens. The *FREPs* genes encode lectin-like hemolymph polypeptides that can precipitate soluble antigens derived from trematodes (Adema et al. 1997) and consist of one or two amino-terminal IgSF domains and a carboxyl-terminal fibrinogen domain. These molecules undergo mutations and recombinatorial processes that lead to diversification (Zhang et al. 2004). According to the arms race hypothesis, polymorphic molecular variants expressed by schistosome larvae in intermediate hosts could explain the observed compatibility polymorphism. While some parasites like *Plasmodium falciparum* or *Trypanosoma sp.* have developed a rich repertoire of

mechanisms to generate polymorphic variants, the system that generates diversity in *S. mansoni* is so far unknown. We have previously shown by a comparative proteomics approach (Roger et al. 2008b) that the principal difference between compatible and incompatible strains of *S. mansoni* is the presence of particular *SmPoMuc* protein variants. We have described the principal characteristic of the coding sequence, gene expression patterns and protein localization of *SmPoMuc* (Roger et al. 2008a). We have shown that these proteins are expressed and secreted by miracidia and sporocysts, i.e the larval stages that interact with the mollusk. In addition, we have described their high level of intra- and inter-strain polymorphism. Here, we elucidate the complex cascade of mechanisms that confers polymorphism to *SmPoMuc*. We show that *SmPoMuc* is coded by a multi-gene family. Genes are transcribed in individual-specific manner, and for each gene, multiple splice variants exist. The incidence of this polymorphism on *SmPoMuc* glycosylation status is demonstrated. Our data support the view that *S. mansoni* has evolved a complex hierarchical system that efficiently generates highly polymorphic variants based on a relatively low number of genes.

## **Materials and Methods**

### **Culture of *S.mansoni***

The compatible Brazilian (strain C) and incompatible Guadeloupean (strain IC) strains of *Schistosoma mansoni* were maintained in (i) *Biomphalaria glabrata* strains Bg.Bra and Bg.Gua, respectively and (ii) hamsters (*Mesocricetus auratus*) as described previously (Theron, Pages and Rognon 1997). Adult worms and primary (mother) sporocysts (Sp1) were obtained as previously described (Roger et al. 2008a).

### **Protein extraction, separation and detection**

Two-D gel proteomic analysis was conducted according to procedures developed previously (Roger et al. 2008a; Roger et al. 2008b). Briefly, the total proteome of C and IC sporocysts originating from different hamster livers was extracted using 2D lysis buffer (8 M urea, 40 mM Tris, 4% CHAPS, 60 mM DTT). One hundred µg of protein were separated in the first dimension using 17 cm Ready Strip IPG Strips (Bio-Rad, Hercules, CA). Different pH gradients were used, a pH 3-10 non-linear gradient to have a broad overview of total protein distribution, and a pH 3-6 narrow-range gradient for increased resolution in the *SmPoMuc* region. Isoelectrofocusing (IEF) was performed with voltage gradually increasing to 8000 V for 180 000 Vh at 20°C. Proteins were separated by 12% SDS-PAGE, visualized by silver staining (Patton 2002) and the 2D gels were scanned using a densitometer (GS-800 Calibrated Densitometer, Bio-Rad, Hercules, CA).

### **Chemical deglycosylation and western blotting**

Chemical deglycosylation of *SmPoMuc* proteins was performed using trifluoromethanesulfonic acid (TFMSA) according to a previously described procedure (Edge 2003). Briefly, 40 µg of each sample was treated with TFMSA and ½ volume of anisole and incubate on ice. TFMSA was neutralized with N-ethylmorpholine (NEM) and deglycosylated proteins were precipitated with acetone overnight at -20°C. Protein pellets were re-suspended

in deionised water and Laemmli buffer and separated on a 12 % SDS-PAGE. Proteins were transferred onto a nitrocellulose membrane (Hybond ECL, GE Healthcare) using semi-dry transfer (SemiPhor, Hoefer) and submitted to Western-Blot analysis.

The membrane was blocked with 5% non-fat dry milk in PBST (pH 7.4 PBS buffer containing 0.05% tween 20) overnight at 4 °C and incubated with primary antibody (anti-*SmPoMuc* IgG purified from rabbit) (1/200 in PBST) for 1.5 hours at room temperature. Incubation with secondary antibody (peroxidase conjugated, purified anti rabbit IgG) diluted 1/5000 was done in PBST for 1.5 hours at room temperature. After incubation with each antibody, the membrane was washed 3 times for 30 minutes with agitation in PBST. Detection was realized using ECL reagents (Pierce). The membrane was incubated with peroxidase substrate for 1 minute and exposed to X-ray film (GE Healthcare).

Confirmation of the removal of carbohydrate moieties was assessed by two procedures, Alcian blue staining and lectin blotting. For Alcian blue staining the SDS-PAGE gel was fixed in 7% acetic acid, stained (0.5% Alcian Blue) and de-stained in the same solution. Lectin blots were carried out after protein electrophoresis and transfer to a nitrocellulose membrane as previously and the membrane was incubated in PBST with a specific lectin peroxidase conjugate (concanavalin A for N-glycosylation, and jacalin for O-glycosylation). Chemiluminescent detection was as described previously.

#### **DNA and RNA extraction**

Genomic DNA was extracted from *S. mansoni* adult worms using DNAzol® Reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. BAC clones containing *SmPoMuc* genes (41B11, 62F12, 62J10, 47P6, 51E8 and 45D24) were grown up and BAC DNA preparations carried out as previously described (Le Paslier et al. 2000).

Messenger RNA from individual sporocysts was isolated using the Dynabeads mRNA DIRECT™ Micro Kit (Invitrogen, Carlsbad, CA). Single Sp1 were pipetted directly into the lysis buffer and then treated according to the instructions of the supplier.

### **PCR amplification and sequencing of genomic DNA**

PCR amplification of *S. mansoni* adult worm genomic DNA or BAC DNA was performed with the Advantage® 2 PCR Enzyme System (Clontech, Palo Alto, CA). The *SmPoMuc* loci were amplified using *S. mansoni* adult worm genomic DNA, forward primer Intron2/3F1 and reverse primer Exon15R (see Table 1 for primer sequences and PCR cycling conditions) designed in conserved genomic regions namely in the introns upstream of exon 2 and in exon 15, respectively (see Figure 3 for primer positions). PCR products were cloned into pCR4-TOPO (TOPO TA Cloning® kit for sequencing, Invitrogen, Carlsbad, CA). Plasmid DNA was purified using the Wizard® Plus SV Miniprep DNA purification system (Promega, Madison, Wis.). DNA was sequenced using a dideoxy-dye-terminator method (CEQ™ DTCS-Quick Start kit, Beckman Coulter) and a CEQ™ 8000 capillary sequencer (Beckman Coulter) with M13 forward, M13 reverse and specific primers (Table 1). Sequence analysis was performed using Sequencher™ software (Gene Codes Corporation, Ann Arbor, MI).

To identify *SmPoMuc* genes in BACs 41B11, 62F12, 62J10, 47P6, 51E8 and 45D24, PCR amplification was performed using BAC DNA and primers (Table 1) that generate PCR fragments of different lengths for each *SmPoMuc* group. PCR products were separated by electrophoresis in 1% agarose gels, and visualized by staining with ethidium bromide. PCR cycling conditions were one denaturation step of 1 min at 95°C followed by 30 amplification cycles: 95°C for 30s,  $t_A$ °C for 30s and 68°C for a specific elongation time.  $t_A$ °C and elongation times specific for each primer couple are given in Table 1. Cloning and sequencing was performed as described above.

To determine the presence of tandem repeats (TR) of r1, r2 and r1/r2 combinations, PCR was done with primers that bind specifically to either repeat. Amplification with r1 forward and reverse primers reveals r1 TR only, amplification with r2 specific primers shows the presence of r2 TR, and amplification with r1 forward/ r2 reverse and r2 forward/r1 reverse primers indicates the presence of r1/r2 combinations. Primers and PCR conditions used are listed in Table 1.

The copy number of *SmPoMuc* genes was measured by quantitative PCR. Real-Time PCR analysis was performed on genomic DNAs extracted from 3 *S. mansoni* adult clones of both strains. PCR and relative quantification were performed according to previously described procedures (Linzmeier and Ganz 2005, 2006) with a Light Cycler<sup>TM</sup> (Roche Molecular Biochemicals, Germany). Specific primers for real-time quantitative PCR were designed using the Light Cycler Probe Design<sup>TM</sup> Software version 1.0 with an annealing temperature of 60°C. The primers used (Exon7F2 and Exon7R, Table 1) were chosen in a conserved region present in all *SmPoMuc* genes in the intron downstream of exon 7 (see Figure 3 for amplicon location). The single-copy gene used as a reference was Src kinase TK3 (GI:37776868) amplified using primers Src.F1 and Src.R1 (Table 1). Real time quantitative PCR cycling conditions were as previously described (Mitta et al. 2005).

#### **Nested RT-PCR on individual sporocysts of both strains**

Messenger RNAs extracted from individual Sp1 were reverse transcribed by adding the enzyme mix (Superscript II, Invitrogen, Carlsbad, CA) directly to the paramagnetic Dynabeads. Dynabeads and associated cDNA were recovered using the magnetic system, washed twice in 10mM Tris (pH=7.5) and directly submitted to PCR amplification. Primers and cycling conditions used for the first round of PCR (Exon1F12 and Exon15R) and a subsequent nested PCR (NestedExon1F and NestedExon15R) are given in Table 1. All PCR were performed with the Advantage<sup>®</sup> 2 PCR Enzyme System (Clontech, Palo Alto, CA).



PCR products were separated by electrophoresis in 1% agarose gels, visualized by staining with ethidium bromide, and cloned into pCR4-TOPO for sequencing.

#### **Fluorescence In Situ Hybridization (FISH)**

FISH was performed on *S. mansoni* sporocyst metaphase chromosome spreads with BACs 41B11 (*SmPoMuc* group 2, 163 A6) and 45D24 (*SmPoMuc* group 3, 180 B12) using techniques previously described (Hirai and LoVerde 1995; Hirai and Hirai 2004).

#### **Northern Blot**

Northern Blot was performed according to previously described procedure (Mitta et al. 1999).

#### **Southern Blot**

Twenty µg genomic DNA or 1 µg BAC DNA were respectively digested with 40 U or 10 U of EcoRV, EcoRI and BclI. DNA fragments were separated by gel electrophoresis through 0.7% agarose, and stained with ethidium bromide to confirm complete digestion. DNA was transferred to Hybond N+ membranes (Amersham Bioscience) using a vacuum blotter (model 785, Bio-Rad, Hercules, CA). Genomic repeat stretches were revealed using a 1079 bp PCR product (obtained using r2.F2 and r2.R2 primers) labeled with digoxigenin-dUTP by random priming. Hybridization and development of blots were performed with the DIG High Prime DNA Labeling and Detection Starter Kit (Roche). After stripping, the same membranes were hybridized with oligonucleotides specific for r1 and r2 repeats (CTGTTGGTTCGCTCAATGCATA, GTGACCTCGCATCAGACAAAC, respectively) 5'-labeled by DIG (Eurogentec). Positive fragments were revealed by the NBT/BCIP color reaction (Roche).

#### ***In-silico* analysis**

Complete CDS corresponding to the three groups of *S. mansoni* mucin-like proteins (groups 1, 2, 3, (Roger et al. 2008a)) were used in Blast searches against the *S. mansoni*

genome (assembly version 3.1) at the Sanger Institute ([http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s\\_mansoni](http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s_mansoni)). Matching contigs were retrieved and genomic DNA sequences were aligned with the Sequencher<sup>TM</sup> software (Gene Codes Corporation, Ann Arbor, MI). *SmPoMuc*-containing BACs were identified by a Blast search in the BAC-ends database of the Institute for Genomic Research (TIGR) ([http://www.tigr.org/tdb/e2k1/sma1/map\\_ends.shtml](http://www.tigr.org/tdb/e2k1/sma1/map_ends.shtml)) using conserved parts of the genes.

Sequences corresponding to *SmPoMuc* conserved genomic regions were aligned using Sequencher and manually inspected with the BioEdit Sequence Alignment Editor software (release 7.0.9.0). Parsimony trees were constructed using PAUP (Swofford, D) and robustness was checked by a bootstrap test (1000 replicates). Trees were visualized with TreeView 1.6.6. (<http://darwin.zoology.gla.ac.uk/~rpage/treeviewx/index.html>).

cDNA sequences were codon-aligned to the corresponding *SmPoMuc* amino acid sequences using the PAL2NAL web server (Suyama, Torrents and Bork 2006), and synonymous and non-synonymous substitution rates ( $K_S$  and  $K_N$ ) were calculated essentially as described by Nei and Gojobori (Nei and Gojobori 1986) using SNAP (<http://www.hiv.lanl.gov/content/hiv-db/SNAP/WEBSNAP/SNAP.html>). A test of neutrality was performed with the Neutrality Test 1.2 software ([http://www.hgc.sph.uth.tmc.edu/neutrality\\_test/](http://www.hgc.sph.uth.tmc.edu/neutrality_test/)). Tajima's D test was used to detect deviation of the  $K_S/K_N$  ratios from neutrality (Tajima 1989). *SmPoMuc* genes were annotated using SeqVISTA 1.9 software; paralogous sequence blocks were color-coded and highlighted to visualize recombination between members of the gene family.

Prediction of glycosylation sites in *SmPoMuc* amino acid sequences was performed using the NetOGlyc 3.1 server (<http://www.cbs.dtu.dk/services/NetOGlyc/>) that produces neural network predictions of GalNAc O-glycosylation sites in mucin-like proteins (Julenius et al. 2005).

## Results

### ***SmPoMuc* polymorphism is apparent at the protein and transcript levels.**

We previously investigated the differences in the proteome between two strains of *S. mansoni* that are compatible (C) or incompatible (IC) towards a specific *B. glabrata* strain in a study that identified the *SmPoMuc* group of proteins (Roger et al. 2008b). Evidence for *SmPoMuc* polymorphism was provided by size and charge differences of these proteins in 2D gels (Figure 1A). To better characterize this polymorphism at the protein level, we used different pH gradients in 2-D electrophoresis. In our previous study, we had used a pH 3-10 non linear gradient to obtain a broad overview of the distribution of *SmPoMucs* (Figure 1A) (Roger et al. 2008b). We have now realized a zoom-in gel using a narrow pH 3-6 range to expand the region containing these proteins (Figure 1A). This approach reveals several supplementary spots previously not observed and provides evidence that polymorphism at the protein level was probably underestimated in our previous study (Roger et al. 2008b). In a subsequent exhaustive analysis of *SmPoMuc* transcripts (Roger et al. 2008a) a large number of molecular variants were revealed that were classified into three groups corresponding to the different spot groups identified in the proteomic analysis. The deduced precursor sequences of the different cDNA variants are shown in Figure 1B. They are composed of a signal peptide (22 amino acids in length) followed by a domain containing a variable number ( $n=1$  to  $n\approx 55$ ) of tandem repeats (VNTR) of 9 amino acids. Three different types of repeats were identified: r1, r1' and r2 that were expressed in both *S. mansoni* strains (Figure 1B). For both strains, groups 1 and 2 share common characteristics: they are always associated with r2 tandem repeats, and the number of repeats is highly variable (Figure 1B). Major differences between the strains emerge in the third group of molecular variants that are preferentially associated with r1 and r1' repeats. The same variability in repeat number as the first two groups was observed (Figure 1B) for the C strain. This is also true for the IC strain, but in contrast to C an

additional sub-group exists in this strain with about half the variants containing combinations of the two types of repeats r1 (or r1') and r2 (Figure 1B). This size polymorphism of tandem repeats was also confirmed using Northern blot analysis in which a smear was obtained after hybridization with probes corresponding to *SmPoMuc* of the first group (data not shown).

Finally, we analyzed the expression of the *SmPoMucs* at the level of individual larvae. Consensus oligonucleotides were used to amplify the whole coding sequence of all *SmPoMucs* by nested RT-PCR and revealed a high degree of polymorphism between individuals for the two strains (Figure 1C). This polymorphism was extensively analyzed by sequencing in the present work (see above and Table S1). Taken together, our data give evidence for a remarkable level of polymorphism of the *SmPoMuc* molecules.

#### ***SmPoMuc* genes form a multi-gene family specific for *S. mansoni***

Based on *in-silico* investigations (Blast searches on the *S. mansoni* genome assembly v3.1, [http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s\\_mansoni/omni](http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s_mansoni/omni)), we estimated the number of *SmPoMuc* genes to be ten. Six of them correspond to full length genes (contigs Smp\_contig019963, -030125, -043854, -026239, -037561, -045752) and four of them (contigs Smp\_contig049466, -010496, -045333 and -030128) are truncated genes interrupted by a transposon insertion. To determine the number of genes in this multigene family in our strains of interest, we performed a Southern blot with DNA extracted from adult worms of the C and IC strains (200 pooled individuals for each strain) and observed several bands and a smear for both strains (Figure 2, lanes 2 and 5). Since these results could be due to *SmPoMuc* polymorphism between individuals we next analyzed *SmPoMuc* copy number by quantitative PCR using primers designed from a conserved region of *SmPoMuc* genes (see Figure 3 for the location of the amplicon) on DNA extracted from adult clones from both strains. Copy numbers were obtained by comparison of *SmPoMuc* target genes to a reference gene that was

confirmed to be single-copy per haploid genome (*Schistosoma mansoni* Src kinase TK3, (Kapp et al. 2004)). Our findings indicate that the number of *SmPoMuc* genes varies from 6 to 9 depending on individuals tested for both strains. These results are in agreement with the gene number (6) identified in the *S. mansoni* genome assembly database since the primers used for copy number quantification do not amplify truncated genes.

We then investigated the structure of *SmPoMuc* genes (Figure 3) based on sequences available in the genome assembly database. *SmPoMuc* genes are composed of 15 exons of an average size of 60 bp. Intron average size is approximately 550 bp and varies between *SmPoMuc* genes because of insertion/deletion events.

The striking feature of the 5' variable region that spans exons 1 to 2 is that exon 2 and its flanking introns occur as tandem repeats (Figure 3). These genomic repeats of approximately 1 kb are separated by microsatellites and we discuss below the detailed description and high level of similarity of these genomic repeats between all members of *SmPoMuc* multigene family. This conservation prevented the assembly of the trace files into this part of the *SmPoMuc* genes and explains their frequent incomplete assembly into contigs in the databases.

To investigate the different genes and/or alleles in our strains of interest and the relationships between individual members of this gene family, we performed an analysis of PCR-amplified, subcloned and aligned *SmPoMuc* genomic DNA sequences and constructed a cladogram with PAUP (Figure 4). The genomic sequences used for this analysis correspond to the 3' part of the genes and were obtained using universal primers amplifying *SmPoMuc* genes between the last exon 2 and exon 15 (see Figure 3 for primer positions) in both strains. Respectively 12 and 11 different sequences were obtained for C and IC strains, (GenBank accession numbers EU676572 to EU676594). The unrooted tree option was chosen since even in *S. japonicum*, the closest relative to *S. mansoni*, no sequences with reasonable homology to

*SmPoMuc* exist that would allow their use for rooting. Our analysis shows clearly that the *SmPoMuc* gene family can be divided into four paralogous sequence groups (gr.1-gr.4) that are closely related for both strains (Figure 4). Essentially, these 4 groups were formed by insertion/deletion (indel) events in the non-coding regions and subsequent gene duplications. Group 3 can be divided into 5 sub-groups (sub-gr.3.1a, -3.1b, -3.2, -3.3, -3.4). Sub-gr.3.1a and -3.1b differ only by sequences located in introns. Sub-gr.3.1 and -3.2 share the same structure as the other groups but their sequences are different. Sub-gr.3.3 and -3.4 have undergone large deletion events in intronic and intronic/exonic regions, respectively. Traceable indel events like solo-LTRs, whose formation involves recombination between the LTRs of retrotransposons (Vitte and Panaud 2005), are observed, for example the creation of C-r1/3.4a and C-r1/3.4b in sub-group 3.4 by insertion of a Saci-4 solo-LTR between exons 10 and 11 (Figure 4). We also identified short tandem repeats flanking some deleted sequences (e.g. between exons 3 and 4 in several *SmPoMuc* genes). Traceable indel events are annotated in Figure 4.

A single group of truncated genes was identified. In these genes of the 3.4 sub-group, exons 4, 5, 6 were deleted. Other truncated genes or pseudogenes were not included in the analysis since they were not amplified with the primers used. In the *S. mansoni* genome assembly (version 3.1, Sanger Institute [http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s\\_mansoni](http://www.sanger.ac.uk/cgi-bin/blast/submitblast/s_mansoni)), four truncated genes/pseudogenes could be identified by BLAST searches (contigs Smp\_contig049466, -010496, -045333 and -030128). However, the current assembly status of the *S. mansoni* genome does not allow for reliable statements concerning these loci because the succession of exons is interrupted by transposon insertion events and these transposons are very conserved and repeated frequently in the *S. mansoni* genome.

Three groups of paralogous *SmPoMuc* genes (gr.1, gr2 and sub-gr.3.1) correspond to the three groups of cDNA variants we found in a previous study (Roger et al. 2008a). These three groups include genes that are expressed in both strains (Figure 1B). The genes that belong to the fourth group and sub-gr.3.3 and 3.4 are probably pseudogenes because their transcripts were not detected either in our previous work (Roger et al. 2008a; Roger et al. 2008b) or in the present study. In addition, the gr.4 *SmPoMuc* genes are associated with a different exon 2 (repeat r3, Figure 1B) that was itself never detected in a transcript. The sequence of this particular exon 2 is given in Figure 1B to allow comparison with other cDNA repeats in *SmPoMuc* transcripts.

Finally, another interesting characteristic of these genes is that they all contain an intron-coded ribozyme between exon 9 and 10. We have aligned the different ribozymes obtained for the 23 genes sequenced in both strains with the sequence of the *S. mansoni* hammerhead ribozyme that is the most extensively studied (GenBank accession number : AF036742; (Ferbeyre, Smith and Cedergren 1998; Canny, Jucker and Pardi 2007; Przybilski and Hammann 2007)). The alignment (Figure 5) shows that all the genes possess the promoter elements (boxes A and B) that are essential for transcription by RNA polymerase III (Ferbeyre, Smith and Cedergren 1998). The aligned sequences correspond to natural ribozymes that display the canonical structure of schistosome hammerhead ribozymes consisting of three helices and a catalytic core. However, only the putative ribozymes of the *SmPoMuc* 2 group possess the G12 of the catalytic core that was shown to be essential for activity (Martick and Scott 2006).

### ***SmPoMuc* genes recombine frequently and evolve under selective pressure**

To analyze whether recombination events occur between *SmPoMuc* genes, we annotated the available sequences by a color code that uses a different color for sequence

fragments of less than 95% identity (Figure 4). By visual inspection we identified at least 14 recombination events between the 23 genes amplified by PCR in the 3' constant region amplified by PCR. Recombination break points are evenly distributed along the sequence. We noted that these recombination events can generate mosaic genes, the sequence of which could originate from the different members of the multi-gene family. For example, exons from the gr.4 *SmPoMuc* pseudogenes can be found in several of these mosaic genes belonging to gr.3 (Figure 4). We then investigated whether all *SmPoMuc* genes are under selective pressure and calculated ratios of synonymous to non-synonymous substitutions ( $K_S/K_N$ ) in 15, 71 and 56 subcloned RT-PCR products for the genes of groups 1, 2 and 3 respectively. This analysis was performed on cDNA sequences (Genbank accession numbers EU676503 to EU676571 and EU042599 to EU042636) in the 3'-terminal conserved part of *SmPoMucs*. As shown in Figure 6, the  $K_S/K_N$  ratio is  $>1$  in a large majority of the transcripts (93.1%, 86.8% and 90.8% for groups 1, 2 and 3 respectively), indicating that all genes are under selection. Likewise, Tajima's test for neutrality delivers significant negative D values indicating that a purifying selection acts within the three groups of *SmPoMuc* ( $D_1 = -2.305$ ,  $D_2 = -2.58$ ,  $D_3 = -2.77$  and  $p \text{ value} < 0.05$ ). Therefore, all *SmPoMuc* genes have evolved under selective pressure.

### ***SmPoMuc* tandem repeats have unique features**

As described above, the part of the genes encoding the repeat units in the protein are composed of repetitive units of approximately 1 kb including an exon of 27 bp (corresponding to exon 2) and separated by microsatellites (see Figure 3). These microsatellites are considered as imperfect because they are composed essentially of dinucleotide repeats (GA) whose succession is often interrupted. To analyze the diversity of these genomic repeats, we analyzed all trace records containing exon 2. Their size ranges from 1022 to 1037 bp differing



by the size of the microsatellite. The intronic part of the repeat units is highly conserved, displaying more than 94% identity and prevents assembly of the trace records corresponding to this part of the *SmPoMuc* genes. PCR amplification, cloning and random sequencing (31 clones for each strain, GenBank accession numbers: EU676595 to EU676656) of the DNA of our strains of interest reveals the same size window (about 1070 pb) and the same level of conservation (>93%).

Since assembly of trace records was impossible in this repetitive 5' region of *SmPoMuc* genes, we were unable to determine the number, size and nature of repeats associated with the different *SmPoMuc* genes by *in silico* analysis. To circumvent this problem, we analyzed *SmPoMuc*-containing BACs from the Sm1 library (Le Paslier et al. 2000). The BACs of interest were selected by Blast searches against the TIGR BAC ends database. Positive clones were recovered and digested with *EcoRV* that does not cut within the repeat units. Fragments were size separated by gel electrophoresis and hybridized with (i) a probe directed to a conserved intronic sequence upstream of exon 2, or (ii) probes that distinguish between r1 and r2 repeats. Results are shown in Figure 2. Each band corresponds to an intact stretch of repeat units and two types of restriction (data not shown) and hybridization patterns can be clearly distinguished. BACs 45D24, 47P6 and 51E8 span a genomic region in which at least 6 polymorphic *SmPoMuc* loci are situated (Figure 2, lanes 7-8-9). This result was confirmed by PCR amplification of BACs and sequencing of the PCR products: the number of sequences obtained corresponds to the number of bands observed for each BAC clone on the Southern blot (Table 2). This cluster contains at least 6 tandemly repeated *SmPoMuc* group 3 genes. In addition, all loci of this cluster exclusively contain r1 type repeats (Figure 2, lanes 13-14-15). The second group of BACs contains one or two genes of *SmPoMuc* group 2 (confirmed by PCR and sequencing (Table 2)). BACs 41B11 and 62F12 contain a single *SmPoMuc* locus or multiple monomorphic loci with r2 repeats only (Figure 2,

lanes 23-24). BAC 62J10 contains none of the repeat sequences (Figure 2, lanes 16-22). PCR confirmed the presence of a truncated gene without genomic repetitive units on this BAC. This truncated gene is also present in the two other BACs spanning this genomic region. It corresponds to a duplicated truncated form of the gene *SmPoMuc 2* in the contig Smp\_contig010496 of genome assembly version 3.1.

Since DNA from C and IC strains were digested with a restriction enzyme that does not cut within the repeat units (*EcoRV*), hybridization with a probe directed to the conserved intronic sequence upstream of exon 2 reveals the genomic repeat stretches of all genes present in the DNA. This Southern blot shows that the labeled fragment size never exceeds 20 kb (Figure 2, lanes 2 and 5) corresponding to a maximum of 20 repeat units per gene for both strains. These results are in agreement with the results obtained with the BACs in which the longest fragment corresponding to *SmPoMuc* repeat stretches again never exceeded 20 kb. In addition, most loci present in BACs 45D24, 47P6 and 51E8 contain less than 10 repeat units (Figure 2, lane 7-8-9). This contrasts strikingly with the transcripts in which up to 100 repeat units can be present (see below).

Since we have found combinations of r1 and r2 repeats in transcripts, we tested, using PCR and primers that are specific for these repeats, whether they could be amplified from two neighboring genomic repetitive units. In the case of the BAC clones, this analysis confirmed the results of the Southern Blot: TR stretches are composed of either r1 (Figure 7B, lanes 1-2-3) or r2 (Figure 7A, lanes 5 and 6) but never of both repeat units intermingled (Figure 7C and D, lanes 1-2-3-4-5-6). However, when the same PCR experiments were conducted on genomic DNA from our two strains of interest, intermingled repeats were detected in some *SmPoMuc* genes of both strains Figure 7C and D, lanes 7 and 8). This result is in agreement with cDNA sequencing showing that intermingled repeats are regularly detected in the IC strain but also once in the C strain (see Table S1, individual C-8). These latter results also

show that genes containing intermingled repeats are present in both strains but are seldom expressed in C strain individuals. In addition, we show that intermingled repeats are not detected in *SmPoMuc*-containing BACs from the Sm1 library prepared with DNA of a Puerto Rican strain of *S. mansoni*. Moreover, intermingled repeats are detected neither in contigs from the *S. mansoni* genome assembly nor in ESTs obtained from BH and PR isolates of *S. mansoni* (Verjovski-Almeida et al. 2003). Therefore, intermingled repeats r1 and r2 seem to be a unique feature of our model strains.

Concerning the different genes associated with the two BAC groups, the combination of PCR analysis and sequencing (Table 2), band lengths obtained by Southern blots (combinations of restriction digests with *EcoRV*, Figure 2 and *EcoRI*, data not shown) and *in silico* analysis of genome assemblies permit us to assign the different bands to their corresponding genes: the first group (BACs 45D24, 47P6 and 51E8) spans a genomic area with at least 6 genes in tandem. All these genes belong to group 3: one gene of the 3.1a sub-group containing 7 genomic tandem repeats; two genes of the 3.1b sub-group, one containing 15 repeats and the other 1 repeat; two genes of the 3.3 sub-group containing one or two repeats; and one gene of the 3.4 sub-group containing one repeat only. The second group of BACs (62J10, 41B11 and 62F12) spans a genomic area with at least two group 2 genes: one of them is truncated and does not possess the genomic tandem repeat region; the other contains approximately 20 repeats.

#### ***SmPoMuc* genes are organized in four locations on chromosome 3 and 4**

FISH on metaphase chromosomes of *S. mansoni* with BACs that are representative of each group identified by Southern blot (41B11 and 45D24) revealed the presence of four genomic *SmPoMuc* locations. Hybridization with BAC 41B11 gave strong signals near the centromere regions of chromosomes 2 and 3 and on the long arm of chromosome 4. Two

weaker signals were also detected on the short and on the long arm of chromosome 3 (Figure 8a). BAC 45D24 hybridized to the same regions on chromosomes 3 and 4, but gave no signal at the large heterochromatic pericentromeric and nucleolus organizer regions of chromosome 2 (Figure 8b). Consequently, the signal on chromosome 2 is specifically obtained only for 41B11 and is probably due to repetitive sequences in this BAC and not to the presence of *SmPoMuc* genes. FISH thus indicates the existence of at least four distinct locations of *SmPoMuc* genes in the genome of *S. mansoni*. Differences in signal intensity suggest that the loci near the centromere on chromosome 3 and on the long arm of chromosome 4 could contain more *SmPoMuc* genes than the others, which is in good agreement with Southern blotting results: BACs 45D24, 47P6 and 51E8, derive from a genomic area that contains at least 6 tandemly oriented group 3 *SmPoMuc* genes (Figure 4). The other group of BACs (41B11, 62J10, and 62F12) covers a genomic area containing at least two group 2 genes (Figure 4). *SmPoMuc* genes of groups 1 and 4 were not identified in any of these BACs. Nevertheless, these genes were identified in the *S. mansoni* genome assembly (contigs Smp\_contig019963 and -026239) as well as in our two strains of interest. These results suggest that they may be present in the two other locations identified by FISH.

***SmPoMuc* transcription patterns are highly polymorph, strain- and individual-specific: involvement of expression polymorphism, alternative splicing, aberrant splicing and exon repetition.**

In our previous studies of *SmPoMucs*, polymorphism was investigated at the transcript level. PCR amplification with consensus primers of cDNA pools (obtained after reverse transcription of RNA extracted from one thousand sporocysts) from both *S. mansoni* strains showed distinctive banding patterns after analysis in agarose gels (Roger et al. 2008a). To address the question of whether each individual sporocyst transcribes all strain-specific

*SmPoMuc* loci, or whether expression patterns are individual-specific, RNA was extracted from 11 single sporocysts (from each strain) and nested RT-PCR was performed on each individual. Banding patterns in agarose gels indicate clearly that each individual sporocyst expresses a characteristic subset of *SmPoMuc* genes (Figure 1C). We never detected the same pattern in different individuals, suggesting a high level of transcript polymorphism within the tested *S. mansoni* populations. PCR products of these individuals were subcloned and 20 clones of each individual were sequenced for both strains. The results are summarized in Table S1.

Our results revealed first an extensive expression polymorphism between individuals. Some of them express only genes belonging to one group (see Table S1, individuals C-2, -3, -5, -7, -8, -9, -10 and IC-1, -3, -4, -6, -11), others express genes belonging to two different groups. In this latter case, they express either genes from group 2 and sub-group 3.1 (C-1, -4 and IC-2, -5, -7, -8, -9, -10), or genes from groups 1 and 2 (C-6), or genes from group 1 and sub-group 3.1 (C-11). Of the 11 individuals, none from the IC strain express group 1 genes. This observation is not in agreement with data obtained using cDNA pools from one thousand individuals that showed that genes from this group are indeed expressed in the IC strain ((Roger et al. 2008a) and Figure 1B) and the result obtained on individual parasites is probably due to sampling.

Second, this analysis of cDNA reveals non-classical splicing events that occur frequently in the 3' constant region of *SmPoMuc* transcripts. The frequency of these events is much higher in individuals from the IC strain (10/11) compared to C strain sporocysts (4/11). Some of these events correspond to alternative splicing that leads to exon deletion(s) without modifications of the Open Reading Frame (ORF). These events are observed in transcripts corresponding to group 2 and sub-group 3.1 genes. Deletions involve either several exons (*SmPoMuc*2 and 3.1 variants of C-1 and C-2 and *SmPoMuc*2 variant of IC-5, Table S1), or

only one exon, such as exon 3, (*SmPoMuc2* variant of IC-2, Table S1), 10 (*SmPoMuc3.1* variants of IC-1, Table S1) or 8 (*SmPoMuc2* variant of C-4, IC-7, -9 and -10, Table S1). Other variants display aberrant splicing that changes the ORF and results in premature stop codons. This phenomenon is essentially observed in transcripts corresponding to genes of the second group and is much more frequent in variants cloned from IC strain individuals. Type 1 aberrant splicing described in Figure 9 is observed for *SmPoMuc2* transcripts of 8 out of 11 IC strain individuals, but only for 1 out of 11 individuals of the C strain. In addition, other types of aberrant splicing variants (types 2 to 6) were observed at low frequency. Data are summarized in Figure 9.

Finally, our individual-level analysis of expression patterns of *SmPoMuc* reveals several novel points concerning the 5' VNTR region, the most striking concerning transcript length. The number of tandem repeats (TR) in transcripts of the different *SmPoMuc* genes varies from 1 to 100. In contrast, our data obtained by Southern blotting on genomic and BAC DNA (Figure 2, lanes 2-5, lanes 7-8-9-11-12) indicate a maximum number of only 20 exons 2 in the *SmPoMuc* genes. The presence of more than 20-fold repetitions of exon 2 in over 50% of the *SmPoMuc2* transcript variants and the occurrence of variants that differ only by the number of repeats in some individuals (see Table S1, individual C-3 for example) shows that a trans-splicing mechanism occurs during *SmPoMuc* transcription.

### **Polymorphic *SmPoMuc*-specific repeat patterns and splicing variations provide a basis for polymorphic glycosylation**

The *pI* of *SmPoMuc* proteins predicted from the cDNA sequences varies between 4.3 and 5 and these values are in good agreement with those determined by IEF. In contrast, the calculated molecular weights of 30 to 80 kDa, depending on the variation in length of the repeat domain, do not correspond to the measured values of between 55 and 130 kDa (Figure

1A). In a previous study we showed that *SmPoMucs* are glycosylated and it is therefore probable that the observed molecular weight shift is due to a high degree of glycosylation in the TR (Roger et al. 2008a). Three different types of repeats were identified: r1, r1' and r2 (Figure 1B). Two of them (r1 and r1') are very similar, differing only by 1 amino acid residue and all contain S, T and P residues. Such repetitive structures with similar amino acid compositions were described in different O-glycosylated mucins, and we predicted O-glycosylation of T or S residues in these repeats (a typical feature of mucins) using the NetOGlyc 3.1 server (<http://www.cbs.dtu.dk/services/NetOGlyc/>).

We applied this prediction tool to the different variants obtained in the present study using the amino acid sequences deduced from these different variants. We considered that glycosylation occurs when prediction score is superior to 0.5 as suggested by Julenius *et al.* (personal communication and (Julenius et al. 2005)). The results are summarized in Table S2. They show that the predicted glycosylation status is dependent on the number and type of repeats and the type of arrangement (alternation of repeat types). These results suggest that TR polymorphism and length could be linked to glycosylation polymorphism associated with the different expressed *SmPoMuc* variants. In addition, alternative splicing and aberrant splicing events that delete portions of the C-terminal part of the protein have a marked influence on the glycosylation prediction, for instance in the case of variants IC2/2/25r2.1 and .2 expressed by the IC-2 individual (Table S2). Aberrant splicing in the latter produces a stop codon that shortens the deduced peptide sequence. The consequence for the glycosylation prediction status is radical since the first variant contains no predicted glycosylation while glycosylation is predicted at 22 sites in the second. Because variants from IC strain individuals are more frequently subject to this type of splicing event, the number of predicted glycosylated variants is much higher (59 glycosylated variants out of 84, each containing an

average glycosylation number of  $8.4 \pm 7.6$ ) than that of C individuals (10/41, each containing an average glycosylation number of  $8.44 \pm 8.7$ ).

To determine whether these glycosylation predictions had biological significance we chemically deglycosylated sporocyst extracts using trifluoromethanesulfonic acid (TFMSA) before western blotting with antibodies directed against *SmPoMuc* proteins. Molecular weight shifts related to the loss of carbohydrate chains after deglycosylation were more marked for the IC strain (Figure 10, lane 4) than for the C strain (Figure 10, lane 2), compared to the respective controls. This suggested that the glycosylation prediction obtained following NetOGlyc 3.1 analysis was correct and that *SmPoMuc*s from the incompatible IC strain are indeed more highly glycosylated than those from the compatible C strain. The total removal of carbohydrate moieties from *SmPoMuc* glycoproteins by TFMSA treatment was confirmed by the observation that no bands were detected in treated samples compared to controls on Alcian blue stained gels or following lectin blotting (not shown).



## Discussion

*SmPoMucs* are mucin-like molecules that we recently discovered by a proteomics approach aiming at identifying molecular determinants of compatibility polymorphism in the interaction between *S. mansoni* and its intermediate host *B. glabrata* (Roger et al. 2008b). The comparison of the proteomes of sporocysts (intramolluscan stage) of two *S. mansoni* strains, one compatible with a specific strain of *B. glabrata*, the other incompatible with the same mollusk strain, showed that the principal difference lies in this protein family with the characteristics of mucins. We showed that these proteins are glycosylated, expressed in the apical gland of *S. mansoni* miracidia and sporocysts and are present in their Excretion-Secretion products (Roger et al. 2008a). These molecules are highly polymorphic and we therefore called them *SmPoMucs* for *S. mansoni* polymorphic mucins. In the present study, we have extended the analysis of their polymorphism and show that each individual larva expresses a unique combination of *SmPoMucs* derived from a limited set of genes. This extraordinary level of polymorphism may be linked to (i) gene structure, organization and evolution, and/or (ii) different regulation processes occurring during gene expression. In this study we have elucidated the complex cascade of mechanisms that confer polymorphism to *SmPoMuc*.

*SmPoMucs* are coded by a multi-gene family that contains at least 6 to 9 members in our strains of interest. PCR with consensus primers amplifying the 3' part of the genes revealed a dozen sequences corresponding to different genes/alleles that can be divided into four paralogous sequence groups (gr.1-gr.4). The first three groups correspond to expressed genes for which we detected transcripts. In addition, their corresponding proteins were identified in a previous study (Roger et al. 2008a). Transcripts and proteins from the fourth group and certain subgroups of the third group were never found, suggesting that these genes are probably pseudogenes. We also identified additional pseudogenes corresponding to

truncated forms of *SmPoMuc* in the *S. mansoni* genome assembly database. These latter observations suggest that a large proportion of the genes belonging to this multigene family are non-functional, exactly as would be expected for multigene families that conform to the birth-and-death model of evolution (Nei and Rooney 2005).

We also found evidence for frequent recombination events between pseudogenes, especially from gr.4 and other members of the multigene family (see Figure 4). This suggests that these pseudogenes can provide an additional pool of genetic variability for the generation of new variants through recombination, gene conversion or exon shuffling. This type of variation-generating mechanism was observed for the *Trypanosoma brucei* variant membrane surface glycoprotein (Roth et al. 1989), *Anaplasma marginale* membrane surface proteins (Brayton et al. 2002) and MHC (Doxiadis et al. 2006). The numerous insertion/deletion events identified in *SmPoMucs*, the solo-LTR identified in some genes (gr.2 and sub-gr.3.4, Figure 4), the truncated genes interrupted by retrotransposition events (contigs identified in the genome assembly) and short tandem repeats flanking some deleted sequences (e.g. between exon 3 and 4 in several *SmPoMuc* genes, Figure 4) illustrate these frequent reshaping events occurring in *SmPoMuc* genes. These structural characteristics suggest that retrotransposons could play a central role as mediators of recombination between *SmPoMuc* genes.

FISH experiments revealed that *SmPoMuc* genes are distributed at four locations in the genome of *S. mansoni*. Two of these locations were analysed in detail using the corresponding BACs. Our analysis shows that *SmPoMuc* gr.2 and gr.3 genes are organized in clusters in two distinct genomic locations. The *SmPoMuc2* cluster is composed of at least two genes, one complete containing all exons (1 to 15) and approximately 20 repeats of exon 2, and a truncated gene with no tandem repeats. The *SmPoMuc3* cluster is composed of at least six tandemly organized genes containing 1 to 15 exon 2 repeats. It is noteworthy that some

individuals might possess fewer genes in these latter clusters as we have evidenced a gene copy number variation between individuals (6 to 9 copies of *SmPoMuc* genes per individual). Furthermore, *SmPoMuc* 2 and 3 clusters are associated with a specific exon 2, the first containing only r2 exons and the second only r1 exons, a result confirmed by PCR on BAC clones. Nevertheless, intermingled r1 and r2 repeats were found in transcripts of *SmPoMuc* gr.3 variants of C and IC strains. This suggests that ectopic recombination occurs to generate this genomic level of polymorphism. Different structural elements could explain these frequent recombination events. Genomic repeats (exon 2 and flanking intronic sequences) are conserved within and between all genes of the *SmPoMuc* family. The level of identity (>93%) shows that these genomic repeats do not evolve independently of each other. The molecular process that leads to homogenization of DNA sequences of a given repetitive family is called “concerted evolution” and occurs in ribosomal genes and in certain protein-coding multigene families, such as those encoding histones or ubiquitin (see (Liao 1999), for review). This phenomenon was also previously described within several genes (i) encoding proteins containing tandemly repeated domains and (ii) displaying genomic repeats like those of *SmPoMuc* genes. This is the case for the repetitive part of the single copy gene encoding the coccidioides spherule outer wall glycoprotein (SOWgp). This protein contributes to the virulence of *Coccidioides spp* both by functioning as an adhesion molecule and by modulating the host’s immune response (Hung et al. 2000; Hung et al. 2002). In this SOWgp gene, genomic repeats corresponding to repeated exons and associated introns are also nearly identical and evolve by concerted evolution. This type of evolutionary process can be driven by directional and/or stochastic processes. The two mechanisms that are the principal explanations of concerted evolution in nuclear DNA are gene conversion and unequal crossing over (Nei and Rooney 2005). For unequal crossing over, increases and decreases in repeat number lead to turnover among repeats and, in principle, stochastic fixation of a single

repeat type. According to this model, unequal crossing-over commonly occurs in central regions of the array where repeats can mispair, and unique sequences flanking the repetitive array inhibit exchanges in the edge repeats, i.e. the repeats located at the termini. The involvement of unequal crossing-over was shown for the gene encoding SOWgp (Johannesson et al. 2005) and support for this view comes from both the higher conservation of repeats in the center of the repetitive array, and the polymorphism in repeat number in SOWgp. The phenomenon of unequal crossing-over probably occurs in members of the *SmPoMuc* family because our Southern blot results indicate a difference in repeat number between individuals, differences that occur when unequal crossing-over takes place. Nevertheless, unequal crossing-over cannot explain homogenization of genomic repeats containing exon 2 between genes situated on different loci and chromosomes. In addition, we did not find a gradient of conservation between central and edge repeats. We therefore favour the alternative hypothesis of gene conversion (non-reciprocal transfer of information). The molecular mechanism of gene conversion in multigene families is not well understood, nevertheless several findings indicate that cis-acting sequences can influence this phenomenon. One example is provided by the two early chorionic-gene families, *ErA* and *ErB*, of the silk moth *Bombyx mori* (Hibner, Burke and Eickbush 1991) which are in close proximity on the same chromosome. The genes of the *ErA* family exhibit 96 % sequence identity, whereas those in the *ErB* have only 63 % sequence identity. Sequence analysis suggested that microsatellite-like simple repeats present in the *ErA* family, but not in the *ErB* family, account for the difference in homogenization, because simple sequence repeats can be the sites for initiation of gene conversion (Hibner, Burke and Eickbush 1991). Another well documented example is given by the microsatellite sequences in the human *RNU2* locus that were proposed to play a role in concerted evolution (Liao and Weiner 1995). These observations support the hypothesis that the microsatellites that separate the exon 2 genomic

repeats in the *SmPoMuc* genes are involved in the gene conversion mechanism leading to the concerted evolution of these repeats. Several hypotheses for the mechanism of gene conversion induced by microsatellites have been evoked. These microsatellites composed of purines are polypurine tracts that can adopt a triple helix conformation called H-DNA (see (Bacolla and Wells 2004) for review). Recently, a major role for this kind of conformation (non-B DNA conformation) in chromosomal rearrangements was proposed. Hotspots of rearrangements occurred invariably at nucleotides abutting or within motifs capable of adopting non-B conformations leading to single or double-strands breaks (Bacolla et al. 2004; Wang and Vasquez 2006). The DNA-break repair mechanism could involve invasive DNA replication leading to gene conversion as shown in yeast (Szostak et al. 1983; Paques and Haber 1999). These different mechanisms could occur in *SmPoMuc* genes and explain the similarities observed between the genomic repeats in all members of the multigene family. Nevertheless, this gene conversion phenomenon is restricted to the intronic sequence of these genomic repeats as we have shown that exons are different between clusters. We therefore conclude that conservation of exon differences is due to selection pressure. Interestingly, the same molecular architecture and the same type of evolutionary processes were described for the above-mentioned modular spider silk protein genes. In these genes, repeated structures in the proteins correspond to genomic tandem repeats composed of exons and introns. Once more, in this case genomic repeats are subject to concerted evolution and intron sequences are more homogenized than are the exons that evolve under purifying selection (Hayashi and Lewis 2000). In *SmPoMuc* genes, the combination of concerted evolution acting on all repeats and exon difference conservation between clusters (by purifying selection) allows combinatorial events that we observed at the genomic level (r1 and r2 exons in the same gene) and in cDNA (r1 and r2 repeats in the same variant) in both strains. The reason is probably

ectopic recombination and exon exchange between clusters. A similar phenomenon was described for homologous sequences in somatic plant cells (Puchta 1999).

In addition to the structure of the genes that favour generation of polymorphism on a genomic level, several processes generating further levels of polymorphism occur during the expression of these genes.

First, genes are transcribed in an individual-specific manner. Some individuals express several genes and/or alleles for one group of *SmPoMuc* and others for two groups that may be different between individuals. This observation raises the question of differential transcriptional regulation of the genes belonging to this multigene family. Future studies will address this question.

Second, we found evidence for various post-transcriptional regulation events. Exhaustive analysis of *SmPoMuc* cDNAs reveals numerous alternative splicing and aberrant splicing events in the coding region in the 234 residue C-terminal region of the precursor. Alternative splicing events do not change the ORF and lead to shorter proteins. Aberrant splicing also appears frequently and produces a non-sense codon immediately downstream of the splice sites, leading to truncated proteins. As a first approach to investigate the consequence of these events on gene products, we analyzed the predicted glycosylation of these truncated forms of *SmPoMuc* proteins. An *in silico* approach revealed that these events can have a marked influence on the predicted glycosylation status of *SmPoMuc* proteins (Table S2) since truncated *SmPoMucs* are predicted to be more highly glycosylated than are complete proteins. There are many more truncated forms in the IC strain than in the C strain. Our analysis of the global glycosylation status of *SmPoMucs* by chemical deglycosylation and subsequent western blotting strengthen these predictions of the glycosylation status since molecular weight shifts were larger for IC strain *SmPoMucs* compared to the C strain. These

results support the view that differential splicing events can influence the glycosylation status of *SmPoMucs*.

In the repeated region of the precursor, a difference in repeat number and repeat combination (r1 and r2 together in the same variant) between *SmPoMuc* variants is apparent. These variations also influence the predicted glycosylation status of the repeats (see Table S2). This level of polymorphism is probably generated by two mechanisms. The first acts at the genomic level and is related to (i) unequal crossing-over leading to the contraction/expansion of genomic repeat number or (ii) ectopic recombination leading to the combinatorial polymorphism found in some variants. But these events that act at the genomic level cannot alone explain the polymorphism observed, particularly the length polymorphism of the TR stretch. Indeed, a maximum of 20 genomic repeats is observed in DNA from both strains and on BACs. Nevertheless, 50% of the *SmPoMuc* cDNA variants possess more than 20 repetitive units (25 - 100 repeats). The fact that long repeats in the transcripts are always composed of the same TR type in a given variant suggests that these processes involve alleles (intergenic trans-splicing) or occur in cis (exon repetition). A trans-splicing mechanism was reported for *S. mansoni* (Rajkovic et al. 1990; Davis et al. 1995; Cheng et al. 2006) but its function in *S. mansoni* and the absence of a spliced leader sequence in *SmPoMuc* transcripts suggest that this mechanism is not involved in the phenomenon we observe. Exon repetition was first identified for the rat carnitine octanoyltransferase gene for which two copies of exon 2 were positioned adjacent to one another in some mRNAs while the genomic sequence contained only a single copy (Caudevilla et al. 1998). This mechanism has been further studied (Frantz et al. 1999; Rigatti et al. 2004) and the intervention of complementary intron sequences has been hypothesized (Dixon, Eperon and Samani 2007). Since we detected two complementary sequences of 13 and 12 nucleotides (Figure 3), respectively, in intronic sequences flanking exon 2 (data not shown), this hypothesis can be proposed for *SmPoMuc*

genes. In addition, the presence of hammerhead ribozymes in all *SmPoMuc* genes is intriguing. *S. mansoni* hammerhead ribozymes were extensively studied and shown to catalyze cleavage (De la Pena, Gago and Flores 2003) and ligation (Canny, Jucker and Pardi 2007) of transcripts *in vitro*. Their *in vivo* function is unknown but we can hypothesize that *S. mansoni* hammerhead ribozymes play a role in *SmPoMuc* transcript processing. This hypothesis is strengthened by a structural characteristic revealed by the analysis of *SmPoMuc* genes and transcripts corresponding to gr.2. Group 2 *SmPoMuc* genes are the only ones that possess the G12 of the catalytic core that was shown to be essential for activity (Martick and Scott 2006). The corresponding transcripts more frequently show exon repetition than others (see Table S1).

All the data we present here for the *SmPoMuc* multigene family show that gene structure, genomic organisation, recombination events and different regulation mechanisms during their expression allow the generation of a remarkably high degree of polymorphism from a limited set of genes (see Figure 11 for a schematic representation). This characteristic is unique for this model compared to the expression of polymorphic molecular variants in other parasites. Indeed, in all previously described cases the molecular variants are synthesized from a large set of genes belonging to a multigene family. An example is the case of *Trypanosoma cruzi* surface mucins (see (Buscaglia et al. 2006) for review) that contribute to parasite protection and to the establishment of a persistent infection. The multigene family encoding these proteins comprises 850 genes covering 1% of the parasite genome. Other relevant gene families include the *vsg* or the *var* family responsible for antigenic variation of *Trypanosoma brucei* (see (Taylor and Rudenko 2006; Navarro, Penate and Landeira 2007)) or *P. falciparum* (Ralph and Scherf 2005; Kyes, Kraemer and Smith 2007). *T. brucei* has > 1000 *vsg* genes and pseudogenes and the genome project of *P. falciparum* has identified 59 intact *var* genes.



In the present study, we provide evidence for several events acting at the genomic level and during expression leading to an extraordinary level of *SmPoMucs* polymorphism at inter and intra strain levels. Population studies have shown that success or failure of *B. glabrata/S. mansoni* infection depends on the matched or mismatched status of the host and parasite phenotypes (Theron and Coustau 2005). Interestingly, recent studies have demonstrated that the previously characterized family of *Biomphalaria* FREPs undergo processes of recombination diversification leading to the concomitant existence of a great diversity of FREPs within a single individual (Zhang et al. 2004). Although their function has not yet been totally clarified, evidence suggests that FREPs are capable of binding molecules of foreign origin such as *Echinostoma paraensei* excretory-secretory products (Adema et al. 1997) via their carbohydrate recognition domain. The exposure of host carbohydrate-binding molecules such as FREPs, to their *S. mansoni* carbohydrate ligands could determine the matched or mismatched status of a given *B. glabrata-S. mansoni* combination. In this context, *SmPoMucs* from *S. mansoni* are very promising candidates. These molecules are different between compatible and incompatible strains and are secreted and glycosylated (Roger et al. 2008a; Roger et al. 2008b). In addition, parasite-derived mucins or mucin-like molecules have been extensively described in different protozoan and helminth parasites. They have roles in host recognition, penetration, adhesion and invasion of host cells, immunoprotection, immunomodulation and in the avoidance of host immune processes (Hicks et al. 2000; Theodoropoulos et al. 2001). Among these molecules in parasites, the mucin-like proteins of *T. cruzi* share numerous characteristics with *SmPoMuc* (see (Buscaglia et al. 2006) for review), such as their structure, a high level of glycosylation and polymorphism. A major difference is that these molecules, encoded by the TcMUC I and II genes, like other surface components of parasitic protozoa, are typically anchored to the outer phospholipid layer of the plasma membrane by GPI. Nevertheless, secreted mucins have been characterized in

helminths such as the infective larvae of the parasitic nematode *Toxocara canis* (Loukas et al. 2000). These secreted mucins might simply create an immunological smoke-screen by trapping antigen-antibody complexes away from the parasite (Marin et al. 1992). An attractive hypothesis is that *SmPoMucs* create an immunological smoke-screen able to block pattern recognition receptors (like FREPs), thus avoiding recognition and the host immune reaction. Finally, we show here that *SmPoMucs* display the level of polymorphism we expect for key determinants of the compatibility polymorphism in play between *B. glabrata* and *S. mansoni*. We show that this high level of polymorphism has a consequence in the glycosylation status of *SmPoMuc*. All these data, taken together, strengthen our hypothesis that *SmPoMucs* are key determinants of *S. mansoni/B. glabrata* compatibility polymorphism.

### **Supplementary material**

Supplementary Table S1 and S2 are available at *Molecular Biology and Evolution* online.

### **Acknowledgements**

We thank Dr Guy-Franck Richard (Institut Pasteur, Paris) for his advice concerning gene conversion. We thank Diana Ballen, Jean-François Allienne, Anne Rognon and Bernard Dejean for technical assistance. We also thank the Sanger Institute (Hinxton, United Kingdom) and The Institute for Genome Research (Rockville, Md.) for making available the draft *S. mansoni* genome sequences.

### **Funding**

This work was supported by the ANR (grant 25390 Schistophepigen), CNRS, Inserm and UPVD and the Global COE Program (A06) of the MEXT, Japan. I.M.C. was supported

by the ECOS Nord (grant V06A01). This research was supported by a scholarship to E.R. from the French Ministry of Superior Education and Research (MESR).

**Accession numbers**

Nucleotide sequence data reported in this paper are available in the GenBank™ database under the accession numbers: EU676447 to EU676656.

## References

- Adema CM, Hertel LA, Miller RD and Loker ES. (1997) A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc Natl Acad Sci U S A.* **94**:8691-8696.
- Bacolla A, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeysinghe SS, O'Connell CD, Cooper DN and Wells RD. (2004) Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci U S A.* **101**:14162-14167.
- Bacolla A and Wells RD. (2004) Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem.* **279**:47411-47414.
- Brayton KA, Palmer GH, Lundgren A, Yi J and Barbet AF. (2002) Antigenic variation of *Anaplasma marginale* msp2 occurs by combinatorial gene conversion. *Mol Microbiol.* **43**:1151-1159.
- Buscaglia CA, Campo VA, Frasch AC and Di Noia JM. (2006) *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nat Rev Microbiol.* **4**:229-236.
- Cannon JP, Haire RN, Schnitker N, Mueller MG and Litman GW. (2004) Individual protochordates have unique immune-type receptor repertoires. *Curr Biol.* **14**:R465-466.
- Canny MD, Jucker FM and Pardi A. (2007) Efficient ligation of the *Schistosoma* hammerhead ribozyme. *Biochemistry.* **46**:3826-3834.
- Caudevilla C, Serra D, Miliar A, Codony C, Asins G, Bach M and Hegardt FG. (1998) Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc Natl Acad Sci U S A.* **95**:12185-12190.
- Cheng G, Cohen L, Ndegwa D and Davis RE. (2006) The flatworm spliced leader 3'-terminal AUG as a translation initiator methionine. *J Biol Chem.* **281**:733-743.
- Davis RE, Hardwick C, Tavernier P, Hodgson S and Singh H. (1995) RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, *Schistosoma mansoni*. *J Biol Chem.* **270**:21813-21819.
- De la Pena M, Gago S and Flores R. (2003) Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *Embo J.* **22**:5561-5570.
- Di Noia JM and Neuberger MS. (2007) Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem.* **76**:1-22.
- Dixon RJ, Eperon IC and Samani NJ. (2007) Complementary intron sequence motifs associated with human exon repetition: a role for intragenic, inter-transcript interactions in gene expression. *Bioinformatics.* **23**:150-155.
- Dong Y, Taylor HE and Dimopoulos G. (2006) AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol.* **4**:e229.

- Doxiadis GG, van der Wiel MK, Brok HP, de Groot NG, Otting N, t Hart BA, van Rood JJ and Bontrop RE. (2006) Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a New World primate species. *Proc Natl Acad Sci U S A*. **103**:5864-5868.
- Ebert D. (2008) Host-parasite coevolution: Insights from the Daphnia-parasite model system. *Curr Opin Microbiol*.
- Edge AS. (2003) Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function. *Biochem J*. **376**:339-350.
- Ferbeyre G, Smith JM and Cedergren R. (1998) Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol Cell Biol*. **18**:3880-3888.
- Finlay BB and McFadden G. (2006) Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell*. **124**:767-782.
- Frantz SA, Thiara AS, Lodwick D, Ng LL, Eperon IC and Samani NJ. (1999) Exon repetition in mRNA. *Proc Natl Acad Sci U S A*. **96**:5400-5405.
- Gryseels B, Polman K, Clerinx J and Kestens L. (2006) Human schistosomiasis. *Lancet*. **368**:1106-1118.
- Hayashi CY and Lewis RV. (2000) Molecular architecture and evolution of a modular spider silk protein gene. *Science*. **287**:1477-1479.
- Hertel KJ, Pardi A, Uhlenbeck OC et al. (1992) Numbering system for the hammerhead. *Nucleic Acids Res*. **20**:3252.
- Hibner BL, Burke WD and Eickbush TH. (1991) Sequence identity in an early chorion multigene family is the result of localized gene conversion. *Genetics*. **128**:595-606.
- Hicks SJ, Theodoropoulos G, Carrington SD and Corfield AP. (2000) The role of mucins in host-parasite interactions. Part I- protozoan parasites. *Parasitol Today*. **16**:476-481.
- Hirai H and LoVerde PT. (1995) FISH techniques for constructing physical maps on schistosome chromosomes. *Parasitol Today*. **11**:310-314.
- Hung CY, Ampel NM, Christian L, Seshan KR and Cole GT. (2000) A major cell surface antigen of *Coccidioides immitis* which elicits both humoral and cellular immune responses. *Infect Immun*. **68**:584-593.
- Hung CY, Yu JJ, Seshan KR, Reichard U and Cole GT. (2002) A parasitic phase-specific adhesin of *Coccidioides immitis* contributes to the virulence of this respiratory Fungal pathogen. *Infect Immun*. **70**:3443-3456.
- Johannesson H, Townsend JP, Hung CY, Cole GT and Taylor JW. (2005) Concerted evolution in the repeats of an immunomodulating cell surface protein, SOWgp, of the human pathogenic fungi *Coccidioides immitis* and *C. posadasii*. *Genetics*. **171**:109-117.

- Julenius K, Molgaard A, Gupta R and Brunak S. (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology*. **15**:153-164.
- Kapp K, Knobloch J, Schussler P, Sroka S, Lammers R, Kunz W and Grevelding CG. (2004) The *Schistosoma mansoni* Src kinase TK3 is expressed in the gonads and likely involved in cytoskeletal organization. *Mol Biochem Parasitol*. **138**:171-182.
- Kyes SA, Kraemer SM and Smith JD. (2007) Antigenic variation in *Plasmodium falciparum*: gene organization and regulation of the var multigene family. *Eukaryot Cell*. **6**:1511-1520.
- Le Paslier MC, Pierce RJ, Merlin F, Hirai H, Wu W, Williams DL, Johnston D, LoVerde PT and Le Paslier D. (2000) Construction and characterization of a *Schistosoma mansoni* bacterial artificial chromosome library. *Genomics*. **65**:87-94.
- Liao D. (1999) Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet*. **64**:24-30.
- Liao D and Weiner AM. (1995) Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the RNU2 locus) does not prevent rapid diversification of the (CT)<sub>n</sub>(GA)<sub>n</sub> microsatellite embedded within the U2 repeat unit. *Genomics*. **30**:583-593.
- Linzmeier RM and Ganz T. (2005) Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23. *Genomics*. **86**:423-430.
- Linzmeier RM and Ganz T. (2006) Copy number polymorphisms are not a common feature of innate immune genes. *Genomics*. **88**:122-126.
- Lopez-Rubio JJ, Riviere L and Scherf A. (2007) Shared epigenetic mechanisms control virulence factors in protozoan parasites. *Curr Opin Microbiol*. **10**:560-568.
- Loukas A, Hintz M, Linder D, Mullin NP, Parkinson J, Tetteh KK and Maizels RM. (2000) A family of secreted mucins from the parasitic nematode *Toxocara canis* bears diverse mucin domains but shares similar flanking six-cysteine repeat motifs. *J Biol Chem*. **275**:39600-39607.
- Marin MS, Prieto M, Martin JM, Casais R, Boga JA and Parra F. (1992) Identification and expression of a *Fasciola hepatica* gene encoding a gut antigen protein bearing repetitive sequences. *Mol Biochem Parasitol*. **55**:155-165.
- Martick M and Scott WG. (2006) Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*. **126**:309-320.
- Medzhitov R and Janeway CA, Jr. (1997) Innate immunity: the virtues of a nonclonal system of recognition. *Cell*. **91**:295-298.
- Mitta G, Galinier R, Tisseyre P, Allienne JF, Girerd-Chambaz Y, Guillou F, Bouchut A and Coustau C. (2005) Gene discovery and expression analysis of immune-relevant genes from *Biomphalaria glabrata* hemocytes. *Dev Comp Immunol*. **29**:393-407.

- Mitta G, Vandenbulcke F, Hubert F and Roch P. (1999) Mussel defensins are synthesised and processed in granulocytes then released into the plasma after bacterial challenge. *J Cell Sci.* **112** ( Pt 23):4233-4242.
- Navarro M, Penate X and Landeira D. (2007) Nuclear architecture underlying gene expression in *Trypanosoma brucei*. *Trends Microbiol.* **15**:263-270.
- Nei M and Gojobori T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* **3**:418-426.
- Nei M and Rooney AP. (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet.* **39**:121-152.
- Pancer Z. (2000) Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. *Proc Natl Acad Sci U S A.* **97**:13156-13161.
- Paques F and Haber JE. (1999) Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev.* **63**:349-404.
- Patton WF. (2002) Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci.* **771**:3-31.
- Przybilski R and Hammann C. (2007) Idiosyncratic cleavage and ligation activity of individual hammerhead ribozymes and core sequence variants thereof. *Biol Chem.* **388**:737-741.
- Puchta H. (1999) Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. *Genetics.* **152**:1173-1181.
- Rajkovic A, Davis RE, Simonsen JN and Rottman FM. (1990) A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc Natl Acad Sci U S A.* **87**:8879-8883.
- Ralph SA and Scherf A. (2005) The epigenetic control of antigenic variation in *Plasmodium falciparum*. *Curr Opin Microbiol.* **8**:434-440.
- Reynaud CA, Bertocci B, Dahan A and Weill JC. (1994) Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. *Adv Immunol.* **57**:353-378.
- Rigatti R, Jia JH, Samani NJ and Eperon IC. (2004) Exon repetition: a major pathway for processing mRNA of some genes is allele-specific. *Nucleic Acids Res.* **32**:441-446.
- Roger E, Gourbal B, Grunau C, Pierce RJ, Galinier R and Mitta G. (2008a) Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. *Mol Biochem Parasitol.* **157**:217-227.
- Roger E, Mitta G, Mone Y, Bouchut A, Rognon A, Grunau C, Boissier J, Theron A and Gourbal BE. (2008b) Molecular determinants of compatibility polymorphism in the

- Biomphalaria glabrata*/*Schistosoma mansoni* model: New candidates identified by a global comparative proteomics approach. *Mol Biochem Parasitol.* **157**:205-216.
- Roth C, Bringaud F, Layden RE, Baltz T and Eisen H. (1989) Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. *Proc Natl Acad Sci U S A.* **86**:9375-9379.
- Suyama M, Torrents D and Bork P. (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**:W609-612.
- Szostak JW, Orr-Weaver TL, Rothstein RJ and Stahl FW. (1983) The double-strand-break repair model for recombination. *Cell.* **33**:25-35.
- Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* **123**:585-595.
- Taylor JE and Rudenko G. (2006) Switching trypanosome coats: what's in the wardrobe? *Trends Genet.* **22**:614-620.
- Theodoropoulos G, Hicks SJ, Corfield AP, Miller BG and Carrington SD. (2001) The role of mucins in host-parasite interactions: Part II - helminth parasites. *Trends Parasitol.* **17**:130-135.
- Theron A and Coustau C. (2005) Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *J Helminthol.* **79**:187-191.
- Theron A, Pages JR and Rognon A. (1997) *Schistosoma mansoni*: distribution patterns of miracidia among *Biomphalaria glabrata* snail as related to host susceptibility and sporocyst regulatory processes. *Exp Parasitol.* **85**:1-9.
- Tonegawa S. (1983) Somatic generation of antibody diversity. *Nature.* **302**:575-581.
- Van Valen L. (1974) Molecular evolution as predicted by natural selection. *J Mol Evol.* **3**:89-101.
- Verjovski-Almeida S, DeMarco R, Martins EA et al. (2003) Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nat Genet.* **35**:148-157.
- Vitte C and Panaud O. (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet Genome Res.* **110**:91-107.
- Wang G and Vasquez KM. (2006) Non-B DNA structure-induced genetic instability. *Mutat Res.* **598**:103-119.
- Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, Rebel VI and Schmucker D. (2005) Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science.* **309**:1874-1878.
- Zhang SM, Adema CM, Kepler TB and Loker ES. (2004) Diversification of Ig superfamily genes in an invertebrate. *Science.* **305**:251-254.



**Table 1: Primers used for sequencing, PCR and quantitative-PCR**

<b>Primers</b>		
<b>name</b>	<b>sequence</b>	<b>cycling conditions</b> annealing/extension/cycle
<b>For sequencing</b>		
Intron2/3F2	TTCTGTGTTATATACAACGTG	48°C / 4min / 40
Exon3F	TCCAGAACATTTGAAAACGAG	
Intron3/4R	CACATGCATAGCTAATGTGGTAATG	
Intron3/4F	AAATCGTGTGTTTATGGAATTGACG	
Exon4F	TATCTCTTGAACCATATACACGCGC	
Exon4R	GCGCGTGTATATGGTTCAAGAGATA	
Exon5F	TATTTCTTCTAGAATGTCTGAG	
Exon5R	TAGATAATGTACTGCCCACTTTGTG	
Intron5/6F	ATATGTGCGTCTGCTTTTAACTACG	
Intron6/7F	GCTGTCTCTCGCTAACAATACGACG	
Intron6/7R	ACATTTTCGTCGTATTGTTAGCGAG	
Intron7/8F1	CAGCTTCACATAAATGGAAACAC	
Intron7/8F2	AGTGGTTTACGAAAGTGAGGC	
Intron7/8R	TAGTAACATTGGTCGTTCTGTG	
Intron8/9F	AATGAAATAGTGAAAGAATGTTTCG	
Intron8/9R	CTTTCACTATTTCAATCAACAACG	
Intron9/10F	CATCGCGTTATTCAGTTAGCC	
Intron9/10R	TAAAGGTGGAATATGCCAAACTCAC	
Exon10F	TGAAGCTCAACTCAGTAAGCTGAAC	
Exon10R	AACTCATTATTTTGAATGTTTCAGC	
Exon11R	CTTGTATCGCCTTCGATTCCAATTC	
Exon11/12F	GACAGATTTCGCTTAGTGATGAAG	
Intron11/12R1	CTTCATCACTAAGCGAATCTGTC	
Intron11/12R2	GTTGCCTGAATTCACCATCTC	
Exon14F	TTCTTAGCACTACCCAAAGATGAAC	
Exon14R	TATTTGTTTCATCTTTGGGTAGTGC	
Intron14/15R	GTATAATTCCTAAATATCGC	
Exon15R	TGACACAGAAAACCTGTTAACGATCC	

<b>For PCR amplification</b>		
Exon1F12	GGAAGAATGAACAAGAAAATTATTCTC	65°C / 3min / 40
Exon15R	TGACACAGAAAACCTGTTAACGATCC	
NestedExon1F	TATNTTGCCTGATGATAAG	46°C / 3min / 40
NestedExon15R	ATCATAAACAACACTGAGG	
Intron2/3F1	CACTTGTTTCATAAACACGTGTCTTC	59,5°C / 10min / 40
Exon15R	TGACACAGAAAACCTGTTAACGATCC	
Intron2/3F1	CACTTGTTTCATAAACACGTGTCTTC	60°C / 45s / 30
Exon4R	GCGCGTGTATATGGTTCAAGAGATA	
Intron3/4F	AAATCGTGTGTTTATGGAATTGACG	60°C / 50s / 30
Intron3/4(gr.2)R	ATTCAAATCAGTGATTGGTGTTCAC	
Intron3/4(gr.3.4-	CATGAAAATGGGTTATTTGCTAGTG	

5)R		
Intron3/4(gr.3.4-5)F	CACTAGCAAATAACCCATTTTCATG	60°C / 4min / 30
Intron3/4F	AAATCGTGTGTTTATGGAATTGACG	
Intron9/10R	TAAAGGTGGAATATGCCAAACTCAC	
Intron5/6F	ATATGTGCGTCTGCTTTTAACTACG	60°C / 5min / 30
Exon11R	CTTGTATCGCCTTCGATTCCAATTC	
Intron5/6F	ATATGTGCGTCTGCTTTTAACTACG	60°C / 50s / 30
Intron6/7R	ACATTTTCGTCGTATTGTTAGCGAG	
Exon10F	TGAAGCTCAACTCAGTAAGCTGAAC	60°C / 25s / 30
Exon11R	CTTGTATCGCCTTCGATTCCAATTC	
Exon3F	TCCAGAACATTTGAAAACGAG	58°C / 20s / 30
Intron3/4R	CACATGCATAGCTAATGTGGTAATG	
<b>For intermingled repeats amplification</b>		
r1.F2	GCTCTCACATTTTCAGATGACTAT	60°C / 1min / 30
r1.R2	AACTCACCTGTTGGTTTCGCTC	
r2.F2	TCTCACATTTTCAGGTGACCTC	
r2.R2	AACTCACCTGTGGGTTTGTCTG	
<b>For real time quantitative PCR</b>		
Exon7F2	TATACGGAACAGACATGAGC	
Exon7R	ACATTGGTCGTTTCGTG	
Src.F1	TACGCTACCAACCCTGT	
Src.R1	CAAACCTGCCCTTCTGT	

**Table 2: *SmPoMuc* genes contained in the different BAC clones.**

<b>45D24</b>		<b>47P6</b>		<b>51E8</b>	
<b>TR</b>	<b><i>SmPoMuc</i></b>	<b>TR</b>	<b><i>SmPoMuc</i></b>	<b>TR</b>	<b><i>SmPoMuc</i></b>
7 or 15 r1	<b>gr. 3.1b</b>	7 or 15 r1	<b>gr. 3.1b</b>	7 or 15 r1	<b>gr. 3.1b</b>
7 or 15 r1	<b>gr. 3.1a</b>	7 or 15 r1	<b>gr. 3.1a</b>	7 or 15 r1	<b>gr. 3.1a</b>
1 or 2 r1	<b>gr. 3.3</b>	1 r1	<b>gr.3.4</b>	1 or 2 r1'	<b>gr. 3.3</b>
1 r1'	<b>gr. 3.1b</b>			1 or 2 r1'	<b>gr. 3.3</b>
1 or 2 r1'	<b>gr. 3.3</b>			1 r1	<b>gr.3.4</b>
1 r1	<b>gr.3.4</b>				

<b>62J10</b>		<b>41B11</b>		<b>62F12</b>	
<b>TR</b>	<b><i>SmPoMuc</i></b>	<b>TR</b>	<b><i>SmPoMuc</i></b>	<b>TR</b>	<b><i>SmPoMuc</i></b>
none	<b>Trunc gr.2</b>	≈ 20r2	<b>gr. 2</b>	≈ 20r2	<b>gr. 2</b>
		none	<b>Trunc gr.2</b>	none	<b>Trunc gr.2</b>

**Figure 1**

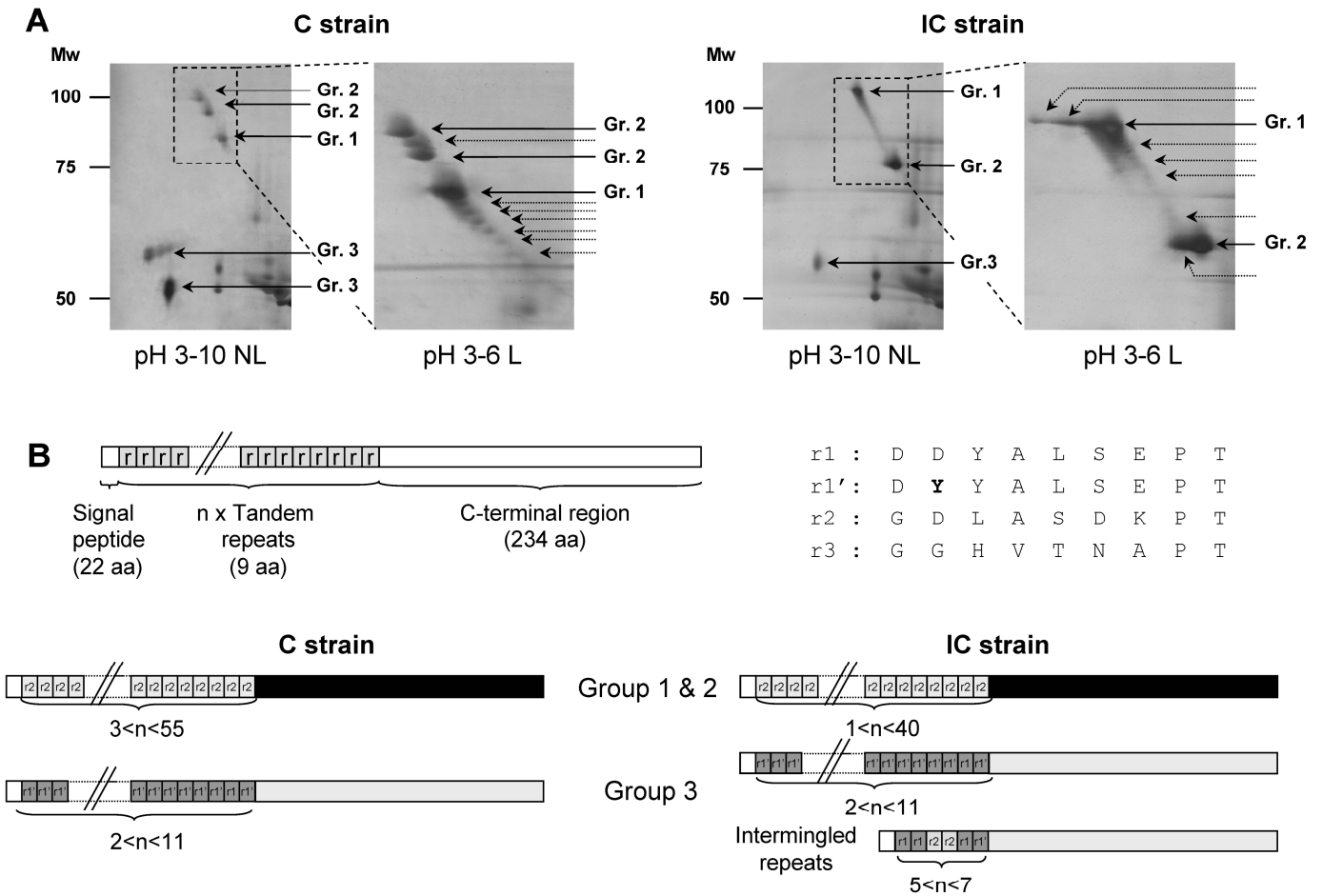


Figure 1

C

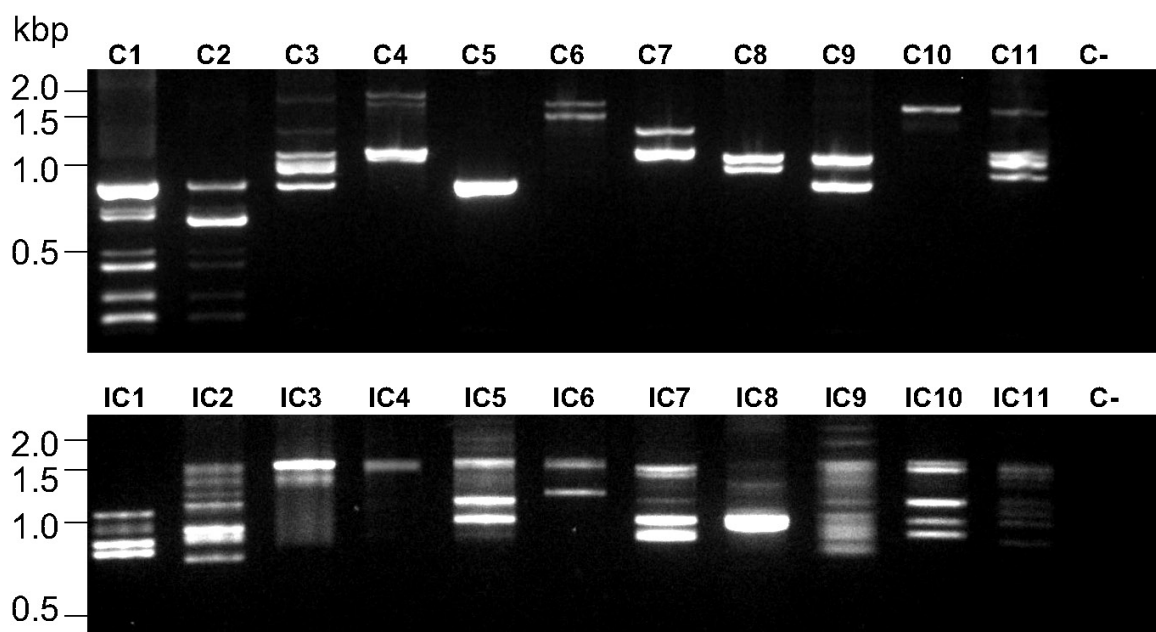


Figure 2

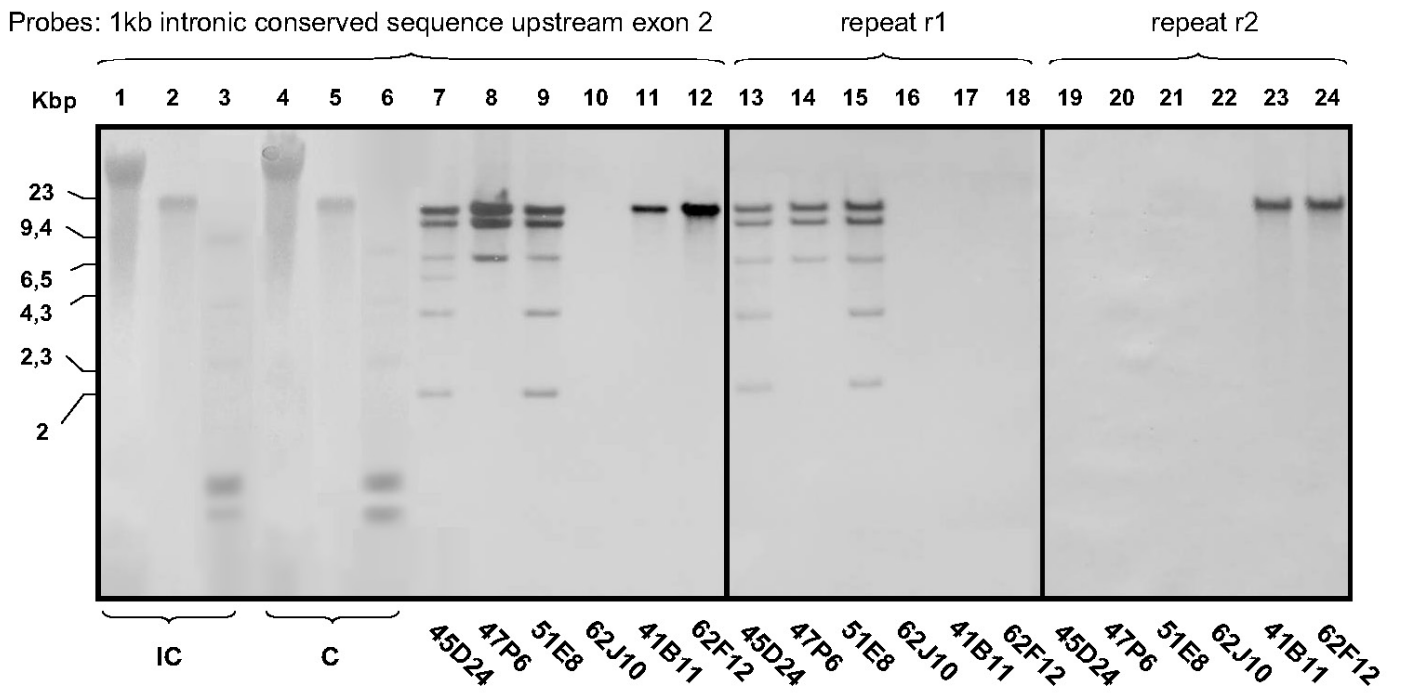
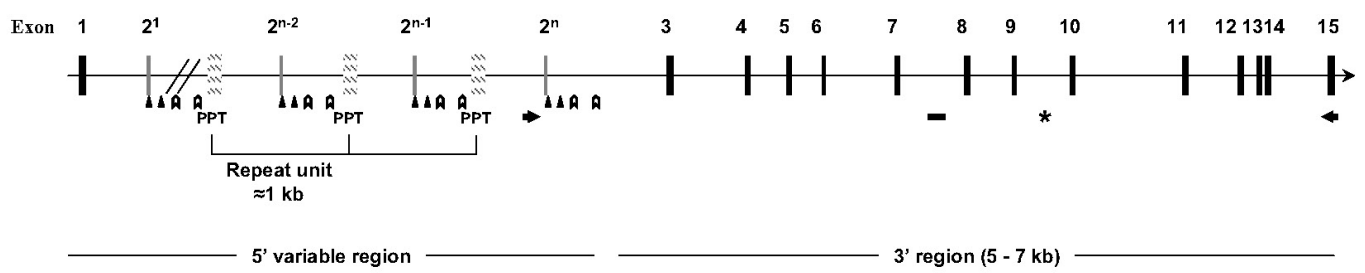


Figure 3



**Figure 4**

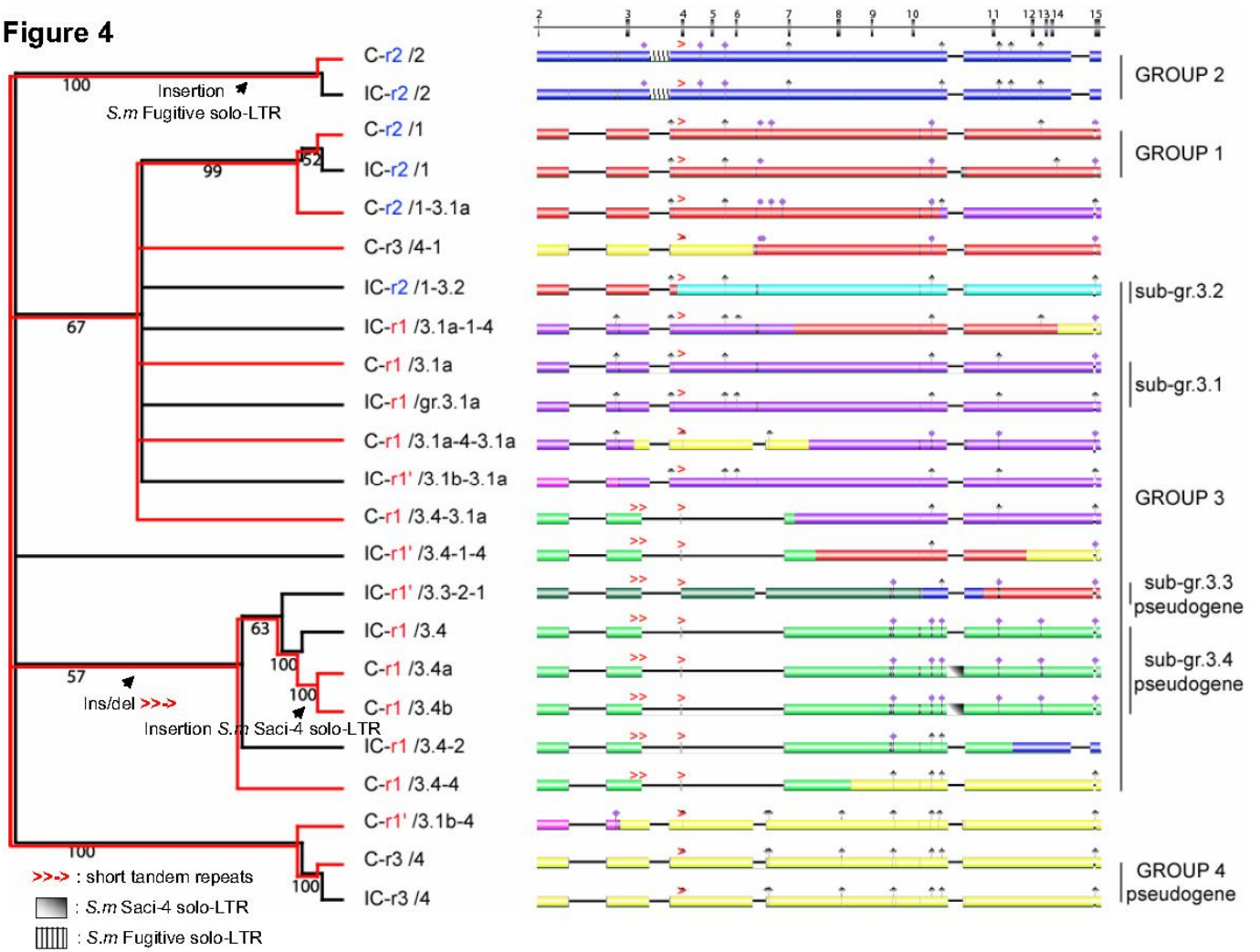
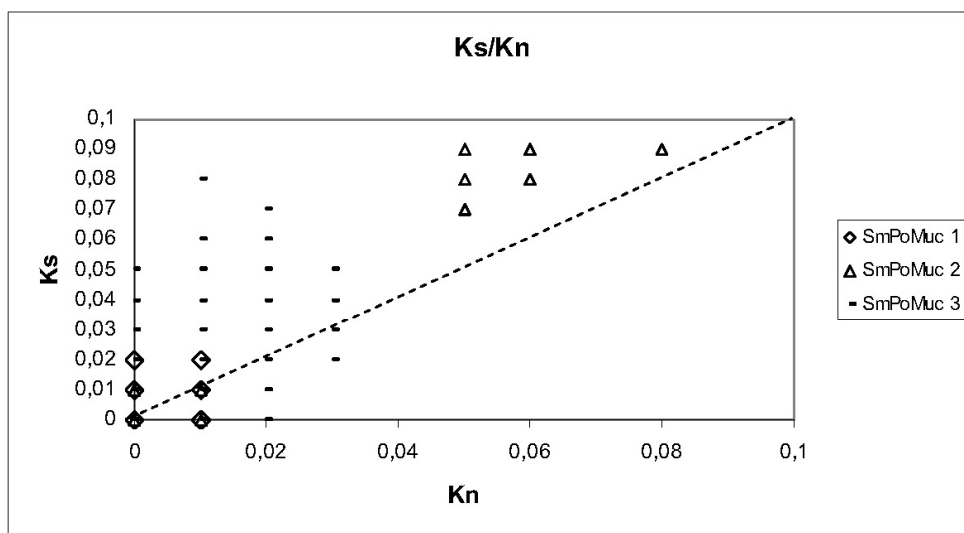






Figure 6



**Figure 7**

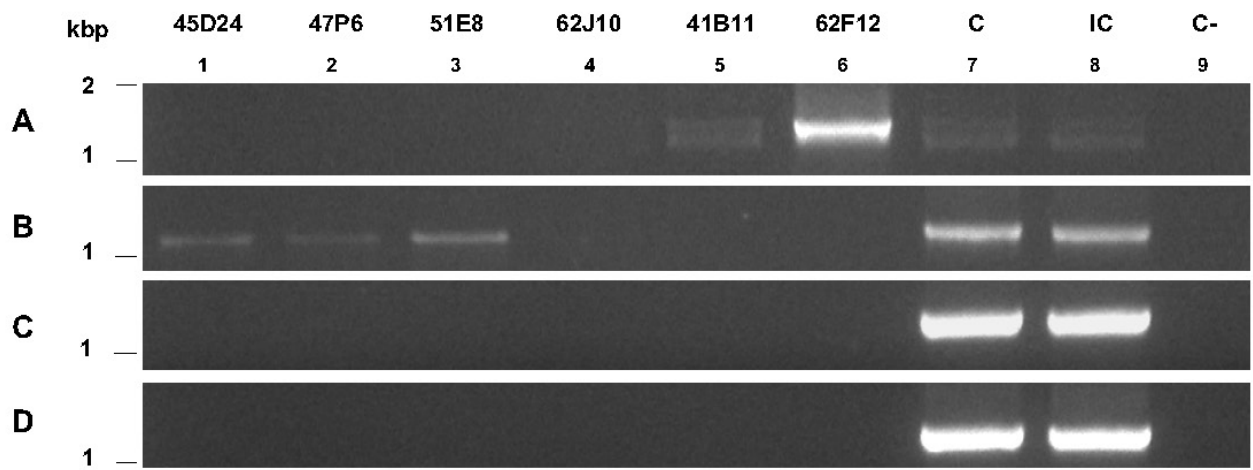


Figure 8

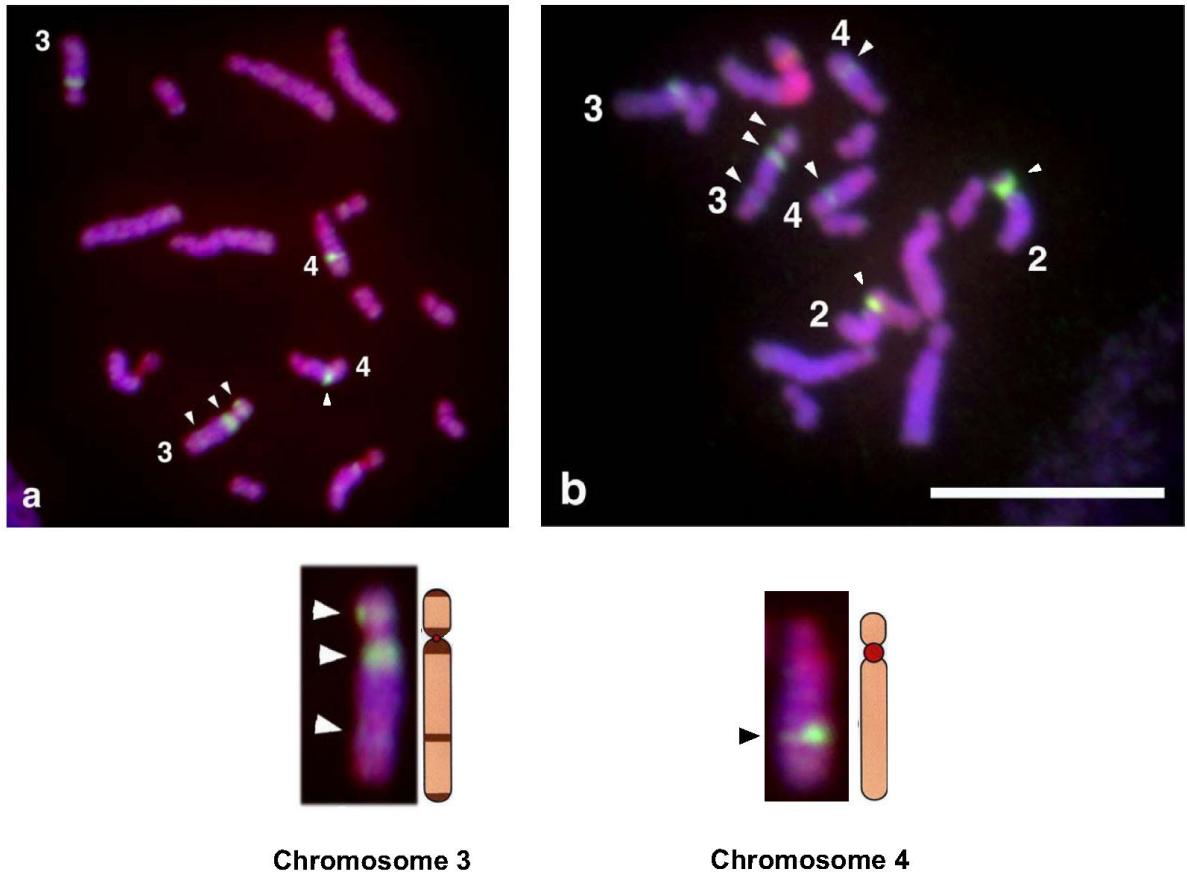
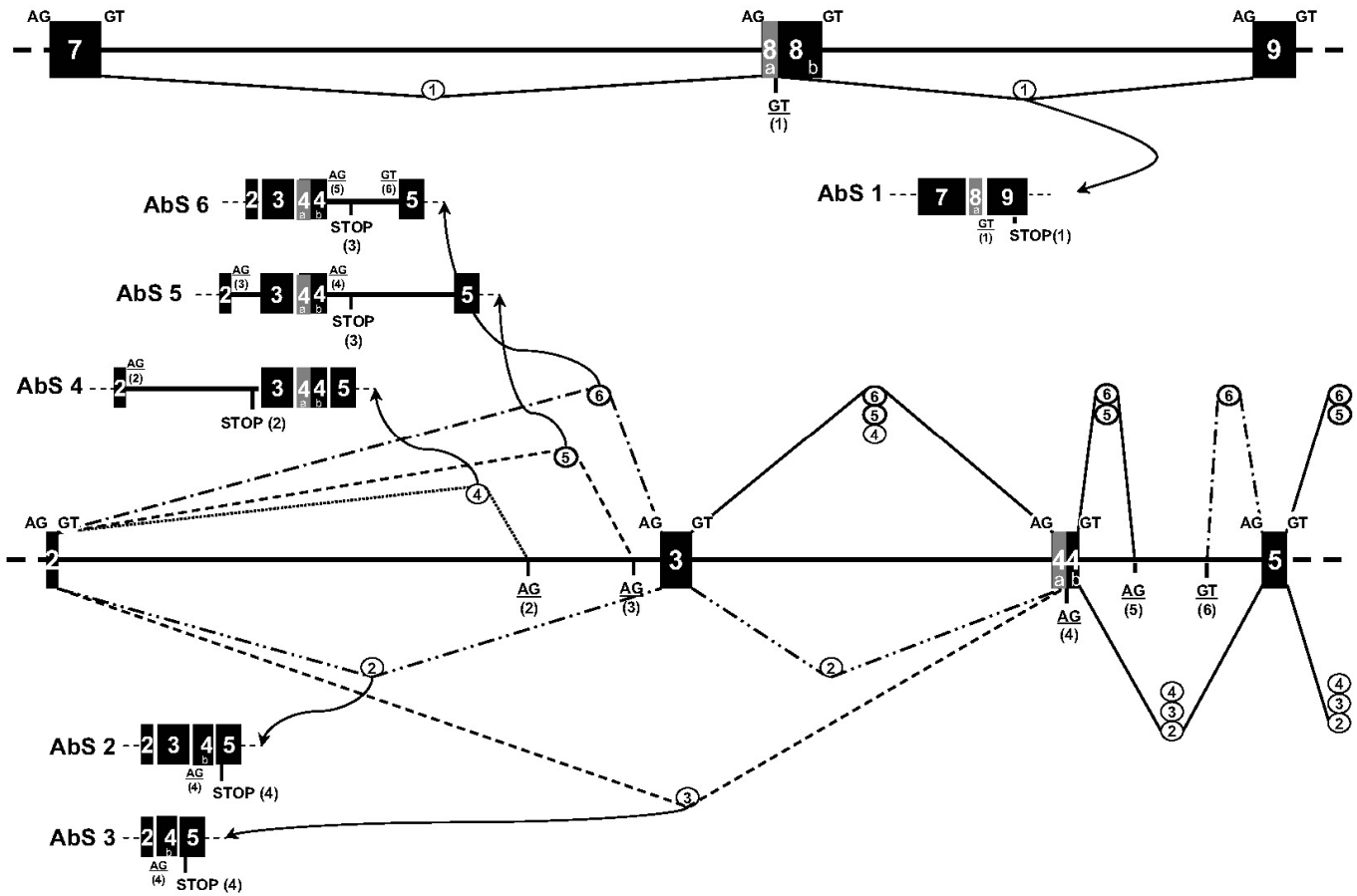


Figure 9



**Figure 10**

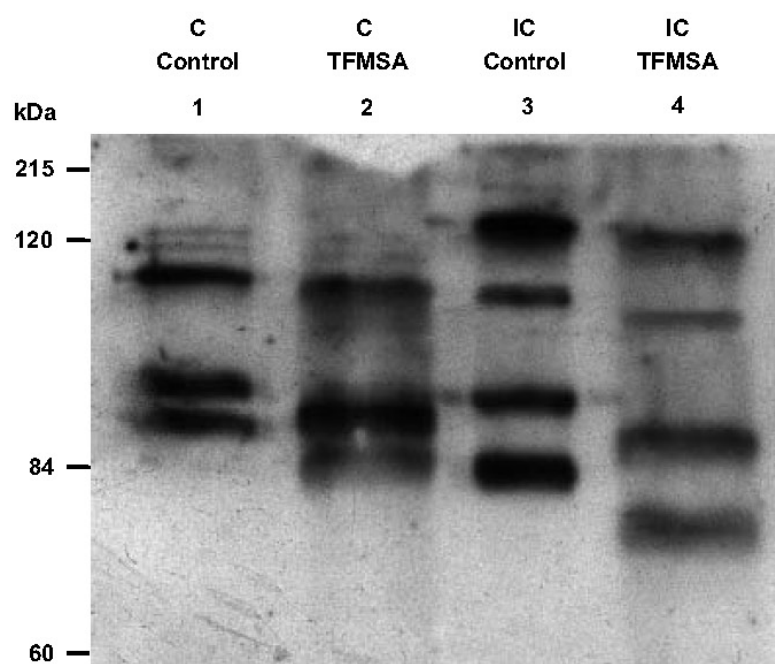
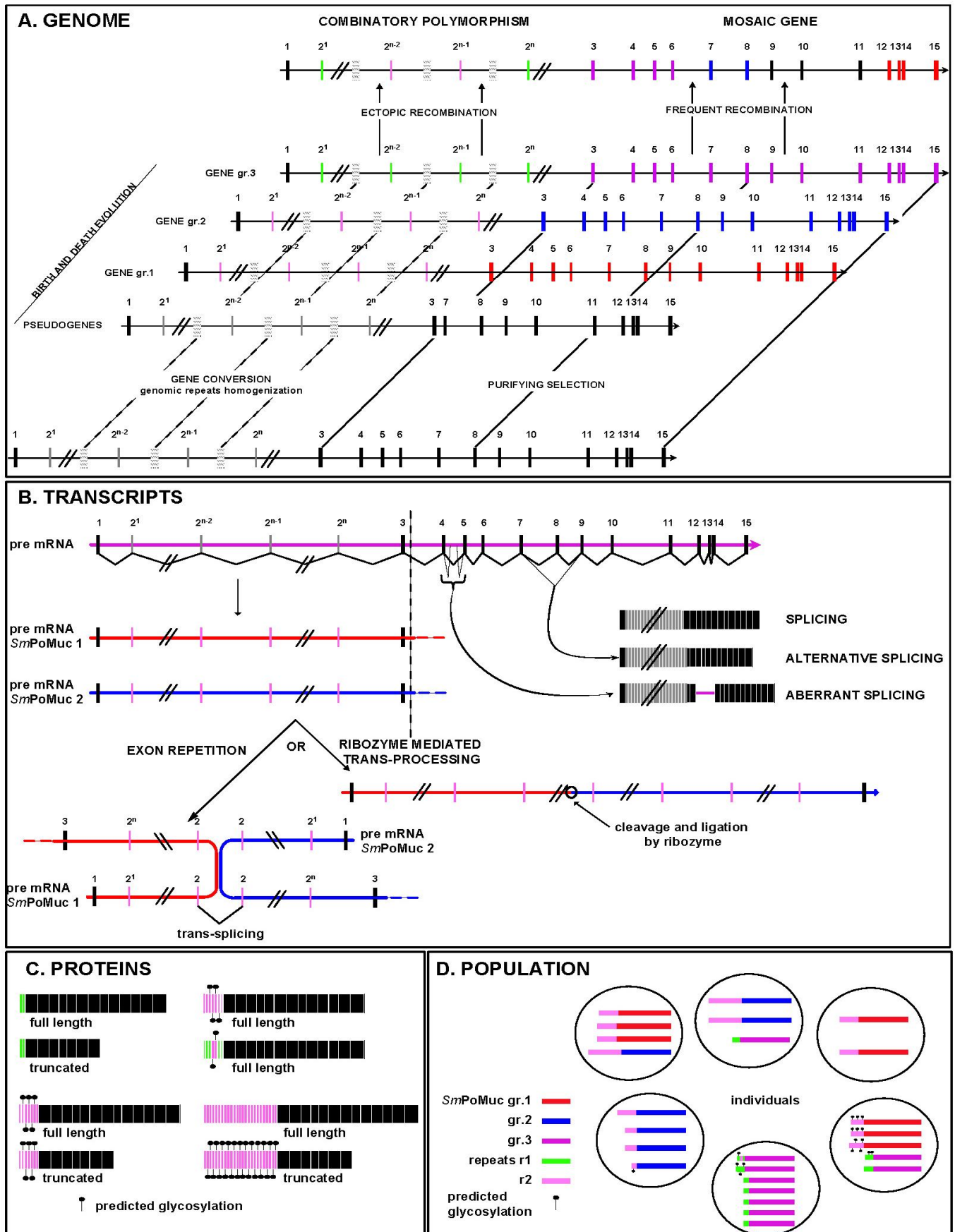


FIGURE 11

**SMPOMUC POLYMORPHISM IS CONTROLLED AT DIFFERENT LEVELS**



**TABLE S1: cDNA variants obtained for 11 individuals from both Compatible and Incompatible strains**

Compatible strain																
ind <sup>a</sup>	SmPoMuc <sup>b</sup> group	TR <sup>c</sup>	exon <sup>d</sup>													splicing <sup>e</sup>
			3	4	5	6	7	8	9	10	11	12	13	14	15	
C - 1	3.1	2r1	x	x	x		x	x	x			x	x	x	x	AS
		ref-3.1	2r1	x	x	x	x	x	x	x	x	x	x	x	x	
		r1										x	x	x	x	AS
		r1										x	x		x	AS
		r1'										x	x	x	x	AS
		r1'										x	x	x	x	AS
	2	7r2										x	x	x	x	AS
		7r2 <sup>o-2-4</sup>										x	x	x	x	AS
C - 2	3.2	r1	x				x	x	x	x	x	x	x	x	x	AS
C - 3	2	≈ 45r2	x	x	x	x	x	// -	STOP	x	x	x	x	x	x	AbS (1)
		≈ 40r2	x	x	x	x	x	// -	STOP	x	x	x	x	x	x	AbS (1)
		≈ 25r2	x	x	x	x	x	// -	STOP	x	x	x	x	x	x	AbS (1)
		11r2 <sup>o-2-4</sup>	x	x	x	x	x	x	x	x	x	x	x	x	x	
C - 4	3.1	9r1'-r1-r1'	x	x	x	x	x	x	x	x	x	x	x	x	x	
	2	≈ 100r2	x	x	x	x	x		x	x	x	x	x	x	x	AS
		≈ 100r2 <sup>o-3</sup>	x	x	x	x	x		x	x	x	x	x	x	AS	
C - 5	3.1	2r1'	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1'	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1'	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1' <sup>o-2</sup>	x	x	x	x	x	x	x	x	x	x	x	x	x	
C - 6	1	≈ 40r2	x	x	x	x	x	x	x	x	x	x	x	x	x	
		≈ 40r2	x	x	x	x	x	x	x	x	x	x	x	x	x	
		≈ 40r2	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2	≈ 80r2	x	x	x	x	x	x	x	x	x	x	x	x	
C - 7	2	ref-2	≈ 25r2	x	x	x	x	x	x	x	x	x	x	x	x	
			≈ 25r2	x	x	x	x	x	x	x	x	x	x	x	x	
			≈ 25r2	x	x	x	x	x	x	x	x	x	x	x	x	
			≈ 25r2	x	x	x	x	x	x	x	x	x	x	x	x	
			≈ 25r2	x	x	x	x	x	x	x	x	x	x	x	x	
			≈ 30r2	x	x	x	x	x	x	x	x	x	x	x	x	x
C - 8	3.1	2r1-2r2 <sup>o-2</sup> -r1-r1'	x	x	x	x	x	x	x	x	x	x	x	x		
C - 9	3.1	2r1'	x	x	x	x	x	x	x	x	x	x	x	x		
		2r1' <sup>o-2</sup>	x	x	x	x	x	x	x	x	x	x	x	x		
		4r1-r1'-4r1	x	x	x	x	x	x	x	x	x	x	x	x	x	
C - 10	1	≈ 40r2	x	x	x	x	x	x	x	x	x	x	x	x		
		≈ 40r2	x	x	x	x	x	x	x	x	x	x	x	x	x	
C - 11	3.1	4r1' <sup>o-2</sup> -r1'-4r1	x	x	x	x	x	x	x	x	x	x	x	x		
			9r1'-r1-r1'	x	x	x	x	x	x	x	x	x	x	x	x	
	1	18r2	x	x	x	x	x	x	x	x	x	x	x	x	x	
		ref-1	15r2	x	x	x	x	x	x	x	x	x	x	x	x	
		15r2	x	x	x	x	x	x	x	x	x	x	x	x		



Incompatible strain																
Ind <sup>a</sup>	SmPoMuc <sup>b</sup> group	IR <sup>c</sup>	exon <sup>d</sup>													splicing <sup>e</sup>
			3	4	5	6	7	8	9	10	11	12	13	14	15	
IC - 1	3.1	9r1 <sup>4-2</sup>	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1-r2-r2*-r1'	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1	x	x	x	x	x	x	x	x	x	x	x	x	x	
		2r1	x	x	x	x	x	x	x	x	x	x	x	x	x	
		r1-r1'	x	x	x	x	x	x	x	x	x	x	x	x	AS	
		r1*-r1' <sup>10</sup>	x	x	x	x	x	x	x	x	x	x	x	x	AS	
		r1-r1' <sup>11</sup>	x	x	x	x	x	x	x	x	x	x	x	x	AS	
IC - 2	2	≈ 25r2*	x	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)
		≈ 25r2	x	x	x	x	x	x	x	x	x	x	x	x	x	
		15r2	x	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)
		6r2	x	x	x	x	x	x	x	x	x	x	x	x	x	
			4r2	x	x	x	x	x	x	x	x	x	x	x	AbS (3)	
	3.1	r1-2r2 <sup>2-2</sup> -r1-r1'	x	x	x	x	x	x	x	x	x	x	x	x	x	
		r1-2r2-r1-r1'	x	x	x	x	x	x	x	x	x	x	x	x	x	
r1-2r2-r1-r1'		x	x	x	x	x	x	x	x	x	x	x	x	x		
		2r1-r1'	x	x	x	x	x	x	x	x	x	x	x			
IC - 3	2	≈ 40r2	x	x	x	x	x	x	x	x	x	x	x	x		
		≈ 30r2	x	x	x	x	x	x	x	x	x	x	x	x		
		≈ 25r2	x	x	x	x	x	x	x	x	x	x	x	x		
		4r2	x	x	x	x	x	x	x	x	x	x	x	x		
IC - 4	2	11r2 <sup>4-1</sup>	x	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)
		13r2 <sup>10</sup>	x	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)
		≈ 25r2	x	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)
		≈ 30r2 <sup>2-1</sup>	x	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)
		10r2	x	x	x	x	x	x	x	x	x	x	x	x		
IC - 5	3.1	10r1	//+	STOP	x	x	x	x	x	x	x	x	x	x	AbS (4)	
		12r1 <sup>4</sup>	x	//+	STOP	x	x	x	x	x	x	x	x	x	AbS (6)	
	ref-3.1	11r1 <sup>10</sup>	//+	x	//+	STOP	x	x	x	x	x	x	x	x	x	AbS (5)
		11r1 <sup>4-3</sup>	x	x	x	x	x	x	x	x	x	x	x	x		
		11r1	x	x	x	x	x	x	x	x	x	x	x	x		
		8r1	x	x	x	x	x	x	x	x	x	x	x	x		
	2	2r1-2r2-r1-r1'	x	x	x	x	x	x	x	x	x	x	x	x		
		2r1-2r2-r1-r1'	x	x	x	x	x	x	x	x	x	x	x	x		
		r2	x	x	x	x	x	x	x	x	x	x	x	x		
			10r2 <sup>10</sup>	x	x	x	x	x	x	x	x	x	x	AS		
IC - 6	2	≈ 25r2 <sup>2-1</sup>	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)	
		≈ 25r2 <sup>4</sup>	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)	
		≈ 25r2	x	x	x	x	x	x	x	x	x	x	x	x		
		≈ 25r2	x	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)	
		15r2	x	x	x	x	//-	STOP	x	x	x	x	x	AbS (1)		
IC - 7	3.1	2r1	x	x	x	x	x	x	x	x	x	x	x			
		2r1	x	x	x	x	x	x	x	x	x	x	x	x		



**TABLE S2: Glycosylation prediction on Tandem Repeats from Compatible and Incompatible cDNA variants**

			Compatible strain																																					
ind <sup>a</sup>	Variant id. <sup>b</sup>	splicing <sup>c</sup>	TR <sup>d</sup>	glycosylation prediction in TR <sup>e</sup>																																				
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35		
C-1	C13.12r1.1	NO	2r1																																					
	C13.12r1.2	AS	2r1																																					
	C13.1r1.1	AS	r1																																					
	C13.1r1.2	AS	r1																																					
	C13.1r1.3	AS	r1*																																					
	C13.1r1.2	AS	r1*																																					
C-2	C12r2.1	AS	7r2																																					
	C12r2.2	AS	7r2																																					
C-3	C32/45r2	AbS (1)	≈ 45r2																																					
	C32/40r2	AbS (1)	≈ 40r2																																					
	C32/25r2	AbS (1)	≈ 25r2																																					
	C32/11r2	NO	11r2																																					
C-4	C43.19r1-r1-r1'	NO	9r1-r1-r1'																																					
	C42/100r2.1	AS	≈ 100r2																																					
	C42/100r2.2	AS	≈ 100r2																																					
C-5	C53.12r1.1	NO	2r1*																																					
	C53.12r1.2	NO	2r1*																																					
	C53.12r1.3	NO	2r1*																																					
	C53.12r1.4	NO	2r1**																																					
C-6	C6/40r2.1	NO	≈ 40r2																																					
	C6/40r2.2	NO	≈ 40r2																																					
	C6/40r2.3	NO	≈ 40r2																																					
	C6/280r2	NO	≈ 80r2																																					
C-7	C7/25r2.1	NO	≈ 25r2																																					
	C7/25r2.2	NO	≈ 25r2																																					
	C7/25r2.3	NO	≈ 25r2																																					
	C7/25r2.4	NO	≈ 25r2																																					
	C7/25r2.5	NO	≈ 25r2																																					
	C7/30r2	NO	≈ 30r2																																					
C-8	C83.12r1-2r2-r1-r1'	NO	2r1-2r2-r1-r1'																																					
C-9	C93.12r1.1	NO	2r1*																																					
	C93.12r1.2	NO	2r1**																																					
	C93.14r1-r1-4r1	NO	4r1-r1-4r1																																					
C-10	C10/40r2.1	NO	≈ 40r2																																					
	C10/40r2.2	NO	≈ 40r2																																					
C-11	C113.14r1-r1-4r1	NO	4r1-r1-4r1																																					
	C113.19r1-r1-r1'	NO	9r1-r1-r1'																																					
	C11/18r2	NO	18r2																																					
	C11/15r2.1	NO	15r2																																					
	C11/15r2.2	NO	15r2																																					

			Incompatible strain																																				
ind <sup>a</sup>	Variant id. <sup>b</sup>	splicing <sup>c</sup>	TR <sup>d</sup>	glycosylation prediction in TR <sup>e</sup>																																			
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	
IC-1	IC13.19r1	NO	9r1**																																				
	IC13.12r1-2r2-r1'	NO	2r1-2r2-r1'																																				
	IC13.12r1.1	NO	2r1																																				
	IC13.12r1.2	NO	2r1																																				
	IC13.12r1.3	NO	2r1																																				
	IC13.12r1.4	NO	2r1																																				
IC-2	IC13.1r1-r1.3	AS	r1-r1*																																				
	IC13.1r1-r1.2	AS	r1-r1*																																				
	IC13.1r1-r1.3	AS	r1-r1*																																				
	IC2/25r2.1	AbS (1)	≈ 25r2																																				
	IC2/25r2.2	NO	≈ 25r2																																				
	IC2/15r2	AbS (1)	15r2																																				
	IC2/26r2	NO	26r2																																				
	IC2/24r2	AbS (3)	4r2																																				
IC-3	IC23.1r1-2r2-r1-r1'	NO	r1-2r2-r1-r1'																																				
	IC23.1r1-2r2-r1-r1'	NO	r1-2r2-r1-r1'																																				
	IC23.1r1-2r2-r1-r1'	NO	r1-2r2-r1-r1'																																				
	IC23.12r1-r1'	NO	2r1-r1'																																				
IC-3	IC3/240r2	NO	≈ 40r2																																				
	IC3/230r2	NO	≈ 30r2																																				



## Figure legends

### Figure 1: *SmPoMuc* polymorphism at the protein and transcript levels

Positional differences between *SmPoMuc* from compatible (C) and incompatible (IC) strains on silver stained 2D-gels shown with a pH 3-10 non linear (NL) gradient or a pH 3-6 linear (L) gradient (A). Arrows indicate positions of spots corresponding to *SmPoMuc*. Dotted arrows indicate supplementary spots found in the present study using the pH 3-6 linear gradient. Molecular weights (Mw) in kDa are indicated. (B) shows the precursor structure and polymorphism of *SmPoMuc* described in a previous study (Roger et al. 2008a). Three kinds of repeats were identified in *SmPoMuc* cDNAs (r1, r1' and r2); the fourth repeat r3 was only identified at the genomic level only in this study. (C) Agarose gel separation of RT-PCR amplicons obtained from 11 individuals (1-11) of both strains (compatible: C and incompatible: IC). Amplification was performed using consensus primers amplifying the complete coding sequence of all *SmPoMuc*. C-: negative control of amplification.

### Figure 2: Southern blot of *S. mansoni* genomic DNA and BACs containing *SmPoMuc* genes.

Southern Blot of adult worm genomic DNA from IC (lanes 1, 2 and 3) and C (lanes 4, 5 and 6) strains and BAC clones 45D24 (lanes 7, 13, 19), 47P6 (lanes 8, 14, 20), 51E8 (lanes 9, 15, 21), 62J10 (lanes 10, 16, 22), 41B11 (lanes 11, 17, 23), 62F12 (lanes 12, 18, 24).

Genomic DNA from both strains is undigested (lanes 1 and 4), digested by EcoRV (lanes 2 and 5) or digested by BclI (lanes 3 and 6). All genomic DNA lanes were hybridized with a DIG labelled probe corresponding to the 1 kbp genomic repeat shared by all *SmPoMuc* genes. BAC DNAs were digested with EcoRV. Lanes 7 to 12, lanes 13 to 18 and lanes 19 to 24 correspond to the same blots hybridized successively with the 1kbp genomic repeat, r1 and r2

probes, respectively. The membrane was stripped between two successive hybridization procedures.

**Figure 3: Schematic representation of a complete *SmPoMuc* gene.** The complete *SmPoMuc* genes are composed of 15 exons. Exon 2 is included in a genomic repeat that can be repeated several times (a maximum of 20 repeats in *SmPoMuc* 2 genes). These genomic repeats of approximately 1 kilobase are separated by imperfect polypurine tracts (PPT). Positions of genomic primers used for *SmPoMuc* gene amplification (Intron2/3F1 – Exon15R) are indicated by arrows. PCR amplicon position used for gene copy number quantification is indicated by a bold line (—) and the position of a ribozyme between exon 9 and 10 is indicated by an asterisk. Triangles and chevrons indicate complementary sequence positions (12 and 13 nucleotides, respectively) identified in introns of the genomic repeats containing exon 2.

**Figure 4: The *SmPoMuc* multigene family is organized in four paralogous groups that frequently recombine.** *SmPoMuc* genomic DNA sequences corresponding to the 3' portion of *SmPoMuc* genes/alleles (exon 2/exon 15) were obtained by long range PCR and aligned to construct a cladogram with PAUP. Tree branches corresponding to C and IC strains are in red and black, respectively. *SmPoMuc* genes are identified as follows: first the strain (C or IC), then the last exon 2 (r1, r1' or r2) and finally the group (1, 2, or 4) or sub-group (3.1a, 3.1b, 3.2, 3.4, 3.5). This analysis reveals four paralogous sequence groups (gr.1-gr.4). In the right-hand part of the figure, a schematic representation of aligned *SmPoMuc* genomic sequences is given. We annotated the sequences by a color code that uses a different color for sequence fragments of less than 95% identity: gr.1 (red), gr.2 (blue), sub-gr.3.1a (purple), sub-gr.3.1b (pink), sub-gr.3.2 (sky-blue), sub-gr.3.3 (dark-green), sub-gr.3.4 (green) and gr.4 (yellow).

Traces of retrotransposon insertion events (solo-LTR) are present in sub-gr.3.4 and gr.2. Large gaps necessary to obtain alignments are represented by dark lines. Short gap (<28 nucleotides) positions are indicated by rhombi. Short tandem repeats are indicated by (>). Frequent recombination events between *SmPoMuc* family members are apparent.

**Figure 5: *SmPoMucs* contain a putative full-length hammerhead ribozyme between exon 9 and 10.** Alignment of putative ribozymes found in all *SmPoMuc* genes with a functional hammerhead ribozyme of *S. mansoni* (Sm5, AF036742). Asterisks indicate conserved positions in the alignment. Boxes A and B delimit sequences necessary for transcription by RNA polymerase III. The catalytic core nucleotides composed of domains I, II and III are underlined. The conserved nucleotides are numbered using the standard convention (Hertel et al. 1992). The nucleotide position corresponding to G12 essential for ribozyme activity (Martick and Scott 2006) is indicated by a dotted arrow. The scissile bond is indicated by an arrow.

**Figure 6:  $K_S/K_N$  comparison of *SmPoMuc* coding sequences.** The analysis was performed using SNAP (see Material and Methods) on 15, 71 and 56 sequences from groups 1, 2 and 3 respectively. The closed rhombi, open triangles and dashed lines are used for a pair of *SmPoMuc* sequences from groups 1, 2 and 3, respectively. The bisecting dotted line corresponds to  $K_S/K_N=1$ .

**Figure 7: Intermingled repeats (r1/r2) are present in C and IC genomic DNA but not in BACs.** PCR experiments were performed on BACs 45D24 – 47P6 – 51E8 – 62J10 – 41B11 – 62F12 (lanes 1 to 6, respectively) and on DNA from C and IC strains (lanes 7 and 8, respectively); lane 9 corresponds to the PCR negative control. Amplicons were separated on

TAE 1% agarose gels and revealed by ethidium bromide staining. The primers used reveal two r2 exons (A), two r1 exons (B), r2r1 exons (C) or r1r2 exons (D) in two successive genomic repeats.

**Figure 8: FISH mapping of *SmPoMuc* BACs clones.** Metaphase chromosome spreads showing positive signals (arrowheads) hybridized with biotinylated *SmPoMuc* BAC clone DNAs. (a) BAC clone 41B11 gave strong signals in the regions near the centromere of chromosome 3 and on the long arm of chromosome 4; two weaker signals were also detected on the short and on the long arm of chromosome 3. (b) BAC clone 45D24 hybridized to the same regions on chromosome 3 and 4, and yielded a strong supplementary signal at the large heterochromatic pericentromeric region of chromosome 2.

**Figure 9: Aberrant splicing events during *SmPoMuc* gene expression.** The six aberrant splicing variants (AbS) obtained at the cDNA level are shown and numbered (AbS 1 to 6). The *SmPoMuc* genomic sequence areas subject to aberrant splicing are shown: introns are represented as thick lines and rectangles represent exons numbered as described in Figure 3. Normal splice donor and acceptor sites are in uppercase above the schematic gene representation. Aberrant splice sites are in uppercase, underlined, numbered (in brackets) below the schematic representation of *SmPoMuc* genes or in AbS 1-6. The different splicing events leading to them are indicated by dotted or full lines linking the different splice sites. These events are identified by circled numbers corresponding to the AbS they produce. Resulting aberrant splicing leads to exclusion (1-2-3) or inclusion (4-5-6) of DNA, leading to frame-shifts that create non-sense codons in all cases. The different aberrant splicing events observed correspond to cDNA variants given in Table S1: AbS 1 (individuals C3/2, IC/2-4-6-



7-8-9-11/2); AbS 2 (individual IC11/2/7r2); AbS 3 (individual IC2/2/4r2); AbS 4 (individual IC5/3.1/10r1); AbS 5 (individual IC5/3.1/11r1) and AbS 6 (individual IC5/3.1/12r1).

**Figure 10: Western blot of *SmPoMuc* proteins from C and IC strain before and after deglycosylation.** *S. mansoni* sporocyst extracts from C (lanes 1-2) and IC (lanes 3-4) strains were treated with TFMSA (lanes 2-4) or not (lanes 1-3) and submitted to a western blotting using anti-*SmPoMuc* antibodies. The shift in molecular weight observed in lanes 2 and 4 is related to the loss of carbohydrate chains associated with *SmPoMuc* proteins.

**Figure 11: Controlled chaos of *SmPoMuc* polymorphism.** *SmPoMuc* polymorphism controlled at the genomic (A), transcript (B), protein (C) and population (D) levels.

## Supplementary material

### Table S1: cDNA variants obtained for 11 individuals from both Compatible and Incompatible strains

<sup>a</sup>Strain Compatible (C) or Incompatible (IC) and individual number (1-11)

<sup>b</sup>ref-(x) = variant used as reference for mutation identification, x refers to *SmPoMuc* group

<sup>c</sup>TR = Tandem Repeats contained in each variant

- number and type of TR is indicated
- °(x) Synonymous mutation (SM) present in the x<sup>th</sup> repeat of the stretch
- \*(x) Non synonymous mutation (NSM) present in the x<sup>th</sup> repeat of the stretch

<sup>d</sup>exon characteristics of each variant:

- exon presence or absence is annotated by x or in grey, respectively
- green, blue, orange and red exons contain SM, NSM, SM+NSM and STOP codon, respectively
- //+ indicates partial intron inclusion
- //- indicates partial exon exclusion

<sup>e</sup>splicing events:

- AS = Alternative Splicing with conservation of ORF
- AbS (x) = Aberrant Splicing; x refer to AbS type described in Figure 9

### Table S2: Glycosylation prediction on Tandem Repeats from Compatible and Incompatible cDNA variants

<sup>a</sup>Strain Compatible (C) or Incompatible (IC) and individual number (1-11)

<sup>b</sup>Variant Id. = Variant Identification (Individual/*SmPoMuc* group/associated tandem repeat stretch). *SmPoMuc* groups are color-coded: gr1, 2, and 3.1 in blue, yellow and purple, respectively

<sup>c</sup>Splicing variant are indicated by a red box:

- AS = Alternative Splicing with conservation of ORF
- AbS (x) = Aberrant Splicing; x refer to AbS type described in Figure 9

<sup>d</sup>TR = Tandem Repeats contained in each variant

- number and type of TR is indicated
- <sup>o</sup>(x) Synonymous mutation (SM) present in the x<sup>th</sup> repeat of the stretch
- <sup>\*</sup>(x) Non synonymous mutation (NSM) present in the x<sup>th</sup> repeat of the stretch

<sup>e</sup>Glycosylation site prediction in tandem repeats from the different variants:

- repeats containing no prediction site are represented as white boxes
- r1 and r2 repeats containing 1 prediction site on T9 are represented as dark grey boxes
- r2 repeats containing 2 prediction sites on S5 and T9 are represented as clear grey boxes

## DISCUSSION

Cette thèse avait pour objectif la mise en évidence de déterminants moléculaires du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria glabrata* (Theron, Coustau, 2005). Nous supposons que le polymorphisme de compatibilité, caractéristique de cette interaction, n'est que le reflet de la dynamique co-évolutive jouant dans ce système. Il serait maintenu par des mécanismes de diversification des molécules clés, telles que les molécules de reconnaissance de type PRR chez l'hôte (Zhang *et al.*, 2004) et les variants moléculaires chez le parasite. Dans notre modèle, cet équilibre dynamique se traduit, en population naturelle, par la présence de certains couples hôtes / parasites compatibles, d'autres étant incompatibles. L'approche protéomique comparative globale développée entre sporocystes issus de souches compatible et incompatible vis-à-vis de la même souche de mollusque a permis de mettre en évidence certaines protéines différenciellement représentées (Roger *et al.*, 2008a). Parmi ces protéines, nous avons identifié une famille de protéines polymorphes de type mucine. Leur étude a été privilégiée car les mucines parasitaires sont connues pour intervenir dans les mécanismes d'évasion immunitaire, d'invasion de l'hôte ou encore dans la protection du parasite (Hicks *et al.*, 2000; Theodoropoulos *et al.*, 2001). Ceci suggère que ces molécules, que nous avons appelées *SmPoMuc*, pourraient être un facteur déterminant de la compatibilité dans l'interaction *S. mansoni* / *B. glabrata*.

Grâce à l'étude approfondie de l'expression des *SmPoMuc* (Roger *et al.*, 2008b), nous avons mis en évidence, pour la première fois, la présence de ce qui pourrait être des variants moléculaires polymorphes chez un parasite métazoaire (*S. mansoni*) interagissant avec un hôte invertébré (*B. glabrata*). En étudiant les bases génétiques génératrices de ce polymorphisme (Roger *et al.*, 2008c), nous avons mis en évidence des mécanismes originaux permettant de générer un haut niveau de polymorphisme, et ceci à partir d'un nombre réduit de gènes. Cette caractéristique est unique pour des variants moléculaires de parasites eucaryotes et contraste avec les variants antigéniques de parasites protozoaires bien connus tels *P. falciparum*, *T. brucei* ou encore *T. cruzi*. Chez ces protozoaires, ces variants sont générés à partir d'un grand répertoire de gènes.

Par ailleurs, comme nous en avons fait état dans l'introduction, plusieurs travaux récents (Cannon *et al.*, 2002; Dong *et al.*, 2006; Zhang *et al.*, 2004) mettent à mal les paradigmes concernant la réponse immunitaire innée des invertébrés. En effet, il semblerait que la reconnaissance, dans le cadre de l'immunité innée, ne repose pas uniquement sur des récepteurs invariants mais, au contraire, mette en jeu des récepteurs immunitaires diversifiés.

Cette diversification et/ou polymorphisme est prédite par les modèles théoriques de co-évolution, tel que celui de la « Reine Rouge », appliqués aux interactions hôte / parasite (Combes, 2000). Toutefois, ces modèles théoriques de co-évolution n'ont été vérifiés jusqu'à présent que dans des interactions impliquant des vertébrés et leurs pathogènes. Nos travaux apportent un élément nouveau quant à la compréhension de la dynamique co-évolutive dans les interactions hôte invertébré / parasite. En effet, nous avons mis en évidence des variants moléculaires polymorphes présents chez un parasite métazoaire (*S. mansoni*) en interaction avec un hôte invertébré (*B. glabrata*). Ceci ouvre la voie de ce qui pourrait ressembler à de la variation « antigénique » chez un parasite d'invertébré. Cela suggère également que les interactions entre hôtes invertébrés et parasites pourraient être soumises aux mêmes contraintes évolutives que les interactions entre hôtes vertébrés et parasites.

# I. LES *SmPoMuc* ET LES GLYCOPROTEINES / MUCINES D'AUTRES SYSTEMES HOTE / PARASITE

## 1. Les *SmPoMuc* au coeur de la compatibilité

La compatibilité de l'interaction entre *Schistosoma mansoni* et son hôte intermédiaire *Biomphalaria glabrata*, et par conséquent le succès parasitaire, résulterait de la capacité du parasite à échapper aux défenses internes de l'hôte. Différents travaux suggèrent que les schistosomes y parviendraient, notamment, grâce à leur capacité à développer une stratégie de type camouflage/mimétisme moléculaire (Abu-Shakra *et al.*, 1999; Damian, 1997; Salzet *et al.*, 2000). Le camouflage moléculaire se traduit par l'adsorption à la surface du parasite de déterminants moléculaires de l'hôte (Van Der Knaap *et al.*, 1985). Le mimétisme moléculaire, quant à lui, consiste à l'expression de déterminants moléculaires de surface commun à ceux du mollusque (Damian, 1987; Yoshino, Bayne, 1983). Dans tous les cas, les épitopes du parasite ne sont pas reconnus et ne provoquent pas de réponse du système interne de défense.

Cependant, lorsque le stade infestant miracidium pénètre dans les tissus du mollusque, il ne dispose pas de la capacité à développer ce type de stratégie. En effet, durant les premières heures de vie intra-molluscale, le parasite va tout d'abord perdre ses plaques ciliées (Pan, 1996). A cet instant, ce parasite qui n'est plus un miracidium mais qui n'est pas encore un sporocyste, est appelé post-miracidium. Il s'agit là du stade le plus vulnérable du cycle de vie de *S. mansoni*. Il est dépourvu d'une surface le protégeant contre le système de défense de l'hôte. Les tissus internes, tels que la lame basale et les muscles, sont directement exposés au système de défense de l'hôte (Pan, 1996). Malgré cela, le parasite, dans le cas d'une interaction compatible, va parvenir à se transformer en sporocyste primaire, qui sera alors pourvu d'un tégument de protection (Pan, 1996). Durant ce processus d'élaboration de l'épithélium du futur sporocyste primaire, une grande quantité de produits excrétés/sécrétés (ES) sont relargués (Lodes, Yoshino, 1989). Une partie est contenue dans des vésicules participant à la formation du nouveau tégument. La caractérisation du contenu protéique des produits ES de *S. mansoni*, a fait l'objet d'une étude à laquelle j'ai participé au laboratoire (Guillou *et al.*, 2007) (cf. *Annexe 1*). Ce travail nous a permis de caractériser des protéines impliquées, notamment, dans la détoxification lors du stress oxydatif induit par les hémocytes (cellules de défenses circulantes du mollusque). Durant le stade post-miracidium, qui dure environ cinq heures, le parasite développerait une stratégie de « détoxification » que

nous avons appelé immunoprotection. Cependant, contrer l'immunité de l'hôte signifie également être capable de bloquer les récepteurs de l'immunité innée du mollusque (pour une interaction compatible versus incompatible). Afin d'identifier ces déterminants moléculaires nous avons développé une approche protéomique comparative globale entre sporocystes issus de souches compatible et incompatible vis-à-vis de la même souche de mollusque. Cette étude a permis de mettre en évidence certaines protéines différenciellement représentées (Roger *et al.*, 2008a). Parmi ces protéines, nous avons identifié une famille de protéines polymorphes de type mucine.

### **1.1. Les SmPoMuc, des glycoprotéines de type mucine**

Les glycoprotéines sont très répandues dans le règne vivant (Montgomery, 1970). Elles sont retrouvées à la surface des microorganismes où elles jouent un rôle important dans l'évasion immunitaire. Les VSG de *Trypanosoma* (cf. i-1.2.2.1 *Les VSG chez Trypanosoma brucei*) constituent un exemple de glycoprotéines de surface au rôle fondamental pour échapper à la réponse de l'hôte.

Les mucines sont des glycoprotéines avec des caractéristiques structurales particulières (Strous, Dekker, 1992). Tout d'abord, elles sont caractérisées par une chaîne polypeptidique composée de 20 à 50% par des résidus glycine (Gly), alanine (Ala), proline (Pro), sérine (Ser) et thréonine (Thr). Ces résidus sont principalement présents au niveau d'unités répétées (UR), de 9 à 80 acides aminés, organisées en tandem. Ces UR, dont le nombre peut aller jusqu'à 100, constituent un domaine central variable (VNTR pour « Variable Number of Tandem Repeats ») (Carlstedt *et al.*, 1985; Strous, Dekker, 1992). Ensuite, les résidus Ser et Thr (Ser/Thr), présents au niveau des VNTR, servent de site de liaison aux chaînes glycaniques. Elles sont liées de façon covalente aux résidus Ser/Thr par l'intermédiaire d'un N-acétylgalactosamine (liaison O-glycosidique) (Hanisch, 2001). Enfin, le niveau de glycosylation est tel, que les carbohydrates peuvent représenter de 50 à 85 % du poids sec des mucines. Les mucines, largement décrites chez les mammifères, peuvent être sécrétées (Gendler, Spicer, 1995; Gum, 1995) ou associées à la membrane (Hilkens *et al.*, 1992; Patton *et al.*, 1995).

Les SmPoMuc identifiées au cours de cette thèse présentent des caractéristiques structurales de mucines (Roger *et al.*, 2008b). En effet, elles sont composées de 25 à 50 % de résidus Gly, Ala, Pro, Ser et Thr. Ces résidus sont essentiellement présents au niveau de la région N-terminale composée d'un domaine VNTR. Ce domaine est constitué d'UR, de 9 acides aminés (de type r1(r1') et/ou r2), qui peuvent être répétées en tandem de 2 à 100 fois.

De plus, ces UR possèdent des acides aminés dont les positions relatives sont favorables aux O-glycosylations (Thanka Christlet, Veluraja, 2001). En effet, la Proline en position -1 par rapport à la Thr, au niveau des UR r1( r1') et r2, semble favorable à la O-glycosylation de cette Thr. Au niveau de l'UR r2, cette même Pro est en position +3 par rapport à la Ser ce qui pourrait favoriser la O-glycosylation de ce résidu (O'connell *et al.*, 1991; Wilson *et al.*, 1991). Enfin, comme nous l'avons mis en évidence, les *SmPoMuc* sont hautement glycosylées au niveau du domaine VNTR (Roger *et al.*, 2008b). Ainsi, pour toutes ces caractéristiques, les *SmPoMuc* peuvent être considérées comme des molécules de type mucine.

## 1.2. Les *SmPoMuc*, polymorphisme et statut de glycosylation

Outre le fait d'être hautement glycosylées, les *SmPoMuc* sont des protéines extrêmement polymorphes. Nous verrons que ces deux caractéristiques ne sont pas indépendantes. Le polymorphisme s'exprime, au niveau intra- et inter-souche mais également au niveau individuel, sur toute la molécule (Roger *et al.*, 2008c). En effet, la région N-terminale, composée du domaine VNTR hautement glycosylé, est caractérisée par un polymorphisme de taille (nombre plus ou moins important d'UR) et de nature des UR (type r1, r2 ou r1 et r2 combinées). La région C-terminale n'est, à priori, pas glycosylée et ne présente pas d'homologie de séquence ni de structure pour des domaines connus. Le polymorphisme de cette région est essentiellement le fruit de mécanismes d'épissages alternatifs ou aberrants (Roger *et al.*, 2008c).

Le polymorphisme de la région N-terminale, semblerait avoir un impact direct sur le statut de glycosylation des *SmPoMuc* (comme prédit par les logiciels de prédiction de glycosylations, (Roger *et al.*, 2008c)). En effet, le nombre, le type et la combinatoire des UR, qui constituent le domaine VNTR contenant les sites de O-glycosylation, sont autant d'éléments qui pourraient influencer le niveau de glycosylation des *SmPoMuc* (Roger *et al.*, 2008c). De plus, nous avons mis en évidence que le polymorphisme de la région C-terminale pourrait également influencer sur ce statut, bien qu'elle ne possède pas de sites potentiels de glycosylation (Roger *et al.*, 2008c). Ainsi, il semblerait que le haut degré de polymorphisme des *SmPoMuc* se répercute sur leur statut de glycosylation. Cela n'est pas sans conséquence, car les carbohydrates jouent un rôle majeur dans les interactions hôtes / parasites. En effet, comme nous le verrons par la suite chez *Coccidioides*, *Trypanosoma cruzi* ou encore *Toxocara canis*, l'immunogénicité des carbohydrates est utilisée pour bloquer les anticorps chez l'hôte vertébré (stratégie « écran de fumée »). Chez *S. mansoni*, les carbohydrates des



*SmPoMuc* pourraient cibler les molécules de reconnaissance de type lectine de l'hôte mollusque. Ainsi, une extrapolation un peu osée, pourrait nous amener à poser l'hypothèse que les *SmPoMuc* pourraient cibler des récepteurs immunitaires diversifiés de *B. glabrata*, tels les FREP (cf. i-II.1.3. *Les FREP chez Biomphalaria glabrata*). Par conséquent, le statut de glycosylation, différent entre souche compatible et incompatible (Roger *et al.*, 2008c), pourrait être à l'origine d'une reconnaissance différentielle par ce type de récepteurs du système immunitaire du mollusque. Ainsi, dans le cas d'une interaction compatible le statut de glycosylation des *SmPoMuc* concorderait (« matched ») avec les molécules de reconnaissance de l'hôte, et bloquerait ces dernières.

De plus, au niveau individuel, vient s'ajouter un polymorphisme d'expression. En effet, chaque larve de *S. mansoni* est capable d'exprimer une combinaison de variant(s) *SmPoMuc* unique qui pourrait être associée à un statut de glycosylation spécifique (Roger *et al.*, 2008c). La compatibilité serait alors testée de façon indépendante pour chaque miracidium lors de l'interaction avec un même hôte (Basch, 1975; Theron, Coustau, 2005). Le succès ou l'échec de l'infection dépendrait de la concordance ou de la non concordance des phénotypes de l'hôte (molécules de reconnaissance) et du parasite (statut de glycosylation des *SmPoMuc*) (modèle de concordance des phénotypes, Figure 19) (Theron, Coustau, 2005).

## **2. Une stratégie « écran de fumée »**

### **2.1. Les *SmPoMuc* chez *S. mansoni***

Les *SmPoMuc* sont exprimées spécifiquement durant les stades en interaction avec le mollusque (miracidium et sporocyste primaire) (Roger *et al.*, 2008b). L'expression est supérieure chez le stade infestant, miracidium. Cette expression « anticipée » permet ainsi au parasite de disposer d'une réserve de *SmPoMuc*, stockée dans la glande apicale, prêt à être sécrétée au moment de l'interaction (Roger *et al.*, 2008b). En effet, la reconnaissance par les hémocytes (1 heure post infestation) et le processus d'encapsulation (4 à 8 heures post infestation) est très rapide dans une interaction incompatible. Ceci suggère que les molécules déterminantes de la compatibilité doivent être présentes constitutivement chez le stade infestant comme nous le montrons (Roger *et al.*, 2008a). De plus, la glande apicale est un organe impliqué dans la sécrétion d'enzymes nécessaires à la pénétration du miracidium dans les tissus du mollusque. Les *SmPoMuc* seraient ainsi libérées de façon concomitante lors de l'entrée du parasite dans son hôte mollusque et pourraient jouer leur rôle de blocage

de la reconnaissance immunitaire. Elles constitueraient ainsi un « écran de fumée » autour du parasite, empêchant la liaison des PRR aux molécules de surface réellement délétères pour le parasite.

Cette hypothèse est étayée par des résultats obtenus dans d'autres interactions où des molécules de type mucine (« mucin-like »), analogues aux *SmPoMuc*, ont été décrites. Nous traiterons des glycoprotéines de surface (SOWgp) des champignons pathogènes *Coccidioides immitis* et *C. posadasii* (cf. d-I.2.2. *Les SOWgp chez Coccidioides immitis et C. posadasii*) (Hung *et al.*, 2007). Nous nous intéresserons également aux parasites protozoaires tel que *T. cruzi* (cf. d-I.2.3. *Les TcSMUG et TcMUC chez Trypanosoma cruzi*) (Hicks *et al.*, 2000) et aux helminthes (Theodoropoulos *et al.*, 2001) tel que le nématode *Toxocara canis* (cf. d-I.2.4. *Les Tc-MUC chez Toxocara canis*).

## **2.2. Les SOWgp chez *Coccidioides immitis* et *C. posadasii***

*Coccidioides immitis* et *C. posadasii* sont des champignons pathogènes des voies respiratoires chez l'Homme, responsables de la fièvre du désert (Galgiani, 1999). La phase saprophyte de *Coccidioides* produit des spores (arthroconidies) qui sont dispersées par le vent et peuvent être inhalées par l'homme. Au niveau des poumons, ces spores (2 x 4 µm) se transforment en sphérules, qui sont des cellules multinucléées. Ces dernières croissent de façon isotropique jusqu'à atteindre une taille de 60 à 100 µm de diamètre. Ces cellules parasitaires volumineuses subissent alors un processus élaboré de compartimentation cytoplasmique. Ce dernier conduit à la formation et la libération d'une multitude d'endospores (de 4 à 10 µm de diamètre) chez l'hôte (Cole, Hung, 2001). Chaque endospore croît alors rapidement de manière isotropique et se différencie en une seconde génération de sphérule. La taille relativement importante des sphérules mûres leur permet sans doute d'échapper à la phagocytose par les neutrophiles, macrophages ou cellules dendritiques. En revanche, les endospores libérées par les sphérules mûres sont beaucoup plus petites et donc sujettes à la phagocytose. Cependant, des études *in vitro* montrent que de nombreuses cellules survivent à l'attaque initiale du système immunitaire inné (Drutz, Huppert, 1983). De plus, les extraits protéiques de la membrane externe des sphérules sont capables d'induire les réponses immunes humorale et cellulaire chez les individus infectés (Hung *et al.*, 2000; Hung *et al.*, 2002). L'analyse de coccidioïdomycoses développées sur le modèle murin révèlent, par ailleurs, qu'une réponse immune de type Th1 est cruciale dans la défense contre les infections coccidioidales (Campbell *et al.*, 2000; Cox, Magee, 1998). Cependant, les

individus infectés développent une pathologie caractérisée par une inflammation persistante et une réponse immune de type Th2 non protectrice et favorable au pathogène (Magee, Cox, 1995; Magee, Cox, 1996).

Un antigène immunodominant, présent à la surface de *Coccidioides* durant la phase parasitaire, serait capable de détourner la réponse immune vers une réponse de type Th2 (Cole *et al.*, 1988a; Cole *et al.*, 1988b). Il s'agit d'une glycoprotéine de la paroi externe, appelée SOWgp (« Sphérule Outer Wall glycoprotein »). Elle est impliquée dans l'adhésion des cellules à l'épithélium respiratoire et joue donc un rôle important dans l'établissement de l'infection. Cette glycoprotéine, dans sa partie N-terminale, est composée d'un peptide signal et d'un propeptide. La partie C-terminale, quant à elle, contient un site potentiel ainsi qu'une séquence signal nécessaire à l'ancrage GPI. Ces régions N- et C-terminales sont conservées entre souches, contrairement à la région centrale qui elle est polymorphe (Hung *et al.*, 2002; Johannesson *et al.*, 2005). En effet, elle est composée d'unités répétées en tandem, dont le nombre peut varier d'une souche à l'autre. Cette variation n'est pas due à un mécanisme d'épissage alternatif mais résulte d'un polymorphisme génétique entre les souches (Hung *et al.*, 2002). Ces répétitions de 41 à 47 acides aminés, sont riches en acides aminés proline, aspartate, tyrosine et glycine. C'est une caractéristique des régions riches en proline ou PRRs (« Proline-Rich Regions »). Les protéines de surfaces ou sécrétées contenant des domaines PRRs jouent un rôle important dans les interactions protéines / protéines et le repliement des protéines (Brady *et al.*, 1998; Kay *et al.*, 2000). De plus, chez *Coccidioides*, ces unités répétées semblent contenir l'épitope immunodominant capable d'induire une réponse de type Th2. Cette glycoprotéine serait donc impliquée dans une stratégie d'immunomodulation de la réponse immunitaire de l'hôte, favorisant la persistance de l'infection (Hung *et al.*, 2002).

Les glycoprotéines SOWgp semblent également impliquées dans une stratégie d'évitement de la reconnaissance par l'hôte. Les opportunités des phagocytes d'ingérer *Coccidioides* sont restreintes aux stades arthroconidies et endospores (Hung *et al.*, 2007). En conséquence le pathogène a développé un mécanisme afin d'éviter la reconnaissance par l'hôte des SOWgp, présentes à la surface des endospores. Des études ont mis en évidence que les SOWgp sont exprimées uniquement durant la phase de croissance isotropique des sphérules (environ pendant les 72 premières heures de la phase parasitaire) (Hung *et al.*, 2002). Durant cette phase de croissance, les SOWgp sont alors localisées à la surface des sphérules mais également dans le cytoplasme, à l'intérieur de vésicules. Ensuite, durant la phase de compartimentation et les premiers stades de différenciation des endospores (à partir de 96 heures), les SOWgp sont localisées au niveau des parois néosynthétisées (Hung *et al.*, 2005). Cependant, une forte réduction du niveau des transcrits ainsi que de la quantité de

SOWgp, est observé durant la phase terminale de maturation des sphérules (Hung *et al.*, 2002). Ainsi, il semblerait que le niveau de production de ces glycoprotéines soit contrôlé par un système de régulation. Un tel système permettrait aux endospores, qui sont vulnérables lorsqu'elles sont libérées, d'éviter d'être reconnue par le système de défense en ne permettant pas la liaison des anticorps à leur surface. Il semblerait qu'une métalloprotéinase (Mpe1), exprimée durant la phase de différenciation des endospores, soit impliquée (Hung *et al.*, 2005). En effet, cette métalloprotéinase est capable de dégrader les SOWgp présentes à la surface des parois néosynthétisées des endospores. Ainsi, les endospores libérées seront dépourvues de l'antigène immunodominant de surface. Cette stratégie va leur permettre d'éviter d'être reconnues par le système de défense de l'hôte. De plus, les produits de dégradation des SOWgp sont toujours immunogènes et sont relargués au moment de la libération des endospores. Ils pourraient alors jouer le rôle d'un « écran de fumée ». En effet, ils permettraient de lier les anticorps et détourneraient ainsi le système immunitaire afin d'induire une réponse de type Th2.

### **2.3. Les TcSMUG et TcMUC chez *Trypanosoma cruzi***

*Trypanosoma cruzi* est l'agent responsable de la maladie de Chagas, troisième grande maladie humaine après la malaria et la schistosomose, qui affecte 18 millions de personnes en Amérique latine (Barrett *et al.*, 2003). Le cycle de vie se déroule chez deux hôtes, un mammifère, dont l'homme, et un insecte vecteur, une punaise hématophage Reduviidae. Au cours d'un repas sanguin, la punaise dépose ses déjections sur la peau de l'hôte piqué. Les trypomastigotes métacycliques qui s'y trouvent pénètrent par la lésion et gagnent la circulation sanguine. Dans les cellules hôtes, ils se transforment en amastigotes et se multiplient par division binaire. Après destruction de la cellule hôte, ils sont libérés dans le sang sous forme trypomastigote. Cette forme peut réinfecter d'autres cellules. Elle peut également être absorbée par la punaise au cours d'une piqûre. Dans ce cas, chez l'insecte vecteur, le parasite se transforme en épimastigote dans l'intestin moyen, et s'y multiplie avant de donner des trypomastigotes métacycliques (Buscaglia, Di Noia, 2003).

Quelque soit le stade parasitaire, hormis la forme intracellulaire amastigote, la surface du trypanosome est recouverte d'une épaisse couche de glycoprotéines. Comme la majorité des composés de surface des protozoaires parasites, ces glycoprotéines sont ancrées à la membrane plasmique par un glycosyl-phosphatidyl-inositol (GPI) (Ferguson *et al.*, 1994). Chez *T. cruzi*, les principales glycoprotéines de surface sont les molécules de type mucine (TcMUC). Elles sont riches en résidus Thr, Ser et Pro, qui sont des sites favorables à

la fixation de chaînes oligosaccharidiques par l'intermédiaire de liaisons O-glycosidiques (Previato *et al.*, 1994; Schenkman *et al.*, 1993). De plus, les TcMUC ont la particularité d'être des accepteurs majeurs d'acide sialique. Ce dernier est transféré d'un sialoglycoconjugué de l'hôte vers les protéines TcMUC du parasite. Ce transfert est réalisé par l'intermédiaire d'une *trans*-sialidase, présente à la surface du parasite et ancrée par un GPI. La sialylation des protéines TcMUC est cruciale pour la viabilité et la propagation du parasite.

Les TcMUC jouent un rôle important dans la protection et l'infectivité du parasite ainsi que dans la modulation de la réponse immune de l'hôte tout au long du cycle de vie de *T. cruzi* (Acosta-Serrano *et al.*, 2001). Toutefois, il est possible de distinguer deux groupes de mucines chez *T. cruzi* : tout d'abord, celles exprimées chez les stades en interaction avec l'insecte (épimastigote et trypomastigote métacyclique) d'une masse moléculaire apparente de 35 à 50 kDa (Previato *et al.*, 1994; Schenkman *et al.*, 1993), et ensuite, celles exprimées lors de l'interaction avec le mammifère (trypomastigote) d'une masse moléculaire apparente de 60 à 200 kDa (Almeida *et al.*, 1994a; Schenkman *et al.*, 1991).

### **2.3.1. Les molécules de type mucine TcSMUG des stades en interaction avec l'insecte**

Les molécules de type mucine, présentes à la surface des stades épimastigotes et trypomastigotes métacycliques, sont codées par une famille de 70 à 80 gènes appelée *TcSMUG* (« *T. cruzi* Small MUCin-like Gene family ») (Di Noia *et al.*, 2000). Cette famille est composée de deux groupes, L pour « Large » et S pour « Small ». Les régions N- et C-terminales des précurseurs, qui correspondent respectivement au peptide signal et à l'ancre GPI, sont conservées entre ces deux groupes. En revanche, l'organisation de la région centrale variable, riche en Thr, est différente. En effet, alors que le groupe L possède de 1 à 3 unités répétées de 17 acides aminés, le groupe S n'en contient aucune. Cependant, ce dernier possède deux domaines composés uniquement d'une succession d'acides aminés Thr. Le premier domaine possède de 17 à 20 acides aminés Thr et le second de 1 à 3 (Di Noia *et al.*, 2000). Ainsi, des différences intervenant soit au niveau du nombre de répétitions au sein du groupe L soit du nombre de Thr dans les successions au sein du groupe S, peuvent être à l'origine d'un certain polymorphisme entre souches de *T. cruzi* (Previato *et al.*, 1994; Previato *et al.*, 1995).

Toutefois, la composition en acides aminés et en carbohydrates des TcSMUG, ainsi que leur structure est quasi identique entre les stades épimastigotes et trypomastigotes métacycliques d'une même souche (Serrano *et al.*, 1995). La seule différence majeure réside

dans la structure lipidique de l'ancre GPI. En effet, chez le stade épimastigote (non infestant), le phosphatidyl-inositol possède majoritairement un alkylacylglycérol. En revanche, chez le stade trypomastigote métacyclique (infestant l'hôte mammifère), celui-ci est remplacé par une céramide (Serrano *et al.*, 1995). Le rôle biologique de cette modification entre ces deux stades peut être corrélé à l'augmentation d'infectivité et la capacité du stade infestant à libérer les mucines dès qu'il envahit une cellule hôte (Schenkman *et al.*, 1993). En effet, les TcSMUG jouent un rôle dans l'attachement et l'invasion des cellules de l'hôte mammifère par le stade trypomastigote métacyclique (Ruiz *et al.*, 1998; Ruiz Rde *et al.*, 1993). De plus, les TcSMUG sont présentes en abondance à la surface ( $2 \times 10^6$  molécules par parasite) et sont résistantes aux protéases et glycosidases (Mortara *et al.*, 1992). Ces propriétés confèrent ainsi au trypanosome une protection contre les protéases qui sont présentes dans le tractus intestinal de l'insecte, plus particulièrement pour le stade épimastigote.

### **2.3.2. Les molécules de type mucine TcMUC des stades en interaction avec le mammifère**

Les molécules de type mucine du stade trypomastigote sont codées par une famille de 500 à 700 gènes appelée *TcMUC* (« *T. cruzi* MUCin ») (Di Noia *et al.*, 1995). Les protéines de cette famille partagent des régions N- et C-terminales homologues, qui correspondent respectivement au peptide signal et à l'ancre GPI. Cependant, des différences au niveau du domaine central, sont à l'origine de leur classification en deux groupes, TcMUCI et II (Buscaglia *et al.*, 2004). La région N-terminale du domaine variable des protéines TcMUCI est composée d'un petit domaine hypervariable suivi de 2 à 10 unités répétées Thr<sub>8</sub>-Lys-Pro<sub>2</sub> en tandem. La région C-terminale de ce domaine est riche en Thr. Le domaine variable des protéines TcMUCII, quant à lui, ne possède pas d'unités répétées Thr<sub>8</sub>-Lys-Pro<sub>2</sub>, mais est riche en résidus Thr, Ser et Pro nécessaires pour les O-glycosylations. Ainsi, la surface d'un même trypanosome est recouverte d'une véritable mosaïque antigénique, contrairement à la majorité des protozoaires parasites qui n'exprime qu'un seul variant à la fois (cf. *i-I.2. Pathogènes*).

Les TcMUC peuvent intervenir de différentes manières dans l'attachement et l'invasion des cellules hôtes mais également dans la modulation ou l'évitement de la réponse immunitaire du mammifère. Tout d'abord, l'hypervariabilité qui caractérise le domaine central semble favoriser l'attachement et l'invasion d'un large spectre de cellules et d'hôtes mammifères par le trypanosome (Di Noia *et al.*, 1998). De plus, l'hypervariabilité de cette région, qui peut être immunogène (Pollevick *et al.*, 2000), permettrait de protéger la forme

trypomastigote extra-cellulaire du système immunitaire de l'hôte. Ensuite, les TcMUC, contrairement au TcSMUG, possèdent des oligosaccharides alpha-galactosylés ( $\alpha$ -gal) hautement immunogènes (Almeida *et al.*, 1994a; Almeida *et al.*, 1994b). Ces antigènes représentent d'ailleurs la cible principale des anticorps trypanolytiques (anti- $\alpha$ -gal). Le niveau de sialylation des TcMUC va alors être déterminant afin de conférer une protection à la forme trypomastigote contre une lyse médiée par les anticorps anti- $\alpha$ -gal. En effet, la présence d'acide sialique semblerait stabiliser les TcMUC entre elles à la surface du parasite. Cela empêcherait leur agrégation par les anticorps anti- $\alpha$ -gal bivalent, et, par conséquent, la déformation de la surface et la lyse du parasite (Pereira-Chioccola *et al.*, 2000). De plus, cette couverture de TcMUC sialylée peut protéger le parasite des enzymes et autres molécules oxydantes utilisées par l'hôte (Pereira-Chioccola, Schenkman, 1999). Ainsi, la sialylation des TcMUC et l'immunogénicité des épitopes des carbohydrates sont utilisés pour bloquer les anticorps anti- $\alpha$ -gal. De plus, le clivage de l'ancre GPI, par une phospholipase C de l'hôte, aura comme conséquence la libération des TcMUC (ou des complexes anticorps / TcMUC) dans l'environnement du parasite. Ce mécanisme de type « écran de fumée » va ainsi permettre au parasite de détourner et d'éviter le système immunitaire de l'hôte.

#### **2.4. Les Tc-MUC chez *Toxocara canis***

*Toxocara canis* est un nématode parasite responsable d'une zoonose cosmopolite. Les hôtes définitifs dans lesquels le cycle de ce parasite peut s'accomplir sont des chiens et autres canidés sauvages ou domestiques. Cependant, les larves manquent de spécificité vis-à-vis de leurs hôtes et sont ainsi capables d'infester un large spectre d'hôte, dont l'homme (Gillespie, 1987; Gillespie *et al.*, 1993). Ces hôtes paraténiques ne permettent pas le développement du parasite. Celui-ci va alors arrêter son développement au stade larvaire et maintenir un métabolisme actif. Ces larves en « attente » sont capables de migrer vers les tissus musculaires et nerveux et sont à l'origine de symptômes viscéraux et oculaires (Glickman, Schantz, 1981). Chez les primates, la larve de *T. canis* peut survivre sous cette forme d'attente durant 9 ans (Beaver, 1966). Ainsi, le parasite semble posséder un mécanisme lui permettant de neutraliser de manière efficace la réponse immunitaire de l'hôte.

En effet, les nématodes parasites ont développé différentes stratégies d'évasion du système immunitaire de l'hôte (Behnke *et al.*, 1992; Maizels *et al.*, 1993). L'une de ces stratégies est la capacité de la larve de *T. canis* à perdre et à renouveler son manteau de

surface qui recouvre et protège la cuticule du parasite. Ce processus a lieu en réponse à la liaison des cellules de défenses ou des anticorps à la larve (Badley *et al.*, 1987; Page *et al.*, 1992b; Smith *et al.*, 1981). Ce manteau de surface est composé majoritairement d'un groupe de glycoprotéines antigéniques. Ces glycoprotéines sont également présentes dans les produits excrétés/sécrétés de *Toxocara*, TES (« *Toxocara* Excretory/Secretory »), et ont été appelées TES-120 (Page *et al.*, 1992a). Ces glycoprotéines possèdent un peptide signal, un domaine central riche en Ser et Thr et un domaine riche en cystéines. Le domaine central est composé d'unités répétées en tandem hautement O-glycosylées (Gems, Maizels, 1996). Ces caractéristiques permettent de classer ces glycoprotéines dans la catégorie des molécules de type mucine. A l'heure actuelle, quatre gènes codant ces protéines de type mucine ont été identifiés (*nmuc-1*, *muc-2*, *-3* et *-4*). Ils codent respectivement quatre protéines de type mucine, Tc-MUC-1 à 4, appartenant à la famille TES-120 (Loukas *et al.*, 2000). Cependant, bien que leur structure soit similaire, des différences au niveau du domaine de type mucine permettent de discriminer les différentes Tc-MUC. En effet, le nombre d'unités répétées constituant ce domaine ainsi que leur séquence en acide aminé est variable (Loukas *et al.*, 2000).

Ainsi, le haut niveau d'expression, le renouvellement continu du manteau de surface et la sécrétion de ces glycoprotéines de type mucine par les stades larvaires « en attente » semblent jouer un rôle dans l'évasion immunitaire (Gems, Maizels, 1996). Elles pourraient donc créer un « écran de fumée » et favoriser la formation des complexes antigènes / anticorps à distance du parasite (Marin *et al.*, 1992). Ainsi, la combinaison d'une surexpression de protéines de type mucine et d'une immunoprotection pourrait permettre à *T. canis* d'éviter la reconnaissance du système immunitaire de l'hôte.

## 2.5. Conclusion

Ces glycoprotéines et molécules de type mucine, jouent donc un rôle dans les interactions hôte / parasite. En effet, elles interviennent, entre autre, dans l'immunoprotection et l'évasion immunitaire. Chez *Coccidioides*, *T. cruzi* et *T. canis*, elles contribuent au développement d'une stratégie « écran de fumée » qui permet au parasite de maintenir une infection chez l'hôte face au système immunitaire. Chez *S. mansoni*, les *SmPoMuc* pourraient jouer le même rôle car elles ont en commun un certain nombre de caractéristiques et sont également présentes au moment le plus critique du développement parasitaire. De plus, le polymorphisme constitue une particularité commune à toutes ces protéines. Cette caractéristique semble être essentielle pour des protéines impliquées dans la



course aux armements entre hôte et parasite. Les *SmPoMuc* de *S. mansoni* présentent un niveau extrêmement élevé de polymorphisme, dont les bases génétiques seront discutées dans la partie suivante.

## II. BASES GENETIQUES DU POLYMORPHISME DES *SmPoMuc*

Les résultats obtenus au travers de cette thèse suggèrent que les *SmPoMuc* pourraient être un facteur déterminant du polymorphisme de compatibilité dans l'interaction *S. mansoni* / *B. glabrata*. Elles pourraient constituer l'élément parasite attendu du modèle de « matching phenotype », caractéristique de l'interaction *B. glabrata* / *S. mansoni*. Dans ce contexte, le haut degré de polymorphisme des *SmPoMuc*, qui s'exprime, au niveau intra- et inter-souche, à la fois sur la partie N- et C-terminale, constitue une particularité très intéressante. Cette deuxième partie de la discussion se focalise sur les éléments structuraux et la régulation de l'expression des gènes codant les *SmPoMuc* susceptibles d'expliquer cet extraordinaire niveau de polymorphisme.

### 1. Le polymorphisme des *SmPoMuc* au niveau génomique

#### 1.1. Evolution des gènes selon le modèle « birth and death »

Le modèle « birth and death » (Nei, Hughes, 1992) est un modèle qui a été proposé afin d'expliquer l'évolution de certaines familles multigéniques, notamment celles en rapport avec le système immunitaire, comme les immunoglobulines (Ota, Nei, 1994) (cf. i-1.1.1.1 *Immunoglobulines et TCR*) et le CMH (Hughes, Nei, 1990) (cf. i-1.1.1.2. *Complexe Majeur d'Histocompatibilité*). Ce modèle prédit la création, perpétuelle et fréquente, de nouveaux gènes par duplication. Ces duplications sont souvent le résultat de « crossing over » inégaux entre gènes (Zhang, 2003). Les gènes dupliqués (ou paralogues) peuvent alors devenir progressivement des gènes non fonctionnels (pseudogènes). En effet, lorsqu'ils ne sont pas sous sélection, ils vont accumuler progressivement des mutations jusqu'à devenir des gènes non fonctionnels (processus de pseudogénéisation, (Lynch, Conery, 2000)). Toutefois, certains gènes dupliqués sont maintenus dans le génome afin de remplir des fonctions spécifiques. Deux mécanismes, empêchant la pseudogénéisation, peuvent être responsables du maintien de ces gènes. Tout d'abord le mécanisme de conversion génique (Palmer, Brayton, 2007), qui correspond à des échanges non réciproques de séquences similaires entre gènes. L'intervention de ces échanges à une fréquence élevée se traduira par une forte similarité de séquence entre les gènes paralogues. Ainsi, ces gènes évoluent de façon dépendante selon le modèle d'évolution concertée (Liao, 1999). Un autre moyen d'empêcher

la divergence des gènes dupliqués est qu'ils soient soumis à une forte sélection purificatrice. Cette sélection agirait principalement (mais pas uniquement) par l'élimination des mutations désavantageuses, impliquant la conservation des mutations neutres (Zhang, 2003).

Ainsi, les familles multigéniques qui évoluent sous le modèle « birth and death » (Nei, Rooney, 2005), sont caractérisées par des gènes et de nombreux pseudogènes organisés en tandem sur le même chromosome. Ce mode d'évolution rapide explique également la faible divergence de séquence entre gènes et pseudogènes. Toutes ces caractéristiques sont partagées par la famille multigénique codant les *SmPoMuc*. En effet, cette famille est constituée de 6 à 9 membres qui peuvent être divisés en quatre groupes de séquences paralogues (gr.1-gr.4) (Roger *et al.*, 2008c). Les trois premiers groupes correspondent à des gènes exprimés. Le quatrième groupe et certains sous-groupes du troisième groupe semblent ne pas être exprimés et sont probablement des pseudogènes. De plus, des pseudogènes tronqués sont également présents. Ces gènes et pseudogènes sont organisés en tandem, au moins au niveau de deux des quatre loci identifiés (Roger *et al.*, 2008c).

La famille multigénique codant les *SmPoMuc* semble donc évoluer sous le modèle de « birth and death ». Dans ce modèle la création / pseudogénéisation fréquente de gènes favorise ainsi la création de nouveaux allèles et donc le polymorphisme génétique.

## **1.2. Recombinaisons fréquentes à l'origine des gènes mosaïques**

Comme nous l'avons énoncé ci-dessus, un gène dupliqué peut maintenir sa fonction dans le génome soit par conversion génique soit sous sélection purificatrice. Pour distinguer ces deux processus, il est possible d'analyser les différences nucléotidiques synonymes (ou silencieuses), entre gènes dupliqués, au niveau des régions codantes (Nei, Gojobori, 1986). En effet, ces différences synonymes sont peu sujettes à la sélection et ne seront pas éliminées par la sélection purificatrice. En revanche, la conversion génique, qui permet l'homogénéisation des séquences entre les gènes paralogues, peut conduire à leur élimination. Cette analyse, réalisée sur les séquences codantes de la région 3' des groupes 1 à 4, révèle que les *SmPoMuc* évoluent sous sélection purificatrice (Roger *et al.*, 2008c).

Cependant, cette évolution de type « birth and death » n'explique pas à elle seule l'extraordinaire degré de polymorphisme des *SmPoMuc*. Ainsi, des événements fréquents de recombinaison semblent intervenir au niveau de la région 3' (Roger *et al.*, 2008c). Ces recombinaisons ont lieu non seulement entre gènes, mais également avec les pseudogènes, conduisant ainsi à la génération de nouveaux gènes « mosaïques ». Les pseudogènes bien que non fonctionnels, peuvent servir de réservoir de séquences et participer ainsi au

polymorphisme des *SmPoMuc*. En effet, ils sont libres d'accumuler des mutations et coexistent dans la même zone génomique que des copies fonctionnelles. Les pseudogènes peuvent alors être déplacés en partie ou dans leur totalité par recombinaison et se trouver dans un contexte génomique qui leur permettrait à nouveau d'être transcrits et donc réactivés. Ce type de mécanisme générateur de variabilité est également observé pour les *vsg* chez *T. brucei* (Roth *et al.*, 1989) (cf. i-I.2.2.1.a. *Conditions nécessaires à la variation antigénique des VSG chez T. brucei*) et les gènes du CMH (Doxiadis *et al.*, 2006) (cf. i-I.1.1.2.b. *Polymorphisme du CMH*). Par ailleurs, chez *S. mansoni*, ce phénomène de remodelage fréquent des gènes codant les *SmPoMuc* se traduit par la présence d'événements d'insertion/délétion encadrés de courtes séquences répétées, de solo-LTR ou encore de gènes interrompus par des rétrotransposons. Ces éléments mobiles représentent 55% des 270 Mb du génome de *S. mansoni* (Brindley *et al.*, 2003). Ils sont également présents dans les régions intergéniques des gènes codant les *SmPoMuc*. Ces rétrotransposons pourraient ainsi être à l'origine de recombinaisons, de réarrangements chromosomiques, et de duplications qui ont abouti à l'émergence de la famille multigénique (Vitte, Panaud, 2005) et également à l'origine du polymorphisme génique de la région 3' des gènes codant les *SmPoMuc*.

### 1.3. Recombinaisons ectopiques et diversité combinatoire

Le polymorphisme observé au niveau du domaine VNTR des *SmPoMuc*, décrit précédemment (cf. d-I.1.2. *Les SmPoMuc, polymorphisme et statut de glycosylation*), se traduit par un polymorphisme de taille et par un polymorphisme combinatoire. Ce dernier se caractérise par la présence des deux types d'UR r1 et r2, combinées au sein du même domaine VNTR. Nous avons également montré la présence de cette combinatoire au niveau génomique (Roger *et al.*, 2008c). Toutefois, l'analyse de l'organisation génomique des gènes codant les *SmPoMuc*, et plus particulièrement de deux loci parmi les quatre identifiés par FISH (Roger *et al.*, 2008c), révèle que ces loci sont associés spécifiquement à l'un des deux types d'UR. Nous suggérons ainsi que cette combinatoire, au niveau génomique, pourrait être générée par des mécanismes de recombinaisons ectopiques entre ces deux loci associés respectivement aux exon 2 codant soit les UR r1 soit r2.

L'analyse de la structure des gènes codant les *SmPoMuc* a révélé certaines caractéristiques, permettant d'expliquer cette combinatoire observée au niveau du domaine VNTR et également au niveau génomique. La région 5' des gènes, codant le domaine VNTR, est constituée d'UR génomiques (URg). Ces URg sont composées d'un exon 2

(codant les UR r1 ou r2) encadré par des séquences introniques (Roger *et al.*, 2008c). Ces introns sont caractérisés par la présence d'un microsatellite imparfait ou « polypurine tract » imparfait. Il s'agit d'une séquence, composée essentiellement de l'enchaînement de dinucléotides (GA)<sub>n</sub>, qui est parfois interrompue (imparfait) par des nucléotides C ou T. De plus, les URg sont très conservées et partagent un niveau de similarité très élevé (93%). C'est une caractéristique de régions géniques répétées qui n'évoluent pas de façon indépendante mais de concert (évolution concertée) (Liao, 1999).

L'évolution concertée, comme nous l'avons énoncée précédemment (cf. d-III.1.1. *Evolution des gènes selon le modèle « birth and death »*), conduit à l'homogénéisation de séquences répétées (Elder, Turner, 1995). Ce phénomène a été observé pour différentes familles multigéniques, telles que la famille codant les ARN ribosomiques (Schlotterer, Tautz, 1994) ou encore les snRNA U2 (« small nuclear ARN U2 ») (Liao, Weiner, 1995). Elle a également été rapportée chez des familles multigéniques codant des protéines, telles que les SOWgp de *Coccidioides*, que nous avons décrits précédemment (cf. d-I.2.2. *Les SOWgp chez Coccidioides immitis et C. posadasii*). La région des gènes codant la région riche en proline des SOWgp est composée d'exons et d'introns constituant des URg organisées en tandem, qui sont extrêmement conservées entre elles (Johannesson *et al.*, 2005). A l'heure actuelle, deux mécanismes principaux sont capables d'expliquer ce type d'évolution concertée, les « crossing over » inégaux (Smith, 1976) et la conversion génique (Hillis *et al.*, 1991). Les évènements de « crossing over » inégaux, interviennent préférentiellement au centre de la région génique contenant les URg. Les URg sur lesquelles se produisent les « crossing over » inégaux peuvent s'associer de façon décalée (« mispairing ») notamment au moment des méioses. Ceci conduit à l'augmentation ou à la contraction du nombre d'URg (Elder, Turner, 1995). Ce phénomène pourrait expliquer le polymorphisme de taille observé au niveau du domaine VNTR des *SmPoMuc*. Ces « crossing over » inégaux conduisent au renouvellement des URg (préférentiellement celles du centre de la région) et en principe à la fixation stochastique d'un type d'URg. Ce cas de figure est observé au niveau des gènes codant les SOWgp, où la similarité entre URg décroît au fur et à mesure que l'on s'éloigne du centre et se rapproche des URg terminales (Johannesson *et al.*, 2005). En revanche, ce n'est pas le cas des gènes codant les *SmPoMuc* où toutes les URg sont similaires à 93%, y compris les URg terminales. Nous supposons donc que le mécanisme de conversion génique, qui lui aussi peut expliquer l'évolution concertée des URg, pourrait jouer dans la région 5' des gènes codant les *SmPoMuc*. En effet, ce mécanisme permet l'échange non réciproque d'ADN entre séquences similaires (Schimenti, 1999). De plus, il permet l'homogénéisation de séquences entre chromosomes

homologues mais également non homologues (Abdulkarim, Hughes, 1996; Amstutz *et al.*, 1985). La conversion génique pourrait ainsi expliquer l'évolution concertée des URg de tous les membres de la famille multigénique codant les *SmPoMuc* (Roger *et al.*, 2008c).

Le mécanisme de conversion génique n'est, à l'heure actuelle, pas élucidé. Cependant, il semblerait que des séquences agissant en *cis* puissent avoir une influence sur ce phénomène. En effet, la présence de séquences répétées de type microsatellite pourrait influencer le degré d'identité partagé par les gènes d'une famille. Cela est très bien illustré chez deux familles de gènes, *ErA* et *ErB*, codant le chorion chez *Bombix mori* (Hibner *et al.*, 1991). Ces deux familles de gènes bien que proches sur le chromosome, présentent des degrés de similarité différents, 96% et 63% respectivement. Cette différence semble être due à la présence de séquences répétées de type microsatellite, qui favoriseraient la conversion génique et donc l'homogénéisation, préférentiellement chez *ErA* (Hibner *et al.*, 1991). Une autre étude a été réalisée sur le locus RNU2 codant les snRNA U2 chez l'homme (Liao, Weiner, 1995). Ces gènes sont organisés en tandem au niveau d'un locus et évoluent de manière concertée (Matera *et al.*, 1990; Westin *et al.*, 1984). Cette étude suggère également l'implication de microsatellites imparfaits en tant que site initiateur de la recombinaison et/ou la conversion génique (Pavelitz *et al.*, 1995). En effet, ces microsatellites peuvent adopter des conformations qui diffèrent de la conformation classique de l'ADN en hélice droite (« B conformation ») (Majumdar, Patel, 2002; Wells, 1988). Plus particulièrement, les « polypurine tracts », que l'on retrouve au niveau des gènes codant les *SmPoMuc*, sont susceptibles d'adopter une conformation en triple hélice (« H-DNA conformation ») (Bacolla, Wells, 2004). Ces conformations en triple hélice sont favorables aux événements de recombinaisons (Biet *et al.*, 2003; Zhurkin *et al.*, 1994) et sont susceptibles d'induire des cassures doubles brins de l'ADN (Bacolla *et al.*, 2004; Wang, Vasquez, 2006). Ces cassures doubles brins sont alors prises en charge par le système de réparation de l'ADN (« Double-Strand-Break Repair »). Chez la levure, ce type de réparation peut se solder par un phénomène de conversion génique (Szostak *et al.*, 1983).

Par ailleurs, cette homogénéisation intervient uniquement au niveau des introns des URg. En effet, les exons 2, codant les UR de type r1 et r2, encadrés par ces introns, sont différents entre loci (Roger *et al.*, 2008c). Cette observation a également été faite lors de l'analyse du gène *Flag* codant un type de soie élastique chez l'araignée *Nephilia clavipes*. Les gènes codant les *SmPoMuc* présentent une architecture et un mode d'évolution similaire à ce gène. Il est composé d'une région constituée d'introns et d'exons répétés en tandem. Dans ce cas également, les séquences des introns sont plus conservées entre elles que les exons, qui semblent, quant à eux, évoluer sous sélection purificatrice (Hayashi, Lewis,

2000). Ainsi, les « polypurine tracts », faciliteraient les réarrangements lors des mitoses et méioses. Ils pourraient être à l'origine du phénomène de conversion génique, permettant l'homogénéisation inter-allélique et inter-locus des URg des gènes codant les *SmPoMuc*. Ces « polypurine tracts » pourraient également être à la base de la diversité combinatoire, observée au niveau des ADN complémentaires, en facilitant les échanges d'exons par des phénomènes de recombinaison ectopique entre les gènes (Puchta, 1999).

## **2. Le polymorphisme des *SmPoMuc* au niveau des transcrits**

Comme nous venons de le décrire, la structure des gènes, leur organisation et leur mode d'évolution sont autant d'éléments explicatifs d'une partie du polymorphisme des *SmPoMuc*. En effet, ils permettent d'expliquer le polymorphisme combinatoire et une partie du polymorphisme de taille du domaine VNTR, ainsi qu'une partie du polymorphisme de la partie C-terminale des *SmPomuc* (gènes mosaïques). Néanmoins, nous avons montré que d'autres *processi* intervenant durant la régulation de l'expression des gènes sont à l'origine d'un niveau supplémentaire de polymorphisme. Cette partie va s'attacher à décrire ces mécanismes.

### **2.1. Épissage alternatif et aberrant au niveau de la région 3'**

Outre le polymorphisme de la partie C-terminale lié à la création par recombinaison de régions 3' mosaïques, une autre forme de polymorphisme, au niveau de cette même région, peut conduire à des *SmPoMuc* tronquées. Ce polymorphisme et ces conséquences sur le statut de glycosylation ont été décrits précédemment (cf. d-I.1.2. *Les SmPoMuc, polymorphisme et statut de glycosylation*). Il est dû à des évènements d'épissages alternatifs ou aberrants (Roger *et al.*, 2008c). L'épissage « classique » des transcrits correspondant aux *SmPoMucs* conduit à des messagers dont le précurseur possède une région C-terminale « complète » de 234 acides aminés. En revanche, des évènements d'épissages alternatifs vont conduire à des régions C-terminales plus courtes (par délétion de certains exons). Par ailleurs, un autre type d'épissage apparaît très fréquemment, notamment au niveau de la souche incompatible. Il s'agit d'un épissage aberrant, qui au lieu d'utiliser les sites donneurs et accepteurs présents à la jonction intron-exon, va utiliser certains sites aberrants. Il en résulte une inclusion ou une exclusion d'ARN par rapport à un épissage dit classique. Ainsi, un site aberrant présent dans un exon se traduira par la perte d'une partie de cet exon. En revanche, un site aberrant présent dans un intron se traduira par une inclusion d'une partie de l'intron. Dans tous les cas, ces « erreurs » d'épissage induisent un décalage du cadre de

lecture ouvert, et par conséquent l'apparition d'un codon STOP très rapidement en aval dans le cadre de lecture. Ainsi, les protéines issues de cet épissage seront tronquées d'une grande partie de leur région C-terminale. Ces événements d'épissage aberrant ont été mis en évidence préférentiellement chez la souche incompatible (Roger *et al.*, 2008c) et semblent se répercuter sur le statut de glycosylation (cf. d-I.1.2. *Les SmPoMuc, polymorphisme et statut de glycosylation*).

## 2.2 *Trans*-épissage et « exon repetition » au niveau de la région 5'

Le polymorphisme de la région VNTR des *SmPoMuc* s'exprime au niveau du nombre d'UR (polymorphisme de taille) et du type d'UR (polymorphisme combinatoire). Ce polymorphisme et ses conséquences sur le statut de glycosylation ont été décrits précédemment (cf. d-I.1.2. *Les SmPoMuc, polymorphisme et statut de glycosylation*). D'une part, les recombinaisons ectopiques, au niveau génomique, expliquent le polymorphisme combinatoire (cf. d-II.1.3. *Recombinaisons ectopiques et diversité combinatoire*). D'autre part, les « crossing over » inégaux (Elder, Turner, 1995), responsables de l'expansion et la contraction des URg au niveau génomique, sont à l'origine, en partie, du polymorphisme de taille. Cependant, ce dernier mécanisme génique ne peut expliquer les différences observées entre le nombre d'URg maximum associées à un gène (20) et le nombre maximum d'UR associées au niveau des ADNc (jusqu'à 100). Ainsi, il semblerait que des mécanismes de *trans*-épissage interviennent au niveau des transcrits et participent à la génération de ce polymorphisme de taille.

Le *trans*-épissage « classique », par opposition au *cis*-épissage (ou épissage classique) (Moore, Sharp, 1993), permet de joindre deux exons appartenant à deux pré-messagers distincts (Konarska *et al.*, 1985; Mayer, Floeter-Winter, 2005). Ce processus intervient couramment chez les organismes unicellulaires et *Caenorhabditis elegans*. Chez ces organismes, les gènes sont organisés en clusters et transcrits de façon polycistronique. Les pré-messagers sont maturés en messagers monocistroniques dont l'extrémité 5' est formée par *trans*-épissage d'une séquence leader (SL RNA) transcrite séparément (Blaxter, Liu, 1996; Frantz *et al.*, 2000; Krause, Hirsh, 1987). Ce mécanisme a également été observé chez *S. mansoni*, mais au niveau de peu de messagers et ne semblent pas jouer de rôle majeur (Davis *et al.*, 1995; Rajkovic *et al.*, 1990). De plus, ce type de mécanisme ne contribue pas à une diversité au niveau des protéines.

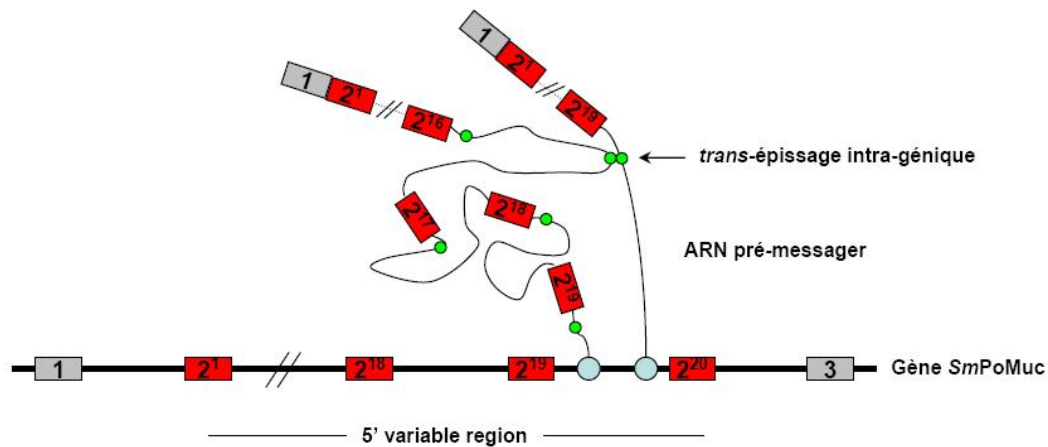
En revanche, chez les mammifères, des phénomènes de *trans*-épissage ont été démontrés *in vitro* (Solnick, 1985). Il existe deux catégories de *trans*-épissage : inter- et



intra-génique (Horiuchi, Aigaki, 2006). Le *trans*-épissage inter-génique génère des ARNm chimères à partir de deux ARN pré-messagers provenant de deux gènes distincts (Li *et al.*, 1999). Le *trans*-épissage intra-génique, quant à lui, se traduit par la présence d'exons répétés au niveau de l'ARNm alors que le gène n'en contient qu'un. Ce phénomène est appelé « exon repetition » et a été décrit pour la première fois chez le rat pour le gène codant la carnitine octanoyltransférase (Caudevilla *et al.*, 1998; Frantz *et al.*, 1999). Ces deux catégories de *trans*-épissages sont des mécanismes qui peuvent expliquer le polymorphisme de taille que l'on observe au niveau du domaine VNTR des *SmPoMuc*. En effet, ils peuvent permettre de relier, au niveau de la région contenant les UR, des ARN pré-messagers différents. De plus, des études récentes ont montré l'intervention de séquences complémentaires dans le processus d'« exon repetition » (Dixon *et al.*, 2007). Ces séquences introniques complémentaires, présentes au niveau de deux ARN pré-messagers, peuvent s'hybrider. Cette hybridation va conduire au rapprochement de sites d'épissages présents sur les deux molécules et ainsi favoriser des phénomènes de *trans*-épissage. Cette hypothèse est fort intéressante car, au niveau des introns constituant la région répétée des gènes codant les *SmPoMuc*, nous retrouvons deux séquences complémentaires de 13 et 12 nucléotides. Ainsi, un tel mécanisme est tout à fait envisageable dans notre cas et pourrait être à l'origine de variants plus long que ceux attendus après épissage classique. Il s'agit de l'hypothèse que nous privilégions à ce jour (Figure 20).

### 2.3. Polymorphisme d'expression individuel spécifique

Nous avons mis en évidence au niveau individuel, une transcription spécifique des gènes pour chaque individu. En effet, chaque larve de *S. mansoni* est capable d'exprimer une combinaison (ou pattern) de variant(s) *SmPoMuc* unique (Roger *et al.*, 2008c). Nous supposons qu'une régulation différentielle génétique ou épigénétique de l'expression selon les individus pourrait intervenir et expliquer ce polymorphisme d'expression. L'intervention de mécanismes épigénétiques est une hypothèse tout à fait séduisante, jamais démontrée chez un parasite métazoaire. En effet, comme nous l'avons décrit dans l'introduction, ce type de régulation est connu chez les parasites protozoaires. Il intervient dans la régulation de l'expression des variants antigéniques de surface chez *T. brucei* (Borst, Ulbert, 2001), *G. lablia* (Kulakova *et al.*, 2006) ou encore *P. falciparum* (Ralph, Scherf, 2005). Nous supposons qu'un contrôle épigénétique chez *S. mansoni* lui permettrait de disposer de nombreux phénotypes (plasticité phénotypique) afin d'assurer la compatibilité avec ses hôtes mollusques.



**Figure 20 : Implication de séquences introniques complémentaires dans le processus d'« exon repetition ».** Représentation schématique du processus de *trans*-épissage intra-génique intervenant au niveau de la région 5' variable des ARN pré-messagers *SmPoMuc*. Les exon 2 des *SmPoMuc* codant les UR sont représentés ici par un carré rouge (l'exposant indique la position de l'exon 2 dans la région répétée). Les séquences introniques complémentaires favorisant le processus d'« exon repetition » sont indiquées par des ronds verts. Ces séquences introniques complémentaires, présentes au niveau de deux ARN pré-messagers (chacun lié à une ARN polymérase, rond bleu), peuvent s'hybrider. Cette hybridation va conduire au rapprochement de sites d'épissages présents sur les deux molécules et ainsi favoriser des phénomènes de *trans*-épissage. Ce mécanisme conduit à des variants contenant un nombre d'UR supérieur à celui présent au niveau génique. Les // indiquent que toutes les UR de la région 5' n'ont pas été représentées.

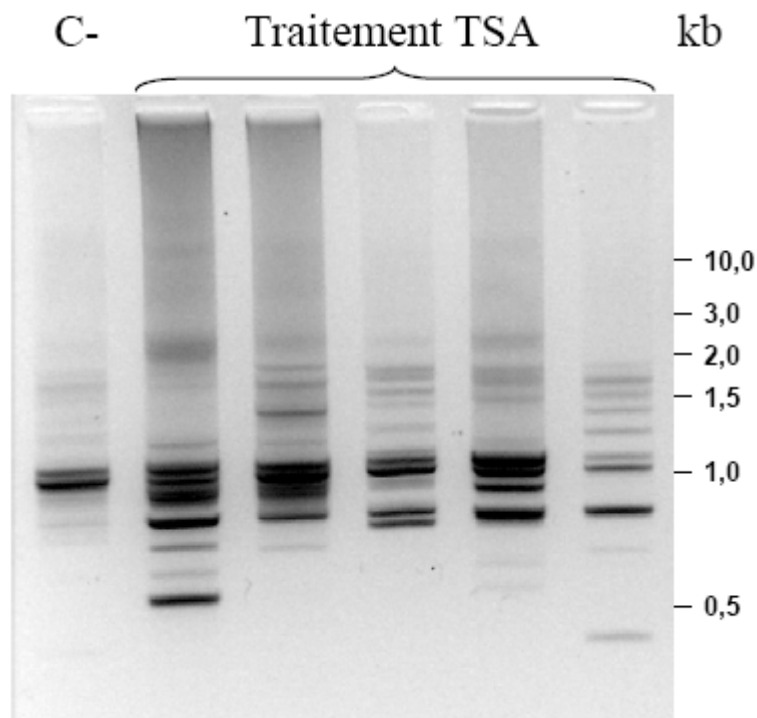
## CONCLUSION ET PERSPECTIVES

Les travaux, développés dans le cadre de cette thèse, ont permis de mettre en évidence, pour la première fois, des variants moléculaires polymorphes (*SmPoMuc*) présents chez un parasite métazoaire (*S. mansoni*) en interaction avec un hôte invertébré (*B. glabrata*). Ces travaux apportent un élément nouveau quant à la compréhension de la dynamique co-évolutive dans les interactions hôte invertébré / parasite. En effet, ces dernières pourraient être soumises aux mêmes contraintes évolutives que les interactions hôte vertébré / parasite. Nos résultats sont donc en accord avec les prédictions des modèles théoriques de co-évolution, tel que celui de la « Reine Rouge », appliqués aux interactions hôte / parasite (Combes, 2000).

Toutefois, afin d'aborder la course aux armements de façon complète dans notre modèle, il est désormais nécessaire de disposer des molécules partenaires des *SmPoMuc* chez *B. glabrata* (opsonines, récepteurs hématocytaires). Dans cette optique, nous avons choisi de développer des expériences de co-immunoprécipitation. Cette approche est développée actuellement au laboratoire dans le cadre de la thèse de Y. Moné. Les régions C-terminales des trois principales *SmPoMuc* (gr.1 à 3) ont d'ores et déjà été produites de façon recombinante et serviront à la production des anticorps anti-*SmPoMuc*. Ces anticorps seront utilisés pour co-immunoprécipiter les protéines plasmatiques et/ou hématocytaires interagissant avec les *SmPoMuc* issus des souches compatible et incompatible. Les protéines du mollusque immunoprécipitées seront caractérisées par une approche protéomique (séparation bi-dimensionnelle) couplée à une analyse en spectrométrie de masse. Ainsi, nous pourrions disposer de molécules « actrices » de la compatibilité dans les deux organismes en interaction. De plus, la fonction des *SmPoMuc* n'est pas élucidée car les analyses fonctionnelles de telles protéines sont pour l'instant irréalisables. Les techniques d'inactivation de gènes (knock-out) n'existent pas pour *S. mansoni* et la technique des ARN interférents est disponible mais n'est pas applicable dans notre cas. En effet, éteindre spécifiquement l'expression de certaines *SmPoMuc*, comme celles qui possèdent la combinatoire dans la souche incompatible, n'est pas réalisable avec les techniques applicables aujourd'hui sur *S. mansoni*. L'identification du partenaire des *SmPoMuc*, par l'approche de co-immunoprécipitation, pourra ainsi nous permettre de progresser et de nous rapprocher au plus près de la fonction potentielle de ces *SmPoMuc*. Dernièrement, de nombreuses études ont montré que la reconnaissance immunitaire chez les invertébrés se faisait via des molécules capables de diversification (Cannon *et al.*, 2002; Dong *et al.*, 2006;

Zhang *et al.*, 2004) et donc que ces phénomènes n'étaient pas exclusivement l'apanage des vertébrés. Ainsi, il a été montré que *B. glabrata* produisait, suite à une infestation, des protéines possédant des domaines de type Immunoglobuline, les FREP. Extrêmement polymorphes, elles sont codées par une famille multigénique. Elles reconnaissent également les résidus glycanes du parasite et précipitent ses antigènes solubles. (Adema *et al.*, 1997; Zhang *et al.*, 2004; Zhang, Loker, 2004). Ainsi, ces molécules sont des candidats de choix en tant que partenaires des *SmPoMuc*. Bien entendu, les études actuellement en cours nous renseigneront sur cet aspect.

Le deuxième axe de recherche que je propose concerne la régulation transcriptionnelle de l'expression des gènes codant les *SmPoMuc*. Les études menées dans le cadre de cette thèse ont montré un polymorphisme d'expression de ces gènes. Il se peut que des mécanismes génétiques et/ou épigénétiques interviennent. Afin d'évaluer la composante génétique, l'analyse comparative des promoteurs des différents loci entre souche compatible et incompatible va être engagée. Si des différences de séquences sont détectées, par cette analyse, la fonctionnalité des promoteurs sera testée en système hétérologue (ANR SCHISTOPHEPIGEN, R. Pierce, Institut Pasteur Lille). La composante épigénétique est évaluée au laboratoire (ANR SCHISTOPHEPIGEN, C. Grunau). D'ores et déjà, des éléments expérimentaux viennent soutenir cette hypothèse. L'utilisation d'une drogue inhibitrice des histones désacétylases (HDAC), induisant des changements épigénétiques, a montré un effet sur la représentativité des transcrits codant les *SmPoMuc* (Figure 21). Ces résultats préliminaires suggèrent que le polymorphisme d'expression des *SmPoMuc* pourrait être, en partie, régulé épigénétiquement et les mécanismes sous-jacents sont actuellement en cours d'investigation.



**Figure 21 : Effet de la trichostatine A (TSA), inhibiteur des histones désacétylases (HDAC), sur le profil d'expression des *SmPoMuc* chez *S. mansoni*.** La modification post-traductionnelle des histones est un porteur d'information épigénétique chez *S. mansoni*. Ces modifications des histones, par acétylation (via les histones acétyl transférases, HAT) ou désacétylation (via les HDAC), sont souvent responsables de phénomènes d'euchromatisation et d'hétérochromatisation, respectivement. Dans le premier cas, les zones du génome sont permises à la transcription, contrairement au second. Ainsi, l'utilisation de la TSA permet d'induire des changements épigénétiques en inhibant les HDAC. Cette drogue a été testée dans notre modèle sur des pools de sporocystes. Des Nested RT- PCR, en utilisant des amorces permettant d'amplifier les ADNc correspondant aux *SmPoMuc* dans leur totalité, ont été réalisées. Les produits PCR obtenus pour les deux conditions (traitée et non traitée) et séparés sur gel d'agarose sont représentés ici. La première piste (C-) correspond au contrôle non traité TSA. Les autres pistes correspondent à la condition traitée TSA. Le marqueur de taille est indiqué à droite en kb. Nous montrons que la TSA modifie fortement le profil d'expression des *SmPoMuc* par rapport au contrôle non traité. Ces premiers résultats suggèrent que le polymorphisme d'expression des *SmPomuc* pourrait donc être régulé épigénétiquement.



## REFERENCES BIBLIOGRAPHIQUES

- [1] ABDULKARIM F., HUGHES D. Homologous recombination between the *tuf* genes of *Salmonella typhimurium*. *J Mol Biol* 1996;260:506-22.
- [2] ABU-SHAKRA M., BUSKILA D., SHOENFELD Y. Molecular mimicry between host and pathogen: examples from parasites and implication. *Immunol Lett* 1999;67:147-52.
- [3] ACOSTA-SERRANO A., ALMEIDA I. C., FREITAS-JUNIOR L. H., YOSHIDA N., SCHENKMAN S. The mucin-like glycoprotein super-family of *Trypanosoma cruzi*: structure and biological roles. *Mol Biochem Parasitol* 2001;114:143-50.
- [4] ADAM R. D. The biology of *Giardia spp.* *Microbiol Rev* 1991;55:706-32.
- [5] ADAMSON M. L., CAIRA J. N. Evolutionary factors influencing the nature of parasite specificity. *Parasitology* 1994;109 Suppl:S85-95.
- [6] ADEMA C. M., LOKER E. S. Specificity and immunology of larval digenean-snail associations. B. Fried and T. K. Graczyk. CRS Press, Boca Raton, Florida 1997.
- [7] ADEMA C. M., HERTEL L. A., MILLER R. D., LOKER E. S. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc Natl Acad Sci U S A* 1997;94:8691-6.
- [8] AGRAWAL A. F., LIVELY C. M. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. *Proc Biol Sci* 2003;270:323-34.
- [9] ALDER M. N., ROGOZIN I. B., IYER L. M., GLAZKO G. V., COOPER M. D., PANCER Z. Diversity and function of adaptive immune receptors in a jawless vertebrate. *Science* 2005;310:1970-3.
- [10] ALMEIDA I. C., FERGUSON M. A., SCHENKMAN S., TRAVASSOS L. R. Lytic anti-alpha-galactosyl antibodies from patients with chronic Chagas' disease recognize novel O-linked oligosaccharides on mucin-like glycosyl-phosphatidylinositol-anchored glycoproteins of *Trypanosoma cruzi*. *Biochem J* 1994a;304 ( Pt 3):793-802.
- [11] ALMEIDA I. C., FERGUSON M. A., SCHENKMAN S., TRAVASSOS L. R. GPI-anchored glycoconjugates from *Trypanosoma cruzi* trypomastigotes are recognized by lytic anti-alpha-galactosyl antibodies isolated from patients with chronic Chagas' disease. *Braz J Med Biol Res* 1994b;27:443-7.
- [12] AMSTUTZ H., MUNZ P., HEYER W. D., LEUPOID U., KOHLI J. Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. *Cell* 1985;40:879-86.
- [13] BACOLLA A., WELLS R. D. Non-B DNA conformations, genomic rearrangements, and human disease. *J Biol Chem* 2004;279:47411-4.

- [14] BACCOLLA A., JAWORSKI A., LARSON J. E., JAKUPCIAK J. P., CHUZHANOVA N., ABEYSINGHE S. S., O'CONNELL C. D., COOPER D. N., WELLS R. D. Breakpoints of gross deletions coincide with non-B DNA conformations. *Proc Natl Acad Sci U S A* 2004;101:14162-7.
- [15] BADLEY J. E., GRIEVE R. B., ROCKEY J. H., GLICKMAN L. T. Immune-mediated adherence of eosinophils to *Toxocara canis* infective larvae: the role of excretory-secretory antigens. *Parasite Immunol* 1987;9:133-43.
- [16] BAKER N. E., MLODZIK M., RUBIN G. M. Spacing differentiation in the developing *Drosophila* eye: a fibrinogen-related lateral inhibitor encoded by scabrous. *Science* 1990;250:1370-7.
- [17] BARBET A. F., KAMPER S. M. The importance of mosaic genes to trypanosome survival. *Parasitol Today* 1993;9:63-6.
- [18] BARBIE V., LEFRANC M. P. The human immunoglobulin kappa variable (IGKV) genes and joining (IGKJ) segments. *Exp Clin Immunogenet* 1998;15:171-83.
- [19] BARRETT M. P., BURCHMORE R. J., STICH A., LAZZARI J. O., FRASCH A. C., CAZZULO J. J., KRISHNA S. The trypanosomiasis. *Lancet* 2003;362:1469-80.
- [20] BARRY J. D., MCCULLOCH R. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv Parasitol* 2001;49:1-70.
- [21] BASCH P. F. An interpretation of snail-trematode infection rates: specificity based on concordance of compatible phenotypes. *Int J Parasitol* 1975;5:449-52.
- [22] BEAVER P. Zoonoses, with particular reference to parasites of veterinary importance. New York: Academic Press 1966.
- [23] BEHNKE J. M., BARNARD C. J., WAKELIN D. Understanding chronic nematode infections: evolutionary considerations, current hypotheses and the way forward. *Int J Parasitol* 1992;22:861-907.
- [24] BERGER S. L. The complex language of chromatin regulation during transcription. *Nature* 2007;447:407-12.



- [25] BERRIMAN M., GHEDIN E., HERTZ-FOWLER C., BLANDIN G., RENAULD H., BARTHOLOMEU D. C., LENNARD N. J., CALER E., HAMLIN N. E., HAAS B., BOHME U., HANNICK L., ASLETT M. A., SHALLOM J., MARCELLO L., HOU L., WICKSTEAD B., ALSMARK U. C., ARROWSMITH C., ATKIN R. J., BARRON A. J., BRINGAUD F., BROOKS K., CARRINGTON M., CHEREVACH I., CHILLINGWORTH T. J., CHURCHER C., CLARK L. N., CORTON C. H., CRONIN A., DAVIES R. M., DOGGETT J., DJIKENG A., FELDBLYUM T., FIELD M. C., FRASER A., GOODHEAD I., HANCE Z., HARPER D., HARRIS B. R., HAUSER H., HOSTETLER J., IVENS A., JAGELS K., JOHNSON D., JOHNSON J., JONES K., KERHORNOU A. X., KOO H., LARKE N., LANDFEAR S., LARKIN C., LEECH V., LINE A., LORD A., MACLEOD A., MOONEY P. J., MOULE S., MARTIN D. M., MORGAN G. W., MUNGALL K., NORBERTCZAK H., ORMOND D., PAI G., PEACOCK C. S., PETERSON J., QUAIL M. A., RABBINOWITSCH E., RAJANDREAM M. A., REITTER C., SALZBERG S. L., SANDERS M., SCHOBEL S., SHARP S., SIMMONDS M., SIMPSON A. J., TALLON L., TURNER C. M., TAIT A., TIVEY A. R., VAN AKEN S., WALKER D., WANLESS D., WANG S., WHITE B., WHITE O., WHITEHEAD S., WOODWARD J., WORTMAN J., ADAMS M. D., EMBLEY T. M., GULL K., ULLU E., BARRY J. D., FAIRLAMB A. H., OPPERDOES F., BARRELL B. G., DONELSON J. E., HALL N., FRASER C. M., MELVILLE S. E., EL-SAYED N. M. The genome of the African trypanosome *Trypanosoma brucei*. *Science* 2005;309:416-22.
- [26] BERRINGTON A. W., TAN Y. C., SRIKHANTA Y., KUIPERS B., VAN DER LEY P., PEAK I. R., JENNINGS M. P. Phase variation in meningococcal lipooligosaccharide biosynthesis genes. *FEMS Immunol Med Microbiol* 2002;34:267-75.
- [27] BIET E., SUN J. S., DUTREIX M. Stimulation of D-loop formation by polypurine/polypyrimidine sequences. *Nucleic Acids Res* 2003;31:1006-12.
- [28] BIGGS B. A., GOOZE L., WYCHERLEY K., WOLLISH W., SOUTHWELL B., LEECH J. H., BROWN G. V. Antigenic variation in *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* 1991;88:9171-4.
- [29] BIGGS B. A., ANDERS R. F., DILLON H. E., DAVERN K. M., MARTIN M., PETERSEN C., BROWN G. V. Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of *Plasmodium falciparum*. *J Immunol* 1992;149:2047-54.
- [30] BIRD A. Perceptions of epigenetics. *Nature* 2007;447:396-8.
- [31] BLACK W. J., SCHWALBE R. S., NACHAMKIN I., CANNON J. G. Characterization of *Neisseria gonorrhoeae* protein II phase variation by use of monoclonal antibodies. *Infect Immun* 1984;45:453-7.
- [32] BLACKARD J. T., COHEN D. E., MAYER K. H. Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. *Clin Infect Dis* 2002;34:1108-14.
- [33] BLAXTER M., LIU L. Nematode spliced leaders--ubiquity, evolution and utility. *Int J Parasitol* 1996;26:1025-33.
- [34] BOGUE M., ROTH D. B. Mechanism of V(D)J recombination. *Curr Opin Immunol* 1996;8:175-80.
- [35] BORST P., ULBERT S. Control of VSG gene expression sites. *Mol Biochem Parasitol* 2001;114:17-27.

- [36] BRADY L. J., CVITKOVITCH D. G., GERIC C. M., ADDISON M. N., JOYCE J. C., CROWLEY P. J., BLEIWEIS A. S. Deletion of the central proline-rich repeat domain results in altered antigenicity and lack of surface expression of the *Streptococcus mutans* P1 adhesin molecule. *Infect Immun* 1998;66:4274-82.
- [37] BRINDLEY P. J., LAHA T., MCMANUS D. P., LOUKAS A. Mobile genetic elements colonizing the genomes of metazoan parasites. *Trends Parasitol* 2003;19:79-87.
- [38] BRUNHAM R. C., PLUMMER F. A., STEPHENS R. S. Bacterial antigenic variation, host immune response, and pathogen-host coevolution. *Infect Immun* 1993;61:2273-6.
- [39] BUSCAGLIA C. A., DI NOIA J. M. *Trypanosoma cruzi* clonal diversity and the epidemiology of Chagas' disease. *Microbes Infect* 2003;5:419-27.
- [40] BUSCAGLIA C. A., CAMPO V. A., DI NOIA J. M., TORRECILHAS A. C., DE MARCHI C. R., FERGUSON M. A., FRASCH A. C., ALMEIDA I. C. The surface coat of the mammal-dwelling infective trypomastigote stage of *Trypanosoma cruzi* is formed by highly diverse immunogenic mucins. *J Biol Chem* 2004;279:15860-9.
- [41] CAMPBELL G., JONES C. S., LOCKYER A. E., HUGHES S., BROWN D., NOBLE L. R., ROLLINSON D. Molecular evidence supports an african affinity of the neotropical freshwater gastropod, *Biomphalaria glabrata*, say 1818, an intermediate host for *Schistosoma mansoni*. *Proc Biol Sci* 2000;267:2351-8.
- [42] CANNON J. P., HAIRE R. N., LITMAN G. W. Identification of diversified genes that contain immunoglobulin-like variable regions in a protochordate. *Nat Immunol* 2002;3:1200-7.
- [43] CANNON J. P., HAIRE R. N., SCHNITKER N., MUELLER M. G., LITMAN G. W. Individual protochordates have unique immune-type receptor repertoires. *Curr Biol* 2004;14:R465-6.
- [44] CARLSTEDT I., SHEEHAN J. K., CORFIELD A. P., GALLAGHER J. T. Mucous glycoproteins: a gel of a problem. *Essays Biochem* 1985;20:40-76.
- [45] CASELLAS R., SHIH T. A., KLEINWIETFELD M., RAKONJAC J., NEMAZEE D., RAJEWSKY K., NUSSENZWEIG M. C. Contribution of receptor editing to the antibody repertoire. *Science* 2001;291:1541-4.
- [46] CAUDEVILLA C., SERRA D., MILIAR A., CODONY C., ASINS G., BACH M., HEGARDT F. G. Natural trans-splicing in carnitine octanoyltransferase pre-mRNAs in rat liver. *Proc Natl Acad Sci U S A* 1998;95:12185-90.
- [47] CHEN B. E., KONDO M., GARNIER A., WATSON F. L., PUETTMANN-HOLGADO R., LAMAR D. R., SCHMUCKER D. The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. *Cell* 2006;125:607-20.
- [48] CHEN Q., FERNANDEZ V., SUNDSTROM A., SCHLICHTERLE M., DATTA S., HAGBLOM P., WAHLGREN M. Developmental selection of var gene expression in *Plasmodium falciparum*. *Nature* 1998;394:392-5.
- [49] CHITSULO L., ENGELS D., MONTRESOR A., SAVIOLI L. The global status of schistosomiasis and its control. *Acta Trop* 2000;77:41-51.

- [50] CHRISTOPHIDES G. K., ZDOBNOV E., BARILLAS-MURY C., BIRNEY E., BLANDIN S., BLASS C., BREY P. T., COLLINS F. H., DANIELLI A., DIMOPOULOS G., HETRU C., HOA N. T., HOFFMANN J. A., KANZOK S. M., LETUNIC I., LEVASHINA E. A., LOUKERIS T. G., LYCETT G., MEISTER S., MICHEL K., MOITA L. F., MULLER H. M., OSTA M. A., PASKEWITZ S. M., REICHHART J. M., RZHETSKY A., TROXLER L., VERNICK K. D., VLACHOU D., VOLZ J., VON MERING C., XU J., ZHENG L., BORK P., KAFATOS F. C. Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 2002;298:159-65.
- [51] COCKELL M., GASSER S. M. Nuclear compartments and gene regulation. *Curr Opin Genet Dev* 1999;9:199-205.
- [52] COLE G. T., SESHAN K. R., FRANCO M., BUKOWNIK E., SUN S. H., HEARN V. M. Isolation and morphology of an immunoreactive outer wall fraction produced by spherules of *Coccidioides immitis*. *Infect Immun* 1988a;56:2686-94.
- [53] COLE G. T., KIRKLAND T. N., FRANCO M., ZHU S., YUAN L., SUN S. H., HEARN V. M. Immunoreactivity of a surface wall fraction produced by spherules of *Coccidioides immitis*. *Infect Immun* 1988b;56:2695-701.
- [54] COLE G. T., HUNG C. Y. The parasitic cell wall of *Coccidioides immitis*. *Med Mycol* 2001;39 Suppl 1:31-40.
- [55] COMBES C. Interactions durables, Ecologie et évolution du parasitisme. Editions Masson 1995.
- [56] COMBES C. [Selective pressure in host-parasite systems]. *J Soc Biol* 2000;194:19-23.
- [57] CONNELL T. D., BLACK W. J., KAWULA T. H., BARRITT D. S., DEMPSEY J. A., KVERNELAND K., JR., STEPHENSON A., SCHEPART B. S., MURPHY G. L., CANNON J. G. Recombination among protein II genes of *Neisseria gonorrhoeae* generates new coding sequences and increases structural variability in the protein II family. *Mol Microbiol* 1988;2:227-36.
- [58] CONSORTIUM M. S. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. *Nature* 1999;401:921-3.
- [59] CONWAY C., MCCULLOCH R., GINGER M. L., ROBINSON N. P., BROWITT A., BARRY J. D. Ku is important for telomere maintenance, but not for differential expression of telomeric VSG genes, in African trypanosomes. *J Biol Chem* 2002;277:21269-77.
- [60] COOPER J. P., WATANABE Y., NURSE P. Fission yeast Taz1 protein is required for meiotic telomere clustering and recombination. *Nature* 1998;392:828-31.
- [61] COUSTAU C., CHEVILLON C., FFRENCH-CONSTANT R. Resistance to xenobiotics and parasites: can we count the cost? *Trends Ecol Evol* 2000;15:378-383.
- [62] COX R. A., MAGEE D. M. Protective immunity in coccidioidomycosis. *Res Immunol* 1998;149:417-28; discussion 506-7.
- [63] CRAYTON M. E., 3RD, POWELL B. C., VISION T. J., GIDDINGS M. C. Tracking the evolution of alternatively spliced exons within the Dscam family. *BMC Evol Biol* 2006;6:16.

- [64] CRISS A. K., KLINE K. A., SEIFERT H. S. The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. *Mol Microbiol* 2005;58:510-9.
- [65] DAI Q., RESTREPO B. I., PORCELLA S. F., RAFFEL S. J., SCHWAN T. G., BARBOUR A. G. Antigenic variation by *Borrelia hermsii* occurs through recombination between extragenic repetitive elements on linear plasmids. *Mol Microbiol* 2006;60:1329-43.
- [66] DAMIAN R. T. Molecular mimicry revisited. *Parasitol Today* 1987;3:263-6.
- [67] DAMIAN R. T. Parasite immune evasion and exploitation: reflections and projections. *Parasitology* 1997;115 Suppl:S169-75.
- [68] DAVIES C. M., WEBSTER J. P., WOOLHOUSE M. E. Trade-offs in the evolution of virulence in an indirectly transmitted macroparasite. *Proc Biol Sci* 2001;268:251-7.
- [69] DAVIS R. E., HARDWICK C., TAVERNIER P., HODGSON S., SINGH H. RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, *Schistosoma mansoni*. *J Biol Chem* 1995;270:21813-9.
- [70] DE GREGORIO E., SPELLMAN P. T., TZOU P., RUBIN G. M., LEMAITRE B. The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. *Embo J* 2002;21:2568-79.
- [71] DE LANGE T., BORST P. Genomic environment of the expression-linked extra copies of genes for surface antigens of *Trypanosoma brucei* resembles the end of a chromosome. *Nature* 1982;299:451-3.
- [72] DE LANGE T. Ending up with the right partner. *Nature* 1998;392:753-4.
- [73] DEITSCH K. W., DEL PINAL A., WELLEMS T. E. Intra-cluster recombination and var transcription switches in the antigenic variation of *Plasmodium falciparum*. *Mol Biochem Parasitol* 1999;101:107-16.
- [74] DESTOUMIEUX D., MUNOZ M., BULET P., BACHERE E. Penaeidins, a family of antimicrobial peptides from penaeid shrimp (Crustacea, Decapoda). *Cell Mol Life Sci* 2000;57:1260-71.
- [75] DI NOIA J. M., SANCHEZ D. O., FRASCH A. C. The protozoan *Trypanosoma cruzi* has a family of genes resembling the mucin genes of mammalian cells. *J Biol Chem* 1995;270:24146-9.
- [76] DI NOIA J. M., D'ORSO I., ASLUND L., SANCHEZ D. O., FRASCH A. C. The *Trypanosoma cruzi* mucin family is transcribed from hundreds of genes having hypervariable regions. *J Biol Chem* 1998;273:10843-50.
- [77] DI NOIA J. M., D'ORSO I., SANCHEZ D. O., FRASCH A. C. AU-rich elements in the 3'-untranslated region of a new mucin-type gene family of *Trypanosoma cruzi* confers mRNA instability and modulates translation efficiency. *J Biol Chem* 2000;275:10218-27.
- [78] DI NOIA J. M., NEUBERGER M. S. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* 2007;76:1-22.

- [79] DIMOPOULOS G., CASAVANT T. L., CHANG S., SCHEETZ T., ROBERTS C., DONOHUE M., SCHULTZ J., BENES V., BORK P., ANSORGE W., SOARES M. B., KAFATOS F. C. *Anopheles gambiae* pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines. *Proc Natl Acad Sci U S A* 2000;97:6619-24.
- [80] DIXON R. J., EPERON I. C., SAMANI N. J. Complementary intron sequence motifs associated with human exon repetition: a role for intragenic, inter-transcript interactions in gene expression. *Bioinformatics* 2007;23:150-5.
- [81] DONELSON J. E. Antigenic variation and the African trypanosome genome. *Acta Trop* 2003;85:391-404.
- [82] DONG Y., TAYLOR H. E., DIMOPOULOS G. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol* 2006;4:e229.
- [83] DOXIADIS G. G., VAN DER WIEL M. K., BROK H. P., DE GROOT N. G., OTTING N., T HART B. A., VAN ROOD J. J., BONTROP R. E. Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a New World primate species. *Proc Natl Acad Sci U S A* 2006;103:5864-8.
- [84] DRUTZ D. J., HUPPERT M. Coccidioidomycosis: factors affecting the host-parasite interaction. *J Infect Dis* 1983;147:372-90.
- [85] DU PASQUIER L. The phylogenetic origin of antigen-specific receptors. *Curr Top Microbiol Immunol* 2000;248:160-85.
- [86] DUNHAM I., SARGENT C. A., TROWSDALE J., CAMPBELL R. D. Molecular mapping of the human major histocompatibility complex by pulsed-field gel electrophoresis. *Proc Natl Acad Sci U S A* 1987;84:7237-41.
- [87] DURASINGH M. T., VOSS T. S., MARTY A. J., DUFFY M. F., GOOD R. T., THOMPSON J. K., FREITAS-JUNIOR L. H., SCHERF A., CRABB B. S., COWMAN A. F. Heterochromatin silencing and locus repositioning linked to regulation of virulence genes in *Plasmodium falciparum*. *Cell* 2005;121:13-24.
- [88] ELDER J. F., JR., TURNER B. J. Concerted evolution of repetitive DNA sequences in eukaryotes. *Q Rev Biol* 1995;70:297-320.
- [89] ELENA S. F., SANJUAN R. Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J Virol* 2005;79:11555-8.
- [90] FERGUSON M. A., BRIMACOMBE J. S., COTTAZ S., FIELD R. A., GUTHER L. S., HOMANS S. W., MCCONVILLE M. J., MEHLERT A., MILNE K. G., RALTON J. E., ET AL. Glycosylphosphatidylinositol molecules of the parasite and the host. *Parasitology* 1994;108 Suppl:S45-54.
- [91] FINLAY B. B., MCFADDEN G. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell* 2006;124:767-82.

- [92] FISCHER K., HORROCKS P., PREUSS M., WIESNER J., WUNSCH S., CAMARGO A. A., LANZER M. Expression of var genes located within polymorphic subtelomeric domains of *Plasmodium falciparum* chromosomes. *Mol Cell Biol* 1997;17:3679-86.
- [93] FLAJNIK M. F., KASAHARA M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. *Immunity* 2001;15:351-62.
- [94] FOLCH G., LEFRANC M. P. The human T cell receptor beta variable (TRBV) genes. *Exp Clin Immunogenet* 2000;17:42-54.
- [95] FOOTE J. Immunology. Isomeric antibodies. *Science* 2003;299:1327-8.
- [96] FRANK M., DEITSCH K. Activation, silencing and mutually exclusive expression within the var gene family of *Plasmodium falciparum*. *Int J Parasitol* 2006;36:975-85.
- [97] FRANK S. A. Specificity versus detectable polymorphism in host-parasite genetics. *Proc Biol Sci* 1993;254:191-7.
- [98] FRANTZ C., EBEL C., PAULUS F., IMBAULT P. Characterization of trans-splicing in Euglenoids. *Curr Genet* 2000;37:349-55.
- [99] FRANTZ S. A., THIARA A. S., LODWICK D., NG L. L., EPERON I. C., SAMANI N. J. Exon repetition in mRNA. *Proc Natl Acad Sci U S A* 1999;96:5400-5.
- [100] FREITAS-JUNIOR L. H., BOTTIUS E., PIRRI L. A., DEITSCH K. W., SCHEIDIG C., GUINET F., NEHRBASS U., WELLEMS T. E., SCHERF A. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. *Nature* 2000;407:1018-22.
- [101] FREITAS-JUNIOR L. H., HERNANDEZ-RIVAS R., RALPH S. A., MONTIEL-CONDADO D., RUVALCABA-SALAZAR O. K., ROJAS-MEZA A. P., MANCIO-SILVA L., LEAL-SILVESTRE R. J., GONTIJO A. M., SHORTE S., SCHERF A. Telomeric heterochromatin propagation and histone acetylation control mutually exclusive expression of antigenic variation genes in malaria parasites. *Cell* 2005;121:25-36.
- [102] GALGANI J. N. Coccidioidomycosis: a regional disease of national importance. Rethinking approaches for control. *Ann Intern Med* 1999;130:293-300.
- [103] GAO F., BAILES E., ROBERTSON D. L., CHEN Y., RODENBURG C. M., MICHAEL S. F., CUMMINS L. B., ARTHUR L. O., PEETERS M., SHAW G. M., SHARP P. M., HAHN B. H. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. *Nature* 1999;397:436-41.
- [104] GARDNER M. J., HALL N., FUNG E., WHITE O., BERRIMAN M., HYMAN R. W., CARLTON J. M., PAIN A., NELSON K. E., BOWMAN S., PAULSEN I. T., JAMES K., EISEN J. A., RUTHERFORD K., SALZBERG S. L., CRAIG A., KYES S., CHAN M. S., NENE V., SHALLOM S. J., SUH B., PETERSON J., ANGIUOLI S., PERTEA M., ALLEN J., SELENGUT J., HAFT D., MATHER M. W., VAIDYA A. B., MARTIN D. M., FAIRLAMB A. H., FRAUNHOLZ M. J., ROOS D. S., RALPH S. A., MCFADDEN G. I., CUMMINGS L. M., SUBRAMANIAN G. M., MUNGALL C., VENTER J. C., CARUCCI D. J., HOFFMAN S. L., NEWBOLD C., DAVIS R. W., FRASER C. M., BARRELL B. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* 2002;419:498-511.

- [105] GAUDIERI S., DAWKINS R. L., HABARA K., KULSKI J. K., GOJOBORI T. SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. *Genome Res* 2000;10:1579-86.
- [106] GELLERT M. V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 2002;71:101-32.
- [107] GEMS D., MAIZELS R. M. An abundantly expressed mucin-like protein from *Toxocara canis* infective larvae: the precursor of the larval surface coat glycoproteins. *Proc Natl Acad Sci U S A* 1996;93:1665-70.
- [108] GENDLER S. J., SPICER A. P. Epithelial mucin genes. *Annu Rev Physiol* 1995;57:607-34.
- [109] GIBBS C. P., REIMANN B. Y., SCHULTZ E., KAUFMANN A., HAAS R., MEYER T. F. Reassortment of pilin genes in *Neisseria gonorrhoeae* occurs by two distinct mechanisms. *Nature* 1989;338:651-2.
- [110] GILBERTSON P., WOTHERSPOON J., RAISON R. L. Evolutionary development of lymphocyte heterogeneity: leucocyte subpopulations in the Pacific hagfish. *Dev Comp Immunol* 1986;10:1-10.
- [111] GILFILLAN S., DIERICH A., LEMEUR M., BENOIST C., MATHIS D. Mice lacking TdT: mature animals with an immature lymphocyte repertoire. *Science* 1993;261:1175-8.
- [112] GILLESPIE S. H. Human toxocariasis. *J Appl Bacteriol* 1987;63:473-9.
- [113] GILLESPIE S. H., DINNING W. J., VOLLER A., CROWCROFT N. S. The spectrum of ocular toxocariasis. *Eye* 1993;7 ( Pt 3):415-8.
- [114] GLICKMAN L. T., SCHANTZ P. M. Epidemiology and pathogenesis of zoonotic toxocariasis. *Epidemiol Rev* 1981;3:230-50.
- [115] GOKUDAN S., MUTA T., TSUDA R., KOORI K., KAWAHARA T., SEKI N., MIZUNOE Y., WAI S. N., IWANAGA S., KAWABATA S. Horseshoe crab acetyl group-recognizing lectins involved in innate immunity are structurally related to fibrinogen. *Proc Natl Acad Sci U S A* 1999;96:10086-91.
- [116] GOODMAN M. F. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* 2002;71:17-50.
- [117] GOWER C. M., WEBSTER J. P. Intraspecific competition and the evolution of virulence in a parasitic trematode. *Evolution Int J Org Evolution* 2005;59:544-53.
- [118] GRAVELEY B. R., KAUR A., GUNNING D., ZIPURSKY S. L., ROWEN L., CLEMENS J. C. The organization and evolution of the dipteran and hymenopteran Down syndrome cell adhesion molecule (Dscam) genes. *Rna* 2004;10:1499-506.
- [119] GUILLOU F., ROGER E., MONE Y., ROGNON A., GRUNAU C., THERON A., MITTA G., COUSTAU C., GOURBAL B. E. Excretory-secretory proteome of larval *Schistosoma mansoni* and *Echinostoma caproni*, two parasites of *Biomphalaria glabrata*. *Mol Biochem Parasitol* 2007;155:45-56.

- [120] GUM J. R., JR. Human mucin glycoproteins: varied structures predict diverse properties and specific functions. *Biochem Soc Trans* 1995;23:795-9.
- [121] HAAS R., MEYER T. F. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. *Cell* 1986;44:107-15.
- [122] HAAS R., SCHWARZ H., MEYER T. F. Release of soluble pilin antigen coupled with gene conversion in *Neisseria gonorrhoeae*. *Proc Natl Acad Sci U S A* 1987;84:9079-83.
- [123] HAAS W., GUI M., HABERL B., STROBEL M. *Miracidia* of *Schistosoma japonicum*: approach and attachment to the snail host. *J Parasitol* 1991;77:509-13.
- [124] HAGBLOM P., SEGAL E., BILLYARD E., SO M. Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae*. *Nature* 1985;315:156-8.
- [125] HAGEN M., FILOSA M. F., YOUSON J. H. The immune response in adult sea lamprey (*Petromyzon marinus* L.): the effect of temperature. *Comp Biochem Physiol A* 1985;82:207-10.
- [126] HAMPSON A. W., MACKENZIE J. S. The influenza viruses. *Med J Aust* 2006;185:S39-43.
- [127] HAMRICK T. S., DEMPSEY J. A., COHEN M. S., CANNON J. G. Antigenic variation of gonococcal pilin expression in vivo: analysis of the strain FA1090 pilin repertoire and identification of the pilS gene copies recombining with pilE during experimental human infection. *Microbiology* 2001;147:839-49.
- [128] HANISCH F. G. O-glycosylation of the mucin type. *Biol Chem* 2001;382:143-9.
- [129] HAYASHI C. Y., LEWIS R. V. Molecular architecture and evolution of a modular spider silk protein gene. *Science* 2000;287:1477-9.
- [130] HERNANDEZ-RIVAS R., MATTEI D., STERKERS Y., PETERSON D. S., WELLEMS T. E., SCHERF A. Expressed var genes are found in *Plasmodium falciparum* subtelomeric regions. *Mol Cell Biol* 1997;17:604-11.
- [131] HERTEL L. A., STRICKER S. A., MONROY F. P., WILSON W. D., LOKER E. S. *Biomphalaria glabrata* hemolymph lectins: binding to bacteria, mammalian erythrocytes, and to sporocysts and rediae of *Echinostoma paraensei*. *J Invertebr Pathol* 1994;64:52-61.
- [132] HIBNER B. L., BURKE W. D., EICKBUSH T. H. Sequence identity in an early chorion multigene family is the result of localized gene conversion. *Genetics* 1991;128:595-606.
- [133] HICKS S. J., THEODOROPOULOS G., CARRINGTON S. D., CORFIELD A. P. The role of mucins in host-parasite interactions. Part I- protozoan parasites. *Parasitol Today* 2000;16:476-81.
- [134] HILDEMANN W. H. Transplantation immunity in fishes: Agnatha, Chondrichthyes and Osteichthyes. *Transplant Proc* 1970;2:253-9.
- [135] HILKENS J., LIGTENBERG M. J., VOS H. L., LITVINOV S. V. Cell membrane-associated mucins and their adhesion-modulating property. *Trends Biochem Sci* 1992;17:359-63.



- [136] HILLIS D. M., MORITZ C., PORTER C. A., BAKER R. J. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science* 1991;251:308-10.
- [137] HOEK R. M., SMIT A. B., FRINGS H., VINK J. M., DE JONG-BRINK M., GERAERTS W. P. A new Ig-superfamily member, molluscan defence molecule (MDM) from *Lymnaea stagnalis*, is down-regulated during parasitosis. *Eur J Immunol* 1996;26:939-44.
- [138] HOFFMANN J. A. The immune response of *Drosophila*. *Nature* 2003;426:33-8.
- [139] HOLMES E. C. The phylogeography of human viruses. *Mol Ecol* 2004;13:745-56.
- [140] HORIUCHI T., AIGAKI T. Alternative trans-splicing: a novel mode of pre-mRNA processing. *Biol Cell* 2006;98:135-40.
- [141] HORN D., CROSS G. A. A developmentally regulated position effect at a telomeric locus in *Trypanosoma brucei*. *Cell* 1995;83:555-61.
- [142] HORN D., CROSS G. A. Position-dependent and promoter-specific regulation of gene expression in *Trypanosoma brucei*. *Embo J* 1997;16:7422-31.
- [143] HORROCKS P., PINCHES R., KYES S., KRIEK N., LEE S., CHRISTODOULOU Z., NEWBOLD C. I. Effect of var gene disruption on switching in *Plasmodium falciparum*. *Mol Microbiol* 2002;45:1131-41.
- [144] HORTON R., WILMING L., RAND V., LOVERING R. C., BRUFORD E. A., KHODIYAR V. K., LUSH M. J., POVEY S., TALBOT C. C., JR., WRIGHT M. W., WAIN H. M., TROWSDALE J., ZIEGLER A., BECK S. Gene map of the extended human MHC. *Nat Rev Genet* 2004;5:889-99.
- [145] HOTEZ P. J., BRINDLEY P. J., BETHONY J. M., KING C. H., PEARCE E. J., JACOBSON J. Helminth infections: the great neglected tropical diseases. *J Clin Invest* 2008;118:1311-21.
- [146] HUGHES A. L., NEI M. Evolutionary relationships of class II major-histocompatibility-complex genes in mammals. *Mol Biol Evol* 1990;7:491-514.
- [147] HUGHES K., WAND M., FOULSTON L., YOUNG R., HARLEY K., TERRY S., ERSFELD K., RUDENKO G. A novel ISWI is involved in VSG expression site downregulation in African trypanosomes. *Embo J* 2007;26:2400-10.
- [148] HUNG C. Y., AMPEL N. M., CHRISTIAN L., SESHAN K. R., COLE G. T. A major cell surface antigen of *Coccidioides immitis* which elicits both humoral and cellular immune responses. *Infect Immun* 2000;68:584-93.
- [149] HUNG C. Y., YU J. J., SESHAN K. R., REICHARD U., COLE G. T. A parasitic phase-specific adhesin of *Coccidioides immitis* contributes to the virulence of this respiratory Fungal pathogen. *Infect Immun* 2002;70:3443-56.
- [150] HUNG C. Y., SESHAN K. R., YU J. J., SCHALLER R., XUE J., BASRUR V., GARDNER M. J., COLE G. T. A metalloproteinase of *Coccidioides posadasii* contributes to evasion of host detection. *Infect Immun* 2005;73:6689-703.
- [151] HUNG C. Y., XUE J., COLE G. T. Virulence mechanisms of coccidioides. *Ann N Y Acad Sci* 2007;1111:225-35.

- [152] JACKSON K. J., GAETA B., SEWELL W., COLLINS A. M. Exonuclease activity and P nucleotide addition in the generation of the expressed immunoglobulin repertoire. *BMC Immunol* 2004;5:19.
- [153] JAMES L. C., ROVERSI P., TAWFIK D. S. Antibody multispecificity mediated by conformational diversity. *Science* 2003;299:1362-7.
- [154] JENUWEIN T., ALLIS C. D. Translating the histone code. *Science* 2001;293:1074-80.
- [155] JERSE A. E., COHEN M. S., DROWN P. M., WHICKER L. G., ISBEY S. F., SEIFERT H. S., CANNON J. G. Multiple gonococcal opacity proteins are expressed during experimental urethral infection in the male. *J Exp Med* 1994;179:911-20.
- [156] JOHANNESSON H., TOWNSEND J. P., HUNG C. Y., COLE G. T., TAYLOR J. W. Concerted evolution in the repeats of an immunomodulating cell surface protein, SOWgp, of the human pathogenic fungi *Coccidioides immitis* and *C. posadasii*. *Genetics* 2005;171:109-17.
- [157] KAY B. K., WILLIAMSON M. P., SUDOL M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *Faseb J* 2000;14:231-41.
- [158] KELLEY J., WALTER L., TROWSDALE J. Comparative genomics of major histocompatibility complexes. *Immunogenetics* 2005;56:683-95.
- [159] KELLOGG D. S., JR., PEACOCK W. L., JR., DEACON W. E., BROWN L., PIRKLE D. I. *Neisseria Gonorrhoeae*. I. Virulence Genetically Linked to Clonal Variation. *J Bacteriol* 1963;85:1274-9.
- [160] KENTER A. L. Class-switch recombination: after the dawn of AID. *Curr Opin Immunol* 2003;15:190-8.
- [161] KLINE K. A., SECHMAN E. V., SKAAR E. P., SEIFERT H. S. Recombination, repair and replication in the pathogenic *Neisseriae*: the 3 R's of molecular genetics of two human-specific bacterial pathogens. *Mol Microbiol* 2003;50:3-13.
- [162] KONARSKA M. M., PADGETT R. A., SHARP P. A. Trans splicing of mRNA precursors in vitro. *Cell* 1985;42:165-71.
- [163] KOUSKOFF V., NEMAZEE D. Role of receptor editing and revision in shaping the B and T lymphocyte repertoire. *Life Sci* 2001;69:1105-13.
- [164] KOUZARIDES T. Chromatin modifications and their function. *Cell* 2007;128:693-705.
- [165] KRAEMER S. M., SMITH J. D. Evidence for the importance of genetic structuring to the structural and functional specialization of the *Plasmodium falciparum* var gene family. *Mol Microbiol* 2003;50:1527-38.
- [166] KRAEMER S. M., KYES S. A., AGGARWAL G., SPRINGER A. L., NELSON S. O., CHRISTODOULOU Z., SMITH L. M., WANG W., LEVIN E., NEWBOLD C. I., MYLER P. J., SMITH J. D. Patterns of gene recombination shape var gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. *BMC Genomics* 2007;8:45.

- [167] KRAUSE M., HIRSH D. A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 1987;49:753-61.
- [168] KULAKOVA L., SINGER S. M., CONRAD J., NASH T. E. Epigenetic mechanisms are involved in the control of *Giardia lamblia* antigenic variation. *Mol Microbiol* 2006;61:1533-42.
- [169] KURACHI S., SONG Z., TAKAGAKI M., YANG Q., WINTER H. C., KURACHI K., GOLDSTEIN I. J. Sialic-acid-binding lectin from the slug *Limax flavus*--cloning, expression of the polypeptide, and tissue localization. *Eur J Biochem* 1998;254:217-22.
- [170] KURTZ J., FRANZ K. Innate defence: evidence for memory in invertebrate immunity. *Nature* 2003;425:37-8.
- [171] KWONG P. D., DOYLE M. L., CASPER D. J., CICALA C., LEAVITT S. A., MAJEED S., STEENBEKE T. D., VENTURI M., CHAIKEN I., FUNG M., KATINGER H., PARREN P. W., ROBINSON J., VAN RYK D., WANG L., BURTON D. R., FREIRE E., WYATT R., SODROSKI J., HENDRICKSON W. A., ARTHOS J. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature* 2002;420:678-82.
- [172] KYES S. A., CHRISTODOULOU Z., RAZA A., HORROCKS P., PINCHES R., ROWE J. A., NEWBOLD C. I. A well-conserved *Plasmodium falciparum* var gene shows an unusual stage-specific transcript pattern. *Mol Microbiol* 2003;48:1339-48.
- [173] KYES S. A., KRAEMER S. M., SMITH J. D. Antigenic variation in *Plasmodium falciparum*: gene organization and regulation of the var multigene family. *Eukaryot Cell* 2007;6:1511-20.
- [174] LAMBELE M., LABROSSE B., ROCH E., MOREAU A., VERRIER B., BARIN F., ROINGEARD P., MAMMANO F., BRAND D. Impact of natural polymorphism within the gp41 cytoplasmic tail of human immunodeficiency virus type 1 on the intracellular distribution of envelope glycoproteins and viral assembly. *J Virol* 2007;81:125-40.
- [175] LANDEIRA D., NAVARRO M. Nuclear repositioning of the VSG promoter during developmental silencing in *Trypanosoma brucei*. *J Cell Biol* 2007;176:133-9.
- [176] LEECH J. H., BARNWELL J. W., MILLER L. H., HOWARD R. J. Identification of a strain-specific malarial antigen exposed on the surface of *Plasmodium falciparum*-infected erythrocytes. *J Exp Med* 1984;159:1567-75.
- [177] LEFRANC M. P. Nomenclature of the human T cell receptor genes. *Curr Protoc Immunol* 2001;Appendix 1:Appendix 10.
- [178] LEONARD P. M., ADEMA C. M., ZHANG S. M., LOKER E. S. Structure of two FREP genes that combine IgSF and fibrinogen domains, with comments on diversity of the FREP gene family in the snail *Biomphalaria glabrata*. *Gene* 2001;269:155-65.
- [179] LEVINSON G., GUTMAN G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 1987;4:203-21.

- [180] LI B. L., LI X. L., DUAN Z. J., LEE O., LIN S., MA Z. M., CHANG C. C., YANG X. Y., PARK J. P., MOHANDAS T. K., NOLL W., CHAN L., CHANG T. Y. Human acyl-CoA:cholesterol acyltransferase-1 (ACAT-1) gene organization and evidence that the 4.3-kilobase ACAT-1 mRNA is produced from two different chromosomes. *J Biol Chem* 1999;274:11060-71.
- [181] LI Z., WOO C. J., IGLESIAS-USSEL M. D., RONAI D., SCHARFF M. D. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev* 2004;18:1-11.
- [182] LIAO D., WEINER A. M. Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the RNU2 locus) does not prevent rapid diversification of the (CT)<sub>n</sub>(GA)<sub>n</sub> microsatellite embedded within the U2 repeat unit. *Genomics* 1995;30:583-93.
- [183] LIAO D. Concerted evolution: molecular mechanism and biological implications. *Am J Hum Genet* 1999;64:24-30.
- [184] LITMAN G. W., FINSTAD F. J., HOWELL J., POLLARA B. W., GOD R. A. The evolution of the immune response. 3. Structural studies of the lamprey immunoglobulin. *J Immunol* 1970;105:1278-85.
- [185] LITMAN G. W., DISHAW L. J., CANNON J. P., HAIRE R. N., RAST J. P. Alternative mechanisms of immune receptor diversity. *Curr Opin Immunol* 2007;19:526-34.
- [186] LITTLE T. J., O'CONNOR B., COLEGRAVE N., WATT K., READ A. F. Maternal transfer of strain-specific immunity in an invertebrate. *Curr Biol* 2003;13:489-92.
- [187] LITTLE T. J., HULTMARK D., READ A. F. Invertebrate immunity and the limits of mechanistic immunology. *Nat Immunol* 2005;6:651-4.
- [188] LODES M. J., YOSHINO T. P. Characterization of excretory-secretory proteins synthesized in vitro by *Schistosoma mansoni* primary sporocysts. *J Parasitol* 1989;75:853-62.
- [189] LOKER E. S., ADEMA C. M., ZHANG S. M., KEPLER T. B. Invertebrate immune systems-not homogeneous, not simple, not well understood. *Immunol Rev* 2004;198:10-24.
- [190] LOMVARDAS S., BARNEA G., PISAPIA D. J., MENDELSON M., KIRKLAND J., AXEL R. Interchromosomal interactions and olfactory receptor choice. *Cell* 2006;126:403-13.
- [191] LOPEZ-RUBIO J. J., GONTIJO A. M., NUNES M. C., ISSAR N., HERNANDEZ RIVAS R., SCHERF A. 5' flanking region of var genes nucleate histone modification patterns linked to phenotypic inheritance of virulence traits in malaria parasites. *Mol Microbiol* 2007;66:1296-305.
- [192] LOUKAS A., HINTZ M., LINDER D., MULLIN N. P., PARKINSON J., TETTEH K. K., MAIZELS R. M. A family of secreted mucins from the parasitic nematode *Toxocara canis* bears diverse mucin domains but shares similar flanking six-cysteine repeat motifs. *J Biol Chem* 2000;275:39600-7.
- [193] LYNCH M., CONERY J. S. The evolutionary fate and consequences of duplicate genes. *Science* 2000;290:1151-5.

- [194] MAGEE D. M., COX R. A. Roles of gamma interferon and interleukin-4 in genetically determined resistance to *Coccidioides immitis*. *Infect Immun* 1995;63:3514-9.
- [195] MAGEE D. M., COX R. A. Interleukin-12 regulation of host defenses against *Coccidioides immitis*. *Infect Immun* 1996;64:3609-13.
- [196] MAIZELS N. Immunoglobulin gene diversification. *Annu Rev Genet* 2005;39:23-46.
- [197] MAIZELS R. M., BUNDY D. A., SELKIRK M. E., SMITH D. F., ANDERSON R. M. Immunological modulation and evasion by helminth parasites in human populations. *Nature* 1993;365:797-805.
- [198] MAIZELS R. M., BALIC A., GOMEZ-ESCOBAR N., NAIR M., TAYLOR M. D., ALLEN J. E. Helminth parasites--masters of regulation. *Immunol Rev* 2004;201:89-116.
- [199] MAJUMDAR A., PATEL D. J. Identifying hydrogen bond alignments in multistranded DNA architectures by NMR. *Acc Chem Res* 2002;35:1-11.
- [200] MARIN M. S., PRIETO M., MARTIN J. M., CASAIS R., BOGA J. A., PARRA F. Identification and expression of a *Fasciola hepatica* gene encoding a gut antigen protein bearing repetitive sequences. *Mol Biochem Parasitol* 1992;55:155-65.
- [201] MATERA A. G., WEINER A. M., SCHMID C. W. Structure and evolution of the U2 small nuclear RNA multigene family in primates: gene amplification under natural selection? *Mol Cell Biol* 1990;10:5876-82.
- [202] MAYER L. W. Rates in vitro changes of gonococcal colony opacity phenotypes. *Infect Immun* 1982;37:481-5.
- [203] MAYER M. G., FLOETER-WINTER L. M. Pre-mRNA trans-splicing: from kinetoplastids to mammals, an easy language for life diversity. *Mem Inst Oswaldo Cruz* 2005;100:501-13.
- [204] MAYER W. E., UINUK-OOL T., TICHY H., GARTLAND L. A., KLEIN J., COOPER M. D. Isolation and characterization of lymphocyte-like cells from a lamprey. *Proc Natl Acad Sci U S A* 2002;99:14350-5.
- [205] MAZET F., SHIMELD S. M. Gene duplication and divergence in the early evolution of vertebrates. *Curr Opin Genet Dev* 2002;12:393-6.
- [206] MEDZHITOV R., JANEWAY C. A., JR. Innate immunity: the virtues of a nonclonal system of recognition. *Cell* 1997;91:295-8.
- [207] MEIER J. T., LEWIS S. M. P nucleotides in V(D)J recombination: a fine-structure analysis. *Mol Cell Biol* 1993;13:1078-92.
- [208] MENDOZA H. L., FAYE I. Physiological aspects of the immunoglobulin superfamily in invertebrates. *Dev Comp Immunol* 1999;23:359-74.
- [209] MEUNIER L. Clonal variation of gene expression as a source of phenotypic diversity in parasitic protozoa. *Trends Parasitol* 2001;17:475-9.
- [210] MEYER T. F., MLAWER N., SO M. Pilus expression in *Neisseria gonorrhoeae* involves chromosomal rearrangement. *Cell* 1982;30:45-52.

- [211] MEYER T. F., BILLYARD E., HAAS R., STORZBACH S., SO M. Pilus genes of *Neisseria gonorrhoeae*: chromosomal organization and DNA sequence. Proc Natl Acad Sci U S A 1984;81:6110-4.
- [212] MEYER T. F., GIBBS C. P., HAAS R. Variation and control of protein expression in *Neisseria*. Annu Rev Microbiol 1990;44:451-77.
- [213] MILLER L. H., BARUCH D. I., MARSH K., DOUMBO O. K. The pathogenic basis of malaria. Nature 2002;415:673-9.
- [214] MONTGOMERY R. Glycoproteins. in: Pigman, W., Horton, D. (Eds), The Carbohydrates: Chemistry and Biochemistry, Vol2B Academic Press, New York, 628-709, 1970.
- [215] MOORE M. J., SHARP P. A. Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature 1993;365:364-8.
- [216] MORTARA R. A., DA SILVA S., ARAGUTH M. F., BLANCO S. A., YOSHIDA N. Polymorphism of the 35- and 50-kilodalton surface glycoconjugates of *Trypanosoma cruzi* metacyclic trypomastigotes. Infect Immun 1992;60:4673-8.
- [217] MULLER N., GOTTSTEIN B. Antigenic variation and the murine immune response to *Giardia lamblia*. Int J Parasitol 1998;28:1829-39.
- [218] MURAMATSU M., SANKARANAND V. S., ANANT S., SUGAI M., KINOSHITA K., DAVIDSON N. O., HONJO T. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J Biol Chem 1999;274:18470-6.
- [219] MURPHY G. L., CONNELL T. D., BARRITT D. S., KOOMEY M., CANNON J. G. Phase variation of gonococcal protein II: regulation of gene expression by slipped-strand mispairing of a repetitive DNA sequence. Cell 1989;56:539-47.
- [220] NASH T. E., BANKS S. M., ALLING D. W., MERRITT J. W., JR., CONRAD J. T. Frequency of variant antigens in *Giardia lamblia*. Exp Parasitol 1990;71:415-21.
- [221] NASH T. E. Surface antigenic variation in *Giardia lamblia*. Mol Microbiol 2002;45:585-90.
- [222] NAVARRO M., CROSS G. A. DNA rearrangements associated with multiple consecutive directed antigenic switches in *Trypanosoma brucei*. Mol Cell Biol 1996;16:3615-25.
- [223] NAVARRO M., CROSS G. A. In situ analysis of a variant surface glycoprotein expression-site promoter region in *Trypanosoma brucei*. Mol Biochem Parasitol 1998;94:53-66.
- [224] NAVARRO M., CROSS G. A., WIRTZ E. *Trypanosoma brucei* variant surface glycoprotein regulation involves coupled activation/inactivation and chromatin remodeling of expression sites. Embo J 1999;18:2265-72.
- [225] NAVARRO M., GULL K. A pol I transcriptional body associated with VSG mono-allelic expression in *Trypanosoma brucei*. Nature 2001;414:759-63.

- [226] NAVARRO M., PENATE X., LANDEIRA D. Nuclear architecture underlying gene expression in *Trypanosoma brucei*. Trends Microbiol 2007;15:263-70.
- [227] NEI M., GOJOBORI T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 1986;3:418-26.
- [228] NEI M., HUGHES A. L. Balanced polymorphism and evolution by birth-and-death process in the MHC loci. Oxford Univ. Press 1992.
- [229] NEI M., ROONEY A. P. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet 2005;39:121-52.
- [230] O'CONNELL B., TABAK L. A., RAMASUBBU N. The influence of flanking sequences on O-glycosylation. Biochem Biophys Res Commun 1991;180:1024-30.
- [231] OETTINGER M. A., SCHATZ D. G., GORKA C., BALTIMORE D. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science 1990;248:1517-23.
- [232] OTA T., NEI M. Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. Mol Biol Evol 1994;11:469-82.
- [233] PAGE A. P., HAMILTON A. J., MAIZELS R. M. *Toxocara canis*: monoclonal antibodies to carbohydrate epitopes of secreted (TES) antigens localize to different secretion-related structures in infective larvae. Exp Parasitol 1992a;75:56-71.
- [234] PAGE A. P., RUDIN W., FLURI E., BLAXTER M. L., MAIZELS R. M. *Toxocara canis*: a labile antigenic surface coat overlying the epicuticle of infective larvae. Exp Parasitol 1992b;75:72-86.
- [235] PALLARES N., FRIPPIAT J. P., GIUDICELLI V., LEFRANC M. P. The human immunoglobulin lambda variable (IGLV) genes and joining (IGLJ) segments. Exp Clin Immunogenet 1998;15:8-18.
- [236] PALLARES N., LEFEBVRE S., CONTET V., MATSUDA F., LEFRANC M. P. The human immunoglobulin heavy variable genes. Exp Clin Immunogenet 1999;16:36-60.
- [237] PALMER G. H., BRAYTON K. A. Gene conversion is a convergent strategy for pathogen antigenic variation. Trends Parasitol 2007;23:408-13.
- [238] PAN S. C. *Schistosoma mansoni*: the ultrastructure of larval morphogenesis in *Biomphalaria glabrata* and of associated host-parasite interactions. Jpn J Med Sci Biol 1996;49:129-49.
- [239] PANCER Z., AMEMIYA C. T., EHRHARDT G. R., CEITLIN J., GARTLAND G. L., COOPER M. D. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature 2004;430:174-80.
- [240] PANCER Z., COOPER M. D. The evolution of adaptive immunity. Annu Rev Immunol 2006;24:497-518.
- [241] PARHAM P., ADAMS E. J., ARNETT K. L. The origins of HLA-A,B,C polymorphism. Immunol Rev 1995;143:141-80.

- [242] PARHAM P., MASSON P. L. Le système immunitaire. De Boeck Université 2003.
- [243] PATTON S., GENDLER S. J., SPICER A. P. The epithelial mucin, MUC1, of milk, mammary gland and other tissues. *Biochim Biophys Acta* 1995;1241:407-23.
- [244] PAVELITZ T., RUSCHE L., MATERA A. G., SCHARF J. M., WEINER A. M. Concerted evolution of the tandem array encoding primate U2 snRNA occurs in situ, without changing the cytological context of the RNU2 locus. *Embo J* 1995;14:169-77.
- [245] PAYS E., VAN ASSEL S., LAURENT M., DARVILLE M., VERVOORT T., VAN MEIRVENNE N., STEINERT M. Gene conversion as a mechanism for antigenic variation in trypanosomes. *Cell* 1983;34:371-81.
- [246] PAYS E., TEBABI P., PAYS A., COQUELET H., REVELARD P., SALMON D., STEINERT M. The genes and transcripts of an antigen gene expression site from *T. brucei*. *Cell* 1989;57:835-45.
- [247] PAYS E., VANHAMME L., PEREZ-MORGA D. Antigenic variation in *Trypanosoma brucei*: facts, challenges and mysteries. *Curr Opin Microbiol* 2004;7:369-74.
- [248] PAYS E. Regulation of antigen gene expression in *Trypanosoma brucei*. *Trends Parasitol* 2005;21:517-20.
- [249] PEREIRA-CHIOCCOLA V. L., SCHENKMAN S. Biological role of *Trypanosoma cruzi* trans-sialidase. *Biochem Soc Trans* 1999;27:516-8.
- [250] PEREIRA-CHIOCCOLA V. L., ACOSTA-SERRANO A., CORREIA DE ALMEIDA I., FERGUSON M. A., SOUTO-PADRON T., RODRIGUES M. M., TRAVASSOS L. R., SCHENKMAN S. Mucin-like molecules form a negatively charged coat that protects *Trypanosoma cruzi* trypomastigotes from killing by human anti-alpha-galactosyl antibodies. *J Cell Sci* 2000;113 ( Pt 7):1299-307.
- [251] PERELSON A. S., NEUMANN A. U., MARKOWITZ M., LEONARD J. M., HO D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 1996;271:1582-6.
- [252] PEREY D. Y., FINSTAD J., POLLARA B., GOOD R. A. Evolution of the immune response. VI. First and second set skin homograft rejections in primitive fishes. *Lab Invest* 1968;19:591-7.
- [253] PLANT L., SUNDQVIST J., ZUGHAIER S., LOVKVIST L., STEPHENS D. S., JONSSON A. B. Lipooligosaccharide structure contributes to multiple steps in the virulence of *Neisseria meningitidis*. *Infect Immun* 2006;74:1360-7.
- [254] PLOTKIN J. B., DUSHOFF J. Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. *Proc Natl Acad Sci U S A* 2003;100:7152-7.
- [255] POLLEVICK G. D., DI NOIA J. M., SALTO M. L., LIMA C., LEGUIZAMON M. S., DE LEDERKREMER R. M., FRASCH A. C. *Trypanosoma cruzi* surface mucins with exposed variant epitopes. *J Biol Chem* 2000;275:27671-80.



- [256] PREVIATO J. O., JONES C., GONCALVES L. P., WAIT R., TRAVASSOS L. R., MENDONCA-PREVIATO L. O-glycosidically linked N-acetylglucosamine-bound oligosaccharides from glycoproteins of *Trypanosoma cruzi*. *Biochem J* 1994;301 ( Pt 1):151-9.
- [257] PREVIATO J. O., JONES C., XAVIER M. T., WAIT R., TRAVASSOS L. R., PARODI A. J., MENDONCA-PREVIATO L. Structural characterization of the major glycosylphosphatidylinositol membrane-anchored glycoprotein from epimastigote forms of *Trypanosoma cruzi* Y-strain. *J Biol Chem* 1995;270:7241-50.
- [258] PUCHTA H. Double-strand break-induced recombination between ectopic homologous sequences in somatic plant cells. *Genetics* 1999;152:1173-81.
- [259] RAISON R. L., HILDEMANN W. H. Immunoglobulin-bearing blood leucocytes in the Pacific hagfish. *Dev Comp Immunol* 1984;8:99-108.
- [260] RAISON R. L., COVERLEY J., HOOK J. W., TOWNS P., WESTON K. M., RAFTOS D. A. A cell-surface opsonic receptor on leucocytes from the phylogenetically primitive vertebrate, *Eptatretus stouti*. *Immunol Cell Biol* 1994;72:326-32.
- [261] RAJKOVIC A., DAVIS R. E., SIMONSEN J. N., ROTTMAN F. M. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc Natl Acad Sci U S A* 1990;87:8879-83.
- [262] RALPH S. A., SCHEIDIG-BENATAR C., SCHERF A. Antigenic variation in *Plasmodium falciparum* is associated with movement of var loci between subnuclear locations. *Proc Natl Acad Sci U S A* 2005;102:5414-9.
- [263] RALPH S. A., SCHERF A. The epigenetic control of antigenic variation in *Plasmodium falciparum*. *Curr Opin Microbiol* 2005;8:434-40.
- [264] READ A. F. The evolution of virulence. *Trends Microbiol* 1994;2:73-6.
- [265] REYNAUD C. A., ANQUEZ V., GRIMAL H., WEILL J. C. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. *Cell* 1987;48:379-88.
- [266] REYNAUD C. A., BERTOCCI B., DAHAN A., WEILL J. C. Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. *Adv Immunol* 1994;57:353-78.
- [267] ROBERTS D. J., CRAIG A. G., BERENDT A. R., PINCHES R., NASH G., MARSH K., NEWBOLD C. I. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* 1992;357:689-92.
- [268] ROBERTSON D. L., SHARP P. M., MCCUTCHAN F. E., HAHN B. H. Recombination in HIV-1. *Nature* 1995;374:124-6.
- [269] ROBINSON J., WALLER M. J., PARHAM P., DE GROOT N., BONTROP R., KENNEDY L. J., STOEHR P., MARSH S. G. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* 2003;31:311-4.

- [270] ROGER E., MITTA G., MONE Y., BOUCHUT A., ROGNON A., GRUNAU C., BOISSIER J., THERON A., GOURBAL B. E. Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata/Schistosoma mansoni* model: new candidates identified by a global comparative proteomics approach. *Mol Biochem Parasitol* 2008a;157:205-16.
- [271] ROGER E., GOURBAL B., GRUNAU C., PIERCE R. J., GALINIER R., MITTA G. Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. *Mol Biochem Parasitol* 2008b;157:217-27.
- [272] ROGER E., GRUNAU C., PIERCE R. J., HIRAI H., GOURBAL B., GALINIER R., EMANS R., CESARI I., COSSEAU C., MITTA G. Polymorphism of Mucins secreted by a metazoan parasite (*Schistosoma mansoni*) interacting with its invertebrate host (*Biomphalaria glabrata*): involvement of recombination, alternative splicing, locus-specific transcription and glycosylation. *Mol Biol Evol* 2008c; *Under Submission*.
- [273] ROGOZIN I. B., IYER L. M., LIANG L., GLAZKO G. V., LISTON V. G., PAVLOV Y. I., ARAVIND L., PANCER Z. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. *Nat Immunol* 2007;8:647-56.
- [274] ROSS A. G., BARTLEY P. B., SLEIGH A. C., OLDS G. R., LI Y., WILLIAMS G. M., MCMANUS D. P. Schistosomiasis. *N Engl J Med* 2002;346:1212-20.
- [275] ROTH C., BRINGAUD F., LAYDEN R. E., BALTZ T., EISEN H. Active late-appearing variable surface antigen genes in *Trypanosoma equiperdum* are constructed entirely from pseudogenes. *Proc Natl Acad Sci U S A* 1989;86:9375-9.
- [276] RUDENKO G., LEE M. G., VAN DER PLOEG L. H. The PARP and VSG genes of *Trypanosoma brucei* do not resemble RNA polymerase II transcription units in sensitivity to Sarkosyl in nuclear run-on assays. *Nucleic Acids Res* 1992;20:303-6.
- [277] RUIZ R. C., FAVORETO S., JR., DORTA M. L., OSHIRO M. E., FERREIRA A. T., MANQUE P. M., YOSHIDA N. Infectivity of *Trypanosoma cruzi* strains is associated with differential expression of surface glycoproteins with differential Ca<sup>2+</sup> signalling activity. *Biochem J* 1998;330 ( Pt 1):505-11.
- [278] RUIZ RDE C., RIGONI V. L., GONZALEZ J., YOSHIDA N. The 35/50 kDa surface antigen of *Trypanosoma cruzi* metacyclic trypomastigotes, an adhesion molecule involved in host cell invasion. *Parasite Immunol* 1993;15:121-5.
- [279] SADD B. M., SCHMID-HEMPEL P. Insect immunity shows specificity in protection upon secondary pathogen exposure. *Curr Biol* 2006;16:1206-10.
- [280] SALZET M., CAPRON A., STEFANO G. B. Molecular crosstalk in host-parasite relationships: schistosome- and leech-host interactions. *Parasitol Today* 2000;16:536-40.
- [281] SAPP K. K., LOKER E. S. Mechanisms underlying digenean-snail specificity: role of miracidial attachment and host plasma factors. *J Parasitol* 2000;86:1012-9.
- [282] SCAVINER D., LEFRANC M. P. The human T cell receptor alpha variable (TRAV) genes. *Exp Clin Immunogenet* 2000;17:83-96.

- [283] SCHEER U., HOCK R. Structure and function of the nucleolus. *Curr Opin Cell Biol* 1999;11:385-90.
- [284] SCHENKMAN S., JIANG M. S., HART G. W., NUSSENZWEIG V. A novel cell surface trans-sialidase of *Trypanosoma cruzi* generates a stage-specific epitope required for invasion of mammalian cells. *Cell* 1991;65:1117-25.
- [285] SCHENKMAN S., FERGUSON M. A., HEISE N., DE ALMEIDA M. L., MORTARA R. A., YOSHIDA N. Mucin-like glycoproteins linked to the membrane by glycosylphosphatidylinositol anchor are the major acceptors of sialic acid in a reaction catalyzed by trans-sialidase in metacyclic forms of *Trypanosoma cruzi*. *Mol Biochem Parasitol* 1993;59:293-303.
- [286] SCHERF A., HERNANDEZ-RIVAS R., BUFFET P., BOTTIUS E., BENATAR C., POUVELLE B., GYSIN J., LANZER M. Antigenic variation in malaria: in situ switching, relaxed and mutually exclusive transcription of var genes during intra-erythrocytic development in *Plasmodium falciparum*. *Embo J* 1998;17:5418-26.
- [287] SCHIMENTI J. C. Mice and the role of unequal recombination in gene-family evolution. *Am J Hum Genet* 1999;64:40-5.
- [288] SCHLOTTERER C., TAUTZ D. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr Biol* 1994;4:777-83.
- [289] SCHMUCKER D., CLEMENS J. C., SHU H., WORBY C. A., XIAO J., MUDA M., DIXON J. E., ZIPURSKY S. L. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 2000;101:671-84.
- [290] SCHNEIDER H., HAMMACK C. A., APICELLA M. A., GRIFFISS J. M. Instability of expression of lipooligosaccharides and their epitopes in *Neisseria gonorrhoeae*. *Infect Immun* 1988;56:942-6.
- [291] SCHOLTISSEK C. Source for influenza pandemics. *Eur J Epidemiol* 1994;10:455-8.
- [292] SCHULENBURG H., BOEHNISCH C., MICHIELS N. K. How do invertebrates generate a highly specific innate immune response? *Mol Immunol* 2007;44:3338-44.
- [293] SCHWEIGER B., ZADOW I., HECKLER R. Antigenic drift and variability of influenza viruses. *Med Microbiol Immunol* 2002;191:133-8.
- [294] SCOTT F. G. *Developmental Biology*, Eighth Edition. 2006.
- [295] SEGAL E., HAGBLOM P., SEIFERT H. S., SO M. Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments. *Proc Natl Acad Sci U S A* 1986;83:2177-81.
- [296] SELSING E., XU B., SIGURDARDOTTIR D. Gene conversion and homologous recombination in murine B cells. *Semin Immunol* 1996;8:151-8.
- [297] SERKIN C. D., SEIFERT H. S. Frequency of pilin antigenic variation in *Neisseria gonorrhoeae*. *J Bacteriol* 1998;180:1955-8.

- [298] SERRANO A. A., SCHENKMAN S., YOSHIDA N., MEHLERT A., RICHARDSON J. M., FERGUSON M. A. The lipid structure of the glycosylphosphatidylinositol-anchored mucin-like sialic acid acceptors of *Trypanosoma cruzi* changes during parasite differentiation from epimastigotes to infective metacyclic trypomastigote forms. *J Biol Chem* 1995;270:27244-53.
- [299] SHARP P. M., ROBERTSON D. L., HAHN B. H. Cross-species transmission and recombination of 'AIDS' viruses. *Philos Trans R Soc Lond B Biol Sci* 1995;349:41-7.
- [300] SHEA C., LEE M. G., VAN DER PLOEG L. H. VSG gene 118 is transcribed from a cotransposed pol I-like promoter. *Cell* 1987;50:603-12.
- [301] SHINTANI S., TERZIC J., SATO A., SARAGA-BABIC M., O'HUIGIN C., TICHY H., KLEIN J. Do lampreys have lymphocytes? The Spi evidence. *Proc Natl Acad Sci U S A* 2000;97:7417-22.
- [302] SMITH G. P. Evolution of repeated DNA sequences by unequal crossover. *Science* 1976;191:528-35.
- [303] SMITH H. V., QUINN R., KUSEL J. R., GIRDWOOD R. W. The effect of temperature and antimetabolites on antibody binding to the outer surface of second stage *Toxocara canis* larvae. *Mol Biochem Parasitol* 1981;4:183-93.
- [304] SNOW R. W., GUERRA C. A., NOOR A. M., MYINT H. Y., HAY S. I. The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* 2005;434:214-7.
- [305] SOLNICK D. Trans splicing of mRNA precursors. *Cell* 1985;42:157-64.
- [306] STARCICH B. R., HAHN B. H., SHAW G. M., MCNEELY P. D., MODROW S., WOLF H., PARKS E. S., PARKS W. P., JOSEPHS S. F., GALLO R. C., ET AL. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* 1986;45:637-48.
- [307] STERN A., MEYER T. F. Common mechanism controlling phase and antigenic variation in pathogenic neisseriae. *Mol Microbiol* 1987;1:5-12.
- [308] STROUS G. J., DEKKER J. Mucin-type glycoproteins. *Crit Rev Biochem Mol Biol* 1992;27:57-92.
- [309] SU X. Z., HEATWOLE V. M., WERTHEIMER S. P., GUINET F., HERRFELDT J. A., PETERSON D. S., RAVETCH J. A., WELLEMS T. E. The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* 1995;82:89-100.
- [310] SUN S. C., LINDSTROM I., BOMAN H. G., FAYE I., SCHMIDT O. Hemolin: an insect-immune protein belonging to the immunoglobulin superfamily. *Science* 1990;250:1729-32.
- [311] SWANSON J. Studies on gonococcus infection. IV. Pili: their role in attachment of gonococci to tissue culture cells. *J Exp Med* 1973;137:571-89.

- [312] SWANSON J. Colony opacity and protein II compositions of gonococci. *Infect Immun* 1982;37:359-68.
- [313] SWANSON J., ROBBINS K., BARRERA O., KOOMEY J. M. Gene conversion variations generate structurally distinct pilin polypeptides in *Neisseria gonorrhoeae*. *J Exp Med* 1987;165:1016-25.
- [314] SZOSTAK J. W., ORR-WEAVER T. L., ROTHSTEIN R. J., STAHL F. W. The double-strand-break repair model for recombination. *Cell* 1983;33:25-35.
- [315] TAYLOR H. M., KYES S. A., NEWBOLD C. I. Var gene diversity in *Plasmodium falciparum* is generated by frequent recombination events. *Mol Biochem Parasitol* 2000;110:391-7.
- [316] TAYLOR J. E., RUDENKO G. Switching trypanosome coats: what's in the wardrobe? *Trends Genet* 2006;22:614-20.
- [317] THANKA CHRISTLET T. H., VELURAJA K. Database analysis of O-glycosylation sites in proteins. *Biophys J* 2001;80:952-60.
- [318] THEODOROPOULOS G., HICKS S. J., CORFIELD A. P., MILLER B. G., CARRINGTON S. D. The role of mucins in host-parasite interactions: Part II - helminth parasites. *Trends Parasitol* 2001;17:130-5.
- [319] THERON A. Chronobiology of schistosome development in the snail host. *Parasitol Today* 1986;2:192-4.
- [320] THERON A., SIRE C., ROGNON A., PRUGNOLLE F., DURAND P. Molecular ecology of *Schistosoma mansoni* transmission inferred from the genetic composition of larval and adult infrapopulations within intermediate and definitive hosts. *Parasitology* 2004;129:571-85.
- [321] THERON A. Les schistosomes et leurs hôtes : apport des marqueurs moléculaires à la connaissance de leur phylogéographie, épidémiologie et écologie. *Bull Soc Zool Fr* 2005;130:205-218.
- [322] THERON A., COUSTAU C. Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *J Helminthol* 2005;79:187-91.
- [323] THOMPSON C. B., NEIMAN P. E. Somatic diversification of the chicken immunoglobulin light chain gene is limited to the rearranged variable gene segment. *Cell* 1987;48:369-78.
- [324] THON G., BALTZ T., EISEN H. Antigenic diversity by the recombination of pseudogenes. *Genes Dev* 1989;3:1247-54.
- [325] TONEGAWA S. Somatic generation of antibody diversity. *Nature* 1983;302:575-81.
- [326] TROWSDALE J., PARHAM P. Mini-review: defense strategies and immunity-related genes. *Eur J Immunol* 2004;34:7-17.
- [327] UDOMSANGPETCH R., WAHLIN B., CARLSON J., BERZINS K., TORII M., AIKAWA M., PERLMANN P., WAHLGREN M. *Plasmodium falciparum*-infected erythrocytes form spontaneous erythrocyte rosettes. *J Exp Med* 1989;169:1835-40.

- [328] UINUK-OOL T., MAYER W. E., SATO A., DONGAK R., COOPER M. D., KLEIN J. Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. *Proc Natl Acad Sci U S A* 2002;99:14356-61.
- [329] VAN DER KNAAP W. P., BOOTS A. M., MEULEMAN E. A., SMINIA T. Search for shared antigens in the schistosome-snail combination *Trichobilharzia ocellata*-*Lymnaea stagnalis*. *Z Parasitenkd* 1985;71:219-26.
- [330] VAN DER KNAAP W. P., LOKER E. S. Immune mechanisms in trematode-snail interactions. *Parasitol Today* 1990;6:175-82.
- [331] VAN DER PLOEG L. H., VALERIO D., DE LANGE T., BERNARDS A., BORST P., GROSVELD F. G. An analysis of cosmid clones of nuclear DNA from *Trypanosoma brucei* shows that the genes for variant surface glycoproteins are clustered in the genome. *Nucleic Acids Res* 1982;10:5905-23.
- [332] VAN DER WOUDE M. W., BAUMLER A. J. Phase and antigenic variation in bacteria. *Clin Microbiol Rev* 2004;17:581-611, table of contents.
- [333] VAN PUTTEN J. P., ROBERTSON B. D. Molecular mechanisms and implications for infection of lipopolysaccharide variation in *Neisseria*. *Mol Microbiol* 1995;16:847-53.
- [334] VAN VALEN L. A New Evolutionary Law. *Evolutionary Theory* 1973;1-30.
- [335] VANHAMME L., PAYS A., TEBABI P., ALEXANDRE S., PAYS E. Specific binding of proteins to the noncoding strand of a crucial element of the variant surface glycoprotein, procyclin, and ribosomal promoters of *Trypanosoma brucei*. *Mol Cell Biol* 1995;15:5598-606.
- [336] VANHAMME L., POELVOORDE P., PAYS A., TEBABI P., VAN XONG H., PAYS E. Differential RNA elongation controls the variant surface glycoprotein gene expression sites of *Trypanosoma brucei*. *Mol Microbiol* 2000;36:328-40.
- [337] VITTE C., PANAUD O. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenet Genome Res* 2005;110:91-107.
- [338] VOGEL C., TEICHMANN S. A., CHOTHIA C. The immunoglobulin superfamily in *Drosophila melanogaster* and *Caenorhabditis elegans* and the evolution of complexity. *Development* 2003;130:6317-28.
- [339] VOSS T. S., HEALER J., MARTY A. J., DUFFY M. F., THOMPSON J. K., BEESON J. G., REEDER J. C., CRABB B. S., COWMAN A. F. A var gene promoter controls allelic exclusion of virulence genes in *Plasmodium falciparum* malaria. *Nature* 2006;439:1004-8.
- [340] WANG G., VASQUEZ K. M. Non-B DNA structure-induced genetic instability. *Mutat Res* 2006;598:103-19.
- [341] WATSON F. L., PUTTMANN-HOLGADO R., THOMAS F., LAMAR D. L., HUGHES M., KONDO M., REBEL V. I., SCHMUCKER D. Extensive diversity of Ig-superfamily proteins in the immune system of insects. *Science* 2005;309:1874-8.

- [342] WEBSTER J. P., WOOLHOUSE M. E. Selection and strain specificity of compatibility between snail intermediate hosts and their parasitic schistosomes. *Evolution* 1998;52:1627-1634.
- [343] WEBSTER J. P., DAVIES C. M. Coevolution and compatibility in the snail-schistosome system. *Parasitology* 2001;123 Suppl:S41-56.
- [344] WEBSTER J. P., GOWER C. M., BLAIR L. Do hosts and parasites coevolve? Empirical support from the *Schistosoma* system. *Am Nat* 2004;164 Suppl 5:S33-51.
- [345] WEBSTER J. P., SHRIVASTAVA J., JOHNSON P. J., BLAIR L. Is host-schistosome coevolution going anywhere? *BMC Evol Biol* 2007;7:91.
- [346] WEBSTER R. G., BEAN W. J., GORMAN O. T., CHAMBERS T. M., KAWAOKA Y. Evolution and ecology of influenza A viruses. *Microbiol Rev* 1992;56:152-79.
- [347] WEILL J. C., REYNAUD C. A. Generation of diversity by post-rearrangement diversification mechanisms: the chicken and the sheep antibody repertoires. 2. Academic Press 1995.
- [348] WEILL J. C., REYNAUD C. A. Rearrangement/hypermutation/gene conversion: when, where and why? *Immunol Today* 1996;17:92-7.
- [349] WELLS R. D. Unusual DNA structures. *J Biol Chem* 1988;263:1095-8.
- [350] WESTIN G., ZABIELSKI J., HAMMARSTROM K., MONSTEIN H. J., BARK C., PETTERSSON U. Clustered genes for human U2 RNA. *Proc Natl Acad Sci U S A* 1984;81:3811-5.
- [351] WILSON I. B., GAVEL Y., VON HEIJNE G. Amino acid distributions around O-linked glycosylation sites. *Biochem J* 1991;275 ( Pt 2):529-34.
- [352] WYATT R., SODROSKI J. The HIV-1 envelope glycoproteins: fusogens, antigens, and immunogens. *Science* 1998;280:1884-8.
- [353] XU X., DOOLITTLE R. F. Presence of a vertebrate fibrinogen-like sequence in an echinoderm. *Proc Natl Acad Sci U S A* 1990;87:2097-101.
- [354] YANG Q. L., GOTSCHLICH E. C. Variation of gonococcal lipooligosaccharide structure is due to alterations in poly-G tracts in *lgt* genes encoding glycosyl transferases. *J Exp Med* 1996;183:323-7.
- [355] YOSHINO T. P., BAYNE C. J. Mimicry of snail host antigens by *miracidia* and primary sporocysts of *Schistosoma mansoni*. *Parasite Immunol* 1983;5:317-28.
- [356] ZAMBON M. C. Epidemiology and pathogenesis of influenza. *J Antimicrob Chemother* 1999;44 Suppl B:3-9.
- [357] ZAPATA A., FANGE R., MATTISSON A., VILLENA A. Plasma cells in adult Atlantic hagfish, *Myxine glutinosa*. *Cell Tissue Res* 1984;235:691-3.
- [358] ZHANG J. Evolution by gene duplication: an update. *TRENDS in Ecology and Evolution* 2003;18:292-298.

- [359] ZHANG Q. Y., DERYCKERE D., LAUER P., KOOMEY M. Gene conversion in *Neisseria gonorrhoeae*: evidence for its role in pilus antigenic variation. Proc Natl Acad Sci U S A 1992;89:5366-70.
- [360] ZHANG S. M., LEONARD P. M., ADEMA C. M., LOKER E. S. Parasite-responsive IgSF members in the snail *Biomphalaria glabrata*: characterization of novel genes with tandemly arranged IgSF domains and a fibrinogen domain. Immunogenetics 2001;53:684-94.
- [361] ZHANG S. M., LOKER E. S. The FREP gene family in the snail *Biomphalaria glabrata*: additional members, and evidence consistent with alternative splicing and FREP retrosequences. Fibrinogen-related proteins. Dev Comp Immunol 2003;27:175-87.
- [362] ZHANG S. M., ADEMA C. M., KEPLER T. B., LOKER E. S. Diversification of Ig superfamily genes in an invertebrate. Science 2004;305:251-4.
- [363] ZHANG S. M., LOKER E. S. Representation of an immune responsive gene family encoding fibrinogen-related proteins in the freshwater mollusc *Biomphalaria glabrata*, an intermediate host for *Schistosoma mansoni*. Gene 2004;341:255-66.
- [364] ZHURKIN V. B., RAGHUNATHAN G., ULYANOV N. B., CAMERINI-OTERO R. D., JERNIGAN R. L. A parallel DNA triplex as a model for the intermediate in homologous recombination. J Mol Biol 1994;239:181-200.
- [365] ZOMERDIJK J. C., OUELLETTE M., TEN ASBROEK A. L., KIEFT R., BOMMER A. M., CLAYTON C. E., BORST P. The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. Embo J 1990;9:2791-801.



## *ANNEXE 1*

*Travaux développés sur le modèle  
Schistosoma mansoni / Biomphalaria glabrata*

## THE COMPATIBILITY POLYMORPHISM IN INVERTEBRATE HOST/TREMATODES INTERACTIONS: RESEARCH OF MOLECULAR DETERMINANTS

BOUCHUT A.\*, ROGER E.\*, GOURBAL B.\*, GRUNAU C.\*, COUSTAU C.\*\* & MITTA G.\*

### Summary:

The co-evolutionary dynamics that exist in many host-parasite interactions sometimes leads to compatibility polymorphism. This phenomenon is well documented in mollusc/trematodes interactions but its molecular base is unknown. In order to identify key molecules involved in this phenomenon, we developed several molecular approaches comparing compatible or incompatible strains of mollusc or parasite. These comparisons led to the identification of numerous candidate genes listed and discussed (some of them) in the present review.

**KEY WORDS :** *Biomphalaria glabrata*, *Schistosoma mansoni*, *Echinostoma caproni*, compatibility polymorphism.

The comprehensive understanding of host-parasite interactions represents a major challenge in evolutionary biology. Because parasites are responsible for substantial deleterious effects, they represent a major driving force for the evolution of their hosts. In parallel, parasites have to co-evolve with their host to avoid elimination. This adaptation of the Red Queen hypothesis to host-parasite systems (Combes, 2000) predicts that an arms race will lead to the potential antagonistic evolution of host resistance and parasite virulence processes leading to the co-evolution of host-parasite systems. In some models, parasite virulence and host defence are in equilibrium in natural populations. This is the case in host-parasite systems for which only certain host and parasite phenotypes are compatible. Such systems constitute models of choice due to the fact that their co-evolutionary dynamics are experimentally accessible permitting the use of comparative approaches between compatible and incompatible strains to identify the molecules playing a key role in the interaction.

The models we have chosen to study are the interaction between the trematodes *Echinostoma caproni* or *Schistosoma mansoni*, the agent of human intestinal

schistosomiasis (Gryseels *et al.*, 2006) and their invertebrate intermediate host, the gastropod mollusc *Biomphalaria glabrata*. In these interactions, a compatibility polymorphism has been evidenced and permits us to select compatible and incompatible strains on which comparative molecular approaches are applicable. In addition, the comparison between these two parasites sharing the same host is interesting because their immune evasion strategy seems to be completely different (Loker & Adema, 1995). Evidences strongly support the hypothesis that the major immune evasion strategy of *S. mansoni* relies on a molecular mimicry mechanism and would therefore avoid recognition as non-self invader by the snail immune system. In contrast to *S. mansoni*, echinostomes are known to strongly interfere with the snail immune response *via* their excretory-secretory (ES) products (Humbert & Coustau, 2001).

The present review focus on the different works we have published the three last years in these two different systems.

### THE INTERACTION BETWEEN *B. GLABRATA* AND *E. CAPRONI*

In this model *B. glabrata*/*E. caproni*, two snail strains of *B. glabrata* were selected in our laboratory (Langand *et al.*, 1998). *E. caproni* miracidia are able to penetrate both compatible (susceptible) and incompatible (resistant) snails. While the parasite undergoes normal development in compatible snails, in incompatible snails it is encapsulated and eliminated. Susceptibility/resistance mechanisms of *B. glabrata* to *E. caproni* are probably inherited in a multigenic fashion (Langand & Morand, 1998) and have been shown to rely on both humoral and cellular factors (Ataev & Coustau, 1999). *In vivo* comparison of *E. caproni* development in both *B. glabrata* strains revealed that, in resistant snails, sporocysts were abnormally developed and degenerated regardless of their level of encapsulation (Ataev & Coustau, 1999). These results suggested that humoral factors are actively involved in the molecular processes

\* Parasitologie Fonctionnelle et Evolutive, UMR 5244, CNRS Université de Perpignan, France.

\*\* Inserm, U 547, Université Lille 2, Institut Pasteur de Lille, IFR 142, France.

Correspondence: Guillaume Mitta, UMR 5244, CNRS Université de Perpignan, 52, avenue Paul Alduy, 66860 Perpignan Cedex, France. Tel.: 33 (0)4 68 66 21 88 – Fax: 33 (0)4 68 66 22 81.

E-mail: mitta@univ-perp.fr

underlying resistance or susceptibility to *E. caproni* (Ataev & Coustau, 1999). As a first approach to investigate the molecular bases of this phenotype difference between strains, we used comparative proteomics with plasma collected from both strains. We revealed 13 plasmatic proteins exhibited significant differences in their apparent representativity (Vergote *et al.*, 2005). The cDNAs corresponding to five of these proteins were identified. They encode proteins belonging to three different classes. The first is a type-2 secreted cystatin that was named Bg type-2 cystatin. The cystatins are cysteine protease inhibitors that control activity of cathepsins, themselves involved in various processes, including immunity (Abrahamson *et al.*, 2003). Two others are isoforms of secreted Calcium Binding proteins of the EF-hand type (Bg CaBP 1 and 2) that are known to play key role in calcium homeostasis. And the two last proteins are two isoforms of an endo-1,4-mannanase that are known to randomly cleave within the  $\beta$ -1,4-mannan main chain of galactomannan, glucomannan, galactoglucomannan and mannan (Xu *et al.*, 2002). All these proteins are secreted, more represented in the plasma of susceptible snails and these differences appear correlated with transcript levels (Vergote *et al.*, 2005). In addition, analysis of their expression in a range of tissues showed that they are expressed by the albumen gland (Vergote *et al.*, 2005). Post-infection expression studies of these five candidate genes showed that the transcript content of all candidates are stable in susceptible snails following infection, while they increase significantly in exposed resistant snails (Vergote *et al.*, 2005). All the differences observed between susceptible and resistant snails could be linked to a differential regulation of genes expressed by the albumen gland. In molluscs, the albumen gland is considered as an accessory sexual gland (Wijdenes *et al.*, 1983). Nevertheless, this organ is also involved in the production of defense effectors like agglutinins in planorbid snails (Michelson & Dubois, 1977; Stein & Basch, 1979; Boswell & Bayne, 1984) or antimicrobial proteins in opothobranchs gastropods (Iijima *et al.*, 1994; Takamatsu *et al.*, 1995). Our results further support the idea that the albumen gland may play a significant role in snail immune response and raise the question of the potential importance of this organ in processes underlying susceptibility/resistance to *E. caproni*.

Regarding the cellular factors, previous studies showed that excretory-secretory (ES) products from in vitro transformed *E. caproni* sporocysts inhibited key defence functions of susceptible host hemocytes such as adhesion and phagocytosis (Humbert & Coustau, 2001). Interestingly, hemocytes from resistant snails remained unaffected by these parasite ES products, suggesting that they exhibit differences compared with hemocytes from susceptible snails (Humbert & Coustau, 2001). We develop two comparative approaches in order to disco-

ver the molecular basis of this phenotype difference. Because such differences were indirectly or directly related to adhesive properties of hemocytes, we developed a first targeted approach to examine the expression of genes directly involved in adhesion processes (Bouchut *et al.*, 2006). In this study, we compared the transcript content of all known hemocytic genes potentially involved in adhesion processes between hemocytes of the two strains. Transcripts of four genes were differentially present between hemocytes from resistant and susceptible snails. These genes encode proteins resembling dermatopontins (two isoforms), matrilin and cadherin (Bouchut *et al.*, 2006). Among them, dermatopontins seem to be promising candidates. These novel dermatopontins (Bg dermatopontin 2 and -3) are probably secreted and the structural features of the precursors suggest that these two molecules share, in addition to dermatopontin general characteristics, some other particular features already observed in two other invertebrate dermatopontins from *Suberites domuncula* (cysteine array, (Schutze *et al.*, 2001)) and *Limulus polyphemus* (additional internal repeat, (Fujii *et al.*, 1992)). The function of these particular structural features has not been elucidated. Nevertheless, it was shown that these two dermatopontins triggered adhesive processes of dissociated cells from *S. domuncula* (Schutze *et al.*, 2001) or hemocytes from *L. polyphemus* (Fujii *et al.*, 1992). These data suggest that *B. glabrata* hemocytic dermatopontins could participate in aggregation processes of hemocytes. Furthermore, we showed in the same work (Bouchut *et al.*, 2006) that the transcript content of these two genes was: *i*) higher in hemocytes from unexposed resistant snails (in comparison with those of susceptible snails, and *ii*) greatly enhanced in resistant snails 48 h after exposure to the parasite. Interestingly, previous histological studies showed that, in *E. caproni*-exposed resistant snails, the hemocytic capsule was mounted at 48/72 h post exposure (Ataev & Coustau, 1999). The fact that the higher level of expression of Bg dermatopontins observed in resistant snails at 48 h after exposure coincides with the capsule formation, further supporting the hypothesis that this gene could play a key role in adhesive processes taking place during capsule formation in resistant snails.

After this first targeted transcriptomic approach developed on hemocytes of both strains, we developed a global proteomic approach (Bouchut *et al.*, 2006). The comparative approach we used to study hemocytes collected from susceptible and resistant snails revealed that 12 out of the 2,552 proteins from hemocytes exhibited significant differences in their apparent abundance. Direct comparison of tags obtained by LC-MS-MS to databases allowed the identification of five of these 12 differentially represented proteins. These proteins were named Bg aldolase, Bg intermediate filament protein, Bg cytidine deaminase, Bg ribosomal protein P1

and Bg histone H4 (see (Bouchut *et al.*, 2006) for discussion of their putative function).

These different approaches taken together allowed identification of several candidate genes and suggested a potential involvement of tissues other than hemocytes in susceptibility/resistance processes (Vergote *et al.*, 2005; Bouchut *et al.*, 2006). In order to investigate this latter point, we compared the transcripts of the two snail strains at the whole body level using a Suppression Subtractive Hybridization (SSH) approach (Bouchut *et al.*, 2007). This approach revealed several genes identified in the above mentioned studies but also new candidates belonging to novel functional groups of interest. The functional groups of – glycolytic enzymes and Calcium Binding Proteins (CaBP) – were identified from *B. glabrata* plasma in the above mentioned proteomic approach (Vergote *et al.*, 2005) and were also identified by SSH. Regarding novel functional groups, the most promising candidates identified in the present study belong to immunity class and were identified in the library containing genes over expressed in resistant snails. For example, we identified a cluster displaying similarities for a defence factor Aplysianin A purified from the albumen gland of *Aplysia kurodai* (Takamatsu *et al.*, 1995). This glycoprotein inhibited the growth of both Gram-positive and -negative bacteria. Another study developed in *Aplysia punctata* has evidenced a molecule called APIT (*A. punctata* Ink Toxin) displaying about 60 % identity to Aplysianin A (Butzke *et al.*, 2005). APIT is an L-amino acid oxidase. This enzyme produces hydrogen peroxide and was shown to lead to a necrosis-like oxidative damage of eukaryotic cells. The Bg Aplysianin A-like we identify in the SSH study could be a defence factor involved in the antiparasitic response of *B. glabrata*. Its higher abundance in resistant snails could explain the efficiency of resistant snail response against *E. caproni*.

## THE INTERACTION BETWEEN *B. GLABRATA* AND *S. MANSONI*

Another model we have chosen to study in our laboratory is the interaction between *S. mansoni*, the agent of human intestinal schistosomiasis (Gryseels *et al.*, 2006) and its invertebrate intermediate host, the gastropod mollusc *B. glabrata*. In this interaction, a compatibility polymorphism has been also evidenced (Theron & Coustau, 2005). In compatible interactions, the parasite penetrates and develops normally within the snail and in incompatible interactions, the larval trematode penetrates but is recognized as non-self, encapsulated and destroyed by the mollusk's internal defense system. The success or failure of an infection does not depend on the snail susceptibi-

lity/resistance status, but on the “matched” or “mismatched” status of the host and parasite phenotypes (“matching phenotype” hypothesis (Theron & Coustau, 2005)).

In this *S. mansoni*/*B. glabrata* interaction, two *S. mansoni* strains are available, one compatible (strain C) and the other incompatible (strain IC) towards the same snail strain. Strain IC miracidia are immediately recognized as non-self and encapsulated in the first hours following penetration (Roger *et al.*, 2007). The compatible ones remain unaffected. These observations suggest constitutive antigenic differences between strains. To identify and characterize the underlying molecular determinants, we conducted a global comparative proteomics approach on primary sporocysts (Sp1) from C and IC strains (Roger *et al.*, 2007). This approach led us to identify a promising candidate belonging to a family of schistosome antigens that share some characteristics with the molecules of the mucin family (named Sm PoMucs). In particular, they display a domain containing a variable number of tandem repeats (VNTR) rich in S, T and P. In a complementary work, we examine precursor structure and report the analysis of expression and polymorphism of this mucin-like family (Roger *et al.*, 2007). The corresponding precursors are composed of three distinct domains: a signal peptide of 22 amino-acids followed by the variable domain containing a variable number of 9-residue tandem repeats and a C-terminal domain of 234 amino-acids. Three types of repeats were identified and named r1, r1' and r2. Three groups of Sm PoMuc cDNAs were identified in both strains. Each group contains cDNAs that are identical in the deduced amino-acid sequence of the 234-residue C-terminal domain. Extensive analysis of polymorphism revealed a crucial difference within and between the IC and C strains. The polymorphism revealed within the two first groups consists in a variable number of r2 repeats (1 to 55). The major difference between strains is observed for the third group of Sm PoMuc. This latter displays a VNTR domain containing only r1 and/or r1' in strain C. In the strain IC, this VNTR combination was observed, but interestingly, a strain-specific subgroup of molecular variants containing both repeats r2 and r1/r1' was also identified.

This VNTR region of Sm PoMuc is of particular interest since in-silico prediction identifies this region as a potential target of glycosylation and we showed that Sm PoMuc proteins are indeed highly glycosylated. The VNTR region is therefore an excellent candidate for polymorphic glycosylation. Differences in the repeat type and number in the molecular variants could be responsible for differences in glycosylation status.

The polymorphism of Sm PoMucs is evidenced at different levels. The first level is due to differences in repeat number within the VNTR region. This phenomenon was observed previously for numerous mucins

from protozoa and helminths (Hicks *et al.*, 2000; Theodoropoulos *et al.*, 2001), but also for other parasitic molecules displaying a VNTR region. Examples are the SOWgp cell surface proteins from the pathogenic fungi *Coccidioides immitis* (Hung *et al.*, 2000) and different proteins from *Plasmodium*, the circumsporozoite proteins (CSP, (Nussenzweig & Nussenzweig, 1985)) or the merozoite surface proteins (Msp-1 and Msp-2, (Holder, 1988; Smythe *et al.*, 1988)). In the case of *Plasmodium* proteins, the repetitive regions seem to be not glycosylated but they are, as in the case of *Sm* PoMuc, highly polymorphic (Rich *et al.*, 1997; Rich & Ayala, 2000) and VNTR regions are present. Polymorphism generated by VNTR is well documented and new alleles can arise by duplication/deletion of the repeated DNA sequences through recombination involving unequal crossing-over (Rich *et al.*, 1997). Our data suggest that a similar process operates to generate the variability of *Sm* PoMuc. The second level of polymorphism is unique to *Sm* PoMuc. We demonstrated a combinatorial polymorphism specific of the IC strain. In order to understand this phenomenon, *Sm* PoMucs genomic structure analysis is underway.

Expression analysis of *Sm* PoMuc (Roger *et al.*, 2007) strengthens the potential of these molecules as key determinants of the compatibility polymorphism in *S. mansoni*/*B. glabrata* interaction. Indeed, *Sm* PoMuc genes are only transcribed in stages interacting with molluscs with a peak of expression in the mollusc infective stage, the miracidium. Furthermore, peptide signal prediction indicates the possible secretion of these molecules. This hypothesis was strengthened by: *i*) immunolocalization of *Sm* PoMuc in the apical gland, and *ii*) the detection of *Sm* PoMuc in the ES products by western blotting. Our data show that these molecules are released in the immediate environment of the miracidia and Sp1. The apical gland is an organ involved in producing secretions necessary for the successful penetration of miracidia into snail tissue (Dresden *et al.*, 1983; Yoshino *et al.*, 1993) and is known to persist for a number of days after miracidium to sporocyst transformation *in vitro* (Basch & DiConza, 1974; Wippersteg *et al.*, 2002) and *in vivo* (Pan, 1980; Pan, 1996). The gland cells could therefore play a role in infestation and post-infestation process in the first hours following the contact between the host and its parasite. The rapid recognition, encapsulation and killing of the IC strain could be explained by the differences evidenced by our studies and we hypothesize that these differences could be responsible for recognition or non-recognition in the case of the IC and C strains, respectively.

Parasite-derived mucin or mucin-like molecules have been extensively described in different protozoan and helminth parasites, but their function has not yet been completely elucidated. They could have roles in host

recognition, penetration, adhesion and invasion of host cells, immunoprotection, immunomodulation and in the avoidance of host immune processes (Hicks *et al.*, 2000; Theodoropoulos *et al.*, 2001). Among these molecules described in parasites, the mucin-like molecules of *Trypanosoma cruzi* have been extensively studied and share numerous characteristics with *Sm* PoMuc (see (Buscaglia *et al.*, 2006) for review), such as their structure, a high level of glycosylation and polymorphism. A major difference is that these molecules encoded by the TcMUC I and II genes, like other surface components of parasitic protozoa, are typically anchored to the outer phospholipid layer of the plasma membrane by GPI. Nevertheless, secreted mucins have been characterized in helminths like the infective larvae of the parasitic nematode *Toxocara canis* (Loukas *et al.*, 2000). These latter secreted mucins might simply create an immunological smoke-screen by generating antigen-antibody complexes away from the parasite (Marin *et al.*, 1992). An attractive hypothesis could be that *Sm* PoMucs could also create an immunological smoke-screen able to block pattern recognition receptors, thus avoiding recognition and the host immune reaction. However, compatibility polymorphism evidenced in our system could alternatively be based on a lock and key mechanism involving *Sm* PoMucs and host recognition partner molecules. In this case positive recognition of the compatible variable antigen would determine success of the infection. Consequently, to further investigate this hypothesis, the potential partners of *Sm* PoMucs will be investigated using co-immunoprecipitation experiments. Diversified host molecules like FREPS (Adema *et al.*, 1997; Zhang *et al.*, 2004) could be ideal candidates. The data obtained in the two last papers presented here describe novel mucin-like proteins from *S. mansoni*. These *Sm* PoMucs are very attractive candidates to explain the compatibility polymorphism occurring in our system. Studies on gene structure, genomic organization and functional characterization are underway and will shed light on the exact role of these proteins in the infection of invertebrate hosts.

## CONCLUSION AND PERSPECTIVE

The different comparative molecular approaches developed in our different experimental systems have highlighted several promising candidates. Nevertheless, these different approaches have also probably revealed genes that are not key determinants of the compatibility we are interested in. Indeed, several of these genes are probably strain markers co-selected with our key genes of interest. The challenge now is to develop functional studies aiming at the selection of actor genes of the compatibility. This challenge requires the development of different methodologies

and tools not available in our systems. This is the challenge in our research field that is at the hour of genomics (genome sequencing and assembly of *S. mansoni* and *B. glabrata*) and is not yet entered in post-genomic time.

## REFERENCES

- ABRAHAMSON M., ALVAREZ-FERNANDEZ M. & NATHANSON C.M. Cystatins. *Biochem. Soc. Symp.*, 2003, 179-199.
- ADEMA C.M., HERTEL L.A., MILLER R.D. & LOKER E.S. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proceedings of the National Academy of Sciences of USA*, 1997, 94, 8691-8696.
- ATAEV G.L. & COUSTAU C. Cellular response to *Echinostoma caproni* infection in *Biomphalaria glabrata* strains selected for susceptibility/resistance. *Developmental and Comparative Immunology*, 1999, 23, 187-198.
- BASCH P.F. & DI CONZA J.J. The miracidium-sporocyst transition in *Schistosoma mansoni*: surface changes *in vitro* with ultrastructural correlation. *Journal of Parasitology*, 1974, 60, 935-941.
- BOSWELL C.A. & BAYNE C.J. Isolation, characterization and functional assessment of a hemagglutinin from the plasma of *Biomphalaria glabrata*, intermediate host of *Schistosoma mansoni*. *Developmental and Comparative Immunology*, 1984, 8, 559-568.
- BOUCHUT A., COUSTAU C., GOURBAL B. & MITTA G. Compatibility in the *Biomphalaria glabrata/Echinostoma caproni* model: new candidate genes evidenced by a suppressive subtractive hybridization approach. *Parasitology*, 2007, 134, 575-188.
- BOUCHUT A., ROGER E., COUSTAU C., GOURBAL B. & MITTA G. Compatibility in the *Biomphalaria glabrata/Echinostoma caproni* model: potential involvement of adhesion genes. *International Journal for Parasitology*, 2006, 36, 175-184.
- BOUCHUT A., SAUTIERE P.E., COUSTAU C. & MITTA G. Compatibility in the *Biomphalaria glabrata/Echinostoma caproni* model: Potential involvement of proteins from hemocytes revealed by a proteomic approach. *Acta Tropica*, 2006, 98, 234-246.
- BUSCAGLIA C.A., CAMPO V.A., FRASCH A.C. & DI NOIA J.M. *Trypanosoma cruzi* surface mucins: host-dependent coat diversity. *Nature Reviews Microbiology*, 2006, 4, 229-236.
- BUTZKE D., HURWITZ R., THIEDE B., GOEDERT S. & RUDEL T. Cloning and biochemical characterization of APIT, a new L-amino acid oxidase from *Aphysia punctata*. *Toxicon*, 2005, 46, 479-89.
- COMBES C. Selective pressure in host-parasite systems. *Journal de la Société de Biologie*, 2000, 194, 19-23.
- DRESDEN M.H., SUNG C.K. & DEELDER A.M. A monoclonal antibody from infected mice to a *Schistosoma mansoni* egg proteinase. *Journal of Immunology*, 1983, 130, 1-3.
- FUJII N., MINETTI C.A., NAKHASI H.L., CHEN S.W., BARBEHENN E., NUNES P.H. & NGUYEN N.Y. Isolation, cDNA cloning, and characterization of an 18-kDa hemagglutinin and amebocyte aggregation factor from *Limulus polyphemus*. *Journal of Biological Chemistry*, 1992, 267, 22452-22459.
- GRYSEELS B., POLMAN K., CLERINX J. & KESTENS L. Human schistosomiasis. *Lancet*, 2006, 368, 1106-1118.
- HICKS S.J., THEODOROPoulos G., CARRINGTON S.D. & CORFIELD A.P. The role of mucins in host-parasite interactions. Part I- protozoan parasites. *Parasitology Today*, 2000, 16, 476-481.
- HOLDER A.A. The precursor to major merozoite surface antigens: structure and role in immunity. *Progress in Allergy*, 1988, 41, 72-97.
- HUMBERT E. & COUSTAU C. Refractoriness of host haemocytes to parasite immunosuppressive factors as a putative resistance mechanism in the *Biomphalaria glabrata-Echinostoma caproni* system. *Parasitology*, 2001, 122, 651-660.
- HUNG C.Y., AMPEL N.M., CHRISTIAN L., SESHAN K.R. & COLE G.T. A major cell surface antigen of *Coccidioides immitis* which elicits both humoral and cellular immune responses. *Infection and Immunity*, 2000, 68, 584-593.
- IJIMA R., KISUGI J. & YAMAZAKI M. Biopolymers from marine invertebrates. XIV. Antifungal property of Dolabellanin A, a putative self-defense molecule of the sea hare, *Dolabella auricularia*. *Biological Pharmaceutical Bulletin*, 1994, 17, 1144-1146.
- LANGAND J., JOURDANE J., COUSTAU C., DELAY B. & MORAND S. Cost of resistance, expressed as a delayed maturity, detected in the host-parasite system *Biomphalaria glabrata/Echinostoma caproni*. *Heredity*, 1998, 80, 320-325.
- LANGAND J. & MORAND S. Heritable non-susceptibility in an allopatric host-parasite system: *Biomphalaria glabrata* (Mollusca)-*Echinostoma caproni* (Platyhelminth, Digenea). *Journal of Parasitology*, 1998, 84, 739-742.
- LOKER E.S. & ADEMA C.M. Schistosomes, Echinostomes and Snails: comparative Immunobiology. *Parasitology Today*, 1995, 11, 120-124.
- LOUKAS A., HINTZ M., LINDER D., MULLIN N. P., PARKINSON J., TETTEH K.K. & MAIZELS R.M. A family of secreted mucins from the parasitic nematode *Toxocara canis* bears diverse mucin domains but shares similar flanking six-cysteine repeat motifs. *Journal of Biological Chemistry*, 2000, 275, 39600-39607.
- MARIN M.S., PRIETO M., MARTIN J.M., CASAIS R., BOGA J.A. & PARRA F. Identification and expression of a *Fasciola hepatica* gene encoding a gut antigen protein bearing repetitive sequences. *Molecular and Biochemical Parasitology*, 1992, 55, 155-165.
- MICHELSON E.H. & DUBOIS L. Agglutinins and lysins in the molluscan family Planorbidae: a survey of hemolymph, egg-masses, and albumen-gland extracts. *Biological Bulletin*, 1977, 153, 219-227.
- NUSSENZWEIG V. & NUSSENZWEIG R.S. Circumsporozoite proteins of malaria parasites. *Cell*, 1985, 42, 401-403.
- PAN S.C. The fine structure of the miracidium of *Schistosoma mansoni*. *Journal of Invertebrate Pathology*, 1980, 36, 307-372.
- PAN S.C. *Schistosoma mansoni*: the ultrastructure of larval morphogenesis in *Biomphalaria glabrata* and of associated host-parasite interactions. *Japanese Journal of Medical Science and Biology*, 1996, 49, 129-149.

- RICH S.M. & AYALA F.J. Population structure and recent evolution of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of USA*, 2000, 97, 6994-7001.
- RICH S.M., HUDSON R.R. & AYALA F.J. *Plasmodium falciparum* antigenic diversity: evidence of clonal population structure. *Proceedings of the National Academy of Sciences of USA*, 1997, 94, 13040-13045.
- ROGER E., GOURBAL B.E., GRUNAU C., PIERCE R., GALINIER R. & MITTA G. Expression analysis of highly polymorphic mucin proteins (Sm PoMuc) from the parasite *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 2007 (in press).
- ROGER E., MITTA G., MONE Y., BOUCHUT A., ROGNON A., GRUNAU C., BOISSIER J., THERON A. & GOURBAL B.E. Molecular determinants of compatibility polymorphism in the *Biomphalaria glabrata*/*Schistosoma mansoni* model: new candidates identified by a global comparative proteomic approach. *Molecular and Biochemical Parasitology*, 2007 (in press).
- SCHUTZE J., SKOROKHOD A., MULLER I.M. & MULLER W.E. Molecular evolution of the metazoan extracellular matrix: cloning and expression of structural proteins from the demospogones *Suberites domuncula* and *Geodia cydonium*. *Journal of Molecular Evolution*, 2001, 53, 402-415.
- SMYTHE J.A., COPPEL R.L., BROWN G.V., RAMASAMY R., KEMP D.J. & ANDERS R.F. Identification of two integral membrane proteins of *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences of USA*, 1988, 85, 5195-5199.
- STEIN P.C. & BASCH P.F. Purification and binding properties of hemagglutinin from *Biomphalaria glabrata*. *Journal of Invertebrate Pathology*, 1979, 33, 10-18.
- TAKAMATSU N., SHIBA T., MURAMOTO K. & KAMIYA H. Molecular cloning of the defense factor in the albumen gland of the sea hare *Aplysia kurodai*. *FEBS Letters*, 1995, 377, 373-376.
- THEODOROPOULOS G., HICKS S.J., CORFIELD A.P., MILLER B.G. & CARRINGTON S.D. The role of mucins in host-parasite interactions: Part II - helminth parasites. *Trends in Parasitology*, 2001, 17, 130-135.
- THERON A. & COUSTAU C. Are *Biomphalaria* snails resistant to *Schistosoma mansoni*? *Journal of Helminthology*, 2005, 79, 187-191.
- VERGOTE D., BOUCHUT A., SAUTIERE P.E., ROGER E., GALINIER R., ROGNON A., COUSTAU C., SALZET M. & MITTA G. Characterisation of proteins differentially present in the plasma of *Biomphalaria glabrata* susceptible or resistant to *Echinostoma caproni*. *International Journal for Parasitology*, 2005, 35, 215-224.
- WIJDENES J., VAN ELK R. & JOOSSE J. Effects of two gonadotropic hormones on polysaccharide synthesis in the albumen gland of *Lymnaea stagnalis*, studied with the organ culture technique. *General and Comparative Endocrinology*, 1983, 51, 263-271.
- WIPPERSTEG V., KAPP K., KUNZ W. & GREVELDING C.G. Characterisation of the cysteine protease ER60 in transgenic *Schistosoma mansoni* larvae. *International Journal for Parasitology*, 2002, 32, 1219-1224.
- XU B., HAGGLUND P., STALBRAND H. & JANSON J.C. Endo-beta-1,4-Mannanases from blue mussel, *Mytilus edulis*: purification, characterization, and mode of action. *Journal of Biotechnology*, 2002, 92, 267-277.
- YOSHINO T.P., LODES M.J., REGE A.A. & CHAPPELL C.L. Proteinase activity in miracidia, transformation excretory-secretory products, and primary sporocysts of *Schistosoma mansoni*. *Journal of Parasitology*, 1993, 79, 23-31.
- ZHANG S.M., ADEMA C.M., KEPLER T.B. & LOKER E.S. Diversification of Ig superfamily genes in an invertebrate. *Science*, 2004, 305, 251-254.



## Excretory–secretory proteome of larval *Schistosoma mansoni* and *Echinostoma caproni*, two parasites of *Biomphalaria glabrata*

François Guillou<sup>a,1</sup>, Emmanuel Roger<sup>a,1</sup>, Yves Moné<sup>a</sup>, Anne Rognon<sup>a</sup>, Christoph Grunau<sup>a</sup>, André Théron<sup>a</sup>, Guillaume Mitta<sup>a</sup>, Christine Coustau<sup>b</sup>, Benjamin E.F. Gourbal<sup>a,\*</sup>

<sup>a</sup> Parasitologie Fonctionnelle et Evolutive, UMR 5244, CNRS Université de Perpignan, 52 Ave Paul Alduy, 66860 Perpignan Cedex, France

<sup>b</sup> U547 INSERM, Institut Pasteur de Lille, 1 rue du Pr. Calmette, 59019 Lille Cedex, France

Received 7 March 2007; received in revised form 22 May 2007; accepted 22 May 2007

Available online 29 May 2007

### Abstract

*Schistosoma mansoni* and *Echinostoma caproni* are two trematode species that use different strategies (mimicry and immunosuppression, respectively) to interfere with the snail innate immune system. Parasites excretory–secretory (ES) products have been shown to play a key role in these host–parasite immune interactions. However, they remain largely uncharacterized in larval trematodes. We developed a global proteomic approach to characterize the ES proteome of *S. mansoni* and *E. caproni* primary sporocysts. In ES products of both parasites, we found proteins involved in reactive oxygen species scavenging, glycolysis, signalling or calcium binding (superoxide dismutase Cu/Zn; glutathione *S*-transferase; aldo-keto-reductase; triose-phosphate isomerase; glyceraldehyde-3-phosphate dehydrogenase; aldolase, enolase, MICAL-like, calreticulin). According to their predicted functions, we propose a model in which these proteins (i) are involved in antioxidant activity, (ii) prevent hemocyte encapsulation process or (iii) favor invasion and migration of sporocysts in host tissues. These results suggest that *S. mansoni* and *E. caproni* sporocysts develop a strong immune protection during the first hours of infection giving them enough time to build up a long lasting immune evasion strategy relying on molecular mimicry or immunosuppression, respectively.

© 2007 Elsevier B.V. All rights reserved.

**Keywords:** *Schistosoma mansoni*; *Echinostoma caproni*; Sporocyst; *Biomphalaria glabrata*; Excretory–secretory proteome; Immune protection

### 1. Introduction

Digenetic trematodes have complex life cycles. At their adult stage, worms parasitize vertebrate definitive hosts, whereas larval stages (sporocysts and/or rediae) always develop in mollusk intermediate hosts. For each developmental stage, parasites need to penetrate into the host, develop, multiply sexually or asexually and finally propagate in the environment to complete their life cycle [1,2]. Parasites therefore face many challenges such as obtaining enough energy to maintain, grow and multiply, and being able to evade the internal defense system of the host [3–5].

The mollusk immune response is mediated by humoral and/or cell components [6,7]. The major effectors of immunity in mollusks are the circulating hemocytes. These phagocytic cells are able to distinguish non-self material and to encapsulate and eliminate invaders such as parasite sporocysts [3,8].

*Schistosoma mansoni* and *Echinostoma caproni* are two trematode parasites that infect the same intermediate host, the snail *Biomphalaria glabrata*. In order to evade the mollusks immune response, these trematode species use different strategies.

The evasion strategy of *S. mansoni* is not well understood. Some authors described the mimicry ability of *S. mansoni* based on antigenic similarities between parasite components and host tissues [9–12]. In contrast to *S. mansoni*, echinostomes interfere with the snail immune system in an immunosuppressive way [13–15]. Various studies showed that hemocytes exposed to echinostome ES products lose several defense functions such as adhesion, phagocytosis and encapsulation, as well as cytotoxic activity [16–20]. Moreover, in co-infection experiments,

\* Corresponding author at: Parasitologie Fonctionnelle et Evolutive, UMR 5244, CNRS, Biologie & Ecologie Tropicale et Méditerranéenne, Université de Perpignan, 52 Ave Paul Alduy, 66860 Perpignan Cedex, France. Tel.: +33 4 68 66 20 50; fax: +33 4 68 66 22 81.

E-mail address: [benjamin.gourbal@univ-perp.fr](mailto:benjamin.gourbal@univ-perp.fr) (B.E.F. Gourbal).

<sup>1</sup> The first two authors have participated in this work in the same substantive way and thus should be considered equally.



the immune suppression activity of echinostome permits the survival of schistosomes in *Schistosoma*-resistant host strains [13].

The evasion strategies of trematodes are related to their ability to produce excretory/secretory products. These contain immuno-regulatory molecules that can interfere with host innate immunity [10,11]. While in the ES of adult worms a few molecules have been characterized that interact with the vertebrate host [3,18,20–23], until now less is known about the ES for invertebrate infection.

We undertook a two-dimensional gel electrophoresis and mass spectrometry approach to (i) characterize the ES proteome of both parasites, (ii) gain insight into the molecular basis of the sporocysts immune evasive strategies during the first hour of interaction, and (iii) understand the molecular cross-talk between *S. mansoni* or *E. caproni* sporocysts and the snail host *B. glabrata*. Our data revealed for the first time that the ES products recovered during the first 24 h in culture contain families of molecules that have the same function in both parasites. These molecules are very likely to be involved in the immune protective strategy.

## 2. Materials and methods

### 2.1. Parasites excretory–secretory products

The parasite *E. caproni* originate from Egypt [24], and a Brazilian (Bra) strain of *S. mansoni* were maintained in Bg.Bra snails, mice (SWISS OF1 stock) and hamsters (*Mesocricetus auratus*) as described previously [20,25–27].

Miracidia from *S. mansoni* Bra were hatched from eggs axenically recovered from 50-day infected hamster livers [25]. Miracidia from *E. caproni* were recovered from eggs laid *in vitro* according to a previously described procedure [20]. For both parasites, after three washes in sterile pond water containing an antibiotic/antimycotic mixture (penicillin 100 units/ml, streptomycin 0.1 mg/ml, amphotericin 0.025 µg/ml; Sigma), miracidia were placed into 50 ml tissue-culture plates containing sterile Chernin's balanced salt solution (CBSS) [28] and the antibiotic/antimycotic mixture at 26 °C under atmospheric conditions. Full transformation of miracidia to primary (mother) sporocysts (Sp1) occurred within the first 24 h in culture [22]. After 24 h, the transformed sporocysts were spun down ( $600 \times g$  for 5 min) in a 15 ml centrifuge tube to pellet the Sp1. The supernatant fraction containing the ES products was recovered and filtered through a 0.22 µm membrane. Both the ES products and the sporocyst pellets were kept at –80 °C until use.

### 2.2. Protein separation, detection and identification

ES products were precipitated using half volume of acetone containing 50% trichloroacetic acid and suspended in same volume of 2D lysis buffer (8M urea, 40 mM Tris, 4% CHAPS, 60 mM DTT). ES products were recovered from 120,700 and 425,000 larvae for *S. mansoni* and *E. caproni*, respectively. ES products recovered from 120,700 sporocysts for each species were used for the proteomics analysis. For the first dimension,

ES products were loaded on 17 cm Ready IPG Strips, pH 3–10 nonlinear gradient (Bio-Rad).

Total proteome extraction was realized using 10,000 Sp1 incubated during 2 h at room temperature in 2D lyses buffer. Cell lysate was centrifuged at  $15,000 \times g$  to pellet remaining cells and nuclear material. Protein concentration in the supernatant was assayed using the 2D Quant Kit (Amersham), 80 µg of protein were loaded on the first dimension, on 17 cm Ready Strips IPG Strips, pH 3–10 nonlinear gradient (Bio-Rad). Isoelectrofocusing (IEF) was performed as previously described [29] with IEF raising gradually until 8000 V and running until 90,000 Vh at 20 °C.

Proteins were separated in second dimension SDS-PAGE and proteins were visualized by Silver staining according to a method compatible with mass spectrometry analysis [30]. 2D gels were scanned using a densitometer (GS-800 Calibrated Densitometer, Bio-Rad) and the pictures were analyzed using the PDQuest 7.4.0 software (Bio-Rad). Gel plugs containing the proteins of interest were excised from the gel, processed and characterized by nanoscale capillary liquid chromatography–tandem mass spectrometry (nanoLC–MS/MS) as described previously [29,31]. NanoLC–MS/MS analysis of the digested proteins was performed using a capillary LC system coupled to a LCQ Deca XP<sup>+</sup> Thermo-electron tandem mass spectrometer.

The mass data recorded during nanoLC–MS/MS analysis were processed and converted into “Merge”. DTA peak lists format prior to searching with the search engine MASCOT (<http://www.matrixscience.com/>) against protein or nucleotide sequence databases, available through the National Centre for Biotechnology Information website ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). One or two missed cleavages per peptide were allowed and some variable modifications were taken into account in the search such as carbamidomethylation of cysteine, oxidation of methionine, histidine or tryptophan, and phosphorylation of serine, threonine or tyrosine. Searches were performed without constraining protein molecular weight or isoelectric point. In this study, we validated an identification when the protein was identified by at least two peptides with at least an overall MASCOT score greater than 50 ( $p < 0.05$ ) [32].

### 2.3. Protein analysis

When possible, the sequence of the gene encoding the proteins of interest was identified by querying peptidic amino acid sequences against database (GenBank dbEST of *S. mansoni* 162 506 sequences, of *E. caproni* 298 sequences) using the TBLASTN (<http://www.ncbi.nlm.nih.gov/blast/>). Sequences were imported in fasta format and were then assembled in contigs using the Sequencher<sup>TM</sup> 4.5 software (Gene Codes Corporation). Consensus sequences were used to detect the open reading frame (ORF) and the possible isoforms of the corresponding protein.

Primary structure analysis and specific domain searches were performed using the ProtFun 2.2 and SecretomeP 2.0 programs (<http://www.cbs.dtu.dk/services/>). Expected molecular weights ( $M_w$ ) and isoelectric points ( $pI$ ) were calculated using

Table 1  
Proteins identified in the ES proteome of *Schistosoma mansoni* primary sporocyst

Function	Spot #	Protein name	Species	Accession # <sup>a</sup>	$M_w$ (kDa) (exp./obs.) <sup>b</sup>	pI (exp./obs.) <sup>b</sup>	Coverage (%)	Peptides #	Peptides sequence data	Mascot score
Glycolytic enzymes	12	Triose phosphate isomerase	<i>S. mansoni</i>	P48501	28/28	7.63/7.6	5	1	IYGGSVTAANCK	78
	13	Triose phosphate isomerase	<i>S. mansoni</i>	P48501	28/28	7.63/8	27	4	GAFTGEISPAAMK IYGGSVTAANCK NIFGEDELAIEK	306
Antioxidant enzymes	14	Triose phosphate isomerase	<i>S. mansoni</i>	P48501	28/28	7.63/8.5	13	2	VVYAYEPVWAIGTGK GAFTGEISPAAMK	152
	35	Glyceraldehyde-3-phosphate dehydrogenase	<i>S. mansoni</i>	P20287	36/38	8.16/8.6	7	2	IYGGSVTAANCK AGISLNNNFVK	43
	36	Glyceraldehyde-3-phosphate dehydrogenase	<i>S. mansoni</i>	P20287	36/38	8.16/9	7	3	GAMQNIIPASTGAAK AGISLNNNFVK GAMQNIIPASTGAAK, GAMQNIIPASTGAAK <sup>c</sup>	74
Antioxidant enzymes	22	Superoxide dismutase Cu/Zn	<i>S. mansoni</i>	AAC14467	16/16	6.09/6.2	44	7	AVCVMTGTAGVK <sup>d</sup> AVCVMTGTAGVK <sup>d,e</sup> AVCVMTGTAGVK <sup>d,e</sup> TMVHENEEDLGR TMVHENEEDLGR TMVHENEEDLGR <sup>c</sup> TMVHENEEDLGR <sup>c</sup> FTQETDNGPVHVHAEFSGLK	251
	23	Superoxide dismutase Cu/Zn	<i>S. mansoni</i>	AAC14467	16/16	6.09/6.3	44	7	HVGDIGNYVAGADGNVYNAIDK AVCVMTGTAGVK <sup>d</sup> AVCVMTGTAGVK <sup>d,e</sup> TMVHENEEDLGR TMVHENEEDLGR TMVHENEEDLGR TMVHENEEDLGR <sup>c</sup> TMVHENEEDLGR <sup>c</sup> FTQETDNGPVHVHAEFSGLK	252
Antioxidant enzymes	24	Superoxide dismutase Cu/Zn	<i>S. mansoni</i>	AAC14467	16/16.5	6.09/6.4	38	6	HVGDIGNYVAGADGNVYNAIDK LAGVYGLAAE <sup>f</sup> AVCVMTGTAGVK <sup>d</sup> AVCVMTGTAGVK <sup>d,e</sup> TMVHENEEDLGR TMVHENEEDLGR TMVHENEEDLGR <sup>c</sup> TMVHENEEDLGR <sup>c</sup> FTQETDNGPVHVHAEFSGLK	211
	33	Glutathione S-transferase	<i>S. japonicum</i>	AAA29888	25/25	7.68/8.2	4	2	HVGDIGNYVAGADGNVYNAIDK LQSMAIR MFEDRLCHK	55
Calcium binding	11	Calreticulin	<i>S. mansoni</i>	Q06814	45/51	4.69/4.5	6	1	SPYDPIEDLGLK	55
Other functions	16	Similar to trematode hemoglobin	<i>S. japonicum</i>	AAP06216	-/22	-/7.8	4	2	IFQDSLPIFIK IFNLLKIQQNK	50
	21	Histone H4	<i>S. mansoni</i>	AAG25601	-/17	-/6.1	24	2	DNIQGITPAIR TVTAMDVYVYALK <sup>c</sup>	105

<sup>a</sup> Accession number in Genbank database.  
<sup>b</sup> Expected and observed molecular weight ( $M_w$ ) and isoelectric point (pI) for each identification.  
<sup>c</sup> Oxidation of methionine.  
<sup>d</sup> Carbamidomethylation of cysteine.

Table 2  
Proteins identified in the ES proteome of *Echinostoma caproni* primary sporocyst

Function	Spot #	Protein name	Species	Accession # <sup>a</sup>	M <sub>w</sub> (kDa) (exp./obs.) <sup>b</sup>	pI (exp./obs.) <sup>b</sup>	Coverage (%)	Peptides #	Peptides: sequence data	Mascot score
Glycolytic enzymes	32	Triose phosphate isomerase	<i>E. michelini</i>	AA06239	~30	~6.5	NA	1	IYGGSVTAANCK <sup>d</sup>	53
	33	Triose phosphate isomerase	<i>E. michelini</i>	AA06239	~30	~6.9	NA	1	IYGGSVTAANCK <sup>d</sup>	82
	34	Triose phosphate isomerase	<i>O. turkestanicum</i>	AAZ57433	~30	~7.6	12	3	MSGSRK KFFVGGNWK	141
Aldolase	19	Aldolase	<i>S. japonicum</i>	AAZ27354	~41	~8	NA	2	IYGGSVTAANCK <sup>d</sup> GILAADESTAIMGK	88
	20	Fbp aldolase	<i>S. mansoni</i>	AF026805	~41	~8.5	10	4	GILAADESTAIMGK <sup>c</sup> NIIPGIK VTEQVLAIFYK	199
Fbp-aldolase	21	Fbp-aldolase	<i>S. mansoni</i>	AF026805	~41	~9	10	3	GILAADESTAIMGK NIIPGIK VTEQVLAIFYK	117
	21	NADH dehydrogenase chain 5	<i>D. blaphoides</i>	BAA28956	~41	~9	NA	3	GILAADESTAIMGK <sup>c</sup> IALSTLSQLGLMK <sup>e</sup> ALLEMCAGAIHNLK <sup>c</sup>	62
Phosphoenol pyruvate hydratase	71	Phosphoenol pyruvate hydratase	<i>E. hepatica</i>	A53665	~46	~6.3	16	5	KIALSTLSQLGLMK <sup>e</sup> YDLDFK GNPTEVDVTTAK IQVGGDLIVTNPLR SGETEDNHADLVVGLR AAVPSGASTGVHEALELR	328
	60	Superoxide dismutase Cu/Zn	<i>F. hepatica</i>	AA030361	~17	~5.9	10	2	AMVVHENEDDLGR AMVVHENEDDLGR <sup>c</sup>	52
Superoxide dismutase Cu/Zn	61	Superoxide dismutase Cu/Zn	<i>F. hepatica</i>	AA030361	~17	~6.1	15	3	AMVVHENEDDLGR AMVVHENEDDLGR <sup>c</sup> AMVVHENEDDLGR <sup>c</sup>	123
	62	Superoxide dismutase Cu/Zn	<i>F. hepatica</i>	AA030361	~17	~6.15	15	4	AMVVHENEDDLGR AMVVHENEDDLGR AMVVHENEDDLGR <sup>c</sup> AMVVHENEDDLGR <sup>c</sup>	110
Superoxide dismutase Cu/Zn	63	Superoxide dismutase Cu/Zn	<i>F. hepatica</i>	AA030361	~17	~6.2	15	4	AMVVHENEDDLGR AMVVHENEDDLGR AMVVHENEDDLGR <sup>c</sup> AMVVHENEDDLGR <sup>c</sup>	99
	23	Aldo-keto-reductase	<i>G. morstans</i>	BX568329	~35	~6.15	NA	3	MVVHENEDDLGRGGHEQSK EDLEVTSK REDLFTSK SIGLSNFNKR	99 148
Signalling molecule	39	MICAL-like	<i>A. gambiae</i>	EAL40681	~28	~9.3	NA	2	EKEEALITR TDSKKEKEEALITR	53
	15	Actin	<i>T. spiralis</i>	AA184892	~41	~5.0	NA	2	GYSFTTAAER CPESLFQPSFLGMESCQIHE	54
Actin 1	35	Actin 1	<i>L. obitica</i>	AAV91408	~32	~8.2	NA	5	DLTDDYLMK GYSFTTAAER SYELPDQGVITIGNER VAPEEHPVLLTEAPLNPK KDLVANTVMSGGTTMYPGIADR	196
	51	Acetyl coA transferase	<i>A. franciscana</i>	AA068833	~19	~4.8	NA	3	EGLK NVTVCHHIEGK KVLPGKIVGSFCIGSK	51

Table 2 (Continued)

Function	Spot #	Protein name	Species	Accession # <sup>a</sup>	M <sub>w</sub> (kDa) (exp./obs.) <sup>b</sup>	pI (exp./obs.) <sup>b</sup>	Coverage (%)	Peptides #	Peptides: sequence data	Mascot score
	57	RNA recognition motif	<i>A. gambiae</i>	EAL41325	-124	-11.7	NA	2	MREHIQQIR KSVFRDILWSMK <sup>c</sup>	76
	58	Histone H1	<i>C. doratals</i>	AAA63154	-18	-5.8	11	2	AEKPKSDKPK <sup>c</sup> FVA-AQYKYDVQK	54
	59	Ubiquinone cyt C chaperone	<i>S. japonicum</i>	AAW24871	-17.5	-6.1	7	2	MNLQMLR <sup>c</sup> NVIPLSTR <sup>c</sup>	55
	8	Hypothetical protein	<i>C. brigssae</i>	CBG23420	-68	-5.3	NA	2	KEGLKLIK GKEATSDLK <sup>c</sup>	57
	13	Hypothetical protein	<i>C. brigssae</i>	CBG23411	-55	-4.6	NA	3	SSGSK <sup>c</sup> LKITELEIK <sup>c</sup> TYGGPGTSSSGR <sup>c</sup>	58
	1 + 1'	Hypothetical protein	<i>C. brigssae</i>	CBG23411	-90	-5.3	NA	2	LKITELEIK <sup>c</sup> RASHLEEYAMQQRV <sup>c</sup>	59

NA: not available.

<sup>a</sup> Accession number in Genbank database.

<sup>b</sup> Expected and observed molecular weight (M<sub>w</sub>) and isoelectric point (pI) for each identification.

<sup>c</sup> Oxidation of methionine.

<sup>d</sup> Carbamidomethylation of cysteine.

<sup>e</sup> Phosphorylation of serine and/or threonine.

the Scansite program ([http://scansite.mit.edu/calc\\_mw\\_pi.html](http://scansite.mit.edu/calc_mw_pi.html)) (see Tables 1 and 2).

### 3. Results

#### 3.1. Excretory–secretory proteome of *S. mansoni* and *E. caproni* primary sporocysts

In the proteomic approach developed in this study we compared the excretory–secretory (ES) proteome to the whole proteome profile of sporocysts (Sp1) for both parasites. To our knowledge, such a comparison of the ES proteome with the whole proteome electrophoresis profiles was never done. It is of particular importance to show the absence of contamination of the ES by non-secreted proteins like membrane or soluble proteins. In our study, most proteins of the Sp1 proteome were not present in the ES proteome, and proteins were strongly enriched in ES extracts (Fig. 1).

The most abundant spots of the ES proteome (25 and 71 for *S. mansoni* and *E. caproni*, respectively) were analyzed by nanoLC–MS/MS (Fig. 1C and F). After comparing the experimental peptide sequence tags to databases, we found identifications for 13 spots in *S. mansoni* and 23 for *E. caproni* (Tables 1 and 2). For both parasites, names of proteins identified in databases, species and Genbank accession numbers are shown in Tables 1 and 2. The number and sequence of identified peptides, the Mascot score and the coverage of the protein are also reported. When possible, the molecular weight (M<sub>w</sub>) and the isoelectric point (pI) expected values were calculated using the Scansite program and these values were compared to the observed values on the gel.

As shown in Fig. 1 and Tables 1 and 2, SOD and TPI (for both parasites) and GAPDH (for *S. mansoni*) were present in more than one position in the gel. Shifts in pI or molecular weight values could be due to the presence of isoforms of the protein or to post-translational modifications. For *E. caproni*, the restricted number of GenBank dbEST sequences (298) did not allow to answer this question. For *S. mansoni* more than 160,000 EST sequences were available in GenBank and allowed further analysis. All *S. mansoni* express sequence tags (EST) corresponding to these proteins were imported and assembled separately using the Sequencher™ 4.5 software. For *S. mansoni* SOD, TPI and GAPDH, respectively, all imported EST formed a single contig and the deduced consensus sequence indicate that shifts in pI values observed are not related to the presence of isoforms. Thus, pI shifts are due to post-translational modifications (such as glycosylation) resulting in a basification of the three proteins (Fig. 1 and Table 1).

Among proteins identified for both parasites and listed in Tables 1 and 2, the first functional group corresponds to proteins involved in metabolism functions, mainly glycolysis. The second group corresponds to functions involved in defense pathways with proteins related to oxidative stress scavengers. For *S. mansoni* we found a third group of proteins involved in calcium binding. Finally, other groups, less abundant, were related to amino acid and mitochondrial metabolism, cell structure or are of unknown function.

### 3.1.1. *S. mansoni* excretory/secretory proteome

Among the identified proteins, several are particularly important in the context of host–parasite interactions. Glycolytic enzymes, antioxidant enzymes and calcium binding proteins (Table 1) are therefore described in more detail in the following paragraphs.

### 3.1.2. In the first group, two glycolytic enzymes were characterized in the ES proteome of *S. mansoni*

Triose phosphate isomerase (TPI) (spots 12–14) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (spots 35, 36) (Fig. 1C). TPI is encoded by a single copy gene [33]. This was confirmed by database analysis (ESTs and genome). The alignment of the protein ORF with TPis from

various other organisms permitted the identification of amino acids involved in the active site of the protein (data not shown). All TPI specific domains are present in the *S. mansoni* TPI [34]. The SecretomeP 2.0 analysis predicted the secretion of the TPI through non-classical pathways (NN-score >0.5).

Regarding GAPDH (spots 35, 36) (Fig. 1C), alignment of the protein ORF with GAPDHs from various other organisms allowed us to identify amino acids involved in the catalytic domain, the cysteine 153 of the active site, and the domains associated with the binding of NAD<sup>+</sup> [35] (data not shown). SecretomeP 2.0 analysis did not find evidence of a signal peptide of secretion or the non-classical secretion of the GAPDH (NN-score <0.5).

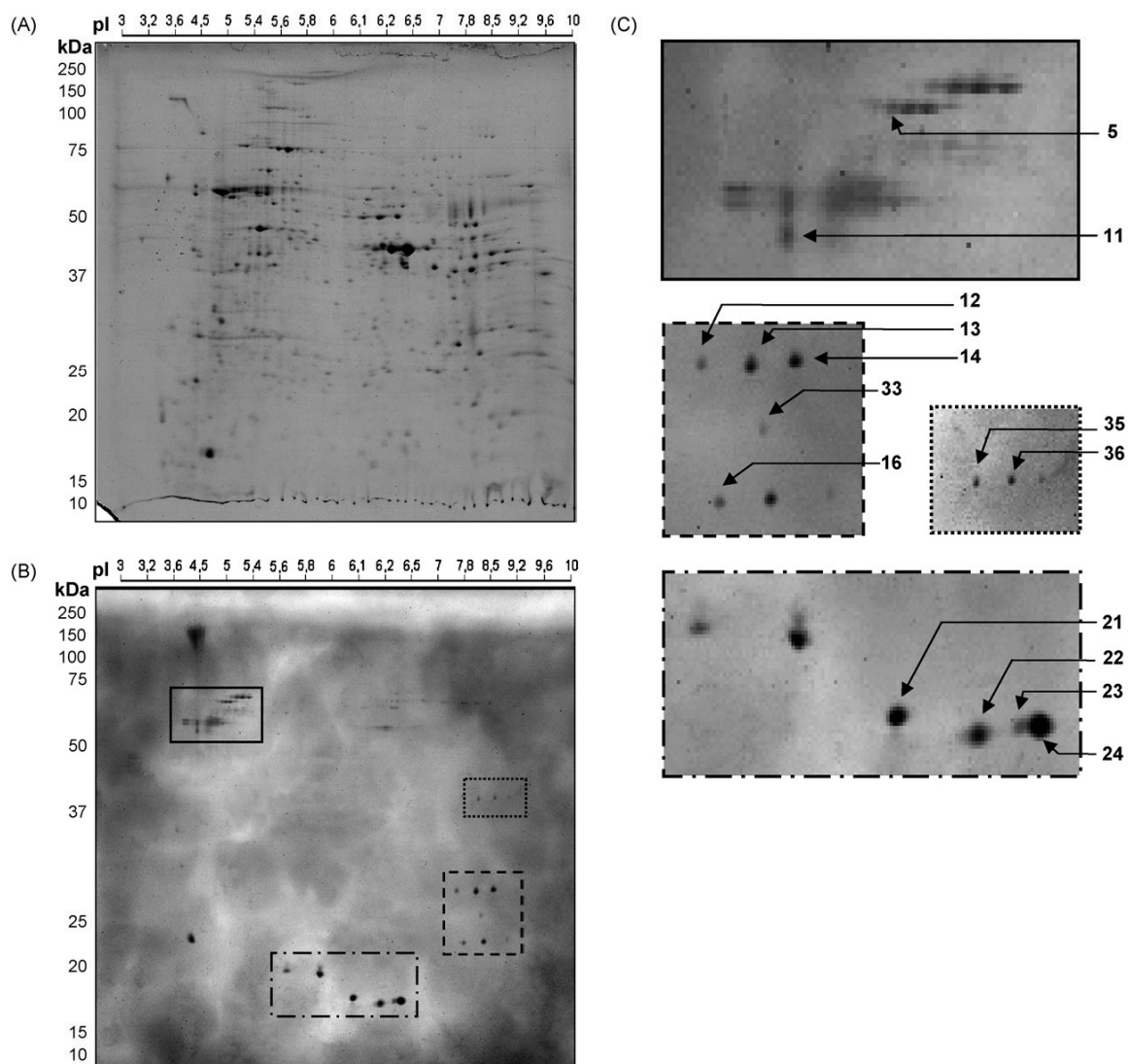


Fig. 1. Proteomic profiles of the total proteome (A) and excretory–secretory proteome (B) of *Schistosoma mansoni* primary sporocyst (C) high magnification of gel area showing the excretory–secretory proteome spots characterized by mass spectrometry (see Table 1). Proteomic profiles of the total proteome (D) and excretory–secretory proteome (E) of *Echinostoma caproni* primary sporocyst (F) high magnification of gel area showing the excretory–secretory proteome spots characterized by mass spectrometry (see Table 2). *x*-Axis is the isoelectric point scale (pI). *y*-Axis is the molecular weight scale in kDa. Proteins were focused using a 3–10 nonlinear pH range as the first dimension and a 12% polyacrylamide gel electrophoresis (SDS-PAGE) as the second dimension.

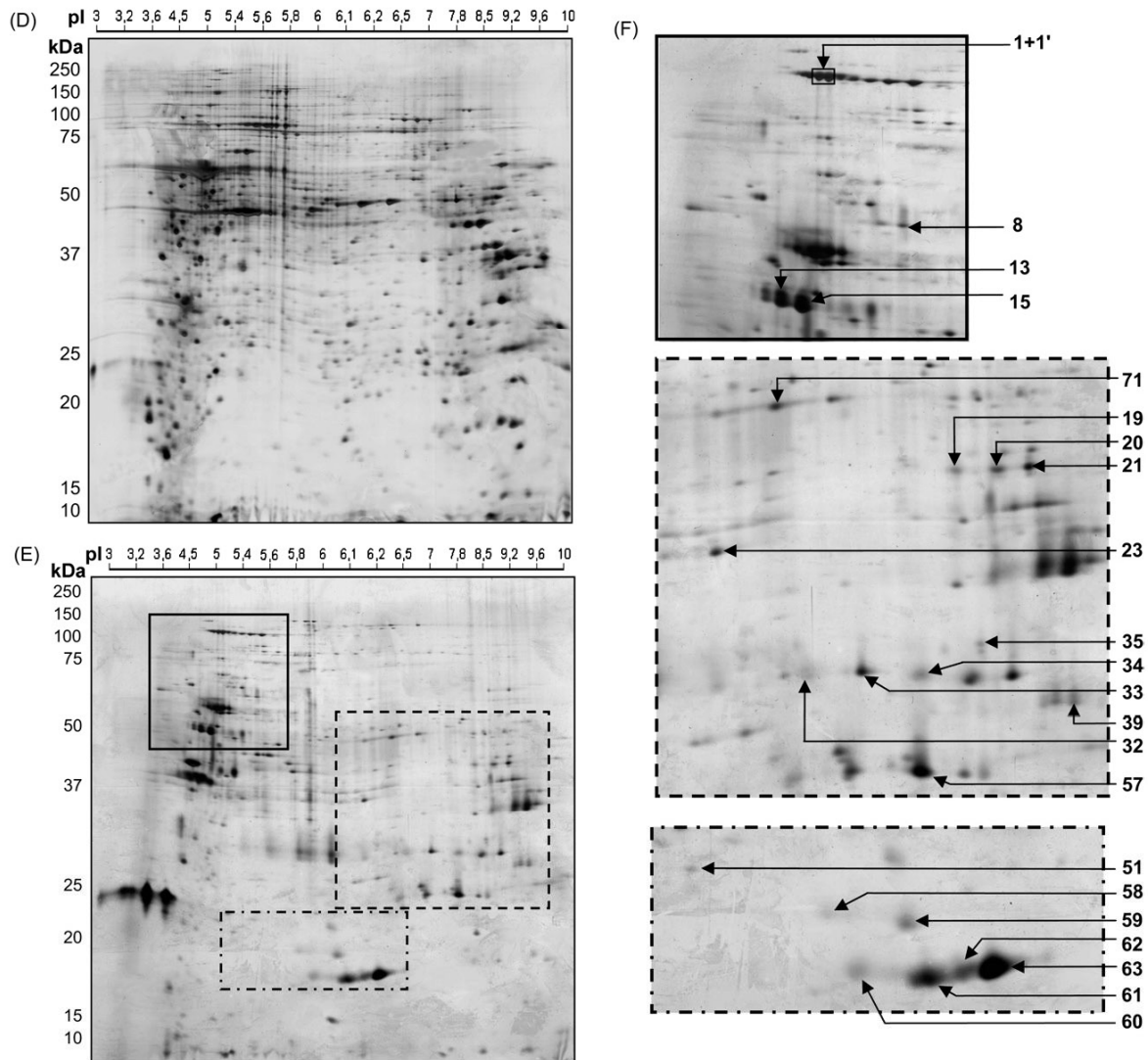


Fig. 1. (Continued).

TPI and GAPDH are two enzymes involved in two related steps of the Embden–Meyerhof–Parnas pathway. Dihydroxyacetone phosphate is converted to glyceraldehyde 3-phosphate by an isomerization catalyzed by TPI. In the subsequent reaction the glyceraldehyde 3-phosphate generates a high-potential phosphorylated compound, 1,3-bisphosphoglycerate. This compound is formed by the action of the GAPDH. The reaction produces NADH.

Superoxide dismutase Cu/Zn (SOD) (spots 22–24) and glutathione *S*-transferase (GST) (spot 33) (Fig. 1C) are involved in the detoxification of oxidative stress. Alignment of the SOD ORF with SODs from various other organisms allowed for the identification of amino acids involved in the active site of the protein (data not shown). Amino acids involved in the tertiary structure of the protein (C57, C146), in copper/zinc binding (H45, H47, H62, H70, H79, D82, H119) and in maintaining active site three-dimensional structure (G43, G60, P65, G81,

G137, G140) [36] are present. The SecretomeP 2.0 analysis predicted the non-classical secretion of the SOD (NN-score > 0.5). Regarding GST (spot 33) (Fig. 1C), the calculated monoisotopic mass of 25 kDa and the expected *pI* of 7.68 were in good agreement with the position of the protein in the gel (Fig. 1B, Table 1). Alignment of the protein ORF with GSTs from various other organisms permitted identification of the amino acids involved in the glutathione binding (Y7, W8, W46, N59, L60) and the amino acids involved in the active site structure (Y7, G12) [37] (data not shown). The presence of a signal peptide of secretion or the non-classical secretion of the GST could not be predicted by SecretomeP 2.0 analysis (NN-score < 0.5). SOD and GST were both involved in oxidative stress scavenging and their specific actions will be described in the discussion of the paper.

Finally, one calreticulin (spot 11) (Fig. 1C) was characterized. The protein is involved in chaperoning and regulation of

Ca(2+) homeostasis. Calreticulin affects more than 40 cellular functions [38]. Calreticulins are highly versatile lectin-like chaperones that participate in the synthesis of a variety of molecules including ion channels, surface receptors, integrins and transporters. The calreticulin identified here had a calculated monoisotopic mass of 45 kDa and an expected pI of 4.69. This is in good agreement with the position of the protein in the gel (Table 1). In order to obtain the complete coding sequence of the calreticulin encoding gene, all the *S. mansoni* ESTs corresponding to this protein were imported and assembled using the Sequencher™ 4.5 software. All calreticulin specific domains [38] were present in this *S. mansoni* calreticulin [39,40]. The SecretomeP 2.0 analysis predicted the presence of a signal peptide of secretion. This is in agreement with our find that the parasite actively secretes a functional calreticulin.

### 3.1.3. *E. caproni* excretory/secretory proteome

Among the proteins identified in the ES proteome of *E. caproni* sporocysts, three groups of proteins have particularly attracted our attention. One is implicated in the glycolysis pathway, the second is implicated in the detoxification of oxidative stress and the third in a signaling pathway (Table 2).

The first group consists of three glycolytic enzymes that intervene in the Embden–Meyerhof–Parnas pathway.

- (i) Fructose-1,6-bisphosphate aldolase (FBPA) (spots 19–21) (Fig. 1F) catalyses the cleavage and generation of 3-carbon units from the 6-carbon sugar. This enzyme catalyzes the formation of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate from fructose-1,6-bisphosphate. The glyceraldehyde 3-phosphate generated in this reaction can be metabolized in the following steps of the glycolytic pathway, but dihydroxyacetone phosphate cannot. To be available for reaction, dihydroxyacetone phosphate needs to be isomerized into glyceraldehyde 3-phosphate.
- (ii) Triose-phosphate isomerase (TPI) (spots 32–34) (Fig. 1F) catalyzes this isomerization and follows the FBPA during glycolysis.
- (iii) Enolase or phosphopyruvate hydratase (spot 71) (Fig. 1F) intervenes at the end of glycolysis to produce phosphoenolpyruvate from a dehydration of 2-phosphoglycerate. Solely a partial sequence of the enolase was known for *E. caproni*. To complete the partial sequences (Genbank accession numbers U85825 and AF274056), we characterized the complete nucleotide sequence of this protein by Racc PCR on cDNA (Genbank accession number DQ869009). All domains and residues essential to the enzymatic activity were present in *E. caproni* enolase [41,42].

In the second group, we describe proteins implicated in the detoxification of oxidative stress:

- (i) Superoxide dismutase Cu/Zn (SOD) (spots 60–63) (Fig. 1F) are considered as reactive oxygen species (ROS) detoxifying enzymes. They catalyze a reduction in hydrogen peroxide and molecular oxygen. Characterized for many helminthes, this enzyme is well known to be secreted by parasites [43].

- (ii) Aldo-keto reductase (AKR) (spot 23) (Fig. 1F). The classification of the AKR superfamily is based on a nonexclusive preference for a specific substrate [44]. The AKR identified here is more similar to AKR in the first family: aldose reductase NADPH dependant, catalyzing the reduction of carbonyls compounds (e.g., aldehyde, carbohydrates, polycyclic aromatic hydrocarbons). They are associated particularly with the detoxification of aldehydes compounds [45–47].

The third group includes MICAL-like protein (spot 39) (Fig. 1F), a signaling molecule involved in regulation of axon steering [48]. MICAL can have other activities not well described yet [49].

## 4. Discussion

Parasitic helminths are known to suppress or modulate the immune response of their vertebrate hosts [5,23,50]. But only recently excretory–secretory (ES) products of important helminths such as schistosomes or echinostomes have been investigated at the molecular level. These characterizations were done almost exclusively for the adult stages and were so far not fully investigated for the larval part of the life cycle of these parasites. ES products of *S. mansoni* cercariae were shown to play key roles in facilitating skin invasion and evading the early innate immune response [51]. ES products of adult schistosome worms were described to counteract host adaptive response [23,52]. Immune relevant molecules were also identified in ES products from adult *Echinostoma friedi* [53,54].

For the larval stage of the parasite, *in vitro* studies have shown immune activity in ES of trematoda sporocysts [15,16,20,22]. However, the minute size of sporocysts and the difficulty to collect larval ES products prevented earlier investigations of sporocysts proteome. Here we performed for the first time a proteomic analysis of sporocysts ES proteome. We used two species that are natural parasites of *B. glabrata*, which use different immune-evasion strategies. Intriguingly, two groups of proteins were widely represented in both ES proteomes: proteins involved in the detoxification of oxidative stress and enzymes implicated in the Embden–Meyerhof–Parnas pathway.

Several studies have shown that reactive oxygen species (ROS), like hydrogen peroxide, produced by hemocytes play a crucial role in the killing of parasites [6,8,55–57]. Parasite elimination capacity is directly correlated with capacity of ROS production [57]. Among these molecules, the most toxic form for sporocysts is hydrogen peroxide (H<sub>2</sub>O<sub>2</sub>), probably due to its capacity to cross cell membranes [8,58].

Trematodes are exposed to ROS throughout their life cycle. Oxidative stress scavengers were described in schistosomula and adult worms [59], and were found in the ES products of cercariae and adults of schistosomes and echinostomes [51–53]. In addition, an antioxidant activity of *S. mansoni* ES products has been reported [60]. These earlier observations and our experimental data fit well into a model in which the proteins characterized in this work function as antioxidants and protect sporocysts from oxidative damage.

The Cu-Zn SOD identified in ES proteome of *S. mansoni* and *E. caproni* catalyzes the dismutation of superoxide anion ( $O_2^-$ ) into hydrogen peroxide ( $H_2O_2$ ) and  $O_2$ .  $H_2O_2$  is the most toxic oxygen species for *S. mansoni* sporocysts [8]. Thus, the fact that the parasite enhances the production of  $H_2O_2$  by the secretion of SOD was intriguing. But, in addition to the usual superoxide dismutation activity, Cu-Zn SOD has a peroxidative function that utilizes its own dismutation product,  $H_2O_2$ , as a substrate [61,62]. Experimental results have shown that, during these reactions, free radicals, such as the hydroxyl radical ( $\bullet OH$ ) are generated [61,63]. SOD can therefore represent the first line of cellular antioxidant defense of the parasite. Using its peroxidative function, SOD can inactivate  $H_2O_2$  and produce  $\bullet OH$  that is less toxic for the sporocysts [8]. The second line of cellular defense involves the glutathione-S-transferase (GST), which detoxifies the membrane lipid peroxidation induced by the ROS like  $\bullet OH$ . GST (identified in *S. mansoni* ES products) neutralizes cytotoxic by-products of lipid peroxidation arising from reactive oxygen species acting on cell membrane [64]. GST catalyses the reduction of lipid peroxidation by-products by the use of reduced glutathione (GSH) producing oxidized glutathione (GSSG) [65].

Moreover, this lipid peroxidation results in the formation of aldehydes (4-hydroxy-2-nonenal; malondialdehyde) that cause protein modifications like carbonylation. Aldo-keto reductase (AKR) in ES products of *E. caproni* can be an efficient catalyst for reducing phospholipid aldehydes. AKR is an antioxidant enzyme that removes and detoxifies products of lipid peroxidation. The enzyme reduced phospholipids with acyl, alkyl, and alkenyl linkages at the *sn*-1 position and with different head groups at the *sn*-3 position, suggesting that it recognizes a broad range of phospholipids aldehydes [66].

The second group of ES proteins widely contains enzymes involved in the glycolysis pathway (triose phosphate isomerase (TPI), glyceraldehydes-3-phosphate dehydrogenase (GAPDH), aldolase, enolase). These enzymes have already been identified in ES products from *S. mansoni* cercariae [51] and in ES from *S. bovis* adult worms [51,52]. GAPDH was identified in ES products from adult *E. friedi* [53] and enolase was identified in ES products from adult *Fasciola hepatica* [54]. This enzyme is also identified at the surface of fungal pathogen [67,68], filarial parasites [69] and mammals [70].

The presence of glycolytic enzymes in the extracellular environment of the parasite is surprising. One might argue that this could be a contamination of the ES products by cytosolic compounds. However, we carefully confirmed that this is not the case (see result section). In addition, glycolytic enzymes have interesting properties that are often overlooked. TPI and GAPDH are involved in two successive reactions of the glycolysis involving the reduction of a molecule of  $NAD^+$  to  $NADH$ . These results strongly support the view that these enzymes of glycolysis account for the synthesis of  $NADH$  that could be involved in the glutathione redox cycle [71]. The  $NADH$  generated can be used as a reducing agent to reduce oxidized glutathione (GSSG). GSSG accumulates due to the GST activity that catalyses the reduction of lipid peroxidation by the use of reduced glutathione (GSH) producing oxidized glutathione (GSSG) [65].

Glycolytic enzymes identified in ES proteome of *E. caproni* let us hypothesize that the production of pyruvate can be implicated in anti-oxidative processes. Indeed, for many parasites, glycolysis is involved in the struggle against oxidative stress [72,73]. Pyruvate can detoxify hydrogen peroxide ( $H_2O_2$ ), when enzymes normally involved in this detoxification pathway are overwhelmed or absent [74–77]. This is done by pyruvate decarboxylation (a non-enzymatic reaction) in presence of ROS [78]. Thus, pyruvate seems to offer a good protection against oxidation [75,76,79,80].

Interestingly, enolase and GAPDH, two other glycolytic enzymes identified in the ES proteome of *E. caproni* and *S. mansoni*, were also described as being able to bind plasminogen and thus facilitate invasion and migration in host tissue [69,81,82].

Finally, two proteins identified in parasite ES proteome can be involved in parasite immune evasive strategies and particularly in inhibition of haemocytes adhesion and phagocytosis: MICAL-like and calreticulin.

First, MICAL-like identified in *E. caproni* ES proteome is a flavoprotein monooxygenase [48,49] that plays a key role in repulsive neuronal guidance [83]. ES products of *E. caproni* sporocysts were shown to cause the repulsion and the loss of haemocytes adhesion [55,20,84]. This activity can be related to the properties described for MICAL of drosophila. MICAL regulates axon steering through a local increase in the concentration of reactive oxygen species (ROS) and in particular hydrogen peroxide [85]. As we described herein it can thus be useful for the parasite to have ROS scavenging enzymes to protect himself against the potential indirect cytotoxic effect of MICAL through  $H_2O_2$ .

The second protein is the calreticulin (CRT) identified in the ES proteome of *S. mansoni* primary sporocysts. This protein is highly conserved and plays a role in many cellular functions [38]. CRT controls  $Ca^{2+}$  homeostasis [86,87]. The proline-rich and highly charged P-domain of CRT contains a high-affinity, low capacity  $Ca^{2+}$ -binding site and the negatively charged C-terminal domain contains a low affinity, high capacity  $Ca^{2+}$ -binding site. As cellular adhesion and spreading is calcium-dependant, we hypothesize that secretion of a protein that can modify the extracellular concentration of  $Ca^{2+}$  prevents the hemocytes encapsulation process [29,40]. It was recently reported that a parasitoid calreticulin inhibits hemocyte spreading behavior, thus preventing encapsulation of the developing parasitoid [88]. This highly conserved and multifunctional CRT has been identified and isolated from many parasites and contributes to various aspect of host–parasite interactions and the modulation of the host response by the parasite [39,40,88–90].

Our proteomic approach provides new insights into the understanding of *S. mansoni* and *E. caproni* sporocyst immune evasive strategy. Interestingly, most proteins evidenced here belong to the same families both in *S. mansoni* and *E. caproni*. In addition, these types of proteins have been demonstrated to be involved in immune evasion in other host–parasite systems interacting with invertebrate or vertebrate hosts [51–54,59,81,88]. This is the second main result of our study: the comparison of proteins involved in immune evasion in different systems suggests that



parasites use a limited and conserved set of molecules to protect themselves from, or to alter host immune responses.

Here, we identified, for the first time, proteins produced by the larval stage of *S. mansoni* or *E. caproni* that are likely to play a role in the molecular cross-talk with the host during the first hours following infection. The exact role and function of these proteins will be investigated in future studies, but their predicted function is informative about host–parasite interaction processes and allows for the establishment of a plausible hypothesis: after the miracidium penetrates its snail host during the first hours after infection, the ciliated epidermal plates detach and epidermal ridges start to expand forming a new tegument [91]. This young post-miracidium parasite is the most vulnerable stage of the intramolluscan parasite since it is devoid of a protective surface covering [91]. The new tegument is formed with the help of the increasing number of membrane-bound vesicles that attach to the outer membrane. They become part of the membrane of the newly formed tegument, releasing at the same time their contents (ES products) into the snail tissue [91,92]. During the first 24 h of culture, sporocysts release the greatest amount of ES proteins [92]. These ES products must therefore be involved in an early immunoprotective strategy that protects the parasite during his most vulnerable stage. Indeed, ES proteome of these two trematodes originated from the same number of sporocysts. However, more proteins were observed in *E. caproni* ES profile (Fig. 1). If both need immuno-protection during the first hour of infection, *E. caproni* did not develop secondarily a mimetic strategy. Thus he needs to use much more proteins than *S. mansoni* to counteract host innate immunity. Differences in immunobiological interactions with the host could perhaps explain spot number differences between the excretory–secretory proteomes of these two trematodes.

Trematode parasites may use different strategies, in a sequential manner, to realize their life cycle in their intermediate snail host. During the first 24 h of infection, sporocysts may develop an active strategy to protect themselves against haemocyte-mediated cytotoxicity. This permits the formation of the new tegument by the vulnerable post-miracidium stage. The formation of this protective membrane is a vital step for the parasite to secondarily develop a mimetic or immunosuppressive strategy according to the species, thus avoiding the recognition and elimination of the parasite for the rest of its intramolluscan cycle [9–11,15,16,19].

We propose a new model for the complex strategy of trematodes to survive in their intermediate host: (i) an early immunoprotection followed by (ii) a delayed active strategy. To our knowledge, the succession of these two different strategies has not been hypothesized so far in snail-trematode systems. Such a scenario is consistent with the results reported here but needs to be sustained by functional studies in the future.

## Acknowledgments

The authors thank Michael S. Blouin for his valuable insights and constructive comments during the preparation of this paper. This work was supported by the CNRS. The mass spectrometry facility used in this study was funded by the European Com-

munity (FEDER), the Region Nord-Pas de Calais (France), the CNRS, the Genopole of Lille and the University of Sciences and Technologies of Lille.

## References

- [1] Davies SJ, McKerrow JH. Developmental plasticity in schistosomes and other helminths. *Int J Parasitol* 2003;33:1277–84.
- [2] Parker GA, Chubb JC, Ball MA, Roberts GN. Evolution of complex life cycles in helminth parasites. *Nature* 2003;425:480–4.
- [3] de Jong-Brink M, Bergamin-Sassen M, Solis Soto M. Multiple strategies of schistosomes to meet their requirements in the intermediate snail host. *Parasitology* 2001;123(Suppl):S129–41.
- [4] Meuleman EA, Bayne CJ, van der Knaap WP. Immunological aspects of snail–trematode interactions. *Prog Clin Biol Res* 1987;233:113–27.
- [5] van der Knaap WP, Loker ES. Immune mechanisms in trematode–snail interactions. *Parasitol Today* 1990;6:175–82.
- [6] Bayne CJ, Hahn UK, Bender RC. Mechanisms of molluscan host resistance and of parasite strategies for survival. *Parasitology* 2001;123(Suppl):S159–67.
- [7] Ataev GL, Coustau C. Cellular response to *Echinostoma caproni* infection in *Biomphalaria glabrata* strains selected for susceptibility/resistance. *Dev Comp Immunol* 1999;23:187–98.
- [8] Hahn UK, Bender RC, Bayne CJ. Involvement of nitric oxide in killing of *Schistosoma mansoni* sporocysts by hemocytes from resistant *Biomphalaria glabrata*. *J Parasitol* 2001;87:778–85.
- [9] Abu-Shakra M, Buskila D, Shoenfeld Y. Molecular mimicry between host and pathogen: examples from parasites and implication. *Immunol Lett* 1999;67:147–52.
- [10] Damian RT. Parasite immune evasion and exploitation: reflections and projections. *Parasitology* 1997;115(Suppl):S169–75.
- [11] Salzet M, Capron A, Stefano GB. Molecular crosstalk in host–parasite relationships: schistosome– and leech–host interactions. *Parasitol Today* 2000;16:536–40.
- [12] Daniel BE, Preston TM, Southgate VR. The in vitro transformation of the miracidium to the mother sporocyst of *Schistosoma margrebowiei*; changes in the parasite surface and implications for interactions with snail plasma factors. *Parasitology* 1992;104(Pt 1):41–9.
- [13] Lie KJ, Heyneman D. Studies on resistance in snails: interference by nonirradiated echinostome larvae with natural resistance to *Schistosoma mansoni* in *Biomphalaria glabrata*. *J Invertebr Pathol* 1977;29:118–25.
- [14] Lie KJ, Heyneman D. Acquired resistance to echinostomes in four *Biomphalaria glabrata* strains. *Int J Parasitol* 1979;9:533–7.
- [15] Lie KJ, Jeong KH, Heyneman D. Further characterization of acquired resistance in *Biomphalaria glabrata*. *J Parasitol* 1982;68:529–31.
- [16] Loker ES, Bayne CJ, Yui MA. *Echinostoma paraensei*: hemocytes of *Biomphalaria glabrata* as targets of echinostome mediated interference with host snail resistance to *Schistosoma mansoni*. *Exp Parasitol* 1986;62:149–54.
- [17] Noda S, Loker ES. Effects of infection with *Echinostoma paraensei* on the circulating haemocyte population of the host snail *Biomphalaria glabrata*. *Parasitology* 1989;98(Pt 1):35–41.
- [18] Loker ES, Cimino DF, Hertel LA. Excretory–secretory products of *Echinostoma paraensei* sporocysts mediate interference with *Biomphalaria glabrata* hemocyte functions. *J Parasitol* 1992;78:104–15.
- [19] Sapp KK, Loker ES. Mechanisms underlying digenean–snail specificity: role of miracidial attachment and host plasma factors. *J Parasitol* 2000;86:1012–9.
- [20] Humbert E, Coustau C. Refractoriness of host haemocytes to parasite immunosuppressive factors as a putative resistance mechanism in the *Biomphalaria glabrata*–*Echinostoma caproni* system. *Parasitology* 2001;122:651–60.
- [21] Loker ES, Adema CM. Schistosomes, Echinostomes and snails: comparative immunobiology. *Parasitology today* 1995;11:120–4.
- [22] Coustau C, Mita G, Dissous C, et al. *Schistosoma mansoni* and *Echinostoma caproni* excretory–secretory products differentially affect gene expression in *Biomphalaria glabrata* embryonic cells. *Parasitology* 2003;127:533–42.

- [23] Loverde PT. Do antioxidants play a role in schistosome host–parasite interactions? *Parasitol Today* 1998;14:284–9.
- [24] Jeyarasasingam U, Heyneman D, Lim HK, Mansour N. Life cycle of a new echinostome from Egypt, *E. liei* sp. nov. (Trematoda; Echinostomatidae). *Parasitology* 1972;65:203–22.
- [25] Dissous C, Dissous C, Capron A. Isolation and characterization of surface antigens from *Schistosoma mansoni* schistosomula. *Mol Biochem Parasitol* 1981;3:215–25.
- [26] Languard J, Jourdan J, Coustau C, Delay B, Morand S. Cost of resistance, expressed as a delayed maturity, detected in the host–parasite system *Biomphalaria glabrata*/*Echinostoma caproni*. *Heredity* 1998;80:320–5.
- [27] Trouvé S, Renaud F, Durand P, Jourdan J. Selfing and outcrossing in a parasitic hermaphrodite helminth (Trematoda, Echinostomatidae). *Heredity* 1996;77:1–8.
- [28] Chemin E. Observations on hearts explanted in vitro from the snail *Australorbis glabratus*. *J Parasitol* 1963;49:353–64.
- [29] Vergote D, Bouchut A, Sautiere PE, et al. Characterisation of proteins differentially present in the plasma of *Biomphalaria glabrata* susceptible or resistant to *Echinostoma caproni*. *Int J Parasitol* 2005;35:215–24.
- [30] Shevchenko A, Wilm M, Vorm O, Mann M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 1996;68:850–8.
- [31] Bouchut A, Sautiere PE, Coustau C, Mitta G. Compatibility in the *Biomphalaria glabrata*/*Echinostoma caproni* model: potential involvement of proteins from hemocytes revealed by a proteomic approach. *Acta Trop* 2006;98:234–46.
- [32] Kayser JP, Vallet JL, Cerny RL. Defining parameters for homology-tolerant database searching. *J Biomol Tech* 2004;15:285–95.
- [33] dos Reis MG, Davis RE, Singh H, Skelly PJ, Shoemaker CB. Characterization of the *Schistosoma mansoni* gene encoding the glycolytic enzyme, triosephosphate isomerase. *Mol Biochem Parasitol* 1993;59:235–42.
- [34] Shoemaker C, Gross A, Gebremichael A, Harn D. cDNA cloning and functional expression of the *Schistosoma mansoni* protective antigen triosephosphate isomerase. *Proc Natl Acad Sci USA* 1992;89:1842–6.
- [35] Jenkins JL, Tanner JJ. High-resolution structure of human D-glyceraldehyde-3-phosphate dehydrogenase. *Acta Crystallogr D Biol Crystallogr* 2006;62:290–301.
- [36] Cardoso RM, Silva CH, Ulian de Araujo AP, Tanaka T, Tanaka M, Garratt RC. Structure of the cytosolic Cu, Zn superoxide dismutase from *Schistosoma mansoni*. *Acta Crystallogr D Biol Crystallogr* 2004;60:1569–78.
- [37] Dirr H, Reinemer P, Huber R. X-ray crystal structures of cytosolic glutathione S-transferases. Implications for protein architecture, substrate recognition and catalytic function. *Eur J Biochem* 1994;220:645–61.
- [38] Michalak M, Corbett EF, Mesaali N, Nakamura K, Opas M. Calreticulin: one protein, one gene, many functions. *Biochem J* 1999;344(Pt 2):281–92.
- [39] Khalife J, Pierce RJ, Godin C, Capron A. Cloning and sequencing of the gene encoding *Schistosoma mansoni* calreticulin. *Mol Biochem Parasitol* 1993;62:313–5.
- [40] Khalife J, Liu JL, Pierce R, Porchet E, Godin C, Capron A. Characterization and localization of *Schistosoma mansoni* calreticulin Sm58. *Parasitology* 1994;108(Pt 5):527–32.
- [41] Lebioda L, Stec B, Brewer JM. The structure of yeast enolase at 2.25-Å resolution. An 8-fold beta + alpha-barrel with a novel beta beta alpha alpha (beta alpha) 6 topology. *J Biol Chem* 1989;264:3685–93.
- [42] Reed GH, Poyner RR, Larsen IM, Wedekind JE, Rayment I. Structural and mechanistic studies of enolase. *Curr Opin Struct Biol* 1996;6:736–43.
- [43] Kim TS, Jung Y, Na BK, Kim KS, Chung PR. Molecular cloning and expression of Cu/Zn-containing superoxide dismutase from *Fasciola hepatica*. *Infect Immun* 2000;68:3941–8.
- [44] Wernuth B. Aldo-keto reductases. *Prog Clin Biol Res* 1985;174:209–30.
- [45] Inoue S, Sharma RC, Schimke RT, Simoni RD. Cellular detoxification of tripeptidyl aldehydes by an aldo-keto reductase. *J Biol Chem* 1993;268:5894–8.
- [46] Picklo Sr MJ, Olson SJ, Hayes JD, Markesbery WR, Montine TJ. Elevation of AKR7A2 (succinic semialdehyde reductase) in neurodegenerative disease. *Brain Res* 2001;916:229–38.
- [47] Zhang D, Tai LK, Wong LL, Chiu LL, Sethi SK, Koay ES. Proteomic study reveals that proteins involved in metabolic and detoxification pathways are highly expressed in HER-2/neu-positive breast cancer. *Mol Cell Proteomics* 2005;4:1686–96.
- [48] Nadella M, Bianchet MA, Gabelli SB, Barrila J, Amzel LM. Structure and activity of the axon guidance protein MICAL. *Proc Natl Acad Sci USA* 2005;102:16830–5.
- [49] Pasterkamp RJ, Dai HN, Terman JR, et al. MICAL flavoprotein monooxygenases: expression during neural development and following spinal cord injuries in the rat. *Mol Cell Neurosci* 2006;31:52–69.
- [50] Maizels RM, Balic A, Gomez-Escobar N, Nair M, Taylor MD, Allen JE. Helminth parasites—masters of regulation. *Immunol Rev* 2004;201:89–116.
- [51] Knudsen GM, Medzihradzky KF, Lim KC, Hansell E, McKerrow JH. Proteomic analysis of *Schistosoma mansoni* cercarial secretions. *Mol Cell Proteomics* 2005;4:1862–75.
- [52] Perez-Sanchez R, Ramajo-Hernandez A, Ramajo-Martin V, Oleaga A. Proteomic analysis of the tegument and excretory–secretory products of adult *Schistosoma bovis* worms. *Proteomics* 2006;6(Suppl 1):S226–36.
- [53] Bernal D, Carpena I, Espert AM, et al. Identification of proteins in excretory/secretory extracts of *Echinostoma friedi* (Trematoda) from chronic and acute infections. *Proteomics* 2006;6:2835–43.
- [54] Bernal D, de la Rubia JE, Carrasco-Abad AM, Toledo R, Mas-Coma S, Marcilla A. Identification of enolase as a plasminogen-binding protein in excretory–secretory products of *Fasciola hepatica*. *FEBS Lett* 2004;563:203–6.
- [55] Adema CM, van Deutekom-Mulder EC, van der Knaap WP, Sminia T. Schistosomicidal activities of *Lymnaea stagnalis* haemocytes: the role of oxygen radicals. *Parasitology* 1994;109(Pt 4):479–85.
- [56] Dikkeboom R, Bayne CJ, van der Knaap WP, Tijnagel JM. Possible role of reactive forms of oxygen in *in vitro* killing of *Schistosoma mansoni* sporocysts by hemocytes of *Lymnaea stagnalis*. *Parasitol Res* 1988;75:148–54.
- [57] Bender RC, Broderick EJ, Goodall CP, Bayne CJ. Respiratory burst of *Biomphalaria glabrata* hemocytes: *Schistosoma mansoni*-resistant snails produce more extracellular H<sub>2</sub>O<sub>2</sub> than susceptible snails. *J Parasitol* 2005;91:275–9.
- [58] Bienert GP, Schjoerring JK, Jahn TP. Membrane transport of hydrogen peroxide. *Biochim Biophys Acta* 2006;1758:994–1003.
- [59] Mei H, LoVerde PT. *Schistosoma mansoni*: the developmental regulation and immunolocalization of antioxidant enzymes. *Exp Parasitol* 1997;86:69–78.
- [60] Connors VA, Lodes MJ, Yoshino TP. Identification of a *Schistosoma mansoni* sporocyst excretory–secretory antioxidant molecule and its effect on superoxide production by *Biomphalaria glabrata* hemocytes. *J Invertebr Pathol* 1991;58:387–95.
- [61] Yim MB, Kang JH, Yim HS, Kwak HS, Chock PB, Stadtman ER. A gain-of-function of an amyotrophic lateral sclerosis-associated Cu, Zn-superoxide dismutase mutant: An enhancement of free radical formation due to a decrease in K<sub>m</sub> for hydrogen peroxide. *Proc Natl Acad Sci USA* 1996;93:5709–14.
- [62] Kim SM, Kang JH. Peroxidative activity of human Cu, Zn-superoxide dismutase. *Mol Cell* 1997;7:120–4.
- [63] Yim MB, Chock PB, Stadtman ER. Enzyme function of copper, zinc superoxide dismutase as a free radical generator. *J Biol Chem* 1993;268:4099–105.
- [64] Zelck UE, Von Janowsky B. Antioxidant enzymes in intramolluscan *Schistosoma mansoni* and ROS-induced changes in expression. *Parasitology* 2004;128:493–501.
- [65] Tew KD, Ronai Z. GST function in drug and stress response. *Drug Resist Updat* 1999;2:143–7.
- [66] Srivastava S, Spite M, Trent JO, West MB, Ahmed Y, Bhatnagar A. Aldose reductase-catalyzed reduction of aldehyde phospholipids. *J Biol Chem* 2004;279:53395–406.
- [67] Fox D, Smulian AG. Plasminogen-binding activity of enolase in the opportunistic pathogen *Pneumocystis carinii*. *Med Mycol* 2001;39:495–507.
- [68] Jong AY, Chen SH, Stins MF, Kim KS, Tuan TL, Huang SH. Binding of *Candida albicans* enolase to plasmin(ogen) results in enhanced invasion of human brain microvascular endothelial cells. *J Med Microbiol* 2003;52:615–22.

- [69] Jolodar A, Fischer P, Bergmann S, Buttner DW, Hammerschmidt S, Brattig NW. Molecular cloning of an alpha-enolase from the human filarial parasite *Onchocerca volvulus* that binds human plasminogen. *Biochim Biophys Acta* 2003;1627:111–20.
- [70] Lopez-Alemayn R, Suelves M, Munoz-Canoves P. Plasmin generation dependent on alpha-enolase-type plasminogen receptor is required for myogenesis. *Thromb Haemost* 2003;90:724–33.
- [71] Kum-Tatt L, Tan IK, Seet AM. A new colorimetric method for the determination of NADH/NADPH dependent glutathione reductase in erythrocytes and in plasma. *Clin Chim Acta* 1975;58:101–8.
- [72] Roth Jr E. *Plasmodium falciparum* carbohydrate metabolism: a connection between host cell and parasite. *Blood Cells* 1990;16:453–60, discussion 61–6.
- [73] Cordeiro AT, Michels PA, Delboni LF, Thiemann OH. The crystal structure of glucose-6-phosphate isomerase from *Leishmania mexicana* reveals novel active site features. *Eur J Biochem* 2004;271:2765–72.
- [74] Salahudeen AK, Clark EC, Nath KA. Hydrogen peroxide-induced renal injury. A protective role for pyruvate *in vitro* and *in vivo*. *J Clin Invest* 1991;88:1886–93.
- [75] Marcengill MB, Puri S, Puri SK, et al. Antioxidant effects of pyruvate in isolated rat hearts. *J Mol Cell Cardiol* 1995;27:2059–67.
- [76] Bassenge E, Sommer O, Schwemmer M, Bunger R. Antioxidant pyruvate inhibits cardiac formation of reactive oxygen species through changes in redox state. *Am J Physiol Heart Circ Physiol* 2000;279:H2431–8.
- [77] Olivencia-Yurvati AH, Blair JL, Baig M, Mallet RT. Pyruvate-enhanced cardioprotection during surgery with cardiopulmonary bypass. *J Cardiothorac Vasc Anesth* 2003;17:715–20.
- [78] Constantopoulos G, Barranger JA. Nonenzymatic decarboxylation of pyruvate. *Anal Biochem* 1984;139:353–8.
- [79] Mallet RT, Sun J, Knott EM, Sharma AB, Olivencia-Yurvati AH. Metabolic cardioprotection by pyruvate: recent progress. *Exp Biol Med* (Maywood) 2005;230:435–43.
- [80] Biagini GA, Park JH, Lloyd D, Edwards MR. The antioxidant potential of pyruvate in the amitochondriate diplomonads *Giardia intestinalis* and *Hexamita inflata*. *Microbiology* 2001;147:3359–65.
- [81] Kolberg J, Aase A, Bergmann S, et al. *Streptococcus pneumoniae* enolase is important for plasminogen binding despite low abundance of enolase protein on the bacterial cell surface. *Microbiology* 2006;152:1307–17.
- [82] Bergmann S, Rohde M, Hammerschmidt S. Glyceraldehyde-3-phosphate dehydrogenase of *Streptococcus pneumoniae* is a surface-displayed plasminogen-binding protein. *Infect Immun* 2004;72:2416–9.
- [83] Terman JR, Mao T, Pasterkamp RJ, Yu HH, Kolodkin AL. MICALs, a family of conserved flavoprotein oxidoreductases, function in plexin-mediated axonal repulsion. *Cell* 2002;109:887–900.
- [84] DeGaffe G, Loker ES. Susceptibility of *Biomphalaria glabrata* to infection with *Echinostoma paraensei*: correlation with the effect of parasite secretory-excretory products on host hemocyte spreading. *J Invertebr Pathol* 1998;71:64–72.
- [85] Ventura A, Pelicci PG. Semaphorins: green light for redox signaling? *Sci STKE* 2002, 2002:PE44.
- [86] Clapham DE. Calcium signaling. *Cell* 1995;80:259–68.
- [87] Berridge MJ. Neuronal calcium signaling. *Neuron* 1998;21:13–26.
- [88] Zhang G, Schmidt O, Asgari S. A calreticulin-like protein from endoparasitoid venom fluid is involved in host hemocyte inactivation. *Dev Comp Immunol* 2006;30:756–64.
- [89] Ferreira V, Molina MC, Valck C, et al. Role of calreticulin from parasites in its interaction with vertebrate hosts. *Mol Immunol* 2004;40:1279–91.
- [90] Nakhasi HL, Pogue GP, Duncan RC, et al. Implications of calreticulin function in parasite biology. *Parasitol today* 1998;14:157–60.
- [91] Pan SC. *Schistosoma mansoni*: the ultrastructure of larval morphogenesis in *Biomphalaria glabrata* and of associated host-parasite interactions. *Jpn J Med Sci Biol* 1996;49:129–49.
- [92] Lodes MJ, Yoshino TP. Characterization of excretory-secretory proteins synthesized *in vitro* by *Schistosoma mansoni* primary sporocysts. *J Parasitol* 1989;75:853–62.



## **ANNEXE 2**

### ***Travaux développés sur le modèle Biomphalaria glabrata / Echinostoma caproni***

Les travaux réalisés au cours de cette thèse, sur le modèle *Schistosoma mansoni* / *Biomphalaria glabrata*, ont été développés dans le cadre d'un projet sur l'étude de la compatibilité dans les interactions mollusque / trématode. Dans ce projet, un autre axe de recherche, réalisé dans le cadre de la thèse de A. Bouchut, a été développé sur le modèle *Biomphalaria glabrata* / *Echinostoma caproni*. Ces deux modèles permettent d'un point de vue fondamental de comparer deux interactions tout à fait différentes concernant deux espèces de trématodes parasites, *S. mansoni* et *E. caproni* qui présentent un même hôte intermédiaire mollusque, *B. glabrata*. En effet, *E. caproni* provoque chez le mollusque une très forte immuno-suppression du système de défense interne contrairement à *S. mansoni*. Ce travail sur le modèle *B. glabrata* / *E. caproni* avait pour objectif de caractériser les déterminants moléculaires responsables de l'échec ou de la réussite de l'infestation en développant des approches moléculaires comparatives entre souches de *B. glabrata* totalement compatibles (EAF-S) et totalement incompatibles (CB-R) à *E. caproni*. J'ai pu participer, dans ce contexte, au cours de mon DEA et de ma première année de thèse, à certains travaux, qui sont présentés dans cette annexe.



# Compatibility in the *Biomphalaria glabrata*/*Echinostoma caproni* model: Potential involvement of adhesion genes<sup>☆</sup>

A. Bouchut, E. Roger, C. Coustau, B. Gourbal, G. Mitta \*

*Parasitologie Fonctionnelle et Evolutive, UMR 5555, CNRS Université de Perpignan, 52 Ave Paul Alduy, Perpignan Cedex 66860, France*

Received 12 July 2005; received in revised form 16 September 2005; accepted 20 September 2005

## Abstract

Because susceptibility or resistance of *Biomphalaria glabrata* to the trematode *Echinostoma caproni* correlates with differential hemocytic adhesive properties, we compared the expression of genes involved in adhesion processes between hemocytes from susceptible and resistant snails. Quantitative reverse transcriptase-PCR analysis revealed four genes whose transcripts were differentially represented between hemocytes from resistant and susceptible snails. These genes encode two dermatopontin-like, one matrilin-like and one cadherin-like proteins. Expression analyses performed following parasite exposure suggested that dermatopontins may be involved in the compatibility differences between these strains. We also investigated expression levels on whole snails of different genes potentially involved in extracellular matrix structure or coagulation. Our results support the hypothesis that susceptible snails possess a hemolymph coagulation-like system that is more potent than that of resistant snails. This system may prevent hemocyte migration towards the parasite larvae and therefore facilitate parasite settlement in susceptible snails.

© 2005 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

**Keywords:** *Biomphalaria glabrata*; *Echinostoma caproni*; Host–parasite interactions; Adhesion; Resistance; Hemocytes

## 1. Introduction

Snail-trematode immunobiological interactions have been studied in the freshwater snail *Biomphalaria glabrata* following infection by the human blood fluke *Schistosoma mansoni* and by the echinostome species, *Echinostoma paraensei* and *Echinostoma caproni*. In these different models, compatibility polymorphism was highlighted (Langand and Morand, 1998). In order to elucidate the mechanisms underlying differences in compatibility, numerous comparative molecular studies between susceptible and resistant strains of molluscs were undertaken. These studies revealed that numerous genes were differentially expressed between strains of *B. glabrata* susceptible or resistant to *S. mansoni* or

*E. caproni* (Miller et al., 2001; Raghavan et al., 2003; Lockyer et al., 2004; Nowak et al., 2004; Vergote et al., 2005).

In our model *B. glabrata*/*E. caproni*, two snail strains of *B. glabrata* were selected (Langand et al., 1998). *Echinostoma caproni* miracidia are able to penetrate both susceptible and resistant snails. While the parasite undergoes normal development in susceptible snails, in resistant snails it is encapsulated and eliminated. Susceptibility/resistance mechanisms of *B. glabrata* to *E. caproni* are probably inherited in a multigenic fashion (Langand and Morand, 1998) and have been shown to rely on both cellular and humoral factors (Ataev and Coustau, 1999). Regarding humoral factors, five proteins differentially represented in the plasma of susceptible and resistant snails were recently characterised, using a proteomic approach (Vergote et al., 2005). Regarding the cellular factors, studies showed that excretory–secretory (ES) products from in vitro transformed *E. caproni* sporocysts inhibited key defence functions of susceptible host hemocytes such as adhesion and phagocytosis (Humbert and Coustau, 2001). Interestingly, hemocytes from resistant snails remained unaffected by these parasite ES products, suggesting that they exhibit constitutive differences compared with hemocytes from susceptible snails (Humbert and Coustau, 2001). Because such differences were

<sup>☆</sup> Nucleotide sequence data reported in this paper are available in the GenBank™ database under the accession numbers: DQ114425, DQ114426, DQ113388, DQ113389, DQ113390, DQ113391, DQ113392, DQ113393, DQ113394, DQ113395, DQ113396, DQ113397, DQ113398, DQ113400, DQ113401, DQ113402, DQ113403, DQ113404, DQ113405, DQ113406.

\* Corresponding author. Tel.: +33 4 6866 2188; fax: +33 4 6866 2281.

E-mail address: [mitta@univ-perp.fr](mailto:mitta@univ-perp.fr) (G. Mitta).

indirectly or directly related to adhesive properties of hemocytes, we developed a targeted approach to examine the expression of genes directly involved in adhesion processes.

In the present study, we compared the transcript content of all known hemocytic genes potentially involved in adhesion processes between hemocytes of the two strains. We selected the differentially represented transcripts and characterised their corresponding open reading frames (ORF). Finally, we analysed their expression after exposure to the parasite.

## 2. Material and methods

### 2.1. Database search and sequence analysis

The *B. glabrata* nucleotide sequences available on GenBank (db EST of GenBank included) in December 2004 (6989 sequences) were imported in a fasta format. The sequences were assembled in clusters using the Sequencher™ software (Gene Codes Corporation). Consensus sequences from clusters were submitted to database searches using BLASTX program (<http://blast.genome.jp/>). Specific domain searches were performed using the RPS-BLAST program (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) or the ScanProsite program (<http://www.expasy.org/tools/scanprosite/>). Additional adhesion genes present in singletons were searched for using the homologies description of the different sequence submitters.

Putative signal peptides were detected using PSORTII prediction software (<http://psort.nibb.ac.jp/form2.html>). The hydrophobicity of Bg dermatopontins was investigated using the Kyte and Doolittle method (Kyte and Doolittle, 1982) and the ProtScale software (<http://www.expasy.org/tools/protscale.html>).

### 2.2. *Biomphalaria glabrata* strains

*Biomphalaria glabrata* snails used in this study belonged either to the unselected Bg.Bra stock originating from Brazil (Guillou et al., 2004), or to the EAF-S or CB-R strains selected, respectively, for their susceptibility or resistance to *E. caproni* infection (Langand et al., 1998). At the time of the study, the percentages of adult snails susceptible to *E. caproni* were: Bg.Bra = 70%, EAF-S = 98% and CB-R = 1%.

The parasite *E. caproni* (Jeyarasingam et al., 1972) was maintained in Bg.Bra snails and mice (SWISS OF1 stock) as described previously (Trouvé et al., 1996). For one experiment, adult EAF and CB snails (measuring 9–14 mm in shell diameter) were exposed to *E. caproni* according to previously described procedures (Guillou et al., 2004).

The Swiss OF1 mice used for this study came from our animal house. Our laboratory has received the permit No. A 66040 for experiments on animals from both French Ministry of Agriculture and Fisheries and French Educational Ministry of Research and Technology. Housing, breeding and animal care of the mice followed the ethical requirements of our country. The experimenter possesses the official certificate for animal experimentation delivered by both ministries

(Decree no. 87-848 19th October 1987; number of the authorisation 007083).

### 2.3. Hemocyte recovery and RNA extraction

Hemolymph was extracted from the headfoot of 200 EAF-S and CB-R snails according to standard procedure (Sminia and Barendsen, 1980) and immediately placed on ice. Approximately, 10 ml of hemolymph were recovered for each strain and centrifuged at 2000xg for 10 min at 4 °C. The pellets of hemocytes were immediately used for total RNA extraction using the NucleoSpin® RNA II Kit according to the manufacturer's instructions (Macherey-Nagel). Total RNA from whole body was extracted according to previously described procedures (Guillou et al., 2004).

### 2.4. Quantitative reverse transcriptase PCR (Q-PCR)

Q-PCR analysis of adhesion gene expression was performed on total RNAs extracted from hemocytes and whole bodies from unexposed and/or exposed (to *E. caproni*) EAF-S and CB-R snails. Reverse transcription was performed according to previously described procedures (Guillou et al., 2004). Specific primers for Q-PCR were edited using the LightCycler Probe Design Software version 1.0 (Roche Diagnostics). PCR reactions were set up according to the LightCycler Manual (Roche Diagnostics). The following LightCycler run protocol was used: denaturation program (95 °C, 10 min), amplification and quantification programs repeated 40 times (95 °C for 15 s, 60 °C for 5 s, 72 °C for 16 s), melting curve program (60–95 °C with a heating rate of 0.1 °C/s and continuous fluorescence measurement), and a cooling step to 40 °C. Amplification of single highly specific products was verified (melting curve analysis). For each reaction, the crossing point (CP) was determined using the 'Fit Point Method' of the LightCycler Software 3.3 (Roche Diagnostics). PCR reactions were performed in triplicate and the mean values of the CPs were calculated. For each mRNA to be analysed, the absence of contaminating genomic DNA was verified by running a no-RT control using primers for the ribosomal protein S19 (S19, accession number CK988928).

Amplification efficiencies (*E*) of each PCR product were determined according to previously described procedures (Guillou et al., 2004). For each sample, the level of transcription of the target gene (Tg) was compared with the transcription of the ribosomal protein S19. The transcription ratio (*R*) was calculated according to the formula:

$$R = (E_{Tg})^{CPTg} / (E_{S19})^{CPS19}$$

Statistical significance was assessed using the Mann-Whitney test for non-paired series and the Wilcoxon test for paired series. *P* values <0.05 were considered as significant. All tests were done using STATVIEW v. 4.5 statistic software package.

2.5. PCR and rapid amplification of cDNA ends (RACE)-PCR

In order to investigate potential differences in cDNA sequences of the various genes of interest, several specific forward and reverse primers were used in this study (Table 1, specific primers used for Q-PCR are also given).

*Biomphalaria glabrata* cDNAs were used to amplify the various fragments. PCR was performed in 50 µl mixtures containing 2 µl of first-strand cDNA using the DyNAzymeEXT™ kit (Finnzymes). PCR cycling conditions were 30 s at 94 °C, 1 min at 60 °C and 1 min 30 s at 72 °C for 40 cycles. For each mRNA to be analysed, the absence of contaminating DNA was verified by running a no-RT control using the same specific primers. PCR products were cloned using TOPO TA

Cloning® (Invitrogen Life Technologies) with the pCR®4-TOPO® vector and chemically competent TOP10F One Shot® *Escherichia coli*. Plasmid DNA of the positive recombinant clones was purified using QIAGEN® Plasmid Miniprep Kit (QIAGEN). For each amplified fragment, six clones were sequenced in both directions using dideoxy sequencing method (CEQ™ DTCS-Quick Start kit, Beckman Coulter) and a CEQ™ 8000 apparatus (Beckman Coulter). M13 forward, M13 reverse and specific primers were used for sequencing. Sequences were obtained using the CEQ™ 8000 sequence analysis software. Sequence analysis was performed using Sequencher™ software (Gene Codes Corporation).

In order to clone the 5'- or 3'-ends of the cDNAs, we used the GeneRacer™ Kit (Invitrogen) following the manufacturer's

Table 1  
Primers used for PCR and quantitative (Q)-PCR

Primers used for Q-PCR		Gene (GenBank™ accession numbers)
Forward	Reverse	
AGACGACCACTCACTG	CCAATGGCTTGGGTCTC	Bg cadherin like-1 (CV022372, -4)
ACCCACCACCCCTTTAG	TCTGCGTAGGCTTTCT	Bg cadherin like-3 (CK656861)
ACACCAAAGACTTTACTTCAG	GTTGTCTCTAATCAGGTTGACATA	
TCCGAGTCTCAATCATTAGC	CATATACTGGATGGCCCT	Bg matrilin (AF486289)
GCAAGAAAAGGTACACACT	AGTTGAAGGGCTGGTC	Bg dermatopontin 2 (CK988741, -817, CK989009, -525)
CCACGACAACAGAAAGG	CTAGTGGTCCGTCAAAATC	Bg dermatopontin 3 (CK989092, -409, -721, -676, -715, CK990133)
ACAGCAGAGGATGGTG	GCCATACGCTTCATCTTT	Bg α-integrin like 2 (CK988729, -945, CK989314)
GCTGACAGAGACCAGG	GAAAAGACTCCACGCC	
GAGAACTGCCAGTCA	GGTCTCTGGCACATCA	Bg VWD domain containing protein 1 (CN655005, CN779676, -716, -707, CV042526)
TTCCCGTCTGGAAC	CCTCTGAGGAGGTACATT	Bg VWD domain containing protein2 (CV548583, -258, -542, -237, -539, -566, CO870398)
AAAGACGTACCGAGGAT	CACTGAAAGCTAGAGCAC	Bg VWA domain containing protein 2 (CK988803, CK989503, -678, -931)
CCAATGTGTTTGGAGATCGG	TGACCACAGCTACTCG	Bg VWA domain containing protein 3 (CK989946, -671, -217, -773, -438, CK988907)
GGCTTTGTTCAGGAC	TCCATCTGGTATTGAGAGT	Bg VWC domain containing protein3 (CK149358, -489)
CCCAGGAAGACTCTGC	GGGTGGCTCTACAAT	Bg-spondin domain containing protein (CK989024, -796, CK988997)
TGCTCCTACAGTCAGGAA	GTCTGAGCTAACGCCAC	Bg-thrombospondin domain containing protein (CV548567, -042547, CN779718, -445838, CK149156, -085011, CD760632)
ATTACCTACACCAGCATCT	TGTCCAGGCTCAACAC	Bg multiple EGF and TSP domain containing protein (CV548577, -50, CO870431, -389, CN445784)
GGAATTGTGACGCTG	ACATCCATTAGTGCAACC	Bg chitin binding, multiple EGF and TSP domain containing protein (CN445780, -65, -71, CO635925)
CTACCTAACCCAACACTCAG	CTCCATCTTTGCTTGATGT	FREP 3.3 (AF515461)
<i>Primers used for cDNA cloning</i>		
Forward	Reverse	Gene
GTTATTGAACCCAACACTGGC	ACAAAGTTGTAAATAGAGCTC	Bg cadherin like-1
TCCTGATGAAGGCCTAAATGG	AACTAGGCTCTGAGGCATTCTG	Bg cadherin like-3
AACAGAGACTTCAAGACAGCC	TAACTTTGGTCCAGCCAGGTC	Bg matrilin
CTCTTGGATCTGGAAAAATCAA	CAATGCAATCGGCTTTTAGTCTA	Bg dermatopontin 2
TAGACTTCACTACTGACAACA	TGCTTTCAATAATCCTTCCATC	Bg dermatopontin 3
GCTCCAGTAAGCGTTGTGCTG	TTCCATGCTGGACTACACACC	Bg α-integrin like-2
AGGTAGCCACACAGCAC	GGCTTTACTGAGGGTCT	Bg VWD domain containing protein 1
CTCTACTTTGTGGATCAGCTG	TTTCAGCCCATGTTGACGGC	Bg VWD domain containing protein 2
TAACTGTCTAGCGCTGATCC	TCGTCACTCCAGTCAGGTAAG	Bg VWA domain containing protein 2
AACCACGTTAATACAAGACAT	AAAATTACCAGAGCCGC	Bg VWA domain containing protein 3
AAGGACGCATGTGAGGGAAGC	ATGCTGGTACTGCCTGGACAG	Bg VWC domain containing protein 3
TTGTCTGCTCCTGACAAGCGG	TTCACAGACTTCTTGACCGC	Bg-spondin domain containing protein
AATGCACTTGTGCAACGGTG	GTGATGAGTTCCAATCTGCCG	Bg thrombospondin domain containing protein
TGTTCTTGTCCCAAGTGTGAC	GTTCACAAGTTGGACATTGGC	Bg multiple EGF and TSP domain containing protein
CAATTGTGACTACCCTCCAC	AGGACATCCGATTGGAATCC	Bg chitin binding, multiple EGF and TSP domain containing protein
CTGGATGTTATCAGATCTC	AGATTGTAGTCTGGAGGTT	FREP 3.3



instructions. The resulting PCR products were ligated and sequenced as described above.

### 3. Results

#### 3.1. Database search for adhesion genes and expression analysis

In order to examine a maximum of genes potentially involved in adhesion processes, we selected genes displaying similarities with: (i) cell adhesion molecules (CAM) like

cadherins, immunoglobulin superfamily molecules, selectins, integrins and mucins; (ii) extracellular matrix (ECM) molecules mediating cell–matrix or matrix–matrix interactions like collagen, thrombospondin, laminin or fibronectin; and (iii) intracellular proteins linking CAM and cytoskeleton and cell signalling proteins involved in the regulation of cytoskeleton structure and CAM activation. Table S1 (online) in Appendix shows the different genes retained for the analysis and the detected domains of the corresponding proteins.

As a first approach to screen differences in these genes, we investigated their presence in transcripts representativity in

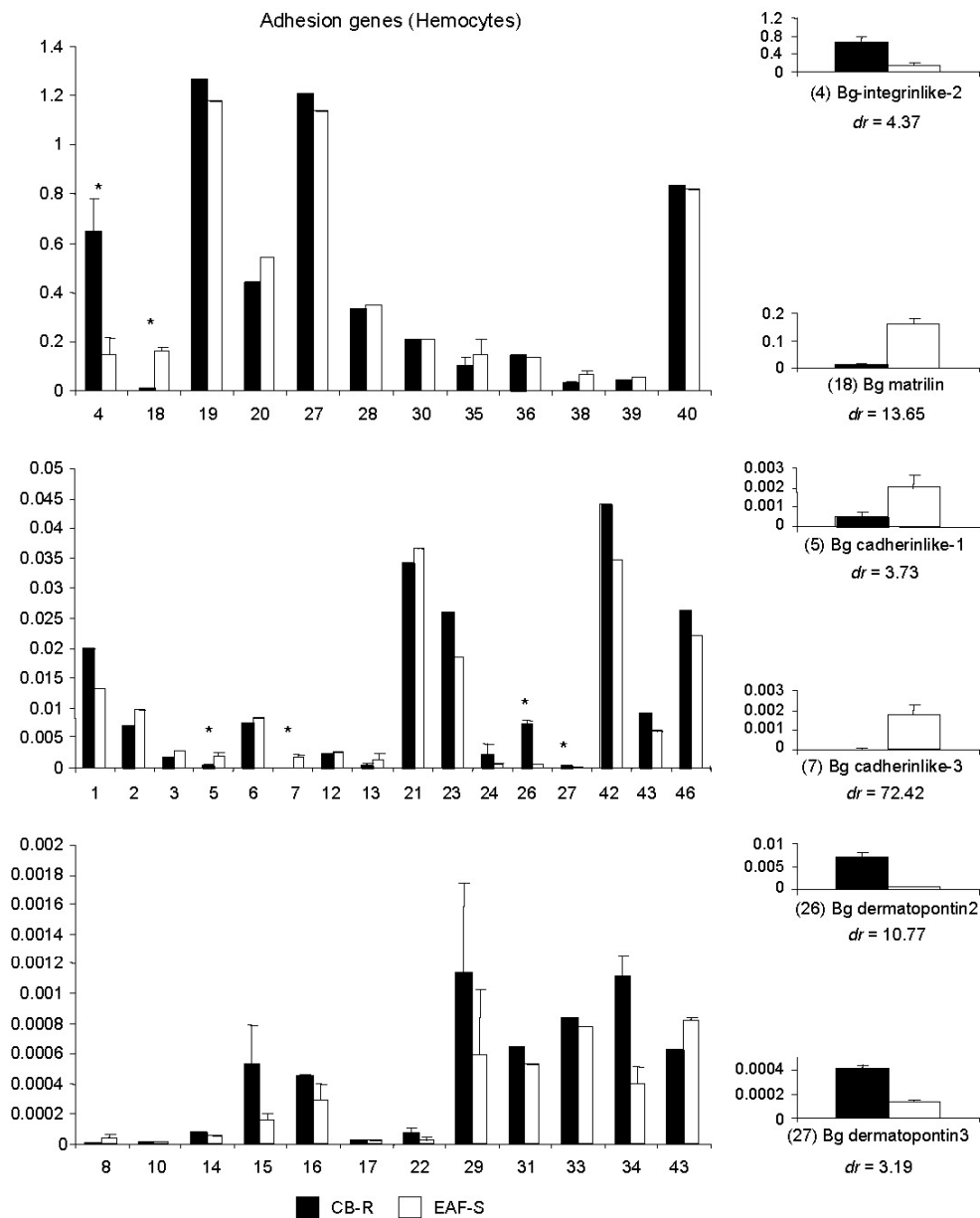


Fig. 1. Ratios of transcripts corresponding to genes potentially involved in adhesion processes in hemocytes from resistant CB-R and susceptible EAF-S snails. Ratios were determined using real-time quantitative PCR and are expressed relative to S19 expression levels (ratio/S19). Each histogram is the average value of triplicates  $\pm$  SD. dr: differential representativity of transcript between strains. \*Genes showing differences in transcript levels that were significant ( $P < 0.05$ ) and greater than three-fold.

Table 2

GenBank accession numbers of cDNAs corresponding to Bg cadherin like-1, Bg matrilin, Bg dermatopontin-2 and -3 characterised from hemocytes of CB and EAF snails

Gene	CB strain	EAF strain
Bg cadherin like 3	DQ114425	DQ114426
Bg integrin like 2	DQ113388	DQ113389
Bg cadherin like 1	DQ113390	DQ113391
Bg matrilin	DQ113392	DQ113393
Bg dermatopontin-2	DQ113394	DQ113395
Bg dermatopontin-3	DQ113396	DQ113397

hemocytes from the two strains. Representation of the transcripts corresponding to the 46 genes was analysed, using Q-PCR, on total RNA extracted from hemocytes of 200 snails (from each strain). Among the 46 transcripts analysed, six showed significant differences ( $P < 0.05$ ) in transcript levels that were greater than three-fold (results shown in Fig. 1).

3.2. cDNA characterisation of the genes displaying differential hemocytic transcript content in hemocytes from susceptible and resistant snails

In order to validate the differences in transcript content between hemocytes of susceptible and resistant snails, we verified that PCR primers were not edited on sites containing mutations between the two strains. cDNA sequences from six EAF-S susceptible snails and six CB-R resistant snails were obtained. Comparisons of sequences revealed that there were no mutations on sequences used to amplify Bg cadherin like-1, Bg matrilin, Bg dermatopontin-2 and -3. On the other hand, consistent point mutations between EAF-S and CB-R snails on

primer sequences for Bg  $\alpha$ -integrin like-2 and Bg cadherin like-3 (see Table 2 for accession numbers in Genbank). In order to verify if these mutations have had an effect on the analysis of transcript levels, new primers were designed and Q-PCR was performed again. Differences in transcript levels could not be observed anymore for these last two genes (data not shown) and these genes were therefore removed from the analysis.

In addition, 5' and 3'-RACE PCR were performed to attempt characterise the complete ORFs of the genes encoding Bg cadherin like-1, Bg matrilin, Bg dermatopontin-2 and -3 from the two strains: complete characterisation was achieved for the three last genes. Only partial characterisation was obtained for the Bg cadherin like-1. The corresponding sequences were deposited in Genbank and the corresponding accession numbers are given in Table 2. Regarding the cDNA sequences of Bg cadherin like-1, Bg matrilin, Bg dermatopontin-2 and -3, several differences were observed between the cDNAs obtained from the two strains. While these differences do not change the amino-acid structure of the Bg dermatopontins corresponding precursors (data not shown), several differences in the amino-acid sequence of the precursors were observed between strains for Bg cadherin like-1 and Bg matrilin (Fig. 2A and B, respectively).

cDNAs corresponding to Bg dermatopontin-2 and -3 encode precursors displaying high degrees of similarity (86% identical amino acids, Fig. 3A). They encode two isoforms of 168 and 166 amino acids, respectively, and are likely to be secreted as predicted using PSORTIII prediction software. These two isoforms seem to be expressed in whole individuals of both CB-R or EAF-S tested (Fig. 3B). Comparison of these sequences to protein databases showed that they were very

A	BgCad-1CB	VIEPNTGRIYLRRPLTESTNNAAYTLVVRATDSGNPQKSNDTDVIVTIQRAERPVFNISLA
	BgCad-1EAF	VIEPNTGRIYLRRPLTESTNNAAYTLVVRATDSGNPQKSNDTDVIVTIKRAERPVFIIINQE * * **
	BgCad-1CB	TISIPETQAIGSSIYNF
	BgCad-1EAF	TISIPETQAIGSSIYNF
B	BgMatCB	<b>MKYLAA SVLLQVGLTLAF</b> SLPNIARLSRREVVEDKGDHILICRQEPLELGIVLDS <del>SS</del> SLR
	BgMatEAF	<b>MKYLAA SVLLQVGLTLAF</b> SLPNIARLSRREVVEDKGDHILICRQEPLELGIVLDS <del>SS</del> SLR
	BgMatCB	NRDFKTAIKFLQEFLSQYEISSDPNGVVRVSIISYGKGIYPEIGFNLTYYDTKDEVIEAIG
	BgMatEAF	NRDFKTAIKFLQEFLSQYEISSDPNGVVRVSIISYGKGIYPEIGFNLTYYDTKDEVIEAIG
	BgMatCB	<u>RIPHKAGLR</u> TDTGRAIQYMHEAQLANGVVRPGVTKVSIVITDGNSQEWKLTKEAAEEARK
	BgMatEAF	<u>KIPHKAGLR</u> TDTGRAIQYMHEAQLANGVVRPGVTKVSIVITDGNSQEWKLTKEAAEEARK *
	BgMatCB	DNIVMFAIGVGTDIRNSELLNIAGDQSRVTKVDNYNQLSSIKESLAHQTCFVQEKTTTTP
	BgMatEAF	DNIVMFAIGVGTDIRNSELLNIAGDQSRVTKVDNYNQLSSIKESLAHQTCFVQEKTTTTP
	BgMatCB	VPMDEPCGVSDPADIFFVFS PANLGS DGLWATSF IGLTVDTEDLNQGF RYGVLSGSCP D
	BgMatEAF	EPMDEPCGVSDPADIFFVFS PANLGS DGLWATSF IGLTVDTEDLNQGF RYGVLSGSCP D *
	BgMatCB	DAGFNLDAYSTVEGIKNRLESYERNKVQELVKTLADDGFTTARGARANSRKIAVI IAAGK
	BgMatEAF	DAGFNLDAYSTVEGIKNRLESYERNKVQELVKTLADDGFTTARGARANSRKIAVI IAAGK
	BgMatCB	KTSELQDEVAALVAQNIQVFVADPADKGLTLEGATILSGRS SKLQAEAFVSALCPKRS
	BgMatEAF	KTSELQDEVAALVAQNIQVFVADPADKGLTLEGATILSGRS SKLQAEAFVSALCPKRS

Fig. 2. Deduced amino acids sequences and alignment of (A) Bg cadherin-like1 (partial coding sequence (CDS)) and of (B) Bg matrilin (complete CDS) obtained from the CB-R and EAF-S strains. Positions where amino acids are different between strains are marked with an asterisk. Signal sequences are in bold. The two von Willebrand domains are highlighted in grey and the metal ion-dependent adhesion site motif is underlined.



Fig. 3. (A) Alignment of invertebrate and vertebrate dermatopontin sequences and (B) transcript detection of *Bg* dermatopontin-2 and -3 in EAF-S and CB-R snails. Bgderm-1, -2 and -3: *Bg* dermatopontin-1, -2 and -3 from *Biomphalaria glabrata*; Limulusderm: dermatopontin from *Limulus polyphemus*; Suberitesderm: dermatopontin from *Suberites domuncula*; Bosderm: dermatopontin from *Bos taurus*. (A) The N-terminal signal sequences of the precursors are in bold and underlined. Cysteines are in bold. The three internal sequence repeats which are conserved in all dermatopontins are underlined. The three additional internal sequence repeats which are shared between *Bg* dermatopontin-2 and -3 and *Limulus* dermatopontin are boxed. (B) The transcription of the two dermatopontins characterised in this study (Bgderm-2 and -3) is analysed on two individuals of each strain by reverse transcriptase (RT)-PCR using specific primers discriminating these two isoforms (+). For each mRNA to be analysed, the absence of contaminating DNA was verified by running a no-RT control using the same specific primers (-).

similar to an *N*-glycosylated protein characterised from the organic shell matrix of *B. glabrata* ( $E$ -value  $7 \times 10^{-39}$  and  $3 \times 10^{-41}$ , respectively, accession number P83553 (Marxen et al., 2003)), named *B. glabrata* dermatopontin (annotated Bg dermatopontin-1 in Fig. 3A). In addition, these two isoforms display similarities with dermatopontins from the invertebrates *Limulus polyphemus* (accession number A44312,  $E$ -value  $3 \times 10^{-32}$  and  $4 \times 10^{-32}$  for isoforms 2 and 3, respectively (Fujii et al., 1992)) and *Suberites domuncula* (accession number CAC38786,  $E$ -value  $3 \times 10^{-30}$  and  $6 \times 10^{-30}$ , respectively (Schutze et al., 2001)). Furthermore, Bg dermatopontins display the cysteine array which is conserved in the other dermatopontins from vertebrates and invertebrates (see Fig. 3A). In contrast to the snail and sponge dermatopontins, both of which contain eight cysteines, the vertebrate (bovine, Neame et al., 1989)) and *Limulus* dermatopontins (Fujii et al., 1992) contain an additional pair of cysteines (Fig. 3A). Bg dermatopontins share the three internal sequence repeats which are conserved in all dermatopontins: D-R-X-W/F-X-F/Y//M-X<sub>1-2</sub>-C. In addition, another internal repeat unit was

detected in Bg dermatopontins-2 and -3 and in *Limulus* dermatopontin, V-N-D/S-F/W-D. This last consensus sequence was also repeated three times in these three precursors (Fig. 3A). As predicted by the Kyte and Doolittle method (Kyte and Doolittle, 1982) using the ProtScale software, these two internal repeating units are found in the sequence at positions of highest hydrophilicity (data not shown), which are likely exposed to the surface of the molecule.

cDNA corresponding to Bg matrilin encodes a precursor of 418 amino acids. The analysis of the primary structure by PSORTII prediction software reveals that it has a secretory protein-like structure with a 19-amino-acid signal sequence. Comparison of this sequence to protein databases showed that it is very similar to a cartilage matrix protein precursor, the Matrilin-1, from *Mus musculus* ( $E$ -value  $1 \times 10^{-20}$ , accession number Q80VN5). This last protein belongs to a family of fibril-forming vertebrate ECM proteins with four paralogues in the human genome (Deak et al., 1999). With the exception of matrilin 3, which has a single von Willebrand (VWA) domain, matrilins contain two VWA domains that flank a variable

number of EGF domains. These VWA domains are involved in protein–protein interactions contributing to the formation of the collagen-independent filamentous network (Chen et al., 1999). The matrilin from *B. glabrata* (Bg matrilin) also contains two VWA domains but no EGF domain was detected (Fig. 2). In addition, we detected a Metal ion-dependent adhesion site (MIDAS motif) within the first VWA domain of our protein (Fig. 2). The MIDAS motif may consist of five amino acids (DXSXS,T, and D) that contribute to adhesive properties of many VWA domains (Michishita et al., 1993; Goodman and Bajt, 1996; Puzon-McLaughlin and Takada, 1996).

### 3.3. Transcription of the candidate genes in susceptible and resistant snails after immune-challenge

In order to further investigate the potential involvement of the candidate genes (Bg cadherin like-1, Bg matrilin, Bg dermatopontin-2 and -3) in the differential susceptibility of strains, we compared their expression at various times after exposure to *E. caproni*. The kinetic rates of transcription of the four genes of interest were analysed, using Q-PCR, on total RNA extracted from pools of 10 snails at 0, 3, 6, 12, 24, 48 and 96 h p.i.

Transcription of these four genes increased regularly following infection in resistant snails (Fig. 4). Maximum increases were observed 48 h post-exposure: transcript corresponding to Bg cadherin like-1, Bg matrilin, Bg dermatopontin-2 and -3 are 3.2, 3.4, 24 and 4 times more abundant, respectively, compared with controls. All increases are significant ( $P < 0.05$ ). In susceptible snails, Bg cadherin like-1, Bg dermatopontin-2 and -3 kinetics appeared irregular: increases and decreases are alternatively observed. In addition, note that a significant difference ( $P < 0.05$ ) in transcript level of Bg dermatopontin-2 between resistant and susceptible snails is observed 48 h p.i. (expression ratio of 0.050 vs 0.013, respectively).

### 3.4. Genes potentially involved in coagulation or ECM in non-exposed snails (whole body) of the two strains

The higher level of Bg matrilin transcripts in hemocytes of the susceptible individuals (13.65 times more than in the hemocytes of resistant individuals, Fig. 1), suggests that the susceptible snails may possess a more potent hemolymph coagulation-like system. Such a system may block hemocyte migration towards the parasite in the first hours following infection (see discussion). In order to investigate this hypothesis, we measured the transcript levels of the genes corresponding to proteins potentially involved in coagulation processes or ECM. Because expression of such genes may not be restricted to hemocytes, we analysed total RNA extracted from whole snails. We focussed our attention on genes corresponding to proteins containing Von Willebrand domains (gene numbers 16–25 and 33–35, Appendix), coagulation factor domains (gene numbers 14 and 15, Appendix), trypsin-like serine domains (gene numbers 28, Appendix), spondin and thrombospondin domains (gene numbers 29 and 30, 34, 35, respectively, Appendix), fibronectin domains (gene number 31, Appendix), fibrinogen-related domains (gene numbers 36–38, Appendix) and laminin domains (gene numbers 39 and 40, Appendix). Among the 23 transcripts analysed, four were not detected (genes numbers 19, 28, 32 and 33, Appendix) and 10 showed differences in transcript levels that were greater than three-fold (Fig. 5). Among these latter genes, nine showed higher transcript level in susceptible snails with significant differences ( $P < 0.05$ ) varying between 3.83- and 23.57-fold (Fig. 5). Only one, the Bg-spondin domain containing protein, displayed a transcript content that was higher (4.33 more) in resistant snails (Fig. 5).

In order to validate the differences in transcript content between RNA from susceptible and resistant snails, we verified if PCR primers were not edited on sites containing mutations between the two strains. cDNA sequences from six EAF-S snails and six CB-R snails were obtained by PCR using primers

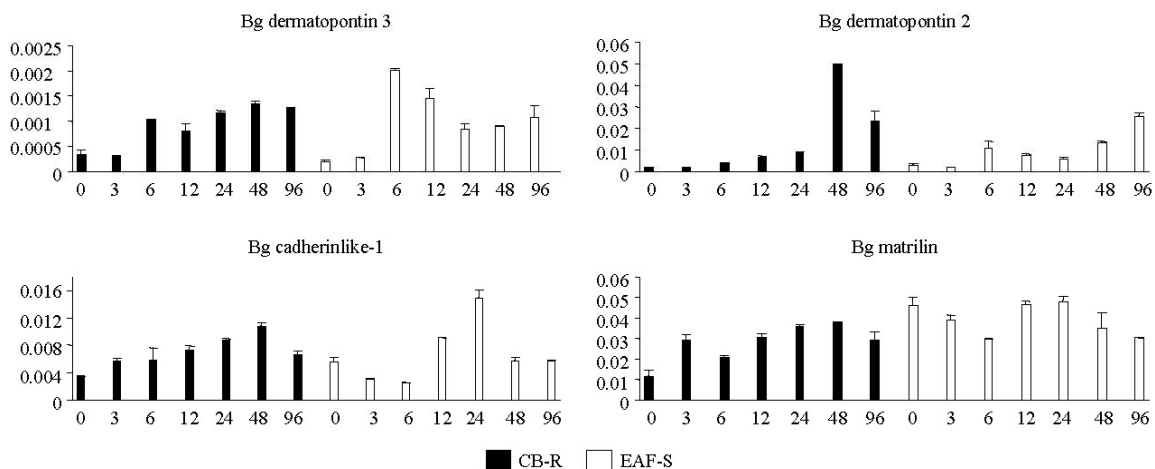


Fig. 4. Ratios of transcript levels corresponding to *Bg* dermatopontin-2 and -3, *Bg* cadherin like-1 and *Bg* matrilin in the resistant CB-R and susceptible EAF-S infected snails at different times post-exposure. Ratios were determined using real-time quantitative PCR and are expressed relative to S19 expression levels. Each histogram is the average value of triplicates  $\pm$  SD.

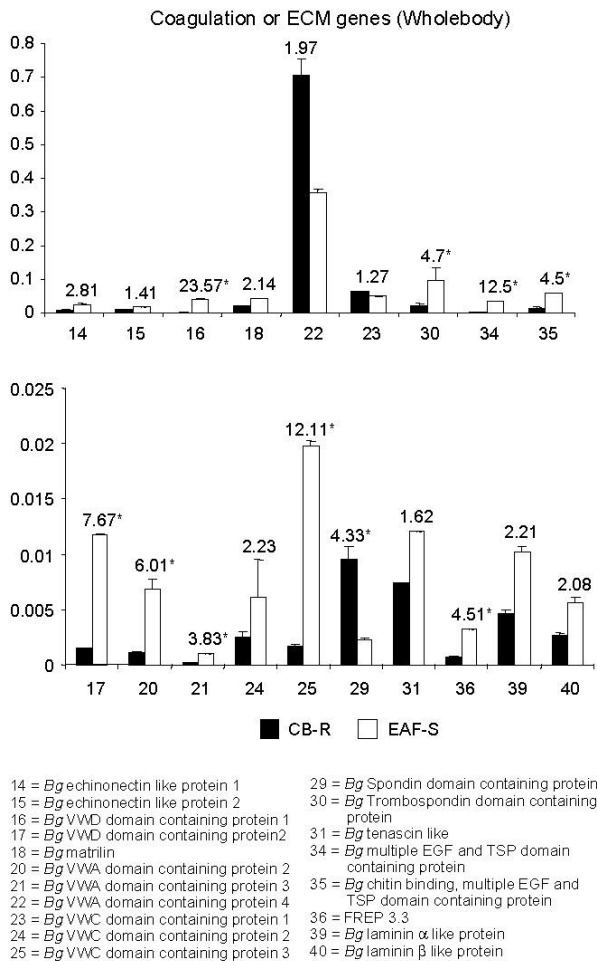


Fig. 5. Ratios of transcript levels corresponding to genes potentially involved in extracellular matrix (ECM) structure or coagulation processes in whole body resistant CB-R and susceptible EAF-S snails. Ratios were determined using quantitative PCR and are expressed relative to S19 levels. Each histogram is the average value of triplicates  $\pm$  SD. Differential levels of the transcripts between strains is indicated above the histograms. \*Genes showing differences in transcript levels that were significant ( $P < 0.05$ ) and greater than three-fold.

flanking the PCR primers used for quantitative PCR and sequenced (accession numbers, Table 3). Comparisons of sequences revealed no mutations in sequences used for Q-PCR.

Table 3

GenBank accession numbers of cDNAs corresponding genes characterised from whole body of CB and EAF snails

Gene	CB/EAF strains
Bg VWD domain containing protein1	DQ113398
Bg VWD domain containing protein2	DQ113400
Bg VWA domain containing protein2	DQ113401
Bg VWA domain containing protein3	DQ113402
Bg VWC domain containing protein3	DQ113403
Bg-spondin domain containing protein	DQ113404
Bg-thrombospondin domain containing protein	DQ113405
Bg multiple EGF and TSP domain containing protein	DQ113406
Bg chitin binding, multiple EGF and TSP domain containing protein	DQ113407
FREP 3.3	DQ113408

#### 4. Discussion

In order to investigate the molecular basis of compatibility in the *E. caproni*/*B. glabrata* model, we developed a comparative molecular approach to examine gene expression in the two strains of mollusc. As we observed previously there were some constitutive differences in adhesion capabilities of the hemocytes from the two snail strains (Humbert and Coustau, 2001), we compared the transcript content of a series of genes potentially involved in adhesion processes between hemocytes from susceptible and resistant snails. Transcripts of four genes were differentially present between hemocytes from resistant and susceptible snails. These genes encode proteins resembling dermatopontins (two), matrilin and cadherin.

The characterisation of the complete coding sequence (CDS) of dermatopontins (Bg dermatopontin-2 and -3) revealed that the precursors contain signal peptides suggesting that they could be constitutively secreted or stored and released from hemocytic granules. The structural features of the precursors suggest that these two new dermatopontins (Bg dermatopontin-2 and -3) share, in addition to dermatopontin general characteristics, some other particular features already observed in two other invertebrate dermatopontins from *S. domuncula* (cysteine array (Schutze et al., 2001)) and *L. polyphemus* (additional internal repeat (Fujii et al., 1992)). The function of these particular structural features has not been elucidated. Nevertheless, it was shown that these two dermatopontins triggered adhesive processes of dissociated cells from *S. domuncula* (Schutze et al., 2001) or hemocytes from *L. polyphemus* (Fujii et al., 1992). In addition, horseshoe crab dermatopontin is produced by hemocytes (as Bg dermatopontin-2 and -3) and stored in large secretory granules (Fujii et al., 1992). These features taken together suggest that *B. glabrata* hemocytic dermatopontins could participate in aggregation processes of hemocytes. Furthermore, we showed that the transcript content of these two genes was: (i) higher in hemocytes from unexposed resistant snails (in comparison with those of susceptible snails, Fig. 1); and (ii) greatly enhanced in resistant snails 48 h after exposure to the parasite (Fig. 4). Interestingly, previous histological studies showed that, in *E. caproni*-exposed resistant snails, the hemocytic capsule was mounted at 48 h post exposure (Ataev and Coustau, 1999). The fact that the higher level of expression of Bg dermatopontin-2 observed in resistant snails at 48 h after exposure (expression ratio of 0.050 vs 0.013 for susceptible) coincides with the capsule formation, further supporting the hypothesis that this gene may participate in adhesive processes taking place during capsule formation. Future studies will address this question by analysing gene expression using in situ hybridisation and determining whether or not Bg dermatopontin is over-expressed in capsule-forming hemocytes.

The characterisation of Bg matrilin showed that the precursor contains a signal peptide suggesting that it could also be constitutively secreted or stored in hemocytic granules. Analysis of the precursor suggests that this protein is a novel VWA domain containing protein. VWA domains are involved in cell adhesion, ECM structure and more generally in

multiprotein complex formation (Whittaker and Hynes, 2002). Bg matrilin shows similarities with matrilins from vertebrates which also contain two VWA domains and are involved in connecting cartilage components to form an integrated matrix network by forming a collagen-independent filamentous network (Chen et al., 1999). The function of Bg matrilin has yet to be elucidated. Nevertheless, our results allow us to hypothesise that this protein could be produced and secreted by hemocytes to participate in an induced filamentous network in the hemolymph of *B. glabrata* exposed to *E. caproni*. Such a network has already been described by Matricon-Gondran and Letocart (1999). The latter showed that infection by *E. caproni* induced the rapid appearance of tubular double-helical filaments in the hemolymph of *B. glabrata*. They showed that these filaments form bundles of a few parallel units associated with fine fibrils. They showed that hemocytes: (i) adhere to these bundles by means of the filopodia tip or by small areas of the cell surface; and (ii) become temporally sedentary after interaction. As we showed that the transcription of the gene corresponding to Bg matrilin is higher in naive susceptible snails (in comparison with resistant, Figs. 1 and 4), these susceptible snails could possess a more potent hemolymph coagulation system which could prevent hemocyte migration towards the parasite larvae and therefore facilitate parasite settlement in susceptible snails. This hypothesis is strengthened by our expression results from entire snails showing that a series of genes involved in coagulation processes or ECM structure have higher transcript levels in susceptible snails (Fig. 5). Future studies will focus on expression localisation of these different genes by in situ hybridisation and immunohistochemistry. This approach will be aimed at characterising, in situ, the potential involvement of these proteins in coagulation processes.

In addition to the expression studies, we observed a variability in the sequences obtained from hemocytes between the two strains. Among the six genes studied, numerous mutations were observed between sequences obtained from hemocytes of CB-R and EAF-S (data not shown, see Table 2 for accession numbers). Interestingly, such point mutations were not observed on transcripts recovered from whole snail bodies (10 genes). Recently, intense diversification through point mutations and recombinatorial processes was demonstrated for the fibrinogen-related proteins (FREPs) of *B. glabrata* (Zhang et al., 2004). Although the diversity observed in the present study on hemocyte adhesion genes only relies on point mutations, it raises the question of the existence of a gene wide diversification process taking place preferentially (or exclusively) in hemocytes of *B. glabrata*.

In conclusion, the present study has been successful in identifying adhesion genes that were differentially expressed in hemocytes and whole bodies of *E. caproni*-susceptible and -resistant *B. glabrata*. Complete molecular characterisation of these genes provided informative data on the potential functions of these proteins. According to these predicted functions, hypotheses were developed to explain their involvement in susceptibility/resistance processes and led to the conclusion that a hemolymph coagulation-like system may

play a critical role in the processes of susceptibility/resistance. Future functional studies will now focus on the role of these genes in the host–parasite interaction.

### Acknowledgements

This work was supported by the CNRS. We are grateful to Pierre Tisseyre (FR2577 CNRS) for performing DNA sequencing. We are indebted to Yannick Bueno for technical assistance.

### Appendix. Supplementary Material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ijpara.2005.09.009

### References

- Ataev, G.L., Coustau, C., 1999. Cellular response to *Echinostoma caproni* infection in *Biomphalaria glabrata* strains selected for susceptibility/resistance. *Dev. Comp. Immunol.* 23, 187–198.
- Chen, Q., Zhang, Y., Johnson, D.M., Goetinck, P.F., 1999. Assembly of a novel cartilage matrix protein filamentous network: molecular basis of differential requirement of von Willebrand factor A domains. *Mol. Biol. Cell* 10, 2149–2162.
- Deak, F., Wagener, R., Kiss, I., Paulsson, M., 1999. The matrilins: a novel family of oligomeric extracellular matrix proteins. *Matrix Biol.* 18, 55–64.
- Fujii, N., Minetti, C.A., Nakhasi, H.L., Chen, S.W., Barbehenn, E., Nunes, P.H., Nguyen, N.Y., 1992. Isolation, cDNA cloning, and characterization of an 18-kDa hemagglutinin and amebocyte aggregation factor from *Limulus polyphemus*. *J. Biol. Chem.* 267, 22452–22459.
- Goodman, T.G., Bajt, M.L., 1996. Identifying the putative metal ion-dependent adhesion site in the beta2 (CD18) subunit required for alphaLbeta2 and alphaMbeta2 ligand interactions. *J. Biol. Chem.* 271, 23729–23736.
- Guillou, F., Mitta, G., Dissous, C., Pierce, R., Coustau, C., 2004. Use of individual polymorphism to validate potential functional markers: case of a candidate lectin (BgSel) differentially expressed in susceptible and resistant strains of *Biomphalaria glabrata*. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* 138, 175–181.
- Humbert, E., Coustau, C., 2001. Refractoriness of host haemocytes to parasite immunosuppressive factors as a putative resistance mechanism in the *Biomphalaria glabrata*–*Echinostoma caproni* system. *Parasitology* 122, 651–660.
- Jeyarasigam, U., Heyneman, D., Lim, H.K., Mansour, N., 1972. Life cycle of a new echinostome from Egypt, *E. liei* sp. nov. (Trematoda; Echinostomatidae). *Parasitology* 65, 203–222.
- Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157, 105–132.
- Langand, J., Morand, S., 1998. Heritable non-susceptibility in an allopatric host–parasite system: *Biomphalaria glabrata* (Mollusca)–*Echinostoma caproni* (platyhelminth Digenea). *J. Parasitol.* 84, 739–742.
- Langand, J., Jourdan, J., Coustau, C., Delay, B., Morand, S., 1998. Cost of resistance, expressed as a delayed maturity, detected in the host–parasite system *Biomphalaria glabrata*/Echinostoma caproni. *Heredity* 80, 320–325.
- Lockyer, A.E., Noble, L.R., Rollinson, D., Jones, C.S., 2004. *Schistosoma mansoni*: resistant specific infection-induced gene expression in *Biomphalaria glabrata* identified by fluorescent-based differential display. *Exp. Parasitol.* 107, 97–104.
- Marxen, J.C., Nimtz, M., Becker, W., Mann, K., 2003. The major soluble 19.6 kDa protein of the organic shell matrix of the freshwater snail *Biomphalaria glabrata* is an N-glycosylated dermatopontin. *Biochim. Biophys. Acta* 1650, 92–98.

- Matricon-Gondran, M., Letocart, M., 1999. Internal defenses of the snail *Biomphalaria glabrata*. *J. Invertebr. Pathol.* 74, 248–254.
- Michishita, M., Videm, V., Arnaout, M.A., 1993. A novel divalent cation-binding site in the A domain of the beta 2 integrin CR3 (CD11b/CD18) is essential for ligand binding. *Cell* 72, 857–867.
- Miller, A.N., Raghavan, N., FitzGerald, P.C., Lewis, F.A., Knight, M., 2001. Differential gene expression in haemocytes of the snail *Biomphalaria glabrata*: effects of *Schistosoma mansoni* infection. *Int. J. Parasitol.* 31, 687–696.
- Neame, P.J., Choi, H.U., Rosenberg, L.C., 1989. The isolation and primary structure of a 22-kDa extracellular matrix protein from bovine skin. *J. Biol. Chem.* 264, 5474–5479.
- Nowak, T.S., Woodards, A.C., Jung, Y., Adema, C.M., Loker, E.S., 2004. Identification of transcripts generated during the response of resistant *Biomphalaria glabrata* to *Schistosoma mansoni* infection using suppression subtractive hybridization. *J. Parasitol.* 90, 1034–1040.
- Puzon-McLaughlin, W., Takada, Y., 1996. Critical residues for ligand binding in an I domain-like structure of the integrin beta1 subunit. *J. Biol. Chem.* 271, 20438–20443.
- Raghavan, N., Miller, A.N., Gardner, M., FitzGerald, P.C., Kerlavage, A.R., Johnston, D.A., Lewis, F.A., Knight, M., 2003. Comparative gene analysis of *Biomphalaria glabrata* hemocytes pre- and post-exposure to miracidia of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 126, 181–191.
- Schutze, J., Skorokhod, A., Müller, I.M., Müller, W.E., 2001. Molecular evolution of the metazoan extracellular matrix: cloning and expression of structural proteins from the demosponges *Suberites domuncula* and *Geodia cydonium*. *J. Mol. Evol.* 53, 402–415.
- Sminia, T., Barendsen, L., 1980. A comparative morphological and enzyme histochemical study on blood cells of the freshwater snails, *Lymnaea stagnalis*, *Biomphalaria glabrata*, and *Bulinus truncatus*. *J. Morphol.* 165, 31–39.
- Trouvé, S., Renaud, F., Durand, P., Jourdan, J., 1996. Selfing and outcrossing in a parasitic hermaphrodite helminth (*Trematoda Echinostomatidae*). *Heredity* 77, 1–8.
- Vergote, D., Bouchut, A., Sautiere, P.E., Roger, E., Galinier, R., Rognon, A., Coustau, C., Salzet, M., Mitta, G., 2005. Characterisation of proteins differentially present in the plasma of *Biomphalaria glabrata* susceptible or resistant to *Echinostoma caproni*. *Int. J. Parasitol.* 35, 215–224.
- Whittaker, C.A., Hynes, R.O., 2002. Distribution and evolution of von Willebrand/integrin A domains: widely dispersed domains with roles in cell adhesion and elsewhere. *Mol. Biol. Cell* 13, 3369–3387.
- Zhang, S.M., Adema, C.M., Kepler, T.B., Loker, E.S., 2004. Diversification of Ig superfamily genes in an invertebrate. *Science* 305, 251–254.



## Characterisation of proteins differentially present in the plasma of *Biomphalaria glabrata* susceptible or resistant to *Echinostoma caproni*

D. Vergote<sup>b</sup>, A. Bouchut<sup>a</sup>, P.E. Sautière<sup>b</sup>, E. Roger<sup>a</sup>, R. Galinier<sup>a</sup>, A. Rognon<sup>a</sup>,  
C. Coustau<sup>a</sup>, M. Salzet<sup>b</sup>, G. Mitta<sup>a,\*</sup>

<sup>a</sup>Parasitologie Fonctionnelle et Evolutive, UMR 5555, CNRS Université de Perpignan, 52 Ave Paul Alduy, 66 860 Perpignan cedex, France

<sup>b</sup>Neuro-immunologie des Annélides, UMR 8017, CNRS Université des Sciences et Technologies de Lille, Villeneuve d'Ascq, France

Received 21 September 2004; received in revised form 27 October 2004; accepted 3 November 2004

### Abstract

Snail immune responses towards a trematode infection are known to rely on both plasmatic and cellular host factors. As an approach to further investigate the suspected involvement of plasmatic factors in *Biomphalaria glabrata* resistance/susceptibility to *Echinostoma caproni*, we compared protein patterns of plasma collected from susceptible and resistant snails. This proteomic approach revealed that 13 plasmatic proteins exhibited significant differences in their apparent representativity. The genes corresponding to five of them were characterised by a combination of mass spectrometry and molecular cloning. They encode two isoforms of a glycolytic enzyme, two isoforms of a calcium binding protein and an inhibitor of cysteine protease. Furthermore, we investigated gene expression in parasite-exposed or -unexposed snails as well as in various tissues by quantitative PCR. This study showed that: (i) differential representation of plasma proteins between the snail strains was correlated with a differential level of transcripts; (ii) expression of these genes after parasite exposure was differentially regulated in the two strains; and (iii) these genes were expressed predominantly in the albumen gland.

© 2004 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

**Keywords:** *Biomphalaria glabrata*; *Echinostoma caproni*; Host-parasite interactions; Proteomic; Susceptibility; Albumen gland

### 1. Introduction

Snail-trematode immunobiological interactions have been studied on the fresh-water snail *Biomphalaria glabrata* following infection by the human blood fluke *Schistosoma mansoni* and by the echinostome species, *Echinostoma paraensei* and *Echinostoma caproni*. In contrast to *S. mansoni*, these two echinostomes are known to strongly interfere with the snail immune response via their excretory-secretory (ES) products (Loker and Adema, 1995; Humbert and Coustau, 2001). Analysis of interactions between *E. paraensei* ES products and *B. glabrata* plasmatic factors led to the identification of the numerous fibrinogen-related proteins

(FREPs) (Adema et al., 1997; Leonard et al., 2001; Zhang et al., 2001, 2004; Zhang and Loker, 2003). At least some of these extremely diversified molecules are capable of binding and precipitating *E. paraensei* ES products, suggesting that they may play an important role in snail-trematode molecular interactions. In addition to the discovery of FREPs, previous studies provided evidence that plasmatic factors may play an important role in the outcome of a trematode infection. In vivo comparison of *E. caproni* development in *B. glabrata* snails selected for their susceptibility or resistance (Langand et al., 1998) revealed that, in resistant snails, sporocysts were abnormally developed and degenerated regardless of their level of encapsulation (Ataev and Coustau, 1999). These results suggested that, in addition to cellular factors, humoral factors are actively involved in the molecular processes underlying resistance or susceptibility to *E. caproni* (Ataev and Coustau, 1999).

\* Corresponding author. Tel.: +33 4 68 66 2188; fax: +33 4 68 66 2281.  
E-mail address: [mitta@univ-perp.fr](mailto:mitta@univ-perp.fr) (G. Mitta).



Except for the presence of FREPs, virtually nothing is known on the composition of plasma, and plasmatic factors remain uncharacterised.

In the present study, we addressed the question of potential constitutive differences between plasma factors of resistant and susceptible snails. We used, for the first time in this model, a proteomic approach aimed at: (i) identifying proteins differentially represented in the plasmas of susceptible and resistant snails; (ii) characterizing the genes encoding these proteins; and (iii) analyzing their expression p.i. and in a range of tissues.

## 2. Material and methods

### 2.1. *Biomphalaria glabrata* strains and plasma preparation

*Biomphalaria glabrata* snails used in this study belonged either to the unselected Bg.Bra stock originating from Brazil (Guillou et al., 2004), or to the EAF-S or CB-R strains selected, respectively, for their susceptibility or resistance to *E. caproni* (Langand et al., 1998; Humbert and Coustau, 2001). At the time of the study, the percentages of adult snails susceptible to *E. caproni* were: Bg.Bra = 70%, EAF-S = 98% and CB-R = 1%.

Hemolymph was extracted from the headfoot of 1,000 uninfected EAF-S and 1,000 uninfected CB-R snails according to standard procedure (Sminia and Barendsen, 1980) and immediately placed on ice. Approximately 100 ml of hemolymph were recovered from each strain and centrifuged at  $800\times g$  for 10 min at 4 °C. The supernatant (plasma) was collected, centrifuged at  $20,000\times g$  for 20 min at 4 °C and filtered using a stirred ultrafiltration cell with a cut-off 500 kDa (Millipore). The filtered plasma was then stored at –80 °C until use.

### 2.2. Experimental infection

The parasite *E. caproni* (Jeyarasingam et al., 1972) was maintained in Bg.Bra snails and mice (SWISS OF1 stock) as described previously (Trouvé et al., 1996). For one experiment, adult EAF-S and CB-R snails (measuring 9–14 mm in shell diameter) were exposed to *E. caproni* infection according to previously described procedures (Guillou et al., 2004). Ten individuals from each strain were collected at 3, 6, 12, 24, 48 and 96 h post-exposure and frozen in liquid nitrogen until RNA extraction.

### 2.3. Two-dimensional gel electrophoresis

#### 2.3.1. Plasmatic protein extracts

Plasmatic proteins were precipitated by adding four volumes of a solution containing 10% trichloroacetic acid, 0.07%  $\beta$ -mercaptoethanol in acetone to 3 or 9 ml of plasma (for analytical or preparative electrophoresis, respectively). After 1 h at –20 °C, proteins were pelleted by centrifugating

at  $15,000\times g$  for 30 min at 4 °C. Four volumes of acetone containing 0.07%  $\beta$ -mercaptoethanol were added to the pellet and incubated 1 h at –20 °C. After centrifugation ( $15,000\times g$  for 30 min at 4 °C), the pellet was air-dried and incubated 90 min at room temperature in 350  $\mu$ l of ‘lysis’ buffer (9 M urea, 2% Triton X-100 and 10 mM Dithiothreitol (DTT) (Wu et al., 2000).

#### 2.3.2. Isoelectrofocusing (IEF)

IEF for the 2D gel electrophoresis was performed using the IEFCell system™ (Biorad). Immobilised pH 3–10 non-linear gradient (IPG) strips were rehydrated with the 350  $\mu$ l of protein extracts to which 0.2% v/v Bio-lyte (pH 3–10:pH 5–8:pH 7–9, 1:1:2, Biorad) were added. Passive rehydration was performed for 5 h at 20 °C, followed by an active rehydration at 50 V for 14 h at 20 °C. Isoelectrofocusing was raising gradually until 8,000 V and running until 90,000 V h at 20 °C. Strips were then incubated twice with 130 mM DTT in equilibration buffer (6 M urea, 375 mM Tris–HCl, 2% SDS, 20% glycerol, 0.02% Coomassie blue G-250, pH 8.8) for 15 min according to Wu (Wu et al., 2000). Proteins were then carbamidomethylated for 15 min (135 mM iodoacetamide in equilibration buffer).

#### 2.3.3. SDS-PAGE

IPG strips™ were transferred on a 12% polyacrylamide gel containing 0.8% of cross-linker piperazine diacrylamide in a 0.5% agarose low melting point gel. Electrophoresis were performed in Tris/Glycine/SDS buffer according to the conditions defined by O’Farrell (O’Farrell, 1975). Proteins were separated 1 h at 30 V and then at 200 V until Coomassie blue has reached the bottom of the gel. After electrophoresis, gels were silver stained according to a modified protocol of Morrissey (Morrissey, 1981) or Coomassie blue stained according to Neuhoff (Neuhoff et al., 1988).

#### 2.3.4. 2D image analysis

Three gels per condition were analyzed using the Progenesis™ v1.5 software program (Nonlinear dynamics, Newcastle upon Tyne, UK). Analysis protocol included spot detection and filtering, whole image warping on a reference gel, background subtraction, average gel creation, spot matching and volume normalization against total volume of all protein spots present in the gel. Each analysis step was manually validated. Statistical significance was assessed using a Student’s *t*-test.

### 2.4. Protein identification

#### 2.4.1. In-gel digestion for mass spectrometry

Proteins of interest were excised and gel pieces were washed in 25 mM ammonium bicarbonate/50% acetonitrile. After drying, gel pieces were placed for 30 min on ice in 50  $\mu$ l of protease solution (sequence grade modified trypsin, Promega, at 0.02 mg/ml in 25 mM ammonium bicarbonate).

Excess protease solution was then removed and replaced by 50  $\mu$ l of 25 mM ammonium bicarbonate. Digestion was performed at 30 °C overnight. Peptide extraction was performed twice for 15 min with 50% acetonitrile/1% trifluoroacetic acid (TFA) for further Matrix-Assisted Laser Desorption/Ionization-Time of Flight Mass Spectrometry (MALDI-ToF MS) analysis or with 50% acetonitrile/1% formic acid for further ESI-MS-MS (Electro Spray Ionization-Mass spectrometry-Mass spectrometry) analysis. Trypsin digests were then lyophilised in a SpeedVac concentrator (Savant) and resuspended in 5  $\mu$ l of 0.1% TFA or 0.1% formic acid.

#### 2.4.2. MALDI-ToF MS

MALDI-ToF MS analysis of trypsin digests was performed on a Voyager DE Pro™ (Applied Biosystems) in reflector mode at an accelerating voltage of 20 kV. One microliter of trypsin digests was spotted on 1  $\mu$ l of dried  $\alpha$ -cyano-4-hydroxycinnamic acid (HCCA, 15 mg/ml in acetone) and covered with 1  $\mu$ l of HCCA (10 mg/ml in 70% acetonitrile/0.1% TFA). About 300 laser shots were accumulated to obtain the final spectrum. Peptide masses were determined after peak smoothing and internal calibration using the two autolysis trypsin fragment [M + H]<sup>+</sup> = 2211.1 and 842.51 Da.

#### 2.4.3. ESI-MS-MS analysis

After desalting by on reversed-phase C18 Zip-Tip micro-column (Millipore, Bedford, MA, USA), samples were loaded into nanoESsray capillaries (Protana, Odense, Denmark) using a 5  $\mu$ l on-column syringe. Capillaries were inserted into an Applied Biosystems Q-STAR Pulsar™ using an Ion spray source (nanoESI-Q-ToF). Doubly-charged peptides were selected, fragmented by N<sub>2</sub> collision and analysed by ESI-MS-MS. The sequences obtained were manually assigned to search against SWISSPROT, NCBI and GenPept and the division dbEST of GenBank (ESTs of

*B. glabrata*) using MS-Pattern (<http://prospector.ucsf.edu/ucsfhtml4.0/mspattern.htm>).

#### 2.5. RNA extraction

RNA extraction, and reverse transcription were performed from whole body *B. glabrata* according to previously described procedures (Guillou et al., 2004).

#### 2.6. Gene characterization

First-strand cDNAs were prepared from total RNAs with oligo d(T)<sub>17</sub> for primer using the Superscript II™ (Invitrogen). The 3' portion of the cDNAs were amplified by PCR using an oligo d(T)<sub>17</sub> and the different degenerate primers designed from the tags obtained by ESI-MS-MS (Table 1). PCR was performed in 50  $\mu$ l mixtures containing 2  $\mu$ l of first-strand cDNA using the DyNAzymeEXT™ kit (Finnzymes). Primer concentration of 2 and 0.2  $\mu$ M were used for the degenerate primers and the oligo d(T)<sub>17</sub>, respectively. In order to clone the 5'- or 3'-ends of the cDNAs, we used the GeneRacer™ Kit (Invitrogen).

The PCR products were ligated into the pCR2.1-TOPO™ vector according to the manufacturer's instructions (Invitrogen). Dideoxy sequencing reactions of the recombinant plasmids were sequenced using a dideoxy-dye-terminator method (CEQ™ DTCS-Quick Start kit, Beckman Coulter) and a CEQ™ 8000 apparatus (Beckman Coulter). M13 forward and reverse primers were used for sequencing. Sequences were obtained using the CEQ™ 8000 sequence analysis software.

#### 2.7. Real-time PCR

Real-time PCR analysis of gene expression was performed on total RNAs extracted from parasite-exposed

Table 1  
Sequences obtained by ESI-MS-MS and degenerated primers designed for cDNA cloning

Spot	Sequences	Degenerate primers
1	[PND]S[FM]Y[QK][QK][QK]G...T[IL]ND...[IL][FM]AVK HATS[QK]V-	
3	[IL][FM]DEVGD [QK][QK][FM]DSGWR	HTIWTBGAYGARGTIGGIGAY
4	YNEGNAAYENK	GGIAAYGCITAYGARAAYAAR
6	DVDGDGN[IL]EPR	GAYGTIGAYGGIGAYGGIAAY
7	DVDGDGN[IL]EPR	GAYGTIGAYGGIGAYGGIAAY
8	[IL][FM][DA/W][IL]DY	
31	[IL][QK]W[FM][QK]YDD SAR	MARTGGTTYMARTAYGAYGAY
43	ns	
48	V[IL]TP[IL]ATAVK [IL]DG[IL][IL]VDER [FM]VNW[QK]AAA[IL]K	YTIGAYGGIYTIATHGTIGAYGA TTYGTIAAYTGGCARGCIGC
51	ns	
58	[FM][IL]NW[QK]ADA[IL]K DSG[FM][IL]VTVGSWNPK YSYYDT[IL]R [IL]DG[F- M][IL]VDWR	TTYHTIAAYTGGCARGCIGA
137	ns	
138	ns	

ns, no sequence obtained by ESI-MS-MS; [], ambiguous amino acids are in square brackets; M, oxydised methionine; wobble, M = A + C, R = A + G, W = A + T, Y = C + T, H = A + C + T, B = G + T + C, I, inosine.

Table 2  
Specific primers used for real time PCR

Gene	Primer name and sequence
RiboS19	R19F1: TTCTGTTGCTCGCCAC R19R1: CCTGTATTGTCATCCTGTT
SPOT1	S1F1: CCATGTTGAAGATTGCCCTAG SR1: CACCATTCAAGTCAGCGGCG
SPOT58	DisS58F1: GCACTCAGCAGGAGACTAT DisS58R2: GACCATTCCAATTGTAATTTTGG
SPOT48	DisS48F1: GCATTCTGCTGGAGATTTC DisS48R2: TTACACCATTAGAAGTTAAGC
SPOT6	DisS6F2: AACTAACGGAGACAATCA DisS6R2: CTGATTGGTCAAGGAATCC
SPOT7	DisS7F2: CGCCGACGGTGACGG DisS7R2: GCACTGGTCAAGATGGTT

EAF-S and CB-R snails. RNA extraction, reverse transcription, PCR and relative quantitation were performed according to previously described procedures (Guillou et al., 2004) with a Light cycler™ apparatus (Roche Molecular Biochemicals, Germany). Specific primers for real-time quantitative PCR (Table 2) were edited using the Light Cycler Probe Design™ Software version 1.0.

### 3. Results

#### 3.1. Differentially represented proteins in plasma from susceptible and resistant strains

The 2D electrophoretic profiles of plasmatic proteins of CB-R and EAF-S strains is illustrated in Fig. 1A. Image analysis revealed 197 plasmatic proteins on the gels. Among these spots, we selected all spots showing a differential representation greater than 2.5-fold. Statistical analysis revealed that differences observed for the following 13 spots—1, 3, 4, 6, 7, 8, 31, 43, 48, 51, 58, 137 and 138—were significant ( $P < 0.05$ ). Ten of these proteins were more represented in the plasma of susceptible snails and three were more represented in the plasma of resistant snails (Fig. 1B and C).

In order to identify these proteins, we first performed an in-gel digestion with trypsin. Trypsin digest were then analysed by ESI-MS-MS to obtain sequence tags for these proteins (Table 1). These sequences were compared to databases using MS-Pattern software. Sequence tags obtained for spot 1 matched with four ESTs of *B. glabrata*. These ESTs (accession numbers: AA547743, AA547744, AA547745 and AA547746) correspond to the 5' portion of cDNAs displaying high similarities to molecules belonging to the cystatin family of protease inhibitors. For example, AA547745 displayed an *E*-value of  $1 \times 10^{-5}$  for a cystatin of *Drosophila melanogaster* (accession number: S12913). In order to better characterise the gene encoding this protein, RT-PCR experiments were performed using primers designed from the ESTs

(see Fig. 2). From the PCR product sequenced, we designed forward primers which were used to obtain the 3' portion of the cDNA. This cDNA is 459 bp in length and encodes a 117 amino acids precursor (Fig. 2). The deduced amino acids sequence begins with a 18 residue signal peptide. The cleavage site for signal peptidase is most likely located after the serine residue preceding the glutamine in position 19 as predicted by the SignalP 3.0 server (<http://www.cbs.dtu.dk/services/SignalP/>) and PSORT II prediction program (<http://psort.nibb.ac.jp/form2.html>). This signal peptide is directly flanked by a carboxyl-terminal 99 amino acid sequence. This protein has a calculated monoisotopic mass of 10,509.26 Da and a theoretical pI of 6.74, which are in good agreement with the position of the protein in the gel. Theoretical masses of peptides generated by the PeptideMass software (<http://www.expasy.org/tools/peptide-mass.html>) were compared to the experimental peptide mass fingerprint of spot 1 obtained by MALDI-ToF mass spectrometry. Four common fragments were identified covering 49% of the protein (signal peptide excluded). Combined results from cDNA cloning and mass spectrometry therefore confirmed that the spot 1 is a novel *B. glabrata* protein belonging to the type-2 cystatin family (Bg type-2 cystatin, GenBank accession number AY678119).

The sequence tags obtained from the other spots were insufficient to identify the other proteins using the same approach. Consequently, tags obtained from spots 3, 4, 6, 7, 31, 48 and 58 were used to design the degenerate forward primers shown in Table 1. These primers and a d(T)<sub>17</sub> primer were used in RT-PCR experiments to amplify cDNA fragments corresponding to the 3' portion of the genes encoding these different proteins. After cloning and sequencing, reverse primers were designed from these sequence and 5'-RACE PCR experiments were performed to obtain the complete open reading frame. cDNAs corresponding to the spots 6, 7, 48 and 58 were characterised.

cDNAs corresponding to the proteins 6 and 7 (GenBank accession numbers: AY678123 and AY678120) encode precursors displaying high degrees of similarities (Fig. 3). They encode two isoforms containing four EF-hand domains (detected by RPS-BLAST program, <http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) characteristic of Calcium Binding Proteins (CaBPs). These precursors are highly similar to other CaBPs like calmodulins for example. Furthermore, these proteins are probably secreted as predicted by the PSORT II prediction program (<http://psort.nibb.ac.jp/form2.html>) and which is in agreement with the plasmatic location of the proteins. These two proteins were named Bg CaBP 1 (spot 6) and 2 (spot 7).

cDNAs corresponding to the proteins 48 and 58 also encode two highly similar isoforms (Fig. 4). Confrontation of the deduced amino acids sequence to the databank reveal high similarities to an endo-1,4-β-mannanase of the bivalve

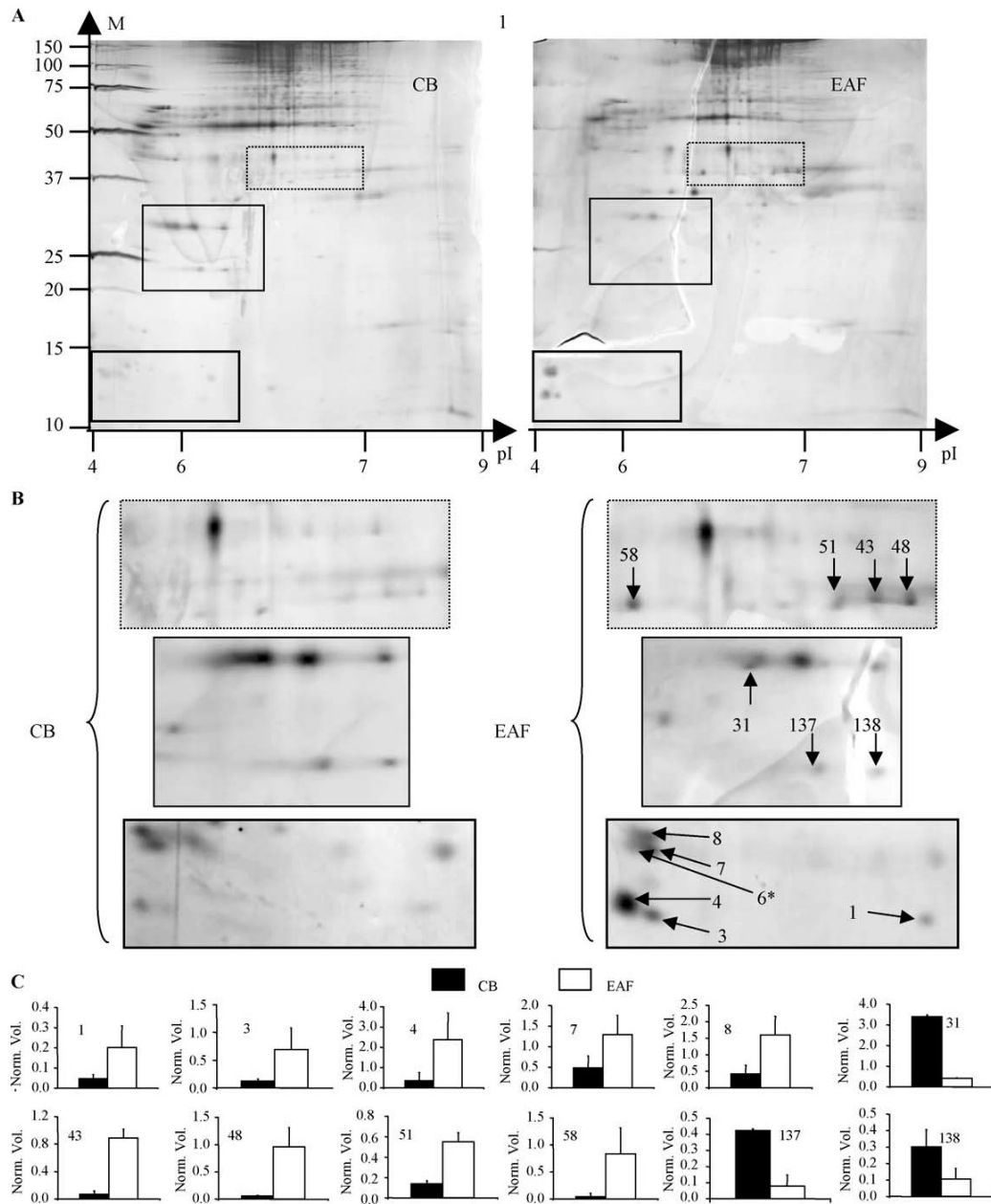


Fig. 1. Comparison of plasmatic protein patterns from uninfected susceptible and resistant snails. (A) Silver-stained 2D gels of plasmatic proteins from *Biomphalaria glabrata* resistant (CB-R) and susceptible (EAF-S) snails. Molecular mass (M) calibration is indicated on the left of the gel and pI calibration on the bottom of the gel. The pI calibration was done with pI standards run in the same conditions on a separate gel. Areas containing differentially represented proteins are boxed. (B) Enlarged views of the areas of interest. Positions of differentially represented proteins are indicated by arrows. \* Spot 6 was not detected on CB-R gels. (C) Normalised volume of plasmatic proteins differentially represented between CB-R and EAF-S quantified with Progenesis™ software. The figure shows the average values obtained from three gels for each condition. \* Significant differences when using the Student *t*-test ( $P < 0.05$ ).

*Mytilus edulis* ( $E$ -value =  $1 \times 10^{-106}$ , Swiss-Prot accession number Q8WPJ2, Xu et al., 2002). These two proteins were named endo-1,4- $\beta$ -mannanase 1 and 2 (GenBank accession numbers: AY678121 (S58) and AY678122 (S48)). Note that S58 is also probably secreted as predicted by the PSORT II prediction program.

### 3.2. Differential plasmatic representation of the proteins is correlated to a differential transcription level between the two strains

The expression ratio of each gene to ribosomal protein (S19) appears substantially higher in susceptible snails

**ccatgttgaagattgccctagttgtgttgctgggtgttgcctttcatttcaagtcagctc**  
 M L K I A L V V L L A V V P F I S S Q L  
 gttggaggtcaggtgccttactactaccctttaaataatgaccccaatgttttgttgcgctc  
 V G G Q V P Y T T T L N D P N V L F A V  
 aaagcaatcaacagcttctatcagcaacaaggggacaacaacctgagaacaggaggtcaaa  
 K A I N S F Y Q Q Q G D N N L R T G V K  
 attgtccacgcccacatctcaggtggtgctggtgccttgccttatagattcaccattcaagtc  
 I V H A T S Q V V A G A L Y R F T I Q V  
**agcggcgggaagcagcaagcgaagattgtacagtagccgtctggagccgccatggctatct**  
 S G G S T N E D C T V A V W S R P W L S  
 ggtaatgaagccaccaactacaggggaacaccatcctgtgtagcatcagcataaataact  
 G N E A T Q L Q G T P S C V A S A  
 gccaaacttatttattgattttttttattataagataagcttaaactatttcaaaata  
aacatttaatacgggaattcaaaaaaaaaaaaaaaaaa

Fig. 2. cDNA (complete coding sequence and 3' untranslated region) corresponding to the protein 1. The deduced amino acids sequence is shown in upper case letters. The precursor displays a signal sequence highlighted in pale grey. The domain shared by all cysteine protease inhibitors is boxed. The amino acids shared by type 2 cystatins are highlighted in dark grey. Fragments of the protein identified by analysis of mass spectrometry data are underlined (mass fingerprint and tags obtained by MALDI-ToF and ESI-MS-MS, respectively). Polyadenylation signal is double-underlined. The primers designed from the ESTs are indicated in italic and bold.

(Fig. 5A). These results are in agreement with the differential representation of the corresponding protein quantified using the Progenesis™ software (Fig. 5B). Therefore, differences in plasmatic protein content appear correlated with differences in expression ratios of the corresponding genes.

### 3.3. Effect of infection on expression ratios

Expression ratio of the different genes were analysed in susceptible and resistant snails at different times (0, 3, 6, 12, 24, 48 and 96 h) post-exposure to *E. caproni* (Fig. 6). Expression ratios of all genes are substantially lower in uninfected resistant than in uninfected susceptible snails. A global analysis of expression ratios during the kinetics of infection reveals a significant increase of transcript levels for the different genes in resistant snails. Indeed, transcripts corresponding to proteins 1, 6, 7, 48 and 58 are 2.8, 16.9, 6.5, 5.4 and 4.8 times more abundant at different times after infection of resistant snails, respectively. In contrast, in susceptible snails, significant decreases are observed for all the genes following parasite exposure.

### 3.4. Comparison of expression ratios in various snail tissues

To investigate tissue transcription of the genes encoding proteins 1, 6, 7, 48 and 58, quantitative PCR analyses were

carried out using total RNA from the various snail tissues representative of the entire snail (Fig. 7). For the five genes, the highest expression ratios were observed in the albumen gland. Weak amounts of transcripts were detected in other tissues, particularly in hemocytes and head-foot for the gene encoding protein 1 and in head-foot for the gene encoding protein 7.

## 4. Discussion

Although, the importance of both humoral and cellular factors in determining the success or failure of a trematode infection has been highlighted for a number of years, past molecular studies mainly focused on the cellular effectors (Miller et al., 2001; Raghavan et al., 2003). Except for the highly diversified family of plasmatic fibrinogen-related proteins (Adema et al., 1997; Zhang et al., 2004), little is currently known on humoral factors.

As an approach to further investigate the suspected involvement of plasmatic factors in *B. glabrata* resistance/susceptibility to *E. caproni*, we compared protein patterns of plasma collected from susceptible and resistant snails. This proteomic approach revealed that 13 out of the 197 plasmatic proteins exhibited significant differences in their apparent representativity.

Characterization of the cDNA corresponding to the spot 1 has been possible by direct comparison of tags obtained by

Spot 7	MFAYALLICLPMLAFGQDFLANQEFNR	<b>FDVDGDGNI</b> EPREV <b>QQYFRRF</b> DADGDGR <b>ISRQE</b>
Spot 6	MFAYALLICLPMLAFGQDFLANQEFNR	<b>FDVDGDGNI</b> EPREV <b>QQYFRRF</b> <b>DTNGDNQISRQE</b>
Spot 7	<b>YRDEVDTHH</b> IN <b>FPEAHR</b> ALLR <b>LFDE</b> IDYDNDNHL <b>DQSDY</b> NT <b>LFNNA</b> DANQNDL <b>VNHNE</b> LR	
Spot 6	<b>YRQEVDT</b> HL <b>NNPETH</b> RALLR <b>LFDE</b> V <b>DYNN</b> D <b>GFLDQSD</b> FN <b>KLF</b> T <b>NAD</b> ANQNDL <b>VNHNE</b> LV	
Spot 7	-----	
Spot 6	TYFHQLTGGPVIG	

Fig. 3. Deduced amino acid sequences and alignment of protein 6 and 7 precursors (complete and partial CDS, respectively). Conserved amino acids are highlighted in grey. Signal sequences are highlighted in dark grey. The four EF-hand domains are boxed. Theoretical tryptic fragments matching with the peptide mass fingerprint obtained after tryptic digestion of these proteins are represented in italic and bold. These matching fragments cover 39 and 56% of the protein 7 and 6, respectively (signal peptide excluded).

Spot58	<b>MKTLITGFLVVLCTLKLVCARLAVSGNQFTYNGQRI</b> FLSGGNLPWIQYAYDFGDHQWDSR
Spot48	-----
Spot58	<b>KGTFENQLTQLKN</b> AGGNSIRLWVHIQGESTPAFDGNGYVTAPDHQGTLINDFKDMLDIAQ
Spot48	-----
Spot58	<b>RHNILVFPPTLWNAAVDQDNSHRLDGFIVDWRK</b> LQSYIDKALVPLASAVR <b>GHPALGAWDIM</b>
Spot48	----- <b>LDGLIVDERK</b> LQSYIDKVLTP <b>LATAVKGHPALGAWDIM</b>
Spot58	<b>NEPEGMINTDISN</b> WDRCYDSTALKNSGAGWAGK <b>KYSYD</b> TLRFINWQ <b>ADAIKNVDSGFLV</b>
Spot48	<b>NEPEGMINPDIGNS</b> DRCYDATALKNSGAGWAA <b>KKYG</b> HDIIRFV <b>NWQAAAIKHVDPGFLV</b>
Spot58	<b>TVGSWNPKSNTDQFGFVDHYSDNCLV</b> KLKPKNGK <b>LDFYQFHTYSYQGNF</b> DNVSPFKHSAG
Spot48	<b>TVGAWNPKSNTDRFGFVDHYS</b> DA <b>CLLKGKPKNGK</b> LDFY <b>QVHSYSYQGNF</b> DNVSPFKHSAG
Spot58	DYGTGKPIVVG <b>EFWEQ</b> GGGMNIDQLFDYVY <b>NHGYAG</b> AWSD <b>LM</b> AHGDNQRGGISHIKNY
Spot48	DFGTGKPIVVG <b>EFWEQ</b> GGGMNINQLFEYVY <b>NHGYAG</b> AWSD <b>LQ</b> AHGANQRGGISHIKGL
Spot58	NW <b>NGQIG</b> INL
Spot48	TS <b>NGVIPI</b> INV

Fig. 4. Deduced amino acid sequences and alignment of partial and complete CDS of proteins 48 and 58, respectively. Conserved amino acids are highlighted in pale grey. Signal sequence of protein 58 is highlighted in dark grey. Theoretical tryptic fragments matching the peptide mass fingerprint obtained after tryptic digestion of these proteins are represented in italic and bold. These matching fragments cover 48 and 45% of the proteins 48 and 58, respectively.

ESI-MS-MS to databases. This cDNA has been identified as encoding a type-2 cystatin. Cystatins are cysteine protease inhibitors that have been reported in all *phyla* of living organisms (Abrahamson et al., 2003). They control activity of cathepsins, themselves involved in various biological processes, including immune processes (Abrahamson et al., 2003). Some parasites such as the nematodes *Brugia malayi*, *Onchocerca volvulus*, *Acanthocheilonema vitea* or *Nippostrongylus brasiliensis* are known to secrete cystatins that inhibit hosts immune responses (Hartmann and Lucius, 2003). Whether the *B. glabrata* cystatin characterised here is involved in regulation of immune processes remains to be elucidated. However, its higher representativity in the plasma from susceptible snails is intriguing as it could reflect a constitutively stronger down-regulation of immunity facilitating parasite establishment in the first hours following infection.

cDNAs corresponding to the spots 6 and 7 are isoforms of a CaBP of the EF-hand type. CaBP play a key role in calcium

signaling and homeostasis. Because the CaBP identified here are extracellular proteins located in the plasma, it is more likely that they participate in calcium homeostasis. However, links between differential representation of these proteins and susceptibility/resistance may exist. A first possibility is that these proteins are involved in regulation of adhesion processes. Cellular adhesion is often calcium-dependant and may be regulated by calcium-dependant molecules such as C-lectins in *B. glabrata* (Duclermortier et al., 1999). Because adhesive processes are optimal for a given concentration range of extracellular calcium (Johansson and Soderhall, 1988), differences in some extracellular CaBP content may result in differences in extracellular calcium concentration and therefore in differential regulation of adhesive processes. A second possibility is that these CaBP are antiparasitic. Such activity has been shown for the human plasmatic CaBP, Calgranuline C. This protein: (i) is one of the initiation factors of *S. mansoni*-induced granulomas (Yang et al., 1997); (ii) is involved in

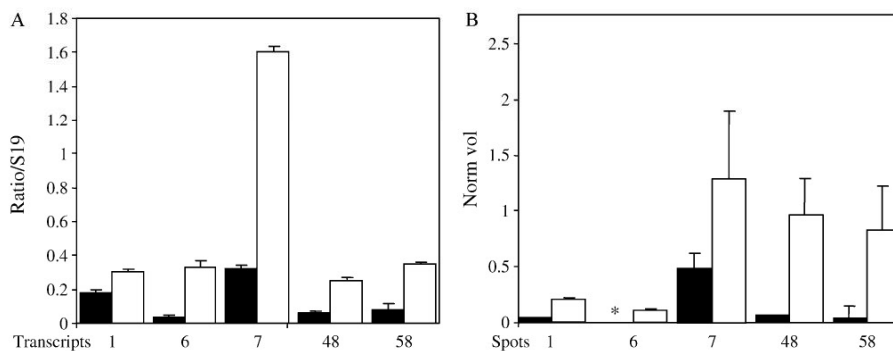


Fig. 5. Differential abundance of transcripts (A) and the corresponding plasmatic proteins (B) in the resistant CB-R (dark boxes) and susceptible EAF-S (light boxes) uninfected snails. (A) Transcript ratios were determined using real-time quantitative PCR and are expressed relative to S19 (ratio/S19). Each histogram is the average value of triplicates. (B) Normalised volumes (volume of each spot/all spot volumes added) of plasmatic proteins calculated with the Progenesis™ software. Histograms represent the average values from three gels. \*, not detected.

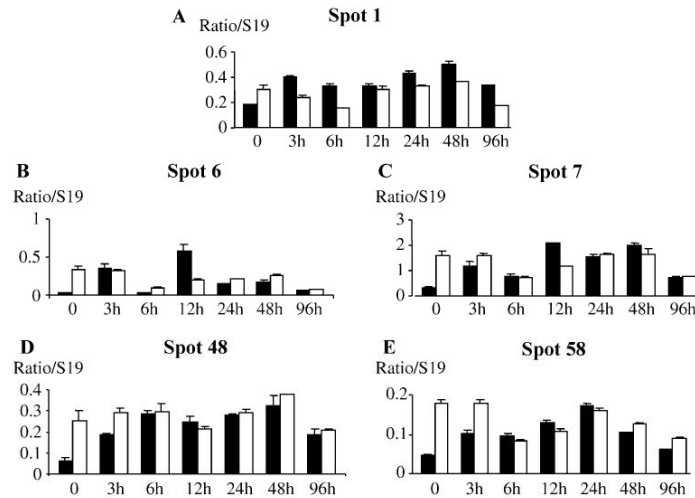


Fig. 6. Ratios of transcripts corresponding to spots 1 (A), 6 (B), 7 (C), 48 (D) and 58 (E) in the resistant CB-R (dark boxes) and susceptible EAF-S (light boxes) infected snails at different times post-exposure. Ratios were determined using real-time quantitative PCR and are expressed relatively to S19 expression levels (ratio/S19). Each histogram is the average value of triplicates  $\pm$  SD.

the inflammatory response against *O. volvulus* (Marti et al., 1996) and; (iii) shows a filaricidal action against *B. malayi* (Gottsch et al., 1999). In our model, the massive induction of the gene encoding protein 6 in resistant snails 12 h p.i. could directly participate to the elimination of this parasite.

Two other cDNAs corresponding to spots 48 and 58 encode two isoforms of an endo-1,4- $\beta$ -mannanase which displays significant similarities with a mannanase from *Mytilus edulis* (Xu et al., 2002). This type of molecule randomly cleaves within the  $\beta$ -1,4-mannan main chain of galactomannan, glucomannan, galactoglucomannan and

mannan. It has been purified from a range of plants, fungi, and bacteria, and has been reported in gastropod snails such as *Pomacea insularis* and *Littorina brevicula* (Yamaura et al., 1997; Yamaura et al., 1996). Mannanases could be involved in the *B. glabrata*/*E. caproni* interactions as carbohydrates are known to participate to helminth immune-evasion strategies (Harnett and Harnett, 2001; Haslam et al., 2001; Khoo and Dell, 2001). For example, previous results suggested that a glycosylated polypeptide was at least partly responsible for the immunosuppressive activity of *E. caproni* ES products on *B. glabrata* hemocytes (Humbert and Coustau, 2001). Such glycosidases could

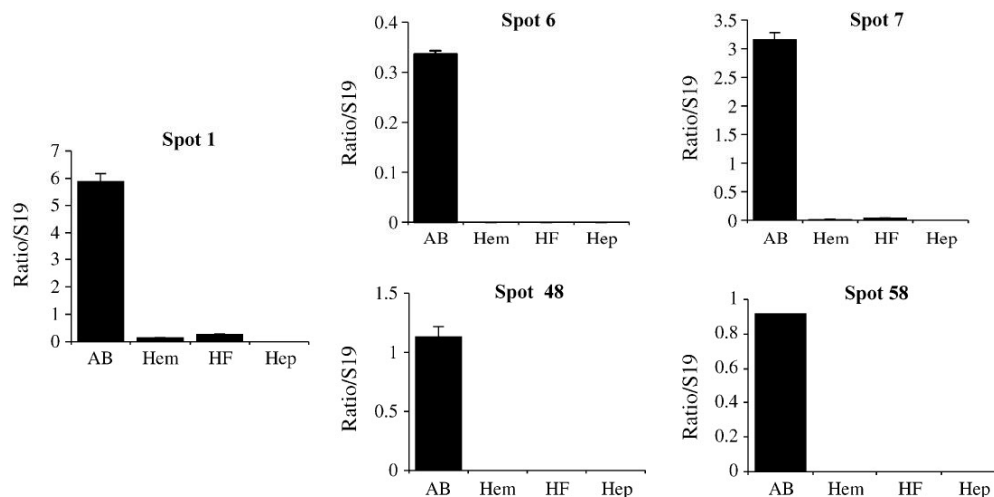


Fig. 7. Ratios of transcripts corresponding to spots 1, 6, 7, 48 and 58 in tissues of *Biomphalaria glabrata*: AB, albumen gland; Hem, hemocytes; HF, head foot; Hep, hepato-pancreas. Ratios were determined using real-time quantitative PCR and are shown relative to S19 levels (ratio/S19). Each histogram is the average value of triplicates  $\pm$  SD.

therefore modify the structure of parasite molecules (whether secreted or surface-exposed) and their interactions with host factors. This hypothesis may be supported by previous observations showing that *S. mansoni*-susceptible *B. glabrata* exhibit a higher  $\beta$ -D-mannosidase activity than resistant snails (Zelck, 1999).

Expression studies of these five candidate genes showed that the transcript content of all candidates are stable in susceptible snails following infection, while they increase significantly in exposed resistant snails. For the gene corresponding to protein 6, we note a significant increase of transcripts level especially 12 h p.i. whereas transcription is quite stable in susceptible snails.

As high transcript levels of these genes are observed in the albumen gland, all the differences observed between susceptible and resistant snails could be linked to a differential regulation of genes expressed by this organ. In molluscs, the albumen gland is considered as an accessory sexual gland involved in the synthesis of various compounds like galactogen necessary for egg development (Wijdenes et al., 1983). Nevertheless, this organ is also involved in the production of defense effectors like agglutinins in planorbid snails (Michelson and Dubois, 1977; Stein and Basch, 1979; Boswell and Bayne, 1984) or antimicrobial proteins in oposthobranchs gastropods (Iijima et al., 1994; Takamatsu et al., 1995). Our results further support the idea that the albumen gland may play a significant role in snail immune response.

In conclusion, this first proteomic approach has been successful in identifying secreted proteins that were differentially represented in the plasma from *E. caproni*-susceptible and—resistant *B. glabrata*. Molecular characterization of the corresponding genes provided informative data on the potential functions of these proteins. According to these predicted functions, various assumptions can be made on their potential involvement in susceptibility/resistance processes. Future functional studies are needed to clarify the role of these genes in the host-parasite interaction. Interestingly, our results also showed that these genes are mainly expressed in the albumen gland and raise the question of the potential importance of this organ in processes underlying susceptibility/resistance to *E. caproni*. In order to elucidate the role of the albumen gland in defense and/or resistance processes, future studies will focus on genes expressed by this organ in susceptible and resistant snails after infection.

## Acknowledgements

This work was supported by the CNRS. The Proteomics facility used in this study was funded by the European Community (FEDER) the Région Nord-Pas de Calais (France), the CNRS, the Génopole of Lille and the Université des Sciences et technologies de Lille.

## References

- Abrahamson, M., Alvarez-Fernandez, M., Nathanson, C.M., 2003. Cystatins. *Biochem. Soc. Symp.*, 179–199.
- Adema, C.M., Hertel, L.A., Miller, R.D., Loker, E.S., 1997. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc. Natl Acad. Sci. USA* 94, 8691–8696.
- Ataev, G.L., Coustau, C., 1999. Cellular response to *Echinostoma caproni* infection in *Biomphalaria glabrata* strains selected for susceptibility/resistance. *Dev. Comp. Immunol.* 23, 187–198.
- Boswell, C.A., Bayne, C.J., 1984. Isolation, characterization and functional assessment of a hemagglutinin from the plasma of *Biomphalaria glabrata*, intermediate host of *Schistosoma mansoni*. *Dev. Comp. Immunol.* 8, 559–568.
- Duclermortier, P., Lardans, V., Serra, E., Trottein, F., Dissous, C., 1999. *Biomphalaria glabrata* embryonic cells express a protein with a domain homologous to the lectin domain of mammalian selectins. *Parasitol. Res.* 85, 481–486.
- Gottsch, J.D., Eisinger, S.W., Liu, S.H., Scott, A.L., 1999. Calgranulin C has filariacidal and filariastatic activity. *Infect. Immun.* 67, 6631–6636.
- Guillou, F., Mitta, G., Dissous, C., Pierce, R., Coustau, C., 2004. Use of individual polymorphism to validate potential functional markers: case of a candidate lectin (BgSel) differentially expressed in susceptible and resistant strains of *Biomphalaria glabrata*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 138, 175–181.
- Harnett, W., Harnett, M.M., 2001. Modulation of the host immune system by phosphorylcholine-containing glycoproteins secreted by parasitic filarial nematodes. *Biochim. Biophys. Acta* 1539, 7–15.
- Hartmann, S., Lucius, R., 2003. Modulation of host immune responses by nematode cystatins. *Int. J. Parasitol.* 33, 1291–1302.
- Haslam, S.M., Morris, H.R., Dell, A., 2001. Mass spectrometric strategies: providing structural clues for helminth glycoproteins. *Trends Parasitol.* 17, 231–235.
- Humbert, E., Coustau, C., 2001. Refractoriness of host haemocytes to parasite immunosuppressive factors as a putative resistance mechanism in the *Biomphalaria glabrata*-*Echinostoma caproni* system. *Parasitology*. 122, 651–660.
- Iijima, R., Kisugi, J., Yamazaki, M., 1994. Biopolymers from marine invertebrates. XIV. Antifungal property of Dolabellin A, a putative self-defense molecule of the sea hare, *Dolabella auricularia*. *Biol. Pharm. Bull.* 17, 1144–1146.
- Jeyarasingam, U., Heyneman, D., Lim, H.K., Mansour, N., 1972. Life cycle of a new echinostome from Egypt, *E. liei* sp. nov. (Trematoda; Echinotomatidae). *Parasitology* 65, 203–222.
- Johansson, M.W., Soderhall, K., 1988. Isolation and purification of a cell adhesion factor from crayfish blood cells. *J. Cell Biol.* 106, 1795–1803.
- Khoo, K.H., Dell, A., 2001. Glycoconjugates from parasitic helminths: structure diversity and immunobiological implications. *Adv. Exp. Med. Biol.* 491, 185–205.
- Langand, J., Jourdan, J., Coustau, C., Delay, B., Morand, S., 1998. Cost of resistance, expressed as a delayed maturity, detected in the host-parasite system *Biomphalaria glabrata*/*Echinostoma caproni*. *Heredity* 80, 320–325.
- Leonard, P.M., Adema, C.M., Zhang, S.M., Loker, E.S., 2001. Structure of two FREP genes that combine IgSF and fibrinogen domains, with comments on diversity of the FREP gene family in the snail *Biomphalaria glabrata*. *Gene* 269, 155–165.
- Loker, E.S., Adema, C.M., 1995. Schistosomes, echinostomes and snails: comparative immunobiology. *Parasitol. Today* 11, 120–124.
- Marti, T., Erttmann, K.D., Gallin, M.Y., 1996. Host-parasite interaction in human onchocerciasis: identification and sequence analysis of a novel human calgranulin. *Biochem. Biophys. Res. Commun.* 221, 454–458.



- Michelson, E.H., Dubois, L., 1977. Agglutinins and lysins in the molluscan family Planorbidae: a survey of hemolymph, egg-masses, and albumen-gland extracts. *Biol. Bull.* 153, 219–227.
- Miller, A.N., Raghavan, N., FitzGerald, P.C., Lewis, F.A., Knight, M., 2001. Differential gene expression in haemocytes of the snail *Biomphalaria glabrata*: effects of *Schistosoma mansoni* infection. *Int. J. Parasitol.* 31, 687–696.
- Morrissey, J.H., 1981. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. *Anal. Biochem.* 117, 307–310.
- Neuhoff, V., Arold, N., Taube, D., Ehrhardt, W., 1988. Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* 9, 255–262.
- O'Farrell, P.H., 1975. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250, 4007–4021.
- Raghavan, N., Miller, A.N., Gardner, M., FitzGerald, P.C., Kerlavage, A.R., Johnston, D.A., Lewis, F.A., Knight, M., 2003. Comparative gene analysis of *Biomphalaria glabrata* hemocytes pre- and post-exposure to miracidia of *Schistosoma mansoni*. *Mol. Biochem. Parasitol.* 126, 181–191.
- Sminia, T., Barendsen, L., 1980. A comparative morphological and enzyme histochemical study on blood cells of the freshwater snails, *Lymnaea stagnalis*, *Biomphalaria glabrata*, and *Bulinus truncatus*. *J. Morphol.* 165, 31–39.
- Stein, P.C., Basch, P.F., 1979. Purification and binding properties of hemagglutinin from *Biomphalaria glabrata*. *J. Invertebr. Pathol.* 33, 10–18.
- Takamatsu, N., Shiba, T., Muramoto, K., Kamiya, H., 1995. Molecular cloning of the defense factor in the albumen gland of the sea hare *Aplysia kurodai*. *Fed. Eur. Biochem. Soc. Lett.* 377, 373–376.
- Trouvé, S., Renaud, F., Durand, P., Jourdan, J., 1996. Selfing and outcrossing in a parasitic hermaphrodite helminth (Trematoda, Echinostomatidae). *Heredity* 77, 1–8.
- Wijdenes, J., Van Elk, R., Joosse, J., 1983. Effects of two gonadotropic hormones on polysaccharide synthesis in the albumen gland of *Lymnaea stagnalis*, studied with the organ culture technique. *Gen. Comp. Endocrinol.* 51, 263–271.
- Wu, X., Ritter, B., Schlattjan, J.H., Lessmann, V., Heumann, R., Dietzel, I.D., 2000. Protein expression patterns of identified neurons and of sprouting cells from the leech central nervous system. *J. Neurobiol.* 44, 320–332.
- Xu, B., Hagglund, P., Stalbrand, H., Janson, J.C., 2002. Endo-beta-1,4-Mannanases from blue mussel, *Mytilus edulis*: purification, characterization, and mode of action. *J. Biotechnol.* 92, 267–277.
- Yamaura, I., Nozaki, Y., Matsumoto, T., Kato, T., 1996. Purification and some properties of an endo-1,4-beta-D-mannanase from a marine mollusc, *Littorina brevicula*. *Biosci. Biotechnol. Biochem.* 60, 674–676.
- Yamaura, I., Koga, T., Matsumoto, T., Kato, T., 1997. Purification and some properties of endo-1,4-beta-D-xylanase from a fresh-water mollusc, *Pomacea insularis* (de Ordigny). *Biosci. Biotechnol. Biochem.* 61, 615–620.
- Yang, T.H., Tzeng, S., Cheng, I., Burnett, M.G., Yoshizawa, Y., Fukuyama, K., Lee, S.C., Epstein, W.L., 1997. Identification of the mouse calcium-binding proteins, MRP 8 and MRP 14, in *Schistosoma mansoni*-induced granulomas: biochemical and functional characterization. *J. Leukoc. Biol.* 61, 258–266.
- Zelck, U.E., 1999. Glycosidase activities in plasma of naive and schistosome-infected *Biomphalaria glabrata* (Gastropoda). *Parasitology* 119 (Pt 6), 563–568.
- Zhang, S.M., Loker, E.S., 2003. The FREP gene family in the snail *Biomphalaria glabrata*: additional members, and evidence consistent with alternative splicing and FREP retrosequences. *Fibrinogen-related proteins. Dev. Comp. Immunol.* 27, 175–187.
- Zhang, S.M., Leonard, P.M., Adema, C.M., Loker, E.S., 2001. Parasite-responsive IgSF members in the snail *Biomphalaria glabrata*: characterization of novel genes with tandemly arranged IgSF domains and a fibrinogen domain. *Immunogenetics* 53, 684–694.
- Zhang, S.M., Adema, C.M., Kepler, T.B., Loker, E.S., 2004. Diversification of Ig superfamily genes in an invertebrate. *Science* 305, 251–254.

---

## Bases génétiques du polymorphisme de compatibilité dans l'interaction *Schistosoma mansoni* / *Biomphalaria glabrata*

---

La dynamique co-évolutive, présente dans les interactions durables hôte invertébré / parasite, conduit dans certains cas à un polymorphisme de compatibilité dont les bases moléculaires ne sont pas connues. Afin d'identifier les déterminants moléculaires de ce polymorphisme de compatibilité dans le modèle *Schistosoma mansoni* / *Biomphalaria glabrata*, nous avons développé une approche protéomique comparative des stades larvaires qui interagissent avec l'hôte invertébré. Ainsi, la comparaison, qualitative et quantitative, des protéomes des souches de parasites compatible et incompatible, nous a permis d'identifier une nouvelle famille d'antigènes chez *S. mansoni*. Les protéines de cette famille partagent les caractéristiques moléculaires des mucines. Elles possèdent un domaine contenant un nombre variable d'unités répétées en tandem (VNTR). Elles sont (i) exprimées spécifiquement dans les stades larvaires interagissant avec le mollusque, (ii) localisées dans la glande apicale des stades *miracidia* et sporocystes et (iii) sécrétées et relarguées avec les produits d'excrétion-sécrétion. Nous avons également montré que ces protéines de type mucine présentent un haut degré de polymorphisme et que des différences importantes existent entre les deux souches de *S. mansoni*. Ces différentes caractéristiques nous ont conduit à nommer les protéines de cette famille « *S. mansoni* Polymorphic Mucin », *SmPoMuc*. Nous avons ensuite montré que les *SmPoMuc* sont codées par une famille multigénique qui évolue selon le modèle de « birth and death ». Les gènes codant les *SmPoMuc* sont transcrits indépendamment, et pour chaque gène, différents transcrits peuvent être obtenus par épissage. De plus, nos résultats suggèrent que ce polymorphisme pourrait avoir des conséquences sur le statut de glycosylation des *SmPoMuc*. Nos résultats supportent l'idée que *S. mansoni* a développé au cours de l'évolution des mécanismes complexes permettant de générer un haut niveau de polymorphisme des *SmPoMuc*. Par ailleurs, le fait que ce « chaos contrôlé » soit généré à partir d'un nombre limité de gènes est unique par rapport à ce qui a été décrit pour des variants polymorphes exprimés par d'autres parasites.

**Mots clés :** *Schistosoma mansoni*, *Biomphalaria glabrata*, interaction hôte / parasite, compatibilité, polymorphisme, mucine, famille multigénique, « birth and death » et évolution concertée, « exon repetition », *SmPoMuc*.

---

## Genetics bases of host-parasite compatibility polymorphism in the *Schistosoma mansoni* / *Biomphalaria glabrata* model

---

The co-evolutionary dynamics that exist in certain invertebrate host / parasite interactions sometimes lead to a compatibility polymorphism, of which the molecular bases are unknown. To identify key molecules involved in this phenomenon in the *Schistosoma mansoni* / *Biomphalaria glabrata* model, we developed a comparative proteomics approach for the larval stages that interact with the invertebrate host. The qualitative and quantitative comparison of the proteomes of compatible and incompatible parasite strains led to the identification of a new family of schistosome antigens that share molecular characteristics with highly glycosylated secreted mucins. In particular, they possess a domain containing a variable number of tandem repeats (VNTR). These proteins are (i) only expressed in larval stages interacting with the mollusc, (ii) located in the apical gland of *miracidia* and sporocysts and (iii) secreted and released in excretion-secretion products. Moreover, we show that these mucin-like proteins display a high degree of polymorphism and that extensive differences are observed between *S. mansoni* strains. These different characteristics led us to name this novel molecular family “*S. mansoni* Polymorphic Mucins” *SmPoMuc*. Next, we show that *SmPoMuc* are coded by a multi-gene family that evolve following the model of “birth and death”. *SmPoMuc* genes are transcribed in an individual-specific manner, and for each gene, multiple splice variants exist. The incidence of this polymorphism on *SmPoMuc* glycosylation status was demonstrated. Our data support the view that *S. mansoni* has evolved a complex cascade of mechanisms – a “controlled chaos” – based on a relatively low number of genes that provides highly polymorphism to *SmPoMuc*.

**Keywords:** *Schistosoma mansoni*, *Biomphalaria glabrata*, host / parasite interaction, compatibility, polymorphism, mucin-like multi-gene family, “birth and death” and concerted evolution, exon repetition, *SmPoMuc*.

---

**DISCIPLINE :** Ecologie – Génétique – Biologie moléculaire – Biochimie      **SPECIALITE :** Parasitologie

---

### LABORATOIRE

Biologie et Ecologie Tropicale et Méditerranéenne  
UMR 5244 CNRS-EPHE-UPVD, Equipe Parasitologie Fonctionnelle et Evolutive  
Université Via Domitia, 52 avenue Paul Alduy  
66860 Perpignan Cedex