



**HAL**  
open science

# Ingénierie de lectines d'invertébrés pour le développement de nouveaux outils de diagnostic en cancérologie.

S. Mathieu

► **To cite this version:**

S. Mathieu. Ingénierie de lectines d'invertébrés pour le développement de nouveaux outils de diagnostic en cancérologie.. Biochimie [q-bio.BM]. Université de Grenoble, 2011. Français. NNT: . tel-00568846v1

**HAL Id: tel-00568846**

**<https://theses.hal.science/tel-00568846v1>**

Submitted on 23 Feb 2011 (v1), last revised 7 Apr 2011 (v2)

**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.

## THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : Chimie et Sciences du Vivant

Arrêté ministériel : 7 août 2006

Présentée par

**Sophie MATHIEU**

Thèse dirigée par **Anne IMBERTY** et  
codirigée par **Annabelle VARROT**

préparée au sein du **Laboratoire CERMAV (Centre de  
Recherche sur les Macromolécules Végétales)**  
dans l'**École Doctorale Chimie et Sciences du Vivant**

# Ingénierie de lectines d'invertébrés pour le développement de nouveaux outils de diagnostic en cancérologie

Thèse soutenue publiquement le **26 janvier 2011**  
devant le jury composé de :

**Mme Julie BOUCKAERT**

CR, Belgique, Rapporteur

**M. Jean-Luc COLL**

DR, France, Président

**Mme Miriam DWEK**

Pr, Royaume-Uni, Rapporteur

**Mme Tania NAKAMURA**

Pr, Brésil, Membre

**Mme Anne IMBERTY**

DR, France, Directrice de thèse

**Mme Annabelle VARROT**

CR, France, Directrice de thèse





A Manon.



**Bientôt cinq ans passés au CERMAV et un nombre important de personnes à remercier...**

*Je remercie tout d'abord le Dr. Julie Bouckaert et le Dr. Miriam Dwek pour avoir accepté de juger ce travail de thèse avec autant d'assiduité. Je remercie également le Dr. Tania Nakamura pour avoir participé à ce jury et le Dr. Jean-Luc Coll pour l'avoir présidé.*

*Je remercie bien évidemment mes deux directrices de thèse, les Dr. Anne Imberty et Dr. Annabelle Varrot, pour l'opportunité qu'elles m'ont offert de travailler sur un sujet aussi passionnant. Elles ont toujours su écouter et prendre en considération mes attentes. Elles ont également fait preuve de patience et compréhension lors de mes activités extra-thèse (monitorat, Glucidoc..., grossesse) et je les remercie chaleureusement pour cela.*

*Je remercie également le Pr. Christelle Breton de m'avoir acceptée en stage de master 1 en 2006. J'ai ainsi pu découvrir concrètement le monde passionnant des glycosciences. Je la remercie également pour l'encadrement qu'elle a su m'apporter lors de mes activités d'enseignement.*

*Je remercie également les personnes avec qui nous avons collaboré. Tout d'abord, à l'Université de Westminster, le Dr. Miriam Dwek pour m'avoir accueillie dans son laboratoire, me permettant ainsi de compléter ma formation par de nouvelles techniques que nous ne disposions pas au laboratoire. Sa patience et sa gentillesse m'ont grandement aidé à supporter le dépaysement. Mais également, le Dr. Bérangère Tissot de l'Imperial College avec qui j'ai découvert la spectrométrie de masse. J'ai également une pensée particulière pour les étudiants français de ces laboratoires, Julien Saint-Guiron et Kévin Canis, qui ont facilité mon intégration dans cette jungle londonienne.*

*Je tiens également à remercier toutes les personnes de l'équipe Glycobiologie Moléculaire qui m'ont aidée durant cette thèse :*

*Catherine Gautier et Valérie Chazalet, qui m'ont chacune enseignée, avec les pédagogies propres, les rudiments de la biochimie et de la biologie moléculaire. Sans elles, tout aurait été tellement plus compliqué. J'ai une pensée particulière pour Valérie avec qui une amitié est née. Je garde en mémoire nos discussions extra-scientifiques sur la vie, les enfants et je quitte maintenant le laboratoire avec la nostalgie de sa présence.*

*Je remercie aussi chaleureusement le Dr. Olivier Lerouxel qui, en plus d'avoir été un collègue plus que sympathique, a toujours pris le temps de m'aider tant pour les manip que pour l'enseignement. Merci aussi au Dr. Aline Thomas pour son aide en modélisation mais aussi pour sa patience et son enthousiasme dans le rush de la fin. On a le sentiment de devenir meilleur à ses côtés.*

*C'est également avec beaucoup d'affections que je tiens à remercier la dream-team des étudiantes de glycobio, Magali Audry, Emilie Lameignère, Sara Fasmer Hansen, qui sont toutes devenues docteur durant mon séjour au Cermav. Mais aussi, le Dr. Claire Crola Da Silva qui était une post-doc très attentive au début de ma thèse et qui est maintenant devenue une amie. Le labo était différent lorsqu'elles étaient là.*

*Je n'oublie évidemment pas Bertrand Blanchard et mes collègues actuels, les Dr. Aymeric Audfray et Nicolas Sapay, Géraldine Ganne, Anita Sarkar et Sumaira Kousar, pour leur bonne humeur et leur patience durant ces longs mois de rédaction...*

*Je garde également une pensée pour tous mes collègues du comité d'organisation du Glucidoc 2009, Samy Cecioni, Fabien Accadbled et Nathalie Favre.*

*Je remercie également le personnel administratif, Isabelle Caldara, Martine Morales, Magali Gardes et Martine Broué qui ont toujours su me simplifier la vie et toutes les personnes du CERMAV que j'ai eu l'occasion de croiser durant ces années. Je quitte maintenant le laboratoire avec beaucoup de nostalgie en vous souhaitant à tous une bonne continuation.*

*Je remercie évidemment ma famille qui a toujours été un soutien sans faille. Merci aussi à ma nounou pour sa flexibilité lors de ces dernières semaines. Ma dernière pensée s'adresse à ma fille Manon qui m'a permis de m'évader, le soir, après une journée de rédaction. La voir grandir apprend à se concentrer sur les choses essentielles...*

# **Sommaire**



## Listes des principales abréviations

## Liste des figures

## Liste des tableaux

### CHAPITRE 1 : Introduction

<b>1. Généralités sur la glycosylation</b> .....	<b>- 1 -</b>
1.1. La <i>N</i> -glycosylation .....	- 2 -
1.1.1. Structure et diversité des <i>N</i> -glycanes chez les mammifères.....	- 2 -
1.1.2. Biosynthèse des <i>N</i> -glycanes .....	- 3 -
1.2. La <i>O</i> -glycosylation.....	- 4 -
1.2.1. Structure et biosynthèse des <i>O</i> -glycanes de type mucine.....	- 5 -
1.2.2. Les mucines .....	- 6 -
1.2.3. Autres types de <i>O</i> -glycosylation.....	- 7 -
1.3. La <i>C</i> -mannosylation .....	- 10 -
1.4. Autres types de glycoconjugués .....	- 10 -
1.4.1. Les glycolipides .....	- 10 -
1.4.2. Les protéoglycanes.....	- 11 -
1.4.3. Les ancras GPI.....	- 12 -
1.5. Les glycosyltransférases.....	- 13 -
1.5.1. Définition et nomenclature.....	- 13 -
1.5.2. Structures tridimensionnelles de types <i>GT-A</i> et <i>GT-B</i> .....	- 14 -
1.5.3. Stéréochimie de la réaction.....	- 15 -
1.6. Modifications pathologiques de la glycosylation.....	- 16 -
<b>2. Modification de la glycosylation dans le processus cancéreux</b> .....	<b>- 17 -</b>
2.1. Les antigènes T et Tn .....	- 17 -
2.1.1. Biosynthèse dans les différents types de cancers .....	- 17 -
2.1.2. Les antigènes T et Tn sialylés.....	- 19 -
2.1.3. Immunité et antigénicité de ces épitopes.....	- 20 -
2.2. Les antigènes des groupes sanguins .....	- 21 -
2.2.1. Définition des groupes sanguins <i>ABO</i> et <i>Lewis</i> .....	- 21 -
2.2.2. Biosynthèse des antigènes des groupes sanguins.....	- 23 -
2.2.3. Les antigènes du système <i>ABO</i> dans le processus cancéreux .....	- 24 -
2.2.4. Les antigènes du système <i>Lewis</i> dans le processus cancéreux.....	- 26 -
2.2.5. Leurs homologues sialylés dans le processus cancéreux.....	- 26 -

2.3.	Modification des N-glycanes .....	- 28 -
2.3.1.	<i>La structure <math>\beta</math>1-6GlcNAc associée aux métastases</i> .....	- 28 -
2.3.2.	<i>La fucosylation des N-glycanes</i> .....	- 30 -
2.4.	Modifications des autres types de glycoconjugués .....	- 31 -
2.4.1.	<i>Glycolipides et cancer</i> .....	- 31 -
2.4.2.	<i>Protéoglycanes et cancer</i> .....	- 32 -
2.4.3.	<i>Ancres GPI et cancer</i> .....	- 33 -
<b>3.</b>	<b>Lectines et cancer</b> .....	<b>- 34 -</b>
3.1.	Définition et historique des lectines .....	- 34 -
3.1.1.	<i>Définition</i> .....	- 34 -
3.1.2.	<i>Historique</i> .....	- 35 -
3.1.3.	<i>Affinité et spécificité des lectines</i> .....	- 36 -
3.1.4.	<i>Applications des lectines en cancérologie</i> .....	- 38 -
3.2.	Utilisation de la lectine HPA comme outil de diagnostic dans le cancer.....	- 40 -
3.2.1.	<i>HPA, marqueur de cancers à fort pouvoir métastatique</i> .....	- 40 -
3.2.2.	<i>Importance du choix de la méthode de détection</i> .....	- 41 -
3.2.3.	<i>Identification des glycotopes reconnus par HPA</i> .....	- 43 -
<b>4.</b>	<b>La famille des lectines de type H</b> .....	<b>- 45 -</b>
4.1.	HPA.....	- 45 -
4.1.1.	<i>Caractérisation biochimique</i> .....	- 45 -
4.1.2.	<i>Caractérisation structurale</i> .....	- 47 -
4.1.3.	<i>Occurrence des protéines similaires à HPA</i> .....	- 51 -
4.2.	Les discoidines .....	- 52 -
4.2.1.	<i>Généralités</i> .....	- 52 -
4.2.2.	<i>Spécificité et affinité</i> .....	- 53 -
4.2.3.	<i>Caractérisation structurale</i> .....	- 54 -
4.2.4.	<i>Similarité avec HPA</i> .....	- 56 -
4.3.	SLL-2 : lectine du corail <i>Sinularia lochmodes</i> .....	- 57 -
<b>5.</b>	<b>Objectifs du travail de thèse</b> .....	<b>- 58 -</b>

## CHAPITRE 2 : Matériels et Méthodes

<b>1.</b>	<b>Utilisation des lectines pour le marquage de cellules cancéreuses</b> .....	<b>- 61 -</b>
1.1.	Principe de la microscopie confocale.....	- 61 -
1.2.	Traitement des échantillons.....	- 62 -

1.2.1.	<i>Biotinylation de DiscI et DiscII</i> .....	- 62 -
1.2.2.	<i>Protocole de marquage des cellules cancéreuses</i> .....	- 63 -
1.2.3.	<i>Révélation par microscopie confocale</i> .....	- 64 -
<b>2.</b>	<b>Techniques de biologie moléculaire</b> .....	<b>- 64 -</b>
2.1.	Souche et plasmides utilisés.....	- 64 -
2.2.	Stratégies de clonage.....	- 69 -
2.2.1.	<i>Clonage des domaines séparés des DiscI et DiscII</i> .....	- 69 -
2.2.2.	<i>Mutagenèse dirigée de DiscII</i> .....	- 71 -
2.2.3.	<i>Clonage de la lectine SLL2 extraite du corail</i> .....	- 74 -
<b>3.</b>	<b>Expression de protéines sous forme recombinante</b> .....	<b>- 77 -</b>
3.1.	Expression dans la bactérie <i>Escherichia coli</i> .....	- 77 -
3.2.	Expression dans la levure <i>Pichia pastoris</i> .....	- 80 -
<b>4.</b>	<b>Techniques de biochimie</b> .....	<b>- 80 -</b>
4.1.	Purification de protéines.....	- 80 -
4.1.1.	<i>Préparation des extraits bruts</i> .....	- 81 -
4.1.2.	<i>Purification des lectines de Dictyostelium discoideum</i> .....	- 81 -
4.1.3.	<i>Purification des lectines de Sinularia lochmodes SLL-2a et SLL-2a-2</i> .....	- 84 -
4.2.	Techniques de renaturation de protéines.....	- 84 -
4.2.1.	<i>Utilisation d'Augmedium™</i> .....	- 84 -
4.2.2.	<i>Lavage et solubilisation des corps d'inclusion</i> .....	- 84 -
4.2.3.	<i>Renaturation des protéines dénaturées</i> .....	- 86 -
4.3.	Caractérisation biochimique des protéines.....	- 88 -
4.3.1.	<i>Electrophorèse sur gel de polyacrylamide en présence de SDS</i> .....	- 88 -
4.3.2.	<i>Transfert de protéines et immunodétection (Western Blot)</i> .....	- 88 -
4.3.3.	<i>Test d'hémagglutination</i> .....	- 89 -
4.3.4.	<i>DLS : Diffusion dynamique de la lumière</i> .....	- 90 -
4.3.5.	<i>La résonance plasmonique de surface</i> .....	- 90 -
<b>5.</b>	<b>La cristallographie aux rayons X</b> .....	<b>- 93 -</b>
5.1.	La cristallogénèse.....	- 93 -
5.2.	Résolution de structures par cristallographie aux rayons X.....	- 95 -
5.2.1.	<i>La diffraction des rayons X par les cristaux</i> .....	- 95 -
5.2.2.	<i>Facteurs de structure et densité électronique</i> .....	- 97 -
5.2.3.	<i>Le problème des phases</i> .....	- 97 -
5.2.4.	<i>Affinement</i> .....	- 98 -

<b>6. Analyse des glycoépitopes reconnus par HPA.....</b>	<b>- 99 -</b>
6.1. Principe de la spectrométrie de masse.....	- 99 -
6.2. Traitement des échantillons.....	- 101 -
6.2.1. Isolation des glycoépitopes à partir des lignées cellulaires .....	- 101 -
6.2.2. Réduction et carboxyméthylation.....	- 103 -
6.2.3. Digestion peptidique à la trypsine .....	- 104 -
6.2.4. Récupération des N-glycanes par digestion à la PNGase F.....	- 104 -
6.2.5. Récupération des O-glycanes par $\beta$ -élimination réductrice.....	- 104 -
6.2.6. Perméthylation.....	- 105 -
6.2.7. Analyses par spectrométrie de masse MALDI.....	- 105 -

### **CHAPITRE 3 : Utilisation des lectines HPA, DiscI et DiscII pour le marquage de cellules cancéreuses**

<b>1. Mise au point expérimentale avec HPA .....</b>	<b>- 107 -</b>
<b>2. Marquages cellulaires avec DiscI et DiscII .....</b>	<b>- 109 -</b>
2.1. Tests avec les discoidines sur la lignée SW 620 .....	- 109 -
2.2. Marquages des trois lignées cellulaires.....	- 109 -
2.3. Compétition entre DiscI et HPA .....	- 111 -
2.4. Inhibition par des monosaccharides .....	- 112 -
<b>3. Conclusion.....</b>	<b>- 113 -</b>

### **CHAPITRE 4 : Structures et spécificité de DiscI Comparaison avec DiscII et HPA**

<b>1. Production de DiscI sous forme recombinante.....</b>	<b>- 115 -</b>
<b>2. Spécificité des lectines de <i>Dictyostelium discoideum</i>.....</b>	<b>- 118 -</b>
<b>3. Structures cristallographiques de DiscI .....</b>	<b>- 119 -</b>
<b>4. Conclusion.....</b>	<b>- 127 -</b>

### **CHAPITRE 5 : Ingénierie de DiscII**

<b>1. Mutants 1 .....</b>	<b>- 132 -</b>
1.1. Construction des protéines par mutagenèse dirigée .....	- 132 -
1.2. Production sous forme recombinante dans <i>E. coli</i> .....	- 133 -
1.3. Purification par chromatographie sur colonne de nickel.....	- 135 -
1.4. Contrôle de la fonctionnalité des mutants par chromatographie d'affinité.....	- 137 -
1.5. Contrôle de l'organisation oligomérique des protéines mutantes .....	- 138 -

1.6. Stratégie pour la construction de nouveaux mutants.....	- 139 -
<b>2. Mutant 2</b> .....	<b>- 140 -</b>
2.1. Clonage, expression et purification par chromatographie sur colonne de nickel.	- 140
-	
2.2. Contrôle de la fonctionnalité de DiscII-boucleHPA-mutée .....	- 142 -
2.3. Contrôle de l'oligomérisation de la protéine.....	- 143 -
2.4. Résolution de la structure de DiscII-boucleHPAmutée par cristallographie aux rayons X .....	- 143 -
<b>3. Mutant 3 : DiscII-boucleHPAmuté-G247H et mutant 4 : DiscII-boucleHPAmuté-G247H-I171S-F243S</b> .....	<b>- 147 -</b>
<b>4. Mutants 5 et 6 : Construction des protéines hybrides DiscIINter-HPA et DiscIINter-HPA-R163S</b> .....	<b>- 150 -</b>
<b>5. Conclusion</b> .....	<b>- 152 -</b>

## CHAPITRE 6 : Etude fonctionnelle des domaines N-terminaux de DiscI et DiscII

<b>1. Le domaine N-terminal de DiscI</b> .....	<b>- 157 -</b>
<b>2. Le domaine N-terminal de DiscII</b> .....	<b>- 160 -</b>
2.1. Production et purification de DiscII-Nter.....	- 160 -
2.2. Contrôle de la fonctionnalité de DiscII-Nter.....	- 162 -
<b>3. Conclusion</b> .....	<b>- 167 -</b>

## CHAPITRE 7 : Expression de la lectine SLL-2 du corail *Sinularia lochmodes*

<b>1. Clonage des gènes <i>sll-2</i></b> .....	<b>- 171 -</b>
<b>2. Expression dans la bactérie <i>E. coli</i></b> .....	<b>- 173 -</b>
<b>3. Renaturation des corps d'inclusion</b> .....	<b>- 180 -</b>
3.1. Solubilisation des corps d'inclusion.....	- 180 -
3.2. Renaturation des corps d'inclusion solubilisés .....	- 183 -
<b>4. Expression dans la levure <i>Pichia pastoris</i></b> .....	<b>- 187 -</b>
4.1. Clonage dans les vecteurs pPIC-Zα .....	- 187 -
4.2. Conditions d'expression.....	- 191 -
4.3. Discussion .....	- 194 -
<b>5. Conclusion</b> .....	<b>- 195 -</b>

<b>Conclusion générale et perspectives</b> .....	<b>- 198 -</b>
--	----------------

<b>Bibliographie.....</b>	<b>- 201 -</b>
---------------------------	----------------

## **ANNEXE**

<b>Analyse, par spectrométrie de masse, des glycoépitopes reconnus par HPA .....</b>	<b>- 218 -</b>
<b>1. Généralités .....</b>	<b>- 219 -</b>
<b>2. Analyse du milieu de culture .....</b>	<b>- 220 -</b>
<b>3. Analyse des <i>N</i>-glycanes des membranes cellulaires .....</b>	<b>- 222 -</b>
3.1. Lignée cellulaire non métastatique SW 480.....	- 223 -
3.2. La lignée métastatique SW 620.....	- 226 -
3.3. La lignée très métastatique HT 29.....	- 226 -
<b>4. Conclusions .....</b>	<b>- 227 -</b>
<b>5. Perspectives.....</b>	<b>- 228 -</b>
<b>Article .....</b>	<b>- 230 -</b>

# Liste des principales abréviations

---

<b>ADNc</b>	ADN complémentaire
<b>AFP</b>	Alpha-fetoprotéine
<b>AMPc</b>	Adénosine monophosphate cyclique
<b>ARN</b>	Acide ribonucléique
<b>ARNm</b>	ARN messager
<b>BSA</b>	Sérum albumine bovine
<b>CAZy</b>	Carbohydate Active Enzyme
<b>CDG</b>	Congenital disorder of glycosylation
<b>CEA</b>	Carcinoembryogenic antigen
<b>CHAPS</b>	3-[(3-Cholamidopropyl)diméthylammonio]-1-propanesulfonate
<b>CMP</b>	Cytosine monophosphate
<b>CRDs</b>	Carbohydrate recognition domains
<b>CS</b>	Chondroïtine sulfate
<b>Cy5</b>	Cyanine 5
<b>DAB</b>	Diaminobenzidine
<b>DBA</b>	<i>Dolichos biflorus</i> agglutinin
<b>DHB</b>	Acide 2,5-dihydrobenzoïque
<b>DMF</b>	Diméthylformamide
<b>DMSO</b>	Diméthylsulfoxyde
<b>DO</b>	Densité optique
<b>DS</b>	Dermatane sulfate
<b>DTT</b>	Dithiothréitol
<b>EDTA</b>	Acide éthylènediaminetétraacétique
<b>EGF</b>	Epidermal Growth Factor ou Facteur de croissance épidermique
<b>FITC</b>	Fluorescein-5-isothiocyanate
<b>FNR</b>	Fraction Non Retenue
<b>Fuc</b>	Fucose
<b>FUT</b>	Fucosyltransférase
<b>GAG</b>	Glycosaminoglycane
<b>Gal</b>	Galactose
<b>GalNAc</b>	<i>N</i> -acétylgalactosamine
<b>GdnCl</b>	Hydrochlorure de guanidinium
<b>GDP</b>	Guanosine diphosphate
<b>GlcNAc</b>	<i>N</i> -acétylglucosamine
<b>GnT</b>	<i>N</i> -acétylglucosaminyltransférase
<b>GPI</b>	Glycosyl phosphatidylinositol
<b>GSH</b>	Glutathion réduit
<b>GSSG</b>	Glutathion oxydé
<b>GST</b>	Glutathion S-Transferase
<b>HAA</b>	<i>Helix aspersa</i> agglutinin
<b>HPA</b>	<i>Helix pomatia</i> agglutinin
<b>HS</b>	Héparane sulfate
<b>IPTG</b>	Isopropyl- $\beta$ -D-thiogalactopyranoside
<b>IMAC</b>	Immobilized Metal ion Affinity Chromatography
<b>kDa</b>	kilodalton

<b>KS</b>	Kératane sulfate
<b>LB</b>	Luria Bertani
<b>LCA</b>	<i>Lens culinaris</i> agglutinin
<b>Le</b>	Lewis
<b>LFA</b>	<i>Limax flavus</i> agglutinin
<b>MAA</b>	<i>Maackia amurensis</i> agglutinin
<b>MALDI</b>	Matrix-assisted laser desorption/ionization
<b>Man</b>	Mannose
<b>ME</b>	Mercaptoéthanol
<b>MEC</b>	Milieu extracellulaire
<b>MGYC</b>	Milieu Glycérol Yeast Culture
<b>MES</b>	Acide 2-[N-Morpholino]éthanesulfonique]
<b>MMP</b>	Matrix metalloproteinase
<b>MMYC</b>	Milieu Méthanol Yeast Culture
<b>MPA</b>	<i>Maclura pomifera</i> agglutinin
<b>MPD</b>	2-methyl-2,4-pentandiol
<b>NeuAc</b>	Acide <i>N</i> -acétylneuraminique
<b>OGT</b>	O- <i>N</i> -acétylglucosaminyltransférase
<b>PAA</b>	Polyacrylamide
<b>PBS</b>	Phosphate Buffered Saline
<b>PCR</b>	Polymerase chaine reaction
<b>PDB</b>	Protein Data Bank
<b>PEG</b>	Polyéthylène glycol
<b>PHA-L</b>	<i>Phaseolus vulgaris</i> leukoagglutinating lectin
<b>PI</b>	Phosphatidyl-inositol
<b>pI</b>	Point isoélectrique
<b>PMT</b>	Dolichyl phosphate D-mannose : protein <i>O</i> -mannosyltransférase
<b>ppGalNAc-T</b>	UDP- <i>N</i> - $\alpha$ -galactosamine : polypeptide <i>N</i> -acétylgalactosaminyltransférase
<b>PSA</b>	Prostate-specific antigen
<b>RACE</b>	Rapid Amplification of cDNA Ends
<b>RE</b>	Réticulum endoplasmique
<b>RMN</b>	Résonance magnétique nucléaire
<b>RPS</b>	Résonance plasmonique de surface
<b>RT-PCR</b>	Reverse Transcriptase PCR
<b>SBA</b>	Soybean agglutinin
<b>Scid</b>	Severe combined immunodeficiency
<b>SDS</b>	Sodium dodecyl sulfate
<b>SDS-PAGE</b>	SDS polyacrylamide gel electrophoresis
<b>TBS</b>	Tris Buffered Saline
<b>TCA</b>	Acide trichloroacétique
<b>TEV</b>	Tobacco Ech Virus
<b>TOF</b>	Time-Of-Flight
<b>Tris</b>	2-amino-[hydroxyméthyl]-1,3-propanediol
<b>TRITC</b>	Tetramethyl Rhodamine Isothiocyanate
<b>UDP</b>	Uridine diphosphate
<b>UEA</b>	<i>Ulex europaeus</i> agglutinin
<b>VVA</b>	<i>Vicia villosa</i> agglutinin
<b>WGA</b>	Wheat germ agglutinin
<b>YNB</b>	Yeast Nitrogen Base

# Liste des figures

---

- Figure 1.1. Schéma d'une membrane cellulaire présentant différentes structures glycaniques (glycoprotéines et glycolipides)..... - 1 -
- Figure 1.2. Exemples des trois types de *N*-glycanes chez l'homme : type oligomannosique, type complexe et type hybride..... - 3 -
- Figure 1.3. Voie de synthèse des *N*-glycanes dans le réticulum endoplasmique et l'appareil de Golgi (Helenius and Aebi 2001). ..... - 4 -
- Figure 1.4. Structures des noyaux identifiés à ce jour dans les chaînes *O*-glycaniques de type mucine..... - 5 -
- Figure 1.5. Schéma de la structure d'un *O*-glycane de type mucine. .... - 6 -
- Figure 1.6. Représentation schématique de la glycosylation de type *O*-GlcNAc (Golks and Guerini 2008). ..... - 7 -
- Figure 1.7. Schéma des voies de biosynthèse des glycosphingolipides chez les mammifères (Wennekes *et al.* 2009). ..... - 11 -
- Figure 1.8. Biosynthèse schématisée des glycosaminoglycanes (Hacker *et al.* 2005).. - 12 -
- Figure 1.9. Structure d'une ancre GPI (Paulick and Bertozzi 2008). ..... - 13 -
- Figure 1.10. Représentation du repliement de type GT-A observé pour l'enzyme SpsA de *Bacillus subtilis* (PDB 1qgq) (A.) et de leur motif DxD en interaction avec le nucléotide-sucré via l'ion  $Mn^{2+}$  (B.)..... - 14 -
- Figure 1.11. Représentation du repliement de type GT-B observé chez la  $\beta$ -glucosyltransférase du bactériophage T4 (PDB 1jg7). ..... - 15 -
- Figure 1.12. Synthèse des liaisons glycosidiques par inversion ou rétention de configuration du carbone anomérique (Lairson *et al.* 2008). ..... - 16 -
- Figure 1.13. Représentation schématique de la biosynthèse des antigènes T et Tn et leurs homologues sialylés..... - 18 -
- Figure 1.14. Représentation schématique des épitopes des groupes sanguins des systèmes ABO et Lewis. .... - 22 -
- Figure 1.15. Voies de biosynthèse des antigènes des groupes sanguins ABH et Lewis basées sur A. le précurseur de type 1 et B. les précurseurs de type 3/4. .... - 24 -

Figure 1.16.	Schéma de biosynthèse des antigènes Lewis a et x sialylés dans les cancers du sein.....	- 27 -
Figure 1.17.	<i>N</i> -acétylglucosaminyltransférases impliquées dans la biosynthèse des <i>N</i> -glycanes de type complexe.....	- 29 -
Figure 1.18.	Schéma résumant les rôles antagonistes des enzymes GnT-III et GnT-V..	- 30 -
Figure 1.19.	Rôle des glycoconjugués situés à la surface cellulaire.....	- 35 -
Figure 1.20.	Marquages positif (A) et négatif (B) par HPA biotinylé d'adénocarcinomes du poumon. Grossissement x200 (Laack <i>et al.</i> 2002). .....	- 41 -
Figure 1.21.	Différentes méthodes de marquage des épitopes cancéreux par HPA. ....	- 42 -
Figure 1.22.	Analyse, par glycan array, de la spécificité d'HPA (Sanchez <i>et al.</i> 2006). ..	- 46 -
Figure 1.23.	Représentation en dessin de la structure cristallographique d'HPA. ....	- 48 -
Figure 1.24.	Structure cristallographique du complexe entre HPA et le GalNAc.....	- 49 -
Figure 1.25.	Site de reconnaissance d'HPA complexée avec les antigènes cancéreux... ..	- 50 -
Figure 1.26.	Alignement des séquences peptidiques d'HPA, de CH-III, lectine de l'escargot <i>C. hortensis</i> , des trois isogènes de SLL-2, lectine du corail <i>S. lochmodes</i> et des discoidines de l'amibe <i>D. discoideum</i> . .....	- 51 -
Figure 1.27.	Développement du cycle de vie de l'amibe <i>Dictyostelium discoideum</i> . ....	- 52 -
Figure 1.28.	Structure tridimensionnelle du trimère de DiscII.....	- 55 -
Figure 1.29.	Site de reconnaissance de DiscII complexée avec les résidus GalNAc (A) et $\beta$ MeGal (B).....	- 56 -
Figure 1.30.	A. Superposition du domaine C-terminal de DiscII (vert, rouge, bleu) avec HPA (jaune, magenta, cyan). B. Alignement de séquence basé sur la superposition structurale.....	- 57 -
Figure 2.1.	Principe du microscope confocale.....	- 62 -
Figure 2.2.	Plasmides utilisés pour l'expression de protéines dans la bactérie <i>E. coli</i> ..	- 66 -
Figure 2.3.	Plasmides utilisés pour l'expression de protéines recombinantes dans la levure <i>Pichia pastoris</i> . .....	- 67 -
Figure 2.4.	Principe de la mutagenèse dirigée GeneTailor™ Site-Directed Mutagenesis System, développée par Invitrogen.....	- 71 -
Figure 2.5.	Principe de la PCR chevauchante appliquée à la formation de la protéine hybride DiscII <sup>nter</sup> -HPA.....	- 73 -
Figure 2.6.	Principe de la 3'RACE (Invitrogen).....	- 74 -
Figure 2.7.	Le test d'hémagglutination sur microplaque de titration .....	- 90 -

Figure 2.8.	Principe du système Biacore®. ....	- 92 -
Figure 2.9.	Diagramme de phase d'une protéine. ....	- 94 -
Figure 2.10.	Système de cristallisation par « diffusion en phase gazeuse par goutte pendante ».....	- 94 -
Figure 2.11.	A. Maille élémentaire et paramètres à trois dimensions. B. Les 14 réseaux de Bravais. ....	- 96 -
Figure 2.12.	Faisceau de rayons X diffracté par un cristal donnant naissance au cliché de diffraction .....	- 97 -
Figure 2.13.	A. Structure d'un spectromètre de masse. B. Trajectoire d'un ion dans un analyseur à temps de vol.....	- 101 -
Figure 2.14.	Schéma détaillant les principales étapes d'une analyse glycomique par spectrométrie de masse. ....	- 103 -
Figure 3.1.	Marquage des lignées cellulaires de cancer colorectal SW 620 et HT 29 par la lectine HPA.....	- 108 -
Figure 3.2.	Marquage de la lignée cellulaire de cancer colorectal SW 620 par A. DiscI et B. DiscII.....	- 109 -
Figure 3.3.	Marquage des lignées cellulaires de cancers colorectaux non métastatiques SW 480 et métastatiques SW 620 et HT 29 par DiscI et HPA et révélation par microscopie confocale. ....	- 110 -
Figure 3.4.	Marquage des lignées cellulaires de cancers colorectaux métastatiques SW 620 et HT 29 par DiscI et HPA et révélation par microscopie confocale. ....	- 112 -
Figure 4.1.	Profils d'élution obtenus lors des purifications de DiscI par chromatographie sur colonne de nickel. ....	- 117 -
Figure 4.2.	Gel SDS-PAGE 12% acrylamide coloré au bleu de Coomassie après expression de DiscI dans Rosetta 2 pLysS et purification sur colonne de nickel.....	- 117 -
Figure 4.3.	Principe du Glycan Arrays. ....	- 118 -
Figure 4.4.	Superposition des résultats de puces à sucre obtenus avec DiscI (rouge) et DiscII (vert) à la même concentration. ....	- 119 -
Figure 4.5.	Structure cristallographique du trimère de DiscI. ....	- 121 -
Figure 4.6.	Influence de l'EDTA sur DiscI contrôlée par hémagglutination. ....	- 122 -

Figure 4.7.	A. Superposition du site de reconnaissance du GalNAc dans DiscI (cyan/marine), DiscII (vert/jaune) et HPA (saumon/magenta), B. Représentation du site de reconnaissance de DiscI pour le GalNAc.....	- 123 -
Figure 4.8.	Liaisons des disaccharides dans le domaine C-terminal de DiscI.....	- 125 -
Figure 4.9.	Représentation de surface des sites de reconnaissance de DiscI (A.), DiscII (B.) et HPA (C.).....	- 128 -
Figure 4.10.	Représentation des structures cristallographiques des sites de reconnaissance des lectines de type H. ....	- 129 -
Figure 5.1.	Superposition des structures cristallographiques d'HPA (cyan, magenta et jaune) et de DiscII (bleu, rouge et vert), complexées avec du GalNAc. ...	- 132 -
Figure 5.2.	Alignement des séquences peptidiques du domaine C-terminal de DiscII et HPA, basé sur les homologies structurales.....	- 133 -
Figure 5.3.	Gels SDS-PAGE coloré au bleu de Coomassie après expression de DiscII-Cter-boucleHPA dans Rosetta 2 pLysS. ....	- 134 -
Figure 5.4.	Profils d'élution obtenus lors des différentes purifications de DiscII-Cter-boucleHPA par chromatographie sur colonne de nickel. ....	- 136 -
Figure 5.5.	Contrôle de la fonctionnalité de DiscII-boucleHPA par chromatographie d'affinité sur colonne de GalNAc (A.).....	- 137 -
Figure 5.6.	Contrôle de l'oligomérisation de DiscII-Cter-boucleHPA par chromatographie d'exclusion stérique sur colonne Superdex 75. ....	- 139 -
Figure 5.7.	A. Superposition du domaine C-terminal de DiscII (vert, rouge, bleu) avec HPA (jaune, magenta, cyan) et B. Alignement de séquences entre HPA et les domaines C-terminaux de DiscII et DiscII-boucleHPA-mutée.....	- 140 -
Figure 5.8.	Profils d'élution après purification sur colonne de nickel de DiscII-boucleHPA-mutée exprimée dans A. Arctic Express RIL, B. BL21 (DE3) C41 et C. Rosetta 2 pLysS .....	- 141 -
Figure 5.9.	Gel SDS-PAGE 12% acrylamide coloré au bleu de Coomassie après expression de DiscII-boucleHPA-mutée dans Rosetta 2 pLysS et purification sur colonne de nickel.....	- 142 -
Figure 5.10.	Contrôle de la fonctionnalité de DiscII-boucleHPA-mutée grâce au test d'hémagglutination.....	- 143 -

Figure 5.11.	A. Superposition des structures cristallographiques d'une face de liaison au GalNAc de DiscII-boucleHPA mutée (en rose), de DiscII (en vert) et d'HPA (en bleu). B. Boucles de spécificité de DiscII-boucleHPA mutée (en rose) et d'HPA (en bleu).....	- 145 -
Figure 5.12.	Superposition des structures cristallographiques des monomères de DiscII-boucleHPA muté (en rose) et d'HPA (en bleu).....	- 146 -
Figure 5.13.	Profils d'élution de la purification par chromatographie sur colonne de nickel de A. DiscII-boucleHPA mutée-G247H, B. DiscII-Cter-boucleHPA mutée-G247H, C. DiscII-boucleHPA mutée-G247H-I171S-F243S et D. DiscII-Cter-boucleHPA mutée-G247H-I171S-F243S.....	- 149 -
Figure 5.14.	Séquence de la protéine hybride DiscII-Nter-HPA.....	- 150 -
Figure 5.15.	A. Gel SDS-PAGE 12% acrylamide coloré au bleu de Coomassie après expression de DiscII-Nter-HPA dans Rosetta gami 2 pLysS et purification sur colonne de nickel (B.) et sur colonne GalNAc-agarose (C.).....	- 151 -
Figure 6.1.	Représentations schématiques des trimères du domaine N-terminal des discoidines (A.), de la lectine AAA complexée au fucose (PDB 1K12) (Bianchet <i>et al.</i> 2002) (B.) et de la lectine F complexée au fucose (PDB 3CQO) (Bianchet <i>et al.</i> ) (C.).....	- 156 -
Figure 6.2.	Superposition de site de reconnaissance du domaine DS. ....	- 156 -
Figure 6.3.	Expression de DiscI-Nter-P153 dans <i>E. coli</i> Rosetta 2 (DE3) pLysS. A. Analyse par SDS-PAGE 15% et B. Immunoblot. ....	- 158 -
Figure 6.4.	Contrôle de l'expression de DiscI-Nter-Q155 et DiscI-Nter-S157 dans Rosetta 2 (DE3) pLysS, par analyse SDS-PAGE 15%. ....	- 159 -
Figure 6.5.	Contrôle de l'expression de DiscII-Nter dans Rosetta 2 (DE3) pLysS, par A. analyse SDS-PAGE 15% et B. immunoblot.....	- 161 -
Figure 6.6.	Profils d'élution obtenus lors de deux purifications différentes sur colonne de nickel du domaine N-terminal de DiscII (A. et B.) et analyse SDS-PAGE 15% (C.).....	- 162 -
Figure 6.7.	Contrôle de la fonctionnalité de DiscII-Nter grâce au test d'hémagglutination..	- 162 -
Figure 6.8.	Comparaison des structures des monosaccharides de configuration <i>L-galacto</i> et <i>D-manno</i> .....	- 163 -

Figure 6.9.	Mesure d'interaction entre DiscII-Nter et les sucres PAA- $\alpha$ -L-Fuc et PAA- $\alpha$ -D-Man par RPS.....	- 164 -
Figure 6.10.	Variation de l'intensité d'interaction entre le PAA- $\alpha$ -D-Man (A.), le PAA- $\alpha$ -L-Fuc (B.) en fonction de la concentration en protéine DiscII-Nter par RPS.-	165 -
Figure 6.11.	Compétition entre des inhibiteurs en solution et les sucres PAA-biotinylés, respectivement PAA- $\alpha$ -D-Man (A.) et PAA- $\alpha$ -L-Fuc (B.) pour la fixation de DiscII-Nter (c=0,5 $\mu$ M), évaluée par RPS.....	- 167 -
Figure 7.1.	Alignement de séquences peptidiques des lectines de type H, HPA, DiscI et DiscII et SLL-2a. ....	- 171 -
Figure 7.2.	Alignement de séquences nucléotidiques entre les gènes synthétiques, <i>sll-2a</i> et <i>sll-2a-2</i> et le gène sauvage, issu de (Jimbo <i>et al.</i> 2005). ....	- 172 -
Figure 7.3.	Purification de SLL-2a clonée dans le vecteur pProEx-HtB et exprimée dans la souche BL21 (DE3). ....	- 175 -
Figure 7.4.	A. Gel SDS-PAGE 15 % acrylamide coloré au bleu de Coomassie après expression de SLL-2a clonée dans le vecteur pET-TEV et exprimée dans les BL21 (DE3), B. Immunomarquage correspondant.....	- 176 -
Figure 7.5.	Gel SDS-PAGE 15 % acrylamide coloré au bleu de Coomassie et immunomarquage après expression de SLL-2a-2/GST.....	- 178 -
Figure 7.6.	A. Renaturation des corps d'inclusion de SLL-2a par chromatographie sur colonne de nickel. B. Gel SDS-PAGE 15 % acrylamide coloré au bleu de Coomassie après expression de SLL-2a clonée dans le vecteur pET-TEV.-	184 -
Figure 7.7.	Protocoles de renaturation utilisés pour les lectines CHA-II et SLL-2a... -	186 -
Figure 7.8.	Schéma des constructions utilisées pour produire SLL-2a et SLL-2a-2 chez <i>P. pastoris</i> . ....	- 188 -
Figure 7.9.	Schémas des recombinaisons homologues possibles dans le vecteur pPICZ- $\alpha$ chez la levure <i>Pichia pastoris</i> .....	- 189 -
Figure 7.10.	Analyse sur gel d'agarose 1,5% de l'amplification des gènes <i>sll-2a</i> et..... <i>sll-2a-2</i> clonés dans pPIC-Z $\alpha$ -HisFlag à partir de l'ADN génomique de la levure. ....	- 190 -

- Figure 7.11. Analyse sur gel SDS-PAGE 15% acrylamide coloré au bleu de Coomassie des surnageants de culture de l'expression de SLL-2a (A.) et SLL-2a-2 (B.) (clonées dans pPICZ $\alpha$ -A) dans *P. pastoris*, après trois jours d'induction au méthanol 0,5%. ..... - 192 -
- Figure 7.12. Courbes de croissance des levures *P. pastoris* transformées avec SLL-2a et SLL-2a-2 dans le milieu MMYC, à 30°C..... - 194 -
- Figure 7.13. Alignements des séquences peptidiques des trois isogènes codant pour la lectine SLL-2 identifiés dans le corail *Sinularia lochmodes*. ..... - 196 -

## **Annexes**

- Figure 1. Analyse, par spectrométrie de masse, des *N*-glycanes extraits du sérum de veau fœtal commercialisé par Sigma-Aldrich. .... - 221 -
- Figure 2. Analyse, par spectrométrie de masse, des *N*-glycanes extraits des glycoprotéines membranaires des lignées de cancer colorectal non métastatique SW 480 et métastatique SW 620..... - 225 -



# Liste des tableaux

---

Tableau 1.1.	Principales structures de glycosphingolipides cancéreux chez l'homme....	- 32 -
Tableau 1.2.	Récapitulatif des structures tridimensionnelles de lectines.....	- 37 -
Tableau 1.3.	Principales lectines utilisées pour la mise en évidence de glycotopes en oncologie.....	- 39 -
Tableau 1.4.	Comparaison entre les propriétés biochimiques des DiscI et DiscII.....	- 53 -
Tableau 2.1.	Description des lignées cellulaires utilisées dans cette étude. ....	- 63 -
Tableau 2.2.	Récapitulatif des souches et plasmides utilisés dans la partie biologie moléculaire de cette étude. ....	- 65 -
Tableau 2.3.	Récapitulatif des amorces utilisées pour cette étude.....	- 69 -
Tableau 2.4.	Récapitulatif des clonages des gènes SLL2a et SLL2a-2. ....	- 75 -
Tableau 2.5.	Récapitulatif des constructions des gène SLL2a et SLL2a-2 obtenues par digestion enzymatique. ....	- 76 -
Tableau 2.6.	Récapitulatif des différentes souches <i>d'E. coli</i> utilisées pour l'expression de protéines sous forme recombinante. ....	- 79 -
Tableau 2.7.	Récapitulatif des différentes chromatographies d'exclusion stérique testées durant cette étude.....	- 83 -
Tableau 2.8.	Récapitulatif des tampons de solubilisation utilisés pour le refolding des lectines de <i>Sinularia lochmodes</i> .....	- 85 -
Tableau 2.9.	Récapitulatif des tampons de renaturation utilisés dans cette étude. ....	- 86 -
Tableau 2.10.	Tampons disponibles dans le kit de renaturation QuickFold™ Protein Refolding commercialisé par Athena ES.....	- 87 -
Tableau 2.11.	Conditions expérimentales pour les mesures d'interaction entre DiscII-Nter et les PAA-Man et PAA-Fuc par RPS.....	- 93 -
Tableau 3.1.	Caractéristiques des lignées de cancers colorectaux utilisées.....	- 107 -

Tableau 4.1.	Récapitulatif des conditions expérimentales et de données pour la résolution des structures cristallographiques de DiscI, native et en complexe avec le GalNAc et les disaccharides Gal $\beta$ 1-3GalNAc et GalNAc $\beta$ 1-3Gal.....	- 120 -
Tableau 4.2.	Récapitulatif des contacts établis entre DiscI et ses ligands disaccharidiques dans les structures cristallographiques.....	- 126 -
Tableau 5.1.	Récapitulatif des conditions d'expression dans <i>E. coli</i> des mutants DiscII-boucleHPA, DiscII-Cter-boucleHPA et DiscII-Cter-N257.....	- 133 -
Tableau 5.2.	Récapitulatif des différents mutants de DiscII construits.....	- 153 -
Tableau 6.1.	Récapitulatif des conditions d'expression dans Rosetta 2 (DE3) pLysS des domaines N-terminaux de DiscI. ....	- 157 -
Tableau 6.2.	Récapitulatif des conditions d'expression dans Rosetta 2 (DE3) pLysS du domaine N-terminal de DiscII. ....	- 160 -
Tableau 6.3.	Récapitulatif des composés testés comme inhibiteurs de l'interaction DiscII-Nter et les sucres PAA-biotinylés, PAA- $\alpha$ -D-Man (A.) et PAA- $\alpha$ -L-Fuc (B.) en RPS. ....	- 166 -
Tableau 7.1.	Conditions d'expression testées pour l'expression de SLL-2a dans la bactérie <i>E. coli</i> .....	- 174 -
Tableau 7.2.	Conditions d'expression testées pour l'expression de SLL-2a-2 dans la bactérie <i>E. coli</i> .....	- 177 -
Tableau 7.3.	Récapitulatif des conditions de solubilisation des corps d'inclusion. ....	- 181 -
Tableau 7.4.	Récapitulatif des conditions de solubilisation des corps d'inclusion de SLL-2a, clonée dans le vecteur pET-TEV et exprimée dans la souche BL21 (DE3).-	182
	-	
Tableau 7.5.	Tailles attendues après amplification de <i>sll-2</i> et <i>sll-2a-2</i> avec les amorces 5' AOX1 / 3' AOX1 et pPicZa-F / -R sur l'ADN génomique de <i>P. pastoris</i> .-	190
	-	

# **Chapitre 1**

## **Introduction**



---

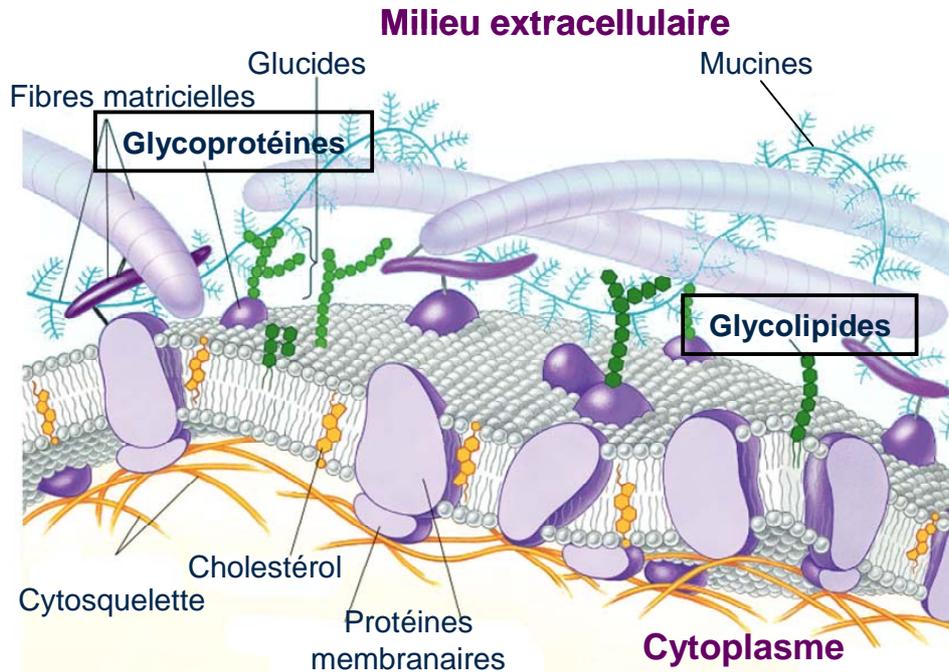
# Chapitre 1

## Introduction

---

### 1. Généralités sur la glycosylation

La glycosylation est la modification co- et post-traductionnelle la plus abondante dans le règne vivant. Elle produit un important répertoire de glucides (ou hydrates de carbone), généralement liés de manière covalente à des protéines ou à des lipides, regroupés alors sous le terme de glycoconjugués. Ils forment une épaisse couche, qui tapisse les membranes externes des cellules, désignée sous le terme « glycocalice ». Elle a pour rôle de protéger la cellule des agressions chimiques et/ou mécaniques (**Figure 1.1.**).



**Figure 1.1. Schéma d'une membrane cellulaire présentant différentes structures glycaniques (glycoprotéines et glycolipides).**

D'après Copyright© 2002 Pearson Education, Inc., publishing as Benjamin Cummings

Les glucides, aussi appelés glycanes, participent à de nombreux processus biologiques indispensables à la vie de la cellule tels que l'adhésion et le trafic cellulaire, la discrimination du soi et du non-soi, l'activation de récepteurs, la transduction du signal ou encore l'endocytose (Ohtsubo and Marth 2006).

Deux grandes familles de glycoprotéines ont été définies selon la nature de la liaison covalente établie entre le glycane et la partie peptidique : les *N*- et les *O*-glycoprotéines. Ces deux types de glycosylation peuvent se retrouver simultanément sur une même protéine. La *C*-mannosylation, découverte il y a une quinzaine d'années, complète maintenant les exemples de glycosylation rencontrés dans la cellule. Il est estimé qu'environ 70% des protéines humaines sont glycosylées (Apweiler *et al.* 1999).

## 1.1. La *N*-glycosylation

### 1.1.1. Structure et diversité des *N*-glycanes chez les mammifères

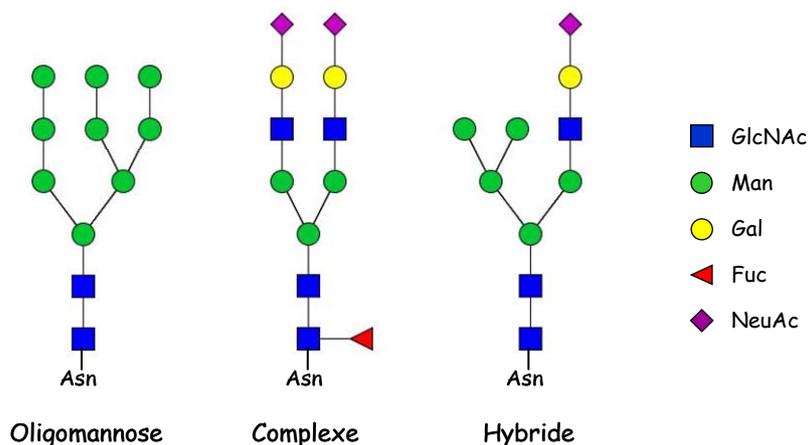
La *N*-glycosylation est une modification co- et post-traductionnelle qui débute dans le réticulum endoplasmique (RE) et qui se poursuit dans l'appareil de Golgi. Elle consiste en l'établissement d'un lien amine sur une asparagine d'une protéine appartenant à la séquence consensus N-X-S/T où X peut être n'importe quel acide aminé excepté une proline. Ces séquences, appelées aussi séquons, sont souvent repérées comme des sites potentiels de *N*-glycosylation même si toutes les asparagines appartenant à des séquons ne sont pas glycosylées. Les *N*-glycanes sont des appendices hydrophiles constitués d'une structure pentasaccharidique commune formée de deux résidus *N*-acétylglucosamine (GlcNAc) et de trois résidus mannose (Man) : Man<sub>3</sub>GlcNAc<sub>2</sub>, appelée le noyau (Kornfeld *et al.* 1978 ; Li *et al.* 1978 ; Tabas and Kornfeld 1978) (revues (Hubbard and Ivatt 1981 ; Kornfeld and Kornfeld 1985 ; Molinari 2007). Il est ensuite utilisé comme substrat pour la diversification des *N*-glycanes dans l'appareil de Golgi.

Globalement, trois grandes catégories de *N*-glycanes peuvent être identifiées (**Figure 1.2.**) :

- Le type oligomannosidique correspond aux structures de type Man<sub>3</sub>GlcNAc<sub>2</sub>, additionnés de résidus mannoses.

- Le type complexe est une structure bi-antennée dans laquelle le noyau pentasaccharidique est substitué par des résidus GlcNAc et galactose (Gal) et se termine par un acide sialique (NeuAc). Un résidu fucose (Fuc) peut également être lié au premier résidu GlcNAc.

- Le type hybride ou mixte possède généralement les caractéristiques des *N*-glycanes de type oligomannosidique sur une branche et ceux de type complexe sur l'autre branche.



**Figure 1.2. Exemples des trois types de *N*-glycanes chez l'homme : type oligomannosique, type complexe et type hybride.**

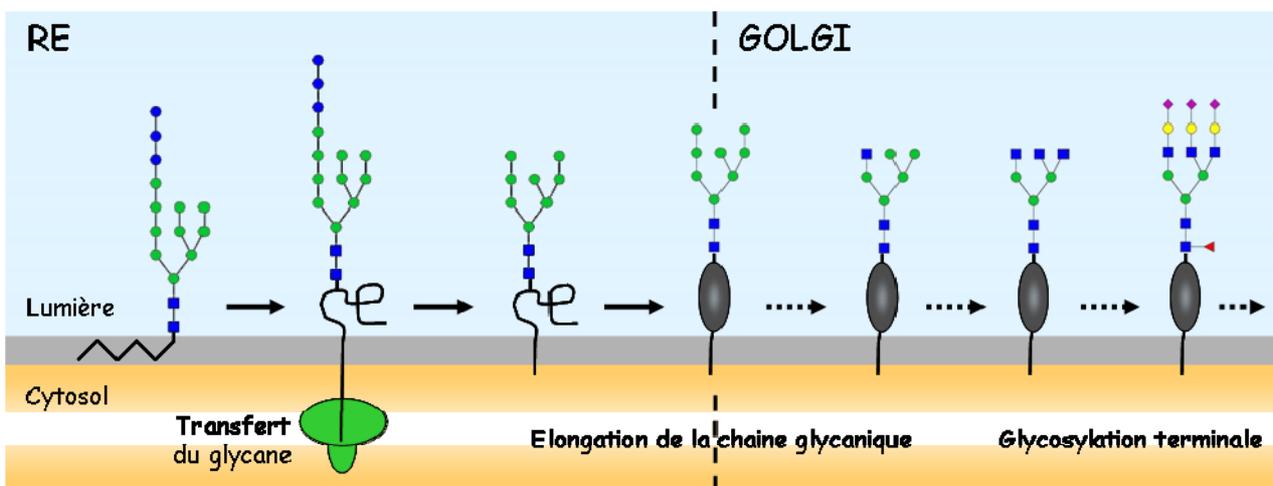
La plupart des *N*-glycanes sécrétés ou présents à la surface cellulaire sont de type complexe. Les carrés bleus schématisent les résidus GlcNAc, les ronds verts et jaunes schématisent respectivement les résidus Man et Gal, le triangle rouge schématise un résidu Fuc et NeuAc est schématisé par un losange violet. Les images ont été réalisées sur la plateforme GlycoWork Bench (Ceroni *et al.* 2007).

### 1.1.2. Biosynthèse des *N*-glycanes

La synthèse des *N*-glycanes dans les cellules de mammifères débute par l'assemblage d'un dodécaoligosaccharide, de structure  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ , sur un précurseur lipidique, le dolicholpyrophosphate dans le réticulum endoplasmique. Ce cœur oligosaccharidique est alors transféré d'un bloc du transporteur lipidique sur la chaîne polypeptidique naissante via l'établissement d'un lien *N*-glycosidique sur une asparagine appartenant à un séquen. Immédiatement après le couplage sur la protéine, la chaîne oligosaccharidique est écourtée au niveau des résidus glucose et mannose terminaux par des enzymes du réticulum endoplasmique. La protéine est alors acheminée dans l'appareil de Golgi où elle subit de

nouvelles modifications séquentielles par addition de glycanes notamment (**Figure 1.3.**) (Helenius and Aebi 2001).

Plusieurs sites de *N*-glycosylation sur la même protéine peuvent être glycosylés de manière différente. La séquence peptidique, mais surtout la conformation de la protéine, peut influencer les différentes étapes réticulaires et golgiennes de la *N*-glycosylation. La proportion relative de ces différentes structures varie en fonction des espèces, de l'état physiologique ou pathologique de la cellule (stress, cancer...), du tissu ou de l'individu.



**Figure 1.3. Voie de synthèse des *N*-glycanes dans le réticulum endoplasmique et l'appareil de Golgi (Helenius and Aebi 2001).**

L'ajout de *N*-glycanes a lieu sur une asparagine appartenant à la séquence consensus N-X-S/T où X peut être n'importe quel acide aminé excepté une proline. Les carrés bleus schématisent les résidus GlcNAc, les ronds verts et jaunes schématisent respectivement les résidus Man et Gal, le triangle rouge schématise un résidu Fuc et NeuAc est schématisé par un losange violet. Les images ont été réalisées sur la plateforme GlycoWork Bench (Ceroni *et al.* 2007).

## 1.2. La *O*-glycosylation

Si le rôle des *N*-glycanes est aujourd'hui relativement bien connu, la complexité de la *O*-glycosylation rend son étude plus difficile. En effet, les *O*-glycanes constituent une vaste famille qui peut-être divisée en plusieurs sous-groupes selon la nature du sucre lié à la protéine :

- la *O*-glycosylation de type mucine (Ser/Thr-*O*- $\alpha$ -*N*-acétylgalactosamine (GalNAc))
- la glycosylation de type *O*-GlcNAc (Ser/Thr-*O*- $\beta$ GlcNAc)

- la *O*-fucosylation (Ser/Thr-*O*- $\alpha$ Fuc)
- la *O*-glucosylation (Ser/Thr-*O*- $\alpha$ Glc)
- la *O*-mannosylation (Ser/Thr-*O*- $\alpha$ Man)

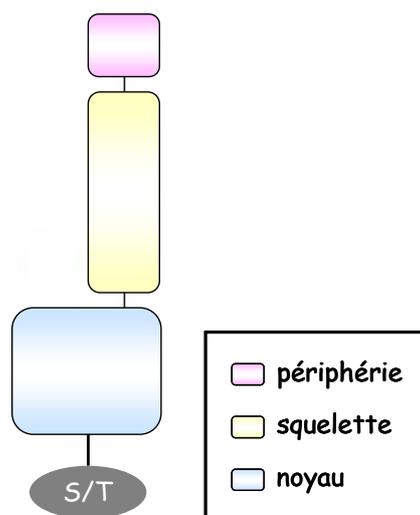
### 1.2.1. Structure et biosynthèse des *O*-glycanes de type mucine

Les *O*-glycanes de type « mucine » ne sont pas préassemblés sur un dérivé dolichol, mais chaque sucre est transféré de manière individuelle et séquentielle à partir d'un nucléotide-sucre. Elle est initiée, dans l'appareil de Golgi, par le transfert d'un résidu GalNAc en configuration  $\alpha$  sur l'hydroxyle de la chaîne latérale d'une sérine ou d'une thréonine (Ser/Thr-*O*- $\alpha$ GalNAc). Une cascade de réactions enzymatiques orchestrées par différentes enzymes allonge ensuite le *O*-glycane de manière à former plusieurs structures de cœur ou noyau, pouvant eux mêmes être modifiés par sialylation, sulfatation ou acétylation (**Figure 1.4.**). Les structures *O*-glycaniques sont classiquement décomposées en trois parties : le noyau (ou cœur), le squelette qui peut être linéaire ou branché et la périphérie (**Figure 1.5.**).



**Figure 1.4. Structures des noyaux identifiés à ce jour dans les chaînes *O*-glycaniques de type mucine.**

Chez les mammifères, sont principalement rencontrés les noyaux 1, 2, 3 et 4. Les noyaux 5 à 8 sont plus marginaux et peuvent être localisés sur des cellules cancéreuses.



**Figure 1.5.** Schéma de la structure d'un *O*-glycane de type mucine.

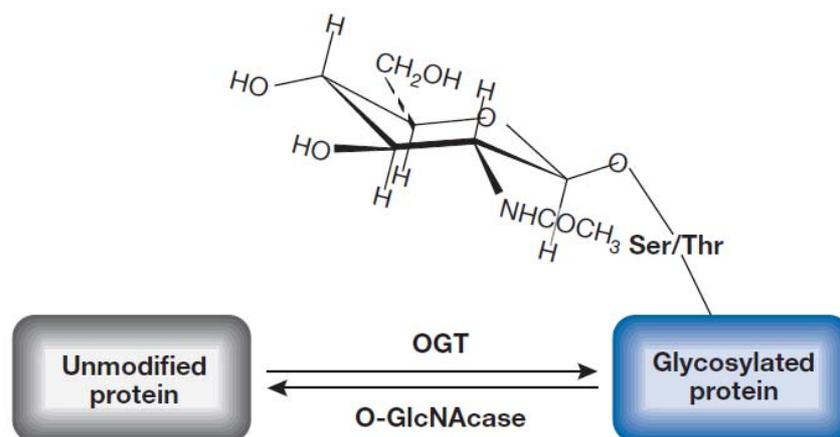
### 1.2.2. Les mucines

Les mucines sont définies comme des glycoprotéines, très fortement *O*-glycosylées. Les chaînes oligosaccharidiques, de plusieurs centaines à plusieurs milliers de kDa, représentent jusqu'à 80% du poids total de la protéine (Lamblin *et al.* 1991). L'abondance de cette *O*-glycosylation est due à la présence de régions tandem extrêmement riches en sérine et thréonine et/ou proline, que l'on appelle communément VNTR pour variable number of tandem regions. L'expression des mucines, sécrétées ou reliées à la membrane, est particulièrement abondante dans les cellules épithéliales, spécialisées dans la production de mucus où elles assurent, au contact de l'environnement extérieur, un rôle de barrière protectrice. Elles participent également au contrôle de l'antigénicité, du système immunitaire et peuvent être impliquées dans les phénomènes d'adhésion cellulaire. Depuis le début des années 1990, vingt gènes appelés MUC codant pour les squelettes protéiques des mucines ont été identifiés (Baldus *et al.* 2004). Même si la majorité des mucines de mammifères possèdent comme structures de cœur, les noyaux 1 et 2 (**Figure 1.4.**), l'élongation des chaînes glycaniques engendre une extrême diversité d'extensions linéaires ou branchées, pouvant être sialylées.

### 1.2.3. Autres types de *O*-glycosylation

#### 1.2.3.1. La glycosylation de type *O*-GlcNAc

Découverte en 1984 par G. Hart (Torres and Hart 1984), la glycosylation de type *O*-GlcNAc est une modification post-traductionnelle dynamique et réversible de nombreuses protéines cytosoliques et nucléaires. Elle consiste en l'addition d'un résidu unique de *N*-acétylglucosamine (GlcNAc) sur une sérine ou une thréonine de la chaîne polypeptidique. Cette glycosylation est régulée par un couple unique d'enzymes : la *O*-*N*-acétylglucosaminyltransférase (OGT) transfère le GlcNAc sur la protéine à partir d'un nucléotide-sucré donneur, l'UDP-GlcNAc (Haltiwanger *et al.* 1990) alors que la *O*-*N*-acétylglucosaminidase (*O*-GlcNAcase) catalyse son hydrolyse (Dong and Hart 1994) (revue par (Ngho and Jones 2008)) (**Figure 1.6.**). Ces deux types de réaction peuvent entrer en compétition avec la réaction de phosphorylation des mêmes résidus sérine ou thréonine ou sur des résidus voisins. À ce jour, plus d'une centaine de protéines modifiées par la glycosylation de type *O*-GlcNAc ont été identifiées, dont des kinases, des phosphatases, des facteurs de transcription ou encore des protéines du cytosquelette mais le rôle de cette glycosylation reste encore mal compris. Elle serait impliquée dans divers processus cellulaires tels que la régulation de la transcription et de la traduction, la protection des protéines vis-à-vis de la dégradation, le trafic intracellulaire, la modulation des interactions protéines-protéines et le système immunitaire (Golks and Guerini 2008).



**Figure 1.6.** Représentation schématique de la glycosylation de type *O*-GlcNAc (Golks and Guerini 2008).

### 1.2.3.2. La *O*-fucosylation et la *O*-glucosylation

Certaines régions des facteurs de croissance épidermique (EGF pour Epidermal Growth Factor) forment des motifs structuraux communs de 30 à 40 acides aminés dont six cystéines sont impliquées dans la formation de trois ponts disulfure conservés. Ces motifs peuvent être *O*-glucosylés et *O*-fucosylés de la manière suivante. L'addition du glucose sur un résidu sérine appartenant à une séquence consensus putative C-X-S-X-P-C est catalysée par une *O*-glucosyltransférase du réticulum endoplasmique appelée Poglut chez les mammifères. Une élongation de la chaîne sur le glucose par un ou deux xyloses liés en  $\alpha$ 1-3 peut éventuellement avoir lieu dans l'appareil de Golgi. La *O*-fucosylation débute par le transfert d'un résidu fucose sur une sérine ou thréonine localisée sur la séquence C-X<sub>3-5</sub>-S/T-C, dans le réticulum endoplasmique, par une *O*-fucosyltransférase 1 appelée Pofut 1 chez les mammifères et OFut1 chez la drosophile. La chaîne peut alors être étendue dans le Golgi pour former un tétrasaccharide NeuAc $\alpha$ 2-6Gal $\beta$ 1-4GlcNAc $\beta$ 1-3Fuc $\alpha$ 1-S/T (Wopereis *et al.* 2006).

Ces deux types de glycosylation sont particulièrement importants dans la voie de signalisation Notch. Présente dans la plupart des organismes multicellulaires, cette voie de signalisation très conservée est fondamentale dans la communication cellulaire, notamment au niveau du développement neuronal ou de la vie embryogénique. Le domaine extracellulaire des récepteurs Notch est composé d'unités de répétition EGF et est donc *O*-glucosylé et *O*-fucosylé. Cette *O*-fucosylation est indispensable à sa fonction puisqu'une perte de l'activité *O*-FucT-1 conduit à des phénotypes, chez la mouche drosophile et la souris, qui ressemblent à ceux observés en absence complète de signalisation Notch et perturbent donc fortement le développement de l'animal (Luther and Haltiwanger 2009).

### 1.2.3.3. La *O*-mannosylation

Identifiée en 1968 chez la levure de boulanger (Sentandreu and Northcote 1968), la *O*-mannosylation consiste, quant à elle, en l'établissement d'une liaison *O*-glycosidique entre un résidu  $\alpha$ -mannosyle et un acide aminé Ser ou Thr de la chaîne peptidique. Catalysée dans le réticulum endoplasmique à partir de dolichol-phosphate mannose (Dol-P-Man) par des dolichyl phosphate D-mannose : protein *O*-mannosyltransférases (PMTs), cette réaction initiale de *O*-mannosylation est relativement bien conservée entre les règnes animal et fongique. La chaîne oligosaccharidique peut, par la suite, être étendue par l'action d'autres glycosyltransférases et diffère selon les espèces. Chez les levures et les champignons, le motif généralement observé est le mannotriose lié en  $\alpha$ 1-2, Man $\alpha$ 1-2Man $\alpha$ 1-2Man $\alpha$ 1-Ser/Thr alors

que chez les mammifères, la grande majorité des glycanes *O*-mannosylés représente des variations du tétrasaccharide NeuAc $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-2Man $\alpha$ 1-Ser/Thr avec des extensions diverses et la présence facultative de fucose lié en  $\alpha$ 1-3 au GlcNAc. Ces structures sont notamment retrouvées sur les protéoglycanes chondroïtine sulfatés et les tissus du cerveau. La *O*-mannosylation assure des rôles très divers suivant l'organisme. Par exemple, chez les champignons, elle participe essentiellement au maintien de l'intégrité de la paroi cellulaire. Chez la levure, elle serait impliquée dans la solubilisation des protéines mal repliées dans le réticulum endoplasmique alors que chez les humains, un problème de *O*-mannosylation est associé à des dystrophies musculaires responsables d'aberrations lourdes au niveau du cerveau et des yeux (Lommel and Strahl 2009).

#### 1.2.3.4. Glycosylation sur des acides aminés hydroxylés

La *O*-glycosylation peut également avoir lieu sur d'autres résidus qu'une sérine ou une thréonine. Ainsi, dans le cas du collagène, des résidus hydroxylysine sont glycosylés, dans le réticulum endoplasmique, par l'ajout d'un résidu Gal, lié en  $\beta$ . La chaîne glycanique peut ensuite être élongée par des Glc, liés en  $\alpha$ 1-2 au Gal précédent (Spiro 1969). Cette glycosylation semble être impliquée dans les fonctions de sécrétion et d'assemblage des collagènes (Sipila *et al.* 2007) mais aussi dans la formation des membranes basales (Ruotsalainen *et al.* 2006).

Les prolines sont également des acides aminés qui peuvent être hydroxylés et glycosylés. Chez l'amibe *Dictyostelium discoideum*, l'hydroxyproline 143 de la protéine Skp1, impliquée dans les phénomènes de polyubiquitinylation des protéines, est glycosylée par l'oligosaccharide suivant : Gal $\alpha$ 1-6Gal $\alpha$ 1Fuc $\alpha$ 1-2Gal $\beta$ 1-3GlcNAc-protéine (Tengumnuay *et al.* 1998 ; Van Der Wel *et al.* 2002). Cette glycosylation est plus généralement rencontrée chez les plantes et les algues vertes, impliquant cette fois des résidus Gal et arabinose (Chrispeels 1970). Les protéines glycosylées sont largement impliquées dans des phénomènes de croissance et de développement.

Ces différents exemples permettent de mettre en évidence la variabilité de structures existant dans la *O*-glycosylation et reflètent ainsi sa grande complexité (Van den Steen *et al.* 1998).

### 1.3. La C-mannosylation

En 1995, un nouveau type de glycosylation, la C-mannosylation, a été identifié sur la protéine humaine ribonucléase 2 (de Beer *et al.* 1995). Elle consiste en l'établissement d'une liaison carbone-carbone entre un résidu  $\alpha$ -mannoside et le carbone 2 du premier tryptophane du séquen W-X-X-W. Ce transfert est catalysé, dans le réticulum endoplasmique, par une enzyme, la C-mannosyltransférase qui utilise, comme pour la O-mannosylation, un dolichol-phosphate-mannose comme donneur (Doucey *et al.* 1998). L'influence de cette glycosylation sur le repliement protéique est peu connue mais l'addition d'un résidu mannose hydrophile sur un acide aminé hydrophobe perturbe sans aucun doute l'état d'enfouissement du tryptophane dans le cœur protéique.

A ce jour, plusieurs protéines C-mannosylées ont été identifiées et même si le rôle de cette glycosylation n'est toujours pas clairement défini, elle semble être impliquée dans les phénomènes de signalisation cellulaire (Luther and Haltiwanger 2009). La C-mannosylation importante de plusieurs protéines appartenant au système du complément laisse également envisager un rôle de cette glycosylation dans la formation et l'assemblage du complexe protéique à la surface des pathogènes (Hofsteenge *et al.* 1999).

La présence de l'enzyme C-mannosyltransférase de manière ubiquitaire chez de nombreux vertébrés et insectes ainsi que l'identification de la séquence consensus dans environ 340 protéines de mammifères laisse présager l'émergence de la C-mannosylation comme glycosylation majeure.

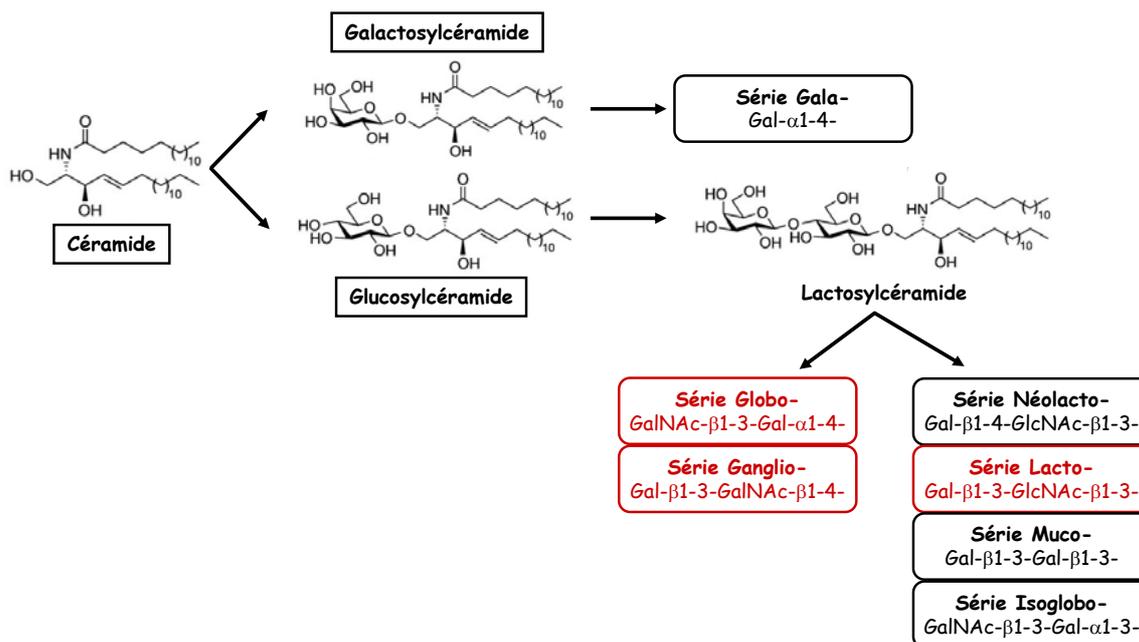
### 1.4. Autres types de glycoconjugués

#### 1.4.1. Les glycolipides

Des motifs glucidiques sont également retrouvés sur les lipides pour former la classe des glycolipides, composée de deux familles selon la nature de la base lipidique : les glycosphingolipides et les glyco-glycérolipides. Ces derniers étant très rares dans le règne animal, ils ne seront pas abordés dans ce manuscrit.

Les glycosphingolipides sont des composés ubiquitaires des membranes cellulaires de mammifères qui résultent de l'association d'une sphingosine et d'un acide gras pour former

un céramide substitué en C3 soit par un résidu galactoside soit par un résidu glucoside formant respectivement le galactosylcéramide (Bosio *et al.* 1996) ou le glucosylcéramide (Ichikawa *et al.* 1996). Un résidu Gal peut encore être ajouté sur ce dernier, donnant naissance au lactosylcéramide, précurseur notamment des séries Globo-, Ganglio- et Lacto- (**Figure 1.7.**) (Takizawa *et al.* 1999). Chez les animaux, ces glycolipides sont localisés dans la membrane externe où ils assurent des rôles variés dans les interactions cellule-cellule et cellule-matrice.



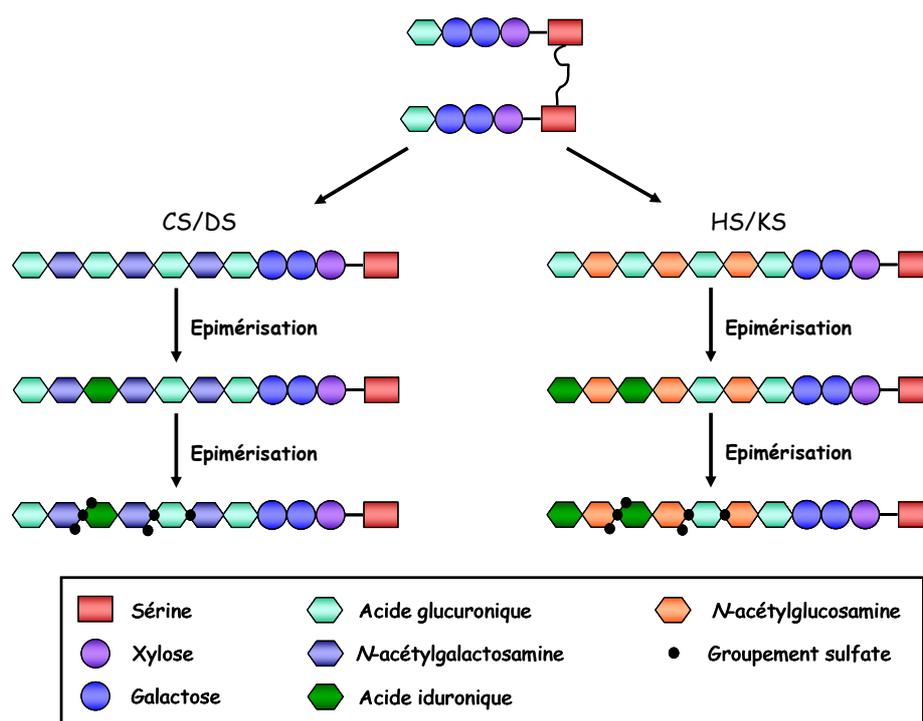
**Figure 1.7.** Schéma des voies de biosynthèse des glycosphingolipides chez les mammifères (Wennekes *et al.* 2009).

#### 1.4.2. Les protéoglycanes

Les protéoglycanes sont des molécules formées d'un cœur protéique sur lequel sont accrochées une ou plusieurs chaînes de glycosaminoglycanes (GAGs). Ces GAGs sont des polysaccharides non branchés, chargés négativement, de haut poids moléculaire, composés d'un squelette d'unités disaccharidiques répétées. Ce dernier est formé d'un sucre aminé (GlcNAc ou GalNAc) et d'un acide uronique (glucuronique ou iduronique). Ces GAGs sont principalement localisés dans la matrice extracellulaire.

Les protéoglycanes sont classés dans différentes catégories selon la nature du motif glycanique : les chondroïtines sulfates (CS) et les dermatanes sulfates (DS) (contenant du

GalNAc), les héparane-sulfates (HS) et les kératane-sulfates (KS) (contenant du GlcNAc). La synthèse des chaînes de GAGs est initiée par l'addition d'un sucre aminé sur un bras espaceur tétrassaccharidique, lui-même attaché à une sérine via un lien *O*-glycosidique (GlcA $\beta$ 1-3Gal $\beta$ 1-3Gal $\beta$ 1-4Xyl $\beta$ 1-*O*-Ser). La chaîne polysaccharidique peut également subir plusieurs modifications telles que la *N*-déacétylation-*N*-sulphatation, la 2-*O*- et 6-*O*-sulphatation et l'épimérisation de l'acide uronique. Ces modifications sont à l'origine de la grande variété rencontrée dans les chaînes de GAGs (Hacker *et al.* 2005) (**Figure 1.8.**).

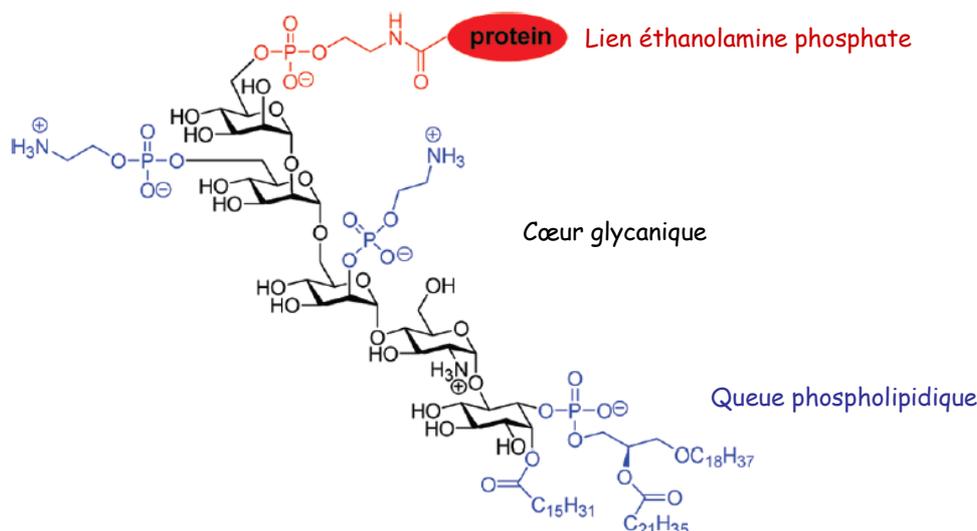


**Figure 1.8.** Biosynthèse schématisée des glycosaminoglycans (Hacker *et al.* 2005).

### 1.4.3. Les ancres GPI

Les protéines membranaires peuvent également être ancrées dans la membrane par l'intermédiaire de « protéines glypiées » ou « protéines à ancre GPI ». Le motif GPI (glycosyl phosphatidylinositol) est présent chez de nombreux eucaryotes dont les mammifères, les plantes, les levures et les protozoaires. La synthèse a lieu dans le réticulum endoplasmique et débute par le transfert d'un GlcNAc sur un phosphatidylinositol (PI), suivi de l'addition de trois résidus mannose pour former un cœur glycanique très conservé (Man $\alpha$ 1-2Man $\alpha$ 1-6Man $\alpha$ 1-4GlcNAc $\alpha$ 1-6PI) et d'un groupement éthanolamine phosphate. L'intégralité de ces groupements est alors transférée en bloc sur l'extrémité C-terminale d'une protéine. Cette

réaction de transfert est parfois appelée glypiation. Une queue phospholipidique attache alors l'ancre GPI à la membrane cellulaire. Le cœur glycanique peut être modifié par différents groupements tels que des sucres (Man, Gal, NeuAc...) ou des groupements éthanolamine phosphate selon les espèces (**Figure 1.9.**) (revue par (Paulick and Bertozzi 2008)).



**Figure 1.9. Structure d'une ancre GPI (Paulick and Bertozzi 2008).**

Les trois domaines d'une ancre GPI sont : (i) le lien éthanolamine phosphate en rouge, (ii) le cœur glycanique conservé en noir et (iii) la queue phospholipidique en bleue. Les appendices en bleu (incluant les lipides de la queue phospholipidique) sont variables selon l'espèce.

## 1.5. Les glycosyltransférases

### 1.5.1. Définition et nomenclature

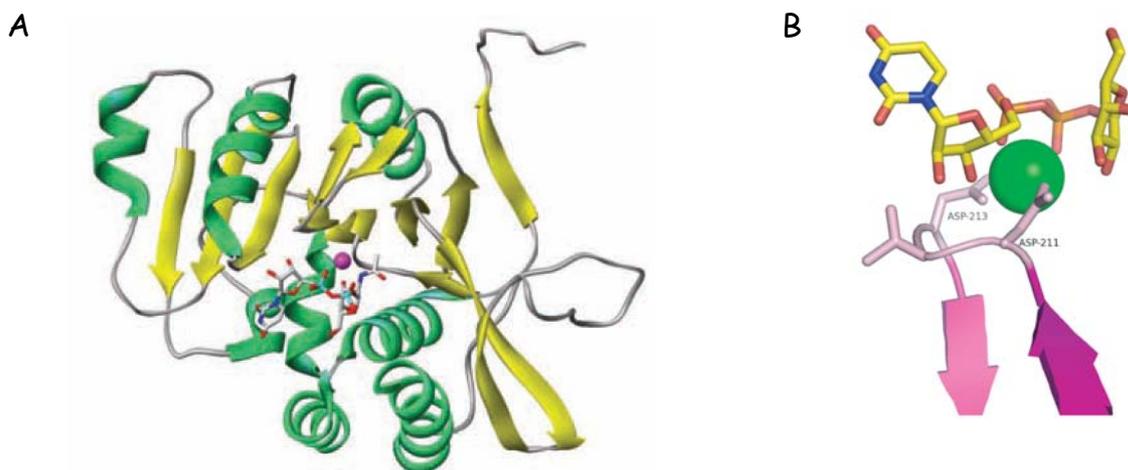
Les glycosyltransférases sont des enzymes capables de synthétiser les différentes structures oligosaccharidiques du vivant. Elles catalysent le transfert d'un monosaccharide activé par des groupements phosphate riches en énergie, sous forme de nucléosides mono- ou diphosphates (UDP-Gal, GDP-Man ou CMP-NeuAc) sur des substrats accepteurs de nature différente (saccharidique, protéique, lipidique...) selon la réaction suivante :



Ces enzymes sont dénommées d'après l'ose qu'elles transfèrent et d'après le type de liaisons qu'elles élaborent. On parlera de galactosyltransférases (GalTs), glucosyltransférases (GlcTs), sialyltransférases (STs)... Intervient également le type de liaison formée par le sucre ainsi que, pour les GalTs, l'anométrie de la liaison ( $\alpha$ 2-GalT,  $\alpha$ 3-GalT,  $\beta$ 3GalT,  $\beta$ 4GalT...). Depuis 1997, ces enzymes sont répertoriées dans la banque de données CAZy (<http://www.cazy.org/>) qui compte aujourd'hui environ 55000 séquences nucléotidiques réparties dans 92 familles désignées par GTx, basées sur leurs similarités de séquence en acides aminés (Cantarel *et al.* 2009).

### 1.5.2. Structures tridimensionnelles de types GT-A et GT-B

A ce jour, 90 structures cristallographiques ont été résolues, révélant deux types de repliement majoritaires, appelés GT-A et GT-B (Breton *et al.* 2006). Le repliement GT-A, décrit pour la première fois pour l'enzyme SpsA de *Bacillus subtilis* (Charnock and Davies 1999), est schématiquement décrit comme un sandwich reposant sur le motif structural de type Rossmann (**Figure 1.10. A**). La plupart des enzymes GT-A possèdent en commun un motif DxD qui coordonne un cation de type  $Mg^{2+}$  ou  $Mn^{2+}$  impliqué dans la fixation des ions phosphate du nucléotide-sucre (**Figure 1.10. B**). Ce motif est indispensable pour l'activité de l'enzyme.

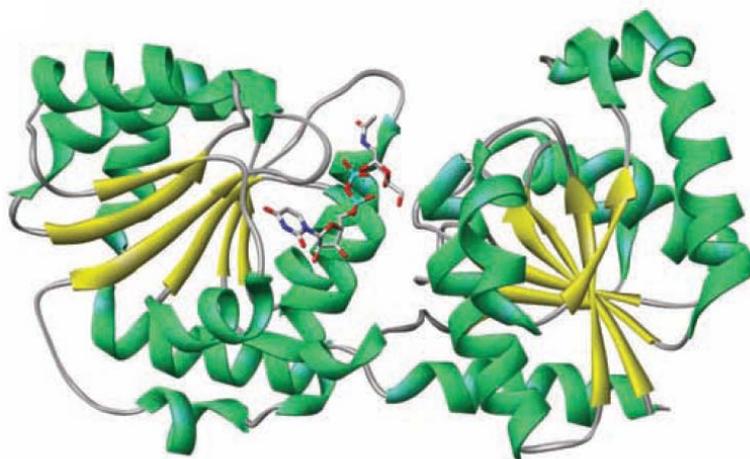


**Figure 1.10.** Représentation du repliement de type GT-A observé pour l'enzyme SpsA de *Bacillus subtilis* (PDB 1qgq) (A.) et de leur motif DxD en interaction avec le nucléotide-sucre via l'ion  $Mn^{2+}$  (B.).

Le deuxième type de repliement, appelé GT-B, a été identifié en 1994 pour la  $\beta$ -glucosyltransférase du bactériophage T4 (Vrieland *et al.* 1994). Ce repliement consiste en deux domaines indépendants de type Rossmann  $\alpha/\beta/\alpha$ , reliés par une zone flexible et terminés par une série de longues hélices  $\alpha$ . Le site catalytique se trouve à l'interface de ces deux domaines (**Figure 1.11.**). Une conservation structurale est observée pour le domaine C-terminal, domaine de liaison au nucléotide-sucre. Le domaine N-terminal est, quant à lui, plus variable et serait impliqué dans la liaison à l'accepteur.

Une troisième famille structurale, considérée comme variants du type de repliement GT-A, a été découverte à partir de la structure d'une sialyltransférase CstII de *Campylobacter jejuni* (Chiu *et al.* 2004). Elle présenterait des similitudes avec la famille des GT-A mais pas de motif DxD.

Récemment, des types de repliement complètement différents ont été observés pour des glycosyltransférases appartenant à deux familles (GT51 et GT66). Elles utilisent des sucres activés par des lipides phosphatés, à la place de nucléotide-sucres (Lovering *et al.* 2007 ; Yuan *et al.* 2007 ; Igura *et al.* 2008).

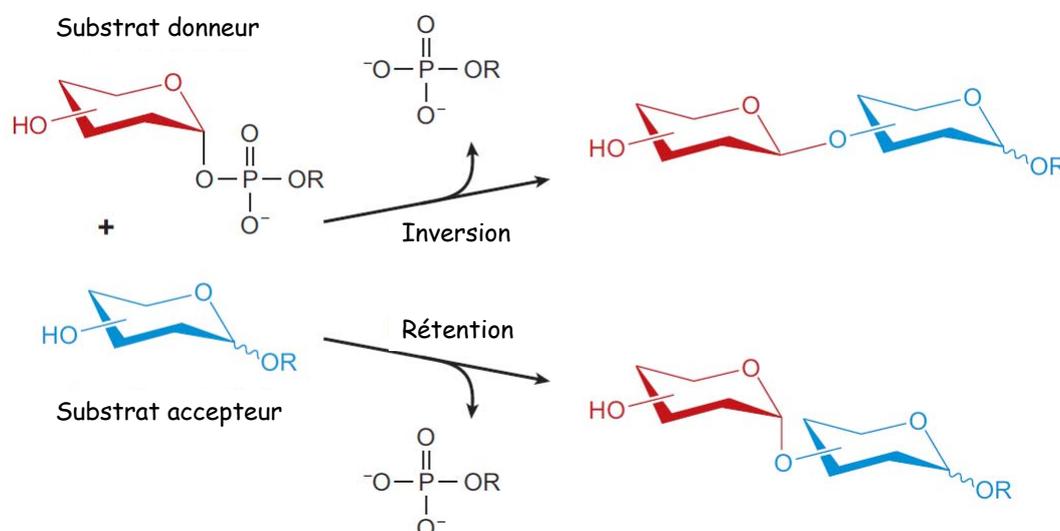


**Figure 1.11. Représentation du repliement de type GT-B observé chez la  $\beta$ -glucosyltransférase du bactériophage T4 (PDB 1jg7).**

### 1.5.3. Stéréochimie de la réaction

La réaction de transfert du sucre s'effectue selon deux mécanismes principaux qui tiennent compte de la stéréochimie du carbone anomérique du sucre lié au nucléotide et de ce même sucre lié à l'accepteur. La réaction de transfert s'accompagne soit d'une rétention, soit

d'une inversion de configuration de ce carbone anomérique (**Figure 1.12.**). Un type de repliement (GT-A ou GT-B) n'est pas strictement associé à un type de réaction (rétention ou inversion) (Lairson *et al.* 2008).



**Figure 1.12. Synthèse des liaisons glycosidiques par inversion ou rétention de configuration du carbone anomérique (Lairson *et al.* 2008).**

## 1.6. Modifications pathologiques de la glycosylation

Le nombre de gènes directement impliqués dans l'assemblage des glycanes est estimé entre 250 et 500, ce qui représente chez l'homme 1% à 2% du génome. Ces « glycoènes » codent pour plusieurs types de protéines différentes comme les glycosyltransférases et sulfotransférases qui agissent directement sur le sucre, les transporteurs de nucléotide-sucre et les synthases, mais aussi les glycosidases et les lectines qui, respectivement, clivent et reconnaissent les sucres. Tous ces glycoènes peuvent subir des modifications, à la fois qualitatives et quantitatives, à l'origine de pathologies parfois très sévères. Ces perturbations peuvent notamment être observées au cours de l'embryogenèse, du vieillissement, du processus inflammatoire et de certaines maladies auto-immunes et sont à l'origine de désordres très graves dont les maladies CDGs («congenital disorders of glycosylation») (Dennis *et al.* 2009).

La transformation d'une cellule saine en cellule cancéreuse s'accompagne également d'une perturbation de la glycosylation de surface. Les modifications les plus fréquentes sont

la perte d'expression ou à l'inverse l'expression excessive de certaines structures, l'expression de structures incomplètes ou tronquées et moins fréquemment, l'apparition de nouvelles structures, généralement identifiées au stade embryonnaire (Varki 1999). Ces modifications peuvent être associées à la formation de métastases, croissance d'une cellule tumorale à distance du site initial par diffusion sanguine ou lymphatique, généralement associée à un mauvais pronostic vital.

## **2. Modification de la glycosylation dans le processus cancéreux**

### **2.1. Les antigènes T et Tn**

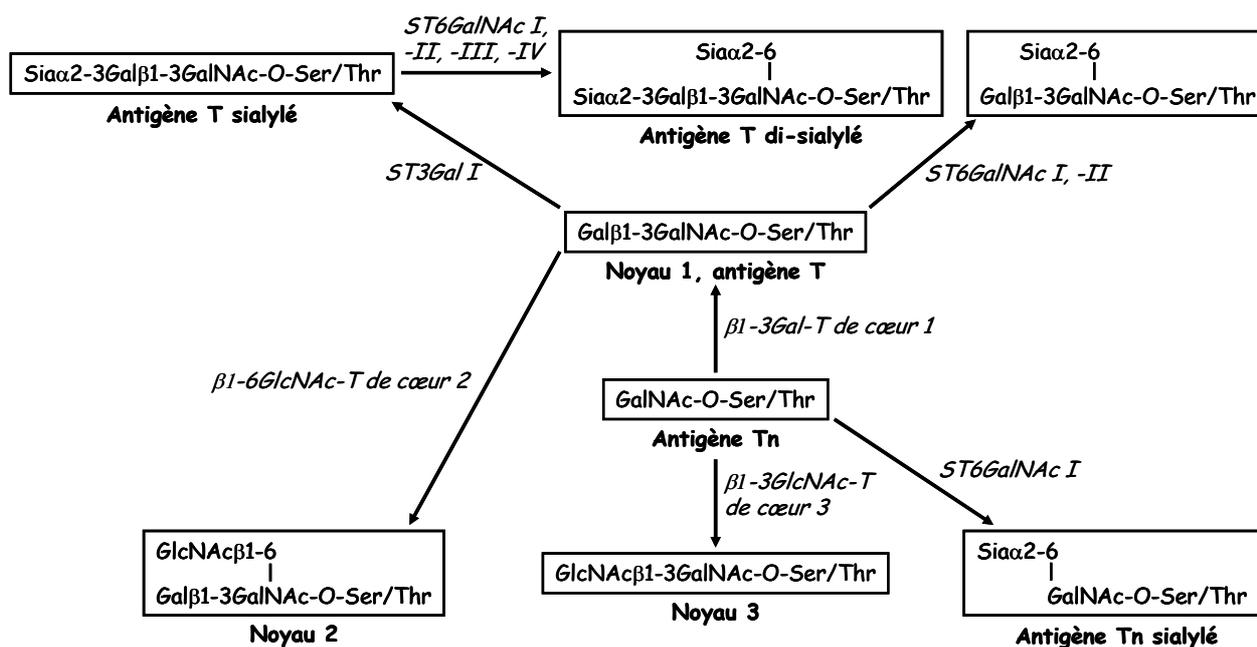
#### **2.1.1. Biosynthèse dans les différents types de cancers**

La *O*-glycosylation de type mucine débute par l'addition d'un résidu GalNAc lié en position  $\alpha$  sur le groupement hydroxyle d'une sérine ou d'une thréonine. Cette liaison est catalysée par une UDP-*N*- $\alpha$ -galactosamine : polypeptide *N*-acétylgalactosaminyltransférase (ppGalNAc-T) et forme l'antigène Tn (GalNAc $\alpha$ 1-Ser/Thr). Le noyau 1 (Gal $\beta$ 1-3GalNAc $\alpha$ 1-Ser/Thr), également dénommé antigène T ou TF pour Thomsen-Friedenreich, est formé par ajout d'un résidu galactoside sur l'antigène Tn par la  $\beta$ 1-3 galactosyltransférase de cœur 1 (**Figure 1.13.**) (Ju *et al.* 2002). Dans les épithelia normaux, ces structures recouvertes par différents types de groupements (acides sialiques, phosphates...) sont utilisées comme précurseurs pour former des *O*-glycanes plus complexes et branchés, tels que les épitopes des groupes sanguins (ABO et Lewis), caractéristiques des mucines. Ces structures non substituées sont, néanmoins, retrouvées durant l'embryogenèse et dans 90% des cancers humains (Springer 1984) tels que les cancers du sein, du colon, de la vessie, de l'utérus, de l'estomac ou de la peau et constituent donc d'importants marqueurs. Dans la majorité des cas, l'apparition de ces structures est associée à la progression du cancer, à l'augmentation de son agressivité et à la formation de métastases (Springer 1997).

Les raisons de ces modifications de *O*-glycosylation restent toujours floues à ce jour mais plusieurs hypothèses peuvent néanmoins tenter d'expliquer la prévalence de ces

antigènes. Une altération dans l'expression de plusieurs enzymes de biosynthèse des *O*-glycanes peut être à l'origine de ces modifications. Les pp-GalNAc-Ts sont regroupées à l'intérieur d'une famille de plus de vingt membres dont quinze ont déjà été caractérisés dans des cellules de mammifères (Ten Hagen *et al.* 2003). Deux enzymes appartenant à cette famille, les ppGalNAc-T3 et -T6 semblent être particulièrement perturbées durant le processus tumoral. Deux études différentes ont démontré qu'elles semblaient surexprimées dans des lignées de cancers du sein alors que leur expression était quasiment silencieuse dans des cellules saines. Ce phénomène pourrait être à l'origine de l'augmentation de l'antigène Tn à la surface des cellules cancéreuses, notamment sur des sites potentiels de *O*-glycosylation normalement inoccupés (Berois *et al.* 2006 ; Brooks *et al.* 2007).

Les enzymes intervenant dans la synthèse d'épitopes plus complexes sont également perturbées dans le cancer. Dans des cellules saines, l'antigène T est normalement rapidement modifié par action de la  $\beta$ 1-6GlcNAc-transférase de cœur 2 qui rajoute un résidu GlcNAc sur le GalNAc du noyau 1 formant ainsi le noyau 2 branché (**Figure 1.13.**).



**Figure 1.13. Représentation schématique de la biosynthèse des antigènes T et Tn et leurs homologues sialylés.**

Le nom des antigènes est écrit en gras et leur structure respective est encadrée. Le nom des enzymes responsables de leur biosynthèse est représenté en italique (Dall'Olio and Chiricolo 2001).

L'expression de cette enzyme semble fortement diminuée dans des tissus cancéreux du sein, exhibant ainsi l'antigène T (Brockhausen et al. 1995). La disponibilité, en quantité plus importante, du substrat de l'enzyme  $\alpha$ 1-3GalT de cœur 1, l'UDP-galactose (Yu 2007) ainsi que la surexpression de son transporteur pourraient également contribuer à l'augmentation de la biosynthèse et l'expression de l'antigène T (Kumamoto et al. 2001). Une légère acidification de l'appareil de Golgi qui favoriserait l'expression de l'antigène T, notamment dans les cellules de cancers du sein et colorectal, a été mise en évidence. D'une manière plus générale, elle pourrait perturber la glycosylation des cellules cancéreuses en modifiant la distribution des glycosyltransférases dans le Golgi (Rivinoja et al. 2006).

### 2.1.2. Les antigènes T et Tn sialylés

Les antigènes T et Tn peuvent également être sialylés pour former respectivement le sialyl-T (ST) et sialyl-Tn (STn) qui sont fortement exprimés à la surface de nombreux cancers (sein, ovaire, colon, estomac...). Ils sont généralement associés à un faible pronostic vital, accompagné de métastases et d'une diminution de la survie des patients (Itzkowitz *et al.* 1990 ; Leivonen *et al.* 2001). L'antigène Tn sialylé est responsable d'une mobilité cellulaire significativement améliorée, d'une adhésion cellulaire diminuée et d'une tumorigénicité augmentée, caractéristiques révélant le caractère agressif de la lignée (Julien *et al.* 2006).

La biosynthèse des antigènes T et Tn sialylés est également résumée dans la **figure 1.13**. La sialylation de l'antigène Tn est réalisée en  $\alpha$ 2-6 du résidu GalNAc par l'enzyme ST6GalNAc I. Dans les cancers du sein et de l'estomac, une très bonne corrélation est observée entre la quantité d'antigène STn exprimée à la surface des cellules et l'abondance d'ARNm codant pour l'enzyme ST6GalNAc I (Marcos *et al.* 2003 ; Sewell *et al.* 2006) alors que ces résultats ne sont pas vérifiés pour les cancers du colon où le niveau d'enzyme ST6GalNAc exprimé est comparable dans les lignées exprimant ou pas l'antigène STn (Brockhausen *et al.* 1998). Trois enzymes utilisent potentiellement l'antigène Tn comme substrat : les  $\beta$ 1-3GalT de cœur 1 et  $\beta$ 1-3GlcNAcT de cœur 3 qui forment respectivement les antigènes T (ou noyau 1) et le noyau 3 et la ST6GalNAc-I qui synthétise le STn et entrent donc en compétition. Pour les cancers du colon et du sein notamment, une faible activité voire une déficience des enzymes  $\beta$ 1-3GalT de cœur 1 et  $\beta$ 1-3GlcNAc-T de cœur 3 explique l'abondance de l'antigène STn (Brockhausen *et al.* 1998).

L'antigène T est sialylé en  $\alpha$ 2-3 sur le résidu Gal principalement par l'action de l'enzyme ST3Gal-I. L'expression de cette enzyme dans différents types de cancers (dont le sein, l'estomac ou la vessie) est corrélée au stade d'avancement de la tumeur et semble être associée au pronostic de la récurrence du cancer (Videira *et al.* 2009). Elle est donc un bon marqueur de la malignité de certains types cancéreux. De manière plus marginale, l'antigène T peut également être mono-sialylé en  $\alpha$ 2-6 sur le résidu GalNAc par deux enzymes potentielles : ST6GalNAc-I et -II qui, comme l'enzyme  $\beta$ 1-6GlcNAc de cœur 2, ont une action limitée dans le cancer. De plus, l'antigène T sialylé peut être lui même le substrat d'autres sialyltransférases pour former un épitope di-sialylé (Dall'Olio and Chiricolo 2001).

La *O*-déacétylation des acides sialiques portés par les mucines des cancers colorectaux semble également expliquer la prévalence de ces épitopes sialylés. En effet, environ 50% des acides sialiques présents à la surface des mucines du colon sont naturellement *O*-acétylés en position 7 et 9 (Rogers *et al.* 1978) et sont alors incapables d'être détectés par certains anticorps. La transformation tumorale s'accompagne généralement d'une forte diminution de cette *O*-acétylation, y compris à des stades précoces, exhibant ainsi ces structures sialylées, qui semblent alors plus abondantes (Corfield *et al.* 1999).

### 2.1.3. Immunité et antigénicité de ces épitopes

Des anticorps anti-T et anti-Tn sont naturellement présents, à un taux constant, chez les patients sains. Ils sont principalement induits par la présence de bactéries dans la flore intestinale durant la petite enfance (Springer 1997). Les antigènes T et Tn étant abondamment exprimés à la surface des cellules cancéreuses, le niveau d'anticorps anti-T et anti-Tn, dans le sérum de patients malades a toujours suscité beaucoup d'intérêt pour les chercheurs et semblent augmenter chez les patients malades. Une étude a démontré que, dans 77% des cas, des carcinomes pouvaient être détectés en titrant la quantité d'anticorps anti-T et anti-Tn des mois voire des années avant que la première biopsie ne se soit révélée positive. De plus, aucun patient, pour lequel un cancer était suspecté et possédant un titre d'anticorps « normal » n'a développé de maladie durant une période moyenne d'observation d'environ cinq ans. La titration de ces anticorps est donc un important outil pronostic (Springer 1997).

Etant masqués par d'autres unités glycaniques dans les tissus sains, la présence des antigènes T et Tn dans les carcinomes est reconnue comme non soi par le système immunitaire de l'hôte. Ces réponses auto-immunes sont cependant insuffisantes pour protéger

---

les patients. L'immunothérapie avec des vaccins anti-cancers dirigés contre des épitopes glycaniques est donc devenue une forme attractive de thérapie, avec des essais très prometteurs chez l'humain (Springer *et al.* 1995 ; Slovin *et al.* 2007).

## **2.2. Les antigènes des groupes sanguins**

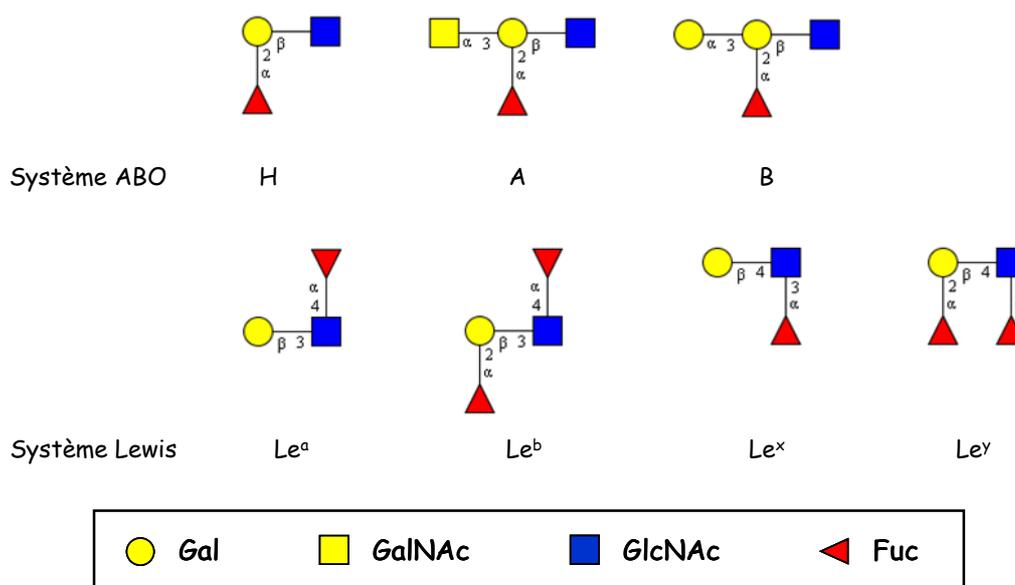
### **2.2.1. Définition des groupes sanguins ABO et Lewis**

Le système ABO a été découvert en 1900 par Karl Landsteiner, de l'université de Vienne, après avoir mélangé les sérums et hématies de six de ses collègues et observé les agglutinations éventuelles (Landsteiner 1900). Une année plus tard, il met en évidence que seulement deux antigènes A et B et deux anticorps anti-A et anti-B permettent d'expliquer l'existence des trois groupes sanguins. Les cellules peuvent présenter soit l'antigène A, soit le B soit aucun des deux. Il démontre également qu'un individu ne présente pas d'anticorps dirigés contre les épitopes présents à la surface de ses propres hématies et que les deux anticorps sont présents dans le sérum de patients de groupe O (Landsteiner 1901). Ces travaux lui ont valu le prix Nobel de médecine en 1930. Le groupe sanguin AB, pour lequel les deux antigènes A et B sont présents à la surface des hématies et dont le sérum ne contient pas d'anticorps, a été découvert en 1902 par Decastello et Sturli (von Decastello 1902), et, en 1924, Berstein démontre que la transmission des groupes sanguins est héréditaire et suit les lois de Mendel (Berstein 1924). En 1948, Morgan et Watkins renomment l'antigène O, H et le définissent comme substance primaire commune à la grande majorité des érythrocytes quelque soit leur groupe sanguin (Morgan and Watkins 1948). Elle est donc précurseur des substances A et B et les cellules O sont en réalité des cellules portant la substance H non recouverte. La découverte du phénotype Bombay (individus qui ne possèdent aucun des antigènes A, B et H à la surface de leurs hématies), en 1953, suggère, pour la première fois, que le caractère H est génétiquement indépendant du système ABO (Bhende *et al.* 1952). Il faudra attendre 1990 et le clonage du gène H (FUT1) pour définitivement définir l'indépendance des gènes H et ABO (Larsen *et al.* 1990).

En 1940, Landsteiner et Harte établissent que les substances actives sont composées de sucres dont le GlcNAc, Gal et Fuc (le GalNAc étant découvert huit ans plus tard (Aminoff and Morgan 1948)) et de protéines (Landsteiner and Harte 1940). Mais l'implication des

sucres comme déterminants antigéniques n'a été démontré que, dix ans plus tard, avec l'inhibition de l'agglutination des globules O par la lectine d'anguille *Anguilla anguilla* agglutinin (AAA) après ajout de  $\alpha$ -Me-L-fucoside (Watkins and Morgan 1952). D'autres expériences d'inhibition ont également permis de révéler la spécificité des groupes sanguins A et B pour le GalNAc et le Gal respectivement (**Figure 1.14.**) (Morgan and Watkins 1953 ; Watkins and Morgan 1955). La découverte, en 1953, de la structure de l'ADN et le décryptage du langage génétique (un gène, gardien de l'information génétique, est traduit en protéine) révèlent que ces épitopes glycaniques sont en réalité un produit secondaire des gènes. Ils sont synthétisés par des enzymes spécifiques, produit direct de l'information génétique (Watson and Crick 1953).

Un autre système de groupes sanguins à épitopes glycaniques, appelé Lewis, est également découvert à la fin des années 1940 (**Figure 1.14.**) (Mourant 1946) (revue par (Watkins 2001)).



**Figure 1.14. Représentation schématique des épitopes des groupes sanguins des systèmes ABO et Lewis.**

Les carrés bleus et jaunes schématisent respectivement les résidus GlcNAc et GalNAc, les ronds jaunes schématisent les résidus Gal, les triangles rouges schématisent les résidus Fuc. Les images ont été réalisées sur la plateforme GlycoWork Bench (Ceroni *et al.* 2007).

Chez l'humain, plus d'une trentaine de groupes sanguins différents ont été identifiés, tels que les systèmes Rhésus, Kell, Duffy... Les déterminants antigéniques ne sont pas nécessairement portés par des résidus glycaniques. Par exemple, la présence de la protéine D

circulante (système Rhésus), suite à une transfusion par exemple, est responsable de l'antigénicité chez les patients Rh négatifs.

La localisation de ces déterminants peut également varier d'un système à l'autre. Les antigènes du système ABO sont localisés à la surface de nombreux tissus. Ce système est appelé tissulaire à la différence du système Rhésus pour qui, la protéine D, se trouve uniquement à la surface des hématies. Ce système est dit érythrocytaire.

La diversité des groupes sanguins existants, d'un individu à l'autre, pourrait être une arme, élaborée pour lutter contre des infections bactériennes ou virales qui s'attaqueraient uniquement à un groupe d'individus présentant un groupe sanguin particulier. Ces considérations restent cependant toujours hypothétiques à ce jour.

### 2.2.2. Biosynthèse des antigènes des groupes sanguins

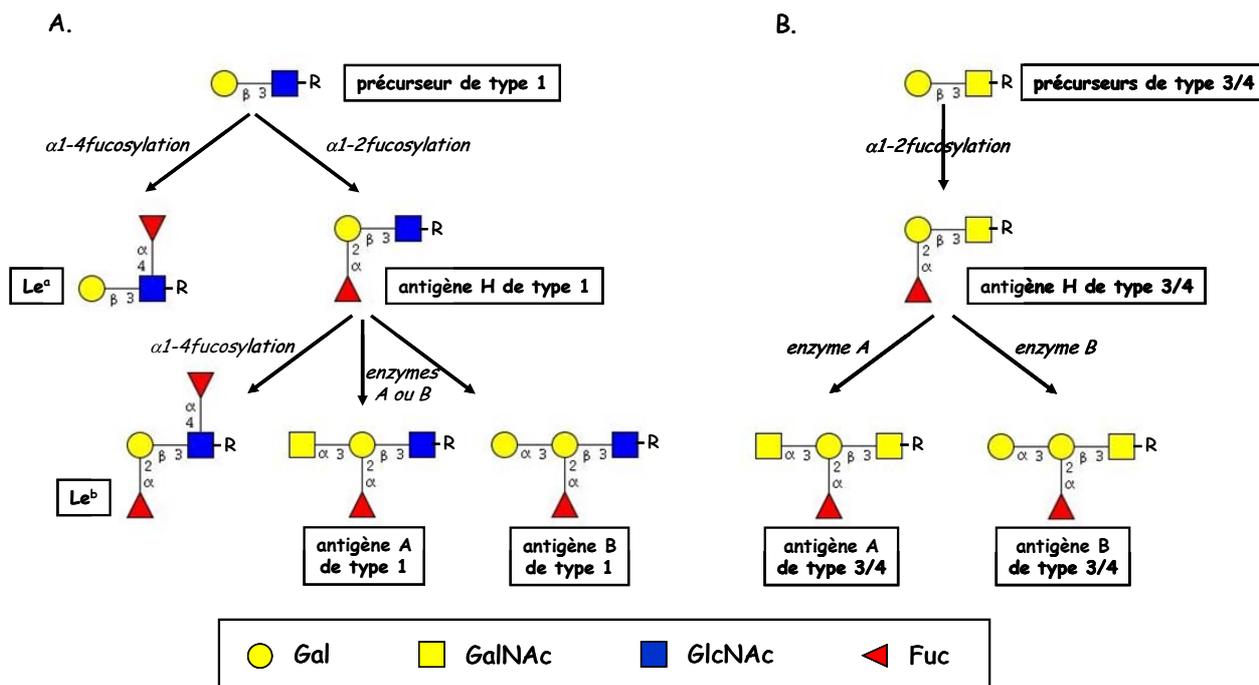
La biosynthèse des antigènes des groupes sanguins peut avoir lieu sur quatre précurseurs différents, qui possèdent tous un résidu  $\beta$ -galactoside terminal :

- type 1 : Gal $\beta$ 1-3GlcNAc $\beta$ 1-R
- type 2 : Gal $\beta$ 1-4GlcNAc $\beta$ 1-R
- type 3 : Gal $\beta$ 1-3GalNAc $\alpha$ 1-R
- type 4 : Gal $\beta$ 1-3GalNAc $\beta$ 1-R

Les précurseurs de type 1 et 2 sont aussi bien retrouvés à la surface de protéines *N*- ou *O*-glycosylées, qu'au niveau des glycolipides de la série lacto- (désignés par la lettre R). La structure de type 3 est exclusivement rencontrée sur les *O*-glycanes et celle de type 4 à la surface des glycolipides des séries globo- et ganglio-. L'addition sur ces précurseurs d'un résidu fucoside lié en  $\alpha$ 1-2 par FUT1 et FUT2 forme l'antigène H, qui correspond au groupe sanguin O. Une fois l'antigène H formé, la synthèse peut se prolonger par ajout de, soit un résidu GalNAc, soit un résidu Gal liés en  $\alpha$ 1-3 sur le Gal de l'antigène H pour former respectivement les antigènes A et B. Cette étape est catalysée par deux enzymes distinctes, une *N*-acétylgalactosaminyltransférase et une galactosyltransférase, codées par deux allèles différents situés sur le même locus. L'addition d'un résidu fucoside lié en  $\alpha$ 1-4 sur le précurseur de type 1 et l'antigène H de type 1 forme respectivement les antigènes dénommés Lewis a ( $Le^a$ ) et Lewis b ( $Le^b$ ). Lorsque le fucose, lié en  $\alpha$ 1-3, est ajouté sur le précurseur de type 2 ou sur l'antigène H de type 2, les antigènes Lewis x ( $Le^x$ ) et Lewis y ( $Le^y$ ) sont formés. Six fucosyltransférases sont capables de catalyser l'addition du fucose en position  $\alpha$ 1-3 ou  $\alpha$ 1-

4 chez l'humain. Les enzymes FUT3 et FUT5 peuvent utiliser comme substrat les précurseurs de type 1 et 2 alors que FUT4, FUT6, FUT7 et FUT9 catalyse l'addition du fucose exclusivement sur le précurseur de type 2 (**Figure 1.15.**) (Le Pendu *et al.* 2001).

Parallèlement aux antigènes T et Tn, l'expression des épitopes des groupes sanguins à la surface des cellules cancéreuses est fortement perturbée durant le processus tumoral. Des modifications divergentes peuvent être constatées suivant le type de cancer et sa localisation mais certaines caractéristiques communes peuvent néanmoins être dégagées.



**Figure 1.15.** Voies de biosynthèse des antigènes des groupes sanguins ABH et Lewis basées sur **A.** le précurseur de type 1 et **B.** les précurseurs de type 3/4.

Le nom de chaque antigène est encadré. **A.** Le même schéma peut être représenté à partir du précurseur de type 2. Dans ce cas,  $Le^x$  remplacera le  $Le^a$  et  $Le^y$  remplacera le  $Le^b$ . **B.** À la différence des précurseurs de type 1 et 2, la fucosylation de l'avant dernier monosaccharide (GalNAc) n'a jamais été décrite. Les carrés bleus schématisent les résidus GlcNAc, les ronds jaunes schématisent les résidus Gal, les triangles rouges schématisent les résidus Fuc. Les images ont été réalisées sur la plateforme GlycoWork Bench (Ceroni *et al.* 2007).

### 2.2.3. Les antigènes du système ABO dans le processus cancéreux

De nombreuses études ont révélé une réduction voire une perte d'expression des antigènes A et B dans la majorité des carcinomes tels que les cancers du sein, des poumons, du colon, de la prostate, de la vessie ou de l'estomac. Cette perte s'accompagne généralement d'une augmentation de l'expression de l'antigène H notamment dans les cancers de la

prostate, du sein, ou de la bouche alors que dans les cancers du poumon ou de l'ovaire, une expression de l'antigène Le<sup>y</sup> est observée suite à l'ajout d'un résidu fucoside sur l'antigène H. Il est important de noter que certains types de carcinomes ne suivent pas ce schéma classique, comme par exemple ceux de la thyroïde ou du foie (Le Pendu *et al.* 2001). La présence des antigènes H et Le<sup>y</sup> serait le résultat d'une synthèse bloquée et non l'action d'enzymes de dégradation des glycanes appelées glycosidases. En effet, plusieurs études ont montré que l'activité des glycosyltransférases A et B, responsables, respectivement, de la synthèse des épitopes de groupes sanguins A et B était fortement réduite durant le processus cancéreux (Orntoft and Wolf 1988). Particulièrement bien référencée pour les cancers de la vessie, la perte d'hétérozygotie des allèles codant pour les glycosyltransférases au niveau du locus 9q34 pourrait être à l'origine de ce phénomène. Ceci consiste en une disparition d'un segment chromosomique, soit par délétion soit par un mécanisme de translocation déséquilibré souvent associée à la perte d'un des deux allèles parentaux (Orlow *et al.* 1998). Ces résultats ont été néanmoins contredit (Meldgaard *et al.* 1995). Une deuxième hypothèse avancée par les chercheurs pour expliquer cette diminution dans l'expression des antigènes A et B est une méthylation excessive de la région promotrice du gène codant pour les glycosyltransférases A et B. Cette modification pourrait inhiber la transcription du gène et jouerait un rôle de régulation dans l'expression des gènes ABO (Kominato *et al.* 1999).

La perte ou la réduction des épitopes A et B dans les cancers humains corrélient généralement avec un degré de malignité élevé et un potentiel métastatique important impliquant un taux de survie des patients faible (Matsumoto *et al.* 1993 ; Hakomori 1999). Des résultats contradictoires ont démontré que la présence de ces antigènes, dans les carcinomes colorectaux, et non leur diminution, serait corrélée à un mauvais pronostic vital (Nakagoe *et al.* 2000). Ils expliquent ces différences par des divergences techniques au niveau des anticorps utilisés ou du mode de conservation des coupes cellulaires. Des résultats similaires ont également été mis en avant pour des cancers de l'ovaire (Welshinger *et al.* 1996) et du sein (Lee *et al.* 1985) où là encore, la présence ou l'absence d'antigènes des groupes sanguins n'était pas corrélée avec la survie des patients.

Un second type d'anomalie a également pu être identifié. Après acquisition de son caractère cancéreux, la cellule modifie sa réactivité vis-à-vis des antigènes ABO. Par exemple, un tissu cancéreux va acquérir une réactivité dite « A », alors que les cellules proviennent d'un patient B ou O (Le Pendu *et al.* 2001). Beaucoup plus marginal, ce phénomène ne sera pas davantage décrit dans ce manuscrit.

#### 2.2.4. Les antigènes du système Lewis dans le processus cancéreux

Durant la formation cancéreuse, l'expression des antigènes Lewis est fortement perturbée mais de manière cancer-dépendant. Par exemple, la réactivité de l'antigène Le<sup>x</sup> est fréquemment augmentée dans les carcinomes du pancréas, des poumons, du rein, de la vessie, du sein alors qu'elle diminue dans ceux de l'endomètre. Concernant le Le<sup>a</sup>, sa fréquence augmente dans les cancers de l'estomac, du pancréas, du poumon, de l'endomètre et de la thyroïde alors qu'elle diminue dans ceux du colon, de la prostate et du sein. Pour le Le<sup>b</sup>, il est surexprimé dans les cancers du colon, de la vessie, de l'endomètre alors qu'il est réprimé dans ceux de l'estomac, du pancréas, des poumons ou du sein. Et enfin, l'expression du Le<sup>y</sup> augmente fortement dans presque tous les types de carcinomes à l'exception de ceux du pancréas et de l'endomètre (Le Pendu *et al.* 2001). Ces exemples permettent de mettre en évidence la grande hétérogénéité concernant la surexpression ou l'inhibition de chaque antigène Lewis dans les différents types de cancers rencontrés.

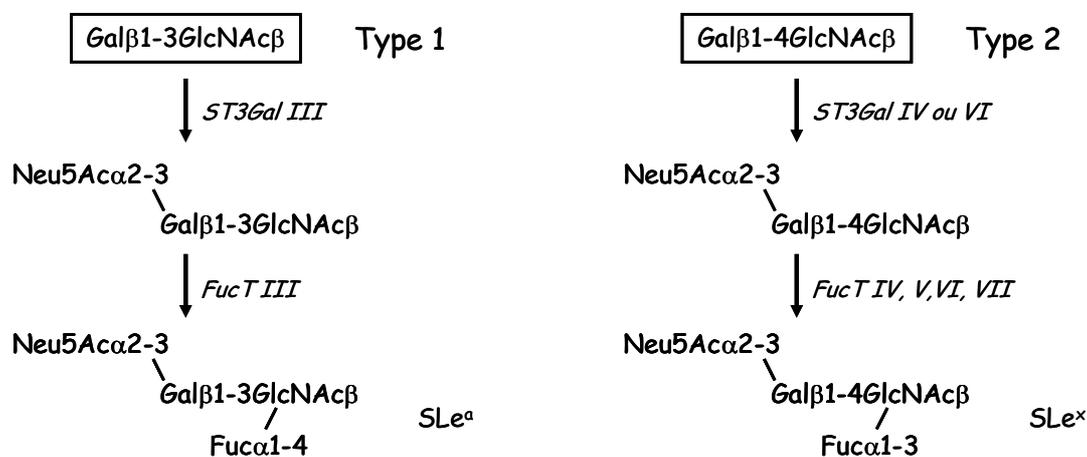
Ces modifications sont dues à une perturbation de l'activité des fucosyltransférases. Par exemple, FUT1 et FUT2 présentent une activité  $\alpha$ 1-2fucosyltransférase plus importante dans les cancers colorectaux favorisant ainsi l'expression des antigènes Le<sup>b</sup> et Le<sup>y</sup> au détriment du Le<sup>a</sup> (Yazawa *et al.* 1993). De plus, FUT3, bien qu'elle possède à l'origine une activité  $\alpha$ 3/4 fucosyltransférase, développerait, dans des conditions malignes, une activité  $\alpha$ 1-2fucosyltransférase inhabituelle (Nakamura *et al.* 1997). Parallèlement à la situation précédente, une augmentation de l'activité  $\alpha$ 1-2fucosyltransférase a également été décrite dans les carcinomes de l'endomètre (Wang *et al.* 1995) alors qu'à l'inverse, une diminution de cette activité a été reportée dans ceux du pancréas (Mas *et al.* 1998). Cette fucosylation aberrante semble influencer le comportement des cellules puisqu'elle participerait à la progression tumorale notamment dans le cancer du colon (Sun *et al.* 1995) et induirait une résistance des cellules de cancer du colon de rat à l'apoptose (Goupille *et al.* 2000).

#### 2.2.5. Leurs homologues sialylés dans le processus cancéreux

L'expression des antigènes Lewis sialylés, en particulier les sialyl-Le<sup>a</sup> (sLe<sup>a</sup>) et sialyl-Le<sup>x</sup> (sLe<sup>x</sup>), est fortement perturbée dans un grand nombre de cancers tels que le sein, le poumon, le colon, le pancréas. Cette augmentation globale de l'expression de ces antigènes est généralement impliquée dans la formation de métastases, réduisant ainsi, de manière

significative, le taux de survie des patients (Narita *et al.* 1993). Les sLe<sup>a</sup> et sLe<sup>x</sup> sont particulièrement importants car ils sont des substrats potentiels des sélectines, molécules impliquées dans les phénomènes d'adhésion cellulaire durant l'inflammation et la migration des lymphocytes. En modifiant leur expression, la cellule cancéreuse perturbe ses capacités d'adhésion et de migration cellulaire, ce qui favorise le développement de métastases et développe parallèlement des mécanismes lui permettant d'échapper au système immunitaire (Ohyama *et al.* 1999).

Cette surexpression est corrélée avec la modification de l'expression des gènes codant pour les glycosyltransférases impliquées dans leur biosynthèse, particulièrement sialyltransférases et fucosyltransférases (**Figure 1.16.**).



**Figure 1.16. Schéma de biosynthèse des antigènes Lewis a et x sialylés dans les cancers du sein.**

Les molécules encadrées représentent les précurseurs et les noms en italique se réfèrent aux glycosyltransférases impliquées. Dans les deux cas, la biosynthèse débute par une étape d' $\alpha$ 2-3 sialylation sur le résidu Gal du précurseur de types 1 ou 2 puis se poursuit par une  $\alpha$ 1-3/4 fucosylation sur le résidu GlcNAc.

Cependant, suivant la localisation du cancer, les enzymes perturbées sont différentes. Par exemple, pour les adénocarcinomes mammaires, l'augmentation de sLe<sup>x</sup> est corrélée à l'augmentation des ARNm de la fucosyltransférase FUT6 (Matsuura *et al.* 1998) alors que, pour les cancers du poumon, les enzymes ST3Gal III et FUT7 semblent impliquées dans cette prévalence (Ogawa *et al.* 1997). Concernant les cancers du colon, une première étude met en relation l'expression de sLe<sup>a</sup> avec l'expression de  $\alpha$ 2-3sialyltransférases actives sur le précurseur de type 1 (Akamatsu *et al.* 1996) alors que des conclusions inverses ont été tirées, dans une autre étude, impliquant plutôt l'action d' $\alpha$ 1-4 fucosyltransférases (Yamada *et al.*

1997). De plus, la quantité d'ARNm d'une même enzyme peut également varier d'une lignée cellulaire à l'autre pour un même type de cancer (Ito *et al.* 1997). Une hétérogénéité est constatée dans l'expression des glycosyltransférases et la régulation de l'expression des antigènes sialylés (Carvalho *et al.* 2010).

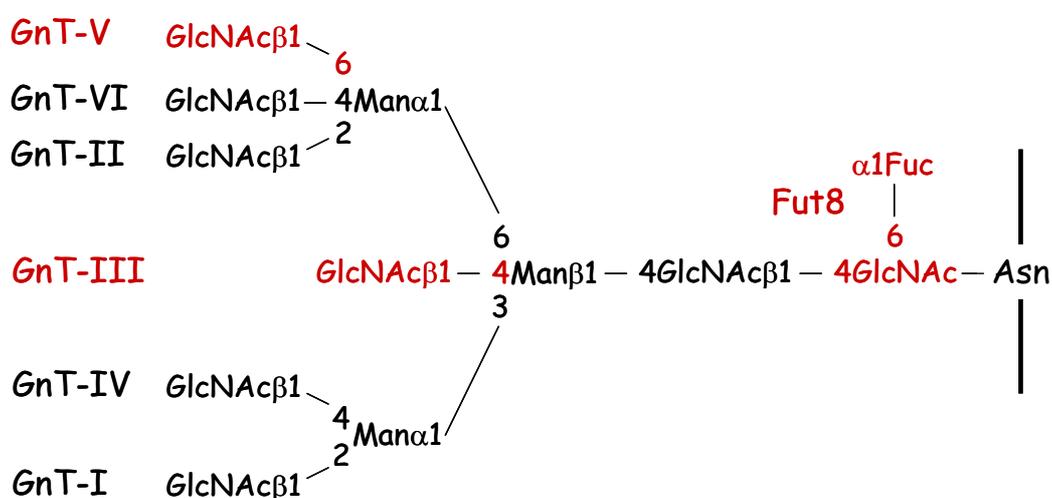
## 2.3. Modification des *N*-glycanes

### 2.3.1. La structure $\beta$ 1-6GlcNAc associée aux métastases

Six *N*-acétylglucosaminyltransférases, désignées par les lettres GnT-I à -VI, sont impliquées dans la biosynthèse de la structure centrale des *N*-glycanes de type complexe (Rademacher *et al.* 1988). Les enzymes *N*-acétylglucosamine- $\alpha$ -mannoside- $\beta$ 2-*N*-acétylglucosaminyltransférases I et II (GnT-I et GnT-II) transfèrent, via une liaison  $\beta$ 1-2, un résidu GlcNAc sur les résidus  $\alpha$ -mannoside liés respectivement en  $\alpha$ 1-3 et  $\alpha$ 1-6 (Harpaz and Schachter 1980). La GnT-III, originellement identifiée dans la membrane de l'oviducte de la poule, catalyse l'addition d'un GlcNAc via un lien  $\beta$ 1-4 sur un  $\beta$ -mannose du cœur trimannosylé (Narasimhan 1982). La GnT-V transfère un résidu GlcNAc sur le groupement hydroxyle du carbone 6 d'un résidu mannoside lié en  $\alpha$  appartenant à l'accepteur minimal GlcNAc $\beta$ 1,2Man $\alpha$ 1,6Man $\beta$ 1 (Cummings *et al.* 1982). Les enzymes GnT-IV et GnT-VI transfèrent, via une liaison  $\beta$ 1-4, un résidu *N*-acétylglucosamine (GlcNAc) sur les résidus Man $\alpha$ 1-3Man et Man $\alpha$ 1-6Man du cœur mannosylé (Gleeson and Schachter 1983) (**Figure 1.17**).

GnT-III et GnT-V sont des enzymes de premier plan dans la progression cancéreuse et la formation des métastases. Depuis plusieurs dizaines d'années, l'activité de l'enzyme GnT-V et donc la présence, de manière abondante, de résidus  $\beta$ 1-6GlcNAc à la surface des glycoprotéines, est directement reliée au caractère fortement métastatique de nombreuses lignées cellulaires animales et humaines (fibroblastes, carcinomes mammaires murins, lymphomes, cellules cancéreuses humaines du colon) (Dennis *et al.* 1987). L'utilisation de la swainsonine, inhibiteur de l' $\alpha$ -mannosidase II qui bloque la voie de biosynthèse des *N*-glycanes avant l'élongation des branchements  $\beta$ 1-6GlcNAc, inhibe la colonisation des organes par des cellules cancéreuses (Dennis 1986 ; Humphries *et al.* 1986). Au niveau de l'organisme entier, les métastases observées dans des souris transgéniques pour lesquelles le

gène codant pour GnT-V a été supprimé, sont très fortement diminuées (Granovsky *et al.* 2000). De plus, cette variation de glycosylation entraîne des modifications importantes au niveau de la cellule tel que la perte de l'inhibition de contact ou la diminution de l'adhésion au substrat (sur fibronectine ou collagène), caractéristiques généralement observées pour des cellules métastatiques (Demetriou *et al.* 1995). Malheureusement, au niveau protéique, la découverte des cibles de GnT-V reste très difficile. Ito *et al.* ont néanmoins pu mettre en évidence que la matriptase, glycoprotéine importante dans la migration cellulaire et fortement associée à l'invasion cancéreuse, était plus résistante à l'auto-digestion et la protéolyse après glycosylation par GnT-V (Ito *et al.* 2006).

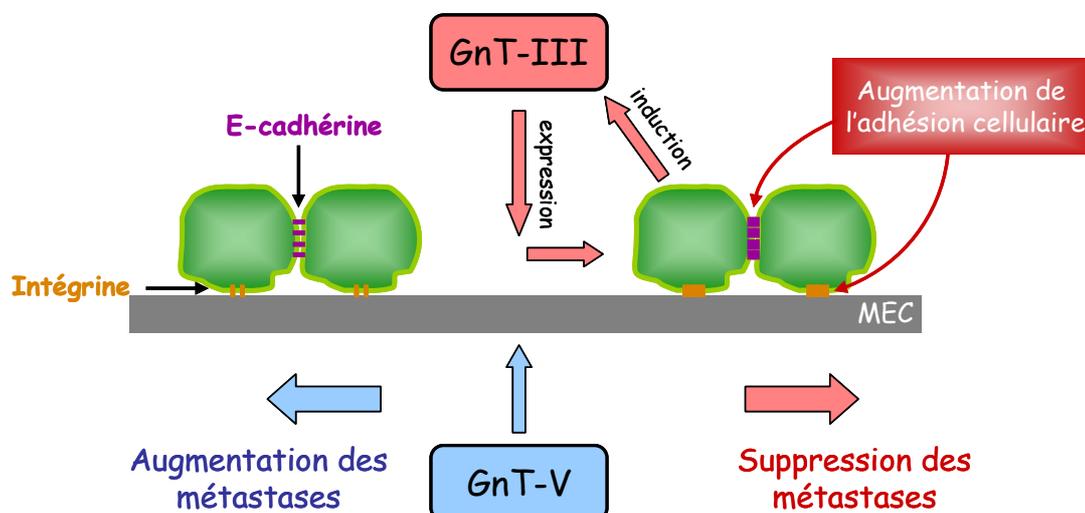


**Figure 1.17. N-acétylglucosaminyltransférases impliquées dans la biosynthèse des N-glycans de type complexe.**

GnT, N-acétylglucosaminyltransférase ; Man, Mannose ; GlcNAc, N-acétylglucosamine ; Fuc, Fuco ; Asn, Asparagine. Ce schéma est tiré de (Taniguchi *et al.* 1999).

A l'inverse, l'enzyme GnT-III semble inhiber la formation des métastases par compétition pour le substrat avec GnT-V. L'ajout d'un résidu GlcNAc en  $\beta 1-4$  sur le  $\beta$ -mannose du cœur trimannosylé par GnT-III entraîne des modifications conformationnelles qui rendent le sucre néo-formé inaccessible comme substrat pour GnT-V (Schachter 1986). La quantité de structures  $\beta 1-6\text{GlcNAc}$  est alors fortement diminuée et le nombre de métastases réduit (Yoshimura *et al.* 1995). GnT-III possède également une activité sur les molécules d'adhésion cellulaires, telles que les intégrines ou les E-cadhérines (Zhao *et al.* 2006). Ces dernières vont réguler, de manière positive, la transcription du gène codant pour GnT-III, qui va alors, de manière rétroactive, glycosylée les E-cadhérines, favorisant ainsi leur résistance à

la protéolyse et leur accumulation dans les contacts cellulaires. Ceci favoriserait l'agrégation des cellules tumorales les unes avec les autres et préviendrait ainsi le détachement de colonies métastatiques (Pinho *et al.* 2009) (**Figure 1.18.**).



**Figure 1.18.** Schéma résumant les rôles antagonistes des enzymes GnT-III et GnT-V.

L'augmentation de l'expression de GnT-V dans les cellules épithéliales a pour conséquence une perte de l'inhibition de contact et une augmentation de la motilité cellulaire. En contrecarrant l'activité de GnT-V, la surexpression de GnT-III provoque une augmentation de l'adhésion cellulaire via les E-cadhérines et intégrines, réduisant ainsi fortement les métastases. GnT, *N*-acétylglucosaminyltransférase ; MEC, Milieu ExtraCellulaire. Ce schéma est inspiré de (Zhao *et al.* 2008).

### 2.3.2. La fucosylation des *N*-glycanes

La fucosylation des glycoconjugués est également une altération majeure chez les patients atteints de cancers (Tatsumura *et al.* 1977). *N*-acétyl- $\beta$ -glucosaminide  $\alpha$ 1-6 fucosyltransférase ou FUT8 (**Figure 1.17.**) est indispensable dans la prolifération et la différenciation cellulaire puisque 70% des souris mutantes qui ne possèdent plus le gène *fut8* meurent trois jours après la naissance et les survivantes présentent d'importants retards de croissance (Wang *et al.* 2005). De nombreuses études ont révélé une augmentation du niveau d'expression de FUT8 dans différentes lignées cellulaires cancéreuses, notamment celles du poumon, du colon, de l'estomac (Miyoshi *et al.* 1997 ; Muinelo-Romay *et al.* 2008), mais aussi des ovaires (Takahashi *et al.* 2000) ou de la thyroïde (Ito *et al.* 2003). L' $\alpha$ -fétoprotéine (AFP) est l'exemple le plus documenté de glycoconjugués modifiés par fucosylation lors d'un processus cancéreux. Une augmentation de l'activité de FUT8 et, par conséquent, une

surexpression de résidus fucoside sur l'unique site de *N*-glycosylation (Asn 232) de l'AFP, au niveau des tissus et dans le sérum, corrobore avec un mauvais diagnostic pour des patients atteints de carcinomes hépatocellulaires et d'autres pathologies chroniques du foie (Khien *et al.* 2001), y compris suite à l'ablation de la tumeur par chirurgie (Hashimoto *et al.* 2004).

Parallèlement à GnT-III, FUT8 possède également une action sur les molécules d'adhésion en augmentant, de manière significative, l'expression des E-cadhérines, notamment dans les cancers colorectaux (Osumi *et al.* 2009). De plus, l' $\alpha$ 1-6 fucosylation régulerait le turnover des E-cadhérines, favorisant ainsi l'adhésion cellule-cellule *in vivo* (Takahashi *et al.* 2009).

Ces structures sont fortement modifiées durant le processus tumoral, engendrant de lourdes conséquences pathologiques, notamment sur la croissance et la survie des cellules, leur habilité à coloniser les tissus environnants et à métastaser ou sur leurs interactions avec les cellules du système immunitaire (Brockhausen 2006).

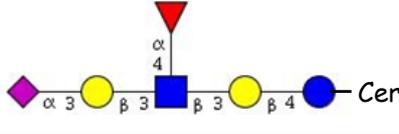
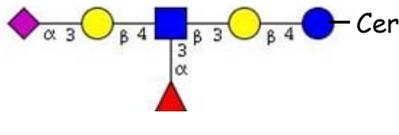
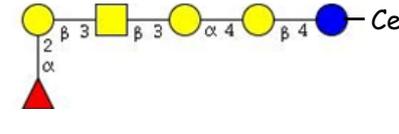
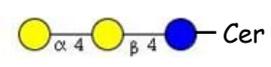
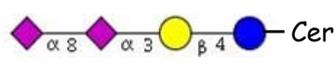
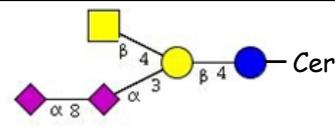
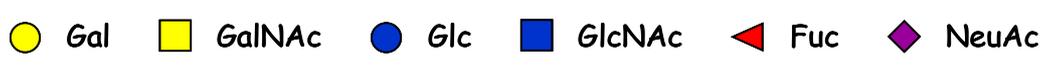
## **2.4. Modifications des autres types de glycoconjugués**

### **2.4.1. Glycolipides et cancer**

L'expression des glycosphingolipides (gangliosides, globosides et lactosides) est également fortement perturbée durant la transformation tumorale. De nouvelles structures, normalement uniquement exprimées dans les stades embryonnaires précoces, apparaissent (**Tableau 1.1.**).

Les principaux exemples sont les gangliosides GD2 et GD3 dans les mélanomes (cancers de la peau) et cancers du cerveau ou Gb3 dans les lymphomes de Burkitt (cancers des organes lymphoïdes) ou cancers ovariens (Hakomori 2000). A l'inverse, d'autres structures peuvent disparaître. Il est cependant très difficile de savoir si l'apparition de ces antigènes lipidiques est causée par une accumulation des précurseurs due à une synthèse incomplète ou au développement de nouvelles voies de biosynthèse provoquée par l'activation de nouveaux gènes de glycosyltransférases (Hakomori 2008). De manière générale, ces modifications influencent le taux de croissance tumorale et jouent un rôle clé dans l'induction de l'invasion et la formation de métastases, selon des mécanismes encore très peu connus. L'expression de ces antigènes lipidiques tumoraux dans le microenvironnement

de la tumeur aide les cellules tumorales à échapper au système immunitaire de l'hôte (Birkle *et al.* 2003).

sLe <sup>a</sup>		Pancréas
sLe <sup>x</sup>		Sein, poumon
Globo-H		Sein, prostate
Gb3		Lymphome, ovaire
GD3		Mélanome, cerveau, sein
GD2		Mélanome, cerveau
		

**Tableau 1.1. Principales structures de glycosphingolipides cancéreux chez l'homme.**

Les carrés bleus schématisent les résidus GlcNAc, les ronds jaunes schématisent les résidus Gal, les triangles rouges schématisent les résidus Fuc. Les images ont été réalisées sur la plateforme GlycoWork Bench (Ceroni *et al.* 2007).

#### 2.4.2. Protéoglycanes et cancer

Même si la plupart des protéoglycanes sont modifiés dans le processus cancéreux, que ce soit au niveau de leur contenu oligosaccharidique ou de leur distribution, les héparanes sulfates (HS) restent les seuls réellement impliqués dans la pathogénicité du cancer. Ils sont naturellement composés de domaines hypersulfatés, communément dénommés domaines S, séparés par des régions faiblement sulfatées, appelées régions N-acétylées ou domaine A. Cette organisation particulière permet d'isoler des sites de reconnaissance spécifiques pour les

différents partenaires d'interaction de ces molécules. Ajoutée à leur grande diversité structurale, ces molécules sont capables de « dialoguer » avec de nombreux partenaires tels que les facteurs de croissance, les chimiokines, les cytokines, des enzymes, la matrice extracellulaire ou des molécules d'adhésion et assurent des rôles variés dans la cellule (Imberty *et al.* 2007).

Comme la majorité des composants cellulaires, les HS subissent des modifications structurales importantes durant la transformation tumorale. Ces perturbations induisent par exemple une diminution d'expression de certains HS (Sanderson *et al.* 2004) ou une modification de la *O*-sulfatation des domaines S (Jayson *et al.* 1998). Ne présentant plus la même structure, ces HS modifiés confèrent alors de nouvelles propriétés à la cellule, dont la transformation cellulaire, la croissance tumorale, la formation de métastases ou encore l'angiogenèse. Ces modifications peuvent également être contrôlées par la cellule elle-même qui s'octroie ainsi, de manière dynamique, de nouvelles propriétés oncogéniques indispensables (Liu *et al.* 2002 ; Sasisekharan *et al.* 2002). Par exemple, durant l'invasion, les cellules tumorales vont sécréter des enzymes capables de dégrader à la fois les composantes protéiques et oligosaccharidiques, dénommées respectivement matrix metalloproteinases et héparanases. Ces dernières vont cliver les HS au niveau de leurs domaines faiblement sulfatés libérant ainsi de courts motifs de sucres de quelques kDa. Les perturbations provoquées alors au niveau de ces HS seront à l'origine d'une augmentation du caractère métastatique du cancer accompagnée d'une diminution du taux de survie post-opératoire des patients (Vlodavsky *et al.* 2008). Les modifications structurales des GAGs sont également à l'origine d'autres perturbations cellulaires telles que le dérèglement du fonctionnement des intégrines, des facteurs de croissance... La liste des conséquences induites par une modification structurale des GAGs dans le cancer n'est évidemment pas exhaustive lorsque l'on considère le nombre important de molécules impliquées dans la reconnaissance des GAGs et ne sera pas davantage développée dans ce manuscrit.

### **2.4.3. Ancres GPI et cancer**

Plusieurs études ont démontré une perte partielle ou totale de protéines à ancre GPI à la surface de cellules cancéreuses (Rege and Hagood 2006), notamment celles à l'origine des lymphomes de Burkitt. Cette perte d'expression peut être expliquée par une perturbation dans la transcription des gènes *PIGL* et *PIGY* impliqués dans les premières étapes de biosynthèse des ancres GPI (Hu *et al.* 2009). Même si les conséquences physiologiques dues à cette

diminution d'expression n'ont pas été clairement définies, elle semble néanmoins associée aux cellules cancéreuses avec un potentiel prolifératif important.

A l'inverse, d'autres protéines à ancre GPI sont fortement surexprimées dans les cancers humains, dont deux appartenant à la famille des matrix metalloproteinases (MMP). Cette famille est composée d'endopeptidases à zinc dont l'implication dans la progression cancéreuse, la régulation des métastases et l'angiogenèse a été démontré depuis longtemps. Deux membres de cette famille MT4- et MT6-MMP sont fortement exprimés dans plusieurs cancers humains (cerveau, colon, prostate ainsi que les gliomes). Elles sont impliquées dans la progression de la maladie et la formation des métastases (Sohail *et al.* 2008).

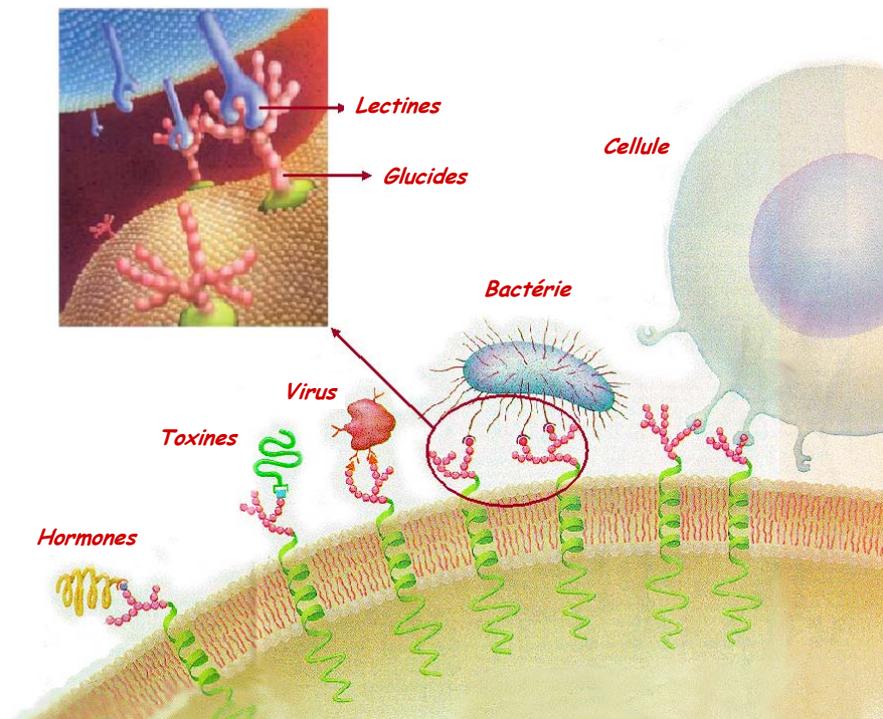
### 3. Lectines et cancer

#### 3.1. Définition et historique des lectines

##### 3.1.1. Définition

Les interactions protéines-glucides jouent un rôle essentiel dans de nombreux phénomènes de reconnaissance et d'adhésion cellulaire. Ces interactions intercellulaires sont notamment assurées par des protéines particulières, d'origine non immunitaire, appelées lectines (Lis and Sharon 1998). Elles ont la capacité de reconnaître et de fixer spécifiquement de manière non covalente et réversible, des structures oligosaccharidiques sans les transformer. Ces protéines sont présentes dans tout le règne vivant avec une très grande diversité structurale et jouent des rôles variés dans la vie sociale des cellules et les interactions avec l'environnement (Sharon and Lis 2004). Elles interviennent notamment dans des phénomènes de reconnaissance hôte-pathogène (l'hémagglutinine du virus *Influenza* est responsable de l'adhésion du virus aux cellules hôte et initie l'infection), dans des mécanismes de défense chez certains invertébrés en agglutinant les pathogènes, dans l'établissement de symbiose entre les bactéries du sol du genre *Rhizobium* et les légumineuses ou encore dans les phénomènes d'immunité innée en favorisant la reconnaissance des leucocytes sur les cellules endothéliales pour initier la réponse immunitaire (**Figure 1.19**). Malgré cette dernière caractéristique, les lectines n'appartiennent

pas au système immunitaire de l'hôte, à la différence des anticorps. Cette distinction permet de souligner la différence entre lectines et immunité innée et anticorps et immunité acquise.



**Figure 1.19. Rôle des glycoconjugués situés à la surface cellulaire.**

Représentation schématique d'exemple d'interactions lectines-sucres.

### 3.1.2. Historique

A la fin du XIX<sup>ème</sup> siècle, plusieurs études mettent en évidence l'existence de protéines capables d'agglutiner naturellement des hématies ou érythrocytes, d'où leur nom d'hémagglutinines. La première description de l'une de ces protéines a été réalisée par Peter Hermann Stillmark durant sa thèse de doctorat en 1888 à l'université de Dorpat (aujourd'hui Tartu en Estonie) où il présenta une hémagglutinine très toxique extraite des graines du ricin (*Ricinus communis*), appelée ricine. En 1919, James B. Sumner, de l'université Cornell (Ithaca, New York), isola à partir du pois (*Canavalia ensiformis*) la première hémagglutinine pure, la concanavaleine A (Sumner 1919). Il fallut patienter presque vingt ans et les travaux de Sumner et Howell en 1936 pour que la spécificité de ces protéines pour les sucres soit mise en évidence avec l'inhibition de l'hémagglutination de la concanavaleine A par du saccharose (Sumner and Howell 1936).

En 1940, environ 40 ans après la découverte des groupes sanguins A, B et O par Karl Landsteiner de l'université de Vienne, William C. Boyd de l'université de Boston et Karl O. Renkonen de l'université d'Helsinki démontrent la spécificité de ces hémagglutinines pour les différents groupes sanguins humains. Cette découverte conduira Boyd et Shapleigh en 1954 à proposer pour ces protéines le nom de lectines, du latin *legere*, qui signifie choisir, reconnaître (Boyd and Shapleigh 1954) (revue par (Sharon and Lis 2004)).

Avec l'introduction des chromatographies d'affinité pour la purification de lectines par Irwin J. Goldstein et Bipin B. L. Agrawal de l'université du Michigan (Agrawal and Goldstein 1965) le nombre de lectines caractérisées a considérablement augmenté ces 40 dernières années. Elles sont aujourd'hui largement utilisées dans différents domaines de la biologie, notamment pour le typage des groupes sanguins, la stimulation mitogénique des lymphocytes ou encore le marquage histochimique de cellules cancéreuses.

### 3.1.3. Affinité et spécificité des lectines

Les lectines reconnaissent de manière spécifique des mono- ou oligosaccharides. Basée sur leur spécificité, les lectines sont classées en cinq groupes différents selon le monosaccharide pour lequel elles présentent la plus forte affinité : i) les lectines à Man, ii) les lectines à Gal/GalNAc, iii) les lectines à GlcNAc, iv) les lectines à Fuc et v) les lectines à NeuAc (Lis and Sharon 1998). Cette reconnaissance est souvent désignée comme la « spécificité primaire » des lectines. Ces monosaccharides et leurs dérivés sont les principaux constituants des épitopes des surfaces cellulaires eucaryotes. A quelques exceptions près, les lectines n'ont pas une très grande affinité pour leurs ligands monosaccharidiques, avec des constantes d'affinité de l'ordre du millimolaire (mM). Cette affinité moyenne est généralement compensée par la multivalence de ces protéines. Elles sont également capables de reconnaître des structures oligosaccharidiques plus complexes, composées de deux, trois voire quatre résidus. Dans ces cas là, les constantes d'affinité peuvent être jusqu'à 1000 fois supérieure (Dam and Brewer 2002). Certaines lectines telles que PHA-L et PHA-E de *Phaseolus vulgaris* possèdent même des spécificités strictes pour des oligosaccharides (Cummings *et al.* 1982).

La sélectivité des lectines vis-à-vis de leurs ligands est relativement importante. Par exemple, les lectines spécifiques du galactose ne réagissent pas avec le glucose, son épimère en C4 ou avec le mannose, épimère du glucose en C2. De la même manière, les lectines spécifiques du mannose ne reconnaissent pas le galactose. Cependant, ces règles ne sont pas

toujours vérifiées. Par exemple, la plupart des lectines qui fixent le galactose, dont l'agglutinine du soja (SBA), reconnaissent aussi son homologue *N*-acétylé (Goldstein and Poretz 1986). Certaines lectines peuvent également reconnaître le mannose et le fucose comme PA-III de *Pseudomonas aeruginosa* (Imberty *et al.* 2004). Ce phénomène est dû à la présence des hydroxyles dans une topologie très similaire. De plus, certaines lectines présentent une spécificité anomérique et peuvent distinguer la configuration en C1 de monosaccharides tels que l'alpha-méthyl D-galactoside et le beta-méthyl D-galactoside.

Les lectines sont des protéines ubiquitaires qui présentent une grande diversité structurale. Le nombre de structures cristallographiques de lectines est en constante augmentation. Plus de 840 structures, principalement obtenues par cristallographie aux rayons X, sont actuellement résolues et répertoriées dans une base de données développée au laboratoire (<http://www.cermav.cnrs.fr/lectines>). La banque de données comprend ainsi 50 familles structurales différentes dont six sont d'origine végétale, vingt d'origine animale, dix d'origine bactérienne, huit familles d'origine fongique, cinq d'origine virale et une famille d'algues (**Tableau 1.2.**).

Origine	Exemples de lectines	Native	Complexé	Total
Plantes	Ricine Concanavaline A	100	219	319
Bactéries	PA-III de <i>P. aeruginosa</i> Toxine de cholera	39	95	134
Animaux	E-selectin Helix pomatia agglutinin	81	172	253
Virus	Hemagglutinine de virus Capside de rotavirus	33	40	73
Champignons	Lectine de mousseron	26	27	53
Algues	Griffithsine	2	7	9
<b>Total</b>		<b>281</b>	<b>560</b>	<b>841</b>

**Tableau 1.2. Récapitulatif des structures tridimensionnelles de lectines**

### 3.1.4. Applications des lectines en cancérologie

Le taux de survie d'un patient atteint d'un cancer est étroitement lié au stade de développement de la maladie au moment du diagnostic. Plus le dépistage est réalisé de manière précoce, meilleures sont les chances de guérison. Ainsi, selon l'institut national du cancer ([www.e-cancer.fr](http://www.e-cancer.fr)), lorsque le cancer colorectal est dépisté et traité prématurément (cancer superficiel dans la paroi), le taux de survie à cinq ans est de 94%. D'où l'intérêt permanent de découvrir de nouveaux marqueurs pour la détection de tous types de cancers.

Des molécules glycosylées sont d'ores et déjà utilisées en médecine pour le diagnostic de certains cancers. Nous pouvons citer parmi les plus répandus les marqueurs CEA (carcinomembryonic antigen) pour le dépistage du cancer colorectal (Gold and Freedman 1965), CA-125 pour le cancer de l'ovaire (Bast *et al.* 1981), PSA (prostate-specific antigen) pour le cancer de la prostate (Stamey *et al.* 1987) ou encore AFP ( $\alpha$ -fetoprotein) pour le cancer du foie (Khien *et al.* 2001). Ces molécules sont des protéines présentes de manière anormalement élevée dans le sérum des patients malades. Ainsi, leur dosage sérique renseigne sur le développement de la maladie et/ou l'éventuel risque de récurrence après une chirurgie pour l'ablation de la tumeur (Park *et al.* 2009). Ces protéines sont toutes glycosylées et leur glycosylation est modifiée lors de la transformation tumorale (Tabares *et al.* 2006). Ces variations de glycosylation pourraient alors être utilisées dans le dépistage du cancer, notamment avec l'utilisation de nouveaux outils comme les lectines. L'AFP, par exemple, possède un site de *N*-glycosylation. Il a été démontré que les fractions d'AFP reconnues par la lectine de lentille *Lens culinaris* agglutinin (LCA), spécifique de l'épitope Fuc $\alpha$ 1-6GlcNAc, sont associées à un faible pronostic vital pour les patients atteints de cancer du foie (Tada *et al.* 2005), ce qui indique que la fucosylation de l'AFP peut être associée à un pronostic préoccupant. Le marqueur PSA, quant à lui, peut également présenter des taux anormalement élevés dans le sérum de patients atteints de pathologies bénignes. En conséquence, en utilisant uniquement le dosage du PSA pour le dépistage du cancer de la prostate, de faux positifs peuvent être révélés. En 2004, Ohyama *et al.* ont néanmoins démontré que la lectine *Maackia amurensis* agglutinin (MAA), qui reconnaît les acides sialiques liés en  $\alpha$ 2-3, présente une affinité plus importante pour le PSA issu de patients atteints de cancer plutôt que celui issu de patients atteints de maladies bénignes de la prostate. Ces résultats sont donc encourageants pour l'utilisation de cette lectine dans le diagnostic du cancer de la prostate (Ohyama *et al.* 2004). D'autres exemples de lectines intéressantes pour le diagnostic peuvent être énumérés.

La lectine de haricot *Phaseolus vulgaris* (PHA-L) reconnaît, de manière spécifique, les *N*-glycanes branchés en  $\beta$ 1-6, particulièrement associés à la malignité du cancer. Une corrélation entre la fixation de PHA-L sur des coupes de cancers du sein et du colon et le mauvais pronostic vital associé a d'ailleurs pu être mis en évidence (Fernandes *et al.* 1991). Deux autres lectines végétales, dénommées *Maclura pomifera* agglutinin (MPA) et *Vicia villosa* agglutinin (VVA) sont respectivement spécifiques des antigènes T (Gal $\beta$ 1-3GalNAc-*O*-Ser/Thr) et Tn (GalNAc-*O*-Ser/Thr). Sachant les conséquences pour le patient du développement de ces épitopes, ces lectines sont donc particulièrement importantes pour le diagnostic de cancers à fort impact pathologique (Korourian *et al.* 2008). Une autre lectine, issue de l'escargot de Bourgogne, appelée *Helix pomatia* agglutinin (HPA), est également spécifique de l'antigène Tn. De nombreuses études ont pu mettre en évidence son caractère prédictif pour l'identification de cancers à faible pronostic vital. Ses caractéristiques seront largement développées dans le paragraphe suivant. Les principales lectines impliquées dans la reconnaissance d'épitopes cancéreux sont résumées dans le **tableau 1.3**.

Lectines	Espèce d'origine	Motif glycanique reconnu
Dolichos biflorus agglutinin (DBA)	Fleur du même nom	GalNAc
<i>Lens culinaris</i> agglutinin (LCA)	La lentille	Fuc $\alpha$ 1-6GlcNAc
Griffonia simplicifolia (GSI-A4)	Fleur du même nom	Gal $\beta$ 1-4GlcNAc
Griffonia simplicifolia (GSI-B4)	Fleur du même nom	Gal
<i>Helix aspersa</i> agglutinin (HAA)	L'escargot petit gris	GalNAc
<i>Helix pomatia</i> agglutinin (HPA)	L'escargot de Bourgogne	$\alpha$ GalNAc
<i>Limex flavus</i> agglutinin (LFA)	La limace	Neu5Ac $\alpha$ 2-3 / $\alpha$ 2-6
<i>Maackia amurensis</i> agglutinin (MAA)	Le maackie de l'amour	Neu5Ac $\alpha$ 2-3Gal
<i>Maclura pomifera</i> agglutinin (MPA)	L'oranger des Osages	Gal $\beta$ 1-3GalNAc
<i>Phaseolus vulgaris</i> leucoagglutinating lectin (PHA-L)	Le haricot	<i>N</i> -glycanes $\beta$ 1-6 branchés
<i>Soybean</i> agglutinin (SBA)	Le soja	GalNAc $\alpha$ et $\beta$
<i>Sambucus nigra</i> agglutinin (SNA-I)	Le sureau noir	Neu5Ac $\alpha$ 2-6
<i>Ulex europaeus</i> agglutinin (UEA-I)	L'ajonc d'Europe	$\alpha$ Fuc
<i>Vicia villosa</i> agglutinin (VVA)	La vesce poilue	GalNAc $\alpha$ et $\beta$
<i>Wheat germ</i> agglutinin (WGA)	Le blé	GlcNAc

**Tableau 1.3. Principales lectines utilisées pour la mise en évidence de glycotopes en oncologie.**

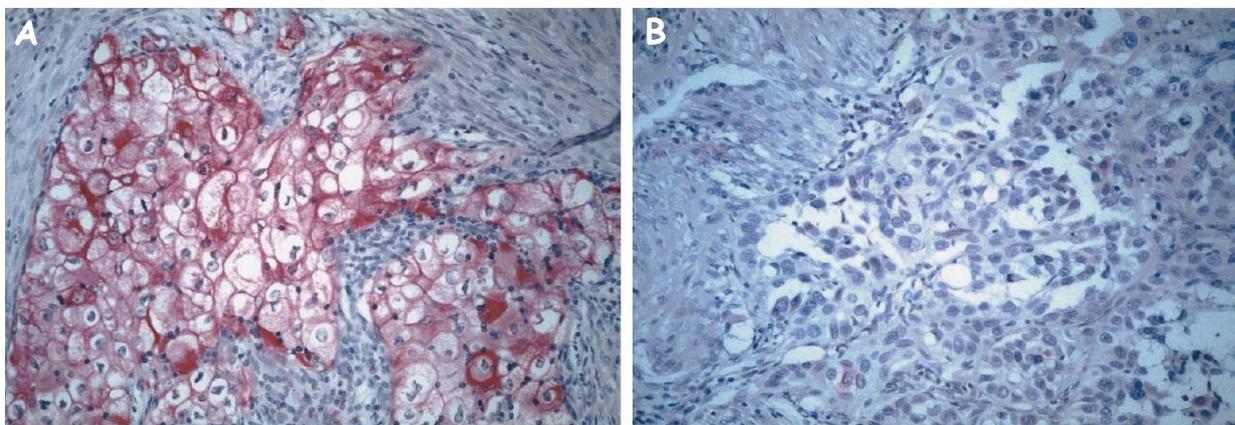
## 3.2. Utilisation de la lectine HPA comme outil de diagnostic dans le cancer

### 3.2.1. HPA, marqueur de cancers à fort pouvoir métastatique

Depuis maintenant plus de vingt ans, la lectine HPA, extraite de l'escargot de Bourgogne est utilisée en histopathologie comme marqueur de plusieurs cancers, dont les adénocarcinomes du sein, de l'estomac, du colon, du poumon mais aussi les mélanomes. Cette lectine a la particularité de reconnaître spécifiquement des glycoépitopes présents à la surface des cancers primaires, associés au développement de métastases notamment dans les ganglions lymphatiques, impliquant, à court terme, un mauvais pronostic vital et un raccourcissement de la vie des patients. Elle est également capable de se fixer sur des épitopes présents sur les cellules ayant déjà métastasées. Ces résultats ont été tout d'abord démontrés, *in vitro* sur des coupes de tissus cancéreux issus de patients dont l'historique de la maladie était connu (**Figure 1.20.**). Que ce soit pour le cancer du sein (Brooks and Leathem 1991), de l'estomac (Kakeji *et al.* 1994), du colon (Ikeda *et al.* 1994), du poumon (Laack *et al.* 2002) mais aussi pour les mélanomes (Thies *et al.* 2001), la plupart des tumeurs marquées positivement par HPA ont la capacité de métastaser soit dans les ganglions lymphatiques soit dans d'autres organes périphériques. En effet, quelque soit la localisation du cancer primaire et des métastases, 80% des cancers à caractère métastatique sont marqués positivement par HPA. Néanmoins, 20% restent insensibles au marquage malgré leur capacité à métastaser (Schumacher *et al.* 1992 ; Brooks and Leathem 1998). Un dérèglement de la machinerie de glycosylation dans les cellules cancéreuses pourrait expliquer ces résultats (Brooks 2000).

La particularité d'HPA à reconnaître les cancers à faible pronostic vital a également pu être vérifiée *in vivo* en injectant des lignées cellulaires cancéreuses métastatiques ou non dans deux systèmes de souris immunodéficientes, les souris nude incapables de produire de lymphocytes T et les souris scid pour severe combined immunodeficiency qui ne produisent aucun lymphocyte mature et fonctionnel. En effet, les lignées cellulaires de mélanomes reconnues par HPA forment plus facilement des métastases dans les poumons des souris nude que celles qui ne sont pas reconnues par la lectine (Kjonniksen *et al.* 1994). Des résultats parallèles ont été également obtenus avec des souris scid. Sur les 26 animaux transplantés avec des lignées cellulaires de sein et de colon marquées positivement à HPA, 23 développent des métastases aux poumons alors que seulement deux animaux sur 28 transplantés avec des

lignées HPA négatives métastasent (Schumacher and Adam 1997). Le marquage chez les souris transplantées par une lignée cellulaire métastatique est moins intense que celui de la même lignée *in vitro* ce qui montre que l'hétérogénéité de la reconnaissance d'HPA est plus importante *in vivo* qu'*in vitro* (Schumacher 1996).



**Figure 1.20. Marquages positif (A) et négatif (B) par HPA biotinylé d'adénocarcinomes du poumon. Grossissement x200 (Laack *et al.* 2002).**

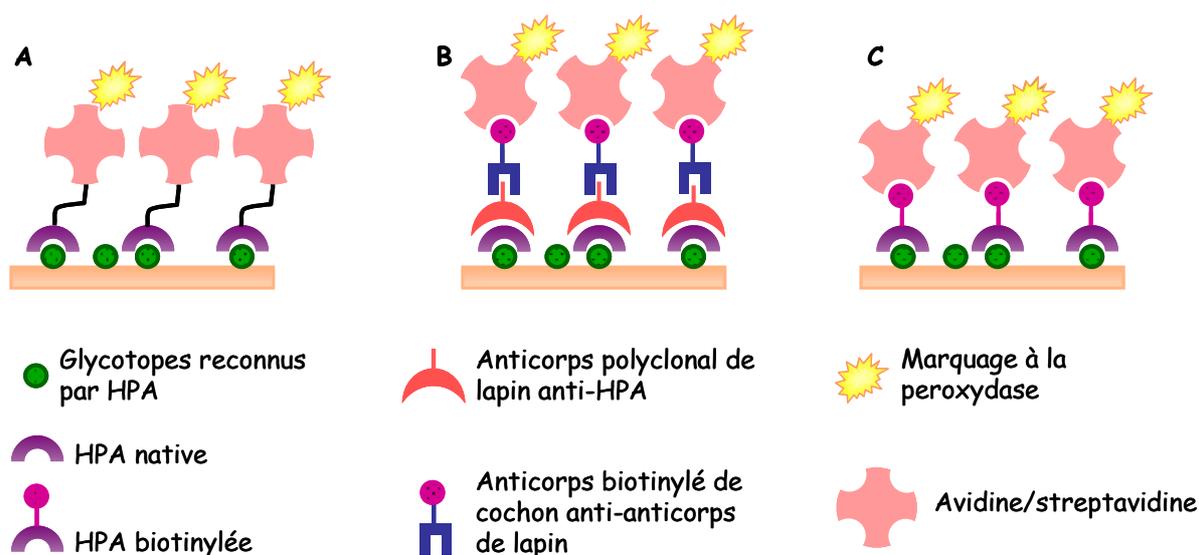
En revanche, aucune association n'a pu être mise en évidence entre d'autres facteurs tels que la taille de la tumeur, l'âge, le sexe ou le groupe sanguin du patient et son pronostic vital (Brooks 2000). Ce dernier point a néanmoins été contredit par Welinder *et al.* en 2009 qui observent une corrélation entre les groupes sanguins des patients et le marquage par HPA de cancers du sein (Welinder *et al.* 2009). Cette étude démontre également des différences de marquage à l'intérieur d'un groupe de patients possédant le même groupe sanguin. Des études complémentaires devront être réalisées pour éclaircir ce point.

### 3.2.2. Importance du choix de la méthode de détection

Malgré le fait qu'un nombre important d'études ait démontré une relation claire entre le marquage par HPA et le faible pronostic vital des patients atteints de cancers, un groupe n'a cependant pas pu vérifier ces résultats (Gusterson 1993). Cet écart pourrait être dû à des différences dans les techniques immunocytochimiques utilisées pour la détection.

Toutes les études qui ont illustrées avec succès une corrélation entre la reconnaissance d'HPA et le faible pronostic ont utilisées soit une lectine native associée à un système de détection indirect utilisant un premier anticorps anti-HPA, un second anticorps biotinylé

dirigé contre le précédent et un complexe streptavidine-peroxydase qui, après ajout du substrat de l'enzyme, donnera une coloration (**Figure 1.21. B**), soit une lectine HPA conjuguée à une petite molécule de biotine, utilisant directement le complexe avidine-peroxydase pour la révélation (**Figure 1.21. C**) (Laack *et al.* 2002). Gusterson *et al.* ont, quant à eux, utilisé une lectine directement couplée au volumineux complexe streptavidine-peroxydase pour une détection directe (**Figure 1.21. A**).



**Figure 1.21. Différentes méthodes de marquage des épitopes cancéreux par HPA.**

**A.** La lectine HPA est directement couplée au complexe volumineux streptavidine-peroxydase. **B.** Après avoir reconnu le glycotope à la surface de la cellule, la lectine HPA native est reconnue par un anticorps polyclonal de lapin anti-HPA, lui-même reconnu par un second anticorps de cochon biotinylé anti-anticorps de lapin. La révélation se fait suite à l'ajout du complexe streptavidine-peroxydase. **C.** La lectine HPA biotinylée est reconnue par le complexe avidine-peroxydase.

Une étude de 1996 a tenté de comprendre ses différences de résultats en marquant une série de 373 cancers primaires de sein selon deux méthodes : la première directe et la seconde indirecte employant le couple d'anticorps spécifiques décrit précédemment (Brooks *et al.* 1996). Les résultats étaient surprenants. Quelque soit la méthode utilisée, 79% des cancers étaient marqués positivement par HPA, 21% n'étant pas reconnus. Cependant, la comparaison de manière plus précise des résultats montre qu'environ 10% des cas « négatifs » par une méthode devenaient « positifs » par l'autre et vice versa et que la méthode indirecte produisait des informations pronostiques plus fiables que la méthode directe. Ces différences ne sont néanmoins pas seulement dues aux variabilités de sensibilité entre les deux méthodes mais aussi à l'utilisation de lectine conjuguée ou non. Le complexe streptavidine-peroxydase est

très volumineux en comparaison à HPA et pourrait alors interférer avec le site de liaison de la lectine modifiant ainsi ses caractéristiques de liaison. Cette technique, malgré sa facilité d'utilisation, semble donc à éviter pour un pronostic valide. En revanche, l'utilisation d'HPA conjuguée à une molécule de biotine beaucoup plus petite que le complexe streptavidine-peroxydase est une méthode aussi sensible et fiable que la méthode indirecte et sans doute moins fastidieuse (Brooks and Wilkinson 2003). Le choix de la méthode de détection pour le marquage de cancers par HPA est donc primordial pour conserver le caractère prédictif de la lectine et s'assurer ainsi de la validité des résultats.

### 3.2.3. Identification des glycotopes reconnus par HPA

Le marquage par HPA à la surface des cellules cancéreuses est, lorsqu'il a lieu, généralement intense et distribué sur une portion substantielle de cellules. La coloration est cytoplasmique et granuleuse avec une localisation marquée du contour des cellules (Schumacher and Adam 1997 ; Saint-Guirons *et al.* 2007). Des études par microscopie confocale ont également révélé un intense marquage de l'appareil de Golgi (Brooks 2000). Quelque soit sa localisation, le marquage peut être inhibé après une incubation en présence de *N*-acétylgalactosamine, résidu monosaccharidique spécifique d'HPA (Brooks and Carter 2001).

Depuis de nombreuses années, les efforts se sont concentrés sur l'identification du ou des glycotopes reconnus par HPA, afin de tenter d'expliquer les perturbations des voies de glycosylation engendrées par les processus de formation des métastases et d'identifier ainsi les étapes clé à cibler pour une meilleure efficacité thérapeutique. Les différentes études publiées utilisaient principalement deux méthodes : d'une part, les techniques d'électrophorèse sur gel de polyacrylamide en condition dénaturante (SDS-PAGE) et WesternBlot où l'anticorps primaire est en réalité la lectine HPA et d'autre part, la purification d'extraits cellulaires directement sur colonne HPA-agarose. Ces publications ont pu mettre en évidence la présence de plusieurs glycoprotéines reconnues par la lectine (Schumacher *et al.* 1995 ; Brooks *et al.* 2001). Très récemment, l'équipe du Dr. Miriam Dwek à Londres a isolé, par purification sur colonne HPA-sepharose, seize protéines membranaires reconnues par HPA à partir de la lignée de cancer colorectal fortement métastatique HT 29. Ces protéines ont été identifiées comme étant les intégrines  $\alpha 6$  et  $\alpha V$  impliquées dans l'angiogenèse, les annexines A2 et A4 assurant un rôle dans les phénomènes d'adhésion cellulaire, d'apoptose et de formation des métastases mais aussi la  $\beta$ -actine, l' $\alpha$ - et  $\beta$ -tubuline,

protéines du cytosquelette impliquées dans la motilité cellulaire et l'invasion métastatique et les cytokératines 8 et 18, retrouvées en complexe avec des protéines impliquées dans les voies anti-apoptotiques telles que Hsp70 et Hsp90 (Saint-Guirons *et al.* 2007).

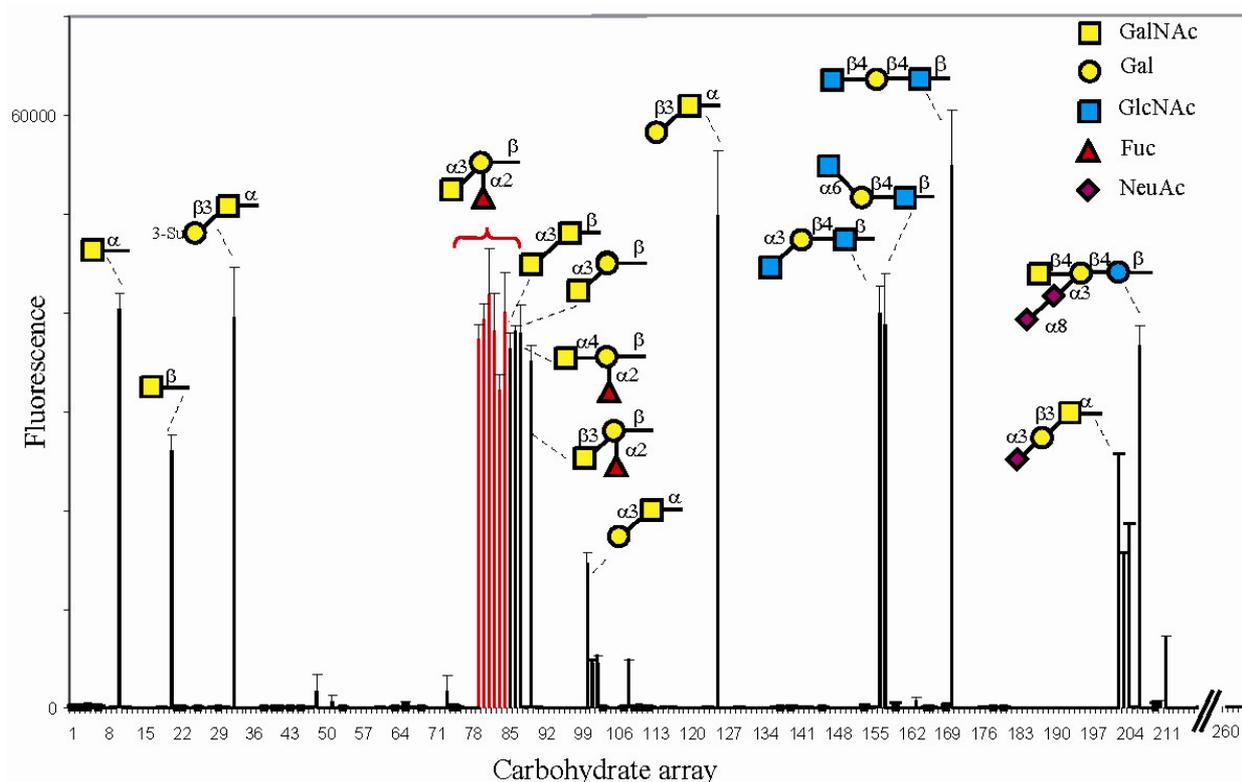
La partie oligosaccharidique des glycoconjugués reconnues par HPA est toujours inconnue à ce jour. Son identification serait pourtant primordiale pour la compréhension des processus de cancérisation et de formation des métastases mais également des phénomènes d'invasion cellulaire (Rye *et al.* 1998). Deux approches ont été principalement développées pour tenter d'identifier ce ou ces glycanes associés à la formation des métastases. La première technique basée sur l'emploi de glycosidases différentes a démontré que les structures reconnues par HPA étaient formées entre autres de résidus de poly-*N*-acétyllactosamine liés sur la protéine via un lien *N*-glycosidique mais le fait que ces glycanes soient portés par un lipide ne peut néanmoins pas être exclu (Ito *et al.* 1996). La seconde approche consiste à marquer les mêmes échantillons cancéreux avec différents lectines aux spécificités relativement proches ou avec des anticorps. Par exemple, le marquage de cellules du sein par les lectines HPA et DBA, spécifiques du GalNAc et par un anticorps dirigé contre l'antigène Tn, épitope longtemps considéré comme ligand exclusif d'HPA ont été comparé (Springer 1989). Les résultats démontrent que même si la spécificité primaire est identique pour les trois protéines, les marquages obtenus sont très différents, HPA reconnaissant davantage d'épitopes que DBA et l'anticorps anti-Tn. Ils soulignent ainsi la capacité d'HPA à reconnaître et différencier des structures plus complexes, probablement formées de plusieurs monosaccharides (Brooks and Leathem 1998). En 2001, une des structures reconnues par HPA a été caractérisée pour la première fois de manière relativement imprécise, grâce à plusieurs étapes de purification. Un oligosaccharide fortement reconnu par HPA, nommé HPA glycan 1 ou HPAGly1, a été isolé à partir de glycoprotéines de lignées de cancers du sein purifiées sur colonne HPA. L'étude de son rayon hydrodynamique a indiqué qu'il serait formé de quatre unités monosaccharidiques neutres et d'un acide sialique (Dwek *et al.* 2001). L'émergence de nouveaux domaines de compétence tels que la glycoprotéomique par exemple, avec l'analyse par spectrométrie de masse de structures glycaniques des glycoconjugués devrait considérablement faciliter l'accession à ces glycanes reconnus par HPA. Ces méthodes nécessitent cependant un arsenal technique et humain perfectionné, ce qui ralentit l'obtention de nouvelles données.

## 4. La famille des lectines de type H

### 4.1. HPA

#### 4.1.1. Caractérisation biochimique

*Helix pomatia* agglutinin est une lectine qui a été identifiée à la fin du XIX<sup>ème</sup> siècle dans l'escargot de Bourgogne (Camus 1899). Principalement localisée dans sa glande à albumine, elle assurerait un rôle dans la fertilisation et la protection des œufs, en agglutinant les éventuels pathogènes. Cette lectine a la particularité d'agglutiner spécifiquement les érythrocytes appartenant au groupe sanguin A alors qu'elle n'a aucun effet sur ceux des groupes sanguins B et O. Sa spécificité pour le résidu *N*-acétylgalactosamine et en particulier l'anomère alpha a été déterminée depuis maintenant près de 50 ans (Hammarstrom and Kabat 1969) mais HPA est également capable de reconnaître des structures plus complexes avec des affinités parfois meilleures que celles observées pour le monosaccharide. Les ligands suivants sont classés par ordre de préférence : antigène de Forssman ( $\alpha$ GalNAc1-3GalNAc-R) > épitope du groupe sanguin A ( $\alpha$ GalNAc1-3[ $\alpha$ Fuc1-2]Gal) > antigène Tn ( $\alpha$ GalNAc-Ser/Thr) > GalNAc > GlcNAc (Wu 1991 ; Sanchez *et al.* 2006). L'étude de sa spécificité, par glycan array, démontre également qu'HPA est capable de reconnaître l'antigène T (Gal $\beta$ 1-3GalNAc-Ser/Thr) (**Figure 1.22.**) (Sanchez *et al.* 2006). Elle a été identifiée comme étant un hexamère d'un poids moléculaire de 79 kDa (Hammarstrom and Kabat 1969). Son fractionnement en au moins douze composants par la technique d'électrofocalisation révèle qu'elle se présente probablement sous différentes glycoformes (Vretblad *et al.* 1979).



**Figure 1.22.** Analyse, par glycan array, de la spécificité d'HPA (Sanchez *et al.* 2006).

Au sein du laboratoire, notre équipe a réussi à cloner le gène codant pour HPA. En effet, grâce à la structure cristallographique obtenue avec la protéine commerciale et détaillée dans le paragraphe suivant, des fragments de séquences peptidiques d'HPA ont pu être déterminées. Des amorces dégénérées spécifiques du N-terminal de la protéine mature ont été conçues et le gène codant pour HPA a pu être isolé à partir des ARN totaux d'escargot. Le gène code pour une protéine mature de 101 acides aminés, précédée d'un peptide signal qui en comporte 20. Le poids moléculaire calculé de la protéine mature est de 11,3 kDa, avec un point isoélectrique (pI) estimé de 8,20. HPA possède un site de *N*-glycosylation au niveau de l'Asn 34. Les analyses par spectrométrie de masse ont révélé plusieurs glycoformes, avec des masses moléculaires correspondant respectivement à un dimère non glycosylé, mono- ou diglycosylé (Sanchez *et al.* 2006). L'écart entre chaque pic correspond à un *N*-glycane de faible poids moléculaire, caractérisé comme étant le cœur pentasaccharidique classique des *N*-glycanes ( $\text{Man}_3\text{GlcNAc}_2$ ), additionné d'un résidu fucose lié en  $\alpha 1-6$  sur le GlcNAc interne et d'un résidu xylose lié en  $\beta 1-2$  sur le mannose branché. Cette structure glucidique est particulièrement représentée sur les  $\alpha$ -hémocyanines d'*Helix pomatia* (van Kuik *et al.* 1985).

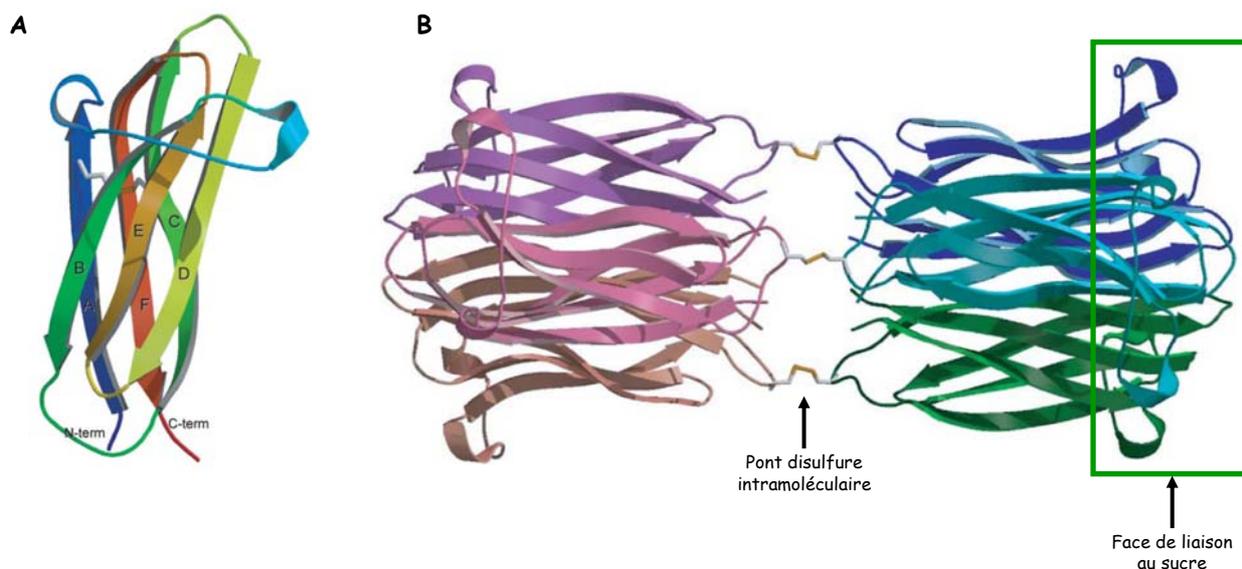
Les paramètres thermodynamiques de l'interaction entre HPA et le GalNAc ont été déterminés par microcalorimétrie isotherme de titration. La constante de dissociation s'élève à  $1,3 \cdot 10^{-4}$  M. Cette liaison est fortement dirigée par l'enthalpie puisque l'énergie libre de l'interaction ( $\Delta G = -22$  kJ/mol) a une contribution enthalpique  $\Delta H$  de -35 kJ/mol contrebalancée par une entropie fortement défavorable  $T\Delta S$  de -13 kJ/mol. Toutes ces valeurs sont caractéristiques des interactions de type lectine-sucre.

#### 4.1.2. Caractérisation structurale

La structure cristallographique d'HPA a été résolue au laboratoire à 2,5 Å pour la structure native (PDB 2CE6) et 1,3 Å pour le complexe avec le GalNAc (PDB 2CCV), dans les groupes d'espace  $P6_322$  et  $H32$ , respectivement (Sanchez *et al.* 2006).

Chaque monomère d'HPA est composé d'un sandwich  $\beta$  formé de six brins antiparallèles. Chaque brin est connecté au suivant par une petite boucle, excepté pour les brins A et B qui sont reliés par une boucle beaucoup plus longue, formant une épingle à l'une des extrémités du sandwich. Les brins A et E sont reliés par un pont disulfure intramoléculaire entre les cystéines 9 et 80 (**Figure 1.23. A.**). Chaque monomère est relié à un autre monomère via l'établissement d'un pont disulfure intermoléculaire entre les cystéines 42 de deux monomères. Cette organisation permet la formation d'un long dimère queue à queue. Associé à deux autres dimères identiques, ils forment une structure hexamérique, correspondant en réalité à un trimère de dimère (**Figure 1.23. B.**). Ce type de repliement n'a jamais été décrit précédemment et a donc donné naissance à une nouvelle famille de lectines dite de type H.

La structure cristallographique obtenue en complexe avec le GalNAc est très similaire à celle de la protéine native. Chaque hexamère présente deux faces de liaison au sucre de part et d'autre de la protéine (**Figure 1.23. B.**).



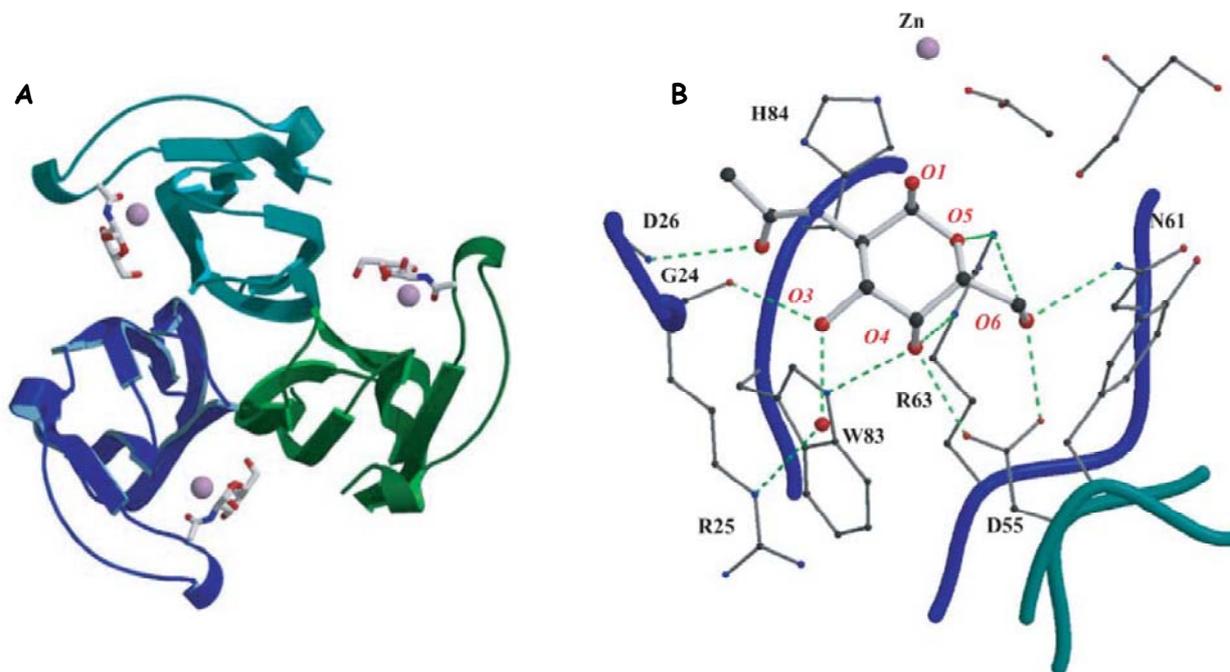
**Figure 1.23. Représentation en dessin de la structure cristallographique d’HPA.**

**A.** Monomère avec visualisation du pont disulfure intramoléculaire entre les brins A et E. **B.** Hexamère formé à partir d’un trimère de dimères.

Le site de liaison au GalNAc est formé d’une poche étroite à l’interface entre deux monomères, au niveau de la grande boucle reliant les brins A et B, des extrémités des brins D et E du même monomère et de celles des brins C et F du monomère adjacent (**Figure 1.24. A.**). Un ion zinc  $Zn^{2+}$ , représenté en rose sur la **figure 1.24**, est également présent au niveau du site de liaison. Il a été uniquement observé dans certains lots de protéines commerciaux et est probablement dû à une exposition des animaux vivants aux ions métalliques, qu’ils ont la capacité d’accumuler. La présence accidentelle de ce cation permet néanmoins de favoriser des contacts intermoléculaires, à l’origine de la modification du groupe d’espace et de l’amélioration de la résolution.

Huit liaisons hydrogène sont observées entre la protéine et son ligand. L’oxygène axial O4 établit trois liaisons hydrogène avec les chaînes latérales des acides aminés Arg 63, Trp 83 et Asp 55 du monomère adjacent, ce qui explique la préférence d’HPA pour la configuration *D-galacto-*. L’Arg 63 est également liée aux oxygènes O5 du cycle du GalNAc et O6. Ce dernier est dirigé à l’intérieur du site et interagit avec les chaînes latérales des résidus Asn 61 et Asp 55 du second monomère. L’oxygène O3 forme des liaisons hydrogène avec l’oxygène de la chaîne principale de l’acide aminé Gly 24 et pont, via une molécule d’eau, la chaîne latérale de l’Arg 25. L’oxygène du carbonyle du groupement *N*-acétylé du GalNAc est lié à l’atome d’azote de la chaîne principale de la Gly 24. En revanche, le groupement méthyle du GalNAc n’est impliqué dans aucun contact hydrophobe. L’His 84 établit quant à elle un

contact important avec le C1 du GalNAc et favorise ainsi la fixation de l'anomère  $\alpha$  par rapport au  $\beta$  (**Figure 1.24. B.**). Des contacts de type Van der Waals ont également lieu entre le groupement  $\text{CH}_2$  du C6 du ligand et la Tyr 89 du monomère voisin.



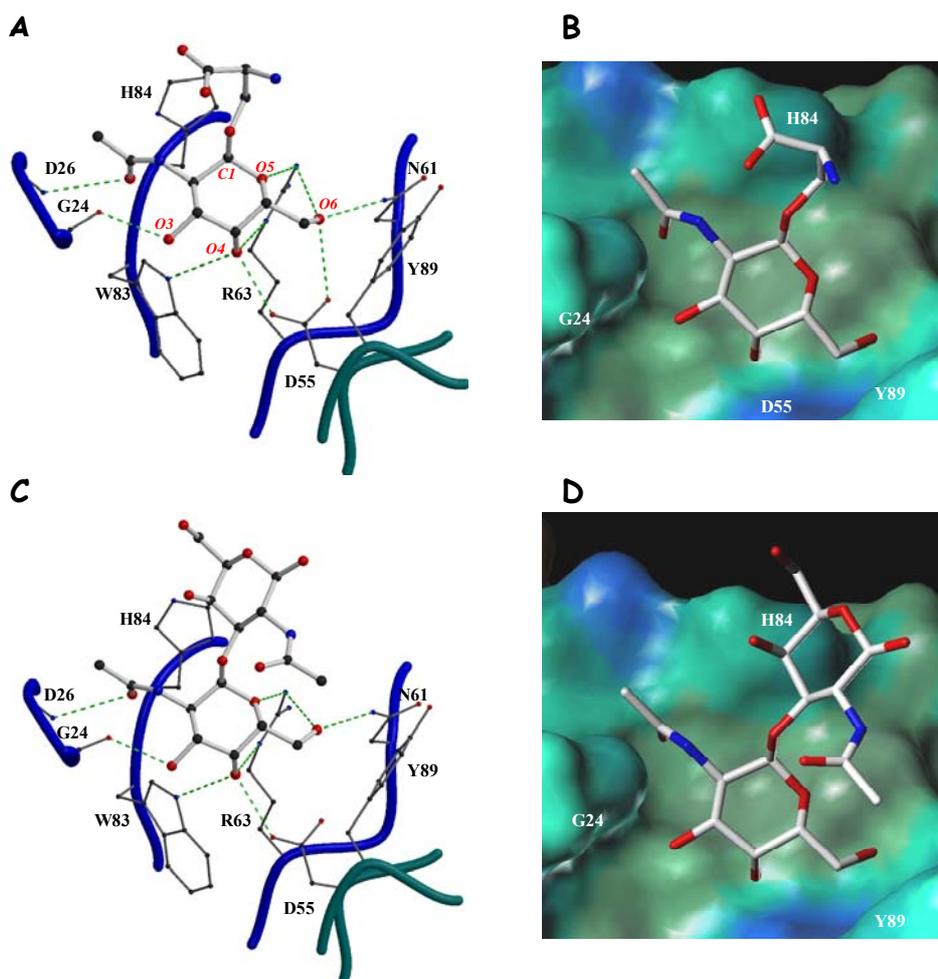
**Figure 1.24. Structure cristallographique du complexe entre HPA et le GalNAc.**

**A.** Représentation d'une face de liaison au sucre. Le GalNAc, représenté en bâtonnets, se situe à l'interface entre deux monomères. La sphère rose symbolise l'ion  $\text{Zn}^{2+}$ . **B.** Représentation du site de liaison. Les liaisons hydrogènes sont matérialisées par des pointillés verts. La sphère rouge représente une molécule d'eau.

Chaque face de liaison présente donc trois sites de liaisons distincts à chaque extrémité de la protéine. Cette organisation particulière confère à la protéine de puissantes propriétés agglutinantes.

Les structures cristallographiques d'HPA en complexe avec les épitopes cancéreux Tn ( $\alpha\text{GalNAc-Ser/Thr}$ ) (PDB 2CGZ) et Forssman ( $\alpha\text{GalNAc1-3GalNAc-R}$ ) (PDB 2CGY) ont également été résolues au laboratoire à une résolution de 2,8 Å, dans le groupe d'espace  $\text{P6}_322$  (Lescar *et al.* 2007). Pour les deux complexes, HPA adopte la même structure hexamérique que celle décrite précédemment. Dans la structure avec l'antigène Tn, en plus des contacts hydrogène et de Van der Waals décrit dans la structure en complexe avec le GalNAc, une forte interaction hydrophobique est créée entre le carbone  $\text{C}\alpha$  du résidu sérine du ligand et la chaîne latérale de l'His 84 (**Figure 1.25. A.**). Cette sérine pointe à l'extérieur du site de liaison, ce qui autorise la reconnaissance de ligands plus longs (**Figure 1.25. B.**).

Parallèlement, dans le complexe entre HPA et l'antigène de Forssman, le disaccharide étend son extrémité réductrice à l'extérieur de la surface du ligand dans une conformation compatible avec la reconnaissance de ligands plus longs (**Figure 1.25. D.**). La seule différence par rapport au réseau de liaisons hydrogène décrit pour l'antigène Tn se situe au niveau de la conformation du C5. Le second résidu GalNAc ne forme aucune liaison hydrogène avec la protéine. Cependant, la face hydrophobique constituée par les groupements CH des C4 et C5 forme des contacts de Van der Waals avec l'His 84 (**Figure 1.25. C.**).



**Figure 1.25. Site de reconnaissance d'HPA complexée avec les antigènes cancéreux.**

**A.** Schéma du réseau de liaisons hydrogènes réalisé avec l'antigène Tn. **B.** Surface de Connoly de ce site. **C.** et **D.** Représentation pour l'antigène de Forssman. La figure est issue de (Lescar *et al.* 2007).

### 4.1.3. Occurrence des protéines similaires à HPA

En recherchant dans les bases de données, plusieurs protéines ont été identifiées comme présentant des similarités de séquences peptidiques avec HPA, élargissant ainsi cette nouvelle famille de lectines de type H (**Figure 1.26.**) (Sanchez *et al.* 2006).

	1	10	20	30	40	50	60														
HPA	RVQ	SGK	IDC	GND	AGWAKVPSDDPGRDNT	RELA	KNIT	FASPYCRPPV	VLLSITQLD	VEQS	Q	NLR	VI	ARL							
CHAI1	RAQ	TGE	IDY	GSD	SSWPKVPSDDPNKDR	REL	VKN	VTFP	PSPYCRPP	KVILSVTT	LDSDH	TK	NL	RYQ	ARL						
SLL2a	RLI	HVS	RCEM	GTSS	HRCWP.....RECD	TSS	DEA	ISFW	FPF	FENT	PKVIVS	FGML	DVDN	SH	NLR	VN	SSA				
SLL2b	RLI	HVS	RCEM	GTSS	HRCWP.....R	PCD	TSS	DEA	ISFW	FPF	FENT	PKVIVS	FGML	DVDN	SN	NLR	VN	SSA			
SLL2c	RLT	AVS	RCEM	GTSS	THRCWP.....S	QCD	TSS	DEA	ISFW	QLP	FENT	PKVIVS	FGML	DVD	STR	NLR	VN	SSA			
DiscI	VTO	VGA	DIYT	GDNC	ALNTGS.....GK	REV	VVP	VKQ	FE	FATL	PKVALN	FDQ	IDCTD	ATN	Q	TRIG	VQP				
DiscII	SVQ	VGE	VSI	GDR	SLNSGT.....GS	RTI	VRH	VK	F	PVE	FLSV	P	IV	SIG	CCK	VD	AHT	DNG	QMR	WE	GKS
	70	80	90	100																	
HPA	YS	VS	PT	GF	KAS	CY	TW	HNT	KV	YS	MS	SI	SW	IS	IE	NY	.				
CHAI1	KS	VS	PS	GF	SAS	SY	TW	HD	TI	YR	MK	LS	WL	AV	ED	.	.				
SLL2a	DD	VT	VG	GF	TL	HYS	WY	TT	IV	WN	YK	LI	WI	IA	CD	.	.				
SLL2b	DD	VT	VG	GF	TL	HYS	WY	TT	IV	WN	YK	LI	WI	IA	CD	.	.				
SLL2c	DD	VDE	EG	GF	TL	HP	NW	ANT	IV	WN	YK	LI	WI	IA	CD	.	.				
DiscI	RN	IT	TK	GF	DC	VP	YTW	NEN	KV	YS	LS	RA	DY	IA	TA	LE	.				
DiscII	EN	IT	TK	GF	DL	TF	ITW	GN	NA	VY	DL	TF	DY	VA	VE	FNN	.				

**Figure 1.26.** Alignement des séquences peptidiques d’HPA, de CH-IIL, lectine de l’escargot *C. hortensis*, des trois isogènes de SLL-2, lectine du corail *S. lochmodes* et des discoidines de l’amibe *D. discoideum*.

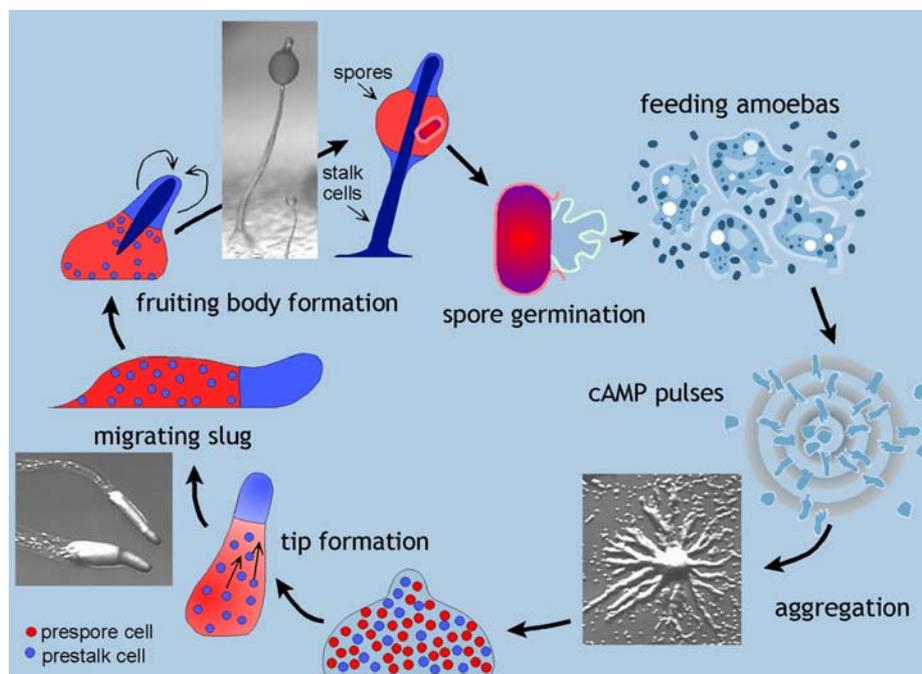
La séquence peptidique qui présente la plus similarité de séquence avec HPA (60%) correspond à la lectine spécifique du GalNAc CH-IIL, extraite de la glande à albumine de l’escargot *Cepaea hortensis* (Gerlach *et al.* 2005). HPA est également homologue avec une autre lectine d’escargot, décrite par le même groupe et appelée *Helix pomatia* lectin (Q575S3\_HELPO). Ces trois lectines ont la même taille et semblent présenter le même arrangement hexamérique puisque les cystéines impliquées dans la formation des ponts disulfure intermoléculaires sont conservées.

SLL-2, lectine isolée à partir du corail *Sinularia lochmodes*, présente également d’importantes similarités avec HPA, en termes de séquence (23% à 25% suivant les isoformes) et de taille (94 acides aminés). Le domaine C-terminal des discoidines, lectines rencontrées chez l’amibe *Dictyostelium discoideum* est également similaire à HPA. Ces deux dernières protéines seront plus largement décrites dans les paragraphes suivants.

## 4.2. Les discoidines

### 4.2.1. Généralités

*Dictyostelium discoideum* est une amibe haploïde du sol, depuis longtemps utilisée par les biologistes comme modèle eucaryote simple. Elle est capable de développer un extraordinaire cycle de vie multicellulaire (**Figure 1.27.**). Dans un environnement favorable en condition alimentaire, l'amibe reste isolée de ses consœurs (stade végétatif), se multiplie par simple division mitotique et se nourrit de bactéries par phagocytose. Dans des conditions plus difficiles, en cas de carence alimentaire notamment, les cellules sont capables de s'assembler sous l'influence d'un gradient d'AMPc pour former des agrégats multicellulaires, capables de se déplacer vers une zone plus favorable. Un corps fructifère, d'environ 2 mm de hauteur, composé d'une tige mince de cellules mortes et d'une masse de spores rassemblées dans une boule de mucus prend alors naissance. La position élevée au sommet de la tige facilite leur dispersion et la colonisation de nouveaux milieux. Lorsque les conditions redeviennent moins hostiles, les spores germent pour produire à nouveau des amibes qui débudent un nouveau cycle végétatif (Annesley and Fisher 2009).



**Figure 1.27.** Développement du cycle de vie de l'amibe *Dictyostelium discoideum*.

L'image est issue de [http://www.lifesci.dundee.ac.uk/groups/pauline\\_schaap/organism.htm](http://www.lifesci.dundee.ac.uk/groups/pauline_schaap/organism.htm).

Le chercheur américain Steven Rosen démontra pour la première fois dans sa thèse de doctorat (Université de Cornell, 1972) que les cellules de *D. discoideum* sont capables d'agglutiner des érythrocytes de moutons lorsqu'elles se différencient en l'absence de nutriments (Rosen *et al.* 1973). Cette hémagglutination n'est pas due à une lectine présente à la surface de *D. discoideum* comme initialement proposé mais à deux aujourd'hui nommées les discoidines I et II (DiscI et DiscII) (Frazier *et al.* 1975), qui présentent 48% d'identité de séquence. Ces protéines présentent des similarités mais sont différentes sur de nombreux points comme la mobilité électrophorétique, la composition en acides aminés, la dépendance au calcium, la spécificité d'hémagglutination envers des érythrocytes de différentes espèces et la distribution spatio-temporelle lors du développement (**Tableau 1.4.**) (Frazier *et al.* 1975 ; Berger and Armant 1982 ; Barondes *et al.* 1983).

	DiscI	DiscII
Spécificité	GalNAc>Galactose	GalNAc~Galactose
Poids moléculaire (Da)	28258,5	28574,7
Point isoélectrique	6,20	6,65
Nombre d'acides aminés	253	257
Coefficient d'extinction	1,50	1,59
Dépendance au calcium	oui	non
Phase d'expression	9 à 12 H après l'induction	Après 12 H
Distribution	Cellules du manteau de la limace et de la tige du corps fructifère	Accumulation dans les vésicules de prés pores
Spécificité d'hémagglutination	Hématies de lapin > mouton	Hématies de mouton > lapin

**Tableau 1.4. Comparaison entre les propriétés biochimiques des DiscI et DiscII.**

#### 4.2.2. Spécificité et affinité

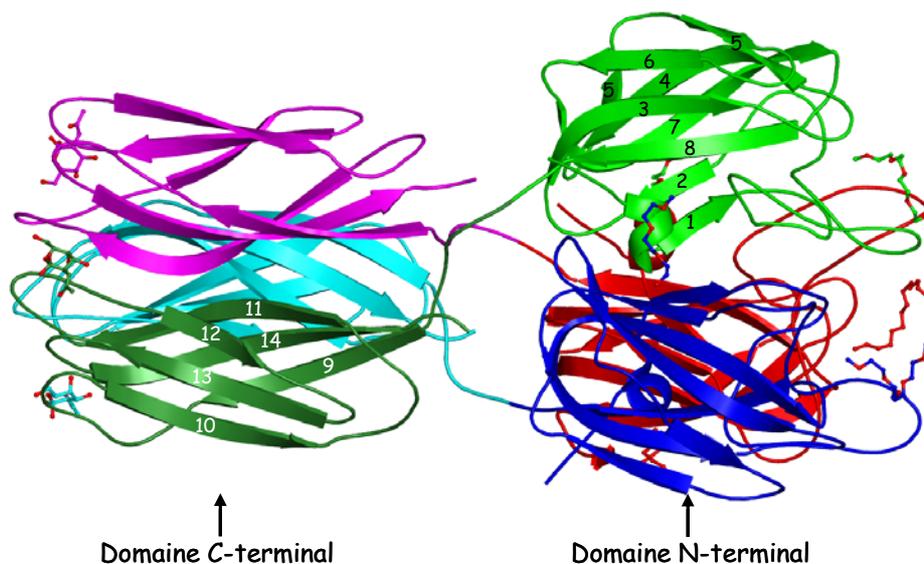
Les gènes codant pour les lectines DiscI et DiscII ont été clonés au laboratoire directement à partir du génome de l'amibe. L'expression de ces deux protéines sous forme recombinante, dans la bactérie *Escherichia coli*, a alors pu être optimisée et leur purification mise au point. Les discoidines forment des trimères en solution. Ces résultats ont fait l'objet

du travail de doctorat de Karoline Saboia Aragão (2008) avec laquelle j'ai collaboré lors de la caractérisation de la DiscI. Ces travaux seront décrits dans un chapitre de la partie Résultats et ont mené à une publication jointe en annexe.

De manière similaire à HPA, l'interaction entre DiscII et les résidus glycaniques GalNAc et  $\beta$ -D-galactoside méthylé ( $\beta$ MeGal) a été quantifiée par microcalorimétrie isotherme de titration. Les constantes de dissociation s'élèvent respectivement à  $1,15 \cdot 10^{-3}$  M et  $0,95 \cdot 10^{-3}$  M, ce qui est environ dix fois plus faible que celle mesurée entre HPA et le GalNAc. Cette liaison est fortement dirigée par l'enthalpie avec des valeurs de  $\Delta H$  de -20,3 kJ/mol et -24 kJ/mol, respectivement, néanmoins contrebalancées par une entropie défavorable  $T\Delta S$  de -3,5 et -6,7 kJ/mol, respectivement (Aragao *et al.* 2008).

#### 4.2.3. Caractérisation structurale

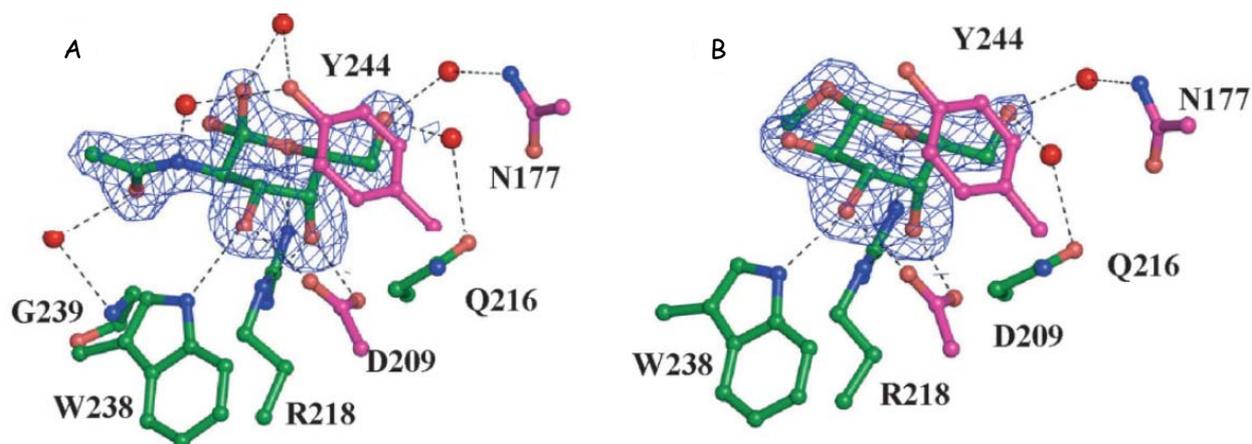
Les structures cristallographiques des DiscI et DiscII natives et complexées avec différents sucres ont été résolues au laboratoire. Les structures obtenues pour DiscI seront détaillées dans la partie Résultats. DiscII a été cristallisée sous sa forme sauvage, directement purifiée à partir de *Dictyostelium discoideum* et sous forme recombinante. La structure de la protéine recombinante a été résolue à 1.8 Å (PDB 2VM9) grâce à un dérivé mercure et celle de la protéine sauvage à 2.4 Å (PDB 2VME) (Aragao *et al.* 2008). Chaque monomère est composé de deux domaines distincts, le domaine N-terminal (1-155) et le C-terminal (163-257), séparés par un bras espaceur court. Le domaine N-terminal, classiquement appelé domaine discoidine (DS), est retrouvé dans de nombreuses protéines notamment impliquées dans l'agrégation ou la signalisation cellulaire. Il est formé d'un sandwich  $\beta$  composé de deux feuillets  $\beta$  antiparallèles de cinq ( $\beta_1$  17-20,  $\beta_2$  63-66,  $\beta_3$  71-79,  $\beta_4$  87-96,  $\beta_5$  102-104) et de trois brins  $\beta$  ( $\beta_6$  121-125,  $\beta_7$  129-142,  $\beta_8$  146-153) assemblés l'un contre l'autre. Le domaine C-terminal présente un repliement en sandwich  $\beta$  caractéristique des lectines de type H composé de six brins antiparallèles ( $\beta_9$  163-170,  $\beta_{10}$  183-190,  $\beta_{11}$  200-210,  $\beta_{12}$  220-224,  $\beta_{13}$  230-237,  $\beta_{14}$  243-254). Il est responsable de l'activité hémagglutinante de la lectine et est spécifique des résidus Gal et GalNAc. Les deux domaines s'entrelacent autour de leur bras espaceur de huit acides aminés pour former un trimère (**Figure 1.28.**).



**Figure 1.28. Structure tridimensionnelle du trimère de DiscII.**

Chaque monomère est représenté par une couleur différente (vert, bleu, rouge). Un résidu GalNAc est visible au niveau du site de reconnaissance du domaine C-terminal.

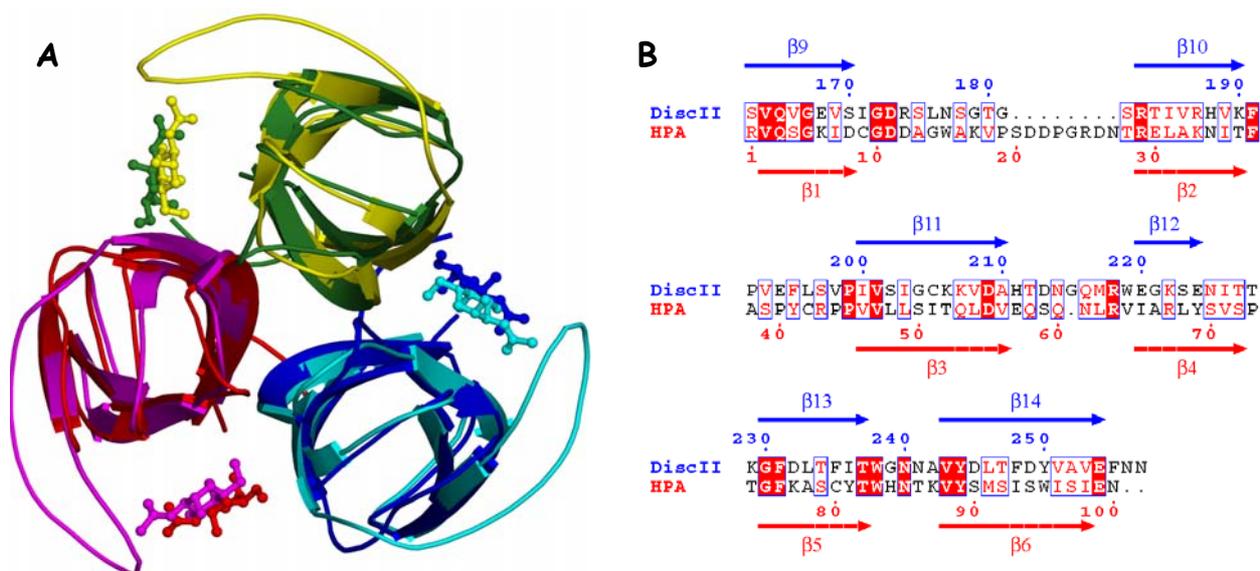
Les structures cristallographiques de DiscII complexée avec les résidus GalNAc (**Figure 1.29. A**) et  $\beta$ MeGal (**Figure 1.29. B**) révèlent la présence d'un ligand par monomère du côté C-terminal. Le site de fixation est à l'interface entre deux monomères et nécessite la formation du trimère. Dans les deux complexes, plusieurs liaisons hydrogène directes sont observées entre la protéine et le ligand : le groupement hydroxyle O3 avec un oxygène de l'Asp 209 et l'azote du Trp 238 appartenant au monomère adjacent, l'O4 avec l'autre oxygène de l'Asp 209 et l'azote NE de l'Arg 218 du monomère voisin et l'O5 avec l'azote NH1 de l'Arg 218 du monomère adjacent. Le groupement hydroxyle O6 forme des liaisons hydrogène avec un azote de l'Asn 177 (monomère adjacent) et un oxygène de la Gln 216 via deux molécules d'eau. L'atome d'azote du groupement *N*-acétyle du GalNAc forme également, via une molécule d'eau, une liaison hydrogène avec le groupement hydroxyle de la Tyr 244 du monomère voisin. Les orientations équatoriales et axiales des hydroxyles O3 et O4 respectivement sont décisives pour la reconnaissance puisqu'elles permettent la formation de liaisons hydrogène cruciales avec l'Asp 209, l'Arg 218 et le Trp 238, acides aminés strictement conservés et responsables de la préférence des lectines de type H pour la configuration *D-galacto*.



**Figure 1.29.** Site de reconnaissance de DiscII complexée avec les résidus GalNAc (A) et  $\beta$ MeGal (B).

#### 4.2.4. Similarité avec HPA

Les trimères d'HPA et du domaine C-terminal de DiscII sont très similaires aussi bien au niveau des séquences peptidiques que de la structure cristallographique (**Figure 1.30.**). La principale différence réside dans la longueur et la séquence de la première boucle de surface. Cette boucle présente une insertion de huit acides aminés chez HPA, dont trois (Gly 24, Arg 25 et Asp 26) sont impliqués dans le réseau de liaisons hydrogène entre la protéine et le GalNAc. Cette boucle a un rôle certain dans la spécificité des lectines de type H et pourrait expliquer la meilleure affinité d'HPA pour le GalNAc en comparaison avec DiscII. De plus, DiscII n'a jamais été testé comme marqueur de cellules cancéreuses. Il serait cependant envisageable que cette boucle soit impliquée dans le caractère prédictif de la lectine dans le diagnostic de cancers à haut pouvoir métastatique.



**Figure 1.30.** A. Superposition du domaine C-terminal de DiscII (vert, rouge, bleu) avec HPA (jaune, magenta, cyan). B. Alignement de séquence basé sur la superposition structurale.

### 4.3. SLL-2 : lectine du corail *Sinularia lochmodes*

Les cnidaires, dont les coraux marins font partie, ont la capacité de former des symbioses avec des dinoflagellates unicellulaires du genre *Symbiodinium*. Cette algue photosynthétique nourrit l'hôte et reçoit en retour une protection et les sources d'azote nécessaires à sa survie. L'association corail/dinoflagellates est indispensable pour la préservation des barrières de coraux, trop souvent endommagées par les phénomènes de blanchiment, perte de leur algue symbiotique entraînant la mort de l'animal. Les lectines semblent activement participer à l'établissement de cette symbiose mais les mécanismes cellulaires et moléculaires à l'origine de cette interaction restent toujours mal connus (Wood-Charlson *et al.* 2006). Leurs études ont donc un impact environnemental de premier ordre pour tenter d'enrayer la disparition massive des barrières de corail.

SLL-2 est une lectine, spécifique du D-galactose localisée dans les nématocystes du corail *Sinularia lochmodes*, véritable usine chimique capable de paralyser ses victimes. Cette lectine a également pu être identifiée à la surface du dinoflagellate symbiotique à l'intérieur du corail (Jimbo *et al.* 2000). Son rôle dans l'établissement de la symbiose n'est toujours pas clairement défini. SLL-2 semble, néanmoins, être capable d'influencer la physiologie des

---

algues symbiotiques en culture, les faisant évoluer d'un stage motile vers un stade agrégé, généralement rencontré au moment de leur intégration par le corail. Aucune action de la lectine n'a cependant été constatée sur des algues non symbiotiques. Localisée dans les nématocystes de l'animal, cette lectine pourrait également servir de sonde. Une fois l'algue intégrée par le corail, la lectine serait relarguée par l'animal permettant l'établissement d'une symbiose spécifique entre une algue et un corail, de manière similaire à ce qui est déjà observé dans les symbioses *Rhizobium*-légumineuses (Koike *et al.* 2004). Ces considérations restent encore hypothétiques à ce jour.

Le gène codant pour SLL-2 a été identifié, par la technique de RT-PCR, à partir de protéine directement extraite de l'animal et purifiée par différentes techniques de chromatographies. Trois gènes différents codant pour trois sous-unités protéines, SLL-2a, -b et -c ont pu être isolés. Ces sous-unités partagent plus de 78% d'identité de séquences les unes en comparaison avec les autres. Un site de *N*-glycosylation a également pu être identifié en position 60. La masse moléculaire de chaque sous-unité est d'environ 10,8 kDa pour la forme déglycosylée et de 13,2 kDa et 16,95 kDa pour les formes glycosylées majeures. Celle de la protéine globale estimée par chromatographie d'exclusion stérique à 122 kDa semble indiquer que SLL-2 serait composée de neuf sous-unités (Jimbo 2005). Il est néanmoins primordial de conserver à l'esprit que le poids moléculaire déterminé par chromatographie d'exclusion stérique n'est qu'une estimation et dépend fortement de la forme et de la glycosylation de la protéine.

SLL-2 est homologue à HPA (25 % de similarité) et au domaine C-terminal des discoidines (33 % de similarité avec DiscI), complétant ainsi la famille des lectines de type H.

## 5. Objectifs du travail de thèse

Le développement d'outils biotechnologiques pour des applications thérapeutiques devient d'un grand intérêt. Le marquage spécifique en histopathologie pourrait notamment faciliter le diagnostic du cancer, améliorant ainsi la survie des patients. Cette approche implique notamment l'identification rapide du caractère métastatique de cellules cancéreuses. Dans ce contexte, l'utilisation de certaines lectines capables de reconnaître spécifiquement les glycanes localisés à la surface des cancers à faible pronostic vital offre donc d'importantes perspectives dans la lutte contre le cancer.

La lectine HPA, de l'escargot *Helix pomatia*, représente un marqueur capable de cibler spécifiquement les cellules à caractère métastatique. Cependant, la variabilité observée dans les résultats selon la méthode histochimique utilisée limite l'intérêt de l'utilisation de cette lectine comme outil d'investigation de routine en oncologie (Brooks *et al.* 1996). Le manque de reproductibilité de certains marquages peut être attribué aux variations dans les lots d'HPA obtenus à partir d'extraits d'escargots. Au laboratoire, des contaminations d'HPA par des métaux (zinc) et par des protéines humaines ont été observées. De plus, l'analyse électrophorétique montre la présence de plusieurs isolectines sur les gels SDS-PAGE. Pour d'une part résoudre les problèmes de pureté et de reproductibilité et d'autre part obtenir des quantités suffisantes de protéines pour s'approcher d'un outil de diagnostic, il est nécessaire d'utiliser une approche de production en système hétérologue. Le gène codant pour HPA a pu être cloné (Sanchez *et al.* 2006), mais à ce jour aucune équipe n'a réussi à produire la protéine active sous forme recombinante malgré les nombreux essais dans la bactérie *Escherichia coli*, la levure *Pichia pastoris*, ou le système baculovirus/cellules d'insectes. L'utilisation des lectines similaires présentes dans d'autres organismes est donc d'un grand intérêt afin de résoudre les problèmes constatés chez HPA.

Les discoidines (DiscI et DiscII) de l'amibe *Dictyostelium discoideum* et SLL-2 du corail *Sinularia lochmodes* présentent des similarités suffisantes avec HPA pour être de bonnes candidates pour une approche d'ingénierie protéique. A ce jour, aucun protocole de production de SLL-2 sous forme recombinante n'a été mis au point. Sa structure cristallographique est toujours inconnue et sa caractérisation biochimique reste relativement sommaire. En revanche, l'expression de DiscII dans *E. coli* et sa purification par des techniques de chromatographies classiques ont été réalisées durant le travail de thèse de K. S. Aragão au laboratoire. La résolution de sa structure cristallographique (Aragao *et al.* 2008) a permis de mettre en évidence une boucle de spécificité chez les lectines de type H. Aucun test de marquage sur cellules cancéreuses n'a été effectué. Il est donc très difficile de conclure quant à l'importance éventuelle de celle-ci chez HPA pour le diagnostic de cancers à faible pronostic vital.

Ces travaux de thèse se divisent donc en différents objectifs :

- Terminer la caractérisation structurale de DiscI initiée durant la thèse de K. S. Aragão au laboratoire.

- Tester DiscI et DiscII comme marqueurs de cancers à faible pronostic vital selon un protocole similaire à celui utilisé pour HPA. Ce travail s'effectue en collaboration avec le Dr. Miriam Dwek de l'université de Westminster à Londres.

- Modifier, par mutagenèse, le gène codant pour DiscII au niveau de la boucle de spécificité de manière à lui conférer la spécificité d'HPA. La protéine chimérique ainsi construite aurait les mêmes propriétés qu'HPA pour le diagnostic de cellules cancéreuses métastatiques, tout en étant produite, en forte quantité, sous forme recombinante dans *E. coli*.

- Mettre au point l'expression sous forme recombinante de SLL-2 et sa purification, grâce à des techniques chromatographiques classiques.

De plus, la ou les structure(s) glycanique(s) complexe(s) reconnue(s) par HPA à la surface des cellules cancéreuses n'est ou ne sont toujours pas caractérisée(s) à ce jour. Leur identification permettrait cependant d'améliorer grandement la compréhension du processus de formation des métastases. Elle pourrait également être à l'origine d'importantes avancées dans le domaine de la prévention de l'apparition de cancers avec le développement d'outils de lutte comme les vaccins. Ainsi, le dernier objectif de cette thèse est également d'identifier, à l'aide de techniques perfectionnées comme la spectrométrie de masse, ce ou ces glycanes, directement extrait(s) et purifié(s) à partir de cellules cancéreuses en culture. Cette étude est réalisée en collaboration avec le Dr. Miriam Dwek de l'université de Westminster à Londres et le Dr. Bérange Tisot de l'Imperial College de Londres. Ces résultats très préliminaires seront présentés en annexe de ce manuscrit.

# **Chapitre 2**

## **Matériels et méthodes**



---

# Chapitre 2

## Matériels et Méthodes

---

### 1. Utilisation des lectines pour le marquage de cellules cancéreuses

Le marquage des lignées cellulaires cancéreuses ainsi que la révélation par microscopie confocale ont été effectués dans le laboratoire du Dr Miriam Dwek, à l'université de Westminster à Londres.

#### 1.1. Principe de la microscopie confocale

La microscopie confocale est une technique de microscopie optique qui permet d'obtenir des images d'objet en trois dimensions. Elle est généralement couplée avec la microscopie à fluorescence. En microscopie, seuls les rayons émis par l'échantillon dans le plan focal de la lentille donnent naissance à une image nette. Ainsi, l'épaisseur d'un échantillon sera à l'origine d'une perte de résolution. La microscopie confocale présente alors l'énorme avantage de sélectionner, grâce à un diaphragme, uniquement les rayons émis dans le plan focal de la lentille. Ils sont alors détectés et intégrés par ordinateur. En positionnant le plan focal de la lentille à différents endroits de l'échantillon, plusieurs coupes d'images sont réalisées et une image globale nette de l'échantillon en 3-D peut être reconstruite (**Figure 2.1.**).

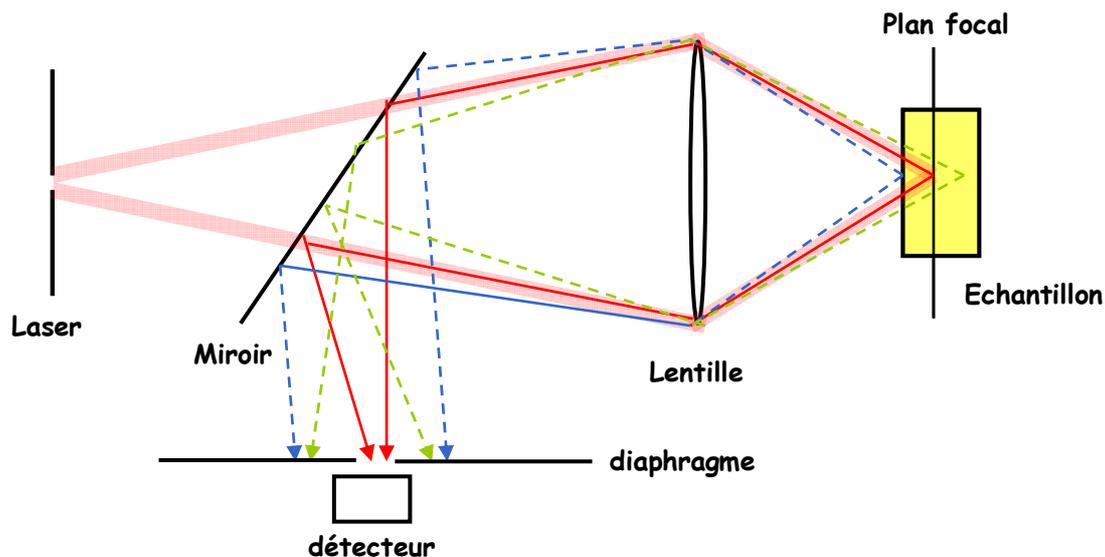


Figure 2.1. Principe du microscope confocale

## 1.2. Traitement des échantillons

### 1.2.1. Biotinylation de DiscI et DiscII

Les lectines DiscI et DiscII recombinantes et HPA commerciale (Sigma) sont biotinylées selon le protocole détaillé par le fournisseur (Pierce). Le réactif utilisé est l'acide biotinamidohexanoyl-6-aminohexanoïque couplé à un ester *N*-hydroxysuccinimide (NHS-Biotin commercialisé par Sigma), préparé extemporanément dans une solution à 10 mM dans du diméthylformamide (DMF). La solution de biotine est incubée 20 fois en excès avec la solution protéique concentrée à 2 mg.ml<sup>-1</sup> dans du PBS, pendant deux heures dans la glace. La protéine biotinylée est récupérée après dessalage sur colonne PD-10 (GE-Healthcare, en suivant les informations du fournisseur) dans le tampon de microscopie Tris-HCl 50 mM, pH 7,4, NaCl 145 mM, MgCl<sub>2</sub> 1 mM, CaCl<sub>2</sub> 1 mM. Elle est ainsi séparée de l'excès de réactif biotinylé n'ayant pas réagi.

### 1.2.2. Protocole de marquage des cellules cancéreuses

Les trois lignées cellulaires d'adénocarcinomes colorectaux utilisées dans cette étude sont décrites dans le **tableau 2.1**.

Nom lignée	Numéro ATCC	Stade tumoral	Genre/Âge/Type
SW480	CCL-228 <sup>TM</sup>	Lignée considérée comme non métastatique, établie à partir du cancer primaire.	Homme de 50 ans de type caucasien
SW620	CCL-227 <sup>TM</sup>	Lignée considérée comme métastatique, établie à partir de métastases dans les ganglions lymphatiques.	Même patient que SW480, un an plus tard
HT-29	HTB-38 <sup>TM</sup>	Lignée considérée comme très fortement métastatique.	Femme de 44 ans de type caucasien

**Tableau 2.1. Description des lignées cellulaires utilisées dans cette étude.**

Les cellules ont été fixées par du formalin, à environ 80% de confluence, dans des plaques de six puits. Elles sont tout d'abord lavées trois fois 3 minutes dans du PBS avant d'être trypsinées (trypsine de pancréas de porc commercialisée par Sigma, à une concentration de 1 mg.ml<sup>-1</sup>) 20 minutes à 37°C pour libérer les antigènes de surface sans décoller les cellules. Une saturation des sites non spécifiques est réalisée par incubation des cellules pendant 30 minutes dans une solution de BSA 5% dans du PBS. L'incubation des solutions de lectines biotinylées à 10 µg.ml<sup>-1</sup> dans le tampon de microscopie se déroule pendant une heure. La spécificité de la reconnaissance des lectines est également contrôlée par une pré-incubation de ces dernières dans une solution de GalNAc à 0,1 M pendant 30 minutes avant la mise en contact avec les cellules. L'incubation de streptavidine conjuguée au fluorophore Fluorescein-5-isothiocyanate (FITC – Zymed) à 4 µg.ml<sup>-1</sup> pendant 30 minutes sera à l'origine de la révélation sous microscope. Pour le marquage avec HPA, une solution d'HPA commerciale couplée au fluorophore Tetramethyl Rhodamine Isothiocyanate (TRITC - Sigma) à 10 µg.ml<sup>-1</sup> a également été testée. A partir de cette étape, les plaques sont conservées à l'abri de la lumière dans du papier aluminium. Les cellules sont ensuite traitées par une solution de ribonucléase A à 100 µg.ml<sup>-1</sup> (Sigma) avant le marquage du noyau par une solution à 1 µM de fluorophore To-Pro-3 (Invitrogen) dans PBS pendant 20 minutes. Chaque étape est réalisée

---

sous une légère agitation à température ambiante et est suivie de trois lavages de 3 minutes chacun dans du PBS.

### 1.2.3. Révélation par microscopie confocale

Les images sont acquises sur un microscope confocal Leica TCS SP2 et collectées à une résolution de 1064 x 1064 pixels avec une vitesse de scan de 400 Hertz (Hz). Des lasers à 543 nm, 633 nm et 480 nm ont été utilisés pour l'excitation de HPAC-TRITC (intensité 47%), de To-Pro-3 (intensité 33%) et FITC (intensité 25%), respectivement. La fluorescence d'émission a été collectée selon les spectres suivants de 565 à 640 nm pour TRITC, de 650 à 700 nm pour To-Pro-3 et de 500 à 535 nm pour FITC.

## 2. Techniques de biologie moléculaire

### 2.1. Souche et plasmides utilisés

La souche et les plasmides utilisés pour toutes les expériences de biologie moléculaire sont présentés dans le **tableau 2.2**. Les vecteurs pET-TEV, pET-M30 et pPICZ $\alpha$ -HisFlag ne sont pas des vecteurs commerciaux mais dérivés de vecteurs existants afin d'apporter à la protéine clonée de nouvelles caractéristiques biochimiques.

Les vecteurs pRSET-A, pET-TEV et pET-M30 sont sous le contrôle transcriptionnel du promoteur fort T7, basé sur la sélectivité et la puissance de l'ARN polymérase du bactériophage T7. Le vecteur pProEx-HtB est, quant à lui, sous l'action du promoteur *trc* (Brosius *et al.* 1985). Dans les deux cas, leur mode de fonctionnement est basé sur celui du promoteur *lac*, très utilisé en biologie et inductible à l'isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG), analogue non hydrolysable du lactose (Terpe 2006).

Souche ou plasmide	Particularités	Références
<b>Souche <i>E. coli</i></b>		
XL1 blue	Cellules résistantes à la tétracycline et traitées chimiquement pour être compétentes à la transformation.	Stratagene
<b>Plasmides</b>		
<b>Expression dans la bactérie <i>E. coli</i></b>		
pRSET-A	Vecteur contenant le promoteur T7, inductible à l'isopropyl- $\beta$ -D-thiogalactopyrannoside (IPTG) et portant le gène de résistance à l'ampicilline. Suivant la stratégie de clonage, la protéine recombinante peut, ou non, être exprimée avec les étiquettes polyhistidine et X-press du côté N-terminal, clivables à l'entérokinase.	Invitrogen
pProEx-Htb	Vecteur sous l'action du promoteur Trc, inductible à l'IPTG et portant le gène de résistance à l'ampicilline. La protéine recombinante est exprimée avec une queue polyhistidine du côté N-terminal, clivable sous l'action de la protéase TEV pour Tobacco Ech Virus.	Life technologies
pET-TEV	Vecteur dérivé du vecteur pET-28a dans lequel le site de coupure à la trombine a été remplacé par celui de la TEV. Vecteur sous l'action du promoteur T7, inductible à l'IPTG et portant le gène de résistance à la kanamycine. La protéine recombinante est exprimée avec une queue polyhistidine du côté N-terminal, clivable sous l'action de la protéase TEV.	N. Tarbouriech (CIBB)
pET-M30	Vecteur dérivé du pBR322, sous l'action du promoteur T7, inductible à l'IPTG et portant le gène de résistance à la kanamycine. La protéine recombinante exprimée est couplée à deux étiquettes différentes du côté N-terminal, une queue polyhistidine et la Glutathion S-Transférase (GST), clivables par la protéase TEV (Smith 2000).	G Stier (EMBL)
<b>Expression dans la levure <i>Pichia pastoris</i></b>		
pPICZ $\alpha$ -A	Vecteur sous l'action du promoteur AOX1, inductible au méthanol et portant le gène de résistance à la zéocine. La protéine recombinante est exprimée sans étiquette. Le facteur $\alpha$ permet la sécrétion de la protéine dans le milieu de culture.	Invitrogen
pPICZ $\alpha$ -HisFlag	Vecteur dérivé du vecteur pPICZ $\alpha$ -C. Deux étiquettes Flag et His <sub>6</sub> sont présentes en N-terminal de la protéine sécrétée.	

**Tableau 2.2. Récapitulatif des souches et plasmides utilisés dans la partie biologique moléculaire de cette étude.**

Le détail des cartes de restriction des différents vecteurs utilisés est représenté **Figure 2.2** et **Figure 2.3**.

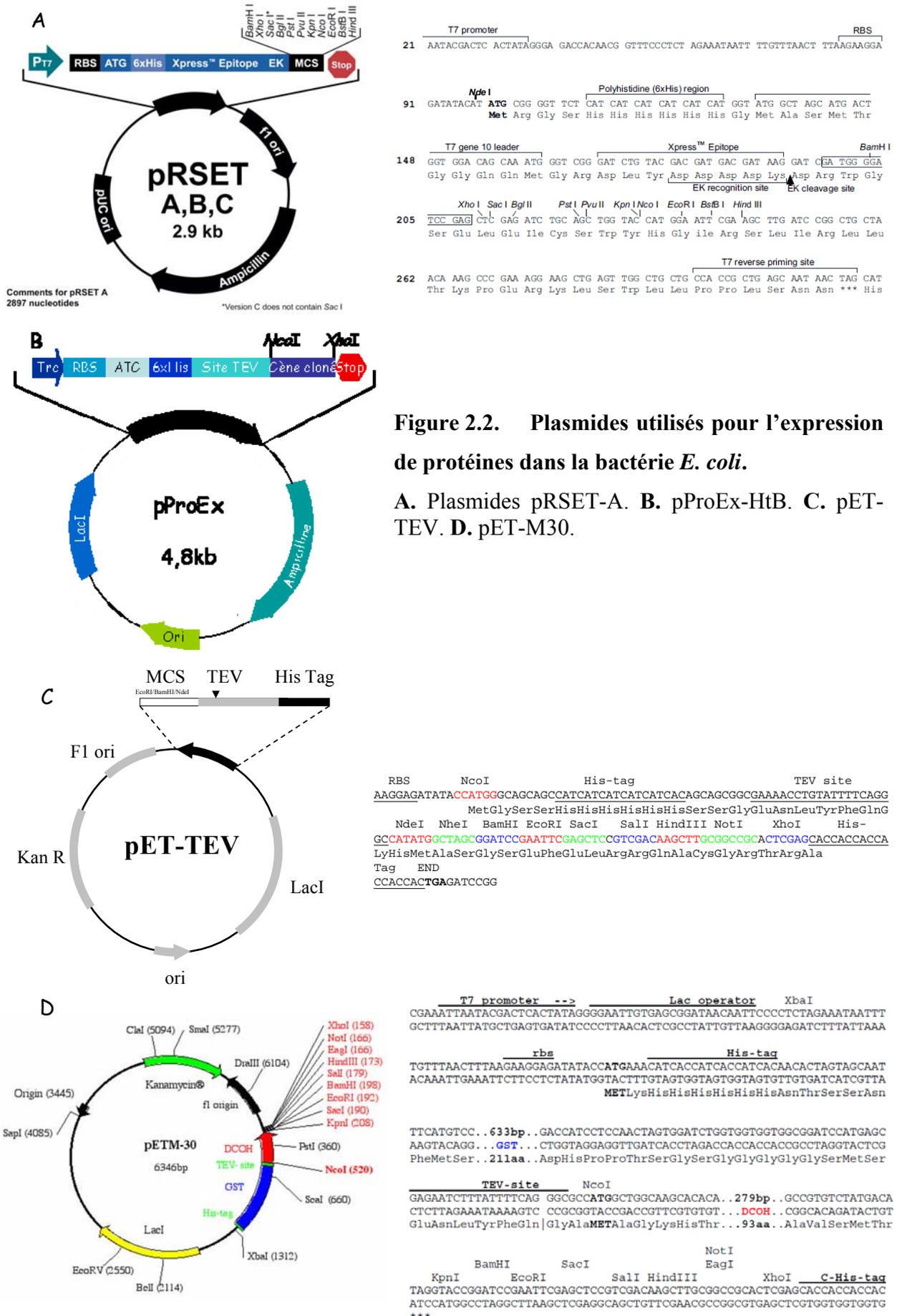
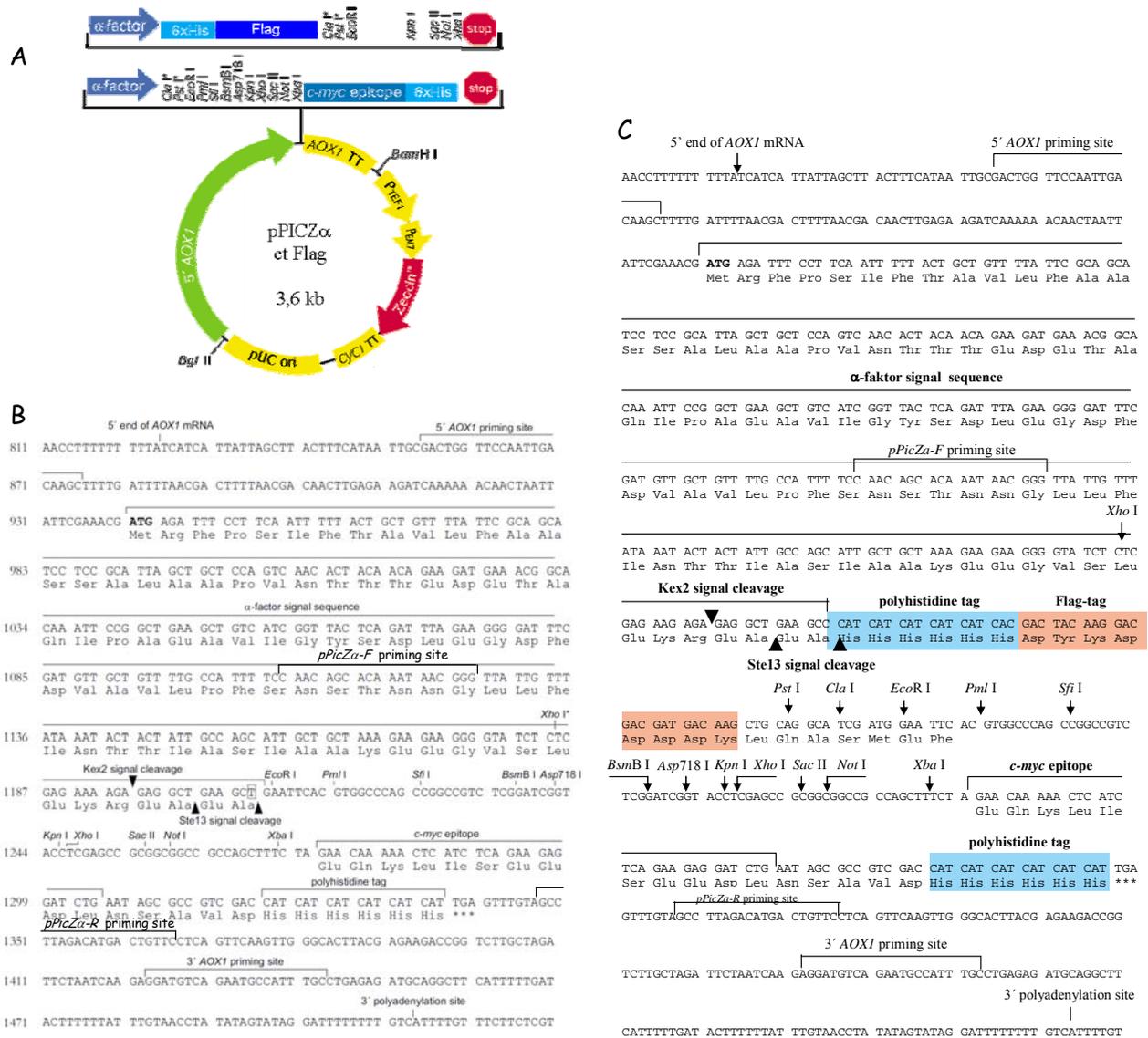


Figure 2.2. Plasmides utilisés pour l’expression de protéines dans la bactérie *E. coli*.

A. Plasmides pRSET-A. B. pProEx-HtB. C. pET-TEV. D. pET-M30.



**Figure 2.3. Plasmides utilisés pour l'expression de protéines recombinantes dans la levure *Pichia pastoris*.**

**A.** Représentation schématique du vecteur pPICZα. **B.** Carte de restriction du vecteur pPICZα-A. **C.** Carte de restriction du vecteur pPICZα-HisFlag.

Le **tableau 2.3** présente les oligonucléotides utilisés pour les différents clonages.

Amorce	Séquence
<b>Lectine de corail : SLL2a</b>	
SLL1	5'- (CT)(CT)(AGCT)CC(AGCT)TT(CT)GA(AG)AA(CT)AC(AGCT) CC-3'
SLL2	5'-GT(AGCT)(AT)(GC)(AGCT)TT(CT)GG(AGTC)ATG(CT)T (AGCT)A-3'
SLL3	5'-CCA(AGT)AT(AGCT)A(AG)(CT)TT(AG)TA(AG)TTCCA-3'
SLL4	5'-(GCT)T(AGCT)(AT)(GC)(AGCT)(AGT)T(AGCT)(AG)(GC) (AGCT)(AC)(ACT)(AGCT)(CT)T(AGCT)GA-3'
SLL5	5'-GA(CT)G(CT)(AGCT)GA(AGCT)(ACT)(AGC)(AGCT)(AT) (GC)(AGCT)(AC)(AG)(AGCT)AA-3'
SLL6	5'-(AGC)(CT)A(AGCT)A(CT)(AGCT)(AG)T(AGCT)(GC)(CT) (AGCT)(GT)(CT)(AGCT)(GT)(AGC)CCA-3'
SLL2a-NcoI	5'-TGCCCATGGCCCGTCTGATTCATGTGAGCCG-3'
SLL2a-XhoI	5'-GCCCTCGAGTTAATCACAGGCAATCCAAATCAG-3'
SLL2a-NdeI	5'-GATCTGCCATATGCGTCTGATTCATGTGAGCCG-3'
SLL2a-BamHI	5'-GCCGGATCCTTAATCACAGGCAATCCAAATCAG-3'
SLL2a-EcoRI	5'-GCGAATTCCGTCTGATTCATGTGAGCCG-3'
SLL2a-XbaI	5'-GCTCTAGATTAATCACAGGCAATCCAAATCAG-3'
SLL2a-PstI	5'-GCCTGCAGCGTCTGATTCATGTGAGCCG-3'
5'AOX1	5'-GACTGGTTCCAATTGACAAGC-3'
3'AOX1	5'-GCAAATGGCATTCTGACATCC-3'
pPICZ $\alpha$ -F	5'-CAACAGCACAAATAACGGG-3'
pPICZ $\alpha$ -R	5'-GAACAGTCATGTCTAAGGC-3'
<b>Lectine de corail : SLL2a-2</b>	
SLL2a2-for-NdeI	5'-GGCATATGGAAATGCCGAAAACCCTGG-3'
SLL2a2-rev- EcoRI	5'-CCGAATTCCTTAATCGCACGCGATCC-3'
SLL2a2-for-XhoI	5'-GGCTCGAGATGGAAATGCCGAAAACCCTGG-3'
SLL2a-2-EcoRI	5'-GCGAATTCATGCCGAAAACCCTGGGCAAACCG-3'
2SLL2a-2-XbaI	5'-GGTCTAGATTAATCGCACGCGATCCAAATCAG-3'
2SLL2a-2-PstI	5'-GCCTGCAGATGCCGAAAACCCTGGGCAAACCG-3'
<b>Lectine de <i>Dictyostelium discoideum</i> : DiscI et DiscII</b>	
DisI NcoI	5'-GTGCCATGGCTACCCAAGGTTTAGTTC-3'
DisI NterXhoI	5'-CACTCGAGTTATGGTTGAGTGTAGAATTCAC-3'
DisI CterNcoI	5'-GTGCCATGGGCTCAGTCACTCAAGTTGG-3'
DisI XhoI	5'-GACTCGAGTTATCCAAAGCGGTAGC-3'
DisI mutforA	5'-GAATTCTACTCAACCAGTACAATAACTCGAGGCATG- 3'
DisI mutrev	5'-TGGTTGAGTGTAGAATTCACATCTTAATGAAATGTG-3'
DisI mutforB	5'- GAATTCTACTCAACCAGTACAAAGCTCATAACTCGAG GCATG-3'

Dis2NcoI	5'-GTGCCATGGCCGTTCCAGCTGGTTC-3'
Dis2revXhoI	5'-GACTCGAGTTATAATGCATATAATTCCCAACGAAG-3'
Dis2forNcoI	5'-GACCATGGCATCAGTCCAAGTTGG-3'
Dis2XhoI	5'-GACTCGAGTTAATTATTAATTC AACAGCAAC-3'
Dis2revCterXhoI	5'-CACTCGAGTTAAAATTC AACAGCAACATAATC-3'
<b>Mutagenèse de la lectine de <i>Dictyostelium discoideum</i> : DiscII</b>	
Dischimfor	5'-AACAGTGGTACTGGTTCGACGACCCAGGCAGAGACAAC TCACGTACGATTGTTTCG-3'
Dischimrev2	5'-ACCAGTACCACTGTTAAGAGATCTATCAC-3'
Mutdisfor	5'-GGTGAAGTTTCAATTGGTAACGATGCCGGCTGGGCCAAA GTCCCGTCTGACGACCCAG-3'
Mutdisrev	5'-ACCAATTGAAACTTCACCAACTTGGACTG-3'
LDD2mG247Hfor	5'-GGTTTTGATTAACTTTTATTACATGGCATAATAATGCA GTTTATG-3'
LDD2mG247Hrev	5'-CATAAACTGCATTATTATGCCATGTAATAAAAAGTTAAAT CAA AACC-3'
LDD2mI171Sfor	5'-GGTGAAGTTTCATCTGGTAACGATGCCGGCTGGG-3'
LDD2mI171Srev	5'-CCCAGCCGGCATCGTTACCAGATGAAACTTCACC-3'
LDD2mF243Sfor	5'-GGTTTTGATTAACTTCTATTACATGGCATAATAATG-3'
LDD2mF243Srev	5'-CATTATTATGCCATGTAATAGAAGTTAAATGAAAACC-3'
<b>Construction et mutagenèse de la protéine hybride : DiscII<sup>Nter</sup>-HPA</b>	
LDD2rev	5'-CATTACCAGTTAAAAGTTATTCAAATCCACGCGTACA GTCGGGAAAATC-3'
HPAfor	5'-GTTAAAAGTTATTCAAATCCACGCGTACAGTCGGGA AAAATCG-3'
HybR163SFor	5'-GTTAAAAGTTATTCAAATCCAAGCGTACAGTCGGG-3'
HybR163SRev	5'-CCCGACTGTACGCTTGGATTTGAATAACTTTTAAAC-3'

**Tableau 2.3. Récapitulatif des amorces utilisées pour cette étude**

Les sites de restriction sont soulignés. Les codons introduisant la mutation dans la séquence originale sont en italique. Les paires de bases écrites entre parenthèses représentent les différentes possibilités pour un même emplacement dans les amorces dégénérées.

## 2.2. Stratégies de clonage

### 2.2.1. Clonage des domaines séparés des DiscI et DiscII

#### 2.2.1.1. Les domaines N- et C-terminaux de DiscI

Le clonage du domaine N-terminal de DiscI a été réalisé selon une méthode de clonage classique.

Afin d'optimiser la production sous forme recombinante dans *E. coli*, trois constructions de longueurs différentes ont été produites pour ce domaine :

- DiscI-Nter-P153 de Met 1 à Pro 153
- DiscI-Nter-Q155 de Met 1 à Gln 155
- DiscI-Nter-S157 de Met 1 à Ser 157

A partir du plasmide pProEx-DiscI, la région codante à DiscI-Nter a été amplifiée par la polymérase HotStar HiFidelity (Qiagen), selon les indications du fournisseur, avec les amorces Dis1NcoI/Dis1NterXhoI. Les régions codantes DiscI-Nter-Q155 et DiscI-Nter-S157 ont été obtenus par mutagenèse dirigée sur la base des instructions fournies avec la polymérase PfuUltraII fusion (Stratagene). Les amorces utilisées sont respectivement Dis1mutforA/Dis1mutrev et Dis1mutforB/Dis1mutrev. Après digestion enzymatique des produits PCR et du vecteur vide pProEx-HtB avec *NcoI* et *XhoI*, une ligation avec le kit TAKARA génère les plasmides pProEx-DiscINter, pProEx-DiscINterQ155, pProEx-DiscINterS157. Après vérification par séquençage, ils sont transformés dans les souches d'expression.

Le clonage du domaine C-terminal de DiscI a été réalisé selon un protocole similaire à celui employé pour DiscI-Nter. Les amorces utilisées étaient Dis1CterNcoI/Dis1XhoI.

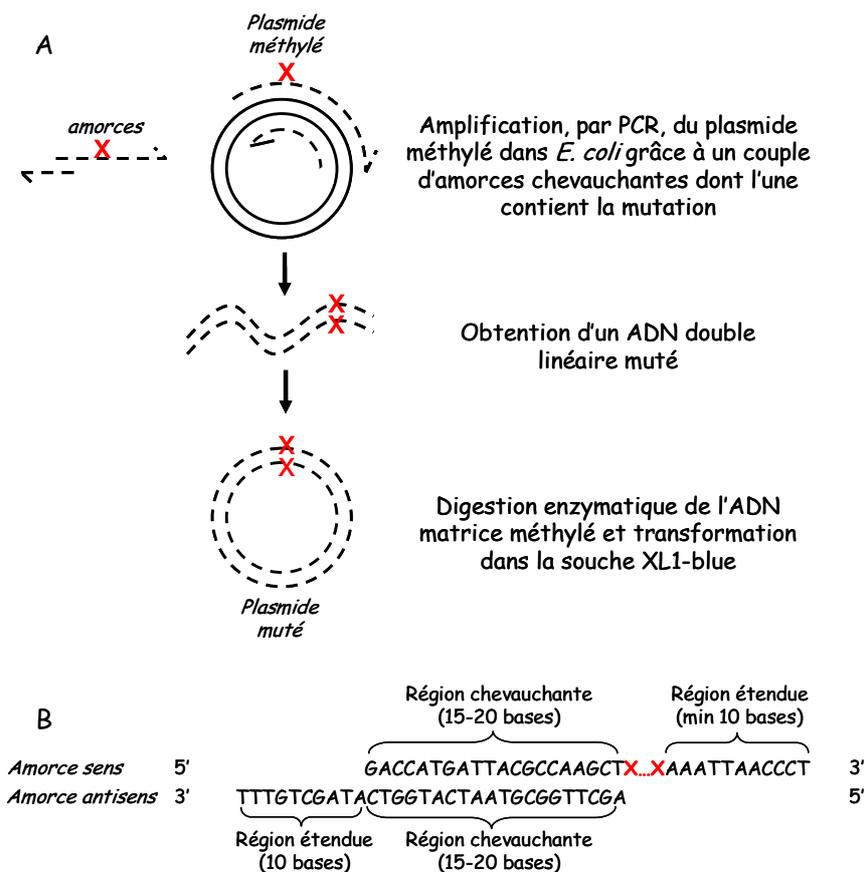
#### 2.2.1.2. Les domaines N- et C-terminaux de DiscII

Parallèlement, les domaines N- et C-terminaux de DiscII ont été clonés séparément selon un protocole identique à celui décrit précédemment. Le domaine N-terminal a été cloné de la Met 1 à la Leu 154. Pour le domaine C-terminal, deux constructions ont été réalisées de la Pro 162 à l'Asn 257 et de la Pro 162 à la Phe 255, respectivement dénommées DiscII-Cter-N257 et DiscII-Cter-F255. A partir du plasmide pProEx-DiscII, les régions codantes DiscII-Nter, DiscII-Cter-N257 et DiscII-Cter-F255 ont été amplifiées par la polymérase HotStar HiFidelity (Qiagen, selon les informations du fournisseur), avec les couples d'amorces Dis2NcoI/Dis2revXhoI, Dis2forNcoI/Dis2XhoI et Dis2forNcoI/Dis2revCterXhoI, respectivement. Les produits PCR obtenus ont été clonés dans le vecteur pProEx-HtB pour générer, respectivement, les plasmides pProEx-DiscIINter, pProEx-DiscIICterN257 et pProEx-DiscIICterN255, vérifiés par séquençage avant d'être transformés dans les souches d'expression.

## 2.2.2. Mutagenèse dirigée de DiscII

### 2.2.2.1. Insertion ou mutation d'un nombre important de bases

L'insertion ou la mutation d'un nombre important de bases nucléotidiques dans un gène a été réalisée selon la méthode GeneTailor™ Site-Directed Mutagenesis System, commercialisée par Invitrogen. Le principe repose sur celui développé pour de la mutagenèse classique, à l'exception du design des amorces. Il est brièvement résumé dans la **figure 2.4**.



**Figure 2.4. Principe de la mutagenèse dirigée GeneTailor™ Site-Directed Mutagenesis System, développée par Invitrogen.**

**A.** Schéma résumé du principe. **B.** Particularités des amorces utilisées. Les bases mutées sont représentées en rouge sur le schéma.

Six mutants de DiscII ont été construits :

- DiscII-boucleHPA : basé sur l'alignement de séquence protéique, ce mutant possède la séquence de DiscII dans laquelle ont été insérés, entre Gly 181 et Ser 182, les huit acides aminés formant la boucle de spécificité d'HPA (Ser-Asp-Asp-Pro-Gly-Arg-Asp-Asn).

---

- DiscII-Cter-boucleHPA correspond au domaine C-terminal, de la Pro 162 à l'Asn 257 du mutant précédent.

- DiscII-boucleHPAmuté : à partir de DiscII-boucleHPA, les neuf acides aminés de DiscII précédant la boucle d'HPA sont mutés par ceux présents dans la séquence d'HPA.

- DiscII-Cter-boucleHPAmuté correspond au domaine C-terminal du mutant précédent.

- DiscII-boucleHPAmuté-G247H : à partir de DiscII-boucleHPAmuté, la Gly 247 de DiscII est mutée en His, comme dans la séquence d'HPA.

- DiscII-boucleHPAmuté-G247H-I171S-F243S : à partir du mutant précédent, les acides aminés Ile 171 et Phe 243 de DiscII sont mutés par des résidus Ser, de manière à former un nouveau réseau de liaisons hydrogène remplaçant le pont disulfure intramoléculaire d'HPA.

La région codante du mutant DiscII-boucleHPA a été amplifiée, à partir de la matrice pProEx-DiscII, par la polymérase PfuTurbo® (Stratagène), avec le couple d'amorces Dischimfor/Dischimrev2. Après digestion enzymatique par *DpnI* (Promega) de l'ADN matriciel méthylé, le produit PCR obtenu a été cloné dans le vecteur pProEx-HtB pour générer le plasmide pProEx-DiscII-boucleHPA, vérifié par séquençage et transformé dans des souches d'expression. Le domaine C-terminal de DiscII-boucleHPA a également été cloné à partir de la matrice précédente en employant le protocole décrit pour le clonage de DiscII-Cter et permet ainsi de générer le plasmide pProEx-DiscII-Cter-boucleHPA.

Les mutants DiscII-boucleHPAmuté et DiscII-Cter-boucleHPAmuté sont construits selon un protocole identique en utilisant la polymérase PfuUltraII fusion (Stratagene) et les amorces Mutdisfor/Mutdisrev. Les plasmides pProEx-DiscII-boucleHPAmuté et pProEx-DiscII-Cter-boucleHPAmuté sont ainsi générés.

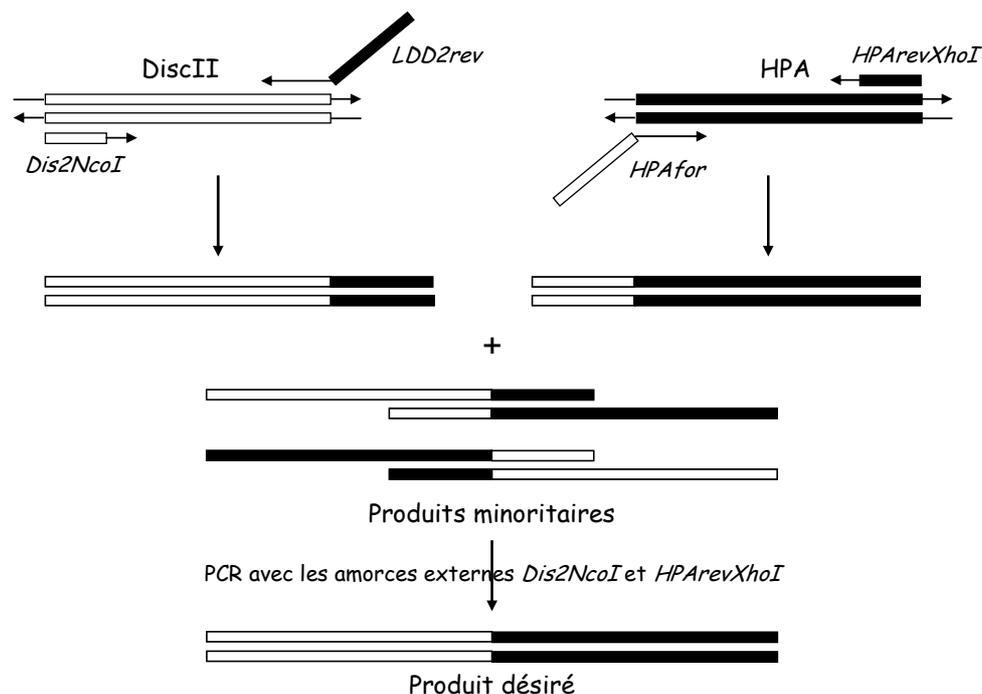
Le mutant DiscII-boucleHPAmuté-G247H est construit selon un protocole de mutagenèse classique à partir de la matrice pProEx-DiscII-boucleHPAmuté et du couple d'amorces LDD2mG247Hfor/ LDD2mG247Hrev. Le plasmide pProEx-DiscII-boucleHPAmuté-G247H est généré.

Le mutant DiscII-boucleHPAmuté-G247H-I171S-F243S est construit à partir du plasmide précédent en deux étapes de réaction PCR, chacune introduisant une mutation, avec les amorces suivantes, LDD2mI171Sfor/ LDD2mI171Srev et LDD2mF243Sfor/ LDD2mF243Srev.

### 2.2.2.2. Construction d'une protéine hybride par PCR chevauchante

Une protéine hybride est également formée dans laquelle le domaine C-terminal de DiscII est remplacé par HPA. Cette construction est réalisée par PCR chevauchante dont le principe est résumé dans la **figure 2.5**.

La première réaction PCR est réalisée à partir des matrices pET-HPA et pProEx-DiscII et les amorces Dis2NcoI/LDD2rev et HPAfor/HPArevXhoI. La polymérase utilisée est la PfuUltraII fusion (Stratagène) selon les informations fournies par le fournisseur. Une seconde amplification conventionnelle, à partir du produit PCR précédent et purifié (50 ng) par extraction sur gel d'agarose 1,5%, est réalisée avec la polymérase HotStar (Qiagen) et les amorces Dis2NcoI et HPArevXhoI. Le produit obtenu est alors cloné dans le vecteur pProEx-HtB pour générer le vecteur pProEx-DiscIINter-HPA, vérifié par séquençage et transformé dans des souches d'expression.



**Figure 2.5. Principe de la PCR chevauchante appliquée à la formation de la protéine hybride DiscIINter-HPA.**

Un mutant DiscIINter-HPA-R163S a également été construit dans lequel le résidu Arg 163 (acide aminé 1 de la séquence HPA) est remplacé par un résidu Ser par mutagenèse dirigée. A partir de la matrice, pProEx-DiscIINter-HPA, la région codant est amplifiée par la

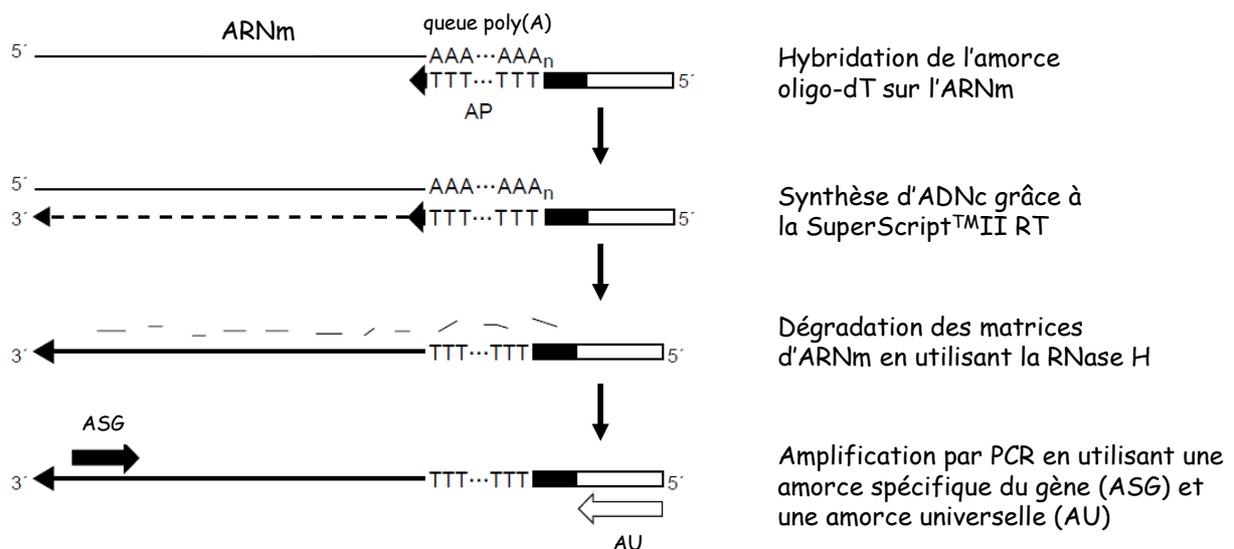
polymérase PfuUltraII fusion (Stratagène) en utilisant les amorces HybR163SFor et HybR163SRev. Le plasmide pProEx-DiscII<sup>Nter</sup>-HPA-R163S est alors généré.

### 2.2.3. Clonage de la lectine SLL2 extraite du corail

#### 2.2.3.1. Clonage

Le clonage du gène *sll2* a débuté avec l'utilisation des ARN totaux extraits d'un corail du genre *Sinularia* en collaboration avec Sylvie Tambutté de l'aquarium de Monaco.

L'isolement du gène est basé sur le principe du système RACE (Rapid Amplification of cDNA Ends) commercialisé par Invitrogen (**Figure 2.6.**). L'amplification par PCR peut utiliser soit des amorces (sens et antisens) spécifiques du gène d'intérêt (ASG) soit une amorce sens spécifique du gène et une amorce antisens dite « universelle » (AU). Les deux stratégies ont été exploitées.



**Figure 2.6. Principe de la 3'RACE (Invitrogen).**

Les amorces sens et antisens dégénérées ont été construites d'après l'alignement de séquences des lectines du corail *Sinularia lochmodes* (Jimbo *et al.* 2005), d'escargot (*Helix pomatia* et *Cepaea hortensis* notamment), d'amibe (*Dictyostelium discoideum*). Huit réactions de PCR ont été testées avec les couples d'amorces suivants SLL1/SLL3, SLL1/AU, SLL2/SLL3, SLL2/AU, SLL4/SLL6, SLL4/AU, SLL5/SLL6 et SLL5/AU. Après purification, le produit PCR est ligué dans le vecteur pGEM-T easy (Promega) et transformé dans la

souche XL1-blue afin d'être amplifié avant d'être séquencé. Toutes ces tentatives sont restées vaines.

Basé sur la séquence nucléotidique publiée (Jimbo *et al.* 2005), deux gènes *sll-2a* et *sll-2a-2*, de longueurs différentes ont alors été commandés à la compagnie GenScript Corporation (VWR Strategic Partner). Leur séquence a été optimisée de manière à éliminer les codons rares présents chez la bactérie *E. coli*, sans pour autant modifier la séquence peptidique. Afin d'optimiser la production sous forme recombinante dans *E. coli* et la levure *P. pastoris*, cinq et six constructions dans des vecteurs différents ont été réalisées pour les protéines SLL2a et SLL2a-2 respectivement. Les huit constructions résumées dans le **tableau 2.4** ont été formées par clonage classique en utilisant la polymérase HotStar HiFidelity (Qiagen) selon les indications du fournisseur. Les trois plasmides restants sont générés par simple digestion enzymatique puis ligation dans le nouveau vecteur d'intérêt et sont résumés dans le **tableau 2.5**.

Protéine	Matrice	Amorce	Vecteur	Plasmide	Expres- -sion
SLL2a	pUC57-SLL2a	SLL2a-NcoI SLL2a-XhoI	pProEx-HtB	pProEx-SLL2a	bactérie
	pUC57-SLL2a	SLL2a-NcoI SLL2a-XhoI	pRSET-A	pRSET-SLL2a	bactérie
	pET-TEV-SLL2a	SLL2a-EcoRI SLL2a-XbaI	pPICZ $\alpha$ -A	pPICZ $\alpha$ -SLL2a	levure
	pET-TEV-SLL2a	SLL2a-PstI SLL2a-XbaI	pPICZ $\alpha$ -HisFlag	pPICZ $\alpha$ Flag-SLL2a	levure
SLL2a-2	pUC57-SLL2a-2	SLL2a2-For-NdeI SLL2a2-Rev-EcoRI	pRSET-A	pRSET-SLL2a-2	bactérie
	pUC57-SLL2a-2	SLL2a2-For-XhoI SLL2a2-Rev-EcoRI	pRSET-A	pRSET-SLL2a-2-His6	bactérie
	pET-M30-SLL2a-2	SLL2a-2-EcoRI 2SLL2a-2-XbaI	pPICZ $\alpha$ -A	pPICZ $\alpha$ -SLL2a-2	levure
	pET-M30-SLL2a-2	2SLL2a-2-PstI 2SLL2a-2-XbaI	pPICZ $\alpha$ -HisFlag	pPICZ $\alpha$ Flag-SLL2a-2	levure

**Tableau 2.4. Récapitulatif des clonages des gènes SLL2a et SLL2a-2.**

Protéine	Enzymes	Insert	Vecteur	Plasmide formé
SLL2a	NdeI/BamHI	pRSET-SLL2a	pET-TEV-CupB6	pET-TEV-SLL2a
SLL2a-2	NcoI/BamHI	pUC57-SLL2a-2	pProEx-HtB vide	pProEx-SLL2a-2
	NcoI/BamHI	pProEx-SLL2a-2	pET-M30 vide	pET-M30-SLL2a-2

**Tableau 2.5. Récapitulatif des constructions des gène SLL2a et SLL2a-2 obtenues par digestion enzymatique.**

#### 2.2.3.2. Electroporation

Après vérification par séquençage, l'ADN d'un clone positif est transformé dans la levure *P. pastoris* par électroporation. Pour cela, le plasmide (20 µg) est linéarisé avec 25 unités de l'enzyme *SstI* (Invitrogen), pendant trois heures à 37°C puis purifié et concentré par SureClean (Bioline).

Deux jours avant l'électroporation, une préculture de nuit de la souche GS 115 de *P. pastoris* est réalisée dans 5 ml de milieu YPD (1% p/v d'extrait de levure, 2% p/v de peptone et 2% p/v de dextrose 20%), à 30°C et 250 rpm. Le lendemain, 60 µl de cette préculture sont ajoutés à 80 ml de YPD. Cette nouvelle culture est mise à 30°C, sous agitation à 250 rpm. Après une nuit, la DO<sub>600nm</sub> doit être autour de 1,5. Successivement par centrifugation à 1500 g, à 4°C pendant 5 minutes et resuspension des culots, les cellules sont lavées dans 80 ml puis 40 ml d'eau froide stérile, puis également dans 3,2 ml de sorbitol 1 M froid et stérile pour enfin être concentrées dans 160 µl de cette dernière solution. La transformation est réalisée avec 5 à 10 µg d'ADN dans au minimum 10 µl d'eau, introduit par électroporation (pulse 1600 V, Eppendorf) dans la levure. La totalité des cellules électroporées est étalée sur milieu YPDS contenant 100 µg.ml<sup>-1</sup> de zéocine. Les boîtes sont incubées quelques jours à 30°C jusqu'à l'apparition des colonies.

#### 2.2.3.3. Analyse de l'ADN génomique

La présence du gène d'intérêt dans le génome de la levure peut être contrôlée par PCR, après extraction de l'ADN génomique. Pour cela, le culot cellulaire correspondant à 1 ou 2 ml de culture est lysé dans 200 µl de tampon (0,1 M Tris-HCl pH 8, 50 mM EDTA, 1% SDS) à 100°C pendant 5 minutes. Après centrifugation une minute à 11000 g, le surnageant est dilué dans 100 µl d'acétate d'ammonium 7 M pH 7 et incubé 5 minutes à 65°C puis dans la glace. Après ajout de 200 µl de chloroforme et centrifugation, le surnageant de la couche

supérieure est précipité avec 250 µl d'isopropanol, incubé 5 minutes à température ambiante et centrifugé 10 minutes à 13000 g. Le culot est alors lavé avec 500 µl d'éthanol 70%, centrifugé 10 minutes à 13000 g et repris dans 40 µl d'eau mQ.

L'ADN extrait est alors amplifié par PCR avec la polymérase GoTaq (Promega) et les deux couples d'amorces 5'AOX1/3'AOX1 et pPICZ $\alpha$ -F/pPICZ $\alpha$ -R, purifié sur gel (kit Gel Extraction de Macherey-Nagel) et analysé par séquençage.

### 3. Expression de protéines sous forme recombinante

#### 3.1. Expression dans la bactérie *Escherichia coli*

De nombreux paramètres peuvent influencer les rendements en protéines : la cellule hôte, la température, la concentration en IPTG, la durée d'induction, la densité optique (DO<sub>600nm</sub>) au moment de l'induction, le vecteur utilisé pour le clonage (Donovan *et al.* 1996 ; Graslund *et al.* 2008). Déterminer les conditions optimales d'expression d'une protéine revient à tester ces paramètres.

Différentes souches bactériennes ont été testées de manière à optimiser l'expression des différentes protéines dans *E. coli*. Cependant, le protocole d'expression reste similaire quelque soit la souche utilisée. Une préculture liquide en milieu Luria Bertani (LB) contenant les antibiotiques appropriés est réalisée à 37°C pendant une nuit sous agitation (180 rpm). La culture est lancée après ajout d'un certain de volume de préculture de manière à obtenir une densité optique à 600nm (DO<sub>600nm</sub>) autour de 0,1. A DO<sub>600nm</sub> comprise entre 0,6 et 1,5 (cela dépend des essais), l'expression des protéines est induite par l'ajout d'IPTG à différentes concentrations et se fait pendant une durée déterminée à différentes températures. Au terme de l'expression, les cellules sont centrifugées 20 minutes à 6000 g. Les culots bactériens sont soit directement utilisés en vue d'une purification de protéines, soit congelés à -20°C. Le niveau d'expression des protéines est vérifié en chargeant sur un gel polyacrylamide dénaturant des extraits totaux d'*E. coli* avant et après induction à l'IPTG. Pour cela, 1 ml de culture bactérienne est prélevé avant et après l'induction des protéines. Les bactéries sont centrifugées 2 minutes à 13000 g et resuspendues dans 50 µl d'eau et 50 µl de tampon dénaturant 2X. Le mélange est alors dénaturé par chauffage 5 minutes à 100°C puis par

sonication pendant 2 minutes, avant d'être déposé sur gel de polyacrylamide en conditions dénaturantes. Les volumes déposés sont calculés pour avoir une quantité équivalente de protéine dans chaque échantillon en tenant compte des facteurs de dilution ou concentration appliqués aux échantillons prélevés.

Les caractéristiques des souches d'expression utilisées pour les différents tests d'expression de protéines recombinantes dans *E. coli* sont résumées dans le **tableau 2.6**.

Souche	Caractéristiques	Référence
BL21 (DE3)	Souche de référence pour l'expression de protéines recombinantes, déficiente en protéases OmpT et Lon. La désignation (DE3) signifie que la souche porte une copie chromosomale du gène de l'ARN polymérase T7 sous le contrôle du promoteur <i>lacUV5</i> et donc inductible à l'IPTG.	Novagen
BL21 (DE3) pLysS	La désignation pLysS indique que la souche porte un plasmide codant pour le lysozyme T7, inhibiteur naturel de l'ARN polymérase T7. L'expression basale de protéine est alors supprimée avant toute induction. Le plasmide est résistant au chloramphénicol.	Novagen
BL21 (DE3) C41	Cette souche, dérivée des BL21(DE3), porte au moins une mutation non caractérisée qui prévient la mort cellulaire due à l'expression de protéines recombinantes toxiques.	Lucigen
BL21 (DE3) C43	Cette souche, dérivée de la précédente, favorise l'expression de protéines toxiques différentes que BL21(DE3) C41.	Lucigen
BL21 Star (DE3) pLysS	La désignation Star révèle que la souche porte un gène <i>rne</i> muté ( <i>rne131</i> ) qui code pour une enzyme RNase E tronquée, qui perd alors ses capacités de dégradation d'ARNm. La stabilité des ARNm produits est alors fortement augmentée.	Invitrogen
Origami (DE3) K12	Cette souche, dérivée de la souche K-12, possède des mutations sur les gènes des enzymes thioredoxin reductase ( <i>trxB</i> ) et glutathione reductases ( <i>gor</i> ), ce qui favorise la formation des ponts disulfure dans le cytoplasme. Ces mutations sont respectivement sélectionnées sur kanamycine et tétracycline.	Novagen
Origami 2 (DE3) K12	La mutation <i>trxB</i> est sélectionnée sur streptomycine.	Novagen
Origami B (DE3)	Cette souche possède les mutations <i>trxB</i> et <i>gor</i> mais dérive du mutant <i>lacZY</i> de BL21. Une mutation sur le gène de la perméase <i>lac</i> provoque une entrée uniforme de l'IPTG dans la cellule, ce qui permet une régulation fine de l'expression protéique en fonction de la concentration de l'inducteur.	Novagen
Rosetta 2 (DE3) pLysS	La souche Rosetta 2, dérivée des BL21 (DE3), porte un vecteur pRARE2, résistant au chloramphénicol, qui permet la synthèse des ARNs de transfert (ARNt) correspondant aux codons AUA, AGG, AGA, CUA, CCC, GGA et CGG, rares chez <i>E. coli</i> .	Novagen
Rosetta gami 2 (DE3) pLysS	Rosetta gami 2 possède les caractéristiques des souches Rosetta 2 et Origami.	Novagen
Arctic Express (DE3) RIL	Cette souche, dérivée des BL21 (DE3), va co-exprimer, grâce à un plasmide résistant à la gentamycine, deux gènes codant pour les protéines chaperones, Cpn10 et Cpn60, de la bactérie <i>Oleispira antartica</i> , qui vont faciliter le repliement de la protéine à basse température. Elle contient également un plasmide supplémentaire, résistant à la streptomycine, qui porte les gènes codant pour les ARNt des codons AGA, AGG, AUA et CUA.	Stratagene

**Tableau 2.6. Récapitulatif des différentes souches d'*E. coli* utilisées pour l'expression de protéines sous forme recombinante.**

### 3.2. Expression dans la levure *Pichia pastoris*

Les colonies de *P. pastoris* ayant intégré le plasmide sont mises en culture dans 5 ml de milieu MGYC (1% yeast extract, 2% peptone, 1% casamino acid, 0,2% biotine, 1% glycérol) à 30°C et sous agitation (250 rpm). Après une nuit, la culture est centrifugée à 2000 g pendant trois minutes. Les culots sont lavés par centrifugation et resuspension dans 5 ml de milieu YNB 1,34% (Yeast Nitrogen Base, Invitrogen) stérile. Ils sont ensuite resuspendus dans 10 ml de milieu inductible MMYC (qui correspond au milieu MGYC dans lequel 0,5% méthanol remplace le glycérol). Les cultures sont incubées à 30°C sous agitation (250 rpm). Du méthanol pur (à différentes concentrations pour l'optimisation) est ajouté quotidiennement pendant 2, 3, 4 ou 5 jours. Des témoins de 2 ml de surnageant de culture avant et après l'induction sont prélevés, précipités à l'acide trichloroacétique (TCA) 100% et analysés sur gel SDS-PAGE. Même si la protéine est sensée être sécrétée dans le milieu de culture, le contenu des levures est tout de même analysé afin de s'assurer de l'absence de protéine recombinante. Les culots de levure correspondant à 1,8 ml de culture sont resuspendus dans le même volume de tampon HEPES-KOH 50 mM pH 7,5, sucrose 0,4 M, MgCl<sub>2</sub> 10 mM puis éclatés au cell disrupter à 2,5 kbars. Après une centrifugation à 12000 g pendant 10 minutes à 4°C, les culots repris dans 900 µl de tampon et les surnageants sont analysés sur gel SDS-PAGE 15%.

## 4. Techniques de biochimie

### 4.1. Purification de protéines

Toutes les chromatographies automatisées effectuées au cours de ces travaux ont été réalisées sur un système ÄKTA-FPLC (Amersham Biosciences). Dans ce système, les paramètres tels que la pression, le pH et la température sont contrôlés. L'élution des protéines est suivie par mesure de la densité optique à 280 nm au cours du temps. Les tampons de purification sont systématiquement filtrés et dégazés grâce à un système de pompe à vide.

#### 4.1.1. Préparation des extraits bruts

Les bactéries culottées sont reprises dans un volume de tampon d'équilibration correspondant à cinq fois leur masse, auquel a été ajouté une goutte d'« Antifoam 204 » (Sigma) et 1 mg de déoxyribonucléase I de pancréas bovin de type IV (Sigma). Les cellules sont éclatées grâce au système (One shot cell Disrupter, Constant System) à 1,8 kbars de pression. Les broyats bactériens sont alors soumis à deux centrifugations successives. La première s'effectue à 12000 g pendant 10 minutes à 4°C. Le culot, appelé fraction 12C, est alors composé de débris cellulaires et des protéines rassemblées sous forme de corps d'inclusion. Le surnageant est de nouveau centrifugé à 50000 g pendant 30 minutes à 4°C. Le surnageant, dénommé fraction 50S et formé des protéines solubles, est filtré sur un système de pompe à vide avant chargement sur colonne. Pour l'analyse des fractions 12C et 50S par électrophorèse sur gel SDS-PAGE, 50 µl d'échantillon sont additionnés à 50 µl de tampon dénaturant 2X avant d'être chauffés 5 minutes à 100°C et soniqués 2 minutes.

#### 4.1.2. Purification des lectines de *Dictyostelium discoideum*

##### 4.1.2.1. Chromatographie sur colonne de Nickel

La présence d'une queue polyhistidine en position N-terminal des protéines recombinantes permet leur purification par chromatographie de type IMAC (Immobilized Metal ion Affinity Chromatography). Le principe de cette chromatographie repose sur les interactions ioniques qui se forment entre les noyaux imidazoles des résidus histidines de la protéine et les charges positives des ions Nickel de la résine. La protéine est ensuite éluée par un mécanisme compétitif en ajoutant des quantités croissantes d'imidazole.

La purification de la protéine se déroule à température ambiante sur colonne HisTrap™ (GE Healthcare) de 1 ml et d'une capacité d'environ 40 mg de protéine-(His)<sub>6</sub>. La colonne est équilibrée dans cinq volumes de tampon d'équilibration : Tris-HCl 20 mM, pH 8,5, NaCl 500 mM, Imidazole 10 mM. L'extrait protéique, dilué dans le même tampon, est chargé sur la colonne à 1 ml.min<sup>-1</sup>. Après un lavage avec le tampon d'équilibration, la protéine est éluée à 1 ml.min<sup>-1</sup> grâce au tampon d'éluion : Tris-HCl 20 mM, pH 8,5, NaCl 500 mM, Imidazole 500 mM, soit par gradient en 20 volumes de colonne, soit par paliers à différentes concentrations d'imidazole.

Le tampon de la protéine est échangé par un tampon sans imidazole et moins concentré en NaCl qui diffère suivant les besoins (cristallogénèse, biotinylation...) grâce à

une colonne de dessalage PD-10 (GE-Healthcare) en suivant les instructions du fabricant. La protéine est alors dosée par spectrophotométrie sur le système Nanodrop ND1000 (Labtech).

L'étiquette polyhistidine peut éventuellement être éliminée par clivage avec la protéase TEV, produite elle aussi sous forme recombinante. La protéine d'intérêt est incubée une nuit à 20°C avec la protéase TEV diluée au 1/20 et avec 2 mM de DTT. Le mélange est ensuite redéposé sur colonne de Nickel afin de séparer les protéines clivées des non clivées. Les protéines sans étiquette seront éluées dans la fraction non retenue (FNR) ; les protéines non digérées, ainsi que la protéase seront retenues sur la colonne grâce à leur étiquette His<sub>6</sub>.

Les fractions correspondant à la lectine sont ensuite rassemblées, concentrées par ultracentrifugation sur colonne Vivaspın4 5kDa (Vivasciences) et conservées à 4°C.

Après en moyenne cinq utilisations successives, la colonne de Nickel est lavée et régénérée selon un protocole détaillé par le fournisseur.

#### 4.1.2.2. Chromatographie d'affinité sur colonne de sucre

La fonctionnalité des lectines purifiées peut être contrôlée par chromatographie d'affinité sur résines Sepharose 4B (Sigma), antigène de Forssman-agarose (Sigma) ou D-GalNAc-agarose (Sigma). L'éluion a lieu par compétition dans un tampon à forte concentration en D-galactose.

Certains tests ont été réalisés en utilisant la technique de purification en batch, directement dans des microtubes de 2 ml. La résine (100 µl) Sepharose 4B ou antigène de Forssman-agarose, déposée dans le microtube, est tout d'abord lavée et équilibrée avec respectivement 500 µl d'H<sub>2</sub>O mQ et 500 µl de tampon d'équilibration : Tris-HCl 20 mM, pH 8,5, NaCl 500 mM, Imidazole 10 mM. Entre chaque étape, un cycle de centrifugation douce (6000 g pendant deux minutes) suivi d'une décantation sur pailleasse pendant deux minutes permet de retirer l'excès de tampon sans perturber la résine. L'échantillon est ensuite incubé pendant 2 heures 30, sous une légère agitation, à 4°C avec la résine. Le surnageant résultant de l'étape de centrifugation constitue la fraction non retenue (FNR). Après deux étapes de lavage (L) successives avec le tampon d'équilibration, la protéine d'intérêt, éventuellement accrochée à la résine est éluée (fraction E) dans un tampon contenant 0,3 M de D-galactose pendant 30 minutes sous une légère agitation.

D'autres tests ont été réalisés dans des colonnes coulées manuellement (1 ml de résine Sepharose 4B), où l'écoulement a lieu par gravité ou des colonnes commerciales, connectées au système ÄKTA-FPLC (5 ml de résine D-GalNAc-agarose). De manière similaire au protocole décrit pour la purification sur colonne de Nickel, la colonne est équilibrée avec le

tampon d'équilibration correspondant au tampon dans lequel évolue la protéine après dessalage. Les différents tampons seront plus largement décrits et analysés dans la partie Résultats. Après chargement sur la colonne et lavage avec le tampon d'équilibration, la protéine est éluée avec un tampon contenant 0,3 M de D-Gal.

Les échantillons correspondant à FNR, L et E sont alors analysés sur gel SDS-PAGE.

#### 4.1.2.3. Chromatographie d'exclusion stérique

Encore appelée tamisage moléculaire ou gel filtration, la chromatographie d'exclusion stérique permet de séparer des molécules en fonction de leur volume et donc de leur poids moléculaire. La phase stationnaire est constituée de billes de résines gonflées dans le solvant utilisé pour l'éluion. Le volume des billes est très finement calibré, ainsi les molécules dont le volume est supérieur à celui des billes ne peuvent y pénétrer et sont éluées rapidement. En revanche, les molécules dont le volume est inférieur à celui des billes y pénètrent d'autant plus qu'il est petit et y subissent des frottements qui les retardent de manière plus ou moins importante.

Plusieurs résines (toutes commercialisées par GE Healthcare) ont été testées et sont répertoriées dans le **tableau 2.7**. L'étape de dessalage décrite dans le paragraphe 4.1.2.1 est une étape de chromatographie d'exclusion stérique par gravité. La colonne PD-10 contient la résine Sephadex G-25 et est particulièrement utilisée pour séparer les protéines des sels.

Les colonnes sont calibrées grâce au kit de protéines standard pour chromatographie par filtration (Amersham). L'échantillon protéique contenu dans un volume de 500  $\mu$ l est déposé sur la colonne pré-équilibrée par le tampon désiré, sous un flux de 0,5 ml.min<sup>-1</sup>.

Nom	Résine	Gamme de séparation pour les protéines globulaires
Sephacryl-S100	Polymères d'allyl-dextran et de N,N' méthylène bisacrylamide	1 à 100 kDa
Superdex 75	Polymères de dextran liés à une matrice d'agarose	3 à 70 kDa
Superdex 200	Polymères de dextran liés à une matrice d'agarose	10 à 600 kDa
Sephadex G-25	Polymères de dextran liés à de l'épichlorohydrine	> à 5 kDa

**Tableau 2.7. Récapitulatif des différentes chromatographies d'exclusion stérique testées durant cette étude.**

### 4.1.3. Purification des lectines de *Sinularia lochmodes* SLL-2a et SLL-2a-2

La purification des lectines SLL-2a et SLL-2a-2 couplées à une étiquette polyhistidine s'effectue selon un schéma similaire aux lectines de *D. discoideum*. Les protéines sont tout d'abord purifiées par chromatographie sur colonne de Nickel puis, leur fonctionnalité est testée par chromatographie d'affinité sur colonne de sucre, soit en batch avec la résine Sepharose 4B (Sigma), soit sur colonne manuelle (Sepharose 4B) ou commerciale (D-GalNAc-agarose (Sigma)). Les protocoles ont été détaillés précédemment pour les lectines de *D. discoideum*.

## 4.2. Techniques de renaturation de protéines

Les protéines SLL-2a et SLL-2a-2 sont majoritairement exprimées sous forme de corps d'inclusion, agrégats protéiques biologiquement inactifs. Différents protocoles ont été testés afin de transformer ces corps d'inclusion en protéines actives.

### 4.2.1. Utilisation d'Augmedium™

L'augmedium™ (Athena ES) est un additif à ajouter au milieu de culture avant l'induction à l'IPTG pour favoriser l'expression de protéines recombinantes solubles. La solution 50 X stérile est ajoutée à la culture à différentes concentrations (diluée 50, 100, 200, 400 et 800 fois) 20 minutes avant l'induction. La proportion de protéines solubles et exprimées sous forme de corps d'inclusion est ensuite contrôlée par analyse SDS-PAGE.

### 4.2.2. Lavage et solubilisation des corps d'inclusion

Les tests de solubilisation des corps d'inclusion ont été réalisés sur les protéines SLL-2a et SLL-2a-2/GST, clonées respectivement dans les vecteurs pET-TEV et pET-M30. De nombreux protocoles ont été tentés. Leurs détails seront plus largement abordés dans la partie Résultats. D'une manière générale, les culots bactériens sont éclatés au cell disrupter à 2 kbars soit directement dans les différents tampons de solubilisation des corps d'inclusion (répertoriés dans le **tableau 2.8.**), soit dans un tampon de lavage des corps d'inclusion : Tris-

HCl 20 mM, pH 8, sucrose 20%, EDTA 5 mM, Triton X-100 1%. Trois à cinq cycles de centrifugation successifs de 10 minutes à 10000 g permettent d'obtenir une protéine agrégée relativement pure, toujours localisée dans le culot. Après lavage, les protéines sont alors solubilisées dans le tampon de solubilisation (**Tableau 2.8.**) deux heures à température ambiante ou une nuit à 4°C et se retrouvent alors dans le surnageant après centrifugation. Les tests de renaturation sont toujours réalisés sur une protéine solubilisée.

Tampon de solubilisation des corps d'inclusion	Concentration en urée (M)
Tris-HCl 20 mM, pH 8, NaCl 500 mM, Imidazole 10 mM, $\beta$ -mercaptoéthanol ( $\beta$ -ME) 10 mM, glycérol 20%	6 M urée
Tris-HCl 20 mM, pH 8,5, NaCl 500 mM, Imidazole 10 mM, $\beta$ -ME 10 mM	6 M urée
Acétate de sodium-acide acétique 20 mM, pH 4, EDTA 2 mM, glutathion réduit (GSH) 1 mM, glutathion oxydé (GSSG) 0,1 mM	0 M urée
	2 M urée
	8 M urée
MES 20 mM, pH 6, EDTA 2 mM, GSH 1 mM, GSSG 0,1 mM	0 M urée
	2 M urée
	8 M urée
Tris-HCl 20 mM, pH 8, EDTA 2 mM, GSH 1 mM, GSSG 0,1 mM	0 M urée
	2 M urée
	4 M urée
	6 M urée
Tris-HCl 20 mM, pH 9, EDTA 2 mM, GSH 1 mM, GSSG 0,1 mM	8 M urée
	2 M urée
	4 M urée
	6 M urée
Tris-HCl 20 mM, pH 10, EDTA 2 mM, GSH 1 mM, GSSG 0,1 mM	8 M urée
	0 M urée
	2 M urée
	4 M urée
	6 M urée
Tris-HCl 20 mM, pH 9, NaCl 150 mM, DTT 0,1 M	8 M urée
Tris-HCl 50 mM, pH 9	6 M urée
	8 M urée

**Tableau 2.8. Récapitulatif des tampons de solubilisation utilisés pour le refolding des lectines de *Sinularia lochmodes***

### 4.2.3. Renaturation des protéines dénaturées

L'étape de renaturation des protéines consiste à diminuer très lentement, la proportion d'agent dénaturant (urée) de manière à ce que la protéine dénaturée adopte le bon repliement tridimensionnel. Deux principales techniques ont été testées dans cette étude. La renaturation sur colonne de nickel est un moyen rapide qui permet de purifier et de renaturer la protéine en une seule étape. La protéine dénaturée est, tout d'abord, accrochée à la résine de nickel via son étiquette polyhistidine à un flux de  $0,5 \text{ ml}\cdot\text{min}^{-1}$ . Elle est ensuite renaturée directement sur la colonne, grâce à l'injection du tampon de renaturation à  $0,1 \text{ ml}\cdot\text{min}^{-1}$  pendant trois heures, avant d'être éluée par compétition avec de l'imidazole dans les conditions précédemment décrites. La seconde méthode, plus classique, consiste à renaturer la protéine dans un bécher par ajout du tampon de renaturation, goutte à goutte, de manière à diluer la solution environ dix fois. Le flux estimé du goutte à goutte réalisé à la pompe péristaltique est de  $0,3$  à  $0,4 \text{ ml}\cdot\text{min}^{-1}$ . Les différents tampons de renaturation utilisés sont référencés dans le **tableau 2.9**.

Tampon de renaturation	Méthode
Tris-HCl 20 mM pH 8,5, NaCl 500 mM, Imidazole 10 mM, GSH 1 mM, GSSG 1 mM	Colonne
Tris-HCl 20 mM pH 8, NaCl 500 mM, Imidazole 10 mM, $\beta$ -ME 5 mM, glycérol 10%	Colonne
Tris-HCl 20 mM pH 7,5, NaCl 150 mM, GlcNAc 10 mM, L-Arginine 0,5 M, glycérol 10%, GSH 5 mM, GSSG 0,5 mM	Bécher
Tris-HCl 50 mM pH 8,5, NaCl 9,6 mM, KCl 0,4 mM, $\text{MgCl}_2$ 2 mM, $\text{CaCl}_2$ 2 mM, L-Arginine 0,5 M, sucrose 0,4 M, Guanidine-HCl 0,75 M, GSH 1 mM, GSSG 0,1 mM	Bécher
PBS	Bécher
Tris-HCl 20 mM pH 8, GalNAc 10 mM	Bécher
Tris-HCl 50 mM pH 8, cystéamine 8 mM, cystamine 4 mM, L-Arginine 0,5 M	Bécher

**Tableau 2.9. Récapitulatif des tampons de renaturation utilisés dans cette étude.**

Afin d'optimiser la recherche des conditions de renaturation, un kit de quinze solutions de renaturation QuickFold™ Protein Refolding, commercialisé par Athena ES a également été testé, selon les indications fournies par le fournisseur. Les solutions sont référencées dans le **tableau 2.10**.

Solution 1	MES 50 mM pH 6, NaCl 9,6 mM, KCl 0,4 mM, MgCl <sub>2</sub> 2 mM, CaCl <sub>2</sub> 2 mM, Guanidine-HCl 0,75 M, Triton X-100 0,5%, DTT 1 mM
Solution 2	MES 50 mM pH 6, NaCl 9,6 mM, KCl 0,4 mM, MgCl <sub>2</sub> 2 mM, CaCl <sub>2</sub> 2 mM, Arginine 0,5 M, PEG 3350 0,05%, GSH 1 mM, GSSG 0,1 mM
Solution 3	MES 50 mM pH 6, NaCl 9,6 mM, KCl 0,4 mM, EDTA 1 mM, sucrose 0,4 M, Guanidine-HCl 0,75 M, Triton X-100 0,5%, PEG 3350 0,05%, DTT 1 mM
Solution 4	MES 50 mM pH 6, NaCl 240 mM, KCl 10 mM, MgCl <sub>2</sub> 2 mM, CaCl <sub>2</sub> 2 mM, Guanidine-HCl 0,75 M, GSH 1 mM, GSSG 0,1 mM
Solution 5	MES 50 mM pH 6, NaCl 240 mM, KCl 10 mM, EDTA 1 mM, sucrose 0,4 M, Guanidine-HCl 0,75 M, DTT 1 mM
Solution 6	MES 50 mM pH 6, NaCl 240 mM, KCl 10 mM, EDTA 1 mM, arginine 0,5 M, sucrose 0,4 M, Triton X-100 0,5%, PEG 3350 0,05%, GSH 1 mM, GSSG 0,1 mM
Solution 7	MES 50 mM pH 6, NaCl 240 mM, KCl 10 mM, MgCl <sub>2</sub> 2 mM, CaCl <sub>2</sub> 2 mM, Guanidine-HCl 0,75 M, PEG 3350 0,05%, DTT 1 mM
Solution 8	Tris-HCl 50 mM pH 8,5, NaCl 9,6 mM, KCl 0,4 mM, MgCl <sub>2</sub> 2 mM, CaCl <sub>2</sub> 2 mM, sucrose 0,4 M, Triton X-100 0,5%, PEG 3350 0,05%, GSH 1 mM, GSSG 0,1 mM
Solution 9	Tris-HCl 50 mM pH 8,5, NaCl 9,6 mM, KCl 0,4 mM, EDTA 1 mM, arginine 0,5 M, Guanidine-HCl 0,75 M, PEG 3350 0,05%, DTT 1 mM
Solution 10	Tris-HCl 50 mM pH 8,5, NaCl 9,6 mM, KCl 0,4 mM, MgCl <sub>2</sub> 2 mM, CaCl <sub>2</sub> 2 mM, arginine 0,5 M, sucrose 0,4 M, Guanidine-HCl 0,75 M, GSH 1 mM, GSSG 0,1 mM
Solution 11	Tris-HCl 50 mM pH 8,5, NaCl 9,6 mM, KCl 0,4 mM, EDTA 1 mM, Triton X-100 0,5%, DTT 1 mM
Solution 12	Tris-HCl 50 mM pH 8,5, NaCl 240 mM, KCl 10 mM, EDTA 1 mM, PEG 3350 0,05%, GSH 1 mM, GSSG 0,1 mM
Solution 13	Tris-HCl 50 mM pH 8,5, NaCl 240 mM, KCl 10 mM, EDTA 1 mM, arginine 0,5 M, Guanidine-HCl 0,75 M, Triton X-100 0,5%, DTT 1 mM
Solution 14	Tris-HCl 50 mM pH 8,5, NaCl 240 mM, KCl 10 mM, MgCl <sub>2</sub> 2 mM, CaCl <sub>2</sub> 2 mM, arginine 0,5 M, sucrose 0,4 M, Guanidine-HCl 0,75 M, Triton X-100 0,5%, PEG 3350 0,05%, GSH 1 mM, GSSG 0,1 mM
Solution 15	Tris-HCl 50 mM pH 8,5, NaCl 240 mM, KCl 10 mM, EDTA 1 mM, MgCl <sub>2</sub> 2 mM, CaCl <sub>2</sub> 2 mM, sucrose 0,4 M, DTT 1 mM

**Tableau 2.10. Tampons disponibles dans le kit de renaturation QuickFold™ Protein Refolding commercialisé par Athena ES.**

La solubilité de la protéine renaturée et sa fonctionnalité sont respectivement contrôlées par analyse sur gel SDS-PAGE et chromatographie d'affinité sur colonne manuelle de sepharose 4B (Sigma).

### 4.3. Caractérisation biochimique des protéines

#### 4.3.1. Electrophorèse sur gel de polyacrylamide en présence de SDS

Cette technique permet de séparer les protéines en fonction de leur masse moléculaire.

Les échantillons protéiques sont préalablement repris dans un tampon Tris-HCl 0,1 M pH 6,8, glycérol 20% (v/v), SDS 4%, DTT 20 mM, bleu de bromophénol 0,012% qui dénature et charge négativement les protéines de manière homogène. La dénaturation est complétée par chauffage à 100°C pendant cinq minutes puis par sonication pendant deux minutes. Les échantillons peuvent également être précipités au TCA 100% afin de les concentrer. Après avoir ajouté 1/10 du volume final de TCA 100% froid, ils sont incubés 15 minutes à 4°C puis centrifugés 15 minutes à 13000 g à 4°C. Le culot protéique est alors repris dans un certain volume d'eau dilué par un équivalent de tampon de dénaturation 2X. 1 µl de NaOH 1 M peut être nécessaire pour augmenter le pH afin de garder la coloration bleue de la solution. Les échantillons sont ensuite traités comme décrit précédemment.

Les protéines sont alors déposées sur le gel de polyacrylamide constitué d'un gel de concentration (partie supérieure) (acrylamide 5%, pH 6,8) et d'un gel de séparation dont le pourcentage d'acrylamide est fonction du poids moléculaire des protéines à séparer. Une fois la migration terminée, le gel est ensuite :

- soit coloré au bleu de Coomassie (Coomassie Brilliant Blue R-250 2,5 g/L (Sigma), éthanol/acide acétique 9 : 2 v/v). L'excès de colorant est ensuite éliminé par lavage avec un mélange d'acide acétique 10% (v/v) et éthanol 45% (v/v).

- soit directement coloré par la solution InstantBlue (Euromedex), sans étape de décoloration. Dans les deux cas précédents, le marqueur de masse moléculaire Unstained Standard (BioRad) est déposé sur le gel parallèlement aux échantillons.

- soit utilisé pour le transfert des protéines sur membrane de nitrocellulose. Dans ce cas, le marqueur de masse moléculaire Dual Color (BioRad) est utilisé.

#### 4.3.2. Transfert de protéines et immunodétection (Western Blot)

Après migration sur gel, les protéines sont électrotransférées sur une membrane de nitrocellulose (Biotrace Pall) à l'aide d'un système de transfert semi-sec (BioRad). Ce système implique la préparation d'un tampon (Tris-HCl 20 mM pH 9,4, glycine 30 mM,

méthanol 20%) pour la cathode dans lequel sera imbibé trois feuilles de papier Whatman et le gel et d'un tampon (Tris-HCl 0,3 M pH 10,4, méthanol 10%) pour l'anode dans lequel la membrane et trois feuilles de papier Whatman seront baignées. Le transfert s'effectue sous courant continu, sous un ampérage correspondant à 0,8 mA que multiplie la surface du gel, pendant une à deux heures.

A la fin du transfert, le gel est coloré afin d'évaluer la présence résiduelle de protéines et contrôler ainsi la qualité du transfert et la membrane est lavée quelques minutes dans du TBS (Tris-HCl 10mM pH 8, NaCl 150mM). Elle subit par la suite les traitements successifs suivant :

- une saturation des sites non spécifiques pendant 45 minutes avec une solution de gélatine 1% dans du TBS, suivie de deux lavages de 5 minutes dans du TBS-T (Tween 20 0,05%)

- une incubation d'une heure en présence d'anticorps primaire (Penta-His, Qiagen) dilué au 1/2000<sup>ème</sup> dans du TBS-T suivie de trois lavages de 5 minutes dans TBS-T.

- une incubation d'une heure en présence d'anticorps secondaire couplé à la peroxydase (goat anti-mouse-Per, Sigma) dilué au 1/2000<sup>ème</sup> dans du TBS-T.

La membrane est ensuite lavée trois fois successivement dans du TBS-T puis rincée dans du TBS avant d'être révélée avec le kit AEC 101 (Sigma) qui contient le substrat chromogène de la peroxydase.

### 4.3.3. Test d'hémagglutination

La mesure d'activité hémagglutinante est le test d'interaction le plus largement utilisé pour la détection des lectines et leur caractérisation (Goldstein 1980 ; Rüdiger 1993). Cette technique repose sur la capacité des lectines à former un réseau avec des érythrocytes par interaction avec leurs glycoconjugués de surface. Les hématies de mammifères sont parmi les plus utilisés (homme, lapin, mouton, porc...). Les tests sont réalisés sur des microplaques de titration et l'agglutination (ou agrégation) des globules rouges par les lectines forme une phase gélatineuse visible à l'œil nu. En absence d'activité hémagglutinante, les hématies sédimentent par gravité au fond des puits.

Pour cela, des dilutions en cascade sont réalisées. Le volume total dans chaque puits est de 50 µl. La concentration en lectine dans le premier puits est de 1 mg.ml<sup>-1</sup>. La protéine est ensuite diluée en cascade, de deux en deux, jusqu'au 22<sup>ème</sup> puits. Puis, 25 µl de la suspension d'érythrocytes de lapin (Biomérieux) à 2% en solution saline de NaCl 150 mM sont ajoutés à

chaque puits, mélangés et incubés à 37°C pendant 30 minutes. La lecture de l'activité hémagglutinante est réalisée directement après cette incubation (**Figure 2.7.**).



**Figure 2.7. Le test d'hémagglutination sur microplaque de titration**

#### 4.3.4. DLS : Diffusion dynamique de la lumière

La diffusion dynamique de la lumière est une technique très puissante pour déterminer la taille des particules diffusantes en solution. Elle fournit ainsi des informations sur la présence ou non d'agrégats, la polydispersité ou l'homogénéité de la solution. Cette technique rapide et facile à utiliser possède comme principal avantage d'être non destructrice. Les échantillons analysés peuvent donc être récupérés en fin de manipulation.

Chaque particule en solution évolue suivant un mouvement aléatoire (mouvement Brownien) en fonction de sa taille. Plus une molécule est volumineuse, moins son déplacement en solution sera rapide. Frappée par une lumière laser, les particules en mouvement vont interagir avec ce faisceau et le disperser dans toutes les directions. L'analyse de la lumière diffusée, recueillie par un détecteur, permettra de déduire le rayon hydrodynamique de la particule, directement proportionnel à sa taille en solution. Les résultats sont alors présentés sous la forme d'une courbe de l'intensité relative de la lumière diffusée en fonction du rayon hydrodynamique des particules. Dans le cas des protéines, il permet ainsi d'élucider leur structure quaternaire. L'appareil disponible au laboratoire est le « Zetasizer nano series » (Malvern Instrument).

#### 4.3.5. La résonance plasmonique de surface

##### 4.3.5.1. Principe

Le système Biacore® repose sur le principe de Résonance Plasmonique de Surface (RPS), pour mesurer, en temps réel, toute interaction biologique sans aucun marquage des molécules. Cette technique permet de déterminer, de manière quantitative, les paramètres cinétiques d'une interaction mais également de cribler, de manière qualitative, les molécules

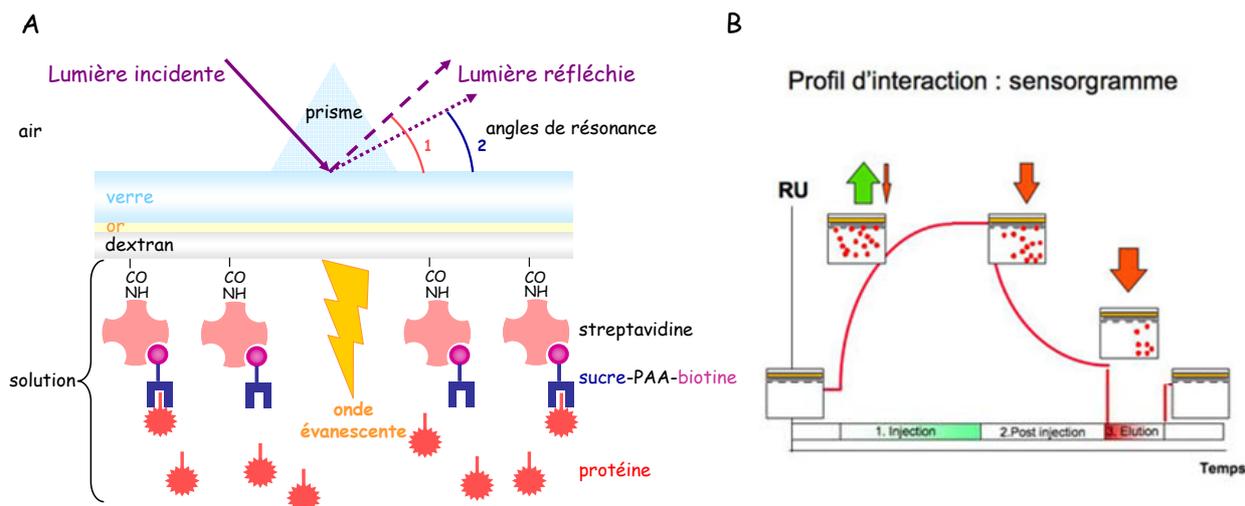
impliquées dans cette interaction. La technique implique la fixation de l'un des partenaires de l'interaction, appelé ligand, sur un biocapteur, la « sensor chip » pendant que l'autre partenaire, l'analyte, passe en solution. Le biocapteur est composé d'une fine couche de verre recouverte d'un film d'or, couplé à différentes fonctions chimiques permettant d'immobiliser le ligand. Lorsqu'un faisceau de lumière monochromatique polarisée illumine, à un angle précis, l'interface entre deux indices de réfraction différents, toute la lumière est réfléchie, aucune réfraction n'est observée. Les électrons délocalisés de la couche périphérique de l'or sont excités par l'onde incidente et entrent en résonance (**Figure 2.8. A.**). Parallèlement, une composante électromagnétique de la lumière incidente, l'onde évanescente, se propage perpendiculairement à la surface dans le milieu d'indice de réfraction le plus faible (la solution). La lumière réfléchie présente alors une chute d'intensité à un angle précis, l'angle de résonance, correspondant à l'absorption d'énergie de l'onde évanescente. Les changements de masse induits par l'association ou la dissociation des complexes ligands-analytes modifient la réfringence du milieu et décalent ainsi la position de cet angle.

Ces variations sont enregistrées sur un sensorgramme (**Figure 2.8. B.**) et quantifiées en unités de résonance (RU), 1000 RU correspondant à un déplacement de l'angle de  $0,1^\circ$  et équivalent à la fixation de 1 ng d'analytes par  $\text{mm}^2$ . Le sensorgramme est composé d'une phase d'injection (association), d'une phase de post-injection (dissociation) avec l'injection de tampon seul et d'une phase de régénération. Les paramètres cinétiques sont déterminés pendant les phases d'association et de dissociation et extraits des données expérimentales à partir de modèles mathématiques intégrés avec le logiciel d'analyse.

#### 4.3.5.2. Préparation du biocapteur

Ces expériences ont été réalisées par Samy Cecioni au National Center for Biomolecular Research dans l'université Masaryk, à Brno en République Tchèque.

Le biocapteur utilisé est appelé CM5 et est formé de quatre canaux indépendants, contenant chacun une matrice de dextrans carboxyméthylés. Les groupements carboxyles, après activation par un mélange de 1-éthyl-3-(3-diméthylaminopropyl) carbodiimide (EDC) et *N*-hydroxysuccinimide (NHS) (Biacore®) sont utilisés pour immobiliser, via un lien amide, la streptavidine présente en solution dans un tampon acétate à pH acide. Une solution d'éthanolamine (Biacore®) est ensuite injectée pour saturer les groupements activés restés libres. Des ligands sucre polyacrylamide-biotine (biotine-PAA-sucre) (Lectinity Holdings, Inc.) différents sont alors fixés sur les quatre canaux du biocapteur grâce à la forte interaction biotine-streptavidine.



**Figure 2.8. Principe du système Biacore®.**

A. Schéma du principe de Résonance Plasmonique de Surface. B. Sensorgramme.

Le principe de la RPS permet de visualiser en temps réel des interactions entre biomolécules non marquées dans un débit continu de tampon. Un des partenaires de l'interaction ou une molécule de capture de ce partenaire est immobilisé sur la surface du biocapteur, les autres partenaires appelés analytes sont injectés à débit constant et circulent dans le flux au contact de la surface du biocapteur. Dans nos conditions, les sucre-polyacrylamide-biotinylés sont accrochés à la surface via un lien biotine-streptavidine, elle-même immobilisée, via un lien amide, à la surface du biocapteur. L'analyte est la protéine DiscII-Nter.

#### 4.3.5.3. Conditions expérimentales

Le tampon de course utilisé est composé de 10 mM HEPES, 150 mM NaCl, 100  $\mu$ M CaCl<sub>2</sub>, 0,005% Tween 20. La régénération de la surface après interaction des biomolécules a été effectuée par des injections de NaCl 1M, NaOH 50mM ou de solutions de sucres (de 10 mM à 100 mM). Tous les inhibiteurs ont été fournis par Elicityl, excepté l' $\alpha$ -Me-fucoside et le sialyl Lewis x qui ont respectivement été commandés chez Sigma et Dextra. Les conditions expérimentales sont brièvement résumées dans le **tableau 2.11**.

Conditions	Premier test	Variation de la concentration en protéines	Cinétique d'inhibition
Température	25°C		
Flux	10 $\mu\text{l}\cdot\text{min}^{-1}$		
Temps de contact	120 s	540 s	540 s
Concentration	Protéines 50 $\mu\text{g}\cdot\text{ml}^{-1}$	Protéines : de 3 à 100 $\mu\text{g}\cdot\text{ml}^{-1}$	Protéines 0,5 $\mu\text{M}$ Inhibiteurs 5 $\mu\text{M}$

**Tableau 2.11. Conditions expérimentales pour les mesures d'interaction entre DiscII-Nter et les PAA-Man et PAA-Fuc par RPS.**

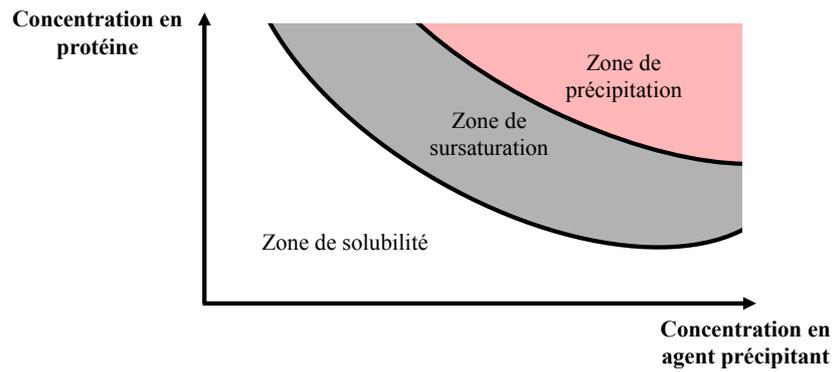
Pour les tests d'inhibitions, les protéines sont mélangées avec les solutions d'inhibiteurs concentrées 10 fois selon un rapport 9 : 1 avant injection sur la surface.

## 5. La cristallographie aux rayons X

La cristallographie aux rayons X est une technique qui permet de déterminer la position de tous les atomes d'un cristal dans l'espace. A la différence de la résonance magnétique nucléaire (RMN), la cristallographie des protéines n'est pas limitée au niveau de la taille des macromolécules étudiées et est donc devenue une technique phare pour la détermination de la structure 3-D des protéines.

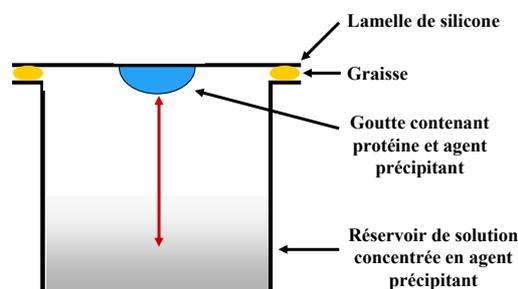
### 5.1. La cristallogenèse

La cristallogenèse consiste à rechercher les conditions optimales dans lesquelles les protéines vont s'organiser dans l'espace sous forme de cristal. Elles passent alors d'un état liquide désordonné à un état solide ordonné. Lors de la cristallisation, la solubilité de la solution protéique diminue, par un procédé d'évaporation contrôlé, jusqu'à atteindre l'état de supersaturation favorable à la formation de cristaux (**Figure 2.9**). Cependant, les protéines représentent des structures biologiques fragiles dont la stabilité dépend de l'environnement dans lequel elles se trouvent (solvant, pH, concentration,...), et le passage progressif d'un état soluble à un état ordonné peut rapidement basculer en une précipitation vers un état agrégé.



**Figure 2.9.** Diagramme de phase d'une protéine.

La méthode utilisée au laboratoire pour la cristallo-genèse est dite par « diffusion en phase gazeuse par goutte pendante » (**Figure 2.10.**). La goutte contient un mélange de la solution protéique pure et d'une solution tamponnée identique à celle présente dans le réservoir (généralement, le volume total de la goutte est de 2  $\mu\text{l}$  (1 : 1, v/v)). Un équilibre s'établit alors entre la goutte et le réservoir par diffusion des espèces volatiles, jusqu'à ce que la tension de vapeur soit identique dans les deux compartiments. L'évaporation d'eau dans la goutte, plus diluée que le réservoir, concentre la protéine et les agents précipitants de manière à atteindre l'état de sursaturation, favorable à la formation de cristaux. Cependant, la méthode reste empirique et les chances de réussite aléatoires tant les paramètres physico-chimiques intervenant sont nombreux. En effet, la solubilité d'une protéine dépend de nombreux paramètres tels que la concentration de la protéine, le pH de la solution qui influence la charge globale de la protéine, la température, la force ionique de la solution, la présence d'ions (calcium, magnésium...), d'agents précipitants (sels, polymères (PEG), alcools...) dont la fonction est de favoriser les interactions protéine-protéine, en brisant la sphère de solvation.



**Figure 2.10.** Système de cristallisation par « diffusion en phase gazeuse par goutte pendante ».

Des tests commerciaux criblant de multiples conditions représentent donc souvent le point de départ des essais. Pour les essais de cristallisation de DiscII-boucleHPA-mutée, les kits testés étaient les « Clear Strategy Screen » (Molecular Dimension Ltd, avec le tampon Tris-HCl 0,1 M, pH 7,5) et le « PACT screen » de Qiagen.

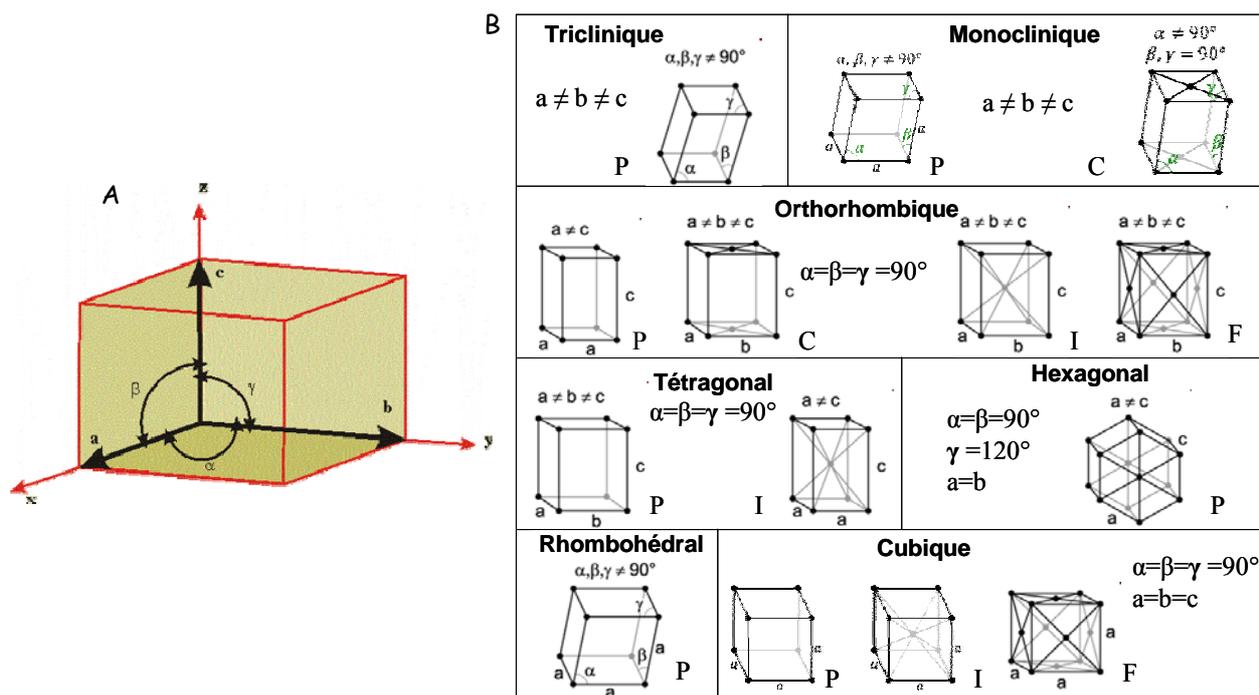
Lors de l'étude de complexe protéine-ligand, deux techniques existent pour l'introduction des ligands dans les cristaux. La méthode de « trempage » consiste à baigner un cristal natif de protéine non complexée dans une solution contenant le ligand, qui va diffuser dans le cristal pour se fixer sur la protéine. La méthode de « co-cristallisation » consiste, quant à elle, à incuber la protéine avec le ligand avant de réaliser les essais de cristallisation. C'est cette méthode qui a été utilisée pour l'obtention des cristaux de DiscI complexée avec les ligands Gal $\beta$ 1-3GalNAc et GalNAc $\beta$ 1-3Gal.

Avant d'être soumis aux rayons X, les cristaux de protéines sont congelés dans l'azote liquide, de manière à limiter les dommages causés par les rayonnements ionisants. Pour cela, un trempage préalable des cristaux dans une solution cryoprotectrice (PEG, glycérol, huile minérale...) est réalisé si nécessaire afin de prévenir la formation de cristaux de glace.

## 5.2. Résolution de structures par cristallographie aux rayons X

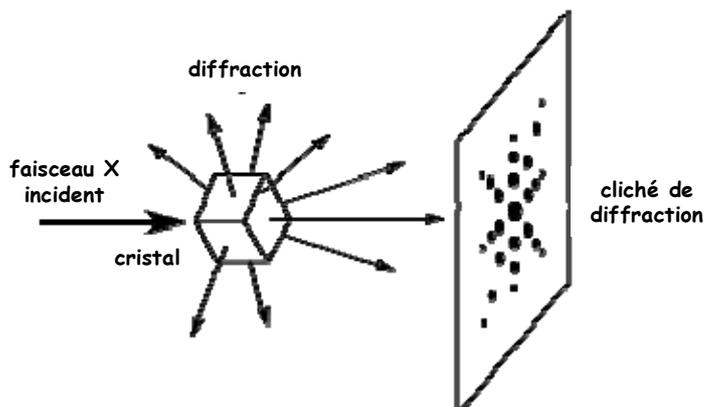
### 5.2.1. La diffraction des rayons X par les cristaux

Un cristal est un objet périodique formé par la répétition à l'identique d'un motif géométrique dans les trois directions de l'espace. Le plus petit volume translaté dans ces trois directions qui permet de générer un cristal entier est appelé maille élémentaire (**Figure 2.11. A.**). Elle est définie par trois vecteurs  $a$ ,  $b$  et  $c$  et trois angles notés  $\alpha$ ,  $\beta$  et  $\gamma$ . Il existe 14 types différents de maille élémentaire pour décrire l'arrangement de points en 3-D encore appelés réseaux de Bravais (**Figure 2.11. B.**). Des opérations de symétries (translations et rotations) peuvent également être appliquées dans la maille. Ces différents éléments définissent le groupe d'espace de la molécule cristallisée.



**Figure 2.11. A. Maille élémentaire et paramètres à trois dimensions. B. Les 14 réseaux de Bravais.**

Le nuage électronique de chaque atome d'un cristal a la capacité de diffracter les rayons X émis par un faisceau incident dans toutes les directions de l'espace. Chaque rayon diffracté, lorsqu'il n'est pas détruit par ces voisins lors d'interférences destructrices, est observé sous forme de tache noire, appelée aussi réflexion sur un cliché de diffraction (**Figure 2.12**). La position de la réflexion indique la direction selon laquelle le rayon a été diffracté par le cristal et son intensité quantifie la force du rayon diffracté. Ces deux paramètres sont indispensables à la reconstruction de l'image des molécules présentes dans le cristal. La direction du faisceau est spécifiée par une série de trois coordonnées ou indices de Miller  $h$ ,  $k$  et  $l$  pour chaque réflexion et son intensité est dénommée  $I_{hkl}$ . Le cliché de diffraction peut être considéré comme l'image inverse du cristal et représente l'espace réciproque. De manière à mesurer les directions et intensités de toutes les réflexions possibles, le cristal est mis en rotation durant la collecte des données. Ainsi, un cliché de diffraction est obtenu pour chaque angle de rotation donné (Rhodes 2006).



**Figure 2.12. Faisceau de rayons X diffracté par un cristal donnant naissance au cliché de diffraction**

### 5.2.2. Facteurs de structure et densité électronique

Chaque réflexion  $hkl$  du réseau réciproque peut être décrite comme une onde (toujours caractérisée par son amplitude et sa phase) résultant de la somme des diffractions de tous les électrons appartenant au plan  $hkl$  correspondant. Cette équation mathématique est alors appelée facteur de structure  $F_{hkl}$ .

$$F_{hkl} = \iiint \rho(x,y,z) e^{2\pi i(hx+ky+lz)} dV$$

La densité électronique  $\rho$  représente la distribution des électrons des molécules ayant diffractées les rayons incidents. Comme les molécules sont ordonnées dans un cristal, la densité électronique peut être décrite par une fonction mathématique périodique. Le graphique correspondant à cette fonction est donc l'image du nuage électronique qui entoure les molécules et est appelé carte de densité électronique. La densité électronique est liée au facteur de structure grâce à la transformée de Fourier. L'intensité  $I_{hkl}$  de chaque réflexion donne l'amplitude de l'onde qui la décrit et la position  $hkl$  précise sa phase. Malheureusement, aucun détecteur n'est capable d'enregistrer sa phase (Rhodes 2006).

### 5.2.3. Le problème des phases

Ce problème peut être détourné par différentes méthodes. Elles utilisent une équation appelée fonction de Patterson qui est indépendante des phases et dépend seulement des

intensités déterminées expérimentalement. C'est la transformée de Fourier des intensités. De manière expérimentale, les phases peuvent être obtenues par la méthode de remplacement isomorphe multiple « MIR multiple isomorphous replacement », où plusieurs atomes dits lourds sont introduits dans le cristal ou par les méthodes SAD « Single anomalous diffraction » ou MAD « Multiwavelength anomalous diffraction » utilisant la diffusion anormale de certains atomes dans le cristal. N'ayant pas été utilisées pour la résolution des structures de DiscI et DiscII-boucleHPA-mutée, ces méthodes ne seront pas davantage développées. En revanche, lorsque la protéine dont la structure reste à résoudre possède un homologue de structure connue (similarité de séquence d'au moins 20%), la méthode du remplacement moléculaire devient la méthode de choix pour résoudre le problème des phases (Rossmann 1962). Elle consiste à placer la molécule modèle dans la maille de la molécule de structure inconnue de façon à utiliser les phases ainsi obtenues comme phases de départ. Pour cela, une recherche en six dimensions de toutes les orientations et translations du modèle est nécessaire. Les programmes de recherche automatique fonctionnent le plus souvent en effectuant deux recherches tridimensionnelles : dans un premier temps une fonction de rotation permet d'orienter la molécule dans la maille puis sa position est déterminée par la fonction de translation. De nos jours des programmes de recherche automatique de plus en plus puissants sont disponibles et permettent, dans le meilleur des cas, d'obtenir une solution en quelques minutes. Pour la résolution des structures cristallographiques de DiscII-boucleHPA-mutée et DiscI en complexe avec les disaccharides Gal $\beta$ 1-3GalNAc et GalNAc $\beta$ 1-3Gal, le programme PHASER a été utilisé (McCoy 2007). Un modèle préliminaire peut alors être obtenu.

#### 5.2.4. Affinement

Grâce au calcul des phases, une première carte de densité électronique est calculée et visualisée à l'aide de logiciels graphiques tels que Coot (Emsley and Cowtan 2004). Pas toujours de qualité suffisante pour déterminer la position exacte de tous les atomes, elle permet toutefois de calculer un nouvel ensemble de phases et l'obtention d'une carte de densité améliorée.

Ce premier modèle est ensuite affiné. Le but est de minimiser les différences entre le facteur de structure calculé à partir du modèle ( $F_c$ ) et le facteur de structure observé expérimentalement ( $F_o$ ). Entre chaque cycle d'affinement, après inspection des cartes de densité électroniques  $2F_o - F_c$  et  $F_o - F_c$ , le modèle est ajusté manuellement. La carte  $2F_o - F_c$

donne la densité électronique pour le modèle observé et la carte Fo-Fc indique les différences de densités entre les cartes calculées et observées. La correspondance entre le modèle et les données expérimentales est mesurée grâce au facteur d'accord cristallographique R qui nous renseigne sur l'accord entre les amplitudes des facteurs de structure calculées et observées et la progression de l'affinement. Plus sa valeur est basse, plus le modèle est en accord avec les données. Le Rfree est un second critère de qualité. Il correspond au facteur R calculé à partir d'une fraction de réflexions (5% à 10% du jeu de données) qui ne sont pas prises en compte au cours de l'affinement. Si la structure n'est pas suraffinée, les facteurs R et Rfree doivent diminuer en parallèle.

Les paramètres structuraux du modèle final sont vérifiés et comparés aux valeurs typiques des structures protéiques avec les logiciels PROCHECK ou WHATIF (Vriend 1990 ; Laskowski 1993). La structure peut alors être déposée dans la base de données internationale RCSB Protein Data Bank (Bernstein *et al.* 1977).

## 6. Analyse des glycoépitopes reconnus par HPA

Cette partie a été réalisée en collaboration avec le Dr. Miriam Dwek de l'université de Westminster à Londres pour les étapes de culture cellulaire et de purification sur colonne HPA et avec le Dr. Bérangère Tissot de l'Imperial College de Londres pour les analyses par spectrométrie de masse.

### 6.1. Principe de la spectrométrie de masse

La spectrométrie de masse (MS) est une méthode d'analyse rapide, sensible et économique en échantillons qui permet d'identifier et de caractériser des molécules d'intérêt selon leur masse. Son principe est basé sur la séparation, sous l'action d'un champ magnétique, de molécules ionisées gazeuses en fonction de leur rapport masse sur charge ( $m/z$ ). Un détecteur placé en fin de course, couplé à un ordinateur permet d'analyser et de tracer, en temps réel, le spectre de l'intensité relative des molécules en fonction du rapport  $m/z$  (**Figure 2.13. A.**). Dès le début des années 1980, cette technique s'est développée dans le

---

but d'analyser les glycanes portés à la surface des glycoconjugués et a donné naissance à une nouvelle discipline, la glycomique.

A ce jour, le nombre de sources d'ionisation et d'analyseurs est très important. La liste suivante ne reprend que les techniques que j'ai rencontrées lors de ces travaux et n'est donc absolument pas exhaustive.

La source d'ionisation la plus connue est probablement le MALDI (matrix-assisted laser desorption/ionization). Dans cette technique, les échantillons sont co-cristallisés en présence d'une matrice capable d'absorber de l'énergie laser, générant ainsi, par vaporisation, des ions de l'échantillon cible. Cette méthode d'ionisation douce produit généralement des ions moléculaires mono chargés où le rapport  $m/z$  est donc équivalent à la masse de l'échantillon. L'ionisation par électrospray (ES) est également très employée. Elle permet, en appliquant un voltage approprié, de vaporiser en aérosol l'échantillon contenu dans un faible flux de liquide. Les microgouttes hautement chargées sont ensuite séchées par des gaz desséchants. Dans ce cas, la distribution de charge est proportionnelle au nombre de groupements ionisables de la molécule. Egalement très employée en glycomique, la spectrométrie de masse en tandem (MS/MS) consiste à sélectionner, après une première ionisation, un ion moléculaire parent, à le fragmenter à nouveau, par collision notamment et à analyser les ions fils ainsi générés. Cette technique permet de déterminer, grâce aux informations connues sur les voies de biosynthèse, la nature des oses présents dans l'échantillon.

La technologie d'analyseurs TOF pour Time-Of-Flight (ou temps de vol) a considérablement amélioré la sensibilité et la résolution des spectres. En effet, plus le « temps de vol » des molécules dans l'analyseur est long, meilleure sera leur séparation en fonction du rapport  $m/z$ . Pour ce faire, une série de miroirs, placés à l'extrémité de l'appareil modifie la trajectoire des molécules ionisées jusqu'au détecteur, placé proche de la source (**Figure 2.13. B.**) (Jang-Lee *et al.* 2006).

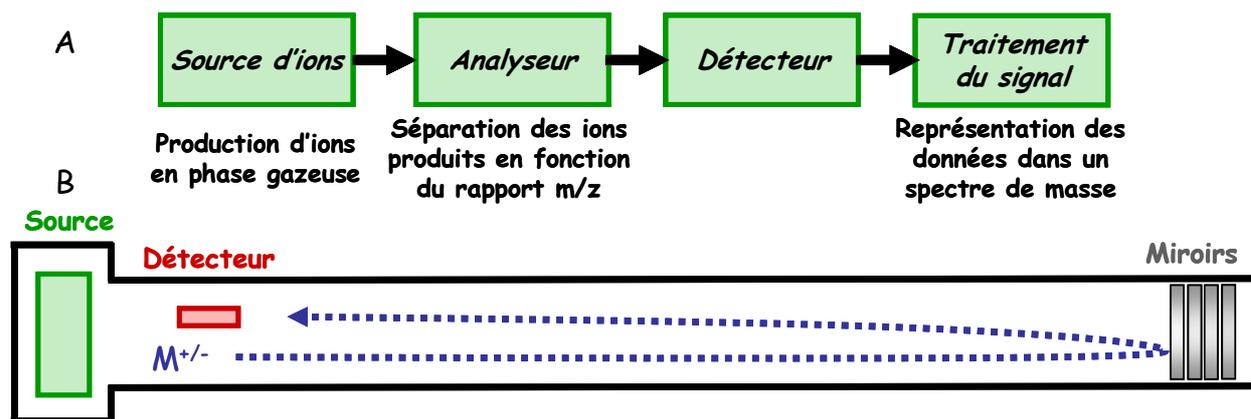


Figure 2.13. A. Structure d'un spectromètre de masse. B. Trajectoire d'un ion dans un analyseur à temps de vol

## 6.2. Traitement des échantillons

### 6.2.1. Isolation des glycoépitopes à partir des lignées cellulaires

#### 6.2.1.1. Préparation des cellules

Trois lignées cellulaires d'adénocarcinomes colorectaux ont été utilisées pour la caractérisation des glycoépitopes reconnus par HPA. Leurs caractéristiques sont présentées dans le **tableau 2.1**. La culture et la préparation des cellules ont été réalisées par le Dr. Miriam Dwek au sein de son laboratoire. Les cellules ont été cultivées selon un protocole classique de culture cellulaire excepté la composition du milieu qui ne contenait pas de sérum de veau fœtal. En effet, riche en protéines glycosylées, il aurait peut-être faussé les résultats de glycomique obtenus par spectrométrie de masse.

Brièvement, les cellules sont éclatées par sonication puis centrifugées à différentes vitesses de manière à récupérer dans le culot les glycoprotéines membranaires. 0,5 ml d'échantillons sont obtenus pour chaque lignée. 200  $\mu$ l sont directement traités en vue de l'analyse par spectrométrie de masse et 200  $\mu$ l sont purifiés par chromatographie d'affinité sur colonne HPA.

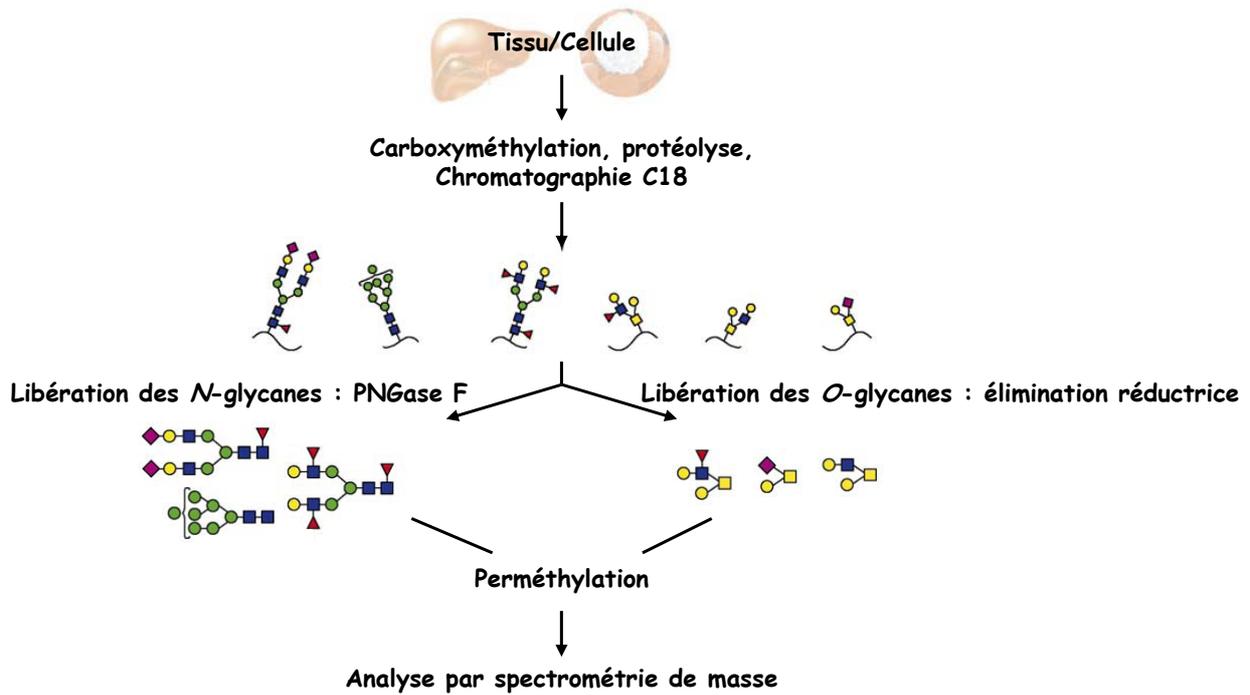
### 6.2.1.2. Purification sur colonne HPA

Après dilution des 200  $\mu$ l d'échantillons dans 300  $\mu$ l de tampon d'équilibration (Tris-HCl 20 mM, pH 7,4, NaCl 0,5 M, CHAPS 0,1% p/v), les protéines membranaires des trois lignées cellulaires sont chargées manuellement sur une colonne de chromatographie d'affinité HPA de 1 ml, équilibrée avec le tampon précédent. La colonne est ensuite lavée avec 5 ml de tampon puis les protéines retenues par HPA sont éluées avec 10 ml de tampon additionné de 250 mM de GlcNAc (tampon d'éluion). Les trois purifications sont réalisées successivement sur la colonne dans l'ordre suivant : SW 480, SW 620 et HT 29.

La concentration des protéines avant chargement sur la colonne et après purification est déterminée par la méthode de Bradford. Les échantillons avant et après purification sont dans deux tampons différents, respectivement le Phosphate Buffered Saline (PBS : NaCl 137 mM, KCl 2,7 mM,  $\text{Na}_2\text{HPO}_4$  10 mM,  $\text{KH}_2\text{PO}_4$  1,76 mM) et le tampon d'éluion. Deux courbes d'étalonnage avec de la sérum albumine bovine (BSA) sont donc réalisées.

La présence des glycoprotéines reconnues par HPA après purification est contrôlée par la méthode DotBlot, basée sur le principe de Western Blot. 10  $\mu$ g de protéines (contenues dans 2  $\mu$ l de tampon) sont déposés sur une membrane de nitrocellulose. Après évaporation complète, la membrane est saturée par une solution de BSA à 5 % dans l'eau pendant une heure sous agitation. La membrane est ensuite incubée pendant deux heures avec une solution d'HPA biotinylée commerciale à 5  $\mu\text{g}\cdot\text{ml}^{-1}$  dans du tampon Tris-buffered Saline-Tween TBS-T (Tris-HCl 50 mM, pH 7,4, NaCl 150 mM, Tween 20 0,05%) puis dans une solution de streptavidine-peroxydase à 2  $\mu\text{g}\cdot\text{ml}^{-1}$  durant une heure. La révélation est réalisée dans une solution de diaminobenzidine (DAB) et de peroxyde d'hydrogène dans du TBS pendant dix minutes. Chaque étape est suivie de trois lavages de 5 minutes dans du TBS-T. Les échantillons après purification sont parallèlement traités comme les extraits protéiques membranaires en vue de l'analyse des glycanes par spectrométrie de masse.

Les principales étapes de l'analyse glycomique par spectrométrie de masse sont reprises dans la **Figure 2.14**.



**Figure 2.14.** Schéma détaillant les principales étapes d'une analyse glycomique par spectrométrie de masse.

Ce schéma est tiré de (Haslam *et al.* 2006).

### 6.2.2. Réduction et carboxyméthylation

Les étapes de réduction et de carboxyméthylation sont respectivement utilisées pour réduire les éventuels ponts disulfures présents dans la protéine et saturés, par des groupements acétyles, les sulfures réactifs des cystéines.

Les échantillons sont réduits dans 0,6 M Tris-acide acétique, pH 8,5, NaCl 150 mM, EDTA 5 mM, CHAPS 1 % p/v, dithiothréitol (DTT) 10 mg.ml<sup>-1</sup>, sous un flux d'azote à 37°C pendant 1 heure 30. La réaction de carboxyméthylation est réalisée dans une solution d'acide iodoacétique (10 fois en excès molaire par rapport au DTT) pendant deux heures à température ambiante dans l'obscurité. Après quatre bains de dialyse dans 4,5 L de bicarbonate d'ammonium 50 mM, pH 8,5, pendant 48 heures, les échantillons sont lyophilisés.

### 6.2.3. Digestion peptidique à la trypsine

Les échantillons sont incubés avec de la trypsine de porc (Promega) selon un ratio de 50 : 1 (p/p) dans un tampon de bicarbonate d'ammonium 50 mM, pH 8,4, à 37°C pendant seize heures puis purifiés sur résine Sep-Pak Oasis HLB Plus (Waters). La résine est tout d'abord équilibrée avec 5 ml de méthanol (MeOH), 5 ml d'acide acétique 5 %, 5 ml de propan-1-ol et trois fois 5 ml d'acide acétique. Puis l'échantillon est chargé et lavé quatre fois avec 5 ml d'acide acétique. L'élution a lieu avec des solutions à teneur croissante en propan-1-ol : 4 ml de propan-1-ol 20%, 40% et 100%. Les produits élués dans chaque solution sont ensuite lyophilisés.

### 6.2.4. Récupération des *N*-glycanes par digestion à la PNGase F

Les fractions éluées dans 20 % et 40 % de propan-1-ol sont combinées et dissoutes dans 150 µl de tampon de bicarbonate d'ammonium 50 mM, pH 8,4 et digérées par cinq unités de l'enzyme Peptide : *N*-glycosidase F (PNGase F – Roche Molecular Biochemicals) à 37°C pendant 24 heures. Cette enzyme a la capacité de cliver les *N*-glycanes des glycoprotéines. Les échantillons sont ensuite lyophilisés et purifiés par chromatographie en phase inverse sur résine Sep-Pak C<sub>18</sub> (Waters) selon un protocole quasi similaire à celui décrit pour la purification précédente. Les *N*-glycanes sont récupérés dans une étape d'élution préliminaire supplémentaire dans 5 ml d'acide acétique 5%. Les fractions éluées à 20% et 40% d'isopropan-1-ol, contenant notamment les *O*-glycanes accrochés sur des fragments peptidiques, sont combinées. Les échantillons contenant les *N*- et *O*-glycanes sont désormais traités séparément.

### 6.2.5. Récupération des *O*-glycanes par β-élimination réductrice

Cette réaction est réalisée dans 400 µl d'une solution de borohydrure de potassium (KBH<sub>4</sub>) (55 mg.ml<sup>-1</sup> dans une solution d'hydroxyde de potassium 0,1 M) à 45°C pendant seize heures. La réaction est stoppée par addition de cinq gouttes d'acide acétique et est suivie d'une chromatographie échangeuse d'ions sur résine Dowex 50W-X8. Après un lavage de la résine avec 15 ml d'acide acétique 5%, l'échantillon chargé est élué avec la solution d'acide acétique 5%, séché sous un flux d'azote et lyophilisé. Les excès de borate sont ensuite

---

éliminés par quatre co-évaporations successives en utilisant 500  $\mu$ l d'une solution d'acide acétique 10% dans MeOH.

#### 6.2.6. Perméthylation

Les *N*- et *O*-glycanes sont perméthylés de manière à augmenter leur stabilité et améliorer la qualité du spectre. Dans un mortier parfaitement sec, dix pastilles d'hydroxyde de sodium (NaOH) sont broyées dans 6 ml de diméthylsulfoxyde (DMSO) de manière à former une pâte laiteuse. 1 ml de cette préparation est rapidement ajouté aux glycanes secs avec 0,5 ml d'iodométhane. La réaction se déroule sous agitation pendant 20 minutes à température ambiante et est stoppée par ajout de quelques gouttes d'eau. Les glycanes perméthylés sont ensuite récupérés par extraction au chloroforme. Environ 2 ml de chloroforme et 3 ml d'eau sont ajoutés. L'échantillon est alors fortement agité puis la phase aqueuse supérieure est éliminée. Cette étape est renouvelée plusieurs fois. Le chloroforme est ensuite séché sous flux d'azote. Les glycanes perméthylés sont alors purifiés sur Sep-Pak C<sub>18</sub>. La résine est lavée successivement avec 5 ml de MeOH, 5 ml d'eau, 5 ml d'acétonitrile 100% et trois fois 5 ml d'eau. L'échantillon, dissous dans 200  $\mu$ l d'un mélange eau/MeOH (1 : 1 v/v) est chargé sur la colonne, lavé avec 5 ml d'eau, élué avec des solutions aqueuses d'acétonitrile à 15%, 35%, 75% et 100% et lyophilisé avant d'être analysé.

#### 6.2.7. Analyses par spectrométrie de masse MALDI

Les échantillons perméthylés sont dissous dans 10  $\mu$ l de MeOH dont 2  $\mu$ l sont mélangés selon un ratio 1 : 1 (v/v) avec la matrice d'acide 2,5-dihydrobenzoïque DHB (20 mg dilués dans un mélange MeOH /eau (70 : 30)).



# Résultats



# Chapitre 3

Utilisation d'HPA, DiscI et  
DiscII pour le marquage  
de cellules cancéreuses



---

## Chapitre 3

# Utilisation des lectines HPA, DiscI et DiscII pour le marquage de cellules cancéreuses

---

HPA, lectine spécifique du résidu GalNAc et isolée à partir de la glande à albumine de l'escargot de Bourgogne, est utilisée depuis plus de vingt ans en histopathologie comme marqueur de cancers métastatiques, associés à un mauvais pronostic vital (Leathem and Brooks 1987). Son utilisation en routine a cependant été limitée par l'impossibilité à produire cette lectine sous forme recombinante.

Les alignements de séquence ont révélé la présence de plusieurs homologues à HPA chez les invertébrés dont les lectines à résidus galactosylés appelées discoidines de l'amibe *Dictyostelium discoideum*.

La microscopie confocale est une technique de microscopie optique qui permet d'obtenir des images très nettes d'échantillons. Elle a notamment été utilisée pour démontrer la spécificité d'HPA dans la reconnaissance de cellules métastatiques issues de lignées de cancers colorectaux (Saint-Guirons *et al.* 2007).

Un des objectifs ce travail de thèse était de tester les discoidines comme marqueur de cellules cancéreuses. A la différence d'HPA, les discoidines sont facilement exprimées sous forme recombinante. La découverte de nouvelles protéines, possédant les mêmes propriétés prédictives qu'HPA, faciliterait le développement de nouveaux outils de diagnostic en cancérologie.

Ces expériences de marquage de lignées cellulaires par des lectines et détection en microscopie confocale à fluorescence ont été réalisées en collaboration avec le Dr Miriam V. Dwek lors de visites dans son laboratoire de l'université de Westminster à Londres.

## 1. Mise au point expérimentale avec HPA

Les trois lignées cellulaires utilisées comme modèle de tissus cancéreux sont issues de cancers colorectaux humains. La lignée SW 480 est non métastatique alors que les lignées SW 620 et HT 29 le sont. Les caractéristiques des cellules sont résumées dans le **tableau 3.1**.

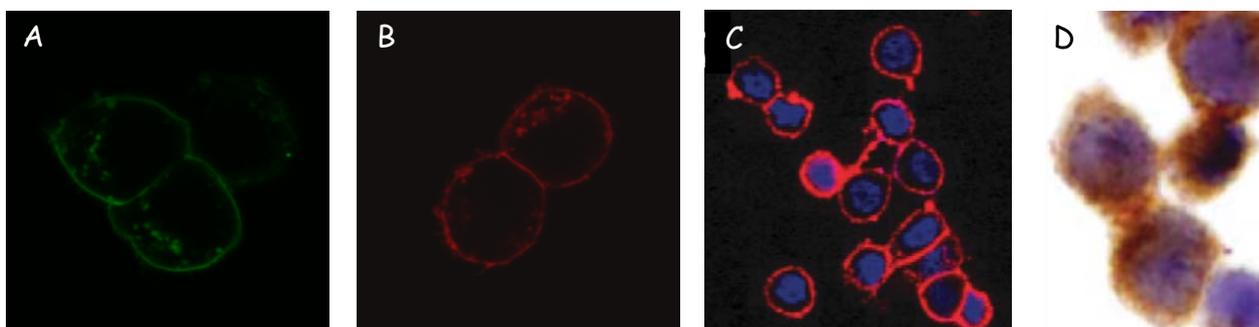
Nom lignée	Numéro ATCC	Stade tumoral	Genre/Âge/Type
SW480	CCL-228 <sup>TM</sup>	Lignée considérée comme non métastatique, établie à partir du cancer primaire.	Homme de 50 ans de type caucasien
SW620	CCL-227 <sup>TM</sup>	Lignée considérée comme métastatique, établie à partir de métastases dans les ganglions lymphatiques.	Même patient que SW480, un an plus tard
HT-29	HTB-38 <sup>TM</sup>	Lignée considérée comme très fortement métastatique.	Femme de 44 ans de type caucasien

**Tableau 3.1. Caractéristiques des lignées de cancers colorectaux utilisées.**

Différents protocoles de marquage de cellules par HPA ont été testés dans la littérature. Le premier, utilisé sur la lignée HT 29, est une méthode de détection colorimétrique qui repose sur l'utilisation d'HPA commerciale biotinylée et d'un conjugué streptavidine-peroxydase. Le produit de la réaction enzymatique est révélé par un chromogène (DAB) qui le colore en brun-rouge (Brooks and Leathem 1998 ; Lescar *et al.* 2007) (**Figure 3.1. D.**). Le second est une méthode directe qui utilise la microscopie confocale à fluorescence. Elle implique l'utilisation d'une solution d'HPA couplée à un fluorophore. Le cyanine-5, de couleur rouge, a été utilisé pour le marquage de la lignée HT 29 (**Figure 3.1. C.**) dans l'article (Lescar *et al.* 2007) alors que j'ai moi-même utilisé le TRITC, également de couleur rouge, pour le marquage de la lignée SW 620 (**Figure 3.1. B.**). Nous avons également testé un mode de révélation indirect. Après fixation d'HPA biotinylée, la révélation est réalisée en utilisant un fluorophore FITC, coloré en vert, couplée à de la streptavidine (**Figure 3.1. A.**). La faible intensité de coloration des images A et B peut davantage être imputée à la qualité des images qu'à un marquage limité de SW 620 par HPA. Quelque soit le mode de

détection utilisé et la lignée considérée, la coloration apparaît, de manière intense, au niveau de la surface cellulaire.

La spécificité du marquage de lignées et tissus métastatiques par HPA a été démontrée de nombreuses fois (Schumacher and Adam 1997 ; Lescar *et al.* 2007 ; Saint-Guirons *et al.* 2007). Cette expérience préliminaire de marquage de cellules avec HPA que j'ai réalisée est utilisée comme contrôle. Ainsi, nous savons que les deux modes de détection (direct et indirect) sont valides et que les cellules ne présentent, à priori, pas de phénotypes aberrants. En effet, la culture cellulaire *in vitro*, particulièrement lorsqu'un processus de cancérisation est présent, peut entraîner une modification de la glycosylation de surface et perturber la reconnaissance des lectines. Avec ces manipulations, nous nous affranchissons d'un problème expérimental. Tous les résultats à venir peuvent donc être considérés comme significatifs.



**Figure 3.1. Marquage des lignées cellulaires de cancer colorectal SW 620 et HT 29 par la lectine HPA.**

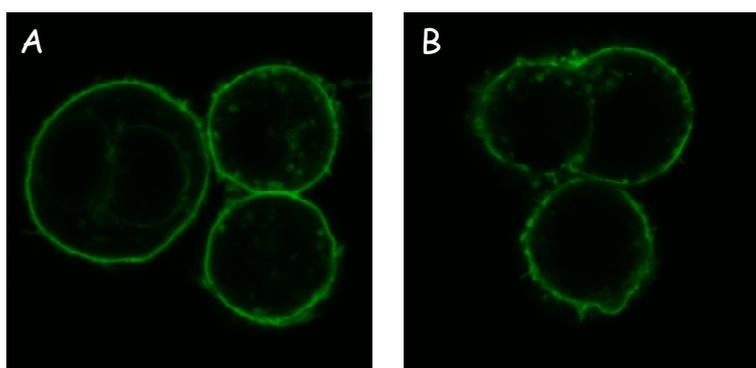
Les images A. et B. ont été acquises au cours de ce travail de thèse. Les images C. et D. sont issues de (Lescar *et al.* 2007).

**A.** Le marquage de la lignée SW 620 est réalisé par une solution d'HPA commerciale biotinylée ( $c = 10 \mu\text{g.ml}^{-1}$ ) et révélé par le fluorophore FITC couplé à la streptavidine ( $c = 4 \mu\text{g.ml}^{-1}$ ). **B.** Le marquage de la lignée SW 620 est directement réalisé par une solution d'HPA commerciale couplée au fluorophore TRITC ( $c = 10 \mu\text{g.ml}^{-1}$ ). **C.** Le marquage de la lignée HT 29 est réalisé par une solution d'HPA couplée au fluorophore Cy5. **D.** Le marquage de la lignée HT 29 est réalisé à partir d'une solution d'HPA commerciale biotinylée et révélé par un complexe streptavidine-peroxydase en présence d' $\text{H}_2\text{O}_2$ .

## 2. Marquages cellulaires avec DiscI et DiscII

### 2.1. Tests avec les discoidines sur la lignée SW 620

Un test préliminaire de marquage de la lignée métastatique SW 620, par microscopie confocale, a été réalisé avec les lectines DiscI et DiscII biotinylées, produites sous forme recombinante (**Figure 3.2.**). Comme pour HPA, une coloration très intense des membranes cellulaires est visible. Une préincubation des lectines dans une solution de GalNAc à 0,1 M inhibe la fixation de ces protéines ; aucun marquage n'est observé (résultats non présentés). Cette expérience démontre que l'interaction entre les discoidines et les épitopes membranaires est bien de type lectine-sucre. Des contrôles positif et négatif ont également été réalisés avec, respectivement, la lectine HPA biotinylée et sans lectine. L'absence de marquage, pour cette dernière condition, démontre la spécificité de la reconnaissance.



**Figure 3.2.** Marquage de la lignée cellulaire de cancer colorectal SW 620 par **A. DiscI** et **B. DiscII**.

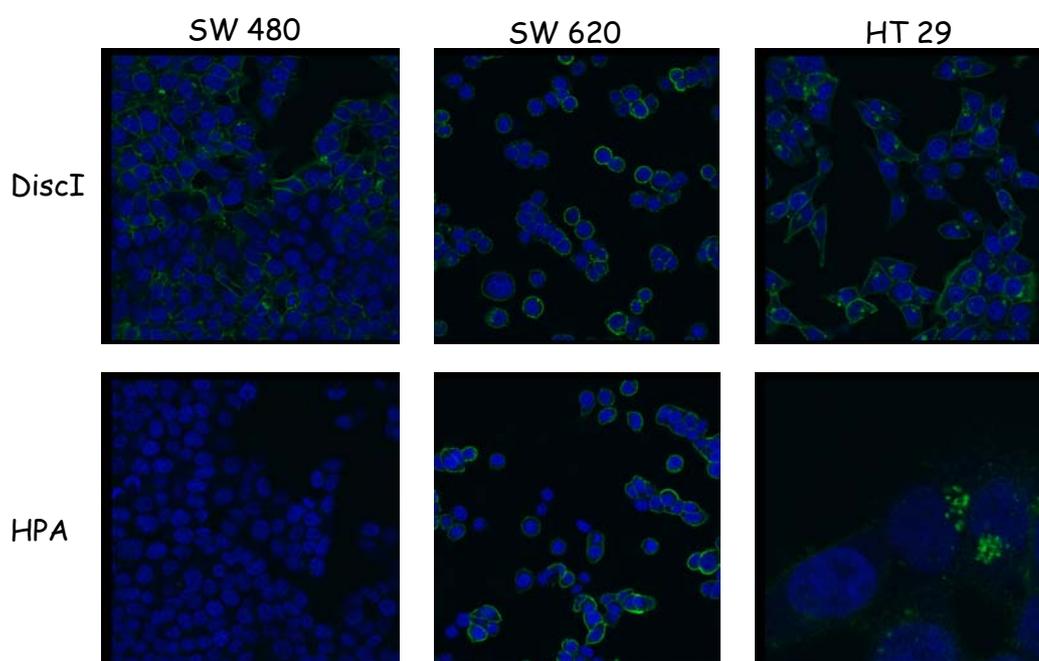
Le marquage est réalisé par une solution de lectines biotinylées ( $c = 10 \mu\text{g.ml}^{-1}$ ) et révélé par le fluorophore FITC couplé à la streptavidine ( $c = 4 \mu\text{g.ml}^{-1}$ ).

### 2.2. Marquages des trois lignées cellulaires

Durant la totalité des manipulations, les lectines sont conservées à 4°C dans leur tampon dit de microscopie à pH 7,4. Le point isoélectrique théorique de DiscII est de 7,2, valeur très proche du pH du tampon, ce qui a entraîné sa précipitation en solution. Aucune expérience complémentaire n'a alors pu être envisagée. En effet, le point isoélectrique

correspond au pH pour lequel la charge globale de la protéine est équivalente à zéro. Aucune répulsion n'est observée entre les différentes molécules qui ont alors tendance à s'agréger et précipiter. Le point isoélectrique estimé de DiscI est de 6,7, valeur suffisamment éloignée du pH du tampon pour ne pas provoquer la précipitation de la protéine.

Selon un protocole identique à celui testé précédemment, les trois lignées cellulaires SW 480, SW 620 et HT 29 ont été marquées par DiscI et HPA (**Figure 3.3.**). Le noyau des cellules est coloré en bleu par le colorant To-Pro-3. Un marquage spécifique par DiscI est visible sur les trois lignées cellulaires testées alors qu'HPA se fixe uniquement aux lignées métastatiques SW 620 et HT 29. DiscI semble donc reconnaître un épitope différent de celui fixé par HPA et présent quelque soit l'avancement du processus de métastases. La résolution des structures cristallographiques de DiscI et HPA en complexe respectivement avec les disaccharides Gal $\beta$ 1-3GalNAc et GalNAc $\beta$ 1-3Gal et l'antigène de Forssman (GalNAc $\alpha$ 1-3GalNAc) révèlent que leurs sites de reconnaissance ne présentent pas la même configuration, celui d'HPA étant plus étroit et profond (Mathieu *et al.* 2010). Ces différences structurales sont à l'origine d'une spécificité propre à chaque lectine et peuvent expliquer les marquages différents observés en microscopie confocale.



**Figure 3.3. Marquage des lignées cellulaires de cancers colorectaux non métastatiques SW 480 et métastatiques SW 620 et HT 29 par DiscI et HPA et révélation par microscopie confocale.**

Le marquage est réalisé par une solution de lectines biotinyllées ( $c = 10 \mu\text{g.ml}^{-1}$ ) et révélé par le fluorophore FITC couplé à la streptavidine ( $c = 4 \mu\text{g.ml}^{-1}$ ). Les noyaux sont marqués par un fluorophore bleu appelé To-Pro-3.

Nous pouvons également constater que le marquage par HPA de la lignée HT 29 est très surprenant. Il se présente sous forme de grains localisés et non plus, de manière homogène, sur l'intégralité de la surface cellulaire (**Figure 3.1.**). La solution de lectine utilisée pour les lignées SW 620 et HT 29 étant identique, le problème proviendrait vraisemblablement des cellules adhérentes sur la plaque. Ce résultat peut être considéré comme non significatif.

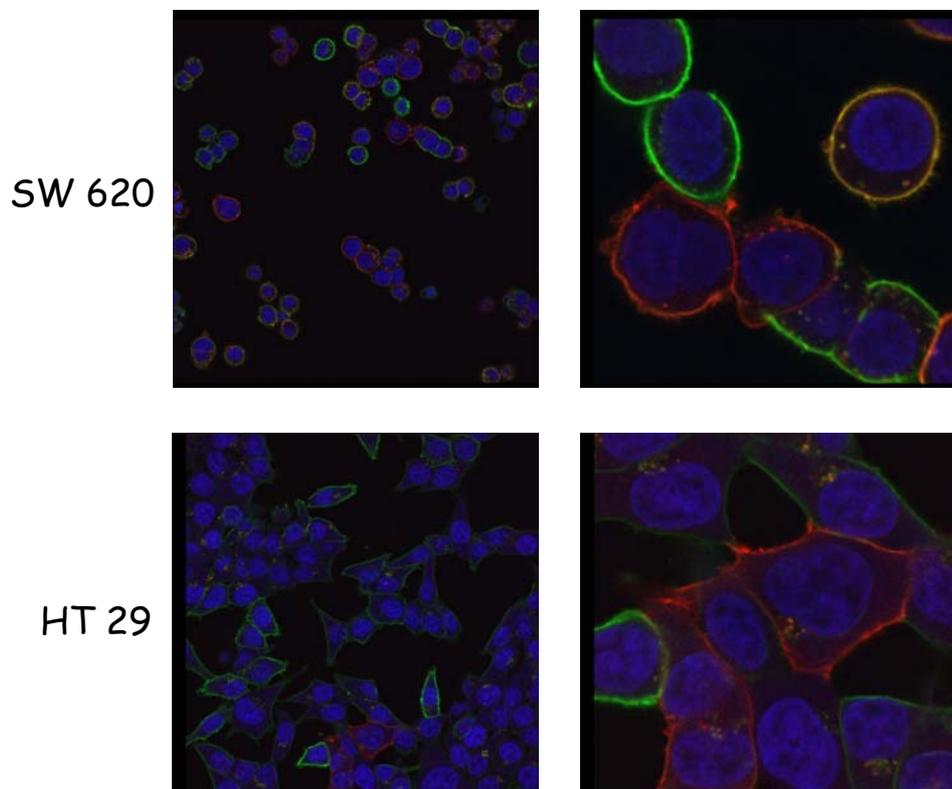
Quelque soit la lectine et/ou la lignée considérées, toutes les cellules ne sont pas marquées. La quantité de cellules reconnues est probablement proportionnelle à la concentration en lectines utilisée, comme cela a été montré précédemment (Lescar *et al.* 2007) et n'est pas nécessairement corrélée à la force de l'interaction entre les lectines et les ligands cellulaires. Cependant, il est également envisageable que toutes les cellules d'une même lignée n'expriment pas nécessairement le ou les antigènes reconnus.

Ces expériences nous permettent de conclure que DiscI ne possède pas la capacité de discriminer spécifiquement les épitopes glycaniques présents sur des cellules métastatiques, comme HPA. Elle reconnaît un ligand présent quelque soit le stade d'avancement de la tumeur alors qu'HPA a la capacité de se fixer sur un épitope uniquement exprimé durant la transformation métastatique. Aux vues de ces résultats préliminaires, les discoidines ne semblent donc pas être des marqueurs tumoraux aussi intéressants qu'HPA.

### 2.3. Compétition entre DiscI et HPA

Un double marquage des lignées métastatiques SW 620 et HT 29 a également été testé avec DiscI biotinylée, révélé par le fluorophore FITC coloré en vert et HPA couplée au fluorophore TRITC coloré en rouge. Pour les deux lignées cellulaires considérées, certaines cellules sont reconnues uniquement par l'une des deux lectines, DiscI si la coloration apparaît verte et HPA si elle est rouge. Une troisième catégorie de cellules, colorées en orange est visible. Elle est reconnue par les deux lectines simultanément (**Figure 3.4.**). La glycosylation de surface des cellules à l'intérieur d'une même lignée n'est pas homogène. Certaines cellules vont uniquement présenter les épitopes reconnus par DiscI ou HPA alors que d'autres semblent exprimer les deux épitopes de manière simultanée. Vraisemblablement, il existe plusieurs sous-populations de cellules qui ne présentent pas la même glycosylation de surface. Ce phénomène peut-être provoqué par la présence de cellules dans différents états cellulaires.

Il est important de noter que le marquage de la lignée HT 29 avec HPA est bien homogène contrairement à ce qui avait été observé **Figure 3.3**. Le problème était donc bien isolé dans l'expérience précédente et les résultats ne doivent donc pas être considérés comme significatifs.



**Figure 3.4. Marquage des lignées cellulaires de cancers colorectaux métastatiques SW 620 et HT 29 par DiscI et HPA et révélation par microscopie confocale.**

Le marquage par DiscI est réalisé par avec solution de lectine biotinyllée ( $c = 10 \mu\text{g.ml}^{-1}$ ) et révélé par le fluorophore FITC couplé à la streptavidine ( $c = 4 \mu\text{g.ml}^{-1}$ ). La coloration apparaît verte. Le marquage d'HPA est réalisé avec une solution de lectine couplée au fluorophore TRITC, coloré en rouge ( $c = 10 \mu\text{g.ml}^{-1}$ ). Les noyaux sont marqués par un fluorophore bleu appelé To-Pro-3.

## 2.4. Inhibition par des monosaccharides

Des expériences d'inhibition ont également été réalisées. Elles consistaient à incuber les solutions de DiscI biotinyllée avec les monosaccharides suivants D-GalNAc, D-GlcNAc, D-Gal et  $\alpha$ -Me-Man à différentes concentrations. Dans un souci d'économie de matériel (solution de lectines), les cellules étaient adhérees sur de petites lamelles en verre et non plus

dans de gros puits en plastique. Le protocole restait néanmoins identique à celui décrit précédemment. Malheureusement, les cellules ne sont pas restées adhérentes au support en verre et aucun marquage n'a pu être visualisé.

### 3. Conclusion

HPA est utilisée depuis plus de vingt ans comme marqueur histopathologique de cellules cancéreuses à fort pouvoir métastatique et faible pronostic vital. Les discoidines, lectines de l'amibe *Dictyostelium discoideum*, sont homologues à HPA. Le but de ce travail consistait à tester le potentiel des discoidines comme marqueurs de cellules cancéreuses.

Malheureusement, DiscII a rapidement précipité et aucun marquage n'a pu être réalisé. En revanche, DiscI est capable de se fixer de manière spécifique sur des glycoépitopes présents à la surface des trois lignées de cancer colorectal alors qu'HPA se fixe uniquement sur les lignées métastatiques. DiscI ne possède donc pas la capacité de discriminer les cellules associées à faible pronostic vital.

La comparaison des structures cristallographiques de DiscI et HPA en complexe respectivement avec les disaccharides GalNAc $\beta$ 1-3Gal et Gal $\beta$ 1-3GalNAc et l'antigène de Forssman ( $\alpha$ GalNAc1-3GalNAc) (Mathieu *et al.* 2010) explique les différences constatées dans le marquage des cellules cancéreuses. Les deux lectines possèdent chacune une spécificité propre qui se reflète dans la reconnaissance des lignées de cellules cancéreuses.

Les épitopes cancéreux reconnus par HPA et DiscI ne sont pas identifiés, ce qui complique l'interprétation des résultats. Un marquage de DiscI sur une lignée de cellules saines du colon aurait permis de contrôler si cette lectine reconnaissait un épitope uniquement présent à la surface de cancers ou non. En effet, à ce stade de la recherche, nous ne pouvons pas affirmer que le ou les épitopes reconnus par DiscI sont uniquement présent après la transformation tumorale des cellules. Certains glycolipides localisés dans les membranes présentent naturellement des structures  $\beta$ 1-3GalNAc (le tétraose de la série Globo- par exemple) (Wennekes *et al.* 2009) et pourraient être des ligands potentiels de DiscI.

La spécificité d'HPA pour les cellules métastatiques peut s'expliquer par la prévalence de certains épitopes  $\alpha$ -GalNAc, comme les antigènes T (Gal $\beta$ 1-3GalNAc $\alpha$ ) et Tn ( $\alpha$ GalNAc) très fortement surexprimés lors du processus de développement de métastases. Cet aspect a

été largement développé dans l'introduction. Alors que la présence de GalNAc lié en  $\beta$ , reconnu par DiscI, à la surface des cellules métastatiques n'a jamais été pas particulièrement identifiée.

L'impossibilité de produire HPA sous forme recombinante limite considérablement son utilisation en histopathologie. La formation de protéines chimères discoidine-HPA permettrait de former des marqueurs recombinants en grande quantité et spécifiques des cellules métastatiques. Cet aspect du travail a été largement développé durant cette thèse et sera détaillé dans le chapitre 5.



# **Chapitre 4**

**Structures et spécificité de DiscI**

**Comparaison avec DiscII**



---

# Chapitre 4

## Structures et spécificité de DiscI

### Comparaison avec DiscII et HPA

---

Les résultats, obtenus en collaboration avec Dr. Karoline S. Aragão, sur les études structurales et fonctionnelles de la discoidine I (DiscI) ont fait l'objet d'une publication qui sera décrite dans ce chapitre.

Sophie V. Mathieu, Karoline S. Aragão, Anne Imberty & Annabelle Varrot. Discoidin I from *Dictyostelium discoideum* and interactions with oligosaccharides: specificity, affinity, crystal structures, and comparison with Discoidin II. *J. Mol. Biol.* 400, 540-554, 2010.

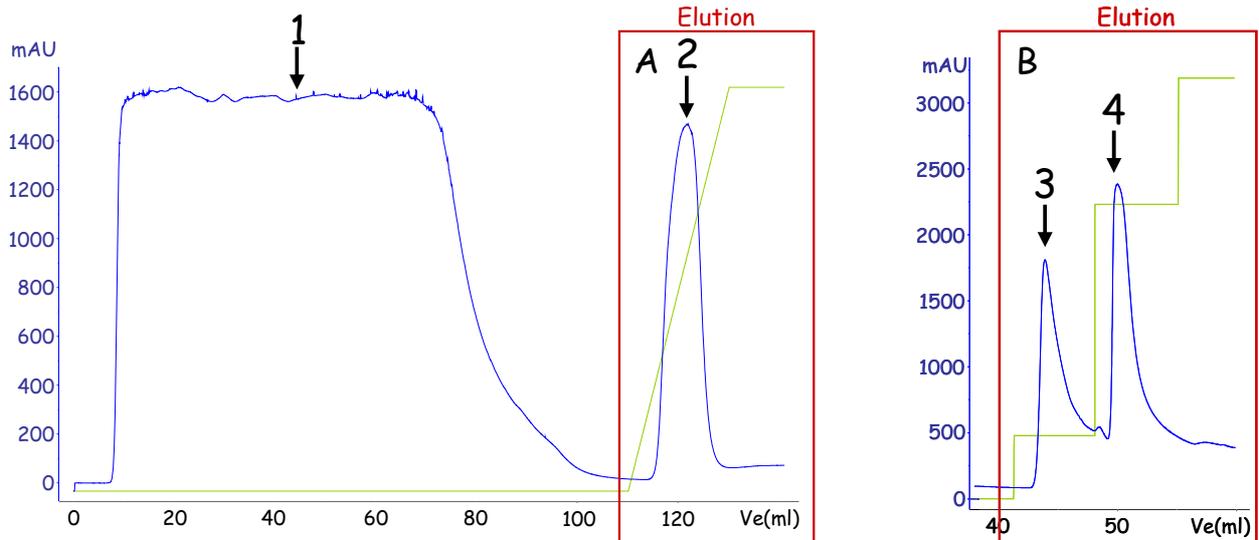
Dans cette partie, mon travail consistait à exprimer et purifier DiscI, à partir de protocoles mis au point par le Dr. K. S. Aragão durant son doctorat, afin d'obtenir une protéine très pure et assez concentrée pour les essais de cristallogénèse. L'obtention de cristaux de DiscI, en complexes avec des résidus mono- et disaccharidiques et la résolution de plusieurs structures cristallographiques ont permis de comprendre le mode d'interaction de cette lectine avec ses ligands et d'expliquer les différences d'affinité mesurées par microcalorimétrie isotherme de titration (ITC). Ce chapitre décrit ces différentes étapes.

## 1. Production de DiscI sous forme recombinante

La protéine DiscI, après clonage dans le vecteur d'expression pProEx-HtB, a été exprimée sous forme recombinante dans la souche bactérienne Rosetta 2 pLysS, qui a la particularité d'exprimer des protéines possédant des codons dits rares chez *E. coli*. De plus, grâce à la présence, à son extrémité N-terminale, d'une étiquette polyhistidine His<sub>6</sub>, clivable par la protéase TEV (Tobacco Etch Virus), DiscI, exprimée sous forme soluble, est ensuite

purifiée par chromatographie de type IMAC (Immobilized Metal Affinity Chromatography) sur résine de nickel (**Figure 4.1.**) et analysée sur gel de polyacrylamide en conditions dénaturantes (**Figure 4.2.**), selon les protocoles décrits dans le chapitre 2. Le tampon d'équilibration est le suivant : Tris-HCl 20 mM, pH 8,5, NaCl 500 mM et Imidazole 10 mM. Selon les indications du fournisseur (GE Healthcare), la rétention des protéines sur la colonne serait favorisée dans un tampon légèrement basique. De plus, la forte concentration en sel limite les effets d'échange d'ions et l'imidazole, à faible concentration, est utilisé pour minimiser les interactions non spécifiques entre les protéines d'*E. coli* et la colonne. L'élution est réalisée soit par gradient linéaire d'imidazole allant de 10 mM à 500 mM (**Figure 4.1. A.**), soit par palier à deux concentrations en imidazole, 75 mM pour décrocher les protéines contaminantes et 350 mM pour éluer la protéine d'intérêt (**Figure 4.1. B.**). La visualisation d'un pic d'élution, quelque soit la méthode utilisée, démontre que la protéine se fixe correctement aux ions nickel via sa queue polyhistidine. Sur le gel SDS-PAGE (**Figure 4.2.**), la présence d'une bande à environ 28 kDa très majoritairement présente dans la fraction soluble (piste 2) en comparaison avec la fraction agrégée (piste 1) prouve que, dans les conditions d'expression testées, la majorité de DiscI s'exprime sous forme soluble. Le pic 1 des chromatogrammes représentent toutes les protéines solubles d'*E. coli* qui n'interagissent pas avec les ions nickel de la colonne et forme la fraction non retenue (piste 3 **Figure 4.2.**). Les pics 3 et 4 représentent respectivement les protéines contaminantes d'*E. coli* qui ont interagit avec la colonne (piste 4) et DiscI pure (piste 5). Le pic 2 est composé, dans les premières fractions, d'un mélange entre les protéines contaminantes et la protéine désirée et, dans les dernières, de DiscI pure. Le rendement moyen est estimé à environ 20 mg de DiscI recombinante purifiée par litre de culture.

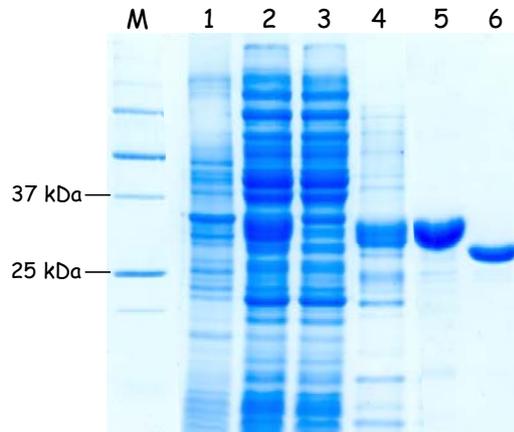
La purification se poursuit ensuite par l'élimination de l'imidazole grâce à une colonne de dessalage PD-10 (GE Healthcare). Cette étape permet de placer la protéine dans un tampon Tris-HCl 20 mM, pH 7,5, NaCl 150 mM, adéquat pour la suite des manipulations. L'étiquette polyhistidine est alors clivée par digestion enzymatique par la protéase TEV pendant une nuit. Une nouvelle étape de purification sur colonne de nickel permet alors de récupérer la protéine sans étiquette dans la fraction non retenue (piste 6 **Figure 4.2.**) alors que la protéine non clivée, la queue polyhistidine et la TEV, elle-même exprimée avec une étiquette His<sub>6</sub>, seront retenues sur les ions nickel de la colonne et éluées en présence d'imidazole.



**Figure 4.1. Profils d'éluion obtenus lors des purifications de *DisCI* par chromatographie sur colonne de nickel.**

**A.** Elution par gradient d'imidazole de 0 à 500 mM, **B.** Elution par palier à 75 mM et 350 mM d'imidazole.

Les courbes bleue et verte représentent respectivement la densité optique à 280 nm (exprimée en mAU) et la concentration d'imidazole appliquée pour éluer les protéines. L'encadré rouge représente les fractions éluées.

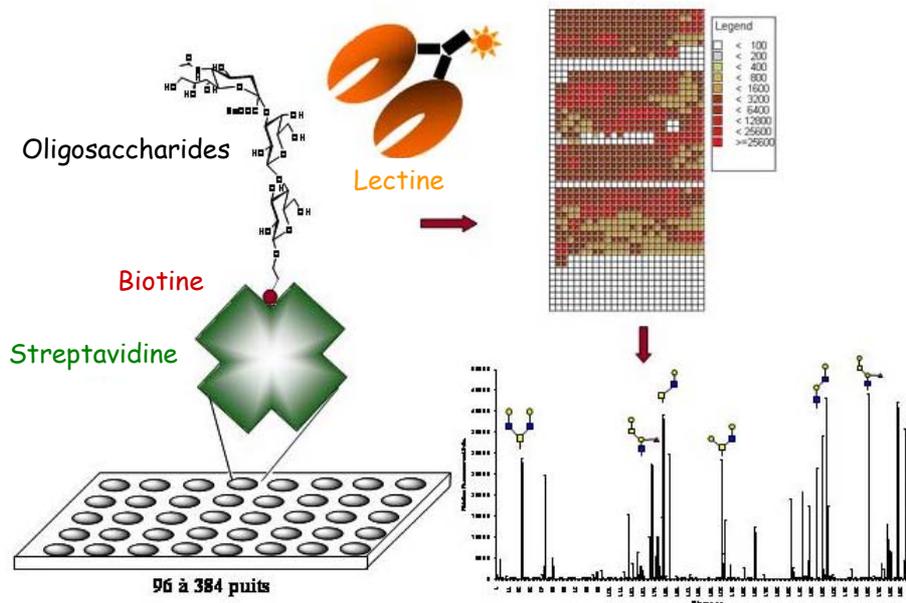


**Figure 4.2. Gel SDS-PAGE 12% acrylamide coloré au bleu de Coomassie après expression de *DisCI* dans Rosetta 2 pLysS et purification sur colonne de nickel.**

Ce gel correspond au chromatogramme B de la **Figure 4.1**. Les pistes 1 et 2 représentent respectivement la fraction contenant la protéine présente sous forme de corps d'inclusion (après centrifugation à 12000 g) et la fraction soluble (après centrifugation à 50000 g) à déposer sur la colonne. La piste 3 représente la fraction non retenue sur la colonne. Les pistes 4 et 5 correspondent aux deux pics majoritaires obtenus lors de l'éluion, respectivement pic 3 et pic 4. La piste 6 représente *DisCI* purifiée après clivage de l'étiquette His<sub>6</sub>.

## 2. Spécificité des lectines de *Dictyostelium discoideum*

La technique des puces à sucres, développée par le « Consortium for Functional Glycomics » (CFG ; <http://www.functionalglycomics.org>) est un outil rapide, mis à la disposition de la communauté scientifique, pour étudier les interactions entre une gamme importante de sucres et des protéines (lectines, enzymes, anticorps), des ARNs, mais aussi des cellules entières ou des particules virales. La détection de l'interaction est permise par le marquage préalable des lectines avec un composé fluorescent comme l'Alexa Fluor 488-TFP (carboxylic acid, 2,3,5,6 tetrafluorophenyl), ou l'utilisation d'anticorps spécifiques (**Figure 4.3.**).

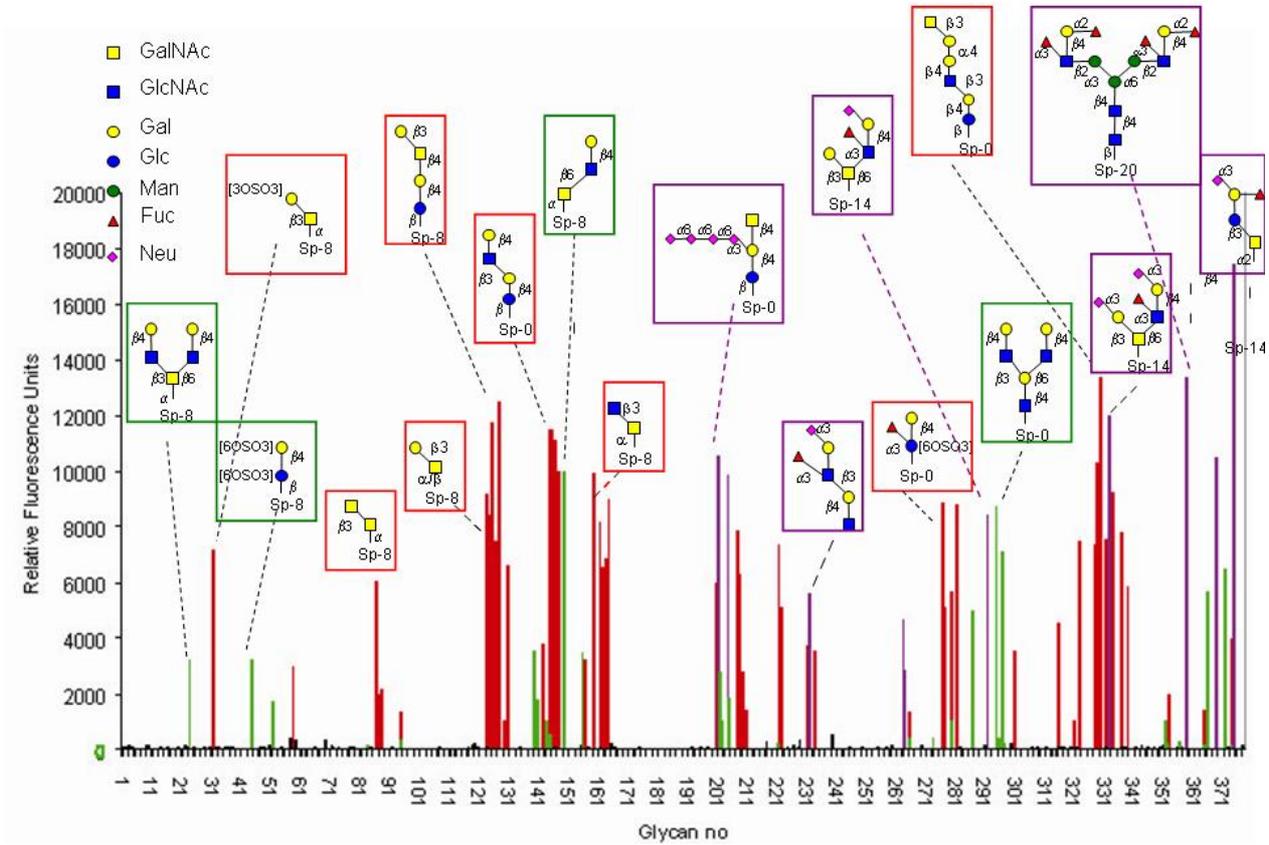


**Figure 4.3. Principe du Glycan Arrays.**

La méthode permet de tester un très grand nombre d'oligosaccharides et exige très peu de lectine (0,1-2  $\mu\text{g}$ ) pour obtenir un signal optimal.

De manière plutôt surprenante, les deux discoidines possèdent peu de spécificités communes. DiscI présente une préférence nette pour les motifs à GalNAc (ou Gal) liés en  $\beta$ 1-3, comme par exemple les disaccharides GlcNAc $\beta$ 1-3GalNAc, Gal $\beta$ 1-3GalNAc ou GalNAc $\beta$ 1-3Gal. Ces résultats ont été confirmés par microcalorimétrie isotherme de titration. DiscII présente un tableau de spécificité plus varié mais semble se lier préférentiellement à des GalNAc portant un LacNAc en position 6 (Gal $\beta$ 1-4GlcNAc $\beta$ 1-6GalNAc) (**Figure 4.4.**). Il

est également très intéressant de relever que les deux lectines reconnaissent des glycanes fucosylés et/ou sialylés. Ces ligands pourraient être reconnus par le domaine N-terminal. Cet aspect sera développé dans le chapitre 6.



**Figure 4.4.** Superposition des résultats de puces à sucre obtenus avec DiscI (rouge) et DiscII (vert) à la même concentration.

Les pics communs aux deux sont colorés en violet.

### 3. Structures cristallographiques de DiscI

Des cristaux de DiscI, obtenus après quelques jours de croissance, ont permis de résoudre la structure cristallographique de cette protéine, complexée avec différents ligands à haute résolution. Les structures cristallographiques de DiscI native (PDB 2W94) et complexée avec le GalNAc (PDB 2W95) ont été résolues par Dr. K. S. Aragão alors que j'ai résolu celles complexées avec les disaccharides Gal $\beta$ 1-3GalNAc (PDB 2WN2) et GalNAc $\beta$ 1-3Gal (PDB 2WN3). La technique utilisée pour résoudre le problème des phases a été le remplacement

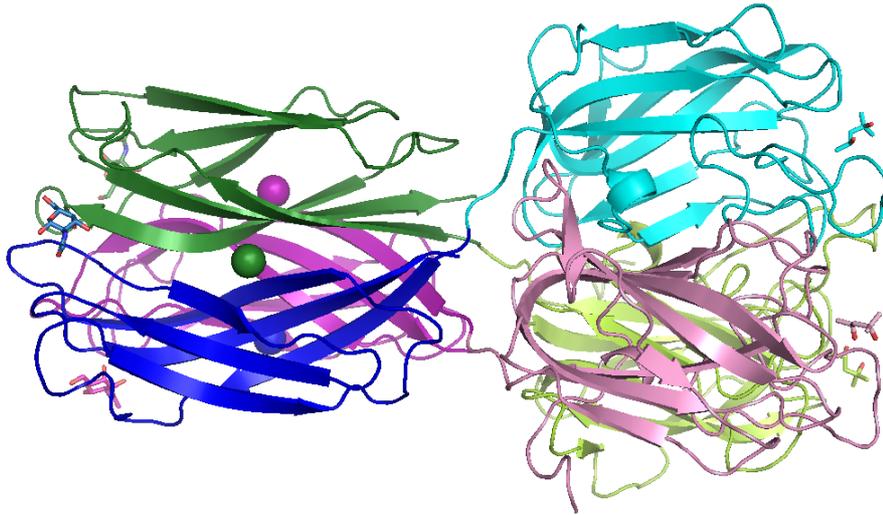
moléculaire à partir de la structure de DiscI/GalNAc. DiscI cristallise dans le groupe d'espace C222<sub>1</sub> où le trimère représente le contenu de l'unité asymétrique (**Tableau 4.1.**).

	DiscI	DiscI/GalNAc	DiscI/Galβ1-3GalNAc	DiscI/GalNAcβ1-3Gal
Conditions de cristallisation	1,5 M Li <sub>2</sub> SO <sub>4</sub> , 0,1 M HEPES pH 7,5 (Solution 16 du kit Hampton crystallization screen I)		1,2 M Li <sub>2</sub> SO <sub>4</sub> , 0,1 M HEPES, pH 7,5	
Insertion du ligand		Trempage dans une solution de GalNAc à 10 mM.	Co-cristallisation (une heure à température ambiante) dans une solution à 1,3 mM.	
Forme des cristaux	bipyramide			
Photos				
Solution de cryo	25% MPD		20% glycérol	
Groupe d'espace	C222 <sub>1</sub>			
Dimensions de la cellule a, b, c (Å)	72,5 ; 274,3 ; 85,9	72,4 ; 274,0 ; 86,1	72,6 ; 274,0 ; 85,8	72,5 ; 273,7 ; 85,8
Résolution	1,8 Å	1,75 Å	1,8 Å	1,6 Å
Code PDB	2W94	2W95	2WN2	2WN3

**Tableau 4.1. Récapitulatif des conditions expérimentales et de données pour la résolution des structures cristallographiques de DiscI, native et en complexe avec le GalNAc et les disaccharides Galβ1-3GalNAc et GalNAcβ1-3Gal.**

De manière similaire à DiscII, l'oligomérisation de DiscI résulte de l'association de trois monomères, chacun composé de deux domaines (**Figure 4.5.**) :

- un domaine N-terminal présentant une topologie classique de domaine discoidine (ou domaine DS).
- un domaine C-terminal typique des lectines de type H, adoptant un repliement de type sandwich à deux feuillets de trois brins  $\beta$  antiparallèles qui s'agencent en *clé grecque*.

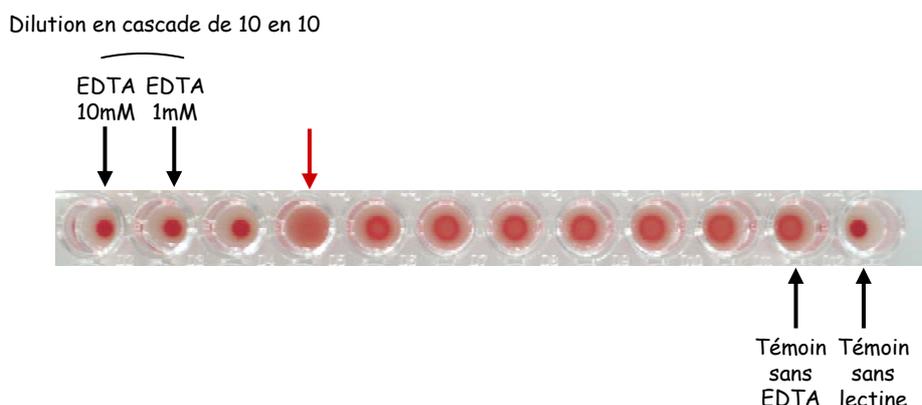


**Figure 4.5. Structure cristallographique du trimère de DiscI.**

Chaque chaîne est colorée par une couleur avec le domaine N-terminal en couleur claire et le C-terminal en foncé. Les molécules de MPD, du côté N-terminal et GalNAc, du côté C-terminal, sont représentées en bâtonnets. Les ions calcium sont représentés par des boules colorées de la même couleur que la chaîne à laquelle ils appartiennent.

La carte de densité électronique révèle également la présence d'ions identifiés comme étant du calcium, situés au milieu du domaine C-terminal à l'interface entre deux monomères. Ils semblent participer à la stabilisation du trimère. Ceci a été vérifié par hémagglutination. La surface des hématies de lapin est tapissée par une couche de sucre. Si la protéine possède une activité hémagglutinante, c'est à dire si c'est une lectine multivalente, un réseau s'établira entre les différents érythrocytes, inhibant leur sédimentation. Une phase gélatineuse est alors visible à l'œil nu (**Figure 4.6. Témoin sans EDTA**). En présence de fortes concentrations d'agents chélatants, comme l'EDTA, DiscI perd ses capacités agglutinantes ; les hématies sédimentent, par gravité, au fond du puits. En revanche, pour des concentrations en EDTA inférieures à 10  $\mu\text{M}$ , l'hémagglutination est de nouveau visible (**Figure 4.6.**). La concentration en EDTA est probablement insuffisante pour provoquer une chélation totale des

calciums et *Discl* reste fonctionnelle. A l'inverse, aucun cation divalent n'est observée dans la structure cristallographique de *DiscII* (Aragao *et al.* 2008). Ces résultats confirment une étude précédente qui révélait déjà la nécessité de cations divalents, notamment le calcium, pour l'activité de *Discl* alors que l'agglutination de *DiscII* n'était pas modulée par la présence d'ions (Alexander *et al.* 1983).



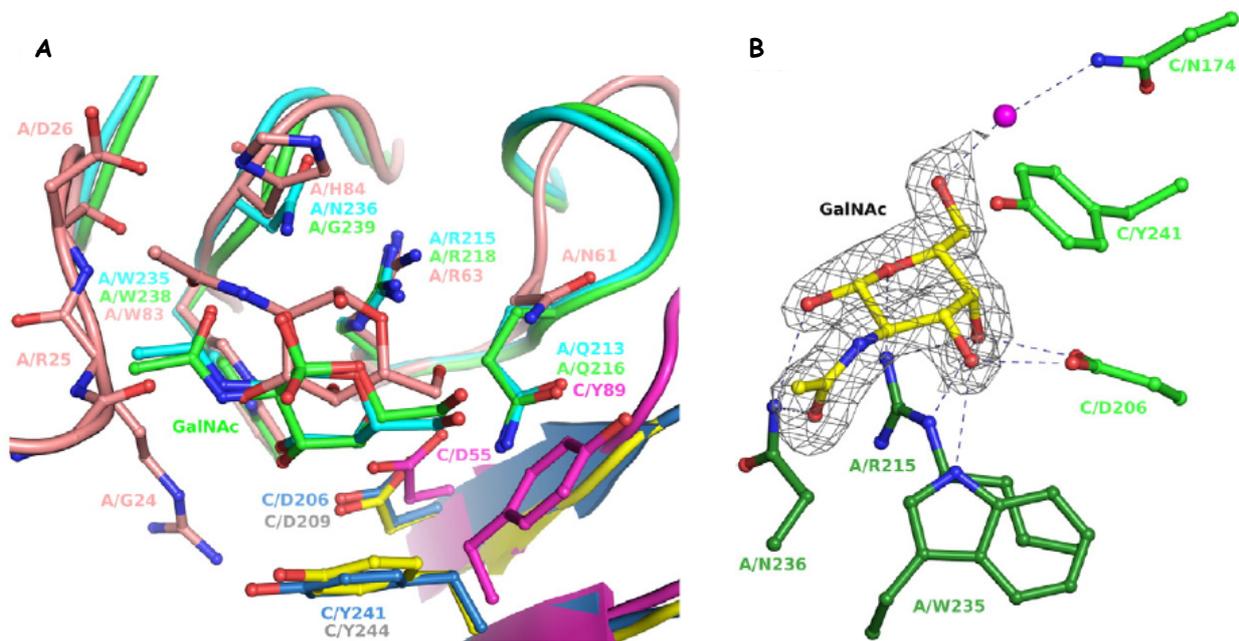
**Figure 4.6. Influence de l'EDTA sur *Discl* contrôlée par hémagglutination.**

La concentration en protéines dans chaque puits est égale à  $0,5 \text{ mg.ml}^{-1}$ , concentration pour laquelle une agglutination des hématies de lapin est observée.

Dans le complexe *Discl*/GalNAc, le monosaccharide, stabilisé par un réseau de liaisons hydrogène, est localisé dans la même position que celle observée précédemment dans les autres lectines de type H (**Figure 4.7. A.**) (Sanchez *et al.* 2006 ; Lescar *et al.* 2007 ; Aragao *et al.* 2008). Le groupement hydroxyle O3 du GalNAc interagit avec l'azote NE1 du Trp 235 et l'oxygène OD2 de l'Asp 206 du monomère adjacent. L'hydroxyle O4 interagit avec l'oxygène OD1 de l'Asp 206 du monomère adjacent et les azotes NE et NH2 de l'Arg 215. L'oxygène du cycle interagit avec l'azote NH2 de l'Arg 215 alors que les hydroxyles O7 et O1 interagissent avec l'azote ND2 de l'Asn 236. Une interaction est également observée entre l'hydroxyle O6 du GalNAc et l'azote ND2 de l'Asn 174. La chaîne latérale de la Tyr 241 du monomère adjacent est impliquée dans une interaction hydrophobe avec les carbones C3, C4, C5 et C6 du cycle du sucre (**Figure 4.7. B.**).

La plupart des interactions observées impliquent les groupements hydroxyles O3 et O4 du sucre, positionnés respectivement en positions équatoriale et axiale. Ces orientations expliquent la spécificité de *Discl* et des autres lectines de type H pour les monosaccharides avec une configuration *D-galacto*. Une autre stéréochimie à ces positions pourrait conduire à un conflit stérique, notamment avec la Tyr 241 et une perte des liaisons hydrogène cruciales

avec la protéine. Les acides aminés impliqués dans ces liaisons sont strictement conservés parmi les lectines de type H (Asp 206, Arg 215 et Trp 235 dans DiscI).



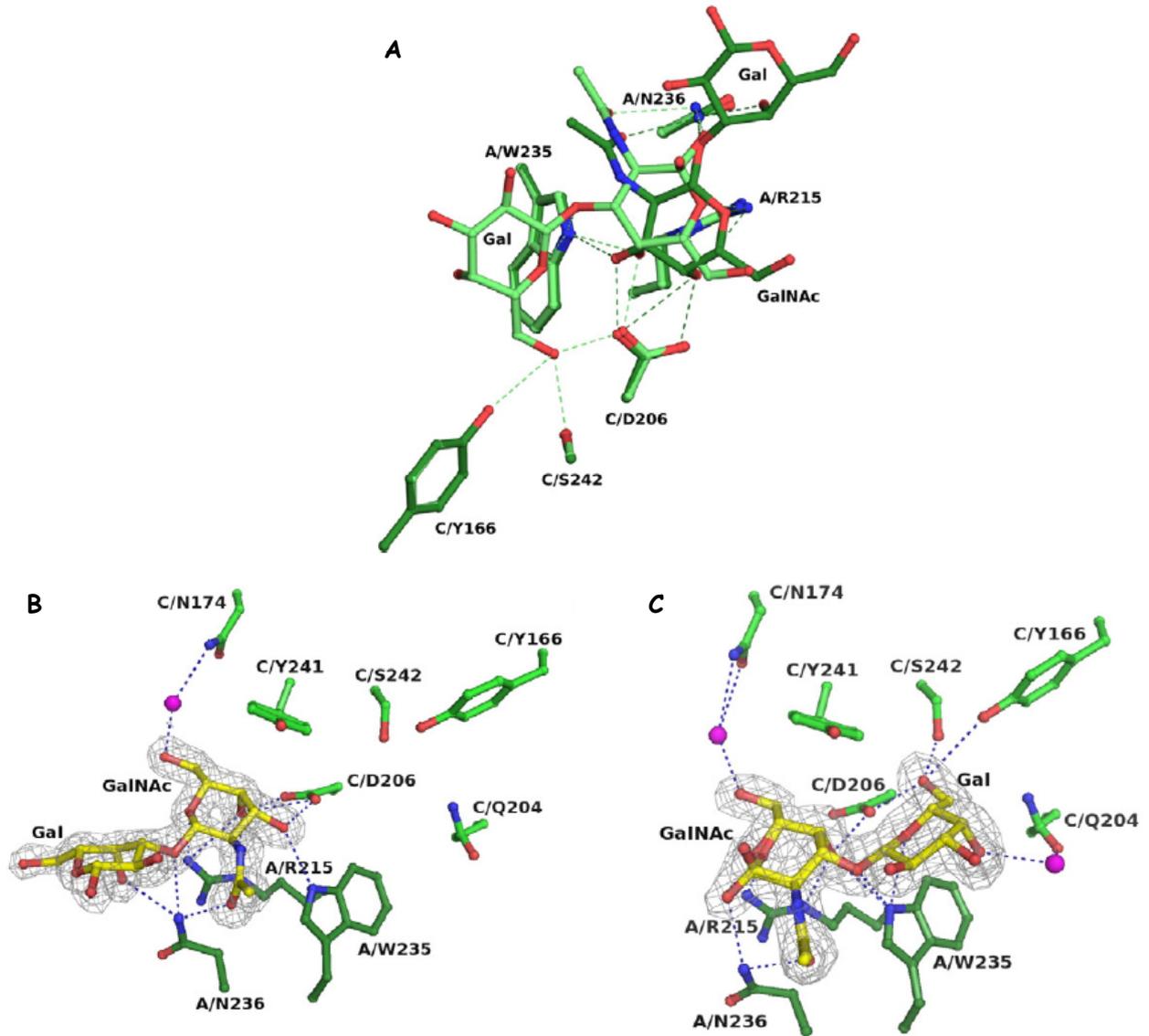
**Figure 4.7.** A. Superposition du site de reconnaissance du GalNAc dans DiscI (cyan/marine), DiscII (vert/jaune) et HPA (saumon/magenta), B. Représentation du site de reconnaissance de DiscI pour le GalNAc.

A. L'annotation des résidus est effectuée selon leur numérotation respective. B. Les liaisons hydrogènes sont représentées en lignes pointillées. La sphère magenta représente une molécule d'eau.

La préférence anomérique de chaque lectine de type H, en complexe avec le GalNAc, est influencée par le résidu qui suit le Trp conservé et localisé dans la boucle de surface. Dans HPA, la chaîne latérale de l'His 84 semble gêner la position de l'anomère  $\beta$ , favorisant ainsi sa spécificité pour l' $\alpha$ GalNAc (Sanchez *et al.* 2006) alors que, dans DiscII, la Gly 239 permet la liaison des deux anomères (Aragao *et al.* 2008). La structure de DiscI/Gal $\beta$ 1-3GalNAc montre que les deux anomères peuvent se lier mais l'Asn 236 favorise néanmoins la liaison de l'anomère  $\beta$ . Ceci a également été vérifié par ITC où DiscI possède une meilleure affinité pour le  $\beta$ MeGal ( $K_d = 2,96$  mM) que pour son anomère  $\alpha$  ( $K_d = 8,39$  mM). Le résidu Asn 236 interagit également avec le groupement *N*-acétyle du GalNAc, ce qui explique la meilleure affinité de DiscI pour le GalNAc ( $K_d = 0,30$  mM) que pour le Gal ( $K_d = 3,77$  mM).

Dans les structures cristallographiques complexées avec les disaccharides GalNAc $\beta$ 1-3Gal et Gal $\beta$ 1-3GalNAc, résolues à 1,6 Å et 1,8 Å, le GalNAc est situé dans une position similaire à celle décrite précédemment. Dans la structure Gal $\beta$ 1-3GalNAc, le GalNAc est cependant positionné différemment. Il subit une rotation de -30°C autour de l'axe qui passe par son carbone anomérique (**Figure 4.8. A.**). Les liaisons hydrogène sont décrites dans le **Tableau 4.2.**

Dans DiscI/GalNAc $\beta$ 1-3Gal, le Gal, quant à lui, est orienté vers l'extérieur de la protéine et forme uniquement deux liaisons hydrogène entre ses groupements hydroxyles O3 et O4 et l'azote ND2 de l'Asn 236 (**Figure 4.8. B.**). Dans la structure DiscI/Gal $\beta$ 1-3GalNAc, aucun changement dans la conformation des acides aminés n'est observé. Les principales modifications concernent les groupements hydroxyles O3 et O4 du GalNAc. La liaison hydrogène entre l'hydroxyle O3 et le Trp 235 est faible et celle avec l'Asp 206 est perdue. Au niveau de l'hydroxyle O4, la perte de la liaison hydrogène avec l'azote NH2 de l'Arg 215 est compensée par une nouvelle formée avec le NE1 du Trp 235. Il interagit maintenant avec l'oxygène OD2 de l'Asp 206 du monomère adjacent plutôt qu'avec l'OD1. L'hydroxyle O6 peut présenter une seconde conformation qui interagit avec l'oxygène OD1 de l'Asp 206 du monomère adjacent. Le mouvement du GalNAc permet une bonne orientation du Gal dans le site de liaison où un nouveau réseau de liaisons hydrogènes est créé. Le groupement hydroxyle O5 interagit alors avec l'azote NE1 du Trp 235, l'O6 avec les oxygènes OG de la Ser 242 du monomère adjacent et OD2 de l'Asp 206 du monomère adjacent et l'O4 avec l'oxygène OE1 de la Gln 204 du monomère voisin via une molécule d'eau. Aucune interaction n'est observée avec les hydroxyles O2 et O3 (**Figure 4.8. C.**). Les interactions supplémentaires avec le Gal expliquent la meilleure affinité de DiscI pour les disaccharides que pour le monosaccharide. La meilleure reconnaissance de DiscI pour le Gal $\beta$ 1-3GalNAc ( $K_d = 27,2 \mu\text{M}$ ) que pour le GalNAc $\beta$ 1-3Gal ( $K_d = 49,4 \mu\text{M}$ ) est lié à l'environnement plus restreint de ce résidu Gal qui forme alors des liaisons hydrogènes plus fortes et plus nombreuses avec la protéine (**Figure 4.8. A.**).



**Figure 4.8. Liaisons des disaccharides dans le domaine C-terminal de DisI.**

**A.** Superposition des interactions des disaccharides avec DisI. Les disaccharides Gal $\beta$ 1-3GalNAc et GalNAc $\beta$ 1-3Gal sont respectivement représentés en vert clair et vert foncé, **B.** Représentation de la densité électronique autour du GalNAc $\beta$ 1-3Gal, **C.** Représentation électronique autour du Gal $\beta$ 1-3GalNAc.

Complexe		Distances (Å)	
Atome du ligand	Atome de la protéine	GalNAcβ1-3Gal	Galβ1-3GalNAc
<b>Site à GalNAc</b>			
O1	Asn 236 ND2		3,03 (0,01)
O7	Asn 236 ND2 H <sub>2</sub> O-Asn 236 N	2,90 (0,03)	3,04 (0,02) 2,73 (0,10)-2,89 (0,07)
O3	Trp 235 NE1 Asp 206 <sup>a</sup> OD2	2,86 (0,05) 2,54 (0,03)	3,06 (0,06)
O4	Asp 206 <sup>a</sup> OD1 Arg 215 NE Arg 215 NH2 Trp 235 NE1 Asp 206 <sup>a</sup> OD2	2,58 (0,04) 2,96 (0,10) 3,17 (0,01)	2,86 (0,04) 3,09 (0,04) 2,60 (0,06)
O5	Arg 215 NH2	2,98 (0,08)	2,91 (0,02)
O6	H <sub>2</sub> O-Asn174 <sup>a</sup> ND2 Asp 206 <sup>a</sup> OD1	2,54 (0,06)-2,95 (0,06)	2,75 (0,13)-3,15 (0,10) 2,50 <sup>b</sup>
<b>Site à Gal</b>			
O3	Asn 236 ND2	3,00 (0,06)	
O4	Asn 236 ND2 H <sub>2</sub> O-Gln204 <sup>a</sup> OE1	2,97 (0,07)	2,90 (0,27)-2,77(0,13)
O5	Trp235 NE1		3,24 (0,04)
O6	Ser 242 <sup>a</sup> OG Asp 206 <sup>a</sup> OD2		2,70 (0,02) 2,72 (0,03)

**Tableau 4.2. Récapitulatif des contacts établis entre Discl et ses ligands disaccharidiques dans les structures cristallographiques.**

Les déviations standards, calculées pour les trois sites de reconnaissances du trimère, sont inscrites entre parenthèses.

<sup>a</sup> Résidu appartenant au monomère adjacent.

<sup>b</sup> Liaison hydrogène pas présente dans chaque monomère.

## 4. Conclusion

L'obtention de DiscI et DiscII sous forme recombinante a grandement facilité leurs études structurales et fonctionnelles. Ces deux lectines homologues de *Dictyostelium discoideum*, sont composées de deux domaines distincts :

- un domaine N-terminal qui sera plus largement développé dans le chapitre 6
- un domaine C-terminal, spécifique des lectines de type H et responsable de la spécificité pour les résidus galactosylés.

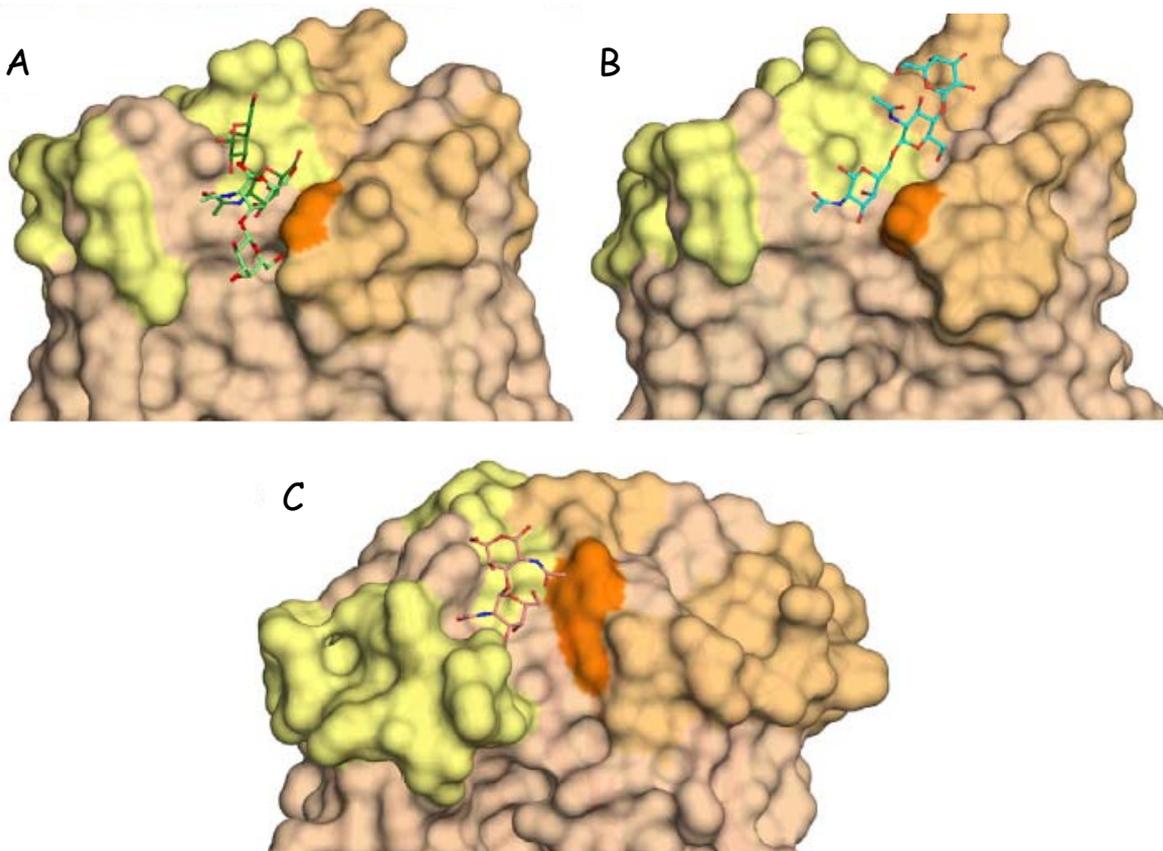
Elles forment des trimères suite à l'enlacement de ces deux domaines l'un autour de l'autre. Pour toutes les lectines de type H, les sites de reconnaissance sont localisés à l'interface entre deux monomères, au niveau d'une face de liaison au sucre. La formation d'un trimère est alors indispensable à la reconnaissance de leur ligand. La partie N-terminale d'un monomère crée avec la partie C-terminale du monomère adjacent un sillon superficiel où viennent interagir les ligands. Dans DiscI, la présence d'ions calcium situés au milieu du domaine C-terminal à l'interface entre deux monomères renforce cette organisation tridimensionnelle.

La résolution des structures cristallographiques de DiscI en complexe avec le monosaccharide GalNAc et les disaccharides Gal $\beta$ 1-3GalNAc et GalNAc $\beta$ 1-3Gal a permis de comprendre le mode d'interaction de cette lectine avec ces ligands et sa préférence pour la configuration D-galacto et l'anomérie  $\beta$ . Les variations d'affinités observées en ITC pour ces ligands sont également facilement expliquées en étudiant les structures 3-D.

Le rôle précis de ces discoidines chez l'animal ainsi que leur ligand naturel ne sont toujours pas clairement définis. Plusieurs hypothèses peuvent néanmoins être émises. Au stade végétatif, lorsque les conditions sont favorables, DiscI est retrouvée dans la membrane extracellulaire de l'amibe où elle reconnaît un glycoconjugué non caractérisé (Cooper and Barondes 1984). En cas de carence alimentaire, les amibes se rassemblent et forment des spores facilitant leur survie et leur dissémination. DiscI pourrait participer à cette agglutination (Barondes *et al.* 1983). La paroi de ces vésicules serait composée de cellulose cristallisée agrémentée de polysaccharides riches en Gal et GalNAc. La présence de substrats naturels pour DiscII, produite majoritairement durant cette phase du cycle de vie de l'amibe laisse supposer une implication de cette lectine dans l'intégrité de ces spores (Cooper *et al.* 1983 ; Erdos 1989 ; West *et al.* 2009). Ces discoidines pourraient alors être impliquées dans la

formation de différents types de parois chez l'animal. Elles pourraient également jouer un rôle dans la reconnaissance de polysaccharides bactériens et participeraient alors aux mécanismes de défense de l'animal en agglutinant les pathogènes. Ce rôle de protection a souvent été identifié pour les lectines d'invertébrés (Zhang *et al.* 2009 ; Rattanaporn and Utarabhand 2010). Mais aussi, elles pourraient être impliquées dans la phagocytose, processus utilisé par les souches sauvages de *Dictyostelium discoideum* pour se nourrir de bactéries (Gotthardt *et al.* 2006 ; Sillo *et al.* 2008). Les discoidines pourraient alors assurer différentes fonctions et reconnaître différents ligands suivant leur localisation, dans le cytoplasme, à l'intérieur de vésicules ou à la surface cellulaire.

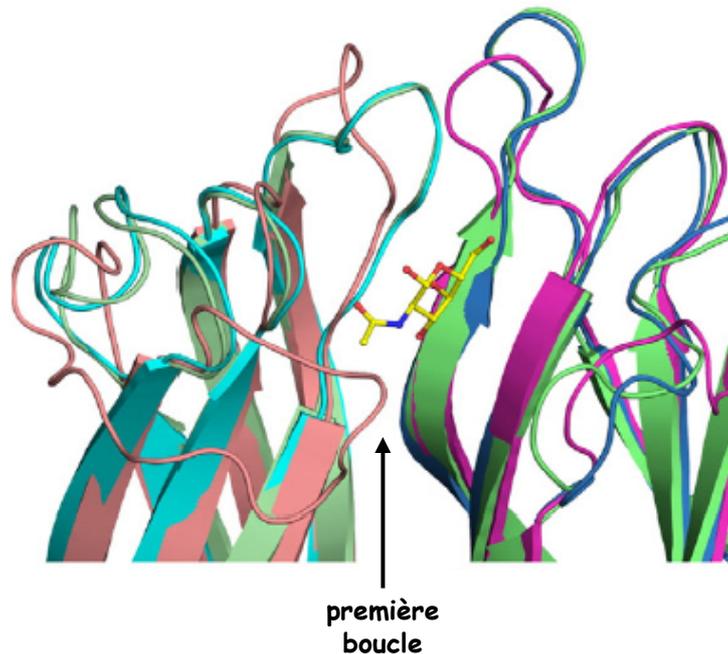
Malgré de fortes ressemblances (organisation trimérique, sites de reconnaissance à l'interface entre deux monomères, préférence pour les ligands en configuration *D-galacto*), les lectines de type H présentent également d'importantes différences, comme en particulier la topologie du site de reconnaissance (**Figure 4.9.**).



**Figure 4.9. Représentation de surface des sites de reconnaissance de DiscI (A.), DiscII (B.) et HPA (C.).**

Les ligands de DiscI, DiscII et HPA sont respectivement les deux disaccharides cristallisés (Gal $\beta$ 1-3GalNAc (vert clair) et GalNAc $\beta$ 1-3Gal (vert foncé)), un trisaccharide modélisé (Gal $\beta$ 1-4GlcNAc $\beta$ 1-6GalNAc) et l'antigène de Forssman (PDB 2CGY).

Les discoidines possèdent un site de reconnaissance en forme de L inversé. Celui de DiscI (**A.**), plus profond, étroit et étendu que celui de DiscII (**B.**), lui permet de reconnaître des oligosaccharides lactosaminiques contrairement à DiscII. En revanche, le dessus du site, plus ouvert dans DiscII est déterminant pour la reconnaissance des oligosaccharides branchés en 1-6 et particulièrement la présence de l'His 211. Dans HPA, l'orientation de la chaîne latérale de la Tyr 89 empêche la liaison des oligosaccharides branchés en 1-6. De plus, la présence de huit acides aminés supplémentaires au niveau de la première boucle (**Figure 4.10.**), conduit à la formation d'un sillon plus étroit, profond et étendu que dans les discoidines (**C.**). Ceci explique la spécificité restreinte de cette lectine pour les GalNAc branchés en 1-3, comme observé par glycan arrays (Sanchez *et al.* 2006). Cette partie de la boucle est également impliquée dans la spécificité d'HPA pour le monosaccharide puisqu'elle entoure et stabilise les groupements *N*-acétyl et O3 du monosaccharide via l'établissement de plusieurs liaisons hydrogène qui favorisent la reconnaissance de ce résidu par rapport au Gal. Dans les discoidines, cette boucle est plus courte, de séquence et de structure différente en comparaison avec HPA, et n'interagit pas avec le monosaccharide.



**Figure 4.10. Représentation des structures cristallographiques des sites de reconnaissance des lectines de type H.**

DiscI, DiscII et HPA sont respectivement représentées en vert clair/vert, en cyan/bleu et en saumon/magenta. Chaque site de reconnaissance est formé par deux monomères chacun coloré d'une couleur. Le résidu GalNAc, représenté en bâtonnets, est localisé dans le site de reconnaissance de DiscII.

La longueur de cette première boucle ainsi que la forme du sillon semblent être primordiales dans le caractère prédictif de cancers à faible pronostic vital. En effet, seule HPA possède la capacité de reconnaître, de manière spécifique, les cellules métastatiques de cancers du colon et du sein notamment (Brooks and Leathem 1991 ; Ikeda *et al.* 1994). Les discoidines, au contraire, se fixent à la surface de lignées cancéreuses, qu'elles soient métastatiques ou pas. Ces résultats ont été précédemment développés dans le chapitre 3. Cette première boucle est donc visiblement fondamentale pour l'utilisation et le développement de ces lectines de type H comme outil de diagnostic en cancérologie.

# **Chapitre 5**

**Ingénierie de DiscII**



---

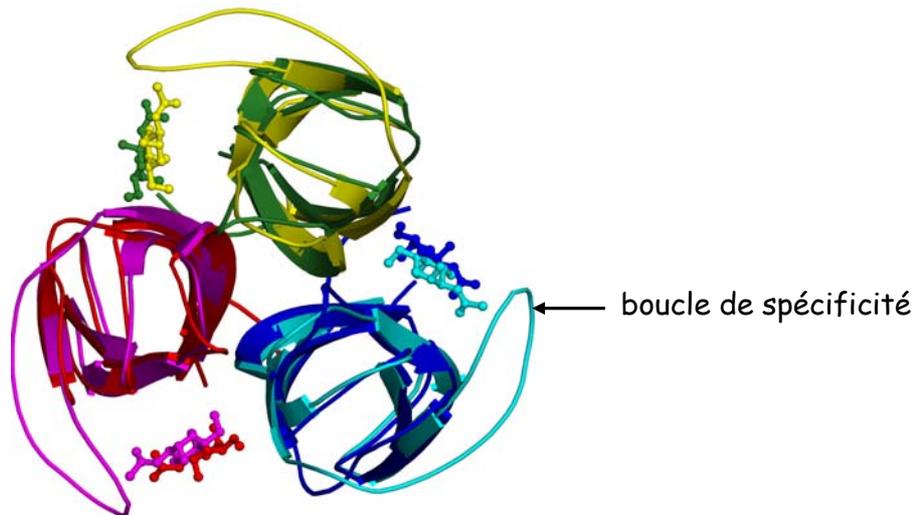
# Chapitre 5

## Ingénierie de DiscII

---

La lectine HPA, de l'escargot *Helix pomatia*, est un marqueur capable de cibler spécifiquement les cellules cancéreuses à caractère métastatique. Cependant, la variabilité des lots commerciaux obtenus par purification à partir d'escargots d'origine non connue et l'impossibilité à la produire sous forme recombinante limite l'intérêt de l'utilisation de cette lectine comme outil d'investigation de routine en oncologie.

HPA possède de fortes similarités de séquence et de structure avec le domaine C-terminal des discoidines (DiscI et II) de l'amibe *Dictyostelium discoideum*. Ces deux lectines ont été exprimées sous forme recombinante au laboratoire et DiscII a rapidement été caractérisée biochimiquement. La structure cristallographique de cette dernière a été résolue durant le travail de doctorat de Karoline S. Aragão (Aragao *et al.* 2008). Les principales différences entre HPA et DiscII résident dans la conformation, la séquence et le nombre d'acides aminés que composent la première boucle de surface appelée boucle de spécificité de ces deux lectines de type H (**Figure 5.1.**). Or, comme nous l'avons constaté dans le chapitre 3, seule HPA est capable de discriminer les glycoépitopes présents à la surface de cellules cancéreuses métastatiques. La longueur de cette boucle semble donc être d'une grande importance dans le caractère prédictif de cette lectine. DiscII étant exprimée sous forme recombinante en quantité importante dans *E. coli*, la stratégie envisagée est de modifier, par mutagenèse, le gène codant pour DiscII au niveau de la boucle de spécificité de manière à lui conférer celle d'HPA. La protéine chimérique ainsi construite aurait les mêmes propriétés prédictives qu'HPA et serait facilement exprimée sous forme recombinante dans *E. coli*. Dans ce chapitre, je détaillerai les différentes protéines chimériques construites et les stratégies mises au point pour l'obtention de mutants fonctionnels.



**Figure 5.1.** Superposition des structures cristallographiques d’HPA (cyan, magenta et jaune) et de DiscII (bleu, rouge et vert), complexées avec du GalNAc.

## 1. Mutants 1

### 1.1. Construction des protéines par mutagenèse dirigée

La construction de la première série de mutants de DiscII a été basée sur les homologies structurales observées entre le domaine C-terminal de DiscII et HPA, selon l’alignement de séquences de la **Figure 5.2**. La boucle d’HPA comporte huit acides aminés supplémentaires (entourés en noir) par rapport à celle de DiscII. Cette séquence a donc été insérée par mutagenèse dirigée (selon le protocole GeneTailor™ Site-Directed Mutagenesis System d’Invitrogen) dans le gène codant pour la protéine DiscII. Le mutant DiscII-boucleHPA est alors construit. Les domaines C-terminaux muté et natif ont également été clonés et sont respectivement dénommés DiscII-Cter-boucleHPA et DiscII-Cter-N257. Chaque gène ainsi obtenu a été cloné dans le vecteur pProEx-HtB, utilisé pour l’expression de DiscII sous forme recombinante dans *E. coli*.

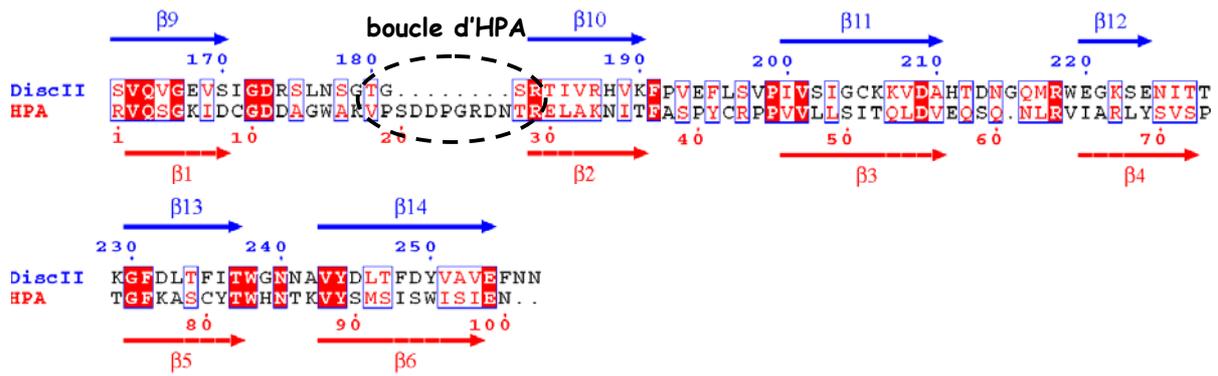


Figure 5.2. Alignement des séquences peptidiques du domaine C-terminal de DiscII et HPA, basé sur les homologies structurales.

## 1.2. Production sous forme recombinante dans *E. coli*

Différentes conditions d'expression dans la bactérie *E. coli* ont été testées afin d'optimiser l'obtention de ces protéines sous forme soluble et fonctionnelle et en quantité importante. Les différents essais sont répertoriés dans le **tableau 5.1**.

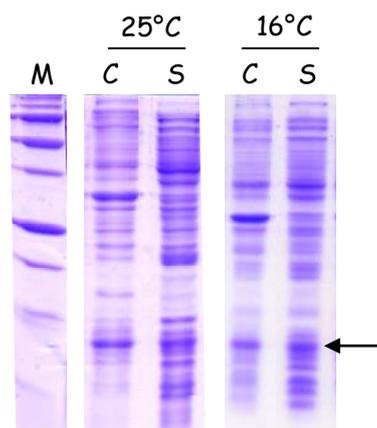
Souche	DO <sub>600</sub> d'induction	IPTG (mM)	Température (°C)	Temps d'induction
<b>DiscII-boucleHPA</b>				
Rosetta 2 pLysS	1,2	0,2	16°C	1 nuit
Arctic Express RIL	0,7	0,2 et 1	16°C	1 nuit
BL21 (DE3) C41 + pRARE 2	0,7	0,2	16°C	1 nuit
BL21 (DE3) C43 + pRARE 2	0,7	0,2	16°C	1 nuit
<b>DiscII-Cter-boucleHPA</b>				
Rosetta 2 pLysS	0,7 et 1,2	0,2	16°C	1 nuit
	0,7	0,2 et 1	25°C	4 heures
Arctic Express RIL	0,7	0,2 et 1	16°C	1 nuit
<b>DiscII-Cter-N257</b>				
Rosetta 2 pLysS	0,7 et 1,2	0,2	16°C	1 nuit
	0,7	0,2 et 1	25°C	4 heures
Arctic Express RIL	0,7	0,2 et 1	16°C	1 nuit

Tableau 5.1. Récapitulatif des conditions d'expression dans *E. coli* des mutants DiscII-boucleHPA, DiscII-Cter-boucleHPA et DiscII-Cter-N257.

L'efficacité des expressions est évaluée par analyse sur gel SDS-PAGE en contrôlant la présence de la protéine d'intérêt avant et après induction et sa proportion sous forme soluble. Les protéines mutantes étant produites en quantité relativement faible, il est parfois difficile de conclure uniquement d'après les analyses SDS-PAGE. Une purification par chromatographie sur colonne de nickel permet alors d'évaluer leur présence et d'estimer leur quantité relative.

Plusieurs conclusions peuvent être tirées de ces différents essais. Tout d'abord, la variation de la température pour l'expression de DiscII-Cter-boucleHPA et DiscII-Cter-N257 ne semble pas modifier la proportion de protéines sous forme soluble. Après une expression à 25°C pendant quatre heures ou une nuit à 16°C, environ 50% de protéine est soluble (**Figure 5.3**). L'expression se déroule néanmoins sur la nuit car, dans ces conditions, l'étape de purification peut débuter directement à la fin de l'expression. A 25°C, la production se termine en fin de journée, ce qui oblige à congeler les culots bactériens et à les décongeler le lendemain pour la purification. La concentration en IPTG de 0,2 mM est suffisante pour induire l'expression de ces protéines.

Considérant uniquement les tentatives réalisées à 16°C pendant une nuit, ni la densité optique au moment de l'induction, ni la concentration en IPTG, ni même la souche utilisée ne semble avoir d'influence sur l'expression des protéines. La souche sélectionnée pour l'expression de ces protéines sous forme recombinante est Rosetta 2 pLysS, spécialisée dans l'expression de protéines présentant des codons dits rares pour la bactérie *E. coli*.



**Figure 5.3. Gels SDS-PAGE coloré au bleu de Coomassie après expression de DiscII-Cter-boucleHPA dans Rosetta 2 pLysS.**

Les pistes C et S correspondent respectivement au culot après centrifugation à 12000 g (corps d'inclusion) et au surnageant après centrifugation à 50000 g (protéines solubles). L'induction est déclenchée par ajout de 0,2 mM d'IPTG et dure 4 heures à 25°C ou une nuit à 16°C. La flèche noire indique la présence de DiscII-Cter-boucleHPA.

### 1.3. Purification par chromatographie sur colonne de nickel

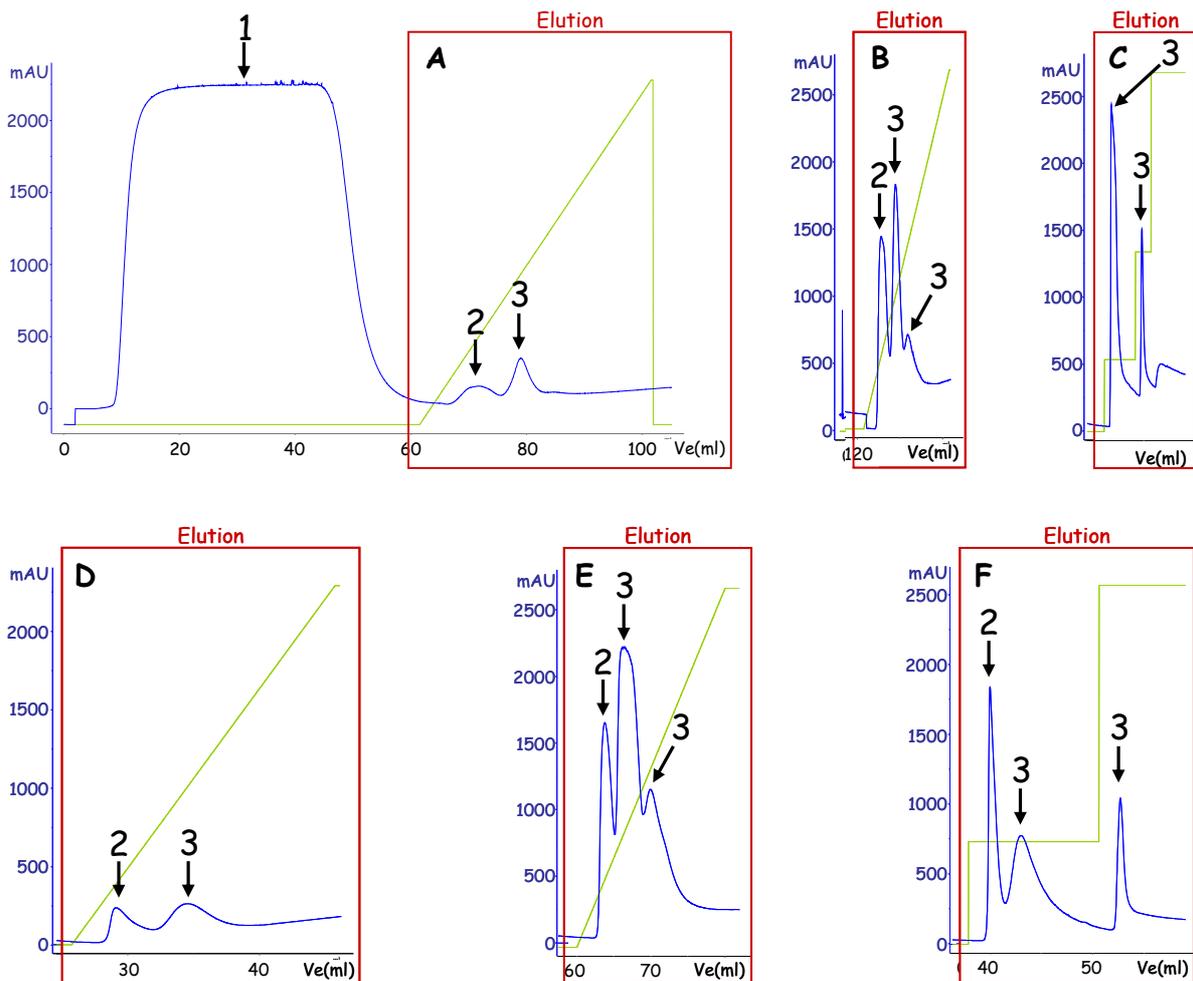
Grâce à la présence d'une étiquette polyhistidine du côté N-terminal, les protéines exprimées sous forme soluble sont ensuite purifiées par chromatographie sur résine de nickel, selon le protocole décrit dans les chapitres précédents. L'élution est réalisée soit par gradient linéaire d'imidazole allant de 10 à 500 mM, soit par palier à deux concentrations en imidazole, 100 mM pour décrocher les protéines contaminantes et 350 mM pour éluer la protéine d'intérêt. La visualisation d'un pic d'élution, quelque soit la méthode utilisée démontre que la protéine se fixe correctement aux ions nickel via sa queue polyhistidine. Cependant, quelque soit la protéine mutante considérée et les conditions d'expression utilisées, cette étape d'élution ne s'est pas révélée très reproductible. Seul un exemple avec la protéine DiscII-Cter-boucleHPA sera décrit dans le manuscrit mais ces résultats ont également été obtenus avec DiscII-boucleHPA et DiscII-Cter-N257.

Les profils d'élution diffèrent d'une manipulation à une autre (**Figure 5.4.**). L'élution des purifications A, B, D et E est réalisée selon un gradient linéaire d'imidazole. Or pour les cas A et D, deux pics sont visibles. Le premier (pic 2) est constitué des protéines contaminantes, possédant une faible affinité pour le nickel et le second (pic 3) est formé par la protéine DiscII-Cter-boucleHPA pure. Dans les cas B et E, trois pics sont visibles dont deux correspondent à la protéine d'intérêt pure. Dans le cas F où l'élution est effectuée par deux paliers à 100 mM et 350 mM d'imidazole, trois pics sont également visibles dont deux correspondent à la protéine d'intérêt pure, élue alors à deux concentrations en imidazole différentes. Deux populations de protéines, identiques après analyse sur gel SDS-PAGE sont alors obtenues. Dans le cas C, où l'élution est également effectué par palier à 100 mM et 250 mM, deux pics sont obtenus. Le premier est formé des protéines contaminantes en mélange avec DiscII-Cter-boucleHPA, présente en quantité relativement importante alors que le second est formé uniquement de la protéine d'intérêt pure. Dans ce cas, la purification n'est pas optimale. Les premiers tests ont été réalisés en utilisant 350 mM d'imidazole pour éluer la protéine alors que 250 mM sont suffisants.

Quelques hypothèses peuvent être mises en avant pour expliquer ces résultats. Durant ces manipulations, certaines constantes n'ont pas forcément été conservées pour tous les essais. Par exemple, le rapport entre le volume de milieu de culture et celui de l'erenmeyer utilisé pour la culture des bactéries, le volume de tampon d'équilibration utilisé pour suspendre les culots bactériens avant éclatement des cellules au cell disrupter (One shot cell

Disrupter, Constant System), la pression de cette machine, qui n'ont pas toujours été conservées durant les tests. Certaines d'entre elles peuvent alors modifier les résultats des purifications. De manière générale, lorsque deux populations de protéines d'intérêt pures étaient obtenues, elles ont été traitées séparément pour la suite des analyses.

La purification se poursuit ensuite par l'élimination de l'imidazole grâce à une colonne de dessalage PD-10 (GE Healthcare) et au clivage de l'étiquette His<sub>6</sub> par la protéase TEV comme décrit précédemment.

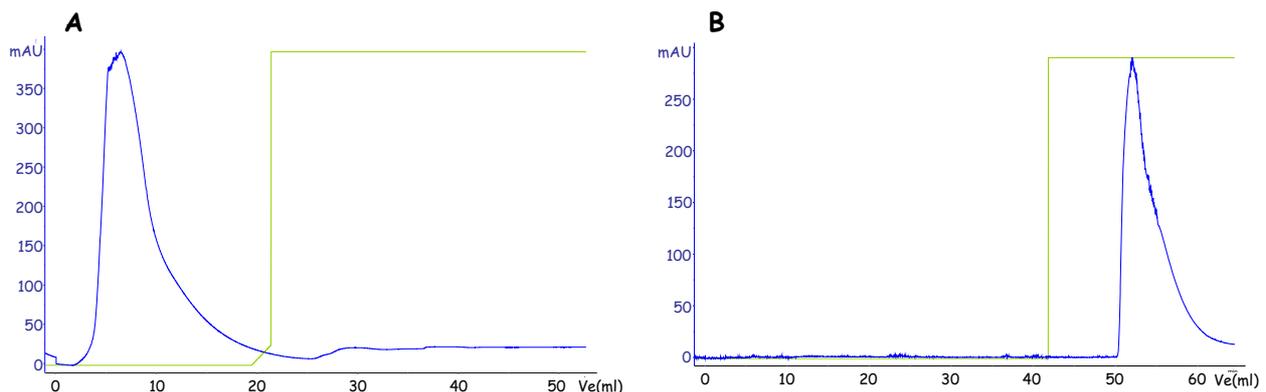


**Figure 5.4. Profils d'éluion obtenus lors des différentes purifications de DiscII-Cter-boucleHPA par chromatographie sur colonne de nickel.**

Les courbes bleue et verte représentent respectivement la densité optique à 280 nm (exprimée en mAU) et le gradient d'imidazole appliqué pour éluer les protéines. L'encadré rouge représente les fractions éluées. Le pic 1 correspond à la fraction non retenue. Le pic 2 est formé des protéines contaminantes qui possèdent une faible affinité pour le nickel ; elles sont éluées à une faible concentration d'imidazole (100 mM). Le pic 3 représente DiscII-boucle HPA pure, éluée autour de 250 mM d'imidazole.

#### 1.4. Contrôle de la fonctionnalité des mutants par chromatographie d'affinité.

Après élimination de l'étiquette polyhistidine, la fonctionnalité des protéines est évaluée par chromatographie d'affinité sur colonne de GalNAc puis analysée sur gel SDS-PAGE. Les chromatogrammes obtenus pour DiscII-boucleHPA (**Figure 5.5. A.**), DiscII-Cter-boucleHPA et DiscII-Cter-N257 sont identiques. Le pic correspondant aux protéines purifiées est identifié dans la fraction non retenue. Aucune protéine n'est éluée après l'ajout d'un tampon d'éluion contenant 0,3 M de galactose (représenté par la courbe verte sur le chromatogramme), ce qui signifie qu'aucune protéine ne s'est fixée sur le résidu GalNAc de la colonne. En revanche pour DiscII native (**Figure 5.5. B.**), que l'on sait fonctionnelle, la protéine est entièrement retenue sur la colonne et est décrochée uniquement après ajout du tampon d'éluion. Aucune des protéines mutantes n'est capable de reconnaître du GalNAc.



**Figure 5.5. Contrôle de la fonctionnalité de DiscII-boucleHPA par chromatographie d'affinité sur colonne de GalNAc (A.)**

B. Chromatogramme contrôle avec DiscII native.

Les courbes bleue et verte représentent respectivement la densité optique à 280 nm (exprimée en mAU) et l'éluion des protéines fixées sur la colonne grâce à un tampon contenant à 0,3 M de galactose.

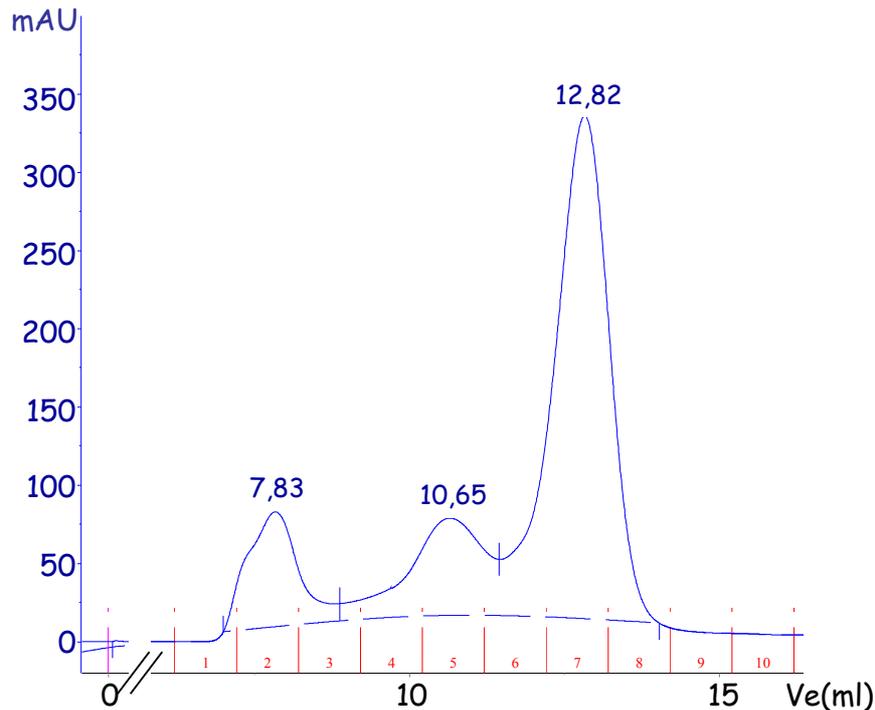
Une purification dite en batch sur résine Sepharose 4B contenant un résidu galactose terminal (Sigma) permet également de contrôler la fonctionnalité de ces lectines. Mais là encore, aucune protéine n'est retenue sur la résine. Elles ne sont donc pas capables de reconnaître le Gal.

Comme nous venons de le constater, le domaine C-terminal de la protéine native, DiscII-Cter-N257, est lui aussi incapable de reconnaître le GalNAc ou le Gal. La non-

fonctionnalité des protéines chimériques n'est donc pas uniquement due aux mutations appliquées au niveau de leur site de reconnaissance mais peut être provoquée par une mauvaise organisation des protéines. Leur oligomérisation a alors été vérifiée par chromatographie d'exclusion stérique.

### 1.5. Contrôle de l'organisation oligomérique des protéines mutantes

Aux vues des résultats précédents, nous avons émis l'hypothèse que l'organisation tridimensionnelle pouvait être à l'origine de la non-fonctionnalité des protéines et de la présence de plusieurs populations de protéines purifiées après chromatographie sur colonne de nickel. En effet, pour les protéines appartenant à la famille des lectines de type H, le site de reconnaissance du sucre est situé à l'interface entre deux monomères. L'organisation trimérique des protéines est donc indispensable pour que la reconnaissance ait lieu. L'oligomérisation de DiscII-Cter-boucleHPA et DiscII-Cter-N257 a alors été contrôlée par la technique de chromatographie d'exclusion stérique. Les tests ont tout d'abord été effectués sur une colonne Superdex 75. Pour DiscII-Cter-boucleHPA, l'élution est composée de trois pics, identifiés par analyse SDS-PAGE comme étant la protéine d'intérêt (**Figure 5.6.**). La courbe de calibration permet d'identifier que le premier correspond au volume mort, c'est-à-dire à la protéine agrégée, le second à une protéine d'environ 51,2 kDa, et le dernier, fortement majoritaire, à une protéine de 20,6 kDa. Le poids moléculaire d'un monomère est de 11,4 kDa. Les pics 2 et 3 pourraient donc correspondre respectivement à la protéine trimérique, présente en très faible quantité et au monomère, majoritairement présent. L'élution de DiscII-Cter-N257 est composée de deux pics, le premier correspondant aux protéines agrégées et le second au monomère. Aucun trimère n'est visible. L'écart important entre les poids moléculaire réels et ceux déterminés expérimentalement s'explique par le fait que DiscII est une molécule fortement allongée et que la courbe de calibration est réalisée avec des protéines standard, toutes globulaires. Dans ce type de chromatographie, la forme de la protéine influence très fortement leur temps de rétention sur colonne. Deux protéines de même poids moléculaire mais de forme différente (globulaire ou allongée) ne seront pas éluées au même volume. La protéine de forme allongée aura un temps de rétention plus faible que la protéine globulaire.



**Figure 5.6. Contrôle de l'oligomérisation de DiscII-Cter-boucleHPA par chromatographie d'exclusion stérique sur colonne Superdex 75.**

Après équilibration de la colonne avec un tampon Tris-HCl 20 mM, pH 7,5, NaCl 500 mM, Gal 10 mM, la protéine DiscII-Cter-boucleHPA est chargée dans un volume de 500  $\mu$ l avec un débit de 0,5 ml.min<sup>-1</sup>. En bleu, le densité optique à 280 nm (en mAU) en fonction du volume d'élution (en ml).

### 1.6. Stratégie pour la construction de nouveaux mutants.

En comparant, de manière plus approfondie, les structures cristallographiques d'HPA et du domaine C-terminal de DiscII, nous avons pu mettre en évidence l'importance des neuf acides aminés précédant la boucle de spécificité (**Figure 5.7. A.**). En effet, ils adoptent une orientation différente dans les deux protéines. Chez HPA, ils présentent une orientation relativement linéaire permettant à la boucle de spécificité de partir vers l'extérieur alors que chez DiscII, un angle bien plus aigu est formé, ce qui orienterait la boucle insérée vers l'intérieur. Cette organisation pourrait être perturbée la construction du trimère et ainsi conduire à la formation d'une protéine non fonctionnelle. Nous avons donc choisi, pour la construction du mutant numéro 2, DiscII-boucleHPA-mutée, de muter les neuf acides aminés de DiscII précédant l'insertion de la boucle et de les remplacer par ceux identifiés chez HPA (**Figure 5.7. B.**).



**Figure 5.7.** A. Superposition du domaine C-terminal de DiscII (vert, rouge, bleu) avec HPA (jaune, magenta, cyan) et B. Alignement de séquences entre HPA et les domaines C-terminaux de DiscII et DiscII-boucleHPA-mutée.

A. Les cercles rouges indiquent les neuf acides aminés mutés dans la deuxième construction. B. Les rectangles jaunes représentent les identités de séquences entre DiscII et DiscII-boucleHPA-mutée. Le rectangle vert indique les séquences identiques entre HPA et DiscII-boucleHPA-mutée.

## 2. Mutant 2

### 2.1. Clonage, expression et purification par chromatographie sur colonne de nickel

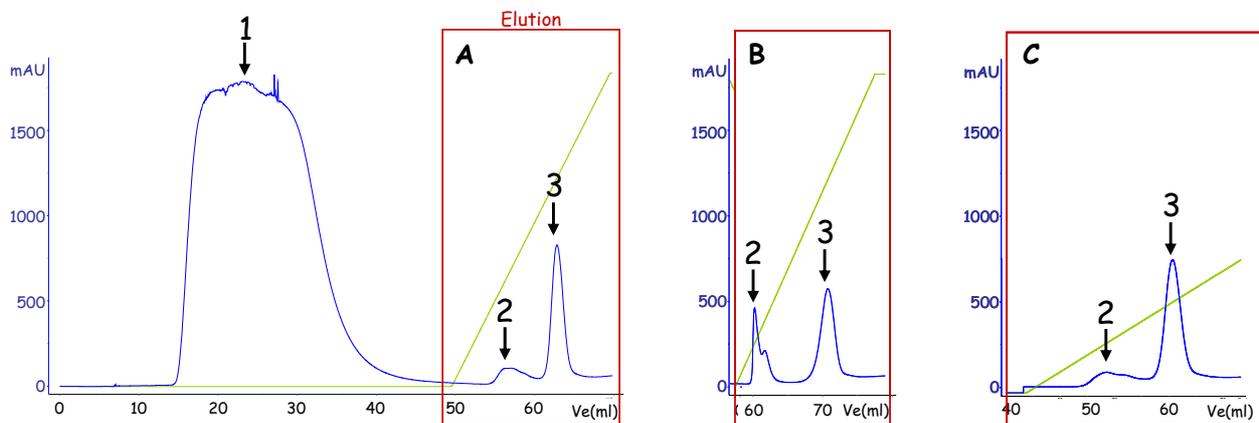
La construction des mutants DiscII-boucleHPA-mutée et DiscII-Cter-boucleHPA-mutée est réalisée selon un protocole identique à celui décrit pour DiscII-boucleHPA et DiscII-Cter-boucleHPA. Ayant émis l'hypothèse que le domaine N-terminal pouvait stabiliser le trimère et ainsi favoriser sa formation, les manipulations ont été uniquement réalisées avec la protéine contenant les deux domaines.

Les souches Rosetta 2 pLysS, Arctic Express RIL et BL21 (DE3) C41 contenant le plasmide pLysSRARE2 ont été testées pour l'expression de DiscII-boucleHPA-mutée. La densité optique à 600 nm était de 0,7 lorsque l'induction par ajout de 0,2 mM d'IPTG a été initiée. Elle s'est poursuivie toute la nuit à 16°C.

Aux vues de la faible reproductibilité des résultats précédant, le protocole de purification par chromatographie sur colonne de nickel a été standardisé. Au terme de

l'induction, les culots bactériens sont pesés puis repris dans cinq volumes (p/v) de tampon d'équilibration utilisé précédemment. Les cellules sont ensuite soumises au cell disrupter (One shot cell Disrupter, Constant System) à une pression de 1,8 kbars puis centrifugées à 50000 g pendant 30 minutes afin d'isoler les protéines solubles qui, après filtration, seront déposées sur la colonne.

La **Figure 5.8** représente les trois profils d'élution obtenus après expression de DiscII-boucleHPA-mutée dans les souches bactériennes précédemment citées. Quelque soit la souche utilisée, la quantité de protéines d'intérêt, évaluée par l'intensité du pic 3, reste relativement identique. La souche Rosetta 2 pLysS (profil C) sera néanmoins privilégiée car le rapport entre la quantité de DiscII-boucleHPA-mutée et celle des protéines contaminantes (pic 2) est le plus important.

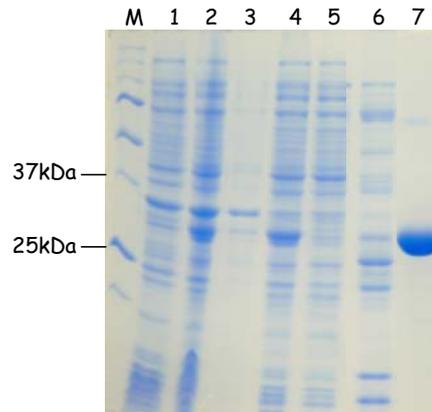


**Figure 5.8. Profils d'élution après purification sur colonne de nickel de DiscII-boucleHPA-mutée exprimée dans A. Arctic Express RIL, B. BL21 (DE3) C41 et C. Rosetta 2 pLysS**

Les courbes bleue et verte représentent respectivement la densité optique à 280 nm (exprimée en mAU) et le gradient d'imidazole appliqué pour éluer les protéines. L'encadré rouge représente les fractions éluées. Le pic 1 correspond à la fraction non retenue. Le pic 2 est formé des protéines contaminantes éluées à 100 mM d'imidazole et le pic 3 représente DiscII-boucleHPA-mutée pure, éluee autour de 350 mM d'imidazole.

L'expression et la purification des protéines peuvent être contrôlées par analyse SDS-PAGE. La **Figure 5.9** représente l'analyse effectuée après expression de DiscII-boucleHPA-mutée dans la souche Rosetta 2 pLysS (chromatogramme C de la Figure 5.6.). Les résultats obtenus avec les souches Arctic Express RIL et BL21 (DE3) C41 sont similaires. La présence d'une bande intense autour de 30 kDa après induction (piste 2) par rapport au témoin avant induction (piste 1), retrouvée majoritairement dans la fraction soluble (piste 4) révèle que

DiscII-boucleHPA-mutée est principalement exprimée sous forme soluble dans ces conditions. Les pistes 5, 6 et 7 représentent respectivement la fraction non retenue sur la colonne de nickel, les protéines contaminantes contenues dans le pic 2 et la protéine d'intérêt élue (pic 3). Cette analyse indique que DiscII-boucleHPA-mutée est abondamment retrouvée dans la fraction élue (pic 3) et que sa pureté est très satisfaisante.



**Figure 5.9. Gel SDS-PAGE 12% acrylamide coloré au bleu de Coomassie après expression de DiscII-boucleHPA-mutée dans Rosetta 2 pLysS et purification sur colonne de nickel.**

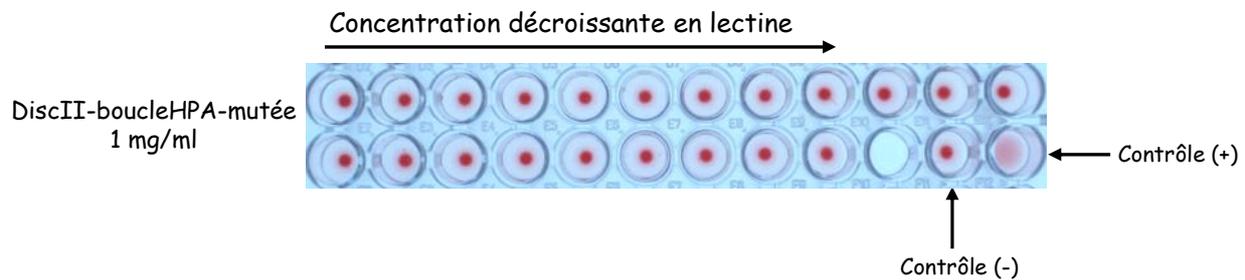
Ce gel correspond au chromatogramme C de la **Figure 5.8**. Les pistes 1 et 2 correspondent aux extraits totaux de la culture utilisée pour la purification, respectivement avant et après induction. La piste 3 représente la fraction contenant la protéine présente sous forme de corps d'inclusion (après centrifugation à 12000 g). La piste 4 représente la fraction soluble (après centrifugation à 50000 g) qui va être déposée sur la colonne. La piste 5 représente la fraction non retenue sur la colonne. Les pistes 6 et 7 correspondent aux deux pics majoritaires obtenus lors de l'élution, respectivement pic 2 et pic 3.

## 2.2. Contrôle de la fonctionnalité de DiscII-boucleHPA-mutée

La fonctionnalité de la protéine est à nouveau contrôlée par chromatographie d'affinité sur colonne à GalNAc. Comme pour le mutant 1, DiscII-boucleHPA-mutée n'est pas retenue sur la résine GalNAc-agarose et ce quelque soit la souche d'expression testée.

Le test d'hémagglutination nous a alors permis de tester sa fonctionnalité sur une importante gamme de glycanes, dans le cas où DiscII-boucleHPA-mutée reconnaîtrait un monosaccharide différent du GalNAc. En effet, la surface des hématies de lapin utilisées pour ce test est tapissée de glycanes très variés. Pour DiscII-boucleHPAmutée, aucune hémagglutination n'est visible, quelque soit la concentration en protéine testée (**Figure 5.10.**).

La protéine est visiblement incapable de reconnaître un résidu de sucre. La validité de ce test a auparavant été testée avec DiscII fonctionnelle.



**Figure 5.10. Contrôle de la fonctionnalité de DiscII-boucleHPA-mutée grâce au test d'hémagglutination.**

Les hématies utilisées proviennent du lapin. Le contrôle (-) ne contient pas de lectine. Le contrôle (+) est réalisée avec la protéine DiscII native à  $1 \text{ mg.ml}^{-1}$ .

### 2.3. Contrôle de l'oligomérisation de la protéine

La diffusion dynamique de la lumière est une technique puissante qui permet de mesurer le rayon hydrodynamique d'une molécule en solution et donc d'obtenir des informations sur sa structure quaternaire. Le volume hydrodynamique de DiscII-boucleHPAmutée et de DiscII est de respectivement 4,65 nm et 4,57 nm, ce qui correspond à des molécules d'environ 120 kDa et 100 kDa. Ces protéines seraient donc bien présentes sous forme trimérique.

### 2.4. Résolution de la structure de DiscII-boucleHPAmutée par cristallographie aux rayons X

Après expression dans Rosetta 2 pLysS et purification par chromatographie sur colonne de nickel, DiscII-boucleHPAmutée n'est pas fonctionnelle même si elle adopte très probablement le bon repliement trimérique. La résolution de sa structure par cristallographie aux rayons X aiderait à comprendre cette non-fonctionnalité et orienterait la construction de nouveaux mutants.

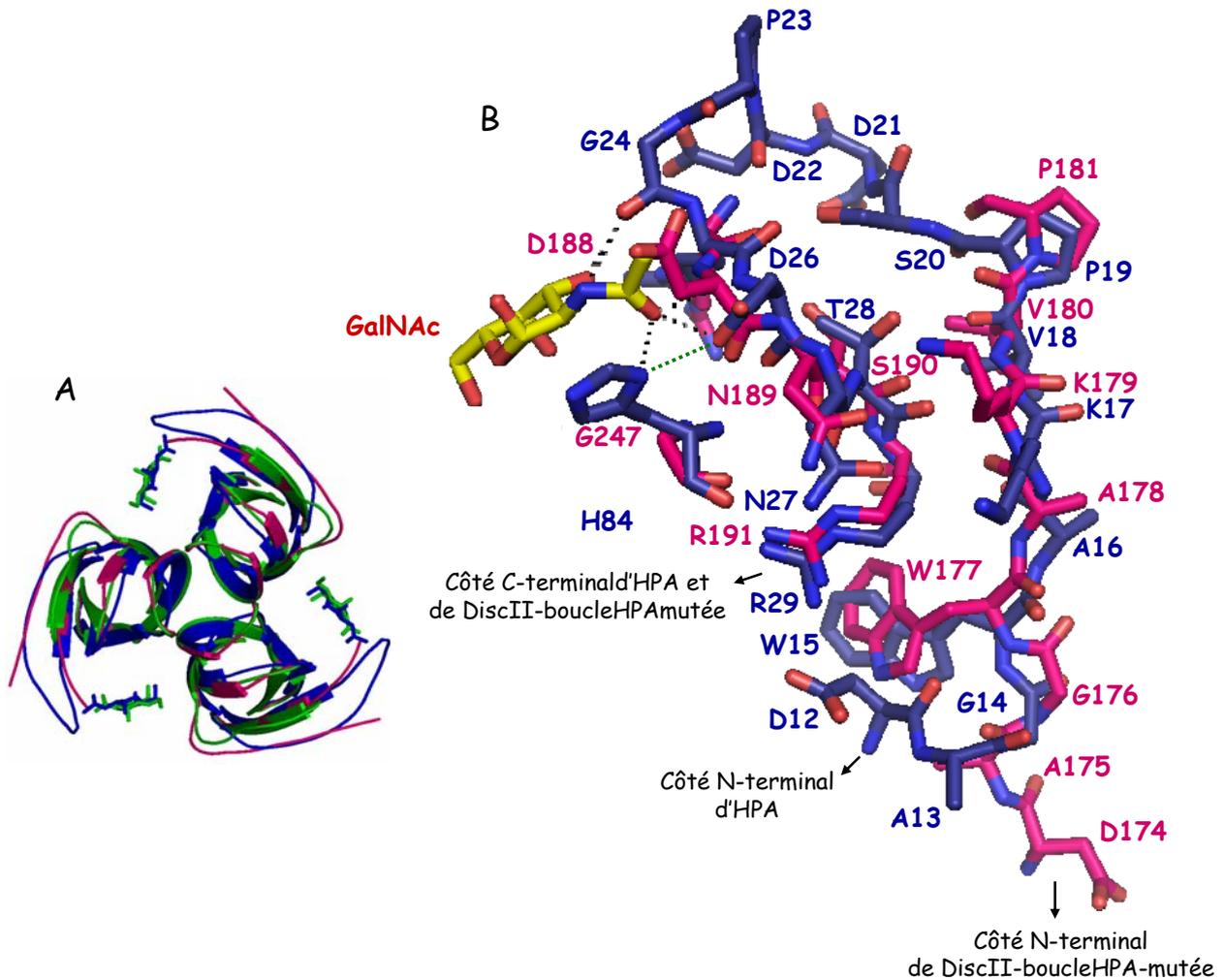
Des cristaux de protéines de forme cubique et bi-pyramidale ont été obtenus dans des solutions de Tris-HCl 100 mM, pH 7,5, formate de sodium 4 M et de HEPES 50 mM, pH 7,5,

$\text{Li}_2\text{SO}_4$  1,2 M, respectivement, à partir d'une solution de protéine à  $5,2 \text{ mg.ml}^{-1}$ . Des essais de co-cristallisation avec une solution de GalNAc et de  $\beta$ -Me-Gal, à 50 mM dans l'eau, ont également été tentés. Après une incubation dans des solutions contenant soit 7,2 M de formate de sodium soit 20% de glycérol (utilisés comme cryo-protectant), les cristaux sont congelés dans de l'azote liquide.

La collecte des données, réalisée à partir des cristaux obtenus dans la solution de formate de sodium, a eu lieu à l'European Synchrotron Radiation Facility (ESRF – Grenoble), sur la ligne de lumière ID23-2. La structure cristallographique de DiscII-boucleHPAmutée a alors été résolue à 1,8 Å. Elle a cristallisé dans le groupe d'espace H32, comme la DiscII native et le trimère est construit autour d'un axe de symétrie d'ordre 3. La structure est résolue par remplacement moléculaire à partir de la structure de DiscII native (PDB 2VM9).

De manière similaire à HPA et DiscII, l'oligomérisation de DiscII-boucleHPAmutée résulte d'une association de trois monomères pour former un trimère. Chaque monomère est formé des deux domaines précédemment décrits. Sur la **Figure 5.11. A** est représentée une superposition des structures cristallographiques des faces de liaisons au sucre (domaine C-terminal) de DiscII-boucleHPAmutée, de DiscII et d'HPA. Ces trois protéines sont globalement organisées de manière équivalente. La boucle de spécificité introduite dans la protéine chimère DiscII-boucleHPAmutée est visible et semble orientée de manière équivalente à ce que l'on observe pour HPA. Malheureusement, l'absence presque totale de densité électronique entre les acides aminés Ser 182 et Gly 186, probablement due à une flexibilité trop importante de la protéine dans le cristal, ne permet pas de construire l'intégralité de cette boucle.

La **Figure 5.11. B** détaille, en représentation en bâtonnets, les acides aminés mutés dans DiscII-boucleHPAmutée et permet donc une comparaison précise avec la structure d'HPA. Concernant les neuf acides aminés (de Asn 173 à Pro 181) précédant l'insertion de la boucle d'HPA, seuls les cinq derniers, de Gly 176 à Pro 181, sont orientés de manière équivalente à HPA. Ceci est d'ailleurs visible sur la Figure 5.11. A. où la boucle de spécificité de DiscII-boucleHPAmuté est davantage orientée comme DiscII que comme HPA. Cependant, ces différences ne semblent pas trop influencer l'orientation des cinq autres acides aminés (de Gly 176 à Pro 181).



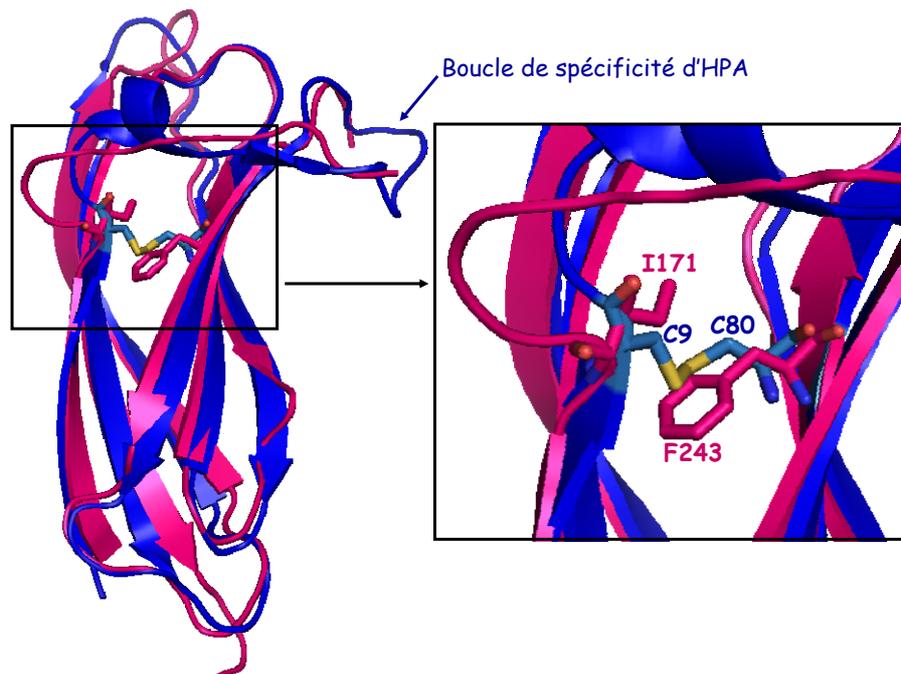
**Figure 5.11. A. Superposition des structures cristallographiques d'une face de liaison au GalNAc de DiscII-boucleHPA mutée (en rose), de DiscII (en vert) et d'HPA (en bleu). B. Boucles de spécificité de DiscII-boucleHPA mutée (en rose) et d'HPA (en bleu).**

A. Les résidus GalNAc représentés en vert et en bleu appartiennent respectivement aux structures de DiscII et d'HPA. B. Le résidu GalNAc appartient à la structure d'HPA. Les pointillés en noir et en vert représentent respectivement les liaisons hydrogènes établies entre HPA et le GalNAc et deux acides aminés d'HPA.

Malheureusement, seuls les trois derniers acides aminés (Arg 187, Asp 188 et Asn 189) de cette boucle de spécificité introduite dans DiscII-boucleHPA mutée ont pu être construits. Et la principale divergence entre les deux structures se situe au niveau de l'Asp 188 (son équivalent chez HPA étant Asp 26). En effet, les chaînes latérale et principale de l'Arg 187, de l'Asn 189, de la Ser 190 et de l'Arg 191 sont orientées de manière parfaitement similaire dans les deux structures. La chaîne latérale de l'Asp 188 dirigée vers l'extérieur provoque un encombrement stérique au niveau du site de reconnaissance du GalNAc, ce qui

pourrait expliquer l'absence de fonctionnalité de la lectine. Dans HPA, l'oxygène de la fonction acide portée par la chaîne latérale de cet acide aminé (Asp 26) forme deux liaisons hydrogène : la première avec l'oxygène du groupement *N*-acétyle du GalNAc et la seconde avec un azote de la chaîne latérale de l'His 84. Cette dernière pourrait être impliquée dans l'orientation de la chaîne latérale de l'Asp 26. De plus, His 84 favorise la stabilisation du GalNAc dans le site de reconnaissance via l'établissement d'une liaison hydrogène avec l'oxygène de son groupe *N*-acétyle et un effet d'empilement des cycles de l'His et du GalNAc. Dans la séquence de DiscII-boucleHPA mutée un résidu Gly 247 se trouve à la place de cette histidine. La troisième construction de protéine implique donc la mutation de Gly 247 en His. La protéine DiscII-boucleHPA mutée G247H est alors construite.

HPA possède également un pont disulfure intramoléculaire établi entre les Cys 9 et 80, qui participe probablement à la stabilisation du monomère de la protéine (**Figure 5.12.**). Situé juste en amont de la boucle de spécificité, il pourrait également être impliqué dans son orientation.



**Figure 5.12. Superposition des structures cristallographiques des monomères de DiscII-boucleHPA muté (en rose) et d'HPA (en bleu).**

Le pont disulfure intramoléculaire présent chez HPA et les acides aminés équivalents chez DiscII-boucleHPA muté sont représentés en bâtonnets.

Un pont disulfure possède une énergie estimée de 5 ou 6 kcal/mol. Sachant qu'une protéine native ou présente dans un état de repliement correct est plus stable de seulement 5 à 10 kcal/mol par rapport à un état dénaturé ou non replié, il est clair qu'un pont disulfure bien formé est impliqué dans la stabilisation des protéines (Zavodszky *et al.* 2001). La déstabilisation provoquée par l'élimination d'un pont disulfure est généralement à l'origine d'une perte de fonction de la protéine même si la structure tridimensionnelle n'en est que très peu affectée (Schulenburg *et al.* 2010). L'insertion de deux résidus cystéines exogènes capables de former un pont disulfure est une stratégie payante pour améliorer la stabilité de certaines protéines produites sous forme recombinante (Pornillos *et al.* 2010 ; Futami *et al.* 2000 ; Hagihara *et al.* 2007 ; Saerens *et al.* 2008). Cette méthode a été appliquée à DiscII-boucleHPA mutée.

A la place des Cys présentes dans HPA, on retrouve, dans la séquence de DiscII-boucleHPA mutée, l'Ile 171 et la Phe 243. Dans le quatrième mutant, ces résidus sont remplacés par deux résidus sérine de manière à établir des liaisons hydrogène, de plus faible énergie qu'un lien covalent soufre-soufre mais probablement plus faciles à exprimer sous forme recombinante. En effet, les difficultés d'expression d'HPA dans la bactérie *E. coli* et la levure *P. pastoris* peuvent être imputées à la mauvaise formation des liens disulfure intra et/ou intermoléculaires. A partir de la protéine chimérique 3, le triple mutant DiscII-boucleHPA mutée-G247H-I171S-F243S est alors construit.

### **3. Mutant 3 : DiscII-boucleHPA muté-G247H et mutant 4 : DiscII-boucleHPA muté-G247H-I171S-F243S**

La construction des mutants DiscII-boucleHPA mutée-G247H et DiscII-Cter-boucleHPA mutée-G247H est réalisée par mutagenèse dirigée classique, à partir des matrices de DiscII-boucleHPA mutée et DiscII-Cter-boucleHPA mutée, respectivement. De manière similaire, les mutants DiscII-boucleHPA mutée-G247H-I171S-F243S et DiscII-Cter-boucleHPA mutée-G247H-I171S-F243S sont construits à partir des matrices DiscII-boucleHPA mutée-G247H et DiscII-Cter-boucleHPA mutée-G247H.

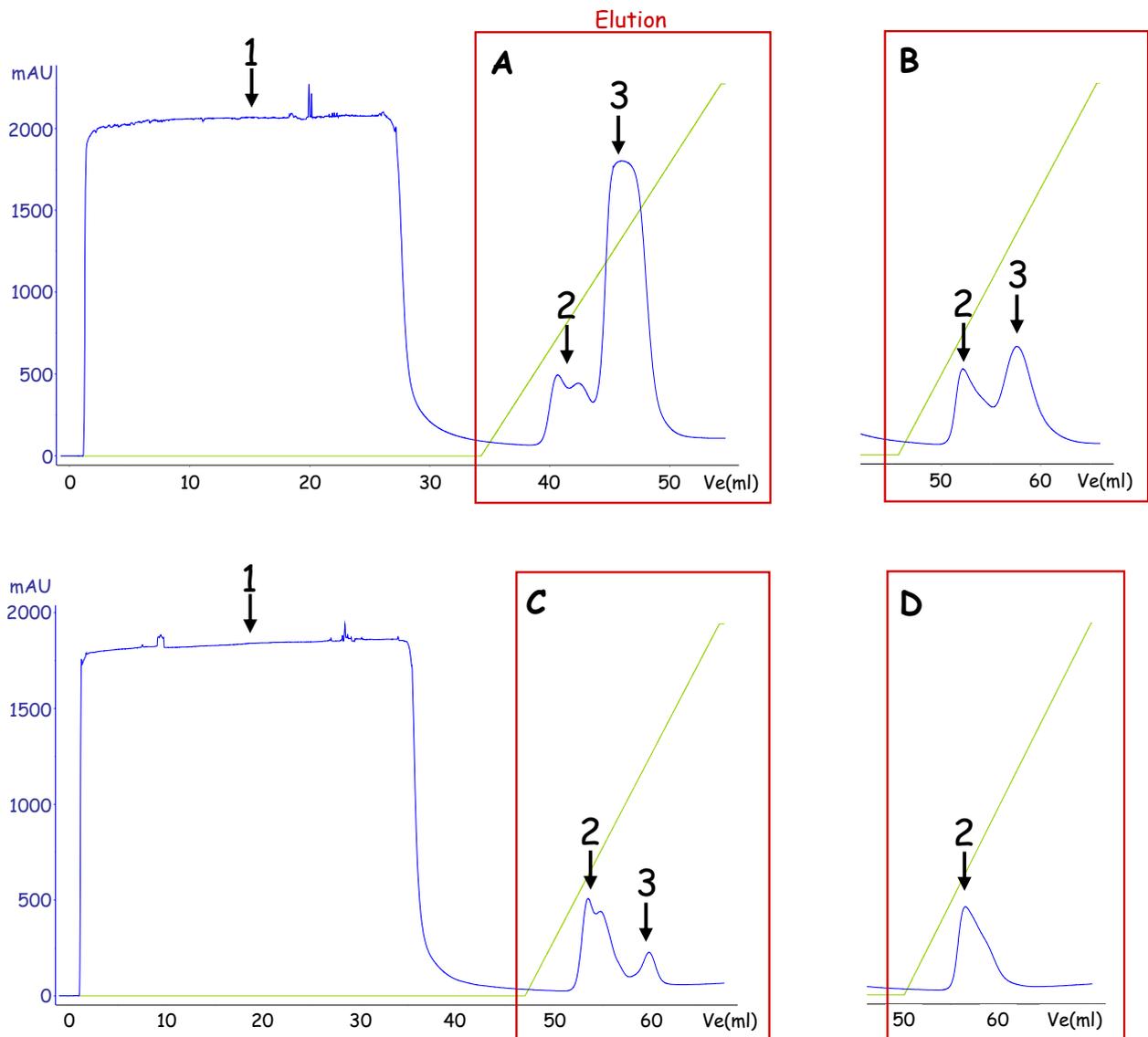
L'expression de ces quatre mutants a été réalisée dans la souche Rosetta 2 pLysS. La densité optique à 600 nm était de 1,2 lorsque l'induction, par ajout de 0,2 mM d'IPTG, a été initiée. Elle s'est poursuivie toute la nuit à 16°C.

Ces protéines sont ensuite purifiées par chromatographie sur colonne de nickel selon le même protocole que celui décrit pour DiscII-boucleHPAmutée. Les chromatogrammes obtenus sont rassemblés dans la **Figure 5.13**. En comparant les profils d'élution obtenus pour DiscII-boucleHPAmutée-G247H (**Figure 5.13. A**) avec DiscII-Cter-boucleHPAmutée-G247H (**Figure 5.13. B.**) et DiscII-boucleHPAmutée-G247H-I171S-F243S (**Figure 5.13. C.**) avec DiscII-Cter-boucleHPAmutée-G247H-I171S-F243S (**Figure 5.13. D.**), il est intéressant de noter que le pic correspondant à la fraction éluée est beaucoup plus important pour les mutants présentant les deux domaines que pour ceux uniquement formés du domaine C-terminal. La présence du domaine N-terminal semble donc être primordiale pour améliorer l'efficacité de l'expression, probablement en stabilisant la formation du trimère.

L'influence des mutations I171S et F243S est, quant à elle, évaluée en comparant les profils d'élution obtenus pour DiscII-boucleHPAmutée-G247H (**A.**) et DiscII-boucleHPAmutée-G247H-I171S-F243S (**C.**). Le pic d'élution pour DiscII-boucleHPAmutée-G247H-I171S-F243S est beaucoup plus petit que celui obtenu avec DiscII-boucleHPAmutée-G247H. Les mutations I171S et F243S défavorisent donc fortement l'expression de ce nouveau mutant. Les mêmes conclusions peuvent être émises en comparant DiscII-Cter-boucleHPAmutée-G247H (**B.**) et DiscII-Cter-boucleHPAmutée-G247H-I171S-F243S (**D.**). L'environnement proche peut être à l'origine de cette altération d'expression. En effet, l'isoleucine et la phénylalanine sont des résidus hydrophobes alors que le groupement hydroxyle de la sérine rend cet acide aminé hydrophile. Enfoui au cœur de la protéine, ce caractère hydrophile peut fortement perturber la stabilité de la protéine et donc nuire à son expression. Il aurait peut-être été intéressant de muter ces résidus I171 et F243 en cystéines et d'exprimer ces protéines dans des souches spécialisées dans la formation de ponts disulfure, telles que les Origami (DE3) ou Rosetta gami (DE3). En effet, mêmes si les groupements thiols (-SH) des cystéines et hydroxyles (-OH) des sérines semblent très similaires, leur caractère hydrophobe est très différent, les cystéines étant fortement hydrophobes par rapport aux sérines (Nagano *et al.* 1999).

Après élimination de l'imidazole par dessalage, la fonctionnalité des quatre mutants est testée par hémagglutination. Aucune activité hémagglutinante n'est détectée, quelque soit la protéine considérée. Ces mutants ne semblent donc pas fonctionnels. L'oligomérisation des protéines n'a pas été testée par DLS. Il est donc difficile de conclure sur les raisons de cette non fonctionnalité des mutants. Elle peut être due à un mauvais repliement protéique qui conduit à la formation incorrecte du site de reconnaissance alors incapable de fixer un résidu

glycanique, quel qu'il soit ou à la non formation du trimère. Une nouvelle stratégie pour l'obtention de protéines chimères est alors envisagée.



**Figure 5.13. Profils d'éluion de la purification par chromatographie sur colonne de nickel de A. DiscII-boucleHPAmutée-G247H, B. DiscII-Cter-boucleHPAmutée-G247H, C. DiscII-boucleHPAmutée-G247H-I171S-F243S et D. DiscII-Cter-boucleHPAmutée-G247H-I171S-F243S.**

Les courbes bleue et verte représentent respectivement la densité optique à 280 nm (exprimée en mAU) et le gradient d'imidazole appliqué pour éluer les protéines. L'encadré rouge représente les fractions éluées. Le pic 1 correspond à la fraction non retenue. Le pic 2 est formé des protéines contaminantes éluées à 100 mM d'imidazole et le pic 3 représente les protéines d'intérêt, éluées autour de 250 mM d'imidazole.

## 4. Mutants 5 et 6 : Construction des protéines hybrides DiscIINter-HPA et DiscIINter-HPA-R163S

Le domaine N-terminal de DiscII semble stabiliser la protéine et favoriser la formation du trimère. De plus, les protéines chimères, malgré les nombreuses tentatives testées restent toujours non fonctionnelles. Nous avons alors construit une protéine hybride composée du domaine N-terminal de DiscII auquel est directement ajoutée HPA (**Figure 5.14.**).

```

1      10      20      30      40      50      60      70
MAVPAGSVSCLANALLNLRSSSTDYNADHGVKNSILNFSNSKDasRFDGSESWSSSVLDKNQFIVAGSDSVKHFVA
80      90      100     110     120     130     140     150
ISTQGRGDHDQWVTSYKLRyTLdNVNwVeyNNGEIINANKDRNSIVTINFNPPiKARsIAIHpQTyNNHISLRWE
160     170     180     190     200     210     220
LYALpVKSySNpRVQSGKIDCGNDAGWAKVPSDDPGRDNTRELAKNITFASpYCRPPVLLSITQLDVEQsQNLr
230     240     250     260
VIARLYSVSPtGfKASCYtWHNTkVYSMSISWISiENy

```

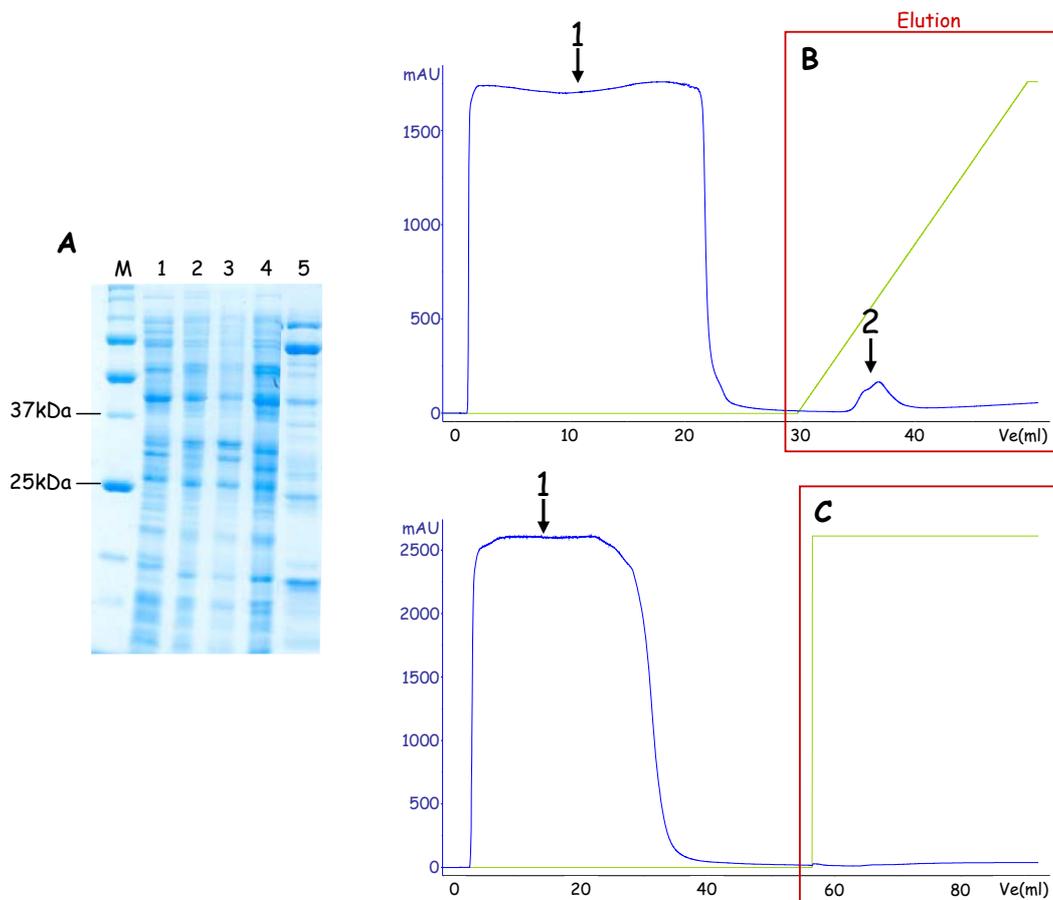
**Figure 5.14. Séquence de la protéine hybride DiscII-Nter-HPA.**

Les séquences écrites en bleu et en noir représentent respectivement le domaine N-terminal de DiscII et son bras espaceur. La séquence d'HPA est inscrite en rouge.

La construction de cet hybride est réalisée par PCR chevauchante à partir des matrices pProEx-DiscII et pET-15-HPA selon le principe développé dans le chapitre 2. Un mutant DiscII-Nter-HPA-R163S est construit par mutagenèse dirigée classique en utilisant comme matrice la protéine hybride précédemment citée. En effet, en étudiant la structure potentielle de cette chimère, il semblait que la chaîne latérale de l'arginine 163 (qui correspond à l'acide aminé 1 dans la séquence d'HPA) présentait un encombrement stérique avec les résidus du bras espaceur. Cet acide aminé est donc remplacé par une sérine, résidu rencontré à cette même place dans la séquence de DiscII entière.

L'expression de ces mutants a été réalisée dans la souche Rosetta gami 2 pLysS. La densité optique à 600 nm était de 0,7 lorsque l'induction, par ajout de 0,2 mM d'IPTG, a été initiée. Elle s'est poursuivie toute la nuit à 16°C. L'analyse de l'expression a été réalisée sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE 12%). Après induction, aucune bande n'est visible à 29 kDa, taille attendue pour cet hybride, ni pour la protéine DiscII-Nter-HPA (**Figure 5.15. A.**) ni pour son mutant DiscII-Nter-HPA-R163S. Le gel SDS-PAGE obtenu pour ce dernier n'est pas présenté dans ce manuscrit. La protéine ne semble pas

être exprimée dans ces conditions. Afin de s'en assurer, une chromatographie sur colonne de nickel (**Figure 5.15 B.**) et une purification sur colonne de GalNAc (**Figure 5.15.**) ont également été testées avec la fraction soluble du mutant DiscII-Nter-HPA. Le seul pic observé lors de l'élution sur colonne de nickel correspond aux protéines contaminantes d'*E. coli*, décrochées avec environ 150 mM d'imidazole (piste 5). Concernant l'élution de la chromatographie d'affinité sur colonne de GalNAc, aucun pic n'est visible. Ces deux nouveaux mutants ne semblent donc pas être exprimés, ni sous forme soluble, ni sous forme de corps d'inclusion.



**Figure 5.15. A. Gel SDS-PAGE 12% acrylamide coloré au bleu de Coomassie après expression de DiscII-NterHPA dans Rosetta gami 2 pLysS et purification sur colonne de nickel (B.) et sur colonne GalNAc-agarose (C.).**

**A.** Les pistes 1 et 2 correspondent aux extraits totaux de la culture utilisée pour la purification, respectivement avant et après induction. Les pistes 3 et 4 représentent les fractions après centrifugation à 12000 g (corps d'inclusion) et à 50000 g (fraction soluble qui va être déposée sur la colonne). La piste 5 représente les protéines présentes le pic 2 (B.)

**B. et C.** Les courbes bleue et verte représentent respectivement la densité optique à 280 nm (exprimée en mAU) et le gradient d'imidazole appliqué pour éluer les protéines. L'encadré rouge représente les fractions éluées. Le pic 1 correspond à la fraction non retenue. Le pic 2 est formé des protéines contaminantes éluées à 100 mM d'imidazole.

## 5. Conclusion

Le but de ce travail de doctorat consistait à développer un marqueur histologique de cellules cancéreuses métastatiques, associées à un faible pronostic vital. HPA possède cette caractéristique mais toutes les tentatives d'expression de cette protéine sous forme recombinante sont restées vaines, quel que soit le système d'expression utilisé. DiscII, à l'inverse, est facilement exprimée dans la bactérie *E. coli* mais ne possède pas ce caractère prédictif. L'idée était donc de construire une protéine chimère qui posséderait les caractéristiques de ces deux lectines. La première stratégie a été de modifier le site de reconnaissance de DiscII pour le transformer en celui présent chez HPA. Plusieurs mutants ont alors été construits. Nos principales tentatives peuvent se résumer ainsi :

- insertion des huit acides aminés qui composent la boucle de spécificité d'HPA dans la séquence de DiscII.

- remplacement des neuf résidus précédents de manière à orienter au mieux cette boucle.

Malheureusement, même si les rendements d'expression obtenus étaient très satisfaisants, les chimères produites n'étaient pas capables de reconnaître et de fixer une structure de sucre. La structure cristallographique de ce deuxième mutant a néanmoins été résolue, ce qui nous a permis d'orienter avec plus de précision les nouvelles mutations à envisager. En effet, basé sur l'observation des différences majeures entre les structures cristallographiques de cette chimère et HPA, nous avons décidé, par la suite, de muter la glycine 247 présente dans le site de reconnaissance des mutants de DiscII en histidine observée chez HPA puis de remplacer les résidus I171 et F243 en sérine de manière à former un réseau de liaisons hydrogènes au niveau du pont disulfure intramoléculaire d'HPA. Les protéines chimères étaient toujours produites sous forme recombinante mais restaient non fonctionnelles.

La seconde stratégie était de construire une protéine hybride possédant le domaine N-terminal et le bras espaceur de DiscII et le domaine C-terminal d'HPA. En effet, les capacités de reconnaissance des épitopes cancéreux d'HPA sont localisées au niveau du domaine C-terminal alors que le N-terminal de DiscII semble favoriser l'oligomérisation trimérique de la protéine, indispensable à sa fonction. Un autre mutant dans lequel la première arginine d'HPA a été remplacé par une sérine à cause d'une suspicion d'encombrement stérique a également été construit. Dans ce deux cas, aucune expression n'a pu être détectée.

Le **tableau 5.2** récapitule les différents mutants construits ainsi que les résultats obtenus.

Nom du mutant	Caractéristiques	Résultat
DiscII-boucleHPA	DiscII entière avec insertion des huit acides aminés que composent la première boucle de surface d'HPA	Expression dans <i>E. coli</i> .
DiscII-Cter-boucleHPA	DiscII-Cter avec insertion des huit acides aminés que composent la première boucle de surface d'HPA	Probablement organisées sous forme de monomères.
DiscII-boucleHPA-mutée	Mutation des neuf acides aminés qui précèdent la boucle d'HPA dans DiscII-boucleHPA	Expression dans <i>E. coli</i> . Visiblement organisée sous forme trimérique. Résolution de sa structure par cristallographie aux rayons X.
DiscII-boucleHPA-mutée-G247H	Mutation de la Gly 247 en His, dans DiscII-boucleHPA-mutée, de manière à favoriser des interactions avec le GalNAc.	Expression dans <i>E. coli</i> .
DiscII-boucleHPA-mutée-G247H-I171S-F243S	Mutations des Ile 171 et Phe 243 en Ser dans DiscII-boucleHPA-mutée-G247H, de manière à provoquer une interaction mimant le pont disulfure intramoléculaire d'HPA.	Expression faible dans <i>E. coli</i> .
Hybride DiscIINter-HPA	Formation d'une protéine hybride entre le domaine N-ter de DiscII et HPA.	Aucune expression détectée.
Hybride DiscIINter-HPA-R163S	Mutation de l'Arg 163 en Ser dans Hybride DiscIINter-HPA, de manière à limiter un éventuel encombrement stérique.	Aucune expression détectée.

**Tableau 5.2. Récapitulatif des différents mutants de DiscII construits.**

Néanmoins, ces expériences ont permis de mettre en évidence l'importance d'un protocole de purification standardisé pour obtenir des résultats reproductibles. Ainsi, un protocole optimisé a été mis au point et les rendements de purification observés grandement améliorés. De plus, nous avons pu constater que la technique de diffusion de la lumière était

d'une grande utilité dans l'étude de l'état d'oligomérisation de lectines. En effet, ces protéines étant capables de reconnaître et fixer des structures oligosaccharidiques, l'emploi de chromatographies d'exclusion stérique avec des colonnes présentant des résines à base de sucre est parfois un peu périlleux. Le test d'héماغglutination est également un moyen rapide de tester la fonctionnalité de lectines multivalentes.

Une étude *in silico*, à partir de la structure cristallographique obtenue avec le mutant 2 DiscII-boucleHPA-muté serait une piste à exploiter pour tenter d'identifier les régions susceptibles de perturber la fonctionnalité de cette protéine. Ainsi, de nouvelles mutations pourraient être construites dans l'espoir d'obtenir des chimères capables de reconnaître une structure de sucre, et notamment le GalNAc, monosaccharide spécifique d'HPA.

Un travail similaire pourrait également être tenté avec la seconde discoidine, DiscI, également homologue à HPA qui présente plus d'affinité que DiscII pour le GalNAc.

# Chapitre 6

Etude fonctionnelle des domaines

N-terminaux de DiscI et DiscII



---

## Chapitre 6

# Etude fonctionnelle des domaines N-terminaux de DiscI et DiscII

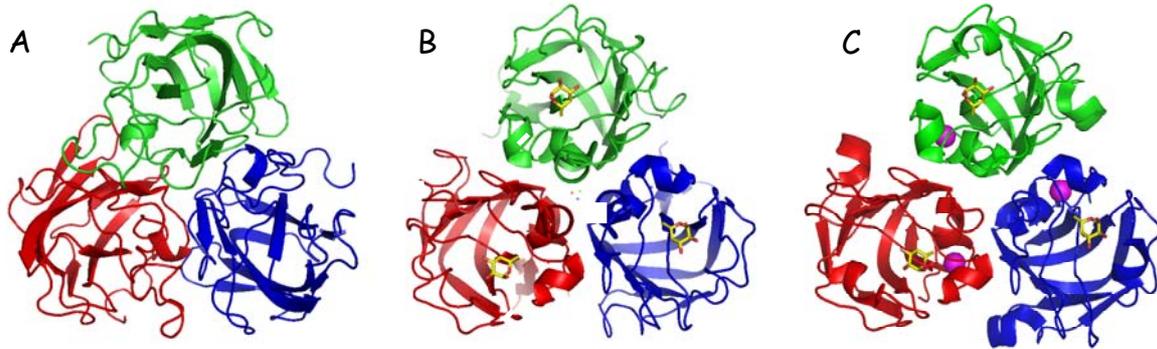
---

Les domaines N-terminaux de DiscI et DiscII, également dénommés domaines discoidine (DS), sont retrouvés dans un grand nombre de protéines eucaryotes et procaryotes telles que des facteurs de coagulation, des enzymes, des récepteurs et des protéines impliquées dans le développement neuronal. Le rôle biologique de ces domaines est basé sur leurs capacités d'interactions avec une variété de molécules telles que les facteurs de croissance, les collagènes, les lipides et phospholipides, et certains glucides (Baumgartner *et al.* 1998). Chez l'amibe *Dictyostelium discoideum*, la fonction précise assurée par ces domaines est toujours inconnue à ce jour.

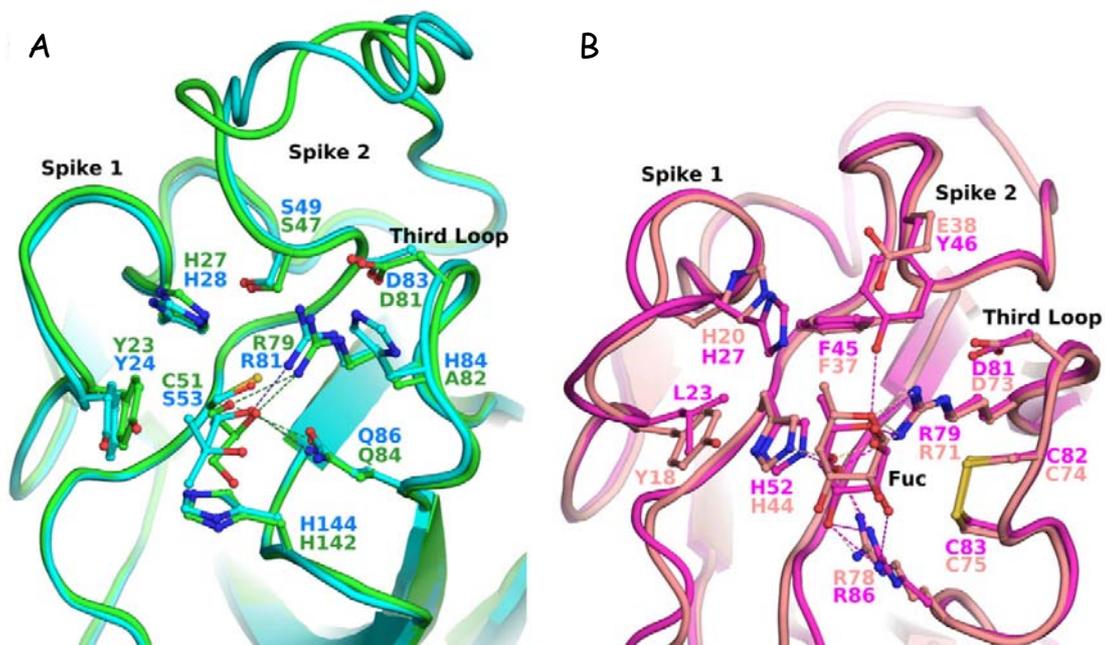
Certaines de ces protéines sont capables de reconnaître des résidus glycaniques. Les domaines N-terminaux de DiscI et DiscII possèdent des homologues structuraux appartenant à la famille de lectine de type F, spécifiques du fucose tels que la lectine AAA de l'anguille *Anguilla anguilla* (PDB 1K12) (Bianchet *et al.* 2002) ou la lectine F du loup de mer (PDB 3CQO) (Bianchet *et al.* 2010). La résolution de ces structures cristallographiques démontre qu'elles adoptent un repliement trimérique. AAA est un homotrimère alors que la lectine F est composée de deux domaines très similaires, formé chacun par l'association de trois monomères liés aux trois autres de manière queue à queue. Ces lectines présentent un site de reconnaissance à fucose sur chaque monomère (**Figure 6.1**).

Les discoidines présentent, elle aussi, un site de reconnaissance, localisé au même endroit que le site à fucose des lectines de type F mais qui abrite des molécules de MPD et glycérol (**Figure 6.2**). Ces molécules sont souvent retrouvées dans les sites de fixation des sucres. L'environnement du ligand dans ces différentes protéines présente de fortes similarités. Un résidu hydrophobe (Tyr 23 dans DiscI, Leu 23 dans AAA, Tyr 18 dans les lectines de type F), qui forme des interactions d'empilement avec le ligand est toujours présent dans la première boucle. La troisième boucle des discoidines, quant à elle, contient un

motif RGD qui favorise l'attachement au substrat dans les cellules de mammifères (Ruoslahti 2003). L'arginine de ce motif est conservée dans les lectines de type F homologues et forme des liaisons hydrogènes avec l'oxygène du cycle et le groupement hydroxyle axial du carbone 4, sélectionnant la configuration *L-galacto* du fucose.



**Figure 6.1.** Représentations schématiques des trimères du domaine N-terminal des discoidines (A.), de la lectine AAA complexée au fucose (PDB 1K12) (Bianchet *et al.* 2002) (B.) et de la lectine F complexée au fucose (PDB 3CQO) (Bianchet *et al.*) (C.). Les sphères magenta représentent des ions calcium.



**Figure 6.2.** Superposition de site de reconnaissance du domaine DS.

**A.** Site de reconnaissance du glycérol/MPD dans DiscI (vert) et DiscII (cyan), respectivement. **B.** Site de reconnaissance du fucose dans AAA (magenta) et lectine F du loup de mer (saumon) placé selon la même orientation que les discoidines.

De plus, la technique des puces à sucres a révélé que DiscI et DiscII reconnaissent également un petit nombre d'oligosaccharides terminés par des résidus fucoside. Ces résultats ont été développés dans le chapitre 4. Le site de liaison du domaine N-terminal des discoidines présente donc toutes les caractéristiques nécessaires pour la fixation de glycanes fucosylés. Les discoidines seraient donc formées de deux domaines dits « lectine » distincts, possédant chacun une spécificité propre.

Afin de vérifier cette hypothèse, j'ai cloné les domaines N-terminaux de DiscI et DiscII de manière isolée. Après expression sous forme recombinante dans *E. coli* et purification par chromatographie sur colonne de nickel, la spécificité de ces protéines a été étudiée en utilisant les tests d'hémagglutination et la technique de résonance plasmonique de surface.

## 1. Le domaine N-terminal de DiscI

Afin d'optimiser la production du domaine N-terminal de DiscI, trois constructions géniques de longueurs différentes ont été réalisées :

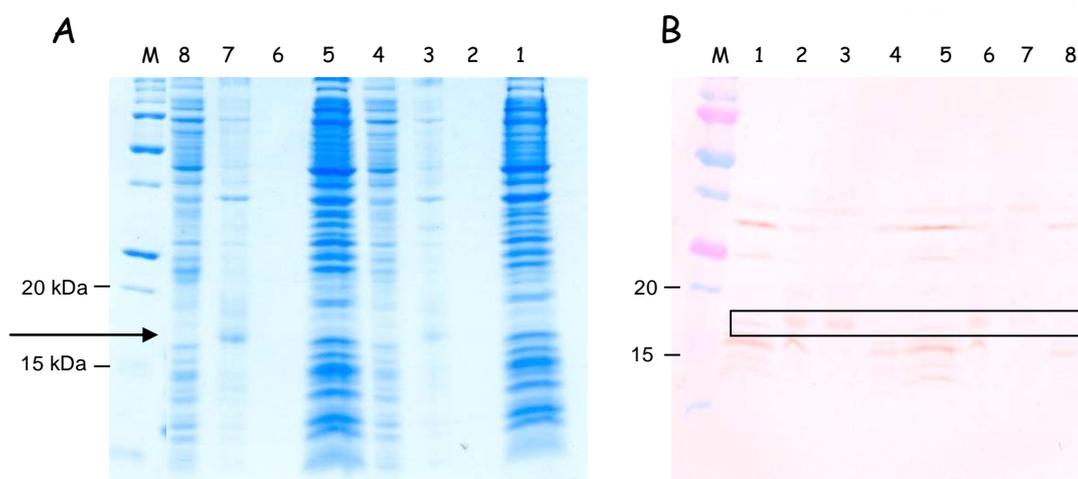
- DiscI-Nter-P153 de Met 1 à Pro 153
- DiscI-Nter-Q155 de Met 1 à Gln 155
- DiscI-Nter-S157 de Met 1 à Ser 157.

Le vecteur utilisé pour le clonage, pProEx-HtB, ainsi que la souche bactérienne, Rosetta 2 (DE3) pLysS sont identiques à ceux utilisés pour l'expression de la protéine entière. Différentes conditions d'expression ont été testées afin d'optimiser l'obtention de ce domaine sous forme soluble et en quantité importante. Les différents essais sont répertoriés dans le **tableau 6.1**.

Protéine	DO <sub>600</sub> d'induction	IPTG (mM)	Température (°C)	Temps d'induction
DiscI-Nter-P153	1	0,2	16°C	1 nuit
	0,7	0,2	16°C	1 nuit
			30°C	4 heures
DiscI-Nter-Q155	0,7	0,2	16°C	1 nuit
DiscI-Nter-S157	0,7	0,2	16°C	1 nuit

**Tableau 6.1. Récapitulatif des conditions d'expression dans Rosetta 2 (DE3) pLysS des domaines N-terminaux de DiscI.**

Les tentatives d'expression de ce domaine ont tout d'abord été réalisées avec la protéine la plus courte, DiscI-Nter-P153. Leur efficacité est évaluée par analyse sur gel SDS-PAGE et Western Blot en contrôlant la présence de la protéine d'intérêt avant et après induction et sa proportion sous forme soluble (**Figure 6.3.**). Sur le gel SDS-PAGE (**Figure 6.3. A.**), les témoins après induction (pistes 2 et 6) n'ont pas été déposés à cause de leur grande viscosité. En revanche, une bande d'environ 17 kDa est présente dans la fraction contenant les protéines sous forme de corps d'inclusion (pistes 3 et 7). Aucune bande n'est visible à ce poids moléculaire dans les fractions solubles (pistes 4 et 8), ce qui laisse supposer que l'intégralité du domaine est produite sous forme de corps d'inclusion. Le marquage spécifique sur l'immunoblot (**Figure 6.3. B.**) confirme que cette protéine est bien la protéine d'intérêt. Deux bandes, très faiblement marquées, sont présentes dans le témoin après induction (pistes 2 et 6) et dans le culot (pistes 3 et 7), pour les deux conditions testées. Les autres bandes visibles, beaucoup plus fines et plus intenses, sont des bandes parasites. Elles ne correspondent pas à un marquage spécifique de DiscI-Nter-P153.



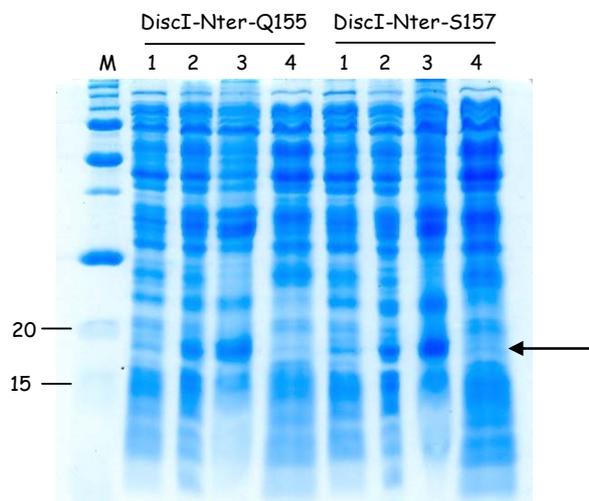
**Figure 6.3. Expression de DiscI-Nter-P153 dans *E. coli* Rosetta 2 (DE3) pLysS. A. Analyse par SDS-PAGE 15% et B. Immunoblot.**

Les pistes 1 et 5 correspondent aux témoins avant induction à, respectivement, 30°C pendant 4 heures et 16°C durant une nuit. Les pistes 2 et 6 représentent les témoins après induction correspondant. Les pistes 3 et 7 contiennent les fractions contenant la protéine sous forme de corps d'inclusion (après centrifugation à 12000 g). Les pistes 4 et 8 représentent les fractions solubles (après centrifugation à 50000 g). Ces dernières sont déposées sur la colonne pour la chromatographie sur colonne de nickel.

Le domaine N-ter de DiscI-P153 serait donc faiblement produit et majoritairement sous forme de corps d'inclusion. Ces résultats ont été confirmés par chromatographie sur colonne de nickel (chromatogramme non présenté) puisque, après élution par un gradient d'imidazole, aucune protéine n'est décrochée. La densité optique d'induction ainsi que la température et la durée d'induction ne semble pas influencer l'expression de ce domaine puisque, quelque soit les conditions considérées les résultats sont identiques.

L'extension de deux et quatre acides aminés du côté C-terminal par rapport à DiscI-Nter-P153, pour former respectivement les mutants DiscI-Nter-Q155 et DiscI-Nter-S157, ne semble pas améliorer l'expression de ce domaine. Sur le gel SDS-PAGE (**Figure 6.4.**), une bande intense est présente à 17 kDa dans le témoin après induction (piste 2), pour les deux protéines considérées. Cette bande est quasiment intégralement retrouvée dans la fraction contenant les corps d'inclusion (piste 3) alors que seulement une très légère bande est visible dans la fraction soluble (piste 4). Ces deux nouveaux domaines sont, comme DiscI-Nter-P153, très majoritairement exprimés sous forme de corps d'inclusion, dans les conditions testées.

Malheureusement, l'impossibilité d'obtenir une protéine sous forme soluble, dans les conditions d'expression testées, limite considérablement les analyses fonctionnelles à réaliser.



**Figure 6.4. Contrôle de l'expression de DiscI-Nter-Q155 et DiscI-Nter-S157 dans Rosetta 2 (DE3) pLysS, par analyse SDS-PAGE 15%.**

Les pistes 1 et 2 correspondent respectivement aux témoins avant et après induction. La piste 3 représente la fraction contenant la protéine sous forme de corps d'inclusion (après centrifugation à 12000 g). La piste 4 représente la fraction soluble (après centrifugation à 50000 g).

## 2. Le domaine N-terminal de DiscII

### 2.1. Production et purification de DiscII-Nter

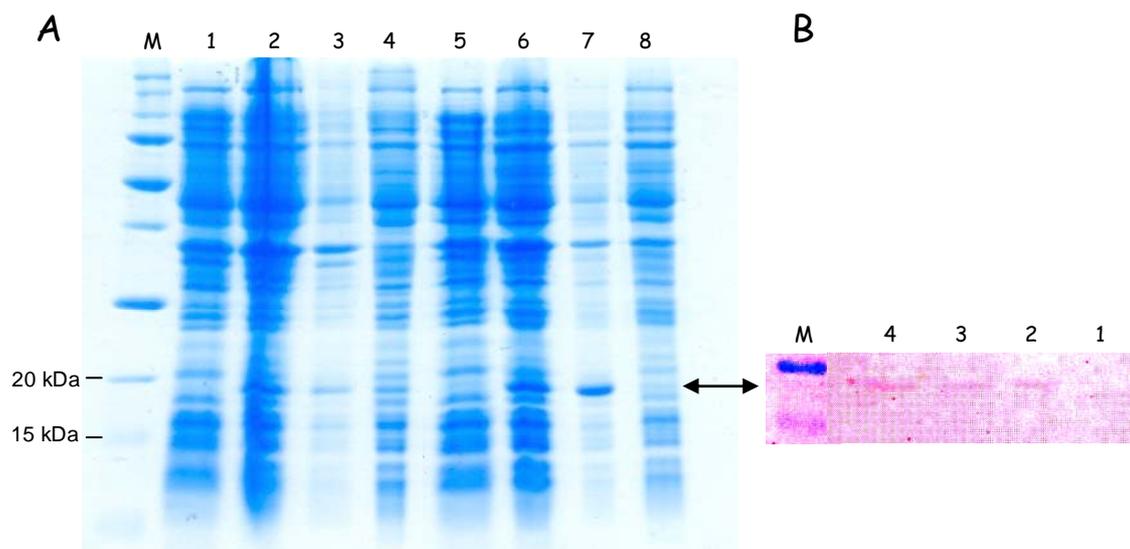
Parallèlement aux manipulations décrites précédemment, le domaine N-terminal de DiscII a également été exprimé sous forme recombinante dans *E. coli*. Un clone, DiscII-Nter, allant de la Met 1 à la Leu 154, a été construit dans le vecteur pProEx-HtB. Trois conditions d'expression différentes, référencées dans le **tableau 6.2**, ont été testées dans la souche Rosetta 2 pLysS. Les résultats obtenus pour l'expression dans les conditions où la densité optique d'induction était de 0,7 sont représentés sur le gel SDS-PAGE **Figure 6.5. A**.

Protéine	DO <sub>600</sub> d'induction	IPTG (mM)	Température (°C)	Temps d'induction
DiscII-Nter	1,2	0,2	16°C	1 nuit
	0,7	0,2	16°C	1 nuit
			30°C	4 heures

**Tableau 6.2. Récapitulatif des conditions d'expression dans Rosetta 2 (DE3) pLysS du domaine N-terminal de DiscII.**

A 19 kDa, une bande apparaît dans le témoin après induction (pistes 2 et 6) alors qu'elle est absente avant induction (pistes 1 et 5), dans les deux conditions testées. Pour les échantillons extraits de l'expression à 16°C sur la nuit, la protéine est présente à la fois sous forme de corps d'inclusion (piste 3) et sous forme soluble (piste 4) alors qu'à 30°C, l'intégralité de la protéine est exprimée sous forme de corps d'inclusion (piste 7). Ce résultat peut s'expliquer par le fait qu'à une température élevée, la croissance bactérienne est rapide. Le nombre de protéines produites sous forme recombinante est alors bien supérieur à celui des chaperones disponibles, limitant ainsi leur bon repliement tridimensionnel. Ces résultats ont été confirmés par Westernblot (**Figure 6.5. B**).

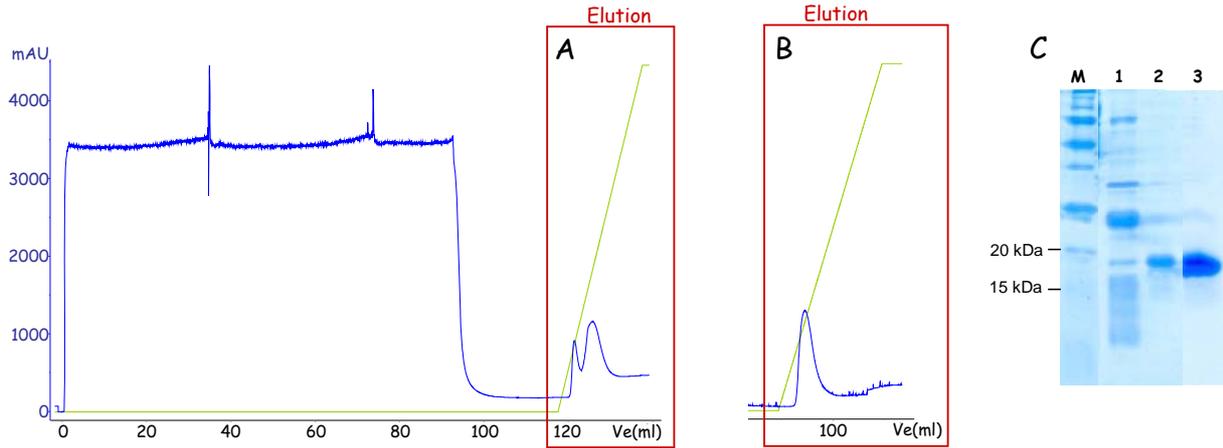
La densité optique d'induction ne semble pas avoir d'influence sur le rendement protéique puisque, au terme de l'induction, la densité bactérienne est identique que l'induction ait eu lieu à une densité de 0,7 ou 1,2. Le rendement n'est pas modifié.



**Figure 6.5. Contrôle de l'expression de DiscII-Nter dans Rosetta 2 (DE3) pLysS, par A. analyse SDS-PAGE 15% et B. immunoblot.**

Les pistes 1 et 5 correspondent aux témoins avant induction à, respectivement, 16°C pendant une nuit et 30°C durant 4 heures. Les pistes 2 et 6 représentent les témoins après induction correspondant. Les pistes 3 et 7 contiennent les fractions contenant la protéine sous forme de corps d'inclusion (après centrifugation à 12000 g). Les pistes 4 et 8 représentent les fractions solubles (après centrifugation à 50000 g).

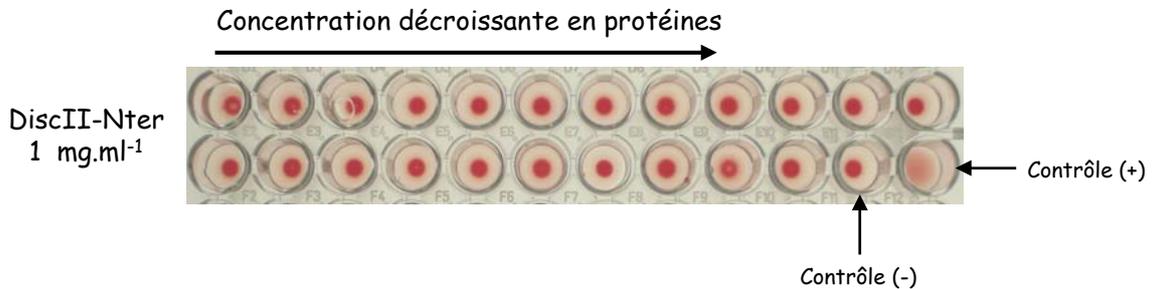
De manière similaire aux autres protéines étudiées, la fraction de DiscII-Nter, exprimée sous forme soluble, est purifiée par chromatographie sur colonne de nickel et analysée par SDS-PAGE (**Figure 6.6.**). Après standardisation du protocole (volume de suspension des cellules utilisée, pression pour éclater les cellules...), l'élution, réalisée par gradient d'imidazole, n'est, là encore, pas très reproductible. Dans un premier test, deux pics distincts à 100 mM et 210 mM d'imidazole (**Figure 6.6. A.**), correspondant respectivement aux protéines contaminantes d'*E. coli*. (**Figure 6.6. C.** piste 1) et à DiscII-Nter, élue de manière relativement pure (**Figure 6.6. C.** piste 2), sont visibles. La fraction dite « contaminante » contient également, en faible quantité, la protéine d'intérêt. Dans un second essai, l'élution donne naissance à un seul pic à 125 mM d'imidazole (**Figure 6.6. B.**) contenant majoritairement la protéine d'intérêt (**Figure 6.6. C.** piste 3).



**Figure 6.6.** Profils d'éluion obtenus lors de deux purifications différentes sur colonne de nickel du domaine N-terminal de DiscII (A. et B.) et analyse SDS-PAGE 15% (C.)

## 2.2. Contrôle de la fonctionnalité de DiscII-Nter

La fonctionnalité de DiscII-Nter a, tout d'abord, été contrôlée en utilisant le test d'hémagglutination. Malheureusement, aucune agglutination n'a été observée (**Figure 6.7.**).



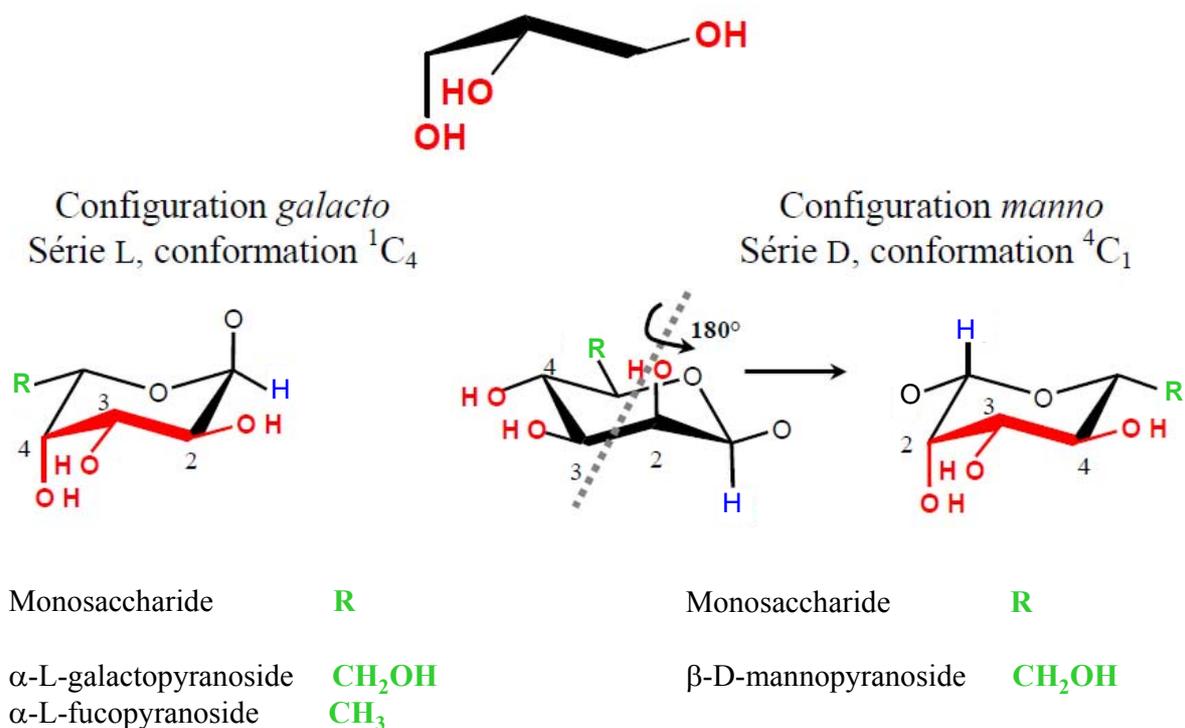
**Figure 6.7.** Contrôle de la fonctionnalité de DiscII-Nter grâce au test d'hémagglutination.

Les hématies utilisées proviennent du lapin. Le contrôle (-) ne contient pas de lectine. Le contrôle (+) est réalisée avec la protéine PA-IL de *P. aeruginosa* à 0,5 mg.ml<sup>-1</sup>.

La résonance plasmonique de surface a alors été utilisée afin de vérifier si DiscII-Nter était capable, comme son homologue structural, la lectine AAA, de reconnaître et de fixer le résidu fucoside. Ces manipulations ont été réalisées par Samy Cecioni au National Center for Biomolecular Research dans l'université Masaryk, à Brno en République Tchèque.

L'interaction a, tout d'abord, eu lieu entre des sucres greffés sur un polymère de polyacrylamide (PAA), immobilisés sur le biocapteur et la protéine qui passe en solution. Le

PAA- $\alpha$ -D-Gal, à priori non reconnu par les lectines dites « à fucose », tapisse le canal 1 et est utilisé comme référence. Les pistes 2 et 4 contiennent respectivement 100% de PAA- $\alpha$ -D-Man et 100% de PAA- $\alpha$ -L-Fuc. La piste 3 est composée d'un mélange équimolaire de ces deux sucres. En effet, le D-mannose et le L-fucose présentent des structures tridimensionnelles relativement proches. Les configurations L-*galacto* (adoptée par le fucose) et D-*manno* présentent deux groupements hydroxyles équatoriaux et un groupement hydroxyle axial lorsque le cycle pyranose est respectivement dans une conformation  ${}^1C_4$  et  ${}^4C_1$ . Une rotation de  $180^\circ$  des monosaccharides de la série D-*manno* positionne les groupements hydroxyles dans la même orientation que ceux de la série L-*galacto* (**Figure 6.8**). Les lectines spécifiques du fucose sont donc généralement capables de reconnaître le mannose avec une affinité moindre, comme par exemple la lectine bactérienne PA-III de *Pseudomonas aeruginosa* (Imberty *et al.* 2004).

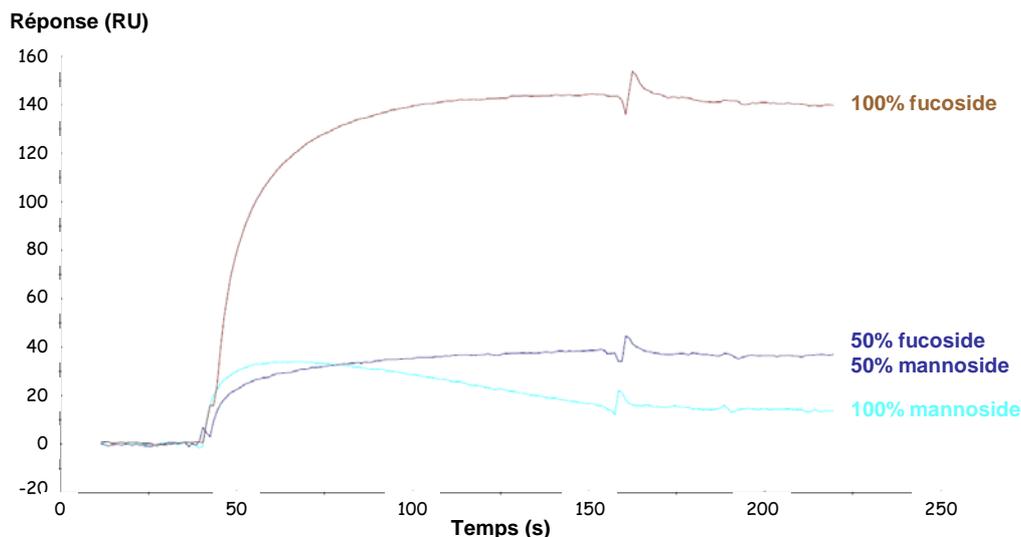


**Figure 6.8.** Comparaison des structures des monosaccharides de configuration L-*galacto* et D-*manno*.

Ce schéma est tiré de la thèse de Charles Sabin, intitulée « La lectine PA-III de *Pseudomonas aeruginosa* : Structure, affinité et spécificité pour des ligands naturels et glycomimétiques. »

Les sensorgrammes obtenus pour une concentration en protéine égale à  $50 \mu\text{g} \cdot \text{ml}^{-1}$  présentent une phase d'association non linéaire qui se termine par un plateau de saturation.

Les valeurs de ce plateau sont de 140 RU pour l'interaction avec le PAA- $\alpha$ -L-Fuc et d'environ 35 RU pour celle avec le mélange PAA- $\alpha$ -L-Fuc / PAA- $\alpha$ -D-Man. Une réponse de 40 RU est également observée sur le canal ne contenant que du PAA- $\alpha$ -D-Man mais la saturation est observée à 0 RU (**Figure 6.9.**). Cette expérience préliminaire semble donc indiquer que le domaine N-terminal de DiscII est capable de reconnaître les résidus  $\alpha$ -fucoside et  $\alpha$ -mannoside, ce dernier avec une affinité moins forte.



**Figure 6.9. Mesure d'interaction entre DiscII-Nter et les sucres PAA- $\alpha$ -L-Fuc et PAA- $\alpha$ -D-Man par RPS.**

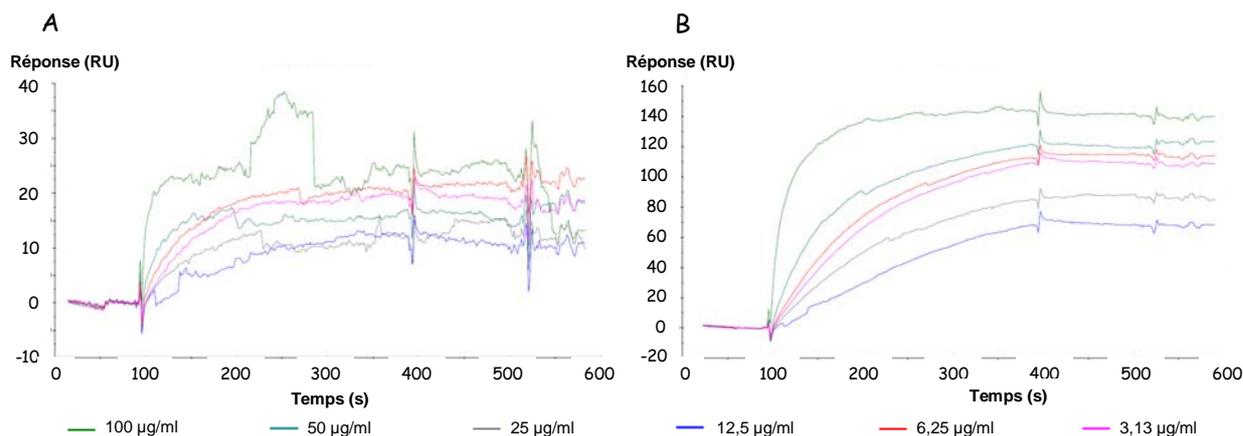
Ces sensorgrammes sont issus de la soustraction de la piste contenant soit 100 de PAA-Man, soit 100% de PAA-Fuc soit un mélange équimolaire des deux à la piste de référence présentant 100% de PAA-Gal, à priori non reconnu par DiscII-Nter.

De manière similaire, l'interaction est, à nouveau, mesurée en faisant varier la concentration en protéine injectée à la surface du biocapteur. La **Figure 6.10** présente les sensorgrammes obtenus sur les pistes contenant 100% de PAA-Man (**A.**) et 100% de PAA-Fuc (**B.**). Dans les deux cas, une réponse est détectée. La réponse maximale obtenue pour la plus grande concentration en protéine ( $100 \mu\text{g}\cdot\text{ml}^{-1}$ ) est d'environ 25 RU pour le PAA-Man et 140 RU pour le PAA-Fuc. La faible intensité de réponse observée pour l'interaction avec le PAA-Man est à l'origine de la qualité médiocre des spectres. Néanmoins, cette expérience permet de confirmer les résultats démontrés précédemment.

En revanche, il est important de noter que l'intensité de la réponse ne décroît pas systématiquement en fonction de la concentration en protéine injectée, comme ce qui pouvait

être attendu. Une véritable optimisation de la quantité de sucre immobilisée ainsi que de la concentration en protéine injectée permettrait probablement de corriger ces résultats surprenants.

La technique de résonance plasmonique de surface a cependant permis de détecter pour la première fois une interaction de type lectine-sucre entre le domaine DiscII-Nter et les résidus monosaccharidiques fucoside et mannoside.



**Figure 6.10. Variation de l'intensité d'interaction entre le PAA- $\alpha$ -D-Man (A.), le PAA- $\alpha$ -L-Fuc (B.) en fonction de la concentration en protéine DiscII-Nter par RPS.**

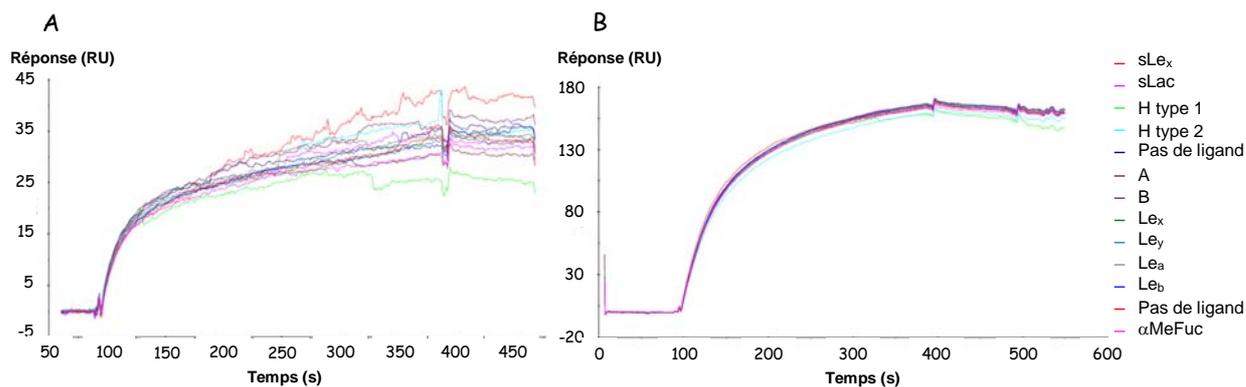
Ces sensorgrammes sont issus de la soustraction des pistes contenant soit 100% de PAA-Man (A.), soit 100% de PAA-Fuc (B.) à la piste de référence présentant 100% de PAA-Gal, à priori non reconnue par DiscII-Nter.

Des cinétiques d'inhibition, dans lesquelles sont mesurées la capacité de chaque composé à inhiber l'interaction entre la protéine et les PAA-sucre immobilisés, ont également été testées. Les inhibiteurs sont alors mis en solution avec DiscII-Nter et injectés sur les PAA-sucre à la surface du biocapteur. Plus l'interaction entre la protéine et l'inhibiteur sera importante, moins la lectine interagira avec les sucres immobilisés et plus la réponse détectée sera faible. Ainsi, les inhibiteurs peuvent être classés en fonction de leur capacité d'inhibition. Dans notre étude préliminaire, les ligands utilisés comme inhibiteur sont référencés dans le **tableau 6.3** et leur concentration est maintenue constante ( $c=5 \mu$ M).

Nom de l'inhibiteur	Abbréviation	Structure de l'inhibiteur
alpha-méthyl-Fucoside	$\alpha$ -Me-Fuc	
Groupe sanguin H de type 1 tétraose	H type 1	Fuc $\alpha$ 1-2Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal
Groupe sanguin H de type 2 tétraose	H type 2	Fuc $\alpha$ 1-2Gal $\beta$ 1-4GlcNAc $\beta$ 1-3Gal
Groupe sanguin A triaose	A	GalNAc $\alpha$ 1-3(Fuc $\alpha$ 1-2)Gal
Groupe sanguin B tétraose	B	Gal $\alpha$ 1-3(Fuc $\alpha$ 1-2)Gal $\beta$ 1-4Glc
Lewis x tétraose	Le <sub>x</sub>	Gal $\beta$ 1-4(Fuc $\alpha$ 1-3)GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc
Sialyl lewis x	sLe <sub>x</sub>	NeuAc $\alpha$ 2-3Gal $\beta$ 1-4(Fuc $\alpha$ 1-3)GlcNAc
Lewis y pentaose	Le <sub>y</sub>	Fuc $\alpha$ 1-2Gal $\beta$ 1-4(Fuc $\alpha$ 1-3)GlcNAc $\beta$ 1-4Gal
Lewis a tétraose	Le <sub>a</sub>	Gal $\beta$ 1-3(Fuc $\alpha$ 1-4)GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc
Lewis b pentaose	Le <sub>b</sub>	Fuc $\alpha$ 1-2Gal $\beta$ 1-3(Fuc $\alpha$ 1-4)GlcNAc $\beta$ 1-3Gal
Sialyl lactose	sLac	NeuAc $\alpha$ 2-3Gal $\beta$ 1-4Glc

**Tableau 6.3. Récapitulatif des composés testés comme inhibiteurs de l'interaction DiscII-Nter et les sucres PAA-biotinylés, PAA- $\alpha$ -D-Man (A.) et PAA- $\alpha$ -L-Fuc (B.) en RPS.**

Les sensorgrammes obtenus (**Figure 6.11.**) avec le PAA-Man (A.) et le PAA-Fuc (B.) présentent la même forme, avec ou sans inhibiteur et quelque soit l'inhibiteur mis en solution. Or, certains inhibiteurs présentent des résidus fucosides disponibles qui pourrait interagir avec DiscII-Nter et provoquer une inhibition. Les conditions utilisées pour réaliser cette expérience ne sont pas optimales. La concentration en inhibiteurs est probablement trop faible pour provoquer une inhibition. Ainsi, il est impossible de conclure sur un éventuel ligand favori pour DiscII-Nter.



**Figure 6.11. Compétition entre des inhibiteurs en solution et les sucres PAA-biotinylés, respectivement PAA- $\alpha$ -D-Man (A.) et PAA- $\alpha$ -L-Fuc (B.) pour la fixation de DiscII-Nter ( $c=0,5 \mu\text{M}$ ), évaluée par RPS.**

Ces sensorgrammes sont issus de la soustraction des pistes contenant soit 100% de PAA-Man (A.), soit 100% de PAA-Fuc (B.) à la piste de référence présentant 100% de PAA-Gal, à priori non reconnue par DiscII-Nter.

### 3. Conclusion

Les discoidines sont des lectines formées de deux domaines distincts : le domaine C-terminal, homologue à la lectine HPA de l'escargot de Bourgogne et spécifique des résidus galactosylés et le domaine N-terminal, homologue à l'agglutinine d'anguille AAA.

Le but de ce travail était de vérifier si, aux vues des homologies structurales avec les lectines de type F, spécifique du fucose, ces domaines N-terminaux étaient capables de reconnaître des structures de sucres et notamment les résidus fucosides.

DiscI-Nter et DiscII-Nter ont donc été clonés de manière isolée et exprimés sous forme recombinante dans la bactérie *E. coli*. Malheureusement, dans les conditions testées, DiscI-Nter s'exprimait très majoritairement sous forme agrégée et aucune expérience complémentaire n'a alors pu être envisagée. En revanche, les conditions d'expression et de purification identiques à celles utilisées pour la protéine entière DiscII nous ont permis d'obtenir le domaine N-terminal de DiscII purifié et sous forme soluble. La comparaison des structures cristallographiques de DiscI et DiscII, démontrent que la Leu 154, dernier acide aminé du domaine N-terminal de DiscII correspond à la Gln 152 dans DiscI. Les tentatives d'expression du domaine N-terminal de cette dernière ont donc été réalisées avec une protéine contenant entre un et cinq résidus supplémentaires par rapport à DiscII-Nter, produite sous

forme soluble. Un nouveau clonage, permettant d'obtenir un gène codant pour la protéine DiscI-Nter-Q152 peut alors être envisagé.

La spécificité de DiscII-Nter a été étudiée par hémagglutination et RPS. Alors que les tests d'hémagglutination n'ont révélé aucune interaction entre ce domaine et les glycoépitopes situés à la surface des hématies, les expériences de résonance plasmonique de surface ont démontré que DiscII-Nter était capable d'interagir avec les résidus fucosides et mannosides immobilisés sur le biocapteur.

Ces expériences nous permettent de tirer plusieurs conclusions :

- le test d'hémagglutination est un moyen rapide de contrôler l'activité agglutinante d'une protéine. Néanmoins, face à un résultat négatif, il est dangereux de conclure trop rapidement à l'absence de propriétés agglutinantes de la protéine ou à une non-fonctionnalité. En effet, DiscII-Nter, qui est une lectine multivalente fonctionnelle spécifique du fucose, n'a interagit avec aucun glycoépitope présent à la surface des hématies. Aux vues de leur grande diversité de structures, il est peu envisageable de penser que ce résultat serait causé par une absence de ligand pour DiscII-Nter. Le résultat négatif du test d'hémagglutination pourrait s'expliquer par la faible affinité entre DiscII-Nter et les glycoépitopes sanguins.

- la RPS a démontré pour la première fois que le domaine N-terminal de DiscII était, comme DiscII-Cter, une lectine mais de spécificité différente. DiscII-Nter reconnaît davantage les résidus fucoside et mannoside, ce dernier avec une affinité moindre. Une optimisation des conditions expérimentales des manipulations de RPS permettrait de déterminer les constantes d'affinité de DiscII-Nter pour ces deux résidus.

DiscI-Nter étant homologue à DiscII-Nter et AAA, il reconnaît probablement également le fucose même si ceci n'a, à ce jour, pas pu être démontré expérimentalement. Les discoidines seraient donc des protéines composées de deux domaines lectines distincts, présentant tous deux des spécificités différentes.

D'autres lectines, procaryotes ou eucaryotes, possèdent également cette singulière propriété. Il a, par exemple, très récemment, été identifié une lectine, dans la bactérie pathogène *Burkholderia cenocepacia*, qui possèdent deux domaines de reconnaissance aux sucres (CRDs pour carbohydrate recognition domains). Le domaine C-terminal, spécifique du mannose et homologue aux lectines BC2L-A de *Burkholderia cenocepacia* et PA-IIL de *Pseudomonas aeruginosa*, serait utilisé pour cibler les glycoconjugués de l'hôte. Le domaine N-terminal est, quant à lui, spécifique des résidus fucosides (notamment les épitopes de groupes sanguins humains H et Lewis), et pourrait être utile à la bactérie pour adhérer aux épithélia particulièrement fucosylés des poumons (Sulak *et al.* 2010). Chez l'humain, la

galectine-9 possède également deux CRDs avec des spécificités différentes. L'étude des structures cristallographiques de ces deux domaines explique ces variations de spécificité et laisse supposer qu'elles seraient à l'origine de fonctions distinctes dans la cellule (Yoshida *et al.* 2010).

Des lectines possédant deux CRDs à spécificité différente existent également chez les invertébrés. Les crevettes *Fenneropenaeus merguensis* et *Fenneropenaeus chinensis* possède chacune une lectine de type C qui présente un premier domaine spécifique du galactose et un second spécifique du mannose. Ces lectines participeraient au système immunitaire de l'animal (Zhang *et al.* 2009 ; Rattanaporn and Utarabhand 2010). Il est donc facilement envisageable de considérer que les discoidines assurent le même rôle chez l'amibe.

Elles pourraient également être impliquées dans la phagocytose, processus utilisé par les souches sauvages de *Dictyostelium discoideum* pour se nourrir de bactéries (Gotthardt *et al.* 2006 ; Sillo *et al.* 2008) ou dans la formation des différents types de parois de l'animal au cours des différents stades du cycle de vie de l'animal, comme ceci a été brièvement détaillé dans le chapitre 4 (Cooper *et al.* 1983 ; Cooper and Barondes 1984 ; Erdos 1989 ; West *et al.* 2009).



# Chapitre 7

Expression de la lectine SLL-2 du  
corail *Sinularia lochmodes*



---

## Chapitre 7

### Expression de la lectine SLL-2 du corail

#### *Sinularia lochmodes*

---

Chez certains coraux, des lectines sont impliquées dans l'établissement d'une symbiose entre l'animal et des algues unicellulaires. Cette algue photosynthétique du genre *Symbiodinium* nourrit l'hôte et reçoit en retour une protection et les sources d'azote qui lui sont vitales. Cette association est indispensable pour la préservation des barrières de coraux, trop souvent endommagées par les phénomènes de blanchiment, perte de leur algue symbiotique entraînant la mort de l'animal (Koike *et al.* 2004 ; Wood-Charlson *et al.* 2006).

La lectine SLL-2, issue du corail *Sinularia lochmodes*, est, comme le domaine C-terminal des discoidines, similaire à HPA (25% de similarité), complétant ainsi la famille des lectines de type H (**Figure 7.1.**). Elle a été purifiée pour la première fois en 2000, par chromatographie d'affinité sur résine Sepharose 4B, à partir d'extrait de corail. Le poids moléculaire global de la protéine, dénommée SLL-2 était estimé à 122 kDa par chromatographie d'exclusion stérique. Des tests d'hémagglutination avec des extraits totaux de corail et de la protéine purifiée ont été réalisés. SLL-2 est capable d'agglutiner des érythrocytes de lapin et de cheval. Cette agglutination peut être inhibée par une pré-incubation de la lectine avec différents mono- et oligosaccharides, tels que le Gal, le GalNAc, le lactose ou le mélibiose (Gal $\alpha$ 1-6Glc). Ces expériences démontrent que SLL-2a est spécifique du monosaccharide D-galactose (Jimbo *et al.* 2000). Un séquençage effectué sur la protéine purifiée a permis de déterminer une partie de sa séquence peptidique et, grâce à la conception d'amorces dégénérées, d'isoler trois séquences d'ADN complémentaire, désignées par *sll-2a*, *-2b* et *-2c*, présentant 78% d'identité. La présence d'un site de N-glycosylation en position 60 a également été identifiée. La masse moléculaire de chaque sous-unité est d'environ 10,8 kDa pour la forme déglycosylée et de 13,2 kDa et 16,95 kDa pour les formes glycosylées majeures. Une hétérogénéité observées au niveau des valeurs des points isoélectriques et/ou

de la glycosylation, a été mise en évidence par des expériences de 2D-PAGE (Jimbo *et al.* 2005).

Le but de ce travail de thèse est de cloner cette lectine, de l'exprimer sous forme recombinante et de la purifier par des techniques de chromatographies classiques. Sa caractérisation biochimique ainsi que la résolution de sa structure cristallographique permettrait de la comparer avec les autres lectines de type H et d'envisager son utilisation en marquage histochimique.

```

DiscII 163 SVQVGEVSI.GDRSLNSGT.....GSRTIIVRHVKEPVEFLSVPIVSI GCKKVD AHTDNGQMRWEGKS 223
DiscI 158 VTQVGAIDIYTGDNALNTGS.....GKREVVPVKFQFEFATLPKVALNFDQIDCTDATNQTRIGVQP 220
HPA 1 RVQSGKIDC.GDDAGWAKVPSDDPGRDNTRELAKNITFASPYCRPPVLLSITQLDVEQSQ.NLRVIARL 68
SLL2a 1 RLIHVSRCEMGTSSHRCWP.....RECDTSSDEAISEFWPPFENTPKVIVSFGMLDVDNSH.NLRVNSA 63

DiscII 224 ENITTKGFDLTFITWGNAVYDLTFDYVAVEFNN 257
DiscI 221 RNIITTKGFDVCFYTWENKVYSLRADYIATALE. 253
HPA 69 YSVSPTGFKASCYTWHNTKVYSMSISWISIENY. 101
SLL2a 64 DDVTVVGFTLHYNSTWYTTIWNWYKLIWIACD... 94
    
```

**Figure 7.1. Alignement de séquences peptidiques des lectines de type H, HPA, DiscI et DiscII et SLL-2a.**

## 1. Clonage des gènes *sll-2*

Dans un premier temps, nous avons tenté d'isoler le gène *sll-2* à partir des ARNs totaux extraits d'un corail du genre *Sinularia*. Sylvie Tambutté de l'aquarium de Monaco nous a fourni l'échantillon d'ARN. La première étape consistait à isoler le gène en utilisant la technique 3'RACE commercialisée par Invitrogen. Le principe de la méthode et les amorces utilisées sont détaillés dans le chapitre 2. Une amplification d'ADN a été observée sur gel d'agarose avec le couple d'amorces SLL4/AU. Après purification, le produit amplifié est cloné dans le vecteur pGEM-T Easy (Promega) et transformé dans des bactéries compétentes. Un marquage blanc/bleu à l'IPTG et X-Gal permet d'isoler quatre clones à priori positifs. Après extraction et purification, l'ADN plasmidique est envoyé à séquencer. Les résultats du séquençage révèlent que la séquence d'ADN complémentaire isolé n'est pas homologue à celle du gène *sll-2*. L'amorce SLL4 est probablement tellement dégénérée qu'une séquence non spécifique a été amplifiée.

Cet échec peut notamment s'expliquer par le fait que les ARN totaux utilisés ne provenaient pas d'un corail appartenant à l'espèce *lochmodes*. En effet, il est visiblement très

difficile de différencier les différentes espèces du genre *Sinularia*. Nos collaborateurs pouvaient uniquement nous assurer que le corail appartenait à ce genre *Sinularia* mais pas forcément à l'espèce *lochmodes*.

Le gène *sll-2* code pour une protéine de 146 acides aminés. Des expériences de séquençage N-terminal de la protéine, purifiée à partir du corail, ont démontré qu'elle débutait au résidu 47 et était donc composée uniquement des 94 acides aminés correspondant au domaine agglutinine des lectines de la famille de type H (**Figure 7.2.**) (Jimbo *et al.* 2005).



**Figure 7.2. Alignement de séquences nucléotidiques entre les gènes synthétiques, *sll-2a* et *sll-2a-2* et le gène sauvage, issu de (Jimbo *et al.* 2005).**

Selon la publication précédente, le gène codant pour la protéine mature débute à la base nucléotidique notée 1. Les 138 bases précédentes codent pour une séquence signal hydrophobe de fonction inconnue. L'astérisque indique le codon stop.

L'étude de la séquence en amont de ce domaine semble indiquer la présence d'un peptide signal. La prédiction d'un site de coupure laisse supposer la conservation d'une séquence de 19 acides aminés, contenant une hélice  $\alpha$  comme élément de structure secondaire en N-terminal de la protéine mature. Ne connaissant ni l'importance ni l'influence de cette hélice sur la solubilité de la protéine, deux gènes, de longueurs différentes, ont alors été synthétisés chimiquement par la compagnie GenScript Corporation (VWR Strategic Partner)

(**Figure 7.2.**). Le premier, *sll-2a*, code, après traduction pour le domaine agglutinine des lectines de type H de 94 acides aminés. Le second, appelé *sll-2a-2*, code pour une protéine de 112 acides aminés qui contient l'hélice  $\alpha$  en N-terminal du domaine lectine. Les séquences de ces deux gènes ont été optimisées de manière à éliminer, après traduction, tous les codons rares chez *E. coli*.

## 2. Expression dans la bactérie *E. coli*

De nombreuses conditions expérimentales ont été testées afin d'optimiser l'expression sous forme soluble de SLL-2a (**Tableau 7.1.**) et SLL-2a-2 (**Tableau 7.2.**) dans la bactérie. A la suite de chaque expérience, nous avons analysé les extraits totaux avant et après induction à l'IPTG, lysé les cellules et recherché la présence de la protéine dans les corps d'inclusion (situés dans le culot après centrifugation à 12000 g) et sous forme soluble (surnageant des échantillons centrifugés à 50000 g).

Toutes les souches d'*E. coli* utilisées contiennent le prophage  $\lambda$ DE3 dont l'un des gènes code pour l'ARN polymérase du bactériophage T7. Ce gène est sous le contrôle du promoteur *lacUV5* inductible à l'IPTG. Ainsi, lorsque ces souches sont transformées par un vecteur d'expression contenant un promoteur fort, la synthèse d'ARNm est accrue. La production de la protéine d'intérêt peut alors atteindre jusqu'à 50% des protéines totales.

Les caractéristiques des différentes souches utilisées sont détaillées dans le chapitre 2. Les résultats obtenus pour chaque tentative d'expression sont très succinctement repris dans les **Tableaux 7.1** et **7.2**. Je ne détaillerai que quelques conditions expérimentales.

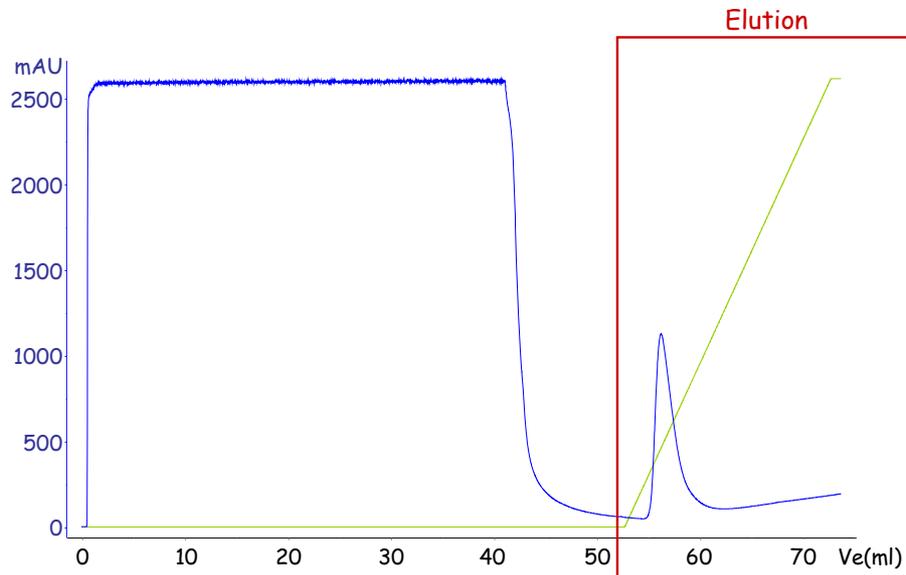
Les premières tentatives d'expression ont été effectuées avec la protéine SLL-2a clonée dans le vecteur pRSET-A. Produite sans étiquette His<sub>6</sub>, la stratégie de purification envisagée était d'utiliser une chromatographie d'affinité de type Sepharose 4B, de manière identique à ce qui avait été effectué avec la protéine sauvage (Jimbo *et al.* 2000). Les résultats d'expression ont été analysés sur gel SDS-PAGE 15%. Une bande apparaît clairement autour de 10 kDa, poids attendu pour le monomère de SLL-2a, dans la fraction contenant les corps d'inclusion séparés par centrifugation à 12000 g. Quelque soit les conditions d'expression testées (température, concentration en IPTG, densité optique, temps d'induction, souche utilisée), aucune différence n'est constatée ni dans la quantité de protéine exprimée ni dans sa

proportion sous forme soluble. L'ajout du plasmide pLysS dans la souche BL21 (DE3) n'améliore pas l'expression de la protéine sous forme soluble.

SLL-2a							
Vecteur	Caractéristique	Souche	Tempé- -rature (°C)	Temps (h)	[IPTG] (mM)	DO <sub>600nm</sub>	Résultats
pRSET -A	Aucune étiquette	BL21 (DE3)	30	3	0,1 ;	0,6 et 1,5	Expression sous forme de corps d'inclusion
			16	16	0,5 et 2		
		BL21 (DE3) pLysS	16	16	0,1 ; 0,5 et 2	0,6 et 1,2	
pProEx -HtB	Etiquette His <sub>6</sub> en N-terminal clivable à la TEV	BL21 (DE3)	16	16	0,5	1	Aucune expression visible. Une purification sur colonne de nickel révèle qu'aucune protéine d'intérêt n'est éluee après ajout d'imidazole.
					0,2	0,8	
		Origami (DE3) K12	16	16	0,1 ; 0,5 et 2	0,6 et 1,2	
		Origami B (DE3)	16	16	0,05	0,6	
		BL21 (DE3) C41	16	16	0,2	1	
		BL21 (DE3) C43	16	16	0,2	1	
pET -TEV	Etiquette His <sub>6</sub> en N-terminal clivable à la TEV	BL21 (DE3)	16	16	0,2 et 1	0,8	Corps d'inclusion. Testée en immunomarquage.
							BL21 (DE3) pLysS
		28	4	0,2	0,8	Corps d'inclusion	
		Rosetta gami 2 (DE3) pLysS	16	16	0,2	0,6	Rien. Testée en immunomarquage
							28
		Origami 2 (DE3) K12	25	5	0,2	0,7	Rien

**Tableau 7.1. Conditions d'expression testées pour l'expression de SLL-2a dans la bactérie *E. coli*.**

Le gène *sll-2a* a alors été cloné dans le vecteur pProEx-HtB, sous l'action du promoteur fort *trc* (Brosius *et al.* 1985). La protéine SLL-2a est exprimée avec une étiquette His<sub>6</sub> du côté N-terminal. Le dépôt sur gel SDS-PAGE 15% des fractions avant et après induction ainsi que des fractions contenant les corps d'inclusion et les protéines solubles (surnageant après centrifugation à 50000 g) ne révèle aucune expression. Afin de contrôler ce résultat, toutes les fractions solubles ont été purifiées par chromatographie sur colonne de nickel. L'élution est effectuée par un gradient de 0 mM à 500 mM d'imidazole. Pour toutes les souches testées, un pic, élué à environ 100 mM, est visible (**Figure 7.3.**). Il correspond aux protéines contaminantes d'*E. coli*. Ce résultat est également retrouvé avec la protéine SLL-2a-2.



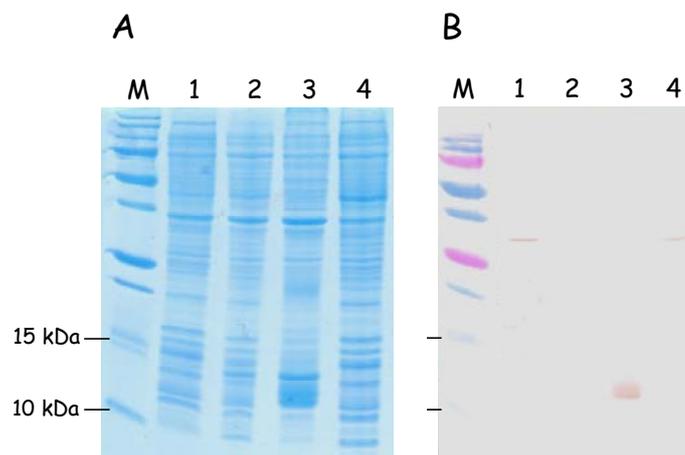
**Figure 7.3. Purification de SLL-2a clonée dans le vecteur pProEx-HtB et exprimée dans la souche BL21 (DE3).**

Les courbes bleue et verte représentent respectivement la densité optique à 280nm (exprimée en mAU) et le gradient d'imidazole appliqué pour éluer les protéines. L'encadre rouge représente les fractions éluées.

Les séquences nucléotidiques étant optimisées pour éviter la présence de codons rares chez *E. coli*, l'absence d'expression ne peut pas être due à un problème de disponibilité des ARNt faiblement présents dans la bactérie. De plus, l'expression de SLL-2a-2 dans la souche Rosetta gami 2 pLysS, qui est capable d'exprimer les codons rares, est aussi nulle. Le vecteur pProEx-HtB, comme tous les vecteurs dérivés du système pET est un vecteur à faible taux de

réplication. Cette caractéristique, couplée à une température d'expression basse, peut expliquer l'absence d'expression de ces protéines.

Pour tester un promoteur plus faible, le gène a été inséré dans le vecteur pET-TEV. SLL-2a exprimée possède une étiquette His<sub>6</sub> en N-terminal. Suivant les conditions et souches testées, elle est soit présente sous forme de corps d'inclusion (**Figure 7.4. A.**) soit absente. Les tests d'expression réalisés à 16°C toute une nuit conduisent également à une absence d'expression (excepté dans BL21 (DE3)), alors que, à une température plus élevée, une expression est clairement visible. Ce résultat n'est pas vérifié pour l'expression dans la souche Origami 2 (DE3) pour qui l'expression est nulle à 25°C mais la croissance bactérienne observée, très faible pour cette souche, peut être à l'origine de ce résultat. Un immunomarquage, avec un anticorps dirigé contre la queue histidine (pentahis), prouve que la protéine présente dans la fraction 12C est bien la protéine d'intérêt (**Figure 7.4. B.**).



**Figure 7.4. A. Gel SDS-PAGE 15 % acrylamide coloré au bleu de Coomassie après expression de SLL-2a clonée dans le vecteur pET-TEV et exprimée dans les BL21 (DE3), B. Immunomarquage correspondant.**

Les pistes 1 et 2 correspondent aux extraits totaux de la culture, respectivement avant et après induction. La piste 3 représente la fraction contenant la protéine présente sous forme de corps d'inclusion (après centrifugation à 12000 g). La piste 4 représente la fraction soluble (après centrifugation à 50000 g) qui, lors d'une purification est déposée sur colonne.

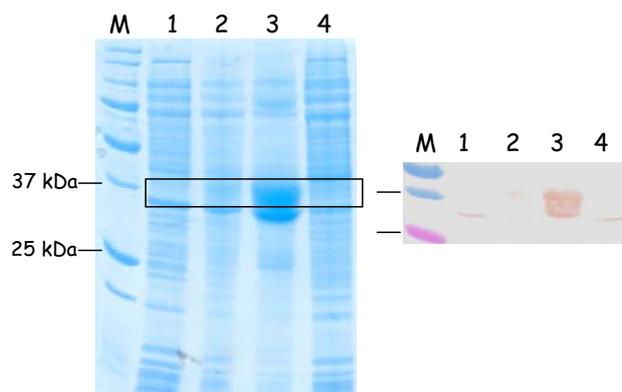
Les tests d'expression réalisés avec la protéine SLL-2a-2 (**Tableau 7.2.**) sont similaires à ceux observés pour SLL-2a.

SLL-2a-2							
Vecteur	Caractéristique	Souche	Température (°C)	Temps (h)	[IPTG] (mM)	DO <sub>600nm</sub>	Résultats
ProEx -HtB	Étiquette His <sub>6</sub> en N-terminal clivable à la TEV	BL21 (DE3)	16	16	0,2	0,8	Aucune expression visible sur gel et par immunomarquage. Une purification sur colonne de nickel révèle qu'aucune protéine d'intérêt n'est éluee après ajout d'imidazole.
		BL21 (DE3) pLysS					
		Origami (DE3) K12					
		BL21 (DE3) C41					
		BL21 (DE3) C43					
		BL21 (DE3) Star pLysS	16	16	0,2	0,6	
		Rosetta gami 2 pLysS					
pET -M30	Étiquettes His <sub>6</sub> et GST en N-terminal clivables à la TEV	BL21 (DE3)	37	3	0,2 et 1	0,7	Corps d'inclusion
			16	16		0,8	
		BL21 (DE3) pLysS	16	16	0,2	0,6	Rien
			28	4		0,8	Corps d'inclusion
		Rosetta gami 2 (DE3) pLysS	16	16	0,2	0,6	Rien
			28	4		0,8	Corps d'inclusion
pRSET -A	Sans et avec étiquette His <sub>6</sub> en N-terminal clivables à l'entérokinase	Origami (DE3) K12	16	16	0,2	0,9	Aucune expression visible ou uniquement sous forme de corps d'inclusion.

**Tableau 7.2. Conditions d'expression testées pour l'expression de SLL-2a-2 dans la bactérie *E. coli*.**

SLL-2a-2, exprimée à partir du vecteur pRSET-A, est localisée dans les corps d'inclusion et apparaît sous la forme d'une bande à 15 kDa sur gel SDS-PAGE 15%. Le vecteur PRSET-A, à la différence des vecteurs dérivés du système pET, est un vecteur avec

un taux de réplication élevé, ce qui explique la détection d'une expression dans les souches Origami (DE3) et Origami (DE3) B, même une nuit à 16°C. Le vecteur pET-M30 permet à la protéine exprimée d'être couplée, en N-terminal, à une étiquette volumineuse, la protéine glutathion-S-transférase (211 acides aminés). L'intérêt majeur de cette fusion est l'expression d'une protéine soluble et stable, produite en grande quantité et facile à purifier grâce à son étiquette His<sub>6</sub> (Smith 2000). Malheureusement, cette stratégie ne s'est pas avérée payante puisque, quelque soit les conditions d'expression testées, la protéine SLL-2a-2/GST est soit exprimée sous forme de corps d'inclusion, soit pas exprimée (**Figure 7.5.**). Les résultats sont identiques à ceux obtenus avec les constructions de SLL-2a dans le vecteur pET-TEV : excepté avec la souche BL21 (DE3), les tests d'expression à 16°C toute une nuit ne permettent pas de détecter la protéine. Seule une augmentation de la température d'expression permet une expression visible.



**Figure 7.5. Gel SDS-PAGE 15 % acrylamide coloré au bleu de Coomassie et immunomarquage après expression de SLL-2a-2/GST.**

Les pistes 1 et 2 correspondent aux extraits totaux de la culture, respectivement avant et après induction. La piste 3 représente la fraction contenant la protéine présente sous forme de corps d'inclusion (après centrifugation à 12000 g). La piste 4 représente la fraction soluble (après centrifugation à 50000 g) qui, lors d'une purification est déposée sur colonne.

L'augmedium™ est un additif commercialisé par AthenaES à ajouter dans le milieu de culture avant l'induction des protéines. Spécialement créée pour les protéines recombinantes exprimées dans les corps d'inclusion, il permet d'induire l'expression de différentes chaperones, favorisant ainsi l'expression, sous forme soluble, des protéines d'intérêt. Malheureusement, la proportion de protéines SLL-2a et SLL-2a-2 solubles n'a pas été améliorée, quelque soit la concentration en augmedium™ utilisée. Les protéines restent intégralement produites sous forme de corps d'inclusion.

Les problèmes de repliement protéique, à l'origine de la formation des corps d'inclusion chez *E. coli*, sont provoqués par deux phénomènes principaux :

- un encombrement important dans le cytoplasme où la concentration en macromolécules peut atteindre jusqu'à 400 mg.ml<sup>-1</sup> (Ellis and Minton 2003)
- une mauvaise formation des ponts disulfure dans le cytoplasme (Kadokura *et al.* 2003). En effet, le cytoplasme bactérien est un compartiment réducteur qui limite fortement la formation des ponts disulfure.

Les séquences peptidiques de SLL-2a et SLL-2a-2 contiennent quatre résidus cystéines (en position 8, 17, 22 et 93), qui peuvent être impliqués dans la formation de ponts disulfure intramoléculaires. Différents serveurs ont été mis au point pour tenter de prédire les liaisons éventuelles entre deux résidus. Les serveurs DISULFIND (Ceroni *et al.* 2006) et DiANNA (Ferre and Clote 2005 ; Ferre and Clote 2005), basés respectivement sur l'analyse de la séquence primaire uniquement et sur les prédictions de structures secondaires, indiquent que deux ponts disulfure seraient formés dans SLL-2a et SLL-2a-2 : le premier établi entre les deux premières cystéines et le second avec les deux dernières. Très récemment, un nouveau serveur, appelé DBCP (Lin and Tseng 2010), utilise, quant à lui, les homologues de structures tertiaires pour prévoir l'établissement de ponts disulfure. Le programme MODELLER prédit les coordonnées X, Y et Z des carbones alpha de chaque résidu de la protéine puis le serveur se charge de calculer les distances séparant chaque paire de cystéines pouvant être formées. Selon DBCP, un seul pont disulfure établi entre les cystéines 17 et 22 serait formé dans la protéine SLL-2a. La probabilité calculée est de 90%. La proximité de deux cystéines dans la séquence peptidique d'une protéine est une caractéristique fréquemment rencontrée dans les protéines comportant des ponts disulfure.

Les souches dites Origami (DE3), mutées au niveau des gènes *gor* et *trxB* codant pour deux enzymes à l'origine du caractère réducteur du cytoplasme, sont donc spécialement conçues pour favoriser la formation des ponts disulfure. Mais leur utilisation n'a pas permis d'obtenir SLL-2a et SLL-2a-2 sous forme soluble.

Dans la littérature, diverses stratégies ont été développées pour favoriser la formation de cette liaison disulfure lors de la synthèse d'une protéine. La première consiste à cloner la protéine d'intérêt de manière à la fusionner, du côté N-terminal, avec la thioredoxine 1, produit du gène *trxA* (LaVallie *et al.* 1993). Cette enzyme compacte, très soluble et stable améliore la solubilité d'un nombre important de protéines synthétisées et favorise la formation des ponts disulfure dans le cytoplasme des souches mutantes *trxB* (Stewart *et al.* 1998). Une grande variété de protéines contenant des ponts disulfure ont alors été exprimées grâce à cette

méthodologie (Jurado *et al.* 2006). Une autre stratégie pour une expression dans le cytoplasme bactérien consiste à introduire la sulfhydryl oxidase Erv1p, enzyme qui utilise l'oxygène moléculaire pour catalyser l'oxydation de dithiols en ponts disulfure, dans le cytoplasme d'*E. coli* (Hatahet *et al.* 2010).

A l'inverse, le périplasma bactérien est un compartiment de choix pour l'expression de protéines comportant des ponts disulfure. Bien plus oxydant que le cytoplasme, il abrite un réseau d'enzymes redox qui catalysent la formation et l'isomérisation des ponts disulfure, telles que les membres du système Dsb, appartenant à la superfamille des thiorédoxines (de Marco 2009). En exprimant une protéine fusionnée avec MalE, protéine appartenant à un transporteur ABC du maltose dans le périplasma, la protéine d'intérêt sera transportée dans ce compartiment cellulaire (Nikaido 1994).

Une autre stratégie consiste à tenter de solubiliser et renaturer les corps d'inclusion obtenus de manière à permettre à la protéine d'intérêt d'adopter un repliement qui lui permet d'assurer sa fonction. C'est cette dernière méthode qui nous avons choisi d'exploiter.

### **3. Renaturation des corps d'inclusion**

La protéine exprimée sous forme de corps d'inclusion dans *E. coli* peut être solubilisée et renaturée *in vitro*, soit par dilution dans un bécher soit directement par chromatographie sur colonne. Les corps d'inclusion sont le résultat d'un déséquilibre entre les réactions de synthèse protéique, de repliement et d'agrégation. Leur formation est provoquée par des interactions hydrophobes relativement stables. La première étape consiste donc à briser ces interactions pour solubiliser les protéines.

#### **3.1. Solubilisation des corps d'inclusion**

Le tampon de solubilisation est généralement composé d'agents dénaturants tels que l'hydrochlorure de guanidinium (GdnCl) ou l'urée, d'agents réducteurs (DTT,  $\beta$ -mercaptoéthanol, couple glutathion réduit/glutathion oxydé (GSH/GSSG)), d'agents chélatants (EDTA), de sels (NaCl, KCl, CaCl<sub>2</sub>, MgCl<sub>2</sub>...) et d'autres composés qui peuvent

stabiliser la protéine et favoriser sa solubilisation (glycérol, PEG, saccharose, détergents...) ou sa purification. Le pH du tampon est également très important (Burgess 2009 ; Fahnert *et al.* 2004).

Des premiers essais de solubilisation des corps d'inclusion de SLL-2a et SLL-2a-2 clonées respectivement dans les vecteurs pET-TEV et pET-ME30 et exprimées dans la souche BL21 (DE3) sont résumés dans le **Tableau 7.3**.

Tampon de solubilisation	Temps d'incubation	Plasmide
Tris-HCl 20 mM, pH 8, NaCl 500 mM, Imidazole 10 mM, Urée 6 M, $\beta$ -mercaptoéthanol ( $\beta$ -ME) 10 mM, glycérol 20%	30 min	pET-M30-SLL-2a-2/GST
	Une nuit	pET-TEV-SLL-2a
Tris-HCl 20 mM, pH 8,5, NaCl 500 mM, Imidazole 10 mM, Urée 6 M, $\beta$ -ME 10 mM	2 heures	pET-M30-SLL-2a-2/GST

**Tableau 7.3. Récapitulatif des conditions de solubilisation des corps d'inclusion.**

La présence d'imidazole s'explique par le fait que la protéine solubilisée devait directement être purifiée par chromatographie sur colonne de nickel et éventuellement renaturée sur colonne. Quelque soit la composition du tampon et le temps d'incubation, la grande majorité des protéines n'est pas solubilisée et s'organise toujours sous forme de corps d'inclusion. Des conditions de solubilisation plus optimales ont été recherchées.

Les essais suivants ont été réalisés avec la protéine SLL-2a clonée dans le vecteur pET-TEV et exprimée dans la souche BL21 (DE3). En effet, le volume important que représente la GST fusionnée à SLL-2a-2 peut éventuellement gêner le bon repliement de la protéine *in vitro*. J'ai préféré travailler sur une protéine courte. Après induction, les cellules sont éclatées et les corps d'inclusion protéiques sont lavés, par cycles de centrifugation successifs, trois à cinq fois, dans le tampon de lyse composé de Tris-HCl 20 mM, pH 8, saccharose 20%, EDTA 5 mM, Triton X-100 1%. L'étape de purification se révèle fondamentale puisque des impuretés associées à la protéine dénaturée peuvent entraver son bon repliement tridimensionnel (Tsumoto *et al.* 2003). Les conditions de solubilisation alors testées sont résumées dans le **Tableau 7.4**.

Composition des différents tampons	Concentration en urée
Acétate de sodium-acide acétique 20 mM, pH 4, EDTA 2 mM, glutathion réduit (GSH) 1 mM, glutathion oxydé (GSSG) 0,1 mM	0 M urée
	2 M urée
	8 M urée
MES 20 mM, pH 6, EDTA 2 mM, GSH 1 mM, GSSG 0,1 mM	0 M urée
	2 M urée
	8 M urée
Tris-HCl 20 mM, pH 8, EDTA 2 mM, GSH 1 mM, GSSG 0,1 mM	0 M urée
	2 M urée
	4 M urée
	6 M urée
	8 M urée
Tris-HCl 20 mM, pH 9, EDTA 2 mM, GSH 1 mM, GSSG 0,1 mM	2 M urée
	4 M urée
	6 M urée
	8 M urée
Tris-HCl 20 mM, pH 10, EDTA 2 mM, GSH 1 mM, GSSG 0,1 mM	0 M urée
	2 M urée
	4 M urée
	6 M urée
	8 M urée

**Tableau 7.4. Récapitulatif des conditions de solubilisation des corps d'inclusion de SLL-2a, clonée dans le vecteur pET-TEV et exprimée dans la souche BL21 (DE3).**

L'analyse est effectuée sur gel SDS-PAGE 15% après deux heures ou une nuit d'incubation dans le tampon de dénaturation. Les deux principales différences dans la composition de ces tampons sont le pH et la concentration en urée. Après analyse sur gel, les conditions retenues pour la solubilisation des corps d'inclusion sont un tampon à pH 9, pH proche de celui des tampons utilisés pour purifier les protéines sur colonne de nickel (pH 8,5) et éloigné du point isoélectrique estimé de SLL-2a ( $pI = 5,7$ ) où la solubilité protéique est minimale et une concentration en urée de 6 M. Deux heures d'incubation sont suffisantes pour solubiliser la majorité des corps d'inclusion dans les conditions décrites ci-dessus.

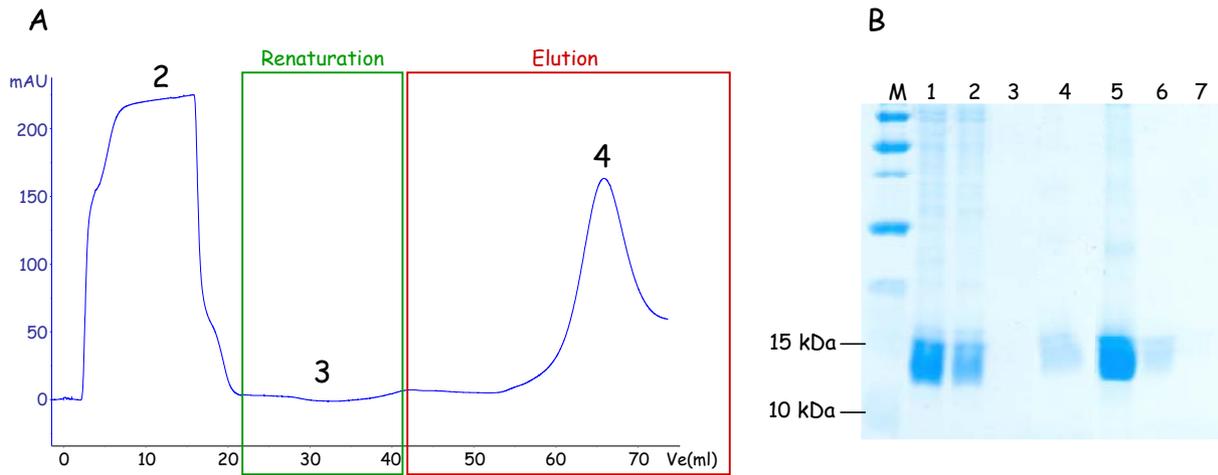
### 3.2. Renaturation des corps d'inclusion solubilisés

Après solubilisation des protéines agrégées, l'étape de renaturation consiste à diminuer la concentration d'agents dénaturants de manière à favoriser le repliement de la protéine. Plusieurs méthodes sont disponibles et peuvent être divisées en deux groupes :

- des méthodes non chromatographiques telles que la dialyse ou la dilution.
- des méthodes chromatographiques dont la chromatographie d'exclusion stérique, la renaturation assistée par des chaperones ou la purification sur colonne (Swietnicki 2006).

La réussite de la renaturation est très empirique et dépend de la protéine. La méthode choisie ainsi que la composition des tampons de solubilisation et de renaturation sont primordiales mais restent très aléatoires. Certains composés qui entrent dans la composition des tampons possèdent des propriétés spécifiques. Le saccharose et les polyols améliorent la stabilité des protéines grâce à l'augmentation de la viscosité du milieu alors que la L-arginine, à la différence de son énantiomère de la série D, possède un effet solubilisant (Hamada *et al.* 2009). Les mécanismes mis en jeu dans cette étape de renaturation sont rarement connus. Il s'agit de diminuer progressivement la concentration en agents dénaturants de manière à favoriser la formation d'un intermédiaire qui commence à se renaturer alors que sa flexibilité et sa solubilité restent importantes (Tsumoto *et al.* 2003).

La première méthode utilisée a été la renaturation par chromatographie sur colonne de nickel (**Figure 7.6. A.**) (Li *et al.* 2004). Après solubilisation de SLL-2a dans le tampon de dénaturation (Tris-HCl 20 mM, pH 9, EDTA 2 mM, GSH 1mM, GSSG 0,1 mM, Urée 6M), la protéine dénaturée (à 0,7 mg.ml<sup>-1</sup>) s'accroche, via son étiquette polyhistidine, sur les ions nickel de la colonne. Ce lien, très stable permet d'utiliser des tampons avec une large gamme de pH contenant les dénaturants GdnCl ou urée présents à forte concentration, des détergents non ioniques (Triton X-100), des sels (NaCl), des agents réducteurs (DTT, couple GSH/GSSG...) et des petites molécules telles que des acides aminés, des polyols, des agents chélatants à faible concentration (Swietnicki 2006). SLL-2a est ensuite renaturée progressivement avec le remplacement du tampon de dénaturation par celui de renaturation (Tris-HCl 20 mM, pH 8, NaCl 500 mM, Imidazole 10 mM,  $\beta$ -ME 5 mM, glycérol 10%) et éluée par un gradient d'imidazole de 0 à 500 mM. Les résultats de cette chromatographie sont analysés sur gel SDS-PAGE 15% (**Figure 7.6. B.**).



**Figure 7.6. A. Renaturation des corps d'inclusion de SLL-2a par chromatographie sur colonne de nickel. B. Gel SDS-PAGE 15 % acrylamide coloré au bleu de Coomassie après expression de SLL-2a clonée dans le vecteur pET-TEV.**

**A.** La courbe bleue représente la densité optique à 280nm (exprimée en mAU). Les encadrés vert et rouge représentent respectivement les étapes de renaturation et d'élué. **B.** La piste 1 correspond au surnageant après lavage et solubilisation des corps d'inclusions. Les pistes 2, 3 et 4 représentent respectivement la fraction non retenue, la fraction durant la renaturation et la fraction éluée. Les échantillons des pistes 5, 6 et 7 sont issus de la purification d'affinité suivante sur Sepharose 4B. Ils correspondent respectivement à la fraction non retenue, le lavage et la fraction éluée avec 0,3 M de galactose.

Le pic du chromatogramme correspondant à la fraction non retenue (pic 2) est formé par SLL-2a qui n'a pas été retenue par les ions nickel de la colonne (piste 2) alors que sa capacité maximale n'était pas atteinte. Le tampon contient 2 mM d'EDTA alors que la concentration maximale préconisée est de 1 mM. Mais, ce phénomène a également été observé pour les autres tests de renaturation sur colonne dans lesquels le tampon ne contenait pas d'agents chélatants. La taille de la colonne ne modifie pas non plus ce résultat. La repurification de la fraction non retenue lors d'une deuxième étape ne permet toujours pas de retenir l'intégralité de SLL-2a. Un équilibre semble s'installer entre la quantité de protéines accrochées sur la colonne et celle non retenue. Aucune recherche bibliographique n'a permis d'expliquer ce résultat. Après renaturation pendant 3 heures sous un flux de  $0,1 \text{ ml} \cdot \text{min}^{-1}$ , la protéine est éluée à environ 350 mM (pic 4 A. et piste 4 B.). Aucun pic n'est visible pendant l'étape de renaturation ; la protéine n'est pas décrochée (encadré vert A. et piste 3 B.). Après élué et contrôle de la présence de SLL-2a sous forme soluble, sa fonctionnalité est testée par chromatographie d'affinité sur Sepharose 4B. Des fractions de 1 ml sont collectées manuellement et analysées sur gel. La protéine est bien présente sous forme soluble, dans le

surnageant de la centrifugation à 50000g mais n'est pas retenue sur la colonne (piste 5 B.). De plus, aucune protéine n'est éluée après l'ajout de 0,3 M de galactose (piste 7 B.). La protéine présente visiblement un repliement tridimensionnel qui ne lui permet pas de retrouver une activité agglutinante. Les conditions de renaturation ne sont pas adaptées.

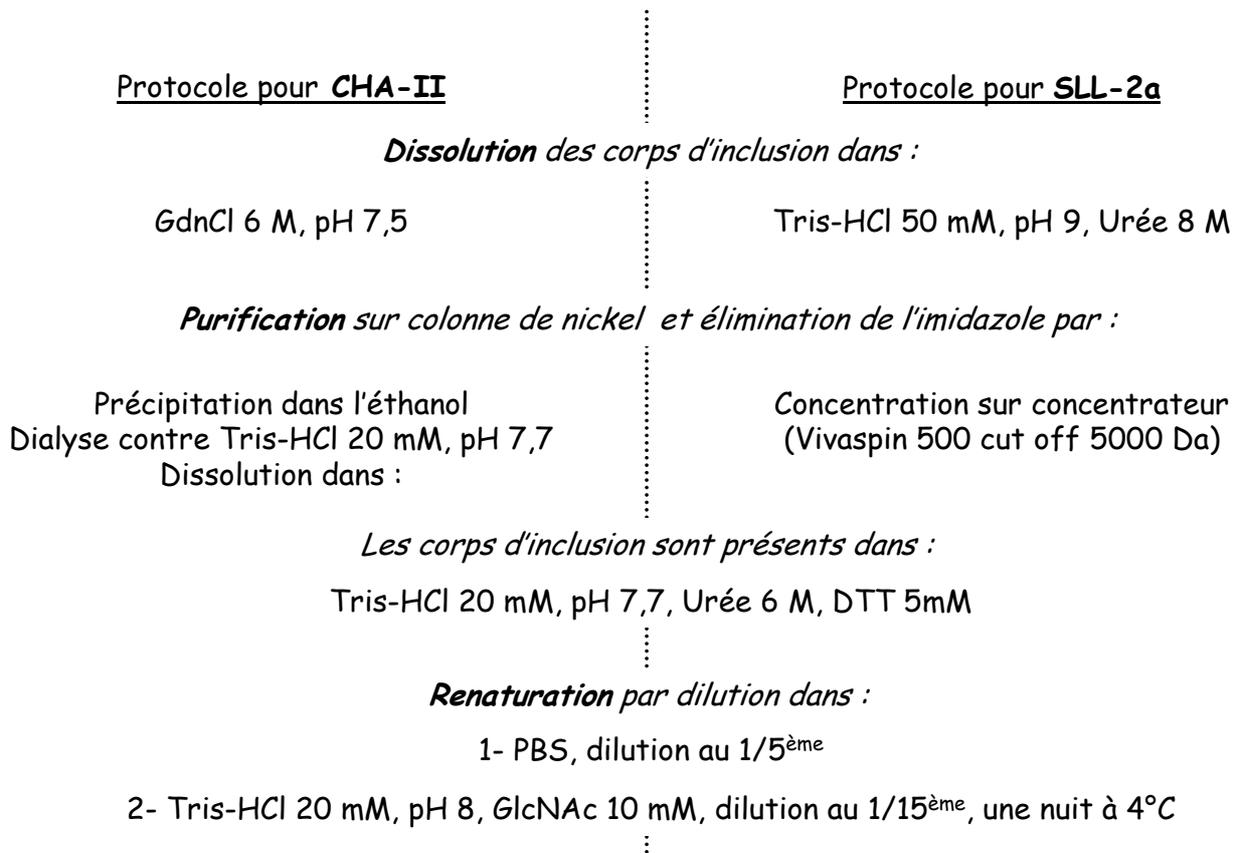
La renaturation par dilution lente a également été testée. Le tampon de renaturation (Tris-HCl 20 mM, pH 7,5, NaCl 150 mM, GlcNAc 10 mM, L-arginine 0,5 M, glycérol 10%, GSH 5 mM, GSSG 0,5 mM) est ajouté au goutte à goutte dans la solution protéique (à 0,5 mg.ml<sup>-1</sup>) de manière à la diluer d'un facteur 10. Une colonne de Sepharose 4B est également utilisée pour contrôler la fonctionnalité de SLL-2a renaturée. La présence d'un substrat potentiel de la lectine ne lui permet pas d'adopter la bonne structure tertiaire et/ou quaternaire. Après analyse sur gel SDS-PAGE 15%, nous constatons que l'intégralité de la protéine renaturée est soluble mais localisée dans la fraction non retenue (résultats non présentés).

Le kit commercialisé par Athena ES a également testé pour la renaturation de SLL-2a. Il est formé de quinze tampons de compositions différentes (détaillées dans le chapitre 2, Tableau 2.10) à utiliser pour diluer la protéine d'intérêt d'un facteur 20 (50 µl de SLL-2a à 1 mg.ml<sup>-1</sup> diluée avec 950 µl de tampon). Après renaturation, la protéine est mise à incuber en présence de résine Sepharose 4B afin de tester sa fonctionnalité. Les échantillons sont ensuite analysés sur gel SDS-PAGE 15% après précipitation au TCA 100%. Le but des kits commerciaux est d'identifier les composés qui favorisent ou, au contraire, desservent la renaturation de la protéine. Quelque soit les conditions testées, SLL-2a est soit intégralement localisée dans le culot, c'est-à-dire, qu'après solubilisation, elle a de nouveau précipité, soit présente dans la fraction non retenue, c'est-à-dire que les conditions de renaturation n'ont pas permis d'obtenir une protéine fonctionnelle. L'analyse des résultats ne permet pas de conclure quant à la présence favorable ou non de certains composés dans le tampon. Une bande semble néanmoins être visible sur gel SDS-PAGE dans la fraction éluée après renaturation dans le tampon 10 (Tris-HCl 50 mM pH 8,5, NaCl 9,6 mM, KCl 0,4 mM, MgCl<sub>2</sub> 2 mM, CaCl<sub>2</sub> 2 mM, L-arginine 0,5 M, saccharose 0,4 M, Guanidine-HCl 0,75 M, GSH 1 mM, GSSG 0,1 mM). Un test similaire avec un tampon de même composition préparée par mes soins ne permet pas de confirmer ce résultat. Aucune protéine renaturée et fonctionnelle n'est présente.

La lectine CHA-II de l'escargot *Cepaea hortensis*, communément appelé « petit gris », est homologue à HPA et spécifique du GalNAc/GlcNAc. Exprimée dans la bactérie *E. coli*, elle est renaturée selon le protocole décrit dans la **Figure 7.7** (Gerlach *et al.* 2005). Un protocole identique a été appliqué à la lectine SLL-2a (**Figure 7.7.**) mais la purification sur colonne Sepharose 4B indique que la protéine, présente dans la fraction non retenue, n'est pas

fonctionnelle. D'autres essais avec, notamment, un tampon avant renaturation ne contenant pas de DTT et du GalNAc à la place du GlcNAc ont également été testés mais les résultats restent inchangés.

La séquence peptidique de CHA-II contient deux cystéines aux positions 12 et 62. Selon le serveur de prédiction de formation de ponts disulfure DBCP (Lin and Tseng 2010), ces résidus ne seraient pas impliqués dans un lien covalent soufre-soufre. L'échec de la renaturation de SLL-2a avec un protocole très similaire à celui employé pour CHA-II est probablement dû à un problème dans la formation des ponts disulfure.



**Figure 7.7. Protocoles de renaturation utilisés pour les lectines CHA-II et SLL-2a.**

Une glycosyltransférase étudiée au laboratoire nécessite également d'une étape de renaturation pour restaurer sa fonction. Son tampon de renaturation, composé principalement de L-arginine et du couple cystéamine/cystamine comme couple rédox ne permet pas à SLL-2a d'adopter un bon repliement tridimensionnel.

La renaturation de protéines exprimées sous forme de corps d'inclusion dans *E. coli* est devenue une méthode de choix pour l'obtention de protéine recombinante. Des protéines

très diverses telles que des lectines, des anticorps, des facteurs de croissance issues d'organismes variés (virus, invertébrés, humains...) ont été obtenues avec cette méthode (Amorim *et al.* 2010 ; Lee *et al.* 2010 ; Gerlach *et al.* 2005). Néanmoins, la recherche des conditions de solubilisation et de renaturation ainsi que la méthode expérimentale à utiliser peut s'avérer être un travail très laborieux. Nous avons alors tenté d'exprimer cette lectine de corail dans un système d'expression eucaryote, la levure *Pichia pastoris*.

## 4. Expression dans la levure *Pichia pastoris*

La levure, système d'expression eucaryote, est une bonne alternative pour les protéines dont l'expression dans la bactérie conduisait à la formation de corps d'inclusion (Daly and Hearn 2005 ; Cregg *et al.* 2009). La levure la plus fréquemment utilisée pour l'expression de protéines recombinantes est appelée *Pichia pastoris*. Cette levure méthylotrophique est capable d'utiliser le méthanol comme seule source de carbone, substrat qu'elle métabolise grâce à l'enzyme alcool oxidase (AOX). Deux gènes *AOX1* et *AOX2* codent pour cette enzyme dans *P. pastoris*. *AOX1* est responsable de la majorité de l'utilisation du méthanol (85%) alors que *AOX2*, son homologue à plus de 97%, consomme le méthanol de manière beaucoup plus lente et donc moins efficace. De plus, *AOX1* est exprimée de manière très abondante dans la levure alors que l'expression du gène *AOX2* est beaucoup faible. Les souches contenant le gène *AOX1*, dont GS 115 utilisée dans cette étude, sont désignées comme Mut<sup>+</sup> alors que les souches ne contenant qu'une copie fonctionnelle du gène *AOX2* possèdent un phénotype Mut<sup>S</sup> (Daly and Hearn 2005).

### 4.1. Clonage dans les vecteurs pPIC-Z $\alpha$

Les vecteurs utilisés pour le clonage sont pPICZ $\alpha$ -A et pPICZ $\alpha$ -HisFlag permettant l'expression d'une protéine sans étiquette et avec une queue His<sub>6</sub> et un épitope Flag respectivement. Ces vecteurs, résistants à la zéocine, contiennent une séquence signal alpha, en amont du site de clonage conduisant à la sécrétion, dans le milieu de culture, de la protéine

synthétisée. Les deux gènes *sll-2a* et *sll-2a-2* ont été clonés dans les deux vecteurs. Le schéma des deux constructions est donné dans la **Figure 7.8**.



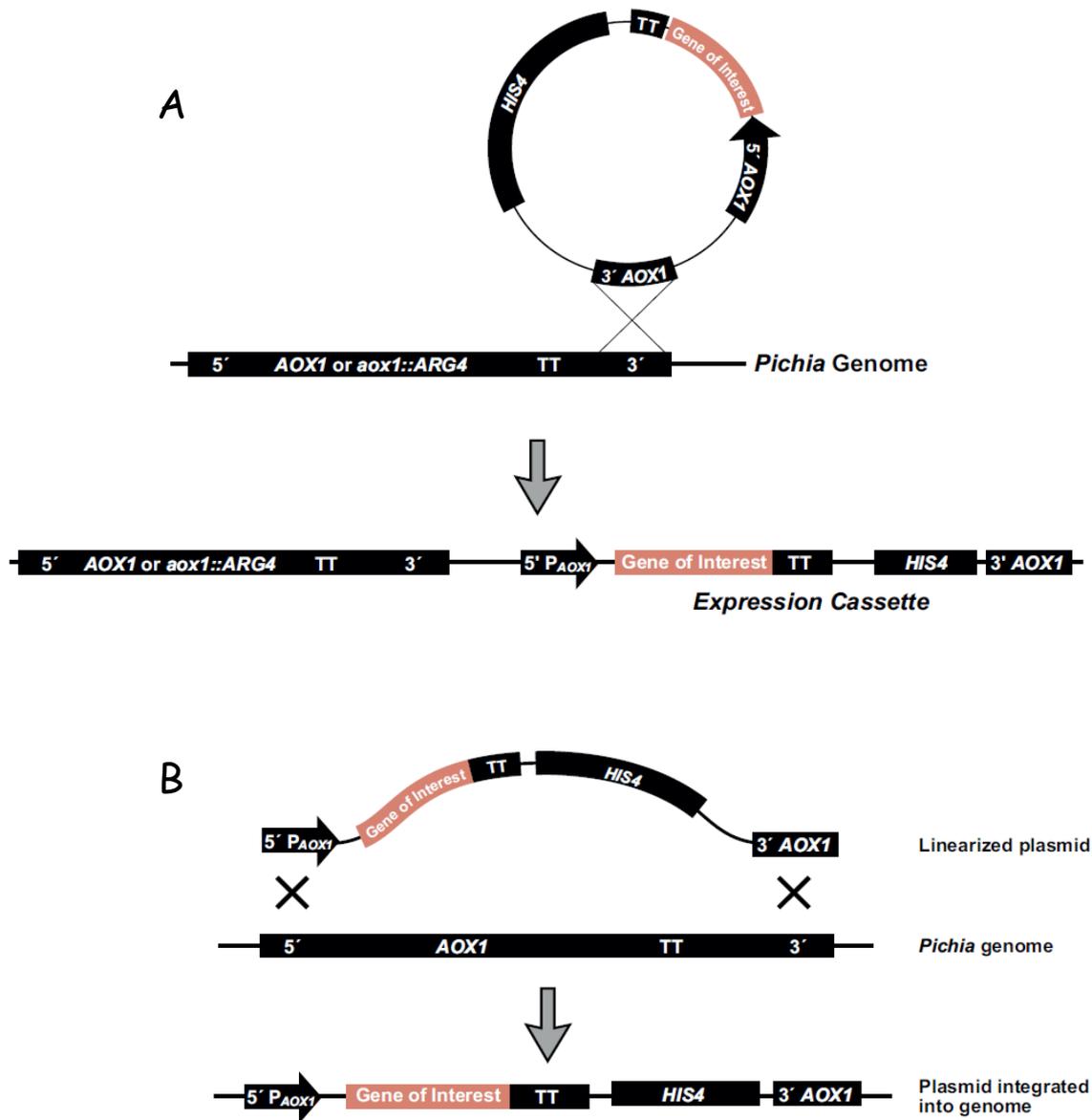
Facteur  $\alpha$  : séquence signal de sécrétion dans le milieu de culture

His<sub>6</sub> : séquence de six histidines

Flag : épitope utilisé pour marquer la protéine en immunoblotting

**Figure 7.8.** Schéma des constructions utilisées pour produire SLL-2a et SLL-2a-2 chez *P. pastoris*.

L'intégration des gènes spécifiques dans le génome de la levure s'effectue par recombinaison homologue après linéarisation du plasmide et génère des transformants de levure stables (**Figure 7.9**). Dans le cas A., la linéarisation donne une recombinaison simple qui permet l'intégration du gène d'intérêt dans une cassette d'expression en conservant le gène *AOX1*. Dans le cas B, la double recombinaison insère le gène d'intérêt à la place de celui d'*AOX1*. Seule *AOX2*, moins abondante et efficace, sera disponible pour métaboliser le méthanol. Des intégrations multiples provoquées par des recombinaisons répétées peuvent également avoir lieu dans environ 1 à 10% des cas (Daly and Hearn 2005 ; Cregg *et al.* 2009). Le site de coupure de l'enzyme de restriction *SstI* est localisé dans la région 5' *AOX1* des vecteurs, utilisée pour la double recombinaison. En choisissant cette enzyme pour la linéarisation du plasmide, nous nous plaçons forcément dans le cas A et les transformants obtenus après électroporation des plasmides sont donc de phénotype Mut<sup>+</sup>.



**Figure 7.9.** Schémas des recombinaisons homologues possibles dans le vecteur pPICZ- $\alpha$  chez la levure *Pichia pastoris*.

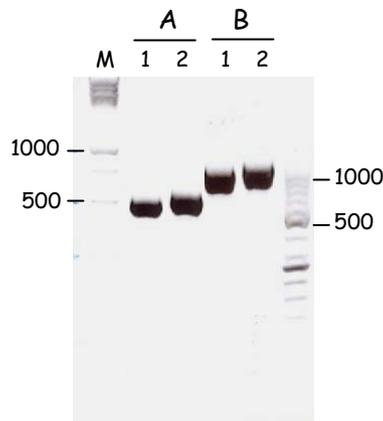
**A.** Simple recombinaison. **B.** Double recombinaison avec suppression du gène *AOX1* codant pour l'alcool oxidase.

Afin de s'assurer de l'intégration des plasmides dans le génome de la levure, une extraction d'ADN génomique suivie d'une amplification par PCR ont été réalisées. Deux couples d'amorces, spécifiques des régions *5' AOX1* / *3' AOX1* et des régions juste en amont et en aval du site de clonage multiple pPicZ $\alpha$ -F et -R (Figure 2.5. du chapitre 2) ont été utilisés. Les produits ont été amplifiés aux tailles attendues (Tableau 7.5. et Figure 7.10.). Nous pouvons néanmoins noter qu'un décalage est visible dans la migration. En effet, deux marqueurs différents ont été utilisés et les standards de 500 pb et 1000 pb n'ont pas migré à la

même distance alors que le front de migration est droit. Aucune explication ne permet d'expliquer ce résultat. Toutefois, après extraction sur gel et purification, une analyse par séquençage a définitivement confirmé l'intégration des plasmides dans le génome de *P. pastoris*.

	Couple d'amorces pPicZ $\alpha$ -F / pPicZ $\alpha$ -R désigné par A	Couple d'amorces 5' AOX1 / 3'AOX1 désigné par B
<b>Clonage dans pPIC-Z<math>\alpha</math>-A</b>		
<i>sll-2a</i>	487 pb	816 pb
<i>sll-2a-2</i>	544 pb	873 pb
<b>Clonage dans pPIC-Z<math>\alpha</math>-HisFlag</b>		
<i>sll-2a</i>	528 pb	858 pb
<i>sll-2a-2</i>	585 pb	915 pb

**Tableau 7.5.** Tailles attendues après amplification de *sll-2* et *sll-2a-2* avec les amorces 5' AOX1 / 3' AOX1 et pPicZ $\alpha$ -F / -R sur l'ADN génomique de *P. pastoris*.



**Figure 7.10.** Analyse sur gel d'agarose 1,5% de l'amplification des gènes *sll-2a* et *sll-2a-2* clonés dans pPIC-Z $\alpha$ -HisFlag à partir de l'ADN génomique de la levure.

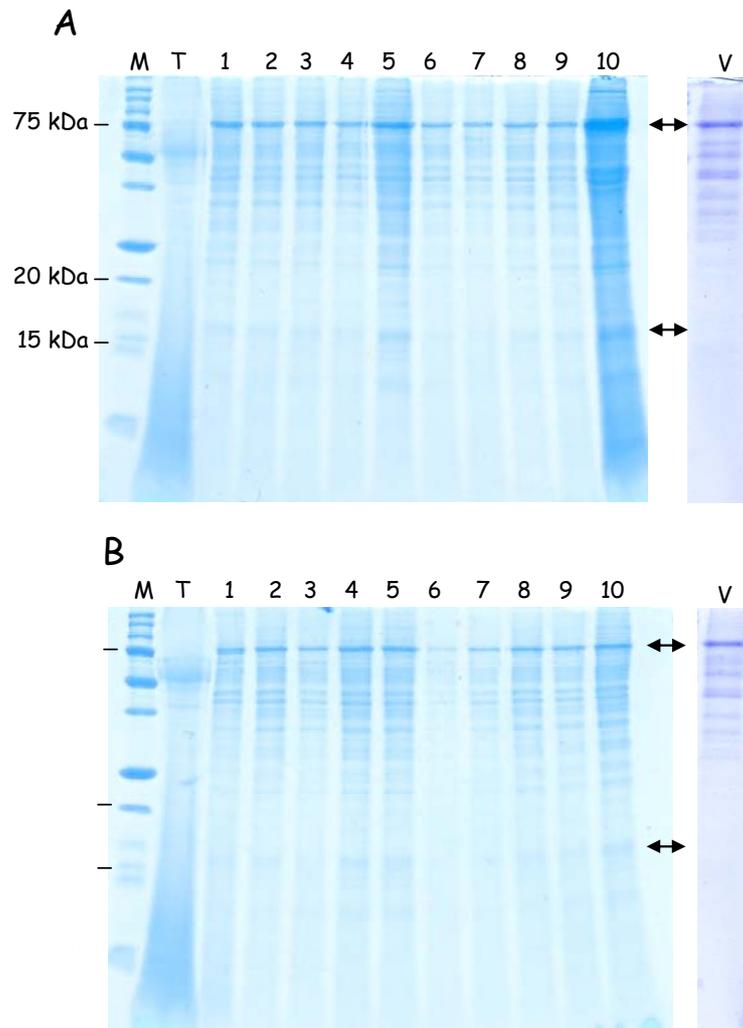
Les pistes 1 et 2 désignent respectivement les gènes *sll-2a* et *sll-2a-2* alors que A et B désignent les amplifications avec les couples d'amorces pPicZ $\alpha$ -F/-R et 5' AOX1 / 3' AOX1.

## 4.2. Conditions d'expression

Clonées dans les vecteurs décrits précédemment, l'expression de SLL-2a et SLL-2a-2 est sous le contrôle du promoteur fort AOX1, inductible au méthanol. La quantité d'enzyme AOX1 peut alors atteindre 30% des protéines cellulaires totales lorsque le méthanol est l'unique source de carbone (Couderc 1980). En effet, le glucose réprime la transcription du gène AOX1, y compris en présence de méthanol. L'utilisation de ce promoteur peut alors conduire à une production importante de protéines recombinantes même si une seule copie du gène est intégrée dans la cassette d'expression.

La première étape de l'expression consiste à sélectionner les transformants les plus productifs après trois jours d'induction à 0,5% de méthanol final. Puis seulement un clone de chaque construction sera testé en production sur cinq jours. Après une préculture sur la nuit dans un milieu contenant du glycérol comme source de carbone, les cellules sont induites dans un milieu MMYC inductible au méthanol à 30°C sous agitation (220 rpm). Trois concentrations en méthanol ont été testées : 0,25%, 0,5% et 1% final. Toutes les 24 heures, une mesure de la densité optique est réalisée et un aliquot de surnageant est prélevé, précipité au TCA 100% et analysé sur gel SDS-PAGE 15%.

La **Figure 7.11** représente les profils électrophorétiques SDS-PAGE 15% de dix clones transformés avec SLL-2a (**A.**) et SLL-2a-2 (**B.**) après trois jours d'induction au méthanol 0,5%. La quantité de protéines sécrétées dans le milieu de culture de *P. pastoris* est relativement faible. L'analyse, sur gel SDS-PAGE, montre donc des bandes peu intenses, surtout dans les bas poids moléculaires. Plusieurs hypothèses peuvent expliquer ces profils surprenants. Tout d'abord, la quantité de protéines totales sécrétées dans le milieu de culture chez *P. pastoris* est faible. La protéine d'intérêt peut représenter jusqu'à 80% des protéines totales sécrétées, ce qui est un atout considérable pour sa purification (Li *et al.* 2007). De plus, nos protéines sont attendues à de faibles poids moléculaires (autour de 10-15 kDa). Les gels sont donc concentrés en acrylamide (15%) pour être plus résolutifs dans la région contenant les protéines de bas poids moléculaires. La comparaison de nos profils avec d'autres obtenus avec des gels moins résolutifs réalisés au laboratoire indique effectivement que la majorité des protéines sécrétées possèdent de hauts poids moléculaires et sont concentrées dans la partie haute des gels. Une détection au nitrate d'argent, plus sensible, aurait néanmoins permis d'accentuer le profil général.



**Figure 7.11. Analyse sur gel SDS-PAGE 15% acrylamide coloré au bleu de Coomassie des surnageants de culture de l'expression de SLL-2a (A.) et SLL-2a-2 (B.) (clonées dans pPICZ $\alpha$ -A) dans *P. pastoris*, après trois jours d'induction au méthanol 0,5%.**

Les pistes T et V correspondent respectivement au témoin avant induction (après la préculture dans le milieu contenant le glycérol) et au témoin après induction avec le plasmide vide (le gène n'a volontairement pas été cloné). Les pistes 1 à 10 correspondent à dix transformants différents.

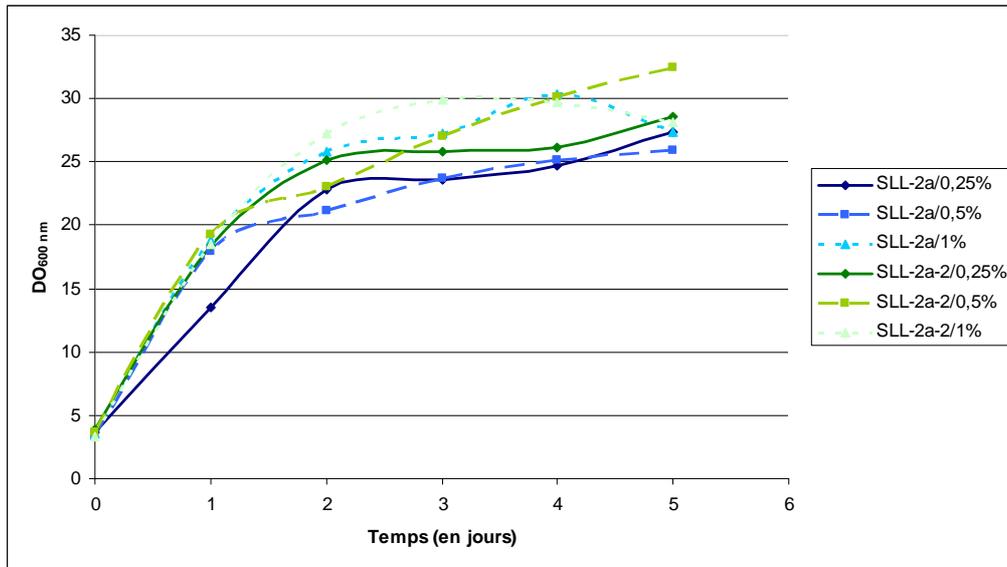
Une bande diffuse est visible à 15,1 kDa et 15,5 kDa sur les gels A et B respectivement. Elle pourrait correspondre à SLL-2a et SLL-2a-2 glycosylées. *P. pastoris* a, en effet, la capacité d'ajouter des résidus mannoses (de huit à quatorze) sur les sites de *N*-glycosylation potentiels des protéines synthétisées (Li *et al.* 2007). SLL-2a-2 possède vingt acides aminés supplémentaires par rapport à SLL-2a, ce qui correspond approximativement à une masse molaire de 2200 Da, perceptible sur gel. Or les deux bandes visibles possèdent sensiblement la même taille. Ces observations laissent envisager que la bande visible dans les

bas poids moléculaires ne correspond pas à nos protéines d'intérêt mais plutôt à une protéine de levure naturellement sécrétée. Ce résultat a également été confirmé par immunomarquage de SLL-2a et SLL-2a-2, exprimées dans pPICZ $\alpha$ -HisFlag, et détectées avec un anticorps spécifique de l'épitope His<sub>6</sub> puisqu'aucune bande n'a été révélée.

Une bande est également visible à 75 kDa dans tous les clones après induction et dans le témoin formé avec le vecteur vide (piste V Figure 7.11.) mais est absente dans le témoin avant induction. Cette bande pourrait correspondre à l'enzyme AOX1 qui a un poids moléculaire, estimé selon la séquence peptidique (AAB57849.1), de 73,9 kDa. La présence de AOX1 après induction nous confirme donc que l'induction a bien fonctionné.

Les résultats décrits précédemment ont été constatés quelque soit le vecteur utilisé, la concentration en méthanol choisie pour l'induction et le temps d'expression.

Les courbes de croissance des levures sur cinq jours (**Figure 7.12.**) sont composées de deux phases : une première partie linéaire où les levures poussent de manière exponentielle pendant deux jours et une phase de saturation où la croissance est très ralentie. Quelque soit la protéine et la concentration en méthanol utilisée pour l'induction, la croissance des levures est similaire. Lors de l'expression d'HPA (non fonctionnelle), la croissance des levures était beaucoup lente et restait exponentielle pendant les cinq jours d'induction (résultats non présentés). Cette courbe de croissance pourrait être représentative de levures en croissance qui n'expriment aucune protéine sous forme recombinante. Toute l'énergie de la cellule est utilisée pour se diviser. A l'inverse, lorsqu'une protéine est exprimée sous forme recombinante, la croissance est beaucoup plus lente puisqu'une partie de l'énergie disponible est réquisitionnée pour la production. Ces courbes de croissance peuvent donc apporter un élément supplémentaire qui semble conforter les résultats précédents : SLL-2a et SLL-2a-2 ne sont pas exprimées (ou très faiblement et donc non détectables) dans ce système d'expression.



**Figure 7.12.** Courbes de croissance des levures *P. pastoris* transformées avec SLL-2a et SLL-2a-2 dans le milieu MMYC, à 30°C.

Afin d'éliminer tout problème survenu dans la sécrétion des protéines dans le milieu de culture, le contenu des levures a été analysé sur gel SDS-PAGE15%, après éclatement au cell disrupter. Les profils obtenus sont très similaires à ceux présentés dans la **Figure 7.11**. Aucune protéine d'intérêt n'est visible dans les bas poids moléculaire.

SLL-2a et SLL-2a-2 ne sont pas exprimées dans le système eucaryote *P. pastoris*, dans les conditions d'expression testées.

### 4.3. Discussion

Il est intéressant de discuter de l'influence du phénotype de la souche de *P. pastoris* utilisée sur le rendement de protéines synthétisées. Lors de l'utilisation de souches Mut<sup>+</sup>, une partie de l'énergie consommée par la cellule est utilisée pour synthétiser l'enzyme AOX1. Lorsque son gène est absent (souche de phénotype Mut<sup>S</sup>), l'induction au méthanol est nettement moins efficace puisque, seule AOX2 est présente mais l'intégralité des protéines synthétisées correspond à la protéine d'intérêt. Le rendement attendu peut alors être meilleur.

Le glucose est un répresseur transcriptionnel du promoteur AOX1. Dans un milieu contenant du glucose, aucune expression n'est observée même en présence de méthanol. Le glycérol, à l'inverse, n'est pas un répresseur et peut être utilisé comme source de carbone. En

utilisant un milieu composé de glycérol et méthanol, les cellules se divisent en consommant le glycérol et la production de protéines est induite par le méthanol. De plus, dans une souche Mut<sup>S</sup>, le méthanol sera très lentement consommé, ce qui prolongera d'autant plus l'induction de la protéine. Ces conditions (souche Mut<sup>S</sup> et milieu composé d'un mélange glycérol/méthanol) se sont révélées optimales pour l'expression d'hormones de croissance humaines (Orman *et al.* 2009).

Ainsi, pour tenter d'améliorer le rendement d'expression de SLL-2a et SLL-2a-2, l'utilisation de la souche KM 71, de phénotype Mut<sup>S</sup>, avec des modifications de la composition du milieu de culture ou non, pourrait être envisagée.

## 5. Conclusion

Le but de ce travail de thèse consistait à trouver les conditions d'expression et de purification de la lectine de corail SLL-2 sous forme soluble et homogène. Appartenant à la famille de lectine de type H, sa production sous forme recombinante aurait grandement facilité ses caractérisations biochimique et structurale, complétant ainsi les données déjà disponibles sur cette famille.

A partir de l'étude de la séquence nucléotidique de la lectine directement issue du corail, deux séquences de gènes ont été synthétisées. La première, nommée *sll-2a*, correspond, selon les alignements de séquence, au domaine agglutinine des lectines de type H. La seconde, *sll-2a-2*, comporte 60 bases supplémentaires en 5' par rapport à *sll-2a*, de manière à conserver une séquence peut-être importante pour le repliement de la protéine.

Deux systèmes d'expression ont été testés. Dans la bactérie *E. coli*, les protéines étaient très majoritairement produites sous forme de corps d'inclusion. Cet état d'agrégation est probablement provoqué par une mauvaise formation du ou des ponts disulfure intramoléculaires proposés et peut-être par la formation de ponts disulfure intermoléculaires. L'utilisation de souches spécialisées, déficientes en réductases, n'a pas amélioré la proportion de protéines exprimées sous forme soluble dans le cytoplasme. Des expériences de dénaturation et renaturation des corps d'inclusion ont alors été envisagées. Sous certaines conditions de renaturation, la protéine était soluble mais incapable de reconnaître du galactose. Elle était donc soit mal repliée soit mal organisée. Appartenant à la famille de lectines de type H, une organisation trimérique est attendue et indispensable à la fixation du



Si les problèmes d'expression dans *E. coli* sont effectivement liés à la formation de cette liaison covalente, le clonage d'un autre gène *sll-2b* ou *sll-2c* n'aurait pas permis d'optimiser la quantité de protéines produite sous forme soluble. La stratégie la plus rapide à mettre en œuvre sera d'exprimer SLL-2a couplée à la thioredoxine dans la bactérie.

L'obtention de SLL-2a, lectine spécifique du galactose, issue du corail *Sinularia lochmodes*, sous forme recombinante aurait grandement facilité sa caractérisation biochimique. La microcalorimétrie isotherme de titration et/ou la résonance plasmonique de surface aurait permis de déterminer les constantes cinétiques et thermodynamiques de l'interaction de cette lectine avec différents ligands glycaniques. La résolution de sa structure tridimensionnelle, par cristallographie aux rayons X par exemple, aurait permis de comparer son type de repliement et son organisation oligomérique avec ses homologues structuraux, HPA et les discoidines, de manière à compléter les données disponibles sur les lectines appartenant à la famille de type H. L'alignement de séquences (**Figure 7.1.**) semble néanmoins indiquer que SLL-2a présenterait une boucle de spécificité plus courte que celle d'HPA. Cette modification, couplée avec la légère différence de spécificité de ces deux protéines (GalNAc pour HPA et Gal pour SLL-2a) laisse présager que la lectine de corail ne serait pas un marqueur de cellules cancéreuses aussi intéressant qu'HPA. Seules des expériences de marquage de cellules par microscopie confocale permettront de répondre à cette question.

# **Conclusion et perspectives**



---

## Chapitre 8

### Conclusion générale et perspectives

---

Mes travaux de thèse se sont inscrits dans le contexte général de l'analyse structure-fonction des lectines, qui constitue une des thématiques de l'équipe de Glycobiologie Moléculaire du CERMAV. Ces recherches m'ont permis d'aborder différents aspects de la biologie moderne tels que la biologie moléculaire et cellulaire, la biochimie, la cristallographie aux rayons X.

Ces travaux peuvent être divisés en trois axes majeurs. Le premier consistait à approfondir l'étude de lectines d'invertébrés appartenant à la famille de type H. En 2006, HPA, lectine à GalNAc extraite de l'escargot de Bourgogne, a été la première à être caractériser biochimiquement et structuralement (Sanchez *et al.* 2006). Les discoidines, DiscI et DiscII, sont deux lectines à Gal/GalNAc de l'amibe *Dictyostelium discoideum*, homologues à HPA. L'étude de leurs spécificités et affinités ainsi que la résolution de leurs structures cristallographiques sous forme native et complexée avec les monosaccharides GalNAc et Gal ont été effectuées durant le travail de doctorat de K. S. Aragão (Aragao *et al.* 2008). J'ai résolu les structures cristallographiques de DiscI complexée avec deux ligands disaccharidiques : le Gal $\beta$ 1-3GalNAc et le GalNAc $\beta$ 1-3Gal et ce travail a mené à publication (Mathieu *et al.* 2010). Les discoidines sont composées de deux domaines distincts : un domaine C-terminal spécifique des lectines de type H et un domaine N-terminal, appelée domaine discoidine. Durant cette thèse, j'ai conforté l'hypothèse que ce dernier serait un domaine agglutinine reconnaissant les résidus fucosylés. En effet, ses homologues structuraux les plus proches appartiennent à la famille des lectines de type F. Ces lectines trimériques sont spécifiques du fucose. Ces discoidines seraient donc des protéines composées de deux domaines lectines de spécificité différente. Dans un avenir proche, l'étude de la spécificité propre de ce domaine, associée à la détermination de constantes d'interaction avec des ligands fucosylés doit être menée. Le rôle de ces lectines, dans l'amibe, n'est toujours pas clairement défini même si l'on sait qu'elles sont impliquées dans le cycle de différenciation et de morphogenèse de l'animal. A plus long terme, il serait intéressant de déterminer la fonction

---

précise de chacun de ces domaines afin de mieux comprendre le cycle cellulaire de *D. discoideum*.

Durant cette thèse, je me suis également attachée à tenter de caractériser une autre lectine de type H issue du corail *Sinularia lochmodes*. Je me suis rapidement heurtée à des difficultés pour l'obtention de la protéine recombinante sous forme soluble. Les tentatives d'expression dans la bactérie *E. coli* et la levure *P. pastoris*, ainsi que des expériences de renaturation, n'ont pas été fructueuses. Il semblerait que la formation de ponts disulfure soit responsable de ces échecs. Des stratégies différentes doivent alors être mises au point afin de produire cette lectine de corail sous forme soluble. Le couplage de la protéine à la thioredoxine est conseillé dans les cas de problèmes de formation de ponts disulfure. Des essais d'expression et de purification sont actuellement en cours. Les caractérisations biochimique et structurale de cette lectine complèteraient les données déjà disponibles sur les lectines de type H. De plus, des expériences de microscopie confocale permettraient de tester le potentiel de cette protéine en tant que marqueur de cellules cancéreuses et d'évaluer ses capacités dans la prédiction de cancers à fort pouvoir métastatique.

Le deuxième axe de cette thèse concerne l'utilisation de ces lectines d'invertébrés dans le développement de nouveaux outils de diagnostic en cancérologie. HPA est utilisée depuis plus de vingt ans comme marqueur d'adénocarcinomes (notamment du sein, du colon, du poumon) à fort pouvoir métastatique et donc faible pronostic vital. J'ai évalué, par microscopie confocale, le potentiel des discoidines comme marqueur cancéreux. A la différence d'HPA, les discoidines ne reconnaissent pas un glycoépitope uniquement présent à la surface de cellules métastatiques et présentent donc un intérêt limité comme marqueur. La principale différence structurale entre ces lectines se situe au niveau de la longueur et de la séquence d'une boucle de spécificité localisée à proximité du site de reconnaissance. HPA présente, entre autres, une insertion de huit acides aminés par rapport aux discoidines.

Toutes les tentatives d'expression d'HPA sous forme recombinante sont restées vaines à ce jour, ce qui limite fortement son développement comme outil de routine en oncologie. Les discoidines, quant à elles, sont exprimées, de manière abondante (20 mg par litre de culture) dans la bactérie *E. coli*. Une partie très importante de ce travail de thèse consistait donc à insérer, par mutagenèse, la séquence correspondant à cette boucle supplémentaire dans DiscII de manière à fabriquer des protéines chimères qui possèderaient le même site de reconnaissance qu'HPA et qui seraient aussi faciles à exprimer que les discoidines. Le but final était de développer de nouveaux marqueurs hybrides qui possèderaient les mêmes caractéristiques prédictives qu'HPA. Plusieurs mutants ont été construits d'après la

comparaison des structures cristallographiques de DiscII et d'HPA. La résolution de la structure cristallographique d'une de ces chimères a permis de contrôler que la boucle de spécificité insérée semblait orientée de manière similaire à HPA. Ces protéines restent néanmoins non fonctionnelles. Une protéine hybride formée du domaine N-terminal de DiscII couplé à HPA a également été construite mais ses tentatives d'expression dans *E. coli* sont restées vaines. Des études par modélisation moléculaire permettraient peut-être d'identifier les zones responsables de cette non fonctionnalité et d'aider à la conception de nouvelles chimères.

La structure du ou des épitopes glycaniques reconnu(s) par HPA à la surface des cellules cancéreuses n'est toujours pas clairement identifiée à ce jour. Sa détermination permettrait pourtant d'apporter des réponses sur les voies biologiques employées par la cellule durant le processus de formation des métastases. La dernière partie de ce travail de thèse consistait donc à tenter d'identifier ces épitopes. L'analyse par spectrométrie de masse des *N*-glycanes, issus de cellules de cancer colorectal métastatiques ou non, démontre un enrichissement de certaines structures après purification sur colonne HPA. Ces travaux restent néanmoins très préliminaires et des expériences complémentaires doivent être réalisées afin de standardiser les protocoles et valider la méthode. Les GalNAc terminaux sont généralement localisés sur des *O*-glycanes. L'étude de ce type de résidu doit, malgré sa complexité plus importante, également être envisagée.

Le taux de guérison d'un patient atteint de cancer est d'autant plus important que le diagnostic de la maladie a été précoce, d'où l'importance de développer rapidement des marqueurs spécifiques des cellules à caractère métastatiques comme HPA. De plus, il serait également envisageable de considérer cette lectine comme vecteur de médicaments. Sa structure à deux domaines similaires permet d'imaginer la construction d'un agent spécifique de cellules métastatiques, sur lequel seraient couplées des molécules capables d'interagir avec ces cellules et de provoquer leur destruction. L'émergence et le développement de nouveaux outils sont donc d'un intérêt majeur en oncologie.



# **Bibliographie**



# Bibliographie

---

- Agrawal, B. B. and I. J. Goldstein (1965). "Specific binding of concanavalin A to cross-linked dextran gels." *Biochem J* **96**(3): 23contd-5c.
- Akamatsu, S., S. Yazawa, T. Tachikawa, T. Furuta, Y. Okaichi, J. Nakamura, T. Asao and Y. Nagamachi (1996). "Alpha 2-->3sialyltransferase associated with the synthesis of CA 19-9 in colorectal tumors." *Cancer* **77**(8 Suppl): 1694-700.
- Alexander, S., A. M. Cibulsky and R. A. Lerner (1983). "Ion dependence of the discoidin I lectin from *Dictyostelium discoideum*." *Differentiation* **24**(3): 209-12.
- Aminoff, D. and W. T. Morgan (1948). "Hexosamine components of the human blood group substances." *Nature* **162**(4119): 579.
- Amorim, J. H., B. F. Porchia, A. Balan, R. C. Cavalcante, S. M. da Costa, A. M. de Barcelos Alves and L. C. de Souza Ferreira (2010). "Refolded dengue virus type 2 NS1 protein expressed in *Escherichia coli* preserves structural and immunological properties of the native protein." *J Virol Methods* **167**(2): 186-92.
- Annesley, S. J. and P. R. Fisher (2009). "*Dictyostelium discoideum*--a model for many reasons." *Mol Cell Biochem* **329**(1-2): 73-91.
- Apweiler, R., H. Hermjakob and N. Sharon (1999). "On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database." *Biochim Biophys Acta* **1473**(1): 4-8.
- Aragao, K. S., M. Satre, A. Imberty and A. Varrot (2008). "Structure determination of Discoidin II from *Dictyostelium discoideum* and carbohydrate binding properties of the lectin domain." *Proteins* **73**(1): 43-52.
- Baldus, S. E., K. Engelmann and F. G. Hanisch (2004). "MUC1 and the MUCs: a family of human mucins with impact in cancer biology." *Crit Rev Clin Lab Sci* **41**(2): 189-231.
- Barondes, S. H., D. N. Cooper and P. L. Haywood-Reid (1983). "Discoidin I and discoidin II are localized differently in developing *Dictyostelium discoideum*." *J Cell Biol* **96**(1): 291-6.
- Bast, R. C., Jr., M. Feeney, H. Lazarus, L. M. Nadler, R. B. Colvin and R. C. Knapp (1981). "Reactivity of a monoclonal antibody with human ovarian carcinoma." *J Clin Invest* **68**(5): 1331-7.
- Baumgartner, S., K. Hofmann, R. Chiquet-Ehrismann and P. Bucher (1998). "The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction." *Protein Sci* **7**(7): 1626-31.
- Berger, E. A. and D. R. Armant (1982). "Discoidins I and II: common and unique regions on two lectins implicated in cell--cell cohesion in *Dictyostelium discoideum*." *Proc Natl Acad Sci U S A* **79**(7): 2162-6.
- Bernstein, F. C., T. F. Koetzle, G. J. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, et al. (1977). "The Protein Data Bank: a computer-based archival file for macromolecular structures." *J Mol Biol* **112**(3): 535-42.
- Berois, N., D. Mazal, L. Ubillos, F. Trajtenberg, A. Nicolas, X. Sastre-Garau, H. Magdelenat and E. Osinaga (2006). "UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 as a new immunohistochemical breast cancer marker." *J Histochem Cytochem* **54**(3): 317-28.

- Berstein, F. (1924). "Ergebnisse einer biostatistischen zusammenfassenden Betrachtung über die erblichen Blutstrukturen des Menschen." Klin Wschr **3**: 1495-1497.
- Bhende, Y. M., C. K. Deshpande, H. M. Bhatia, R. Sanger, R. R. Race, W. T. Morgan and W. M. Watkins (1952). "A "new" blood group character related to the ABO system." Lancet **1**(6714): 903-4.
- Bianchet, M. A., E. W. Odom, G. R. Vasta and L. M. Amzel (2010). "Structure and specificity of a binary tandem domain F-lectin from striped bass (*Morone saxatilis*)." J Mol Biol **401**(2): 239-52.
- Bianchet, M. A., E. W. Odom, G. R. Vasta and L. M. Amzel (2002). "A novel fucose recognition fold involved in innate immunity." Nat Struct Biol **9**(8): 628-34.
- Birkle, S., G. Zeng, L. Gao, R. K. Yu and J. Aubry (2003). "Role of tumor-associated gangliosides in cancer progression." Biochimie **85**(3-4): 455-63.
- Bosio, A., E. Binczek, M. M. Le Beau, A. A. Fernald and W. Stoffel (1996). "The human gene CGT encoding the UDP-galactose ceramide galactosyl transferase (cerebrosidase synthase): cloning, characterization, and assignment to human chromosome 4, band q26." Genomics **34**(1): 69-75.
- Boyd, W. C. and E. Shapleigh (1954). "Specific Precipitating Activity of Plant Agglutinins (Lectins)." Science **119**(3091): 419.
- Breton, C., L. Snajdrova, C. Jeanneau, J. Koca and A. Imberty (2006). "Structures and mechanisms of glycosyltransferases." Glycobiology **16**(2): 29R-37R.
- Brockhausen, I. (2006). "Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions." EMBO Rep **7**(6): 599-604.
- Brockhausen, I., J. Yang, N. Dickinson, S. Ogata and S. H. Itzkowitz (1998). "Enzymatic basis for sialyl-Tn expression in human colon cancer cells." Glycoconj J **15**(6): 595-603.
- Brockhausen, I., J. M. Yang, J. Burchell, C. Whitehouse and J. Taylor-Papadimitriou (1995). "Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells." Eur J Biochem **233**(2): 607-17.
- Brooks, S. A. (2000). "The involvement of Helix pomatia lectin (HPA) binding N-acetylgalactosamine glycans in cancer progression." Histol Histopathol **15**(1): 143-58.
- Brooks, S. A. and T. M. Carter (2001). "N-acetylgalactosamine, N-acetylglucosamine and sialic acid expression in primary breast cancers." Acta Histochem **103**(1): 37-51.
- Brooks, S. A., T. M. Carter, E. P. Bennett, H. Clausen and U. Mandel (2007). "Immunolocalisation of members of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer." Acta Histochem **109**(4): 273-84.
- Brooks, S. A., D. M. Hall and I. Buley (2001). "GalNAc glycoprotein expression by breast cell lines, primary breast cancer and normal breast epithelial membrane." Br J Cancer **85**(7): 1014-22.
- Brooks, S. A. and A. J. Leathem (1991). "Prediction of lymph node involvement in breast cancer by detection of altered glycosylation in the primary tumour." Lancet **338**(8759): 71-4.
- Brooks, S. A. and A. J. Leathem (1998). "Expression of N-acetyl galactosaminylated and sialylated glycans by metastases arising from primary breast cancer." Invasion Metastasis **18**(3): 115-21.
- Brooks, S. A., M. Lymboura, U. Schumacher and A. J. Leathem (1996). "Histochemistry to detect Helix pomatia lectin binding in breast cancer: methodology makes a difference." J Histochem Cytochem **44**(5): 519-24.

- Brooks, S. A. and D. Wilkinson (2003). "Validation of a simple avidin-biotin detection method for *Helix pomatia* lectin (HPA) binding as a prognostic marker in cancer." Acta Histochem **105**(3): 205-12.
- Brosius, J., M. Erfle and J. Storella (1985). "Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity." J Biol Chem **260**(6): 3539-41.
- Bryan, N., K. Andrews, M. J. Loughran, N. P. Rhodes and J. A. Hunt (2010). "Elucidating the Contribution of the Elemental Composition of Fetal Calf Serum Towards Antigenic Expression of Primary Human Umbilical Vein Endothelial Cells In Vitro." Biosci Rep.
- Burgess, R. R. (2009). "Refolding solubilized inclusion body proteins." Methods Enzymol **463**: 259-82.
- Camus, M. L. (1899). "Recherches expérimentales sur une agglutinine produite par la glande de l'albumen chez l'*Helix pomatia*." C R Acad Sci **129**: 233-234.
- Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics." Nucleic Acids Res **37**(Database issue): D233-8.
- Carvalho, A. S., A. Harduin-Lepers, A. Magalhaes, E. Machado, N. Mendes, L. T. Costa, R. Matthiesen, R. Almeida, et al. (2010) "Differential expression of alpha-2,3-sialyltransferases and alpha-1,3/4-fucosyltransferases regulates the levels of sialyl Lewis a and sialyl Lewis x in gastrointestinal carcinoma cells." Int J Biochem Cell Biol **42**(1): 80-9.
- Ceroni, A., A. Dell and S. M. Haslam (2007). "The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures." Source Code Biol Med **2**: 3.
- Ceroni, A., A. Passerini, A. Vullo and P. Frasconi (2006). "DISULFIND: a disulfide bonding state and cysteine connectivity prediction server." Nucleic Acids Res **34**(Web Server issue): W177-81.
- Charnock, S. J. and G. J. Davies (1999). "Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms." Biochemistry **38**(20): 6380-5.
- Chiu, C. P., A. G. Watts, L. L. Lairson, M. Gilbert, D. Lim, W. W. Wakarchuk, S. G. Withers and N. C. Strynadka (2004). "Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analog." Nat Struct Mol Biol **11**(2): 163-70.
- Chrispeels, M. J. (1970). "Biosynthesis of cell wall protein: sequential hydroxylation of proline, glycosylation of hydroxyproline and secretion of the glycoprotein." Biochem Biophys Res Commun **39**(4): 732-7.
- Ciucanu, I., and F. Kerek (1984). "A simple and rapid method for permethylation of carbohydrates." Carbohydrate Res **131**: 209-.
- Ciucanu, I. a. C. E. C. (2003). "Elimination of oxidative degradation during the per-O-methylation of carbohydrates." J Am Chem Soc **125**: 16213-16219.
- Cooper, D. N. and S. H. Barondes (1984). "Colocalization of discoidin-binding ligands with discoidin in developing *Dictyostelium discoideum*." Dev Biol **105**(1): 59-70.
- Cooper, D. N., S. C. Lee and S. H. Barondes (1983). "Discoidin-binding polysaccharide from *Dictyostelium discoideum*." J Biol Chem **258**(14): 8745-50.
- Corfield, A. P., N. Myerscough, B. F. Warren, P. Durdey, C. Paraskeva and R. Schauer (1999). "Reduction of sialic acid O-acetylation in human colonic mucins in the adenoma-carcinoma sequence." Glycoconj J **16**(6): 307-17.
- Couderc, R., and J. Baratti (1980). "Oxidation of methanol by the yeast *Pichia pastoris*. Purification and properties of alcohol oxidase." Agric Biol Chem **44**: 2279-2289.

- Cregg, J. M., I. Tolstorukov, A. Kusari, J. Sunga, K. Madden and T. Chappell (2009). "Expression in the yeast *Pichia pastoris*." Methods Enzymol **463**: 169-89.
- Cummings, R. D., I. S. Trowbridge and S. Kornfeld (1982). "A mouse lymphoma cell line resistant to the leucoagglutinating lectin from *Phaseolus vulgaris* is deficient in UDP-GlcNAc: alpha-D-mannoside beta 1,6 N-acetylglucosaminyltransferase." J Biol Chem **257**(22): 13421-7.
- Dall'Olio, F. and M. Chiricolo (2001). "Sialyltransferases in cancer." Glycoconj J **18**(11-12): 841-50.
- Daly, R. and M. T. Hearn (2005). "Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production." J Mol Recognit **18**(2): 119-38.
- Dam, T. K. and C. F. Brewer (2002). "Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry." Chem Rev **102**(2): 387-429.
- de Beer, T., J. F. Vliegthart, A. Loffler and J. Hofsteenge (1995). "The hexopyranosyl residue that is C-glycosidically linked to the side chain of tryptophan-7 in human RNase Us is alpha-mannopyranose." Biochemistry **34**(37): 11785-9.
- de Leoz, M. L., L. J. Young, H. J. An, S. R. Kronewitter, J. Kim, S. Miyamoto, A. D. Borowsky, H. K. Chew, et al. (2010). "High-mannose glycans are elevated during breast cancer progression." Mol Cell Proteomics.
- de Marco, A. (2009). "Strategies for successful recombinant expression of disulfide bond-dependent proteins in *Escherichia coli*." Microb Cell Fact **8**: 26.
- Demetriou, M., I. R. Nabi, M. Coppolino, S. Dedhar and J. W. Dennis (1995). "Reduced contact-inhibition and substratum adhesion in epithelial cells expressing GlcNAc-transferase V." J Cell Biol **130**(2): 383-92.
- Dennis, J. W. (1986). "Effects of swainsonine and polyinosinic:polycytidylic acid on murine tumor cell growth and metastasis." Cancer Res **46**(10): 5131-6.
- Dennis, J. W., S. Laferte, C. Waghorne, M. L. Breitman and R. S. Kerbel (1987). "Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis." Science **236**(4801): 582-5.
- Dennis, J. W., I. R. Nabi and M. Demetriou (2009). "Metabolism, cell surface organization, and disease." Cell **139**(7): 1229-41.
- Dong, D. L. and G. W. Hart (1994). "Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from rat spleen cytosol." J Biol Chem **269**(30): 19321-30.
- Donovan, R. S., C. W. Robinson and B. R. Glick (1996). "Review: optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter." J Ind Microbiol **16**(3): 145-54.
- Doucey, M. A., D. Hess, R. Cacan and J. Hofsteenge (1998). "Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosphate-mannose as a precursor." Mol Biol Cell **9**(2): 291-300.
- Dwek, M. V., H. A. Ross, A. J. Streets, S. A. Brooks, E. Adam, A. Titcomb, J. V. Woodside, U. Schumacher, et al. (2001). "Helix pomatia agglutinin lectin-binding oligosaccharides of aggressive breast cancer." Int J Cancer **95**(2): 79-85.
- Ellis, R. J. and A. P. Minton (2003). "Cell biology: join the crowd." Nature **425**(6953): 27-8.
- Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.
- Erdos, G. W., and C. M. West (1989). "Formation and organization of the spore coat of *Dictyostelium discoideum*." Exp Mycol **13**: 169-182.
- Fahnert, B., H. Lilie and P. Neubauer (2004). "Inclusion bodies: formation and utilisation." Adv Biochem Eng Biotechnol **89**: 93-142.

- Fernandes, B., U. Sagman, M. Auger, M. Demetrio and J. W. Dennis (1991). "Beta 1-6 branched oligosaccharides as a marker of tumor progression in human breast and colon neoplasia." *Cancer Res* **51**(2): 718-23.
- Ferre, F. and P. Clote (2005). "DiANNA: a web server for disulfide connectivity prediction." *Nucleic Acids Res* **33**(Web Server issue): W230-2.
- Ferre, F. and P. Clote (2005). "Disulfide connectivity prediction using secondary structure information and diresidue frequencies." *Bioinformatics* **21**(10): 2336-46.
- Frazier, W. A., S. D. Rosen, R. W. Reitherman and S. H. Barondes (1975). "Purification and comparison of two developmentally regulated lectins from Dictyostelium discoideum. Discoidin I and II." *J Biol Chem* **250**(19): 7714-21.
- Futami, J., H. Tada, M. Seno, S. Ishikami and H. Yamada (2000). "Stabilization of human RNase 1 by introduction of a disulfide bond between residues 4 and 118." *J Biochem* **128**(2): 245-50.
- Gerlach, D., B. Schlott, U. Zahringer and K. H. Schmidt (2005). "N-acetyl-D-galactosamine/N-acetyl-D-glucosamine--recognizing lectin from the snail *Cepaea hortensis*: purification, chemical characterization, cloning and expression in *E. coli*." *FEMS Immunol Med Microbiol* **43**(2): 223-32.
- Geyer, H. and R. Geyer (2006). "Strategies for analysis of glycoprotein glycosylation." *Biochim Biophys Acta* **1764**(12): 1853-69.
- Gleeson, P. A. and H. Schachter (1983). "Control of glycoprotein synthesis." *J Biol Chem* **258**(10): 6162-73.
- Gold, P. and S. O. Freedman (1965). "Specific carcinoembryonic antigens of the human digestive system." *J Exp Med* **122**(3): 467-81.
- Goldstein, I. J. and R. D. Poretz, Eds. (1986). *The lectins. Properties, functions and applications in biology and medicine*. Orlando, Florida, Academic Press.
- Goldstein, I. J., R. C. Hughes, M. Monsigny, T. Osawa and N. Sharon, N. (1980). "What should be called a lectin?" *Nature* **285**: 66.
- Golks, A. and D. Guerini (2008). "The O-linked N-acetylglucosamine modification in cellular signalling and the immune system. 'Protein modifications: beyond the usual suspects' review series." *EMBO Rep* **9**(8): 748-53.
- Gotthardt, D., V. Blancheteau, A. Bosserhoff, T. Ruppert, M. Delorenzi and T. Soldati (2006). "Proteomics fingerprinting of phagosome maturation and evidence for the role of a Galpha during uptake." *Mol Cell Proteomics* **5**(12): 2228-43.
- Goupille, C., S. Marionneau, V. Bureau, F. Hallouin, M. Meichenin, J. Rocher and J. Le Pendu (2000). "alpha1,2Fucosyltransferase increases resistance to apoptosis of rat colon carcinoma cells." *Glycobiology* **10**(4): 375-82.
- Granovsky, M., J. Fata, J. Pawling, W. J. Muller, R. Khokha and J. W. Dennis (2000). "Suppression of tumor growth and metastasis in Mgat5-deficient mice." *Nat Med* **6**(3): 306-12.
- Graslund, S., P. Nordlund, J. Weigelt, B. M. Hallberg, J. Bray, O. Gileadi, S. Knapp, U. Oppermann, et al. (2008). "Protein production and purification." *Nat Methods* **5**(2): 135-46.
- Gusterson, B. A., and the International (Ludwig) Breast Cancer Study Group (1993). "Prognostic value of Helix pomatia in breast cancer." *Br J Cancer* **68**(1): 146-150.
- Hacker, U., K. Nybakken and N. Perrimon (2005). "Heparan sulphate proteoglycans: the sweet side of development." *Nat Rev Mol Cell Biol* **6**(7): 530-41.
- Hagihara, Y., S. Mine and K. Uegaki (2007). "Stabilization of an immunoglobulin fold domain by an engineered disulfide bond at the buried hydrophobic region." *J Biol Chem* **282**(50): 36489-95.

- Hakomori, S. (1999). "Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer." *Biochim Biophys Acta* **1473**(1): 247-66.
- Hakomori, S. (2000). "Traveling for the glycosphingolipid path." *Glycoconj J* **17**(7-9): 627-47.
- Hakomori, S. I. (2008). "Structure and function of glycosphingolipids and sphingolipids: recollections and future trends." *Biochim Biophys Acta* **1780**(3): 325-46.
- Haltiwanger, R. S., G. D. Holt and G. W. Hart (1990). "Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase." *J Biol Chem* **265**(5): 2563-8.
- Hamada, H., T. Arakawa and K. Shiraki (2009). "Effect of additives on protein aggregation." *Curr Pharm Biotechnol* **10**(4): 400-7.
- Hammarstrom, S. and E. A. Kabat (1969). "Purification and characterization of a blood-group A reactive hemagglutinin from the snail *Helix pomatia* and a study of its combining site." *Biochemistry* **8**(7): 2696-705.
- Harpaz, N. and H. Schachter (1980). "Control of glycoprotein synthesis. Bovine colostrum UDP-N-acetylglucosamine:alpha-D-mannoside beta 2-N-acetylglucosaminyltransferase I. Separation from UDP-N-acetylglucosamine:alpha-D-mannoside beta 2-N-acetylglucosaminyltransferase II, partial purification, and substrate specificity." *J Biol Chem* **255**(10): 4885-93.
- Harvey, D. J. (2005). "Structural determination of N-linked glycans by matrix-assisted laser desorption/ionization and electrospray ionization mass spectrometry." *Proteomics* **5**(7): 1774-86.
- Hashimoto, S., T. Asao, J. Takahashi, Y. Yagihashi, T. Nishimura, A. R. Saniabadi, D. C. Poland, W. van Dijk, et al. (2004). "alpha1-acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis." *Cancer* **101**(12): 2825-36.
- Haslam, S. M., S. J. North and A. Dell (2006). "Mass spectrometric analysis of N- and O-glycosylation of tissues and cells." *Curr Opin Struct Biol* **16**(5): 584-91.
- Hatahet, F., V. D. Nguyen, K. E. Salo and L. W. Ruddock (2010). "Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*." *Microb Cell Fact* **9**: 67.
- Helenius, A. and M. Aebi (2001). "Intracellular functions of N-linked glycans." *Science* **291**(5512): 2364-9.
- Hofsteenge, J., M. Blommers, D. Hess, A. Furmanek and O. Miroshnichenko (1999). "The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues." *J Biol Chem* **274**(46): 32786-94.
- Hu, R., G. L. Mukhina, S. H. Lee, R. J. Jones, P. T. Englund, P. Brown, S. J. Sharkis, J. T. Buckley, et al. (2009). "Silencing of genes required for glycosylphosphatidylinositol anchor biosynthesis in Burkitt lymphoma." *Exp Hematol* **37**(4): 423-434 e2.
- Hubbard, S. C. and R. J. Ivatt (1981). "Synthesis and processing of asparagine-linked oligosaccharides." *Annu Rev Biochem* **50**: 555-83.
- Humphries, M. J., K. Matsumoto, S. L. White and K. Olden (1986). "Oligosaccharide modification by swainsonine treatment inhibits pulmonary colonization by B16-F10 murine melanoma cells." *Proc Natl Acad Sci U S A* **83**(6): 1752-6.
- Ichikawa, S., H. Sakiyama, G. Suzuki, K. I. Hidari and Y. Hirabayashi (1996). "Expression cloning of a cDNA for human ceramide glucosyltransferase that catalyzes the first glycosylation step of glycosphingolipid synthesis." *Proc Natl Acad Sci U S A* **93**(22): 12654.

- Igura, M., N. Maita, J. Kamishikiryo, M. Yamada, T. Obita, K. Maenaka and D. Kohda (2008). "Structure-guided identification of a new catalytic motif of oligosaccharyltransferase." EMBO J **27**(1): 234-43.
- Ikeda, Y., M. Mori, Y. Adachi, T. Matsushima and K. Sugimachi (1994). "Prognostic value of the histochemical expression of helix pomatia agglutinin in advanced colorectal cancer. A univariate and multivariate analysis." Dis Colon Rectum **37**(2): 181-4.
- Imberty, A., H. Lortat-Jacob and S. Perez (2007). "Structural view of glycosaminoglycan-protein interactions." Carbohydr Res **342**(3-4): 430-9.
- Imberty, A., M. wimmerova, E. P. Mitchell and N. Gilboa-Garber (2004). "Structures of the lectins from *Pseudomonas aeruginosa*: insight into the molecular basis for host glycan recognition." Microbes Infect **6**(2): 221-8.
- Ito, H., N. Hiraiwa, M. Sawada-Kasugai, S. Akamatsu, T. Tachikawa, Y. Kasai, S. Akiyama, K. Ito, et al. (1997). "Altered mRNA expression of specific molecular species of fucosyl- and sialyl-transferases in human colorectal cancer tissues." Int J Cancer **71**(4): 556-64.
- Ito, N., S. Imai, S. Haga, C. Nagaike, Y. Morimura and K. Hatake (1996). "Localization of binding sites of *Ulex europaeus* I, *Helix pomatia* and *Griffonia simplicifolia* I-B4 lectins and analysis of their backbone structures by several glycosidases and poly-N-acetyllactosamine-specific lectins in human breast carcinomas." Histochem Cell Biol **106**(3): 331-9.
- Ito, Y., A. Akinaga, K. Yamanaka, T. Nakagawa, A. Kondo, R. B. Dickson, C. Y. Lin, A. Miyauchi, et al. (2006). "Co-expression of matriptase and N-acetylglucosaminyltransferase V in thyroid cancer tissues--its possible role in prolonged stability in vivo by aberrant glycosylation." Glycobiology **16**(5): 368-74.
- Ito, Y., A. Miyauchi, H. Yoshida, T. Uruno, K. Nakano, Y. Takamura, A. Miya, K. Kobayashi, et al. (2003). "Expression of alpha1,6-fucosyltransferase (FUT8) in papillary carcinoma of the thyroid: its linkage to biological aggressiveness and anaplastic transformation." Cancer Lett **200**(2): 167-72.
- Itzkowitz, S. H., E. J. Bloom, W. A. Kokal, G. Modin, S. Hakomori and Y. S. Kim (1990). "Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients." Cancer **66**(9): 1960-6.
- Jang-Lee, J., S. J. North, M. Sutton-Smith, D. Goldberg, M. Panico, H. Morris, S. Haslam and A. Dell (2006). "Glycomic profiling of cells and tissues by mass spectrometry: fingerprinting and sequencing methodologies." Methods Enzymol **415**: 59-86.
- Jayson, G. C., M. Lyon, C. Paraskeva, J. E. Turnbull, J. A. Deakin and J. T. Gallagher (1998). "Heparan sulfate undergoes specific structural changes during the progression from human colon adenoma to carcinoma in vitro." J Biol Chem **273**(1): 51-7.
- Jimbo, M., K. Koike, R. Sakai, K. Muramoto and H. Kamiya (2005). "Cloning and characterization of a lectin from the octocoral *Sinularia lochmodes*." Biochem Biophys Res Commun **330**(1): 157-62.
- Jimbo, M., Koike, K., Sakai, R., Muramoto, K., and Kamiya, H. (2005). "Cloning and characterization of a lectin from the octocoral *Sinularia lochmodes*." Biochem Biophys Res Commun **330**(1): 157-162.
- Jimbo, M., T. Yanohara, K. Koike, R. Sakai, K. Muramoto and H. Kamiya (2000). "The D-galactose-binding lectin of the octocoral *Sinularia lochmodes*: characterization and possible relationship to the symbiotic dinoflagellates." Comp Biochem Physiol B Biochem Mol Biol **125**(2): 227-36.
- Ju, T., K. Brewer, A. D'Souza, R. D. Cummings and W. M. Canfield (2002). "Cloning and expression of human core 1 beta1,3-galactosyltransferase." J Biol Chem **277**(1): 178-86.

- Julien, S., E. Adriaenssens, K. Ottenberg, A. Furlan, G. Courtand, A. S. Vercoutter-Edouart, F. G. Hanisch, P. Delannoy, et al. (2006). "ST6GalNAc I expression in MDA-MB-231 breast cancer cells greatly modifies their O-glycosylation pattern and enhances their tumorigenicity." *Glycobiology* **16**(1): 54-64.
- Jurado, P., V. de Lorenzo and L. A. Fernandez (2006). "Thioredoxin fusions increase folding of single chain Fv antibodies in the cytoplasm of Escherichia coli: evidence that chaperone activity is the prime effect of thioredoxin." *J Mol Biol* **357**(1): 49-61.
- Kadokura, H., F. Katzen and J. Beckwith (2003). "Protein disulfide bond formation in prokaryotes." *Annu Rev Biochem* **72**: 111-35.
- Takeji, Y., Y. Maehara, S. Tsujitani, H. Baba, S. Ohno, A. Watanabe and K. Sugimachi (1994). "Helix pomatia agglutinin binding activity and lymph node metastasis in patients with gastric cancer." *Semin Surg Oncol* **10**(2): 130-4.
- Khien, V. V., H. V. Mao, T. T. Chinh, P. T. Ha, M. H. Bang, B. V. Lac, T. V. Hop, N. A. Tuan, et al. (2001). "Clinical evaluation of lentil lectin-reactive alpha-fetoprotein-L3 in histology-proven hepatocellular carcinoma." *Int J Biol Markers* **16**(2): 105-11.
- Kjonniksen, I., P. D. Rye and O. Fodstad (1994). "Helix pomatia agglutinin binding in human tumour cell lines: correlation with pulmonary metastases in nude mice." *Br J Cancer* **69**(6): 1021-4.
- Koike, K., M. Jimbo, R. Sakai, M. Kaeriyama, K. Muramoto, T. Ogata, T. Maruyama and H. Kamiya (2004). "Octocoral chemical signaling selects and controls dinoflagellate symbionts." *Biol Bull* **207**(2): 80-6.
- Kominato, Y., Y. Hata, H. Takizawa, T. Tsuchiya, J. Tsukada and F. Yamamoto (1999). "Expression of human histo-blood group ABO genes is dependent upon DNA methylation of the promoter region." *J Biol Chem* **274**(52): 37240-50.
- Kornfeld, R. and S. Kornfeld (1985). "Assembly of asparagine-linked oligosaccharides." *Annu Rev Biochem* **54**: 631-64.
- Kornfeld, S., E. Li and I. Tabas (1978). "The synthesis of complex-type oligosaccharides. II. Characterization of the processing intermediates in the synthesis of the complex oligosaccharide units of the vesicular stomatitis virus G protein." *J Biol Chem* **253**(21): 7771-8.
- Korourian, S., E. Siegel, T. Kieber-Emmons and B. Monzavi-Karbassi (2008). "Expression analysis of carbohydrate antigens in ductal carcinoma in situ of the breast by lectin histochemistry." *BMC Cancer* **8**: 136.
- Kumamoto, K., Y. Goto, K. Sekikawa, S. Takenoshita, N. Ishida, M. Kawakita and R. Kannagi (2001). "Increased expression of UDP-galactose transporter messenger RNA in human colon cancer tissues and its implication in synthesis of Thomsen-Friedenreich antigen and sialyl Lewis A/X determinants." *Cancer Res* **61**(11): 4620-7.
- Laack, E., H. Nikbakht, A. Peters, C. Kugler, Y. Jasiewicz, L. Edler, D. K. Hossfeld and U. Schumacher (2002). "Lectin histochemistry of resected adenocarcinoma of the lung: helix pomatia agglutinin binding is an independent prognostic factor." *Am J Pathol* **160**(3): 1001-8.
- Lairson, L. L., B. Henrissat, G. J. Davies and S. G. Withers (2008). "Glycosyltransferases: structures, functions, and mechanisms." *Annu Rev Biochem* **77**: 521-55.
- Lamblin, G., M. Lhermitte, A. Klein, N. Houdret, A. Scharfman, R. Ramphal and P. Roussel (1991). "The carbohydrate diversity of human respiratory mucins: a protection of the underlying mucosa?" *Am Rev Respir Dis* **144**(3 Pt 2): S19-24.
- Landsteiner, K. (1900). "Zur Kenntniss der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe." *Zbl Bakt* **27**: 357-366.
- Landsteiner, K. (1901). "Über Agglunationserscheinungen normalen menschlichen Blutes." *Wein Klin Wschr* **14**: 1132-1134.

- Landsteiner, K. and R. A. Harte (1940). "On Group Specific a Substances : Iv. The Substance from Hog Stomach." J Exp Med **71**(4): 551-62.
- Larsen, R. D., L. K. Ernst, R. P. Nair and J. B. Lowe (1990). "Molecular cloning, sequence, and expression of a human GDP-L-fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase cDNA that can form the H blood group antigen." Proc Natl Acad Sci U S A **87**(17): 6674-8.
- Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J Appl Crystallogr **26**: 283-291.
- LaVallie, E. R., E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel and J. M. McCoy (1993). "A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm." Biotechnology (N Y) **11**(2): 187-93.
- Le Pendu, J., S. Marionneau, A. Cailleau-Thomas, J. Rocher, B. Le Moullac-Vaidye and M. Clement (2001). "ABH and Lewis histo-blood group antigens in cancer." APMIS **109**(1): 9-31.
- Leathem, A. J. and S. A. Brooks (1987). "Predictive value of lectin binding on breast-cancer recurrence and survival." Lancet **1**(8541): 1054-6.
- Lee, A. K., R. A. DeLellis, P. P. Rosen, P. E. Saigo, M. D. Gangi, R. Bagin, S. Groshen and H. J. Wolfe (1985). "ABH blood group isoantigen expression in breast carcinomas--an immunohistochemical evaluation using monoclonal antibodies." Am J Clin Pathol **83**(3): 308-19.
- Lee, I. L., P. S. Li, W. L. Yu and H. H. Shen (2010). "Prokaryotic expression, refolding, and purification of functional human vascular endothelial growth factor isoform 165: Purification procedures and refolding conditions revisited." Protein Expr Purif.
- Leivonen, M., S. Nordling, J. Lundin, K. von Boguslawski and C. Haglund (2001). "STn and prognosis in breast cancer." Oncology **61**(4): 299-305.
- Lescar, J., J. F. Sanchez, A. Audfray, J. L. Coll, C. Breton, E. P. Mitchell and A. Imberty (2007). "Structural basis for recognition of breast and colon cancer epitopes Tn antigen and Forssman disaccharide by Helix pomatia lectin." Glycobiology **17**(10): 1077-83.
- Li, E., I. Tabas and S. Kornfeld (1978). "The synthesis of complex-type oligosaccharides. I. Structure of the lipid-linked oligosaccharide precursor of the complex-type oligosaccharides of the vesicular stomatitis virus G protein." J Biol Chem **253**(21): 7762-70.
- Li, M., Z. G. Su and J. C. Janson (2004). "In vitro protein refolding by chromatographic procedures." Protein Expr Purif **33**(1): 1-10.
- Li, P., A. Anumanthan, X. G. Gao, K. Ilangovan, V. V. Suzara, N. Duzgunes and V. Renugopalakrishnan (2007). "Expression of recombinant proteins in Pichia pastoris." Appl Biochem Biotechnol **142**(2): 105-24.
- Lin, H. H. and L. Y. Tseng (2010). "DBCP: a web server for disulfide bonding connectivity pattern prediction without the prior knowledge of the bonding state of cysteines." Nucleic Acids Res **38** **Suppl**: W503-7.
- Lis, H. and N. Sharon (1998). "Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition." Chem Rev **98**(2): 637-674.
- Liu, D., Z. Shriver, G. Venkataraman, Y. El Shabrawi and R. Sasisekharan (2002). "Tumor cell surface heparan sulfate as cryptic promoters or inhibitors of tumor growth and metastasis." Proc Natl Acad Sci U S A **99**(2): 568-73.
- Lommel, M. and S. Strahl (2009). "Protein O-mannosylation: conserved from bacteria to humans." Glycobiology **19**(8): 816-28.

- Lovering, A. L., L. H. de Castro, D. Lim and N. C. Strynadka (2007). "Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis." *Science* **315**(5817): 1402-5.
- Luther, K. B. and R. S. Haltiwanger (2009). "Role of unusual O-glycans in intercellular signaling." *Int J Biochem Cell Biol* **41**(5): 1011-24.
- Marcos, N. T., A. Cruz, F. Silva, R. Almeida, L. David, U. Mandel, H. Clausen, S. Von Mensdorff-Pouilly, et al. (2003). "Polypeptide GalNAc-transferases, ST6GalNAc-transferase I, and ST3Gal-transferase I expression in gastric carcinoma cell lines." *J Histochem Cytochem* **51**(6): 761-71.
- Mas, E., E. Pasqualini, N. Caillol, A. El Battari, C. Crotte, D. Lombardo and M. O. Sadoulet (1998). "Fucosyltransferase activities in human pancreatic tissue: comparative study between cancer tissues and established tumoral cell lines." *Glycobiology* **8**(6): 605-13.
- Mathieu, S. V., K. S. Aragao, A. Imberty and A. Varrot (2010). "Discoidin I from Dictyostelium discoideum and Interactions with oligosaccharides: specificity, affinity, crystal structures, and comparison with discoidin II." *J Mol Biol* **400**(3): 540-54.
- Matsumoto, H., H. Muramatsu, T. Shimotakahara, M. Yanagi, H. Nishijima, N. Mitani, K. Baba, T. Muramatsu, et al. (1993). "Correlation of expression of ABH blood group carbohydrate antigens with metastatic potential in human lung carcinomas." *Cancer* **72**(1): 75-81.
- Matsuura, N., T. Narita, N. Hiraiwa, M. Hiraiwa, H. Murai, T. Iwase, H. Funahashi, T. Imai, et al. (1998). "Gene expression of fucosyl- and sialyl-transferases which synthesize sialyl Lewisx, the carbohydrate ligands for E-selectin, in human breast cancer." *Int J Oncol* **12**(5): 1157-64.
- McCoy, A. J. (2007). "Solving structures of protein complexes by molecular replacement with Phaser." *Acta Crystallogr D Biol Crystallogr* **63**(Pt 1): 32-41.
- Meldgaard, P., P. H. Johnson, N. C. Langkilde, H. Wolf and T. F. Orntoft (1995). "Loss of ABH antigen expression in bladder cancer is not caused by loss of heterozygosity of the ABO locus." *Int J Cancer* **63**(3): 341-4.
- Miyoshi, E., N. Uozumi, K. Noda, N. Hayashi, M. Hori and N. Taniguchi (1997). "Expression of alpha1-6 fucosyltransferase in rat tissues and human cancer cell lines." *Int J Cancer* **72**(6): 1117-21.
- Molinari, M. (2007). "N-glycan structure dictates extension of protein folding or onset of disposal." *Nat Chem Biol* **3**(6): 313-20.
- Morelle, W., V. Faid, F. Chirat and J. C. Michalski (2009). "Analysis of N- and O-linked glycans from glycoproteins using MALDI-TOF mass spectrometry." *Methods Mol Biol* **534**: 5-21.
- Morgan, W. T. and W. M. Watkins (1948). "The detection of a product of the blood group O gene and the relationship of the so-called O-substance to the agglutinates A and B." *Br J Exp Pathol* **29**(2): 159-73.
- Morgan, W. T. and W. M. Watkins (1953). "The inhibition of the haemagglutinins in plant seeds by human blood group substances and simple sugars." *Br J Exp Pathol* **34**(1): 94-103.
- Mourant, A. E. (1946). "A new human blood group antigen of frequent occurrence." *Nature* **158**: 237.
- Muinelo-Romay, L., C. Vazquez-Martin, S. Villar-Portela, E. Cuevas, E. Gil-Martin and A. Fernandez-Briera (2008). "Expression and enzyme activity of alpha(1,6)fucosyltransferase in human colorectal cancer." *Int J Cancer* **123**(3): 641-6.
- Nagano, N., M. Ota and K. Nishikawa (1999). "Strong hydrophobic nature of cysteine residues in proteins." *FEBS Lett* **458**(1): 69-71.

- Nakagoe, T., A. Nanashima, T. Sawai, T. Tuji, M. Ohbatake, M. Jibiki, H. Yamaguchi, T. Yasutake, et al. (2000). "Expression of blood group antigens A, B and H in carcinoma tissue correlates with a poor prognosis for colorectal cancer patients." J Cancer Res Clin Oncol **126**(7): 375-82.
- Nakamura, J., A. Mogi, T. Asao, Y. Nagamachi and S. Yazawa (1997). "Evidence that the aberrant alpha 1-->2fucosyltransferase found in colorectal carcinoma may be encoded by Fuc-TIII (Le) gene." Anticancer Res **17**(6D): 4563-9.
- Narasimhan, S. (1982). "Control of glycoprotein synthesis. UDP-GlcNAc:glycopeptide beta 4-N-acetylglucosaminyltransferase III, an enzyme in hen oviduct which adds GlcNAc in beta 1-4 linkage to the beta-linked mannose of the trimannosyl core of N-glycosyl oligosaccharides." J Biol Chem **257**(17): 10235-42.
- Narita, T., H. Funahashi, Y. Satoh, T. Watanabe, J. Sakamoto and H. Takagi (1993). "Association of expression of blood group-related carbohydrate antigens with prognosis in breast cancer." Cancer **71**(10): 3044-53.
- Ngoh, G. A. and S. P. Jones (2008). "New insights into metabolic signaling and cell survival: the role of beta-O-linkage of N-acetylglucosamine." J Pharmacol Exp Ther **327**(3): 602-9.
- Nikaido, H. (1994). "Maltose transport system of Escherichia coli: an ABC-type transporter." FEBS Lett **346**(1): 55-8.
- Ogawa, J. I., H. Inoue and S. Koide (1997). "alpha-2,3-Sialyltransferase type 3N and alpha-1,3-fucosyltransferase type VII are related to sialyl Lewis(x) synthesis and patient survival from lung carcinoma." Cancer **79**(9): 1678-85.
- Ohtsubo, K. and J. D. Marth (2006). "Glycosylation in cellular mechanisms of health and disease." Cell **126**(5): 855-67.
- Ohyama, C., M. Hosono, K. Nitta, M. Oh-eda, K. Yoshikawa, T. Habuchi, Y. Arai and M. Fukuda (2004). "Carbohydrate structure and differential binding of prostate specific antigen to Maackia amurensis lectin between prostate cancer and benign prostate hypertrophy." Glycobiology **14**(8): 671-9.
- Ohyama, C., S. Tsuboi and M. Fukuda (1999). "Dual roles of sialyl Lewis X oligosaccharides in tumor metastasis and rejection by natural killer cells." EMBO J **18**(6): 1516-25.
- Orlow, I., L. Lacombe, I. Pellicer, F. Rabbani, R. Delgado, Z. F. Zhang, I. Szijan and C. Cordon-Cardo (1998). "Genotypic and phenotypic characterization of the histoblood group ABO(H) in primary bladder tumors." Int J Cancer **75**(6): 819-24.
- Orman, M. A., P. Calik and T. H. Ozdamar (2009). "The influence of carbon sources on recombinant-human- growth-hormone production by Pichia pastoris is dependent on phenotype: a comparison of Muts and Mut+ strains." Biotechnol Appl Biochem **52**(Pt 3): 245-55.
- Orntoft, T. F. and H. Wolf (1988). "Blood group ABO and Lewis antigens in bladder tumors: correlation between glycosyltransferase activity and antigen expression." APMIS Suppl **4**: 126-33.
- Osumi, D., M. Takahashi, E. Miyoshi, S. Yokoe, S. H. Lee, K. Noda, S. Nakamori, J. Gu, et al. (2009). "Core fucosylation of E-cadherin enhances cell-cell adhesion in human colon carcinoma WiDr cells." Cancer Sci **100**(5): 888-95.
- Park, I. J., G. S. Choi, K. H. Lim, B. M. Kang and S. H. Jun (2009). "Serum carcinoembryonic antigen monitoring after curative resection for colorectal cancer: clinical significance of the preoperative level." Ann Surg Oncol **16**(11): 3087-93.
- Paulick, M. G. and C. R. Bertozzi (2008). "The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins." Biochemistry **47**(27): 6991-7000.

- Pinho, S. S., C. A. Reis, J. Paredes, A. M. Magalhaes, A. C. Ferreira, J. Figueiredo, W. Xiaogang, F. Carneiro, et al. (2009). "The role of N-acetylglucosaminyltransferase III and V in the post-transcriptional modifications of E-cadherin." Hum Mol Genet **18**(14): 2599-608.
- Pornillos, O., B. K. Ganser-Pornillos, S. Banumathi, Y. Hua and M. Yeager (2010) "Disulfide bond stabilization of the hexameric capsomer of human immunodeficiency virus." J Mol Biol **401**(5): 985-95.
- Rademacher, T. W., R. B. Parekh and R. A. Dwek (1988). "Glycobiology." Annu Rev Biochem **57**: 785-838.
- Rattanaporn, O. and P. Utarabhand (2010). "Molecular cloning of a C-type lectin with two CRD domains from the banana shrimp *Fenneropenaeus merguensis*: Early gene up-regulation after *Vibrio harveyi* infection." J Invertebr Pathol.
- Rege, T. A. and J. S. Hagood (2006). "Thy-1 as a regulator of cell-cell and cell-matrix interactions in axon regeneration, apoptosis, adhesion, migration, cancer, and fibrosis." FASEB J **20**(8): 1045-54.
- Rhodes, G., Ed. (2006). Crystallography made crystal clear: a guide for users of macromolecular models, Elsevier.
- Rivinoja, A., N. Kokkonen, I. Kellokumpu and S. Kellokumpu (2006). "Elevated Golgi pH in breast and colorectal cancer cells correlates with the expression of oncofetal carbohydrate T-antigen." J Cell Physiol **208**(1): 167-74.
- Rogers, C. M., K. B. Cooke and M. I. Filipe (1978). "Sialic acids of human large bowel mucosa: o-acetylated variants in normal and malignant states." Gut **19**(7): 587-92.
- Rosen, S. D., J. A. Kafka, D. L. Simpson and S. H. Barondes (1973). "Developmentally regulated, carbohydrate-binding protein in *Dictyostelium discoideum*." Proc Natl Acad Sci U S A **70**(9): 2554-7.
- Rossmann, M. G., and D. M. Blow (1962). "The detection of sub-units within the crystallographic asymmetric unit." Acta Cryst **15**: 24-31.
- Rüdiger, H., Ed. (1993). Purification of plant lectin. Gabius, H. J. G. Berlin, Springer.
- Ruoslahti, E. (2003). "The RGD story: a personal account." Matrix Biol **22**(6): 459-65.
- Ruotsalainen, H., L. Sipila, M. Vapola, R. Sormunen, A. M. Salo, L. Uitto, D. K. Mercer, S. P. Robins, et al. (2006). "Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes." J Cell Sci **119**(Pt 4): 625-35.
- Rye, P. D., O. Fodstad, E. Emilsen and M. Bryne (1998). "Invasion potential and N-acetylgalactosamine expression in a human melanoma model." Int J Cancer **75**(4): 609-14.
- Saerens, D., K. Conrath, J. Govaert and S. Muyldermans (2008). "Disulfide bond introduction for general stabilization of immunoglobulin heavy-chain variable domains." J Mol Biol **377**(2): 478-88.
- Saint-Guirons, J., E. Zeqiraj, U. Schumacher, P. Greenwell and M. Dwek (2007). "Proteome analysis of metastatic colorectal cancer cells recognized by the lectin *Helix pomatia* agglutinin (HPA)." Proteomics **7**(22): 4082-9.
- Sanchez, J. F., J. Lescar, V. Chazalet, A. Audfray, J. Gagnon, R. Alvarez, C. Breton, A. Imberty, et al. (2006). "Biochemical and structural analysis of *Helix pomatia* agglutinin. A hexameric lectin with a novel fold." J Biol Chem **281**(29): 20171-80.
- Sanderson, R. D., Y. Yang, L. J. Suva and T. Kelly (2004). "Heparan sulfate proteoglycans and heparanase--partners in osteolytic tumor growth and metastasis." Matrix Biol **23**(6): 341-52.
- Sasisekharan, R., Z. Shriver, G. Venkataraman and U. Narayanasami (2002). "Roles of heparan-sulphate glycosaminoglycans in cancer." Nat Rev Cancer **2**(7): 521-8.

- Schachter, H. (1986). "Biosynthetic controls that determine the branching and microheterogeneity of protein-bound oligosaccharides." *Adv Exp Med Biol* **205**: 53-85.
- Schulenburg, C., U. Weininger, P. Neumann, H. Meiselbach, M. T. Stubbs, H. Sticht, J. Balbach, R. Ulbrich-Hofmann, et al. (2010) "Impact of the C-terminal disulfide bond on the folding and stability of onconase." *Chembiochem* **11**(7): 978-86.
- Schumacher, U. and E. Adam (1997). "Lectin histochemical HPA-binding pattern of human breast and colon cancers is associated with metastases formation in severe combined immunodeficient mice." *Histochem J* **29**(9): 677-84.
- Schumacher, U., E. Adam, S. A. Brooks and A. J. Leathem (1995). "Lectin-binding properties of human breast cancer cell lines and human milk with particular reference to *Helix pomatia* agglutinin." *J Histochem Cytochem* **43**(3): 275-81.
- Schumacher, U., Adam, E., Brooks, S. A. and Leathem, A. J., Ed. (1996). *The use of Helix pomatia agglutinin in metastases research*. Lectins: Biology, Biochemistry, Clinical Biochemistry. Hellerup (Denmark), Textop.
- Schumacher, U., H. Kretschmar, S. Brooks and A. Leathem (1992). "Helix pomatia lectin binding pattern of brain metastases originating from breast cancers." *Pathol Res Pract* **188**(3): 284-6.
- Sentandreu, R. and D. H. Northcote (1968). "The structure of a glycopeptide isolated from the yeast cell wall." *Biochem J* **109**(3): 419-32.
- Sewell, R., M. Backstrom, M. Dalziel, S. Gschmeissner, H. Karlsson, T. Noll, J. Gatgens, H. Clausen, et al. (2006). "The ST6GalNAc-I sialyltransferase localizes throughout the Golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer." *J Biol Chem* **281**(6): 3586-94.
- Sharon, N. and H. Lis (2004). "History of lectins: from hemagglutinins to biological recognition molecules." *Glycobiology* **14**(11): 53R-62R.
- Sillo, A., G. Bloomfield, A. Balest, A. Balbo, B. Pergolizzi, B. Peracino, J. Skelton, A. Ivens, et al. (2008). "Genome-wide transcriptional changes induced by phagocytosis or growth on bacteria in *Dictyostelium*." *BMC Genomics* **9**: 291.
- Sipila, L., H. Ruotsalainen, R. Sormunen, N. L. Baker, S. R. Lamande, M. Vapola, C. Wang, Y. Sado, et al. (2007). "Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines." *J Biol Chem* **282**(46): 33381-8.
- Slovin, S. F., G. Ragupathi, C. Fernandez, M. Diani, M. P. Jefferson, A. Wilton, W. K. Kelly, M. Morris, et al. (2007). "A polyvalent vaccine for high-risk prostate patients: "are more antigens better?"" *Cancer Immunol Immunother* **56**(12): 1921-30.
- Smith, D. B. (2000). "Generating fusions to glutathione S-transferase for protein studies." *Methods Enzymol* **326**: 254-70.
- Sohail, A., Q. Sun, H. Zhao, M. M. Bernardo, J. A. Cho and R. Fridman (2008). "MT4-(MMP17) and MT6-MMP (MMP25), A unique set of membrane-anchored matrix metalloproteinases: properties and expression in cancer." *Cancer Metastasis Rev* **27**(2): 289-302.
- Spiro, R. G. (1969). "Characterization and quantitative determination of the hydroxylysine-linked carbohydrate units of several collagens." *J Biol Chem* **244**(4): 602-12.
- Springer, G. F. (1984). "T and Tn, general carcinoma autoantigens." *Science* **224**(4654): 1198-206.
- Springer, G. F. (1989). "Tn epitope (N-acetyl-D-galactosamine alpha-O-serine/threonine) density in primary breast carcinoma: a functional predictor of aggressiveness." *Mol Immunol* **26**(1): 1-5.
- Springer, G. F. (1997). "Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy." *J Mol Med* **75**(8): 594-602.

- Springer, G. F., P. R. Desai, B. D. Spencer, H. Tegtmeier, S. C. Carlstedt and E. F. Scanlon (1995). "T/Tn antigen vaccine is effective and safe in preventing recurrence of advanced breast carcinoma." *Cancer Detect Prev* **19**(4): 374-80.
- Stamey, T. A., N. Yang, A. R. Hay, J. E. McNeal, F. S. Freiha and E. Redwine (1987). "Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate." *N Engl J Med* **317**(15): 909-16.
- Stewart, E. J., F. Aslund and J. Beckwith (1998). "Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins." *EMBO J* **17**(19): 5543-50.
- Sulak, O., G. Cioci, M. Delia, M. Lahmann, A. Varrot, A. Imberty and M. Wimmerova (2010). "A TNF-like trimeric lectin domain from Burkholderia cenocepacia with specificity for fucosylated human histo-blood group antigens." *Structure* **18**(1): 59-72.
- Sumner, J. B. (1919). "The globulins of the jack bean, Conavalia ensiformis." *J Biol Chem* **37**: 137-142.
- Sumner, J. B. and S. F. Howell (1936). "Identification of Hemagglutinin of Jack Bean with Concanavalin A." *J Bacteriol* **32**(2): 227-37.
- Sun, J., J. Thurin, H. S. Cooper, P. Wang, M. Mackiewicz, Z. Stepkowski and M. Blaszczyk-Thurin (1995). "Elevated expression of H type GDP-L-fucose:beta-D-galactoside alpha-2-L-fucosyltransferase is associated with human colon adenocarcinoma progression." *Proc Natl Acad Sci U S A* **92**(12): 5724-8.
- Swietnicki, W. (2006). "Folding aggregated proteins into functionally active forms." *Curr Opin Biotechnol* **17**(4): 367-72.
- Tabares, G., C. M. Radcliffe, S. Barrabes, M. Ramirez, R. N. Aleixandre, W. Hoesel, R. A. Dwek, P. M. Rudd, et al. (2006). "Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA." *Glycobiology* **16**(2): 132-45.
- Tabas, I. and S. Kornfeld (1978). "The synthesis of complex-type oligosaccharides. III. Identification of an alpha-D-mannosidase activity involved in a late stage of processing of complex-type oligosaccharides." *J Biol Chem* **253**(21): 7779-86.
- Tada, T., T. Kumada, H. Toyoda, S. Kiriyama, Y. Sone, M. Tanikawa, Y. Hisanaga, S. Kitabatake, et al. (2005). "Relationship between Lens culinaris agglutinin-reactive alpha-fetoprotein and pathologic features of hepatocellular carcinoma." *Liver Int* **25**(4): 848-53.
- Takahashi, M., Y. Kuroki, K. Ohtsubo and N. Taniguchi (2009). "Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins." *Carbohydr Res* **344**(12): 1387-90.
- Takahashi, T., Y. Ikeda, E. Miyoshi, Y. Yaginuma, M. Ishikawa and N. Taniguchi (2000). "alpha1,6fucosyltransferase is highly and specifically expressed in human ovarian serous adenocarcinomas." *Int J Cancer* **88**(6): 914-9.
- Takizawa, M., T. Nomura, E. Wakisaka, N. Yoshizuka, J. Aoki, H. Arai, K. Inoue, M. Hattori, et al. (1999). "cDNA cloning and expression of human lactosylceramide synthase." *Biochim Biophys Acta* **1438**(2): 301-4.
- Taniguchi, N., E. Miyoshi, J. H. Ko, Y. Ikeda and Y. Ihara (1999). "Implication of N-acetylglucosaminyltransferases III and V in cancer: gene regulation and signaling mechanism." *Biochim Biophys Acta* **1455**(2-3): 287-300.
- Tatsumura, T., H. Sato, A. Mori, Y. Komori, K. Yamamoto, G. Fukatani and S. Kuno (1977). "Clinical significance of fucose level in glycoprotein fraction of serum in patients with malignant tumors." *Cancer Res* **37**(11): 4101-3.

- Ten Hagen, K. G., T. A. Fritz and L. A. Tabak (2003). "All in the family: the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases." *Glycobiology* **13**(1): 1R-16R.
- Teng-umnuay, P., H. R. Morris, A. Dell, M. Panico, T. Paxton and C. M. West (1998). "The cytoplasmic F-box binding protein SKP1 contains a novel pentasaccharide linked to hydroxyproline in Dictyostelium." *J Biol Chem* **273**(29): 18242-9.
- Terpe, K. (2006). "Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems." *Appl Microbiol Biotechnol* **72**(2): 211-22.
- Thies, A., I. Moll, J. Berger and U. Schumacher (2001). "Lectin binding to cutaneous malignant melanoma: HPA is associated with metastasis formation." *Br J Cancer* **84**(6): 819-23.
- Torres, C. R. and G. W. Hart (1984). "Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc." *J Biol Chem* **259**(5): 3308-17.
- Tsumoto, K., D. Ejima, I. Kumagai and T. Arakawa (2003). "Practical considerations in refolding proteins from inclusion bodies." *Protein Expr Purif* **28**(1): 1-8.
- Van den Steen, P., P. M. Rudd, R. A. Dwek and G. Opdenakker (1998). "Concepts and principles of O-linked glycosylation." *Crit Rev Biochem Mol Biol* **33**(3): 151-208.
- Van Der Wel, H., H. R. Morris, M. Panico, T. Paxton, A. Dell, L. Kaplan and C. M. West (2002). "Molecular cloning and expression of a UDP-N-acetylglucosamine (GlcNAc):hydroxyproline polypeptide GlcNAc-transferase that modifies Skp1 in the cytoplasm of dictyostelium." *J Biol Chem* **277**(48): 46328-37.
- van Kuik, J. A., H. van Halbeek, J. P. Kamerling and J. F. Vliegenthart (1985). "Primary structure of the low-molecular-weight carbohydrate chains of Helix pomatia alpha-hemocyanin. Xylose as a constituent of N-linked oligosaccharides in an animal glycoprotein." *J Biol Chem* **260**(26): 13984-8.
- Varki, A., R. Cummings, J. Esko, H. Freeze, G. Hart and J. Marth (1999). "Essentials of Glycobiology, 2nd edition." *Consortium of Glycobiology Editors, La Jolla, California.*
- Vercoutter-Edouart, A. S., M. C. Slomianny, O. Dekeyzer-Beseme, J. F. Haeuw and J. C. Michalski (2008). "Glycoproteomics and glycomics investigation of membrane N-glycosylproteins from human colon carcinoma cells." *Proteomics* **8**(16): 3236-56.
- Videira, P. A., M. Correia, N. Malagolini, H. J. Crespo, D. Ligeiro, F. M. Calais, H. Trindade and F. Dall'Olio (2009). "ST3Gal.I sialyltransferase relevance in bladder cancer tissues and cell lines." *BMC Cancer* **9**: 357.
- Vlodavsky, I., M. Elkin, G. Abboud-Jarrous, F. Levi-Adam, L. Fuks, I. Shafat and N. Ilan (2008). "Heparanase: one molecule with multiple functions in cancer progression." *Connect Tissue Res* **49**(3): 207-10.
- von Decastello, A., and Sturli, A. (1902). "Uber die isoagglutinine im serum gesunder und kranker Menschen." *Munchen Medicine Wchnschr* **49**(1090-1095).
- Vretblad, P., R. Hjorth and T. Laas (1979). "The isolectins of Helix pomatia. Separation by isoelectric focusing and preliminary characterization." *Biochim Biophys Acta* **579**(1): 52-61.
- Vrieling, A., W. Ruger, H. P. Driessen and P. S. Freemont (1994). "Crystal structure of the DNA modifying enzyme beta-glucosyltransferase in the presence and absence of the substrate uridine diphosphoglucose." *EMBO J* **13**(15): 3413-22.
- Vriend, G. (1990). "WHAT IF: a molecular modeling and drug design program." *J Mol Graph* **8**(1): 52-6, 29.

- Wang, J. W., R. A. Ambros, P. B. Weber and T. G. Rosano (1995). "Fucosyltransferase and alpha-L-fucosidase activities and fucose levels in normal and malignant endometrial tissue." *Cancer Res* **55**(16): 3654-8.
- Wang, X., S. Inoue, J. Gu, E. Miyoshi, K. Noda, W. Li, Y. Mizuno-Horikawa, M. Nakano, et al. (2005). "Dysregulation of TGF-beta1 receptor activation leads to abnormal lung development and emphysema-like phenotype in core fucose-deficient mice." *Proc Natl Acad Sci U S A* **102**(44): 15791-6.
- Watkins, W. M. (2001). "The ABO blood group system: historical background." *Transfus Med* **11**(4): 243-65.
- Watkins, W. M. and W. T. Morgan (1952). "Neutralization of the anti-H agglutinin in eel serum by simple sugars." *Nature* **169**(4307): 825-6.
- Watkins, W. M. and W. T. Morgan (1955). "Inhibition by simple sugars of enzymes which decompose the blood-group substances." *Nature* **175**(4459): 676-7.
- Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." *Nature* **171**(4356): 737-8.
- Welinder, C., B. Jansson, M. Ferno, H. Olson and B. Baldetorp (2009). "Expression of Helix pomatia lectin binding glycoproteins in women with breast cancer in relationship to their blood group phenotypes." *J Proteome Res* **8**(2): 782-7.
- Welshinger, M., C. L. Finstad, E. Venkatraman, M. G. Federici, S. C. Rubin, J. L. Lewis, Jr. and K. O. Lloyd (1996). "Expression of A, B, and H blood group antigens in epithelial ovarian cancer: relationship to tumor grade and patient survival." *Gynecol Oncol* **62**(1): 106-12.
- Wennekes, T., R. J. van den Berg, R. G. Boot, G. A. van der Marel, H. S. Overkleeft and J. M. Aerts (2009). "Glycosphingolipids--nature, function, and pharmacological modulation." *Angew Chem Int Ed Engl* **48**(47): 8848-69.
- West, C. M., P. Nguyen, H. van der Wel, T. Metcalf, K. R. Sweeney, I. J. Blader and G. W. Erdos (2009). "Dependence of stress resistance on a spore coat heteropolysaccharide in *Dictyostelium*." *Eukaryot Cell* **8**(1): 27-36.
- Wood-Charlson, E. M., L. L. Hollingsworth, D. A. Krupp and V. M. Weis (2006). "Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis." *Cell Microbiol* **8**(12): 1985-93.
- Wopereis, S., D. J. Lefeber, E. Morava and R. A. Wevers (2006). "Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review." *Clin Chem* **52**(4): 574-600.
- Wu, A. M., and Sugii, S. (1991). "Coding and classification of D-galactose, N-acetyl-D-galactosamine, and  $\beta$ -D-Galp-[1-3(4)]- $\beta$ -D-GlcpNAc, specificities of applied lectins." *Carbohydr Res* **213**: 127-143.
- Yamada, N., Y. S. Chung, S. Takatsuka, Y. Arimoto, T. Sawada, T. Dohi and M. Sowa (1997). "Increased sialyl Lewis A expression and fucosyltransferase activity with acquisition of a high metastatic capacity in a colon cancer cell line." *Br J Cancer* **76**(5): 582-7.
- Yazawa, S., J. Nakamura, T. Asao, Y. Nagamachi, M. Sagi, K. L. Matta, T. Tachikawa and M. Akamatsu (1993). "Aberrant alpha 1-->2fucosyltransferases found in human colorectal carcinoma involved in the accumulation of Leb and Y antigens in colorectal tumors." *Jpn J Cancer Res* **84**(9): 989-95.
- Yoshida, H., M. Teraoka, N. Nishi, S. I. Nakakita, T. Nakamura, M. Hirashima and S. Kamitori (2010). "X-ray structures of human galectin-9 C-terminal domain in complexes with a biantennary oligosaccharide and sialyllactose." *J Biol Chem*.

- 
- Yoshimura, M., A. Nishikawa, Y. Ihara, S. Taniguchi and N. Taniguchi (1995). "Suppression of lung metastasis of B16 mouse melanoma by N-acetylglucosaminyltransferase III gene transfection." *Proc Natl Acad Sci U S A* **92**(19): 8754-8.
- Yu, L. G. (2007). "The oncofetal Thomsen-Friedenreich carbohydrate antigen in cancer progression." *Glycoconj J* **24**(8): 411-20.
- Yuan, Y., D. Barrett, Y. Zhang, D. Kahne, P. Sliz and S. Walker (2007). "Crystal structure of a peptidoglycan glycosyltransferase suggests a model for processive glycan chain synthesis." *Proc Natl Acad Sci U S A* **104**(13): 5348-53.
- Zavodszky, M., C. W. Chen, J. K. Huang, M. Zolkiewski, L. Wen and R. Krishnamoorthi (2001). "Disulfide bond effects on protein stability: designed variants of *Cucurbita maxima* trypsin inhibitor-V." *Protein Sci* **10**(1): 149-60.
- Zhang, X. W., W. T. Xu, X. W. Wang, Y. Mu, X. F. Zhao, X. Q. Yu and J. X. Wang (2009). "A novel C-type lectin with two CRD domains from Chinese shrimp *Fenneropenaeus chinensis* functions as a pattern recognition protein." *Mol Immunol* **46**(8-9): 1626-37.
- Zhao, Y., T. Nakagawa, S. Itoh, K. Inamori, T. Isaji, Y. Kariya, A. Kondo, E. Miyoshi, et al. (2006). "N-acetylglucosaminyltransferase III antagonizes the effect of N-acetylglucosaminyltransferase V on alpha3beta1 integrin-mediated cell migration." *J Biol Chem* **281**(43): 32122-30.
- Zhao, Y. Y., M. Takahashi, J. G. Gu, E. Miyoshi, A. Matsumoto, S. Kitazume and N. Taniguchi (2008). "Functional roles of N-glycans in cell signaling and cell adhesion in cancer." *Cancer Sci* **99**(7): 1304-10.



# **Annexe**



# Annexe : Analyse, par spectrométrie de masse, des glycoépitopes reconnus par HPA

---

Depuis plus de vingt ans, la lectine HPA, extraite de l'escargot de Bourgogne, est connue pour sa capacité à reconnaître spécifiquement certains glycotopes à la surface d'adénocarcinomes métastatiques (particulièrement ceux du sein, du colon et du poumon), associés à un très faible pronostic vital. Cette lectine est donc un outil de choix dans le développement de nouveaux marqueurs histopathologiques.

A ce jour, les molécules reconnues par HPA restent toujours à identifier. En 2007, seize glycoprotéines membranaires ont été identifiées sur une lignée de cancer colorectal fortement métastatique, après purification par chromatographie d'affinité sur colonne HPA-sepharose (Saint-Guirons *et al.* 2007). Il s'agirait d'intégrines, d'annexines, de cytokératines, d'actine et de tubuline et seraient notamment impliquées dans les phénomènes d'angiogenèse, d'apoptose, de motilité cellulaire et dans la formation des métastases.

La partie oligosaccharidique des glycoconjugués reconnues par HPA est également toujours inconnue à ce jour. Son identification serait pourtant primordiale pour la compréhension des processus de cancérisation et de formation des métastases. En 2001, une des structures reconnues par HPA a été caractérisée pour la première fois, de manière relativement imprécise, dans des lignées de cancers du sein. Nommée HPA glycan 1 ou HPAgly1, elle serait formée de quatre unités monosaccharidiques neutres et d'un acide sialique (Dwek *et al.* 2001).

La spectrométrie de masse est devenue une technique de pointe dans l'étude des glycanes associés aux glycoconjugués et notamment aux protéines, grâce à son importante sensibilité et à sa haute résolution (Harvey 2005 ; Geyer and Geyer 2006). Cette nouvelle discipline est appelée glycoprotéomique. Lors d'une collaboration avec Dr. Bérange Tissot, dans le laboratoire du Pr. Anne Dell à l'Imperial College et au cours de deux courtes sessions à Londres, j'ai tenté d'identifier le ou les motif(s) glycanique(s) reconnu(s) par HPA sur les lignées cellulaires de cancers colorectaux métastatiques (lignée SW 620 et HT 29) ou non (lignée SW 480). L'analyse par spectrométrie de masse des *N*- et *O*-glycanes extraits des

membranes cellulaires avant et après purification par chromatographie d'affinité sur colonne HPA-sepharose devait permettre d'identifier, de manière plus précise, les glycanes reconnus par la lectine.

## 1. Généralités

Dans un spectromètre de masse, les échantillons sont séparés en fonction de leur rapport masse/charge ( $m/z$ ), exprimé en Thomson (Th). Si la source d'ionisation donne majoritairement naissance à des ions monochargés, ce rapport est équivalent à la masse de la molécule. Pour une meilleure stabilité des échantillons, notamment ceux contenant des résidus d'acides sialiques, et sensibilité de l'appareil, les glycanes étudiés sont généralement perméthylés avant analyse, c'est-à-dire qu'un groupement méthyle vient substituer l'atome d'hydrogène des hydroxyles libres du sucre. La méthode a été développée par (Ciucanu 1984), puis modifiée par (Ciucanu 2003). Le rapport  $m/z$  moyen de chaque monosaccharide est repris dans le **Tableau 1**. Les résidus possédant le même nombre d'atomes (par exemple Glc, Gal et Man ou GlcNAc et GalNAc) ont donc un rapport masse/charge identique. Lors de l'analyse primaire des spectres, la différenciation entre les résidus de même poids moléculaire est impossible et l'attribution des différentes structures oligosaccharidiques est basée sur les règles connues de biosynthèse des glycanes. L'oligosaccharide peut également être désigné selon le nombre de chaque résidu de rapport masse/charge différent présent dans l'échantillon. Ainsi, certains pics seront désignés comme suit  $\text{Hex}_x\text{HexNAc}_y\text{NeuAc}_z$  et représentent des oligosaccharides contenant  $x$  hexoses,  $y$  hexoses *N*-acétylés et  $z$  acides neuraminiques. La différence de masse entre deux ions d'une même série correspond à l'incrément de masse d'un monosaccharide Hex, HexNAc ou NeuAc et permet de renseigner sur la séquence primaire en monosaccharides de l'échantillon.

Le glycanes le plus abondant dans l'échantillon est représenté par le pic le plus grand et présente une intensité relative de 100%. L'intensité des autres pics est déterminée en comparaison avec ce pic majoritaire. La spectrométrie de masse n'est pas une méthode quantitative puisque l'intensité du signal dépend de l'état d'ionisation et de la taille de la molécule. Il n'est donc pas rigoureusement exact de comparer l'intensité de deux pics présents sur le même spectre. Néanmoins, l'efficacité d'ionisation de la source MALDI étant

relativement homogène, les intensités relatives de chaque pic sur un même spectre donnent un ordre d'idées de l'abondance de chaque structure dans l'échantillon.

Symbole	Nom	m/z (Th)	m/z per (Th)
	Fucose (Fuc)	146	174
	Mannose (Man)	162	204
	Glucose (Glc)	162	204
	Galactose (Gal)	162	204
	N-acétylglucosamine (GlcNAc)	203	245
	N-acétylgalactosamine (GalNAc)	203	245
	Acide N-acétylneuraminique (NeuAc)	291	361

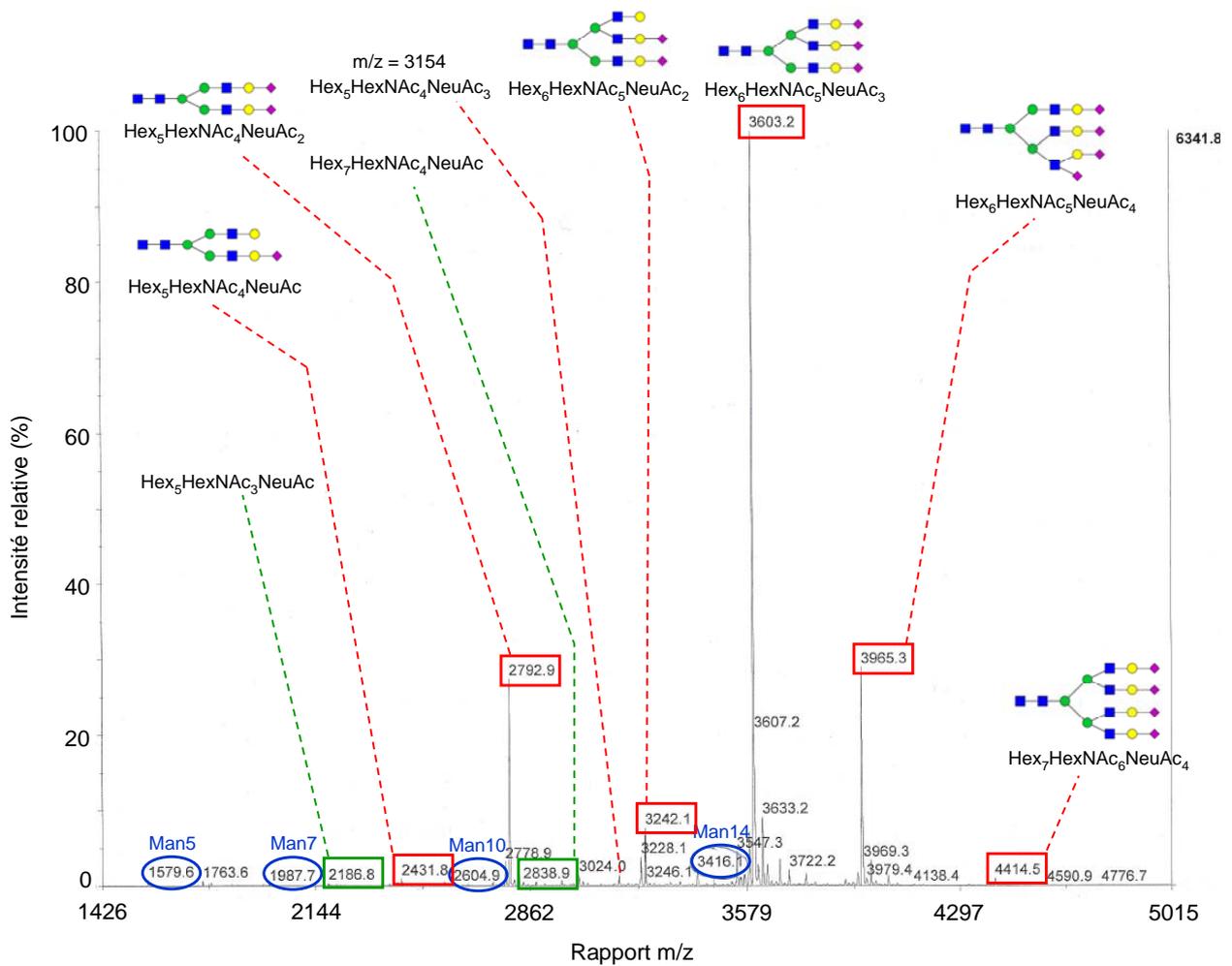
**Tableau 1. Poids moléculaire des principaux résidus glycaniques avant et après perméthylation.**

## 2. Analyse du milieu de culture

Les trois lignées cellulaires utilisées comme modèle de tissus cancéreux sont issues de cancers colorectaux humains. La croissance cellulaire est généralement réalisée dans un milieu complet contenant notamment 10% de sérum de veau fœtal. Ce dernier est composé d'un mélange de protéines (albumine, anticorps...), de minéraux, de sucres, de lipides, de facteurs de croissance et substances cytoestimulantes qui activent les divisions cellulaires. Sa composition détaillée reste néanmoins à préciser et varie suivant le fournisseur (Bryan *et al.* 2010). Certains des composés (protéines et lipides) sont glycosylés et peuvent alors interférer, lors de l'analyse glycomique, avec les glycanes isolés à partir des glycoprotéines membranaires.

La **Figure 1** représente le spectre des N-glycanes présents dans le sérum de veau fœtal déterminé par MALDI. La majorité des pics correspondent à des structures de type complexe, très fréquemment rencontrés dans les cellules de vertébrés (Varki 1999). Les trois pics

principaux, de rapport m/z équivalent à 2793, 3242 et 3603 correspondent à des structures contenant un noyau pentasaccharidique  $\text{Man}_3\text{GlcNAc}_2$  duquel s'étendent deux ou trois antennes *N*-acétyllactosaminiques terminées ou non par des acides sialiques. Le pic avec un rapport m/z de 3965 correspond à la structure triantennée de rapport m/z égal à 3603 sur laquelle un second acide sialique lié au GlcNAc de la chaîne lactosaminique est présent. Cette structure a été confirmée dans (Morelle *et al.* 2009). Le pic avec un rapport m/z de 3154 et de formule  $\text{Hex}_5\text{HexNAc}_4\text{NeuAc}_3$  correspond également probablement à ce type de structure mais avec deux antennes au lieu de trois.



**Figure 1. Analyse, par spectrométrie de masse, des *N*-glycane extraits du sérum de veau fœtal commercialisé par Sigma-Aldrich.**

Les valeurs entourées en bleu ou encadrées en rouge et en vert correspondent respectivement à des structures de type oligomannosidique, complexe et probablement hybride.

Des structures de type oligomannosidique sont également visibles sur ce spectre. Les pics de rapport  $m/z$  de 1580, 1988, 2605 et 3416 correspondent à des structures comportant respectivement 5, 7, 10 et 14 résidus de mannose. Les pics de rapport  $m/z$  égal à 2187 et 2839 correspondent à des structures de formule  $\text{Hex}_5\text{HexNAc}_3\text{NeuAc}$  et  $\text{Hex}_7\text{HexNAc}_4\text{NeuAc}$ , respectivement, probablement de type hybride.

L'analyse des *N*-glycanes de la fœtuine de veau, composant majeur du sérum de veau fœtal, a été réalisée par spectrométrie de masse MALDI-TOF (Morelle *et al.* 2009). Le spectre est composé de cinq pics de rapport  $m/z$  équivalents à 2431, 2792, 3241, 3603 et 3964. Les intensités relatives des glycanes les uns par rapport aux autres sont similaires à ce qui est observé dans la Figure 1. Cependant, aucune structure de type oligomannosidique et hybride n'est représentée. Très faiblement présentes, elles se confondent probablement avec le bruit de fond. Ces résultats confirment néanmoins l'allure du spectre présenté dans la figure 1.

Les pics précédents des *N*-glycanes rencontrés dans le sérum de fœtal, sont des composantes très minoritaires des *N*-glycanes membranaires identifiés à la surface des trois lignées cancéreuses étudiées. Néanmoins, afin de s'affranchir de tous problèmes de contamination des spectres, les cellules ont été entretenues dans un milieu ne contenant pas de sérum de veau fœtal.

### **3. Analyse des *N*-glycanes des membranes cellulaires**

La culture des lignées cellulaires de cancers colorectaux SW 480, SW 620 et HT 29 a été réalisée par Dr Miriam Dwek de l'université de Westminster, en l'absence de sérum de veau fœtal. Une description plus complète de ces cellules a été réalisée dans les chapitres 2 et 3.

Après extraction des protéines membranaires par sonication et centrifugation, les *N*-glycanes sont isolés, perméthylés et analysés par spectrométrie de masse MALDI. Préalablement, une partie d'entre eux est directement purifiée par chromatographie d'affinité sur colonne HPA-sepharose afin d'isoler les *N*-glycanes reconnus par cette lectine.

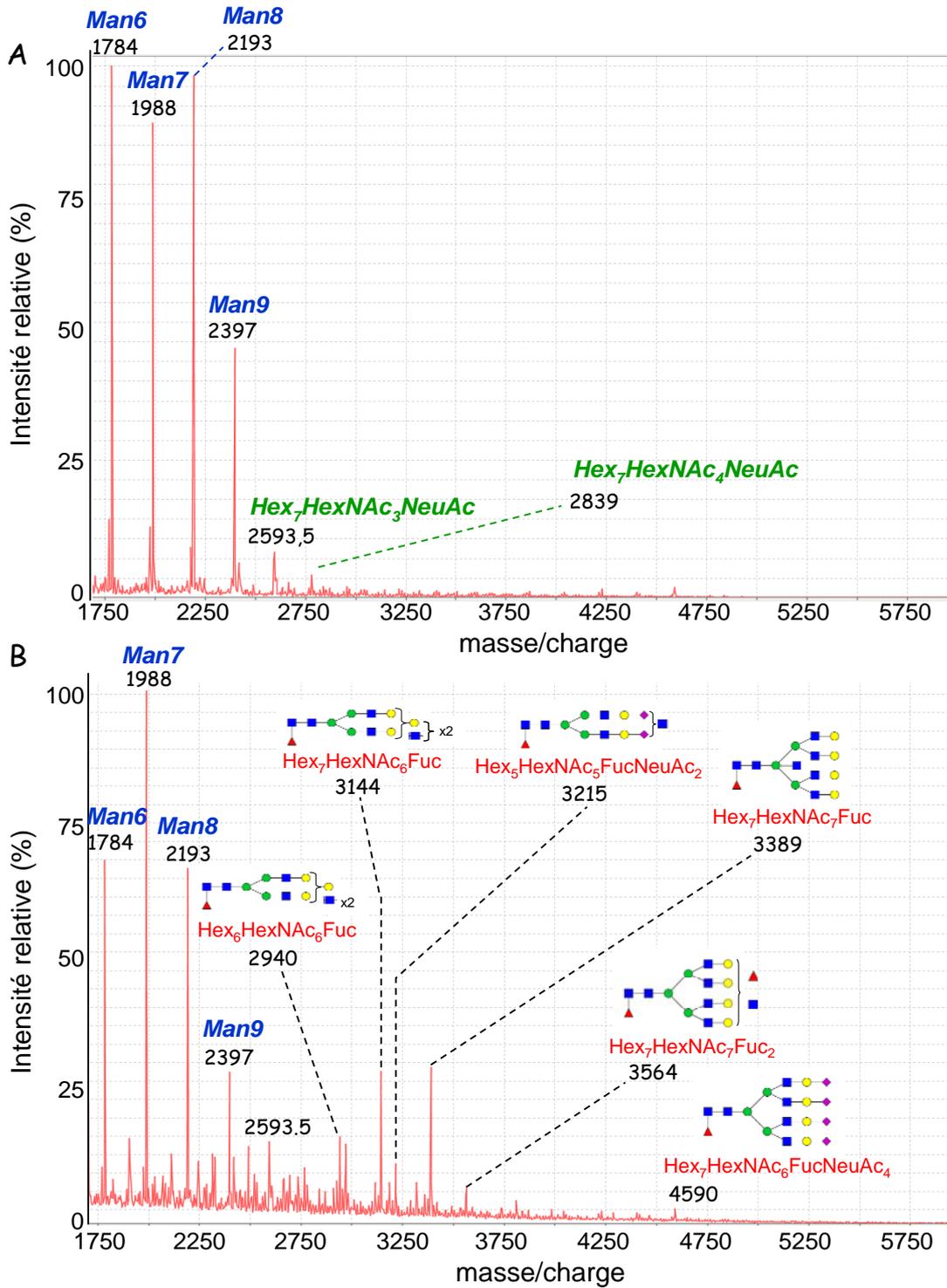
Tous les spectres obtenus, quelque soit la lignée cellulaire considérée, présentent des pics correspondant à des structures de type oligomannosidique ( $\text{Hex}_{6-9}\text{HexNAc}_2$ , de rapport  $m/z$  équivalent à 1784, 1988, 2193, 2397, contenant respectivement 6, 7, 8 et 9 résidus

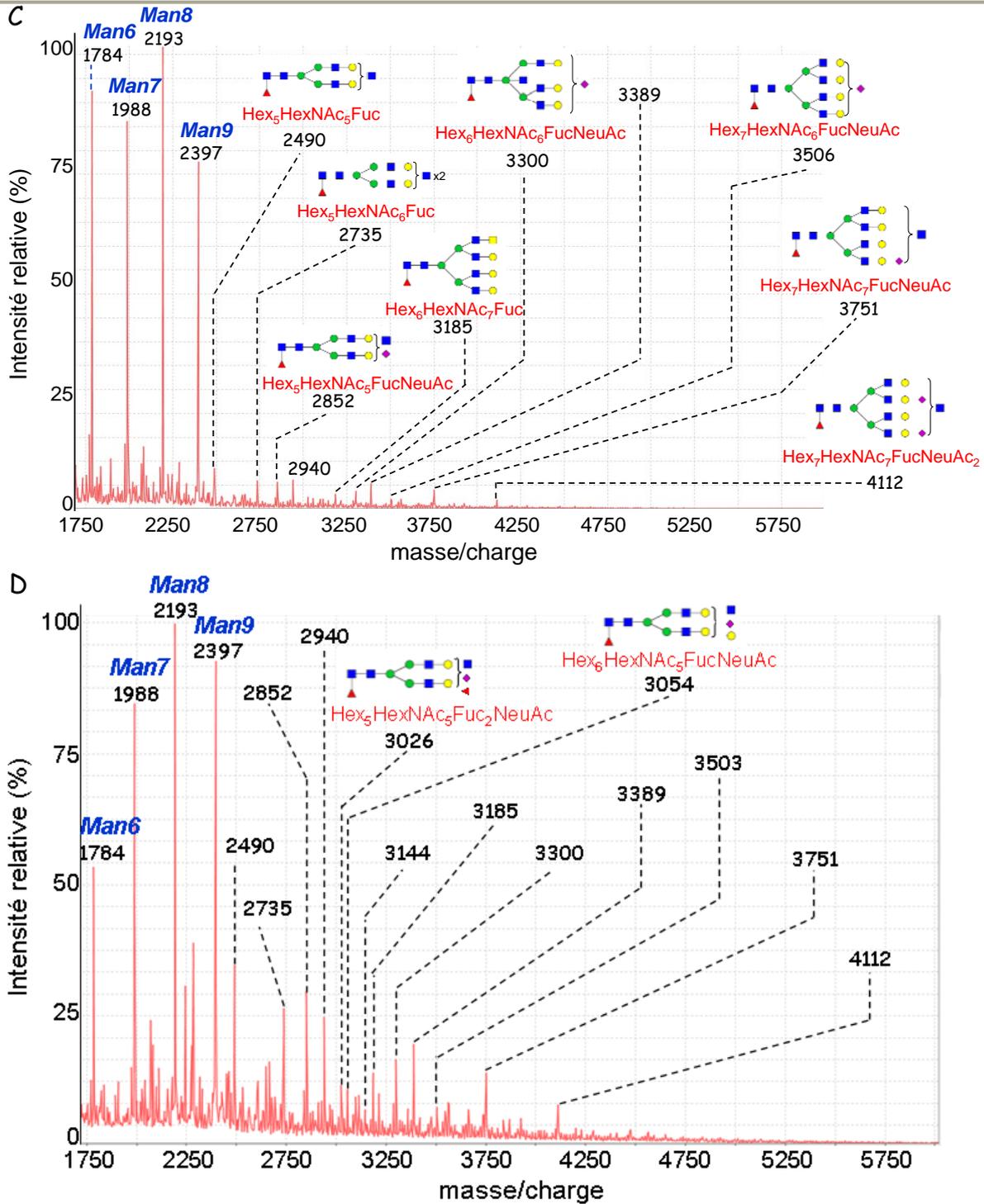
mannose). Sont également présents des pics correspondant à des structures de type hybride ( $\text{Hex}_7\text{HexNAc}_{3-4}\text{NeuAc}$ ) et de type complexe ( $\text{Hex}_{5-7}\text{HexNAc}_{3-7}\text{Fuc}_{0-2}\text{NeuAc}_{0-4}$ ).

### 3.1. Lignée cellulaire non métastatique SW 480

Avant l'étape de purification par chromatographie d'affinité sur colonne HPA-sepharose, le nombre de *N*-glycanes extraits des glycoprotéines membranaires de la lignée SW 480 est relativement faible (**Figure 2. A.**). Les pics les plus abondants correspondent aux structures hautement mannosylées décrites dans le paragraphe précédent. Deux pics de rapport *m/z* de 2593,5 et 2839 sont également visibles sur le spectre. Ils sont respectivement attribués aux structures suivantes  $\text{Hex}_7\text{HexNAc}_3\text{NeuAc}$  et  $\text{Hex}_7\text{HexNAc}_4\text{NeuAc}$ , probablement de type hybride.

Après l'étape de purification, les pics correspondant aux structures hautement mannosylées restent toujours majoritaires (**Figure 2. B.**). La présence de nouveaux pics à des rapports *m/z* supérieurs témoignent d'un enrichissement de l'échantillon par de nouvelles structures. Les pics à 2940 et 3144 correspondent à des *N*-glycanes biantennés de type complexe sur lesquels les chaînes poly *N*-acétyllactosaminiques s'allongent par ajout de GlcNAc et Gal successivement. En effet, la différence de rapport *m/z* entre 3114 et 2940 est égale à 204, masse correspondant à un hexose. Les différentes étapes des voies de biosynthèse laissent présager qu'il s'agit d'un résidu Gal qui sera lié à un GlcNAc. Ces structures peuvent également présenter un nombre d'antennes plus important. Par exemple, le pic de rapport *m/z* égal à 4590 correspond à une structure tetraantennée terminée par quatre résidus d'acides sialiques. Il est intéressant de noter la présence de fucose de cœur lié en  $\alpha$ 2-6 sur le premier GlcNAc ou de GlcNAc bisecté, lié sur le mannose de cœur, sur les structures présentant un rapport *m/z* équivalents à 3564 et 3389 respectivement. Cette dernière a également été déterminée dans la lignée très fortement métastatique du colon HT 29 par l'équipe du Dr Michalski (Vercoutter-Edouart *et al.* 2008).





**Figure 2.** Analyse, par spectrométrie de masse, des *N*-glycane extraits des glycoprotéines membranaires des lignées de cancer colorectal non métastatique SW 480 et métastatique SW 620.

Lignée SW 480 avant (A.) et après (B.) l'étape de purification par chromatographie d'affinité sur colonne HPA-agarose. C. et D. Identique à A. et B. avec la lignée SW 620.

Les valeurs au dessus des pics correspondent au rapport masse/charge du pic. La structure de l'oligosaccharide correspondant au pic est écrite en gras et italique. Les carrés bleus, les ronds verts et jaunes, les triangles rouges et les losanges violets schématisent respectivement les résidus GlcNAc, Man, Gal, Fuc et NeuAc. Les images ont été réalisées sur la plateforme GlycoWork Bench (Ceroni *et al.* 2007).

### 3.2. La lignée métastatique SW 620

Le spectre de l'échantillon SW 620 avant l'étape de purification (**Figure 2. C.**) présente plus de pics que celui obtenu avec la lignée SW 480. L'échantillon est donc globalement plus riche en oligosaccharides que le précédent. Un dosage de Bradford révèle que la concentration en protéines de la lignée SW 620 ( $c = 6,62 \text{ mg.ml}^{-1}$ ) est environ six fois supérieur à celle déterminée pour la lignée SW 480 ( $c = 1,23 \text{ mg.ml}^{-1}$ ). La présence de pics plus nombreux sur le spectre SW 620 par rapport à celui de la lignée SW 480 peut donc d'expliquer par une quantité de matériel de départ plus importante et ne peut, à ce stade de l'étude, pas être imputée à des variations de glycosylation provoquées par la formation de métastases.

Les structures glycaniques déterminées dans la lignée SW 620 sont sensiblement équivalentes à celles identifiées dans la lignée SW 480. Les structures hautement mannosylées restent toujours très fortement majoritaires. Sont également présents des *N*-glycanes de type complexe bi- et tetraantennés principalement, terminées par des résidus Gal, GlcNAc et NeuAc. Les pics de rapport  $m/z$  égal à 3300, 3389, 3751 et 4112 correspondent à des structures possédant un GlcNAc bisecté, toutes retrouvées par l'équipe du Dr. Michalski dans la lignée HT 29 (Vercoutter-Edouart *et al.* 2008). Le détail de chaque structure est annoté sur le spectre (**Figure 2.**).

De manière similaire à la lignée SW 480, un enrichissement des structures est également observé après l'étape de purification pour la lignée SW 620 (**Figure 2. D.**). Deux nouvelles structures (rapport  $m/z$  de 3026 et 3054) apparaissent, formées d'oligosaccharides bi-antennés, comportant probablement des chaînes poly-*N*-acétyllactosaminiques, terminées par des acides sialiques. La structure de rapport  $m/z$  égal 3026 présente également un deuxième résidu Fuc.

### 3.3. La lignée très métastatique HT 29

De manière similaire aux lignées précédentes, les *N*-glycanes de la lignée HT 29 ont également été étudiés. Malheureusement, les spectres obtenus sont inexploitable. Un défaut partiel de perméthylation rend l'attribution des pics très difficiles puisque les rapports  $m/z$  obtenus ne correspondent à aucune structure de *N*-glycanes connue. De plus, pour le spectre

déterminé après l'étape de purification, le manque de matériel déforme la ligne de base. Un problème est probablement survenu lors de l'étape de purification. Cette conclusion a également été vérifiée par Dot Blot. Cette technique, similaire à la révélation par immunoblot, consiste à déposer une goutte d'échantillon sur une membrane de nitrocellulose et à révéler les échantillons d'intérêt par des anticorps ou lectines appropriés. Dans notre cas, 15 µg de protéines, dosées par la méthode de Bradford, et contenues dans une goutte de 2 µl sont déposés sur membrane de nitrocellulose. Après évaporation complète du tampon et saturation de la membrane par une solution de BSA à 5%, la révélation débute par la fixation de la lectine HPA biotinylée. Puis, le complexe streptavidine-peroxydase permettra, après ajout d'un substrat coloré le DAB, d'observer une coloration si des glycanes reconnus par HPA sont présents dans l'échantillon. Cette technique a été réalisée pour les trois échantillons après purification. Une coloration est bien visible pour les échantillons des lignées SW 480 et SW 620 mais la membrane reste blanche pour la lignée HT 29. Or, son marquage détecté par microscopie confocale prouve qu'elle possède de nombreux glycoépitopes reconnus par HPA, tout comme la lignée SW 620 (Saint-Guirons *et al.* 2007). Un problème a donc été rencontré lors de l'étape de purification.

## 4. Conclusions

L'analyse des *N*-glycanes extraits des lignées de cancer colorectal SW 480 et SW 620 démontre que les structures majoritaires sont de types oligomannosidique et complexe, avec, dans ce dernier cas, une prévalence des structures poly-*N*-acétyllactosaminiques. Ces caractéristiques ont également été observées pour le cancer du sein chez la souris et/ou l'homme (de Leoz *et al.* 2010 ; Ito *et al.* 1996). La synthèse des *N*-glycanes semble donc être un phénomène impliqué et perturbée dans les processus de transformation tumorale et de formation des métastases.

HPA, lectine spécifique du GalNAc est, depuis vingt ans, identifiée comme marqueur de cancers métastatiques. Or, après purification des *N*-glycanes des lignées SW 480 et SW 620, seule une structure terminée par un résidu GalNAc est identifiée. La majorité des oligosaccharides présentent des Gal, GlcNAc et NeuAc comme résidu terminal. Même si un large spectre de reconnaissance a été mis en évidence pour cette lectine (Brooks and Leatham 1998 ; Brooks *et al.* 2001 ; Sanchez *et al.* 2006 ; Saint-Guirons *et al.* 2007), nous savons

qu'HPA reconnaît uniquement la lignée métastatique SW 620. Ces résultats ont notamment été présentés dans le chapitre 3. L'enrichissement, après purification par chromatographie d'affinité sur colonne HPA-sepharose, de la lignée SW 480 ainsi que la détection par Dot Blot laisse envisager que la purification n'était peut-être pas spécifique.

La majorité des structures isolées dans la lignée SW 480 est également retrouvée dans les lignées SW 620 et HT 29 par l'équipe du Dr Michalski (Vercoutter-Edouart *et al.* 2008). Néanmoins, certaines structures semblent n'être présentes que dans la lignée SW 620 (par exemple, celles de rapport  $m/z$  égal à 2852, 3026, 3300, 3751, 4111). La concentration de glycoprotéines membranaires, six fois supérieure dans la lignée SW 620 par rapport à la SW 480, explique probablement cette abondance de pics plutôt qu'un dérèglement de glycosylation provoqué par la formation de métastases.

Les résidus GalNAc en position terminale sont rarement observés à la surface des cellules saines. Ils composent généralement les structures de cœur des *O*-glycanes et sont donc recouverts par d'autres monosaccharides. En revanche, ils sont bien plus abondants à la surface de cellules cancéreuses, mais principalement sous forme de *O*-glycanes, comme par exemple dans l'antigène Tn (Springer 1984). L'analyse des *O*-glycanes isolés à la surface des protéines membranaires des lignées de cancer colorectal avait également été réalisée mais les spectres obtenus étaient inexploitable.

## 5. Perspectives

La purification est une étape cruciale de l'expérience qu'il convient d'optimiser. Dans notre cas, les purifications ont été réalisées manuellement sur une colonne HPA-sepharose formée à partir d'une colonne activée par du NHS sur lequel a été directement couplé la lectine (Saint-Guirons *et al.* 2007). Cette colonne avait déjà été utilisée auparavant. De plus, les trois purifications ont eu lieu de manière successive, sur la même colonne dans l'ordre suivant : la lignée SW 480 puis la SW 620 pour terminer avec la HT 29. Les résultats aberrants de DotBlot (purification de glycanes de la lignée SW 480 par HPA et absence de reconnaissance pour la lignée HT 29) ainsi que l'absence de signal exploitable pour l'analyse des *N*-glycanes de la lignée HT 29 laisse supposer l'introduction d'erreurs lors de cette étape. De plus, cette étape a été réalisée directement après la lyse cellulaire. Il serait peut-être plus judicieux de réaliser cette purification après l'étape de clivage des *N*-glycanes à la PNGase F

lorsque l'échantillon est déjà majoritairement débarrassé des autres glycanes et de tous les organites et contaminants cellulaires. Une purification en batch avec de la résine HPA-agarose en solution pourrait également être testée.

Cette étude des *N*-glycanes des trois lignées cancéreuses du colon est très préliminaire. Les résultats obtenus soulèvent de nombreuses interrogations. Afin de tenter d'approfondir ces recherches, plusieurs pistes expérimentales peuvent être testées. Tout d'abord, dans cette étude, afin de ne pas « contaminer » le spectre obtenu par les *N*-glycanes du milieu de culture, les cellules ont été mises en croissance dans un milieu ne contenant pas de sérum de veau fœtal. Nous sommes donc en mesure de nous interroger sur les conséquences de ce manque tant sur la croissance cellulaire que sur l'expression des glycanes membranaires. En effet, si les cellules évoluent dans un milieu pas parfaitement adaptée, sont-elles capables d'induire des modifications dans l'expression de leurs *N*-glycanes membranaires ? Une étude rigoureusement similaire avec des cellules cultivées en milieu complet doit être menée afin de comparer les résultats et mesurer l'influence d'une perturbation de milieu de culture des cellules.

De plus, un effort important doit être apporté à l'étude des *O*-glycanes qui portent généralement des GalNAc terminaux à la surface de cellules cancéreuses et qui sont donc des ligands potentiels d'HPA.

# Article

---



# Discoidin I from *Dictyostelium discoideum* and Interactions with Oligosaccharides: Specificity, Affinity, Crystal Structures, and Comparison with Discoidin II

Sophie V. Mathieu<sup>†</sup>, Karoline Saboia Aragão<sup>†</sup>, Anne Imberty and Annabelle Varrot<sup>\*</sup>

CERMAV-CNRS (affiliated with University Joseph Fourier and ICMG), 601 rue de la Chimie, BP53, F-38041 Grenoble Cedex 09, France

Received 24 March 2010;  
received in revised form  
17 May 2010;  
accepted 17 May 2010  
Available online  
24 May 2010

Discoidin I (DisCI) and discoidin II (DisCII) are *N*-acetylgalactosamine (GalNAc)-binding proteins from *Dictyostelium discoideum*. They consist of two domains: an N-terminal discoidin domain and a C-terminal H-type lectin domain. They were cloned and expressed in high yield in recombinant form in *Escherichia coli*. Although both lectins bind galactose (Gal) and GalNAc, glycan array experiments performed on the recombinant proteins displayed strong differences in their specificity for oligosaccharides. DisCI and DisCII bind preferentially to Gal/GalNAc $\beta$ 1–3Gal/GalNAc-containing and Gal/GalNAc $\beta$ 1–4GlcNAc $\beta$ 1–6Gal/GalNAc-containing glycans, respectively. The affinity of the interaction of DisCI with monosaccharides and disaccharides was evaluated using isothermal titration calorimetry experiments. The three-dimensional structures of native DisCI and its complexes with GalNAc, GalNAc $\beta$ 1–3Gal, and Gal $\beta$ 1–3GalNAc were solved by X-ray crystallography. DisCI forms trimers with involvement of calcium at the monomer interface. The N-terminal discoidin domain presents a structural similarity to F-type lectins such as the eel agglutinin, where an amphiphilic binding pocket suggests possible carbohydrate-binding activity. In the C-terminal H-type lectin domain, the GalNAc residue establishes specific hydrogen bonds that explain the observed affinity ( $K_d = 3 \times 10^{-4}$  M). The different specificities of DisCI and DisCII for oligosaccharides were rationalized from the different structures obtained by either X-ray crystallography or molecular modeling.

© 2010 Elsevier Ltd. All rights reserved.

Edited by I. Wilson

**Keywords:** H-type lectin; F-type lectin; discoidin domain; adhesion; oligosaccharide

## Introduction

Lectins are ubiquitous proteins of nonimmune origin that interact reversibly and specifically with carbohydrates without modifying them. A large panel of lectins has been identified in invertebrates, and considerable attention is focused on their roles in biological recognition. Invertebrate lectins are often involved in specific binding to bacterial polysaccharides, playing a role in the establishment of symbiosis or in innate immunity.<sup>1–3</sup> In other cases, they are involved in self-recognition and cell–cell aggregation, which are required for the building of multicellular organisms such as corals and sponges.<sup>4</sup> They can be useful tools in cancer diagnosis/prognosis as histochemical markers or in cancer therapy, thanks to antitumoral activities.<sup>5–7</sup>

<sup>\*</sup>Corresponding author. E-mail address: varrot@cermav.cnrs.fr.

<sup>†</sup>S.V.M. and K.S.A. contributed equally to this work.

Present address: K. S. Aragão, Laboratório de Moléculas Biologicamente Ativas, Universidade Federal do Ceará, Caixa Postal 6043, Campus do Pici, 60455-970 Fortaleza CE, Brazil.

Abbreviations used: DisCI, discoidin I; DisCII, discoidin II; GalNAc, *N*-acetylgalactosamine; Gal, galactose; DS domain, discoidin domain; GlcNAc, *N*-acetylglucosamine;  $\beta$  MeGal,  $\beta$ -methyl-galactoside; MPD, 2-methyl-2,4-pentanediol; HPA, *Helix pomatia* agglutinin; CBM, carbohydrate-binding module; PDB, Protein Data Bank; ITC, isothermal titration calorimetry; GPS, Gal/GalNAc-rich polysaccharide.

The slime mold *Dictyostelium discoideum* grows as free-living amoebae in the soil, feeding on bacteria. Upon starvation, it undergoes a complex developmental cycle in which about 10,000–50,000 individual amoebae aggregate to form a multicellular fruiting body that is able to produce spores. The aggregation of individual amoeba occurs by chemotaxis to periodic cAMP signals (reviewed by Kessin<sup>8</sup>). During differentiation, numerous new proteins are synthesized, and some of the most abundant of these new products are the two *N*-acetylgalactosamine (GalNAc)-binding proteins discoidin I (DisCI) and discoidin II (DisCII).<sup>9,10</sup> They are virtually undetectable in cells growing on bacteria, but constitute over 1% of cellular proteins in aggregated cells.<sup>9</sup> They present overlapping but distinct sugar specificities.<sup>10</sup> The production of DisCI is prominent in aggregating cells, accumulating intracellularly and upon externalization in multilamellar bodies, while DisCII is prominent during fruiting body formation and is localized in pre-spore vesicles.<sup>11,12</sup> The precise function of these lectins, as well as their biological ligands, remains unknown.

DisCI is polymorphic, with at least three isoforms per strain coded by genes, forming a small, coordinately regulated multigene family.<sup>13,14</sup> The three genes *dscA*, *dscC*, and *dscD* are found on chromosome 2 and are duplicated on the same chromosome in some laboratory strains (AX3 and AX4). They are often used as markers of early development in *D. discoideum*.<sup>15,16</sup> DisCI expression is developmentally regulated at the transcriptional level by several signal transduction pathways involving cAMP, prestarvation factor, and conditioned medium factor.<sup>17–19</sup> Several mutants and antisense transformant analysis showed that DisCI was apparently not necessary for aggregation, since mutants plated at high density can aggregate and form fruiting bodies.<sup>17</sup> The lack of gene product impairs the process of cell streaming in early development, where it seems important in the formation of head-to-tail streams by aggregating cells.<sup>20</sup> DisCI was also shown to be involved in cell substratum attachment and ordered cell migration during aggregation by a fibronectin-like mechanism.<sup>21</sup> It contains the RGD motif, which is the cell attachment site found in a large collection of adhesive proteins.<sup>22,23</sup> The proposed receptor for this RGD sequence is a 67-kDa developmentally regulated cell surface glycoprotein.<sup>24</sup> DisCI also recognizes glycoconjugates that contain GalNAc in the slime coat around aggregates and in multilamellar bodies, since the interaction is completely blocked in the presence of GalNAc.<sup>25–27</sup> The sugar binding site was shown to be different from the cell adhesion site and requires a divalent cation.<sup>24,28</sup>

DisCI is a two-domain protein of 253 amino acids with a higher affinity for GalNAc than for galactose (Gal).<sup>10</sup> It displays a 48% sequence identity with DisCII, whose trimeric structure has been recently solved in unliganded state and in complex with Gal and GalNAc.<sup>29</sup> The N-terminal domain, referred to

as the discoidin domain (DS domain), is a structural and functional motif found in various proteins from both eukaryotic and prokaryotic origins. This domain shows considerable functional diversity via interactions with a wide range of molecules and is mainly involved in cell-surface-mediated regulatory events and glycoconjugates binding.<sup>30–32</sup> The C-terminal domain belongs to the H-type lectin family recently identified in invertebrates such as snails and corals.<sup>33</sup>

We present here the specificity and affinity of DisCI and DisCII for a variety of carbohydrate ligands, as well as the crystal structures of DisCI in unliganded form and in complex with GalNAc, GalNAc $\beta$ 1–3Gal, and Gal $\beta$ 1–3GalNAc. The specificities, affinities, structures, and binding sites of both discoidins are compared and discussed. DisCI and DisCII present interesting differences in terms of oligosaccharide binding, which may be correlated to their differences in localization and their role in the slime mold.

## Results

### Specificity of *D. discoideum* lectins with glycan microarray resource

Recombinant DisCI was obtained with a yield of 20 mg l<sup>-1</sup>, and its molecular weight, checked by mass spectrometry, corresponded to the expected value. Recombinant DisCII was produced as previously described.<sup>29</sup> Both lectins were labeled with Alexa 488 and assayed on Glycan Microarray available at the Consortium for Functional Glycomics, using concentrations varying from 0.01 to 200  $\mu$ g ml<sup>-1</sup>. DisCI gave excellent signals at very low concentrations, and the analysis displayed in Fig. 1a has been obtained for 0.04  $\mu$ g ml<sup>-1</sup>. Strong binding to glycochips was also observed for DisCII, and Fig. 1b has been obtained with 0.5  $\mu$ g ml<sup>-1</sup>. These concentrations were selected to optimize the quality of the figure. Both proteins gave a signal that increased parallelly with the concentration and reached a maximum at approximately 0.5–1  $\mu$ g ml<sup>-1</sup>.

Quite surprisingly, there is almost no superimposition in the binding specificity of the two discoidins. When looking at disaccharides, DisCI binds to several  $\beta$ 1–3-linked disaccharides with a galactose residue at the reducing position, such as GlcNAc $\beta$ 1–3GalNAc (position 159) and Gal $\beta$ 1–3GalNAc (position 123), and binds less strongly to GalNAc $\beta$ 1–3Gal (position 300). Elongation by additional residues or by sulfate groups is well tolerated in position 3 or 4 of nonreducing sugar, while reducing sugar can be only elongated at position 1, and no branching is accepted. Polylactosamine oligosaccharides containing repetitive Gal $\beta$ 1–4GlcNAc $\beta$ 1–3 sequences are high-affinity ligands. In a few cases, such as in ganglioside of the GM2 to GT2 series (positions 201–209), the terminal GalNAc is also recognized.



**Table 1.** Microcalorimetry data

Protein	Ligand	$K_d$ ( $\mu$ M)	$n$	$-\Delta G$ (kJ mol <sup>-1</sup> )	$-\Delta H$ (kJ mol <sup>-1</sup> )	$-T\Delta S$ (kJ mol <sup>-1</sup> )
DiscI	GalNAc	302 ± 27	1 <sup>a</sup>	20.0 ± 0.2	33.4 ± 5.8	13.3 ± 6.0
	Gal	3770 ± 200	1 <sup>a</sup>	13.8 ± 0.1	31 ± 7.4	17.1 ± 7.5
	$\alpha$ -Methyl-galactoside	8390 ± 700	1 <sup>a</sup>	11.8 ± 0.2	42.4 ± 5.9	30.5 ± 6.1
	$\beta$ MeGal	2960 ± 270	1 <sup>a</sup>	14.4 ± 0.2	40.3 ± 2.4	25.9 ± 2.2
	Gal $\beta$ 1-4GlcNAc (LacNAc)	2390 ± 170	1 <sup>a</sup>	15.0 ± 0.2	38.4 ± 2.2	23.3 ± 2.0
	Gal $\beta$ 1-3GalNAc	27.2 ± 4.6	1.03 ± 0.01	26.1 ± 0.4	45.6 ± 1.1	19.5 ± 1.5
	GalNAc $\beta$ 1-3Gal	49.4 ± 1.5	0.81 ± 0.01	24.5 ± 0.3	53.2 ± 2.9	28.81 ± 2.8
DiscII	GalNAc <sup>b</sup>	1115	1 <sup>a</sup>	16.8	20.3	3.5
	$\beta$ MeGal <sup>b</sup>	952	1 <sup>a</sup>	17.4	24.3	6.7
	Gal $\beta$ 1-4GlcNAc (LacNAc)	1221 ± 23	1 <sup>a</sup>	16.7 ± 0.5	37.0 ± 0.3	20.5 ± 0.4
	Gal $\beta$ 1-3GalNAc <sup>c</sup>	1377	1 <sup>a</sup>	16.3	42.3	26.0
	GalNAc $\beta$ 1-3Gal <sup>c</sup>	1113	1 <sup>a</sup>	16.8	45.6	28.8

All values are averaged over two or three experiments. The reported standard deviations are calculated upon averaging of those experiments, unless otherwise specified.

<sup>a</sup> Fixed stoichiometry during the fitting procedure for low-affinity interaction.

<sup>b</sup> Taken from Saboia Aragão *et al.*<sup>29</sup>

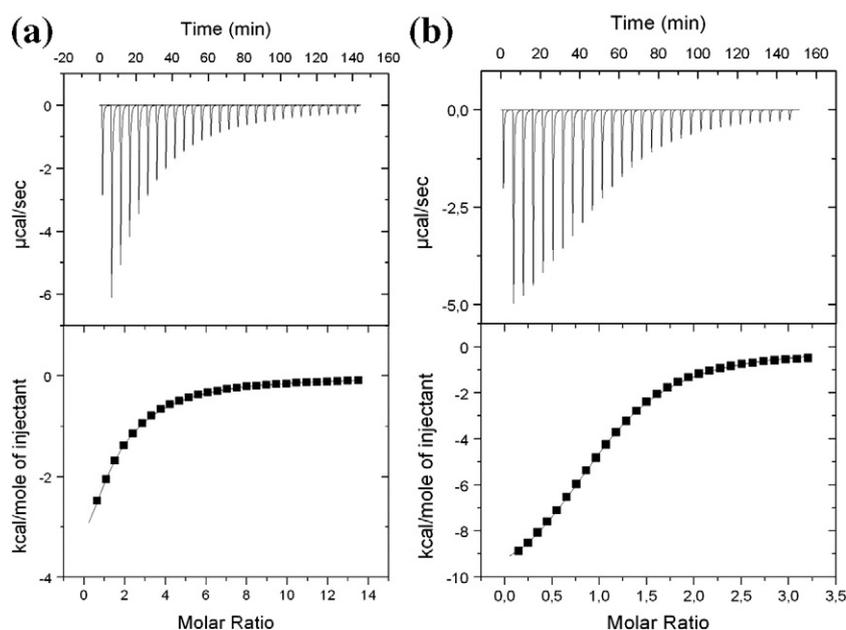
<sup>c</sup> Only one experiment due to lack of material.

In contrast, DiscII does not recognize any disaccharide efficiently. GlcNAc $\beta$ 1-6Gal/GalNAc is the preferred motif, but it is recognized only as part of larger oligosaccharides. The trisaccharide with the highest affinity is Gal $\beta$ 1-4GlcNAc $\beta$ 1-6GalNAc (position 149). In most of the high-affinity oligosaccharides, this trisaccharide is branched with a Gal or *N*-acetylglucosamine (GlcNAc) residue in position 3 of GalNAc. Glycan array data indicate, therefore, that DiscI has a higher affinity for the GalNAc-containing oligosaccharide elongated in position 3, while DiscII prefers a Gal/GalNAc-containing motif substituted in position 6. It could be noted that both lectins bind to large fucosylated and/or sialylated glycans such as positions 358 and 371, but those

could be ligands of the DS domain as discussed below.

#### Affinity studies by isothermal titration microcalorimetry

The affinity of DiscII for GalNAc and  $\beta$ -methyl-galactoside ( $\beta$ MeGal) was previously determined using titration microcalorimetry, leading to dissociation constants of 1.15 and 0.95 mM, respectively.<sup>29</sup> When DiscI is assayed with the same approach, a higher affinity for GalNAc ( $K_d$  = 0.3 mM) but a lower affinity for Gal and  $\beta$ MeGal ( $K_d$  in the range of 3–4 mM) are observed (Table 1). DiscI affinity for disaccharides confirms the preference for  $\beta$ 1-3Gal/



**Fig. 2.** Isothermal titration microcalorimetry. (a) Results for GalNAc monosaccharide (10 mM) binding to DiscI. (b) Results for Gal $\beta$ 1-3GalNAc disaccharide (2 mM) binding to DiscI. Experiments with 0.15 mM protein in 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl buffer solution at 25 °C. Top: Data from 30 automatic injections of sugar (10  $\mu$ l each) into the cell containing DiscI. Bottom: Plot of the total heat released as a function of ligand concentration for the titration shown above (squares). The continuous line represents the best least-squares fit for the obtained data.

GalNAc-linked disaccharides. Among tested disaccharides, Gal $\beta$ 1-3GalNAc is the best ligand, with a  $K_d$  of 27  $\mu$ M, which is a rather high affinity for lectin/glycan interaction. GalNAc $\beta$ 1-3Gal is also a strong ligand, with a  $K_d$  of approximately 50  $\mu$ M, while Gal $\beta$ 1-4GlcNAc (LacNAc) is a poor ligand, with a  $K_d$  in the millimolar range.

The affinity of DiscII for disaccharides was also tested, and it was confirmed that none of the disaccharides cited above displays a high affinity, since all observed dissociation constants were in the millimolar range. The trisaccharidic ligand best recognized by DiscII is not available and therefore could not be tested.

Titration curves of DiscI interacting with GalNAc and Gal $\beta$ 1-3GalNAc are shown in Fig. 2. Thermograms display a large decrease in the exothermic heat of binding while saturation is achieved. For all ligands, the interaction observed upon binding to the lectin was found to be strongly driven by enthalpy as a result of the large number of hydrogen bonds with a strong unfavorable entropy contribution classically attributed either to solvent rearrangement or to loss of ligand conformational flexibility.<sup>34</sup>

### Crystal structure of DiscI

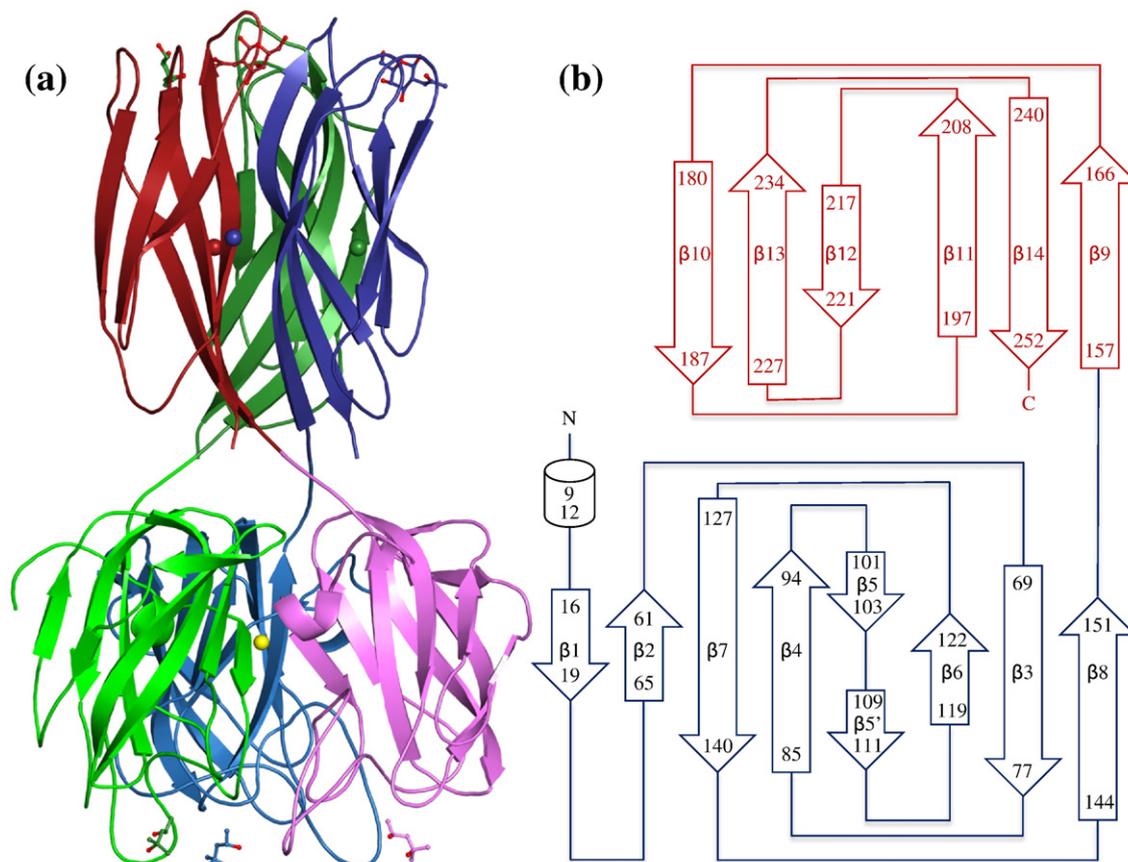
DiscI gives rise to bipyramidal crystals in 4 days, which diffracted to high resolution. They belong to the C222<sub>1</sub> space group with one trimer in the asymmetric unit. The final model of the native structure has a crystallographic  $R$  value of 16.7%, with a corresponding  $R_{free}$  of 20.6% for all observed

data between 28.6 and 1.8 Å resolutions. In this model, all 253 amino acids, as well as some 2-methyl-2,4-pentanediol (MPD) molecules and sulfate, calcium, and nickel ions, could be observed in electron density (refinement statistics in Table 2). In chain A, an additional alanine resulting from the cleavage of the N-terminal His-tag is found. The overall fold of DiscI is comparable to the one observed for DiscII, with formation of a trimer through the intertwining of the two domains of the protein (Fig. 3).<sup>29</sup> The native structures of both recombinant discoidins present an rmsd of 1.2 Å for the 245 aligned C $\alpha$ , as calculated by secondary structure matching.<sup>35</sup>

The N-terminal domain (residues 1–155) presents a classical topology of the DS domain with a  $\beta$ -sandwich consisting of eight antiparallel strands arranged in two  $\beta$ -sheets of five and three  $\beta$ -strands packed against one another to form a distorted barrel (Fig. 3).<sup>36,37</sup> Prominent loops protrude like spikes on the opposite side of the termini, creating a ligand binding site where either a molecule of MPD or a molecule of glycerol, depending on the structure, is observed. At the level of the pseudo-3-fold axis of the trimer, a strong peak of positive density appears between the three His16 side chains. It was modeled as nickel and is probably a residual from affinity chromatography. In disaccharide complexes resulting from another batch of protein, it was modeled as water, since there was no extra positive density and a slight movement of the His16 side chain yielded to coordination distances more relevant to water than to nickel.

**Table 2.** Data collection and refinement statistics

	rDiscI	rDiscI/GalNAc	rDiscI/GalNAc $\beta$ 1-3Gal	rDiscI/Gal $\beta$ 1-3GalNAc
<i>Data collection</i>				
Space group	C222 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>	C222 <sub>1</sub>
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	72.5, 274.3, 85.9	72.4, 274.0, 86.1	72.6, 274.0, 85.8	72.5, 273.7, 85.8
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (outer shell) (Å)	28.6–1.8 (1.84–1.8)	47.4–1.75 (1.79–1.75)	43.7–1.6 (1.67–1.6)	47.3 1.8 (1.92–1.82)
Measured/unique reflections	299,983/75,469	390,844/86,561	362,907/114,648	615,779/76,712
Average multiplicity	4.0 (3.8)	4.5 (4.1)	3.2 (3.0)	8.0 (7.7)
$R_{merge}$	0.070 (0.234)	0.069 (0.389)	0.065 (0.283)	0.094 (0.376)
Completeness (%)	94.8 (94.8)	99.9 (99.9)	99.2 (99.1)	99.7 (98.7)
Average $I/\sigma I$	5.9 (2.0)	8.1 (1.8)	9.8 (2.6)	15.1 (4.7)
<i>Refinement</i>				
$R_{cryst}/R_{free}$	16.7/20.6	15.9/19.3	15.2/18.0	14.1/18.0
Reflections/free reflections	71,665/3802	82,187/4351	108,859/5754	72,846/3849
rmsd bonds (Å)	0.017	0.016	0.017	0.017
rmsd angles (°)	1.5	1.55	1.6	1.6
Protein atoms	1995/1990/1990	1995/1990/1995	2019/2008/2002	2019/2008/2002
<i>B</i> -factors (Å <sup>2</sup> )	17.5/18.0/18.1	14.7/15.3/15.3	12.2/12.6/13.8	12.2/12.6/13.8
Water molecules	348/327/306	371/350/332	409/369/343	384/363/333
<i>B</i> -factors (Å <sup>2</sup> )	30.7/30.9/31.3	29.2/29.1/29.2	26.7/27.0/28.8	25.0/24.0/25.8
Ligand atoms		15/15/15	26/26/26	41/41/41
<i>B</i> -factors (Å <sup>2</sup> )		24.9/23.3/29.2	23.9/28.2/28.5	12.3/13.5/14.7
Heteroatoms	28/14/22	27/33/35	28/12/13	35/19/19
<i>B</i> -factors (Å <sup>2</sup> )	29.0/28.0/26.4	26.9/25.7/26.9	18.9/28.2/24.6	22.1/29.6/28.7
<i>Ramachandran plot</i>				
Most favored regions	88.0	88.0	87.9	87.3
Allowed regions	12.0	12.0	12.0	12.7
Generously allowed regions			0.1	
PDB code	2W94	2W95	2WN3	2WN2



**Fig. 3.** (a) Crystal structure of the Discl trimer colored by chain, with the N-terminal domain in light colors and with the C-terminal domain in dark colors. MPD and GalNAc ligands are represented in balls-and-sticks. Calcium ions are represented as spheres according to chain colors, and nickel ion is represented as a yellow sphere. (b) Schematic representation of Discl topology.

The C-terminal domain presents the immunoglobulin-like  $\beta$ -sandwich fold typical of H-type lectins first observed for *Helix pomatia* agglutinin (HPA) and then for DiscII.<sup>29,33</sup> It consists of six antiparallel  $\beta$ -strands, which form two sheets of three  $\beta$ -strands stacked against one another.<sup>33</sup> Inspection of electron density map also reveals the presence of three ions identified as calcium according to distance values and coordination numbers.<sup>38</sup> Each calcium is found in the middle of the C-terminal domain at the interface between two monomers. It presents seven coordinations: five on the plane, one above the plane, and one under the plane.<sup>38</sup> The contacts involve bidentate Asp246 ( $\beta 14$ ), monodentate Asp164 ( $\beta 9$ ) and Asp203 ( $\beta 11$ ), Gln219 ( $\beta 12$ ) from the neighboring monomer, and two water molecules coordinating the main-chain oxygen atoms of Leu198 and Val200 ( $\beta 11$ ) from the neighboring monomer. The interactions that involve four  $\beta$ -strands over six would help to maintain the trimeric arrangement of the lectin domain. That would explain the dependence for divalent cation on carbohydrate binding and the destabilizing roles of ethylenediaminetetraacetic acid and ethylene glycol bis(*b*-aminoethyl ether) *N,N'*-tetraacetic acid observed for Discl.<sup>28</sup> In DiscII, which is not calcium dependent, the position

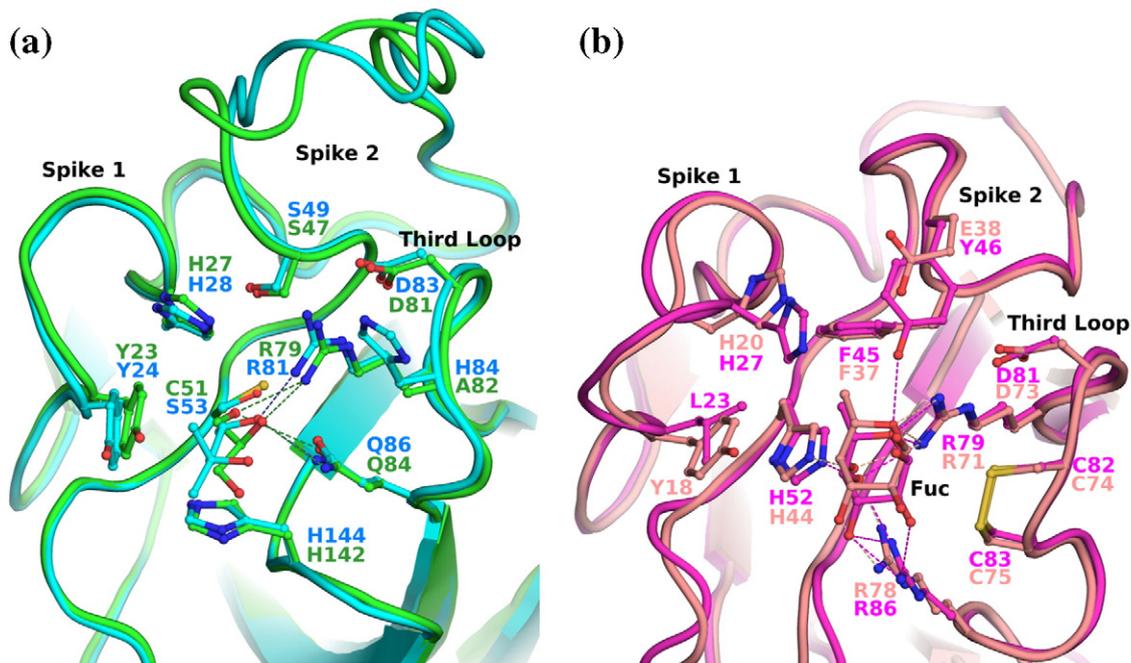
of the calcium ion is occupied by the NZ of Lys206, and other interactions stabilize the trimer.

### DS domain

The N-terminal DS domain of Discl is found in a variety of extracellular and membrane proteins from about 100 eukaryotes and 300 prokaryotes such as coagulation factors, neuropilins, DS domain receptors, or carbohydrate-binding modules (CBMs) associated with enzymes.<sup>30,31</sup> Such domains evolved from a common ancestor to bind very diverse ligands (collagen, lipids, carbohydrates, or growth factors) via a binding site formed by surface loops (spikes) that protrude at the  $\beta$ -barrel opening. The nature and length of those spikes therefore define protein specificity and are responsible for its biological function.

Search of the protein structure database using Dali shows that besides DiscII, coagulation factors, neuropilin, lactadherin, and several CBMs share the DS domain fold.<sup>39</sup> The latter belong to families CBM32 and CBM47 in the CAZy database<sup>‡</sup>, which

<sup>‡</sup> <http://www.cazy.org>

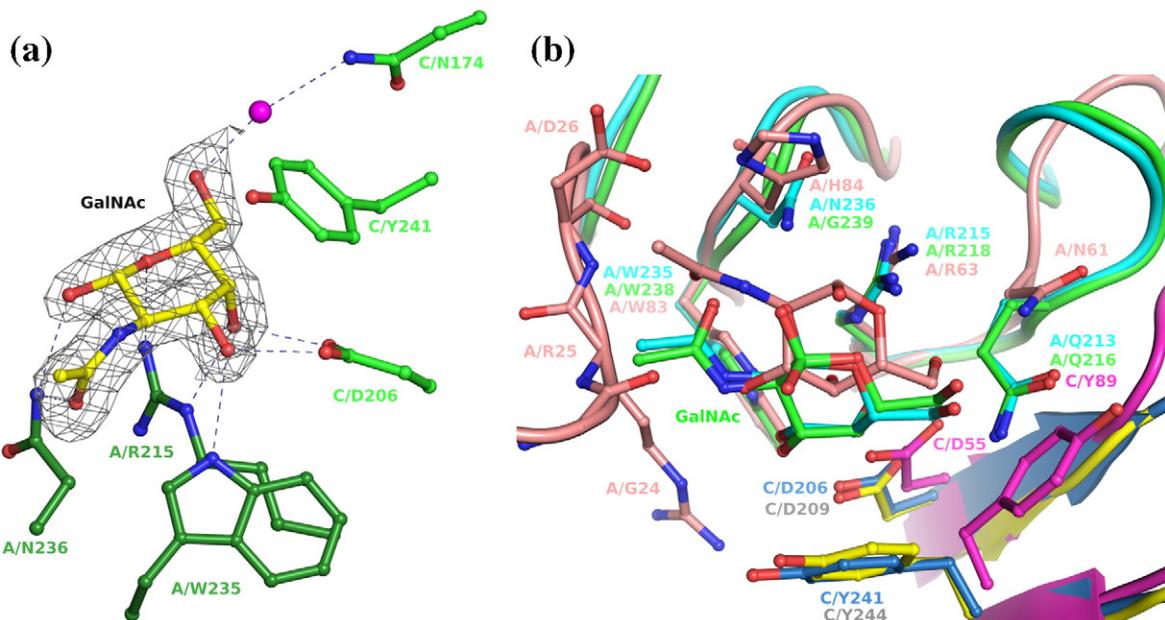


**Fig. 4.** Superposition of the DS domain binding site. (a) Blowup of the glycerol/MPD binding site of DiscI (green) and DiscII (cyan), respectively. (b) Blowup of the fucose binding site of AAA (magenta) and the F-lectin from striped bass (salmon) in the same orientation as discoidins.

bind mainly Gal and fucose, respectively.<sup>40</sup> The structural homologues of discoidins that also form trimers are F-type fucose binding lectins comprising *Anguilla anguilla* agglutinin and F-lectin from striped bass [Protein Data Bank (PDB) code 3CQO].<sup>41</sup>

The overlay of the monomer of those different proteins shows that their carbohydrate binding site

is located at the same position as that of the observed MPD/glycerol binding site of DiscI and DiscII. The ligand environment in AAA, CBMs, and discoidins presents similarity (Fig. 4). A hydrophobic residue (Tyr23 in DiscI, Leu23 in AAA, Tyr18 in F-lectin, and Trp-Tyr in CBMs), which makes stacking interactions with the ligand, is always



**Fig. 5.** Analysis of the GalNAc binding site of H-type lectins. (a) Representation of the maximum likelihood weighted  $2mF_o - DF_c$  electron density around the GalNAc complexed to DiscI contoured at  $1\sigma$  ( $0.41 \text{ e} \text{ \AA}^{-3}$ ). Hydrogen bonds are represented as dotted lines, and waters are represented as magenta spheres. (b) Overlay of the GalNAc binding site of DiscI (cyan/marine), DiscII (green/yellow), and HPA (salmon/magenta). The residues are labeled according to their polypeptide chain and number.

found at the first spike of the first loop. Several charged residues (His27, His142, and Gln84 in DiscI), which would be involved in the hydrogen-bond network, are also observed, but they are different from those of the F-type lectins. At position 82, an alanine is observed in isoform A, contrary to a histidine in the other isoforms and in DiscII. This residue could also play a role in substrate binding, and its possible regulation by phosphorylation has been proposed in DiscII.<sup>29</sup> The third loop in discoidins contains the RGD motif, which promotes cell attachment to substrate in mammalian cells.<sup>23</sup> The arginine of this motif is conserved in homologous F-type lectins and CBMs. In F-type lectins, this residue makes hydrogen bonds with the ring oxygen and an axial hydroxyl group at carbon 4 selecting the L-galacto configuration of fucose (Fig. 4b). The binding site of the DS domain of discoidins presents all characteristics required to bind carbohydrates.

### H-type lectin binding site in complex with carbohydrates

Crystallization studies of complexes with monosaccharide and disaccharides were performed in order to understand the 10-fold increase in DiscI affinity for disaccharides Gal $\beta$ 1-3GalNAc and GalNAc $\beta$ 1-3Gal, when compared to GalNAc monosaccharide. The complex structure of DiscI with GalNAc at 1.75 Å resolution was obtained by soaking, while those with the disaccharides Gal $\beta$ 1-3GalNAc and GalNAc $\beta$ 1-3Gal at 1.8 and 1.6 Å, respectively, were obtained by cocrystallization. All

crystals belong to the C222<sub>1</sub> space group, and data and refinement statistics are summarized in Table 2. Inspection of the initial electron density map revealed for each monomer unambiguous density for the sugars in a shallow groove at the interface between two monomers. In the DiscI/GalNAc $\beta$ 1-3Gal complex, the density for the Gal residue became of good enough quality only after positioning and refinement of the GalNAc moiety.

In the DiscI/GalNAc complex, the monosaccharide is found at the same position as the one observed previously in the other H-type lectins.<sup>29,33,42</sup> It is mainly involved in direct hydrogen bonds with the protein (Fig. 5a). The O3 hydroxyl interacts with Trp235 NE1 and Asp206 OD2 (neighboring monomer); the O4 hydroxyl interacts with Asp206 OD1 (neighboring monomer) and Arg215 NE and NH2; the ring oxygen interacts with Arg215 NH2; and the O7 and O1 hydroxyls interact with Asn236 ND2 (details of distances are listed in Table 3). A single water-mediated interaction is usually observed between the O6 hydroxyl and Asn174 ND2. The Tyr241 side chain (neighboring monomer) is involved in a hydrophobic interaction with part of the sugar ring including C3, C4, C5, and C6. Its orientation is quite different from the one observed in HPA, where its equivalent Tyr89 has rotated by about 145° around the C $\alpha$ -C $\beta$  bond and only stacks against the O6 hydroxymethyl group of GalNAc.

The specificity of DiscI and other H-type lectins for monosaccharides with D-galacto configuration is determined by the fact that most observed interactions involve the O3 and O4 hydroxyls. Those two have to adopt equatorial and axial orientations,

**Table 3.** Contacts between DiscI and its ligand in complex crystal structures, with standard deviations calculated for the three binding sites of the trimer in parentheses

Complex		Distances (Å)		
Sugar atom	Protein atom	GalNAc	GalNAc $\beta$ 1-3Gal	Gal $\beta$ 1-3GalNAc
<i>GalNAc site</i>				
O1	Asn236 ND2	3.15 (0.10)		3.03 (0.01)
O7	Asn236 ND2	3.01 (0.10)	2.90 (0.03)	3.04 (0.02)
	Wat-Asn236 N			2.73 (0.10)-2.89 (0.07)
O3	Trp235 NE1	2.94 (0.09)	2.86 (0.05)	3.06 (0.06)
	Asp206 <sup>a</sup> OD2	2.55 (0.10)	2.54 (0.03)	
O4	Asp206 <sup>a</sup> OD1	2.55 (0.00)	2.58 (0.04)	
	Arg215 NE	2.78 (0.08)	2.96 (0.10)	2.86 (0.04)
	Arg215 NH2	3.15 (0.10)	3.17 (0.01)	
	Trp235 NE1			3.09 (0.04)
	Asp206 <sup>a</sup> OD2			2.60 (0.06)
O5	Arg215 NH2	3.17 (0.03)	2.98 (0.08)	2.91 (0.02)
O6	Wat-Asn174 <sup>a</sup> ND2	2.4 <sup>b</sup> -3.05 <sup>b</sup>	2.54 (0.06)-2.95 (0.06)	2.75 (0.13)-3.15 (0.10)
	Asp206 <sup>a</sup> OD1			2.50 <sup>b</sup>
<i>Gal site</i>				
O3	Asn236 ND2		3.00 (0.06)	
O4	Asn236 ND2		2.97 (0.07)	
	Wat-Gln204 <sup>a</sup> OE1			2.90 (0.27) 2.77 (0.13)
O5	Trp235 NE1			3.24 (0.04)
O6	Ser242 <sup>a</sup> OG			2.70 (0.02)
	Asp206 <sup>a</sup> OD2			2.72 (0.03)

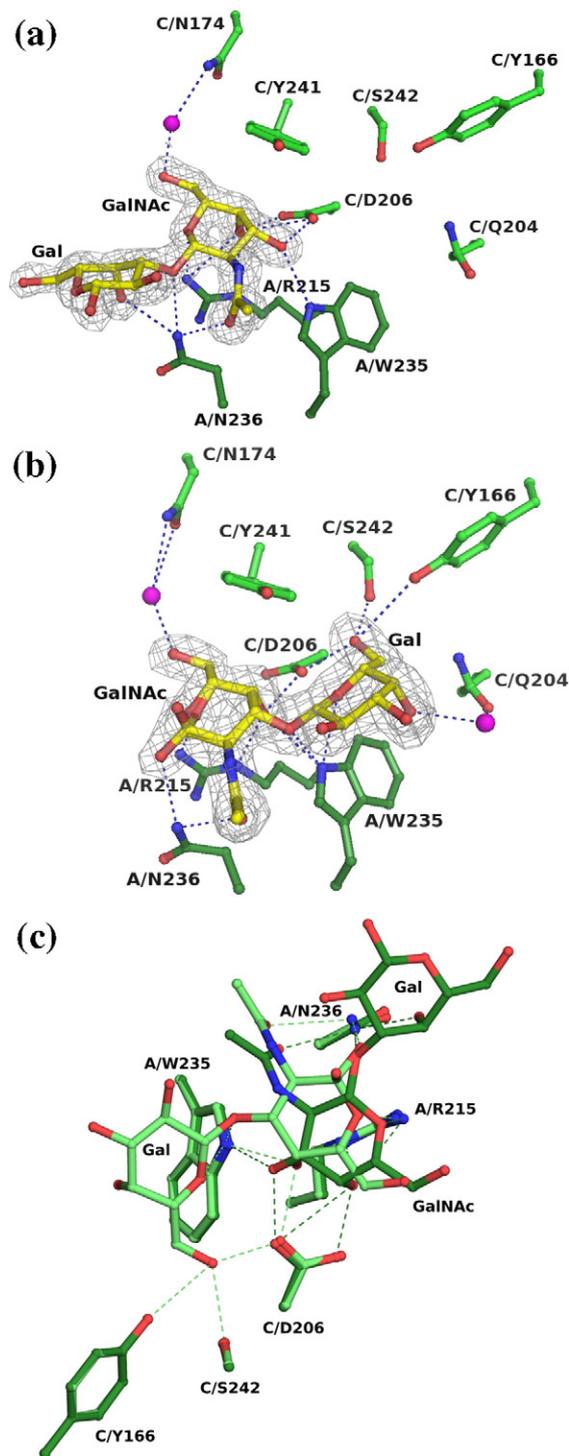
<sup>a</sup> Residue from the neighboring molecule.

<sup>b</sup> Hydrogen bond not present in each monomer.

respectively, for optimal binding, since any other stereochemistry at those positions would lead to steric conflict with Tyr241 and to loss of crucial hydrogen bonds with the protein. The amino acids involved are strictly conserved among H-type lectins (Asp206, Arg215, and Trp235 in DiscI).

Comparison of the structures of H-type lectins in complex with GalNAc indicates that anomeric preference is influenced by the residue following the conserved Trp in the last surface loop (Fig. 5b). In HPA, His84 side chain would hinder the accommodation of the  $\beta$ -anomer, justifying HPA specificity for  $\alpha$ GalNAc.<sup>33</sup> In DiscII, Gly239 allows the binding of both anomers as observed in its GalNAc complex structure.<sup>29</sup> In DiscI, both anomers can bind as observed in the Gal $\beta$ 1-3GalNAc complex structure. Asn236 favors, however, the binding of the  $\beta$ -anomer, as demonstrated in isothermal titration calorimetry (ITC) experiments where DiscI has a higher affinity for  $\beta$ MeGal than for  $\alpha$ -methyl-galactoside (Table 1). Asn236 also interacts with the *N*-acetyl group of GalNAc, explaining the higher affinity of DiscI for GalNAc than for Gal.

In the DiscI/GalNAc $\beta$ 1-3Gal complex structure, the GalNAc moiety is found in the previously described position and presents the same hydrogen-bond network. The Gal moiety is oriented outwards at the protein surface and only makes two hydrogen bonds between the O3 and O4 hydroxyls and the Asn236 ND2 (Fig. 6a). In the DiscI/Gal $\beta$ 1-3GalNAc complex, the GalNAc moiety is also in the primary binding site but in a different orientation, since it is rotated by about  $-30^\circ$  around the anomeric carbon. At the protein level, no changes in amino acids conformation are detected. This mostly modifies the interactions involving the O3 and O4 hydroxyls. The hydrogen bond between the O3 hydroxyl and Trp235 is weaker, while the one with Asp206 is lost. At the level of the O4 hydroxyl, the loss of the hydrogen bond with Arg215 NH2 is compensated for by a new one with Trp235 NE1. It interacts now with the Asp206 OD2 (neighboring monomer), rather than with the OD1 (Fig. 6b). The O6 hydroxyl can present a second conformation, which interacts with the Asp206 OD1 (neighboring monomer). The GalNAc shift allows the good positioning of the Gal moiety in the binding site. The hydrogen-bond network created involves the O5 hydroxyl with Trp235 NE1, the O6 hydroxyl with Ser242 OG (neighboring monomer) and Asp206 OD2 (neighboring monomer), and the O4 hydroxyl with Gln204 OE1 (neighboring monomer) via a water molecule. No interaction is observed for the O2 and O3 hydroxyls. The additional interactions with the Gal moiety explain the higher affinity of DiscI for the disaccharides than for the monosaccharide. Better recognition of Gal $\beta$ 1-3GalNAc by DiscI, compared to GalNAc $\beta$ 1-3Gal, is related to the more enclosed environment of its Gal moiety that makes twice as many and stronger hydrogen bonds with the protein (Fig. 6c).



**Fig. 6.** Disaccharides binding to the DiscI H-type lectin domain. Representation of the maximum likelihood weighted  $2mF_o - DF_c$  electron density around GalNAc $\beta$ 1-3Gal (a) and Gal $\beta$ 1-3GalNAc (b) complexed to DiscI contoured at  $1\sigma$  (0.44 and 0.45  $e \text{ \AA}^{-3}$ , respectively). Hydrogen bonds are represented as dotted lines, and waters are represented as magenta spheres. (c) Overlay of the interactions of the disaccharides with DiscI. The Gal $\beta$ 1-3GalNAc and GalNAc $\beta$ 1-3Gal disaccharides are shown in light green and dark green, respectively. The residues are labeled according to their polypeptide chain and number.

### Molecular modeling of oligosaccharides in DiscII

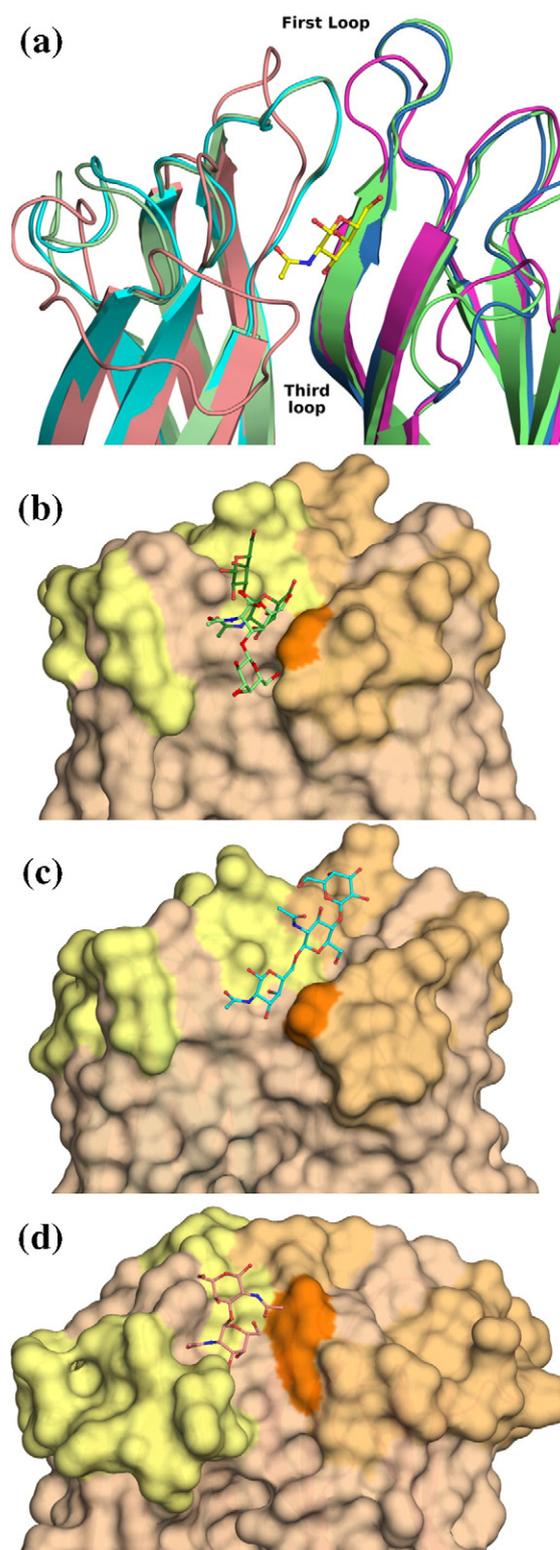
DiscII has no strong affinity for disaccharide (except for sulfated ones), but binds efficiently to the Gal $\beta$ 1-4GlcNAc $\beta$ 1-6 $\beta$ GalNAc trisaccharide present in core 2 and core 4 mucins. Since the trisaccharide was not available for ITC and cocrystallization experiments, it was modeled in the DiscII binding site, with the reducing GalNAc sitting at the position observed in the crystal structures. The trisaccharide is accommodated in a shallow groove at the top of the  $\beta$ -sandwich and extends towards the surface of the lectin trimer. This area is rather hydrophilic, and two hydrogen bonds from the O6 hydroxyl of GlcNAc to the Asn177 side chain and the Ala242 carbonyl group are predicted. Additional stabilizing interactions are established by van der Waals interactions between the nonreducing Gal of the trisaccharide and the aromatic group of His211. The absence of this residue in DiscI (Thr208) may explain its lack of affinity for 1-6-branched GalNAc.

### Specificity loops

The main difference in the overall structure and binding site architecture of H-type lectins is observed in the regions of the first and third surface loops that play a role in tuning specificity and affinity for oligosaccharides. The first loop varies in sequence, length, and structure from one member to the other, while the third one varies in sequence and conformation (Fig. 7a). Their N-terminal part in one monomer creates with their C-terminal part of the neighboring monomer a shallow groove that goes along the side of the  $\beta$ -sandwich and continues across its top. Along the side, it is formed on one side by the N-terminal part of the first loop (residues 167-174 in DiscI, residues 169-177 in DiscII, and residues 9-13 in HPA) and on the other side by its neighboring C-terminal part (residues 177-179 in DiscI, residues 180-182 in DiscII, and residues 17-28 in HPA). At the top of the binding surface, one side entails both ends of the third loop, and the other entails the part of the first loop not listed above.

The binding side has an almost inverted L-shape in discoidins and is deeper, narrower, and more extended in DiscI than in DiscII along the side of  $\beta$ -sandwich (Fig. 7b and c). This allows DiscI to recognize type 3/4 oligosaccharides such as poly-

lactosamine, contrary to DiscII, which lacks affinity for those, since its binding side is far too open on that area to discriminate any sugar. Across the top of the binding surface, the differences in sequence in the third loop are determinants for the recognition of 1,6-oligosaccharides by DiscII, particularly in the presence of His211. The model of Gal $\beta$ 1-4GlcNAc $\beta$ 1-6 $\beta$ GalNAc in the binding site of DiscII



**Fig. 7.** Binding site of H-type lectins. (a) Cartoon representation of the superimposition of DiscI (light green/green), DiscII (cyan/blue), and HPA (salmon/magenta) crystal structures. Zoom on one binding site form by two monomers colored by chain and representation of the GalNAc moiety bound to DiscI in balls-and-sticks. Surface representation of the binding sites of DiscI (b), DiscII (c), and HPA (d). The disaccharides, modeled trisaccharide, and Forssman antigen bound to DiscI, DiscII, and HPA (PDB code 2CGY), respectively, are represented in balls-and-sticks. The Gal $\beta$ 1-3Gal and GalNAc $\beta$ 1-3Gal disaccharides are shown in light green and dark green, respectively.

also illustrates that the O3 of GalNAc is available for branching residues, which could extend the interaction area.

Following the differences in conformation between the third loop and Tyr89, the binding site of HPA extends towards the top of the trimer. The orientation of the Tyr89 side chain prevents the binding of 1,6-branched oligosaccharides similarly to discoidins. The C-terminal end of the first loop is longer by at least seven amino acids in HPA compared to discoidins, leading to a narrower, deeper, and more extended groove with a twist on the side, compared to discoidins (Fig. 7d). This would result in a more restricted specificity for 1,3-branched GalNAc as observed in glycan arrays.<sup>33</sup> This part of the loop is also involved in HPA specificity for the monosaccharide, since it surrounds and stabilizes the *N*-acetyl group and the O3 hydroxyl of GalNAc via several hydrogen bonds that would favor the binding of GalNAc to Gal. In discoidins, it is short and does not interact with the monosaccharide.

## Discussion

Discl and DiscII are two domain proteins assembled as trimers. We reported previously the trimeric structure of DiscII, while the present work focuses on isoform A of Discl. The structures of isoforms C and D of Discl should be almost identical with this one, since they only differ by 13 and 9 amino acids, respectively. These mutations are mainly located at the protein surface, and only one is not found in the DS domain. It is very likely that heterotrimers combining the different isoforms are formed in the slime mold, but no experimental details are available.

The binding sites of the DS domain are very similar for both discoidins, and the nature of its ligand is unknown. The presence of MPD and glycerol molecules, often observed in carbohydrate binding sites, together with the overall similarity to many carbohydrate-binding proteins, suggests that it could bind carbohydrates. According to the glycan array results, fucosylated and/or sialylated N-glycans and O-glycans could be candidates, since they are the only ones recognized by both discoidins and they do not have the characteristics required to be ligands for the H-type lectin domain. Such ligand would be also in agreement with the proposed glycoprotein receptor for the Discl DS domain.<sup>24</sup> The strong resemblance in fold, quaternary structure, and binding site properties with the fucose-binding F-type lectins indicates that the DS domain of discoidins could belong to this lectin family. Specificity studies on the whole proteins and their isolated DS domain are in progress to resolve this issue.

The natural ligands of the H-type lectin domain of Discl and DiscII are still unknown, but the high specificity observed for some oligosaccharides present in the glycan arrays opens the route for

hypothesis. There are not many data on the composition and structures of glycoconjugates present in *D. discoideum*. In addition to classical N-glycans and O-glycans,  $\alpha$ GlcNAc-1-PO<sub>4</sub> phosphoglycosylation has been observed in serine residues of cysteine proteinases, and a complex glycan containing a core tetrasaccharide Gal $\alpha$ 1-3Fuc $\alpha$ 1,2Gal $\beta$ 1,3GlcNAc- has been found attached to hydroxyproline on Skp1, a subunit of ubiquitin ligase protein complex.<sup>43,44</sup> A polysaccharide rich in Gal and GalNAc extracted from *D. discoideum* was proposed to be a DiscII endogenous ligand, since it resembled the spore coat polysaccharide and is produced during late development when DiscII is very prominent and accumulates in pre-spore vesicles.<sup>26</sup> The spores formed upon *D. discoideum* starvation are indeed enclosed in a specialized cell wall, the spore coat, which contains crystalline cellulose together with a Gal/GalNAc-rich polysaccharide (GPS).<sup>45</sup> GPS is stored in pre-spore vesicles prior to secretion and incorporation into the spore coat and the matrix around. It is essential for spore integrity and plays a role in stress resistance.<sup>46</sup> The GPS of the related species *D. mucoideus* has been demonstrated to be a branched polysaccharide with a repeating unit of [-3Gal $\beta$ 1-3(Gal $\beta$ 1-6)GalNAc $\alpha$ 1-].<sup>47</sup> Antisera cross-reactivity indicates that this GPS is related to the GPS of *D. discoideum*. Interestingly, this branched trisaccharide corresponds closely to the glycan with the highest affinity for DiscII in its glycan array. This strengthens the suggestion that GPS would be a DiscII natural ligand.

DiscII appears to be externalized along with glycoconjugates of the spore coat as the spores mature.<sup>11</sup> Similarly, in early development, Discl was found extracellularly around aggregates and slugs in the so-called slime coat, where it binds to an uncharacterized glycoconjugate.<sup>25</sup> The externalization of discoidins, in association with their complementary glycoconjugates during differentiation, lets us suppose a role in the formation of *Dictyostelium* extracellular matrices. Alternatively, discoidins could play a role in the recognition of polysaccharides in bacteria, which are abundant in soil. They could be involved in defense mechanisms by aggregating bacteria, as it has been proposed for the related *H. pomatia* lectin that protects fertilized eggs from infection.<sup>48</sup> They could also play a role in phagocytosis, the process by which wild-type strains of *D. discoideum* feed upon bacteria. A proteomics analysis of the phagosome indicated the presence of DiscII and the three isoforms of Discl.<sup>49</sup> More recently, DNA microarray comparison between the wild-type strain V12M2, which is strictly dependent on phagocytosis, and the axenic strain AX2, which is not, demonstrated that both Discl and DiscII are strongly up-regulated in V12M2, indicating a possible role of these proteins in phagocytosis.<sup>50</sup> Discoidins could have different roles and different ligands depending on their localization, since they are broadly distributed in the cytoplasm, within vesicles, and at the cell surface.

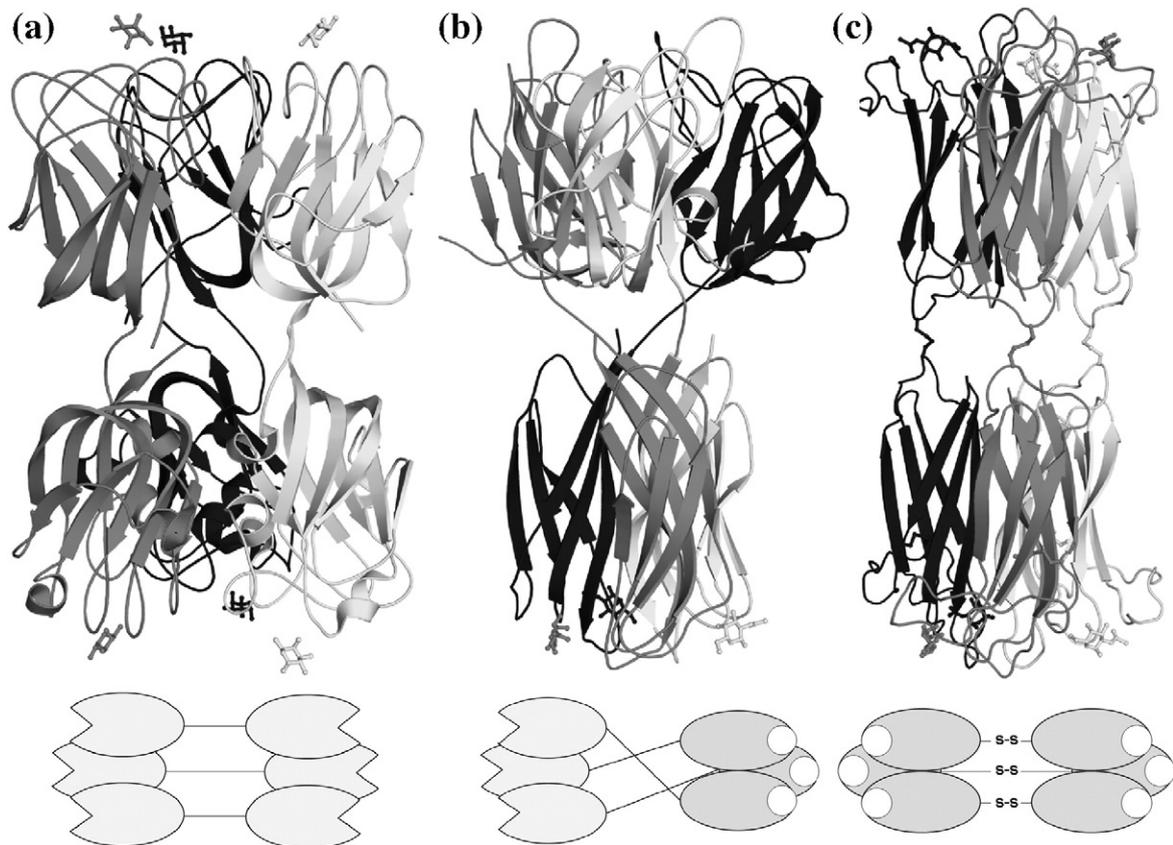
The present study highlighted the differences in specificity of both discoidins from *D. discoideum* proteins, which are spatially and temporally regulated during the life cycle of the slime mold. Structural work rationalized the observed differences in affinity for the lectin domain. The possibility of a carbohydrate binding site in the DS domain is now also opened. In such case, discoidins would be a mix of F-type and H-type lectins with distinct specificities. They present six binding sites as HPA and the striped bass F-lectin. Those trimeric proteins present different domain organizations. HPA creates intermolecular disulfide bridges between two trimers of H-type lectin, and the F-type lectin domain is repeated in tandem in the striped bass F-lectin trimer (PDB code 3CQO), while in discoidins, one domain of each is stabilized by intertwining during trimer formation (Fig. 8).<sup>29,33</sup> This will greatly enhance the multivalency of discoidins, as they will present two binding surfaces on opposite sides of the protein. They would be able to bind simultaneously and nonsimultaneously to their different ligands. Determination of the endogenous ligands of both DS domains will require additional knowledge on *D. discoideum* glycoconjugates. Such study will be necessary to shed some light on the biological function of discoidins, which remains unknown.

## Materials and Methods

### Construction of plasmid for the expression and purification of recombinant DiscI

The *discI* gene of isoform A (P02886) was cloned in-frame with an N-terminal hexahistidine tag (His-tag) and the tobacco etch virus protease cleavage site of the prokaryotic expression vector pProEXHTb (Invitrogen) using polymerase chain reaction and standard molecular biology techniques. The primers used are 5'-GTG CCA TGG CTA CCC AAG GTT TAG TTC-3' and 5'-GAC TCG AGT TAT TCC AAA GCG GTA GC-3'. Sequence analysis was in accordance with the one deposited.

The protein was overexpressed in Rosetta2(DE3)plysS bacterial cells (Novagen) and Luria broth medium. Cells were cultured at 37 °C up to an optical density of 1.2–1.3 at 600 nm before induction with 0.2 mM IPTG overnight at 16 °C and harvested by centrifugation. Each gram of cells was resuspended in 3 ml of buffer A [20 mM Tris-HCl (pH 8.5), 500 mM NaCl, and 10 mM imidazole] before using a cell disrupter (Constant System Ltd.). After centrifugation for 30 min at 40,000g, the supernatant was loaded onto a 1-ml Histrap™ column (GE Healthcare). After a wash step of 5 vol with buffer A, DiscI was eluted with a 10–500 mM imidazole gradient. His-tag was cut by addition of the tobacco etch virus protease at a ratio of 1:100 overnight at 20 °C in the presence of 2 mM DTT. Imidazole was removed using a PD10 column (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl



**Fig. 8.** Cartoon representation of the quaternary structure of the F-lectin from striped bass (PDB code 3CQO), (a) DiscI (PDB code 2CCV), and HPA (PDB code 2CCV) (c). A diagram of the domain layout is drawn beneath each protein.

before reloading on the affinity column. The pure protein without tag eluted in the flow-through was concentrated to around 5.5 mg ml<sup>-1</sup> using Microsep™ 10K (Pall). Each purification step was analyzed on a 12% SDS-PAGE.

### Isothermal titration microcalorimetry analysis

ITC experiments were performed using a Microcal VP-ITC microcalorimeter. All titrations were carried out in 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl at 25 °C. Sugars and proteins were dissolved in the same buffer. Thirty injections of 10 μl of sugar solution at concentrations varying from 2 to 40 mM were added at 5-min intervals to the lectin solution present in the calorimeter cell (1.4478 ml) with stirring at 310 rpm. The protein concentration was 0.15 mM.

Control experiments performed by injection of sugar solution into the microcalorimeter cell in the absence of protein yielded insignificant heats of dilution. Integrated heat effects were analyzed by nonlinear regression using a single-site binding model (Origin 7.0). The experimental data were fitted, and the association constant ( $K_a$ ) and enthalpy of binding ( $\Delta H$ ) were then obtained. The stoichiometry  $n$  was fixed to 1 for low-affinity ligands and fitted to the curve for high-affinity ligands. Other thermodynamic parameters (i.e., changes in free energy  $\Delta G$ , entropy  $\Delta S$ , and number of binding sites per monomer  $n$ ) were calculated from the equation:

$$\Delta G = \Delta H - T\Delta S = -RT \ln K_a$$

where  $T$  is the absolute temperature,  $R=8.32 \text{ J mol}^{-1} \text{ K}^{-1}$ , and  $K$  is the association constant. The experiment was performed at least in duplicate for each ligand.

### Glycan microarray analysis

Recombinant DiscI and DiscII were labeled with Alexa Fluor® 488-TFP (carboxylic acid, 2,3,5,6-tetrafluorophenyl; Invitrogen) in accordance with the manufacturer's instructions and purified on a D-Salt™ polyacrylamide 6000 desalting column (Pierce). The specificity of the two proteins was then analyzed using the printed glycan arrays (version 3.1) of the Consortium for Functional Glycomics§ following the standard procedure.

### Crystallization and structure determination

All crystals were obtained by the hanging-drop-vapor diffusion method using 2-μl drops containing a 50:50 (vol/vol) mix of the protein at 5.5 mg ml<sup>-1</sup> and the reservoir solution at 20 °C. Rod crystals for unliganded experiments were obtained from solution 16 of Hampton Crystallization Screen I (Hampton Research) containing 1.5 M lithium sulfate and 100 mM Na-Hepes (pH 7.5) after almost 1 month. GalNAc was introduced by soaking the crystals for a short time in mother liquor supplemented with 10 mM GalNAc. Crystals of the disaccharide complexes were obtained after 1 week by cocrystallization from a solution containing 100 mM Na-Hepes (pH 7.5) and 1.2 M lithium sulfate. The sugars (13 mM) were incubated with the protein for an hour at room temperature prior to crystallization. Before data collection, the crystals were transferred in a cryoprotecting solution

[either 25% MPD (for unliganded and GalNAc crystals) or 20% glycerol (for disaccharides crystals) was added to the mother liquor], mounted on a rayon fiber loop, and frozen at 100 K. Native and GalNAc complex data were collected at 1.8 and 1.75 Å resolution, respectively, on ID14-2 beamline at the European Synchrotron Radiation Facility (Grenoble, France) using an ADSC Q4 CCD detector (Quantum Corp.). Data on the GalNAcβ1-3Gal and Galβ1-3GalNAc complexes were collected at 1.8 and 1.6 Å resolution, respectively, on ID23-2 beamline at the European Synchrotron Radiation Facility using a MarCCD detector. All data belonged to the C222<sub>1</sub> space group and were processed and reduced using MOSFLM and SCALA.<sup>51</sup> Further computing was performed using the CCP4 suite, unless otherwise stated.<sup>52</sup>

### Structure determination

All DiscI structures were solved by molecular replacement using PHASER.<sup>53</sup> The coordinates of the DiscII monomer (PDB code 2VM9) and of the native DiscI trimer were used as search models for the native and complex structures, respectively.<sup>29</sup> For the refinement of each structure, 5% of the observations were immediately set aside for cross-validation analysis and used to monitor various refinement strategies such as the weighting of geometrical and temperature factor restraints. Manual corrections of the model were performed using Coot<sup>54</sup> and interspersed with cycles of maximum likelihood refinement in REFMAC.<sup>55</sup> Water molecules were added in Coot and checked manually. The quality of the model was assessed in Procheck before its deposition in Brookhaven Protein Data Bank.<sup>56,57</sup> Details of the data and model quality are given in Table 2. All figures were drawn with the 2002 PyMOL Molecular Graphics System (DeLano Scientific, Palo Alto, CA), unless otherwise stated.

### Molecular modeling

The trisaccharide Galβ1-4βGlcNAcβ1-6GalNAc was modeled in the binding site of DiscII using the crystal structure of the complex with GalNAc (PDB code 2VMC) as starting point. The protein moiety was edited with the use of the Sybyl software (Tripos International, St. Louis, MO). Two adjacent lectin domains were considered. All hydrogen atoms were added, partial charges were included in the Pullman library, and the position of hydrogen atoms was optimized with the use of the Tripos force field.<sup>58</sup> The trisaccharide was built using the monosaccharide library available at Glyco3D|| using the PIM parameters and partial charges for carbohydrates.<sup>59</sup> The reducing GalNAc moiety was placed in the binding site by fitting to the GalNAc moiety observed in the crystal structure. Possible orientations of the remaining residues were calculated by systematic search around glycosidic linkages. The lowest energy conformations were further optimized by energy minimization of the full oligosaccharide, and the protein hydrogen atoms with the Tripos force field and the PIM parameters. Energy minimizations were carried out, using the Powell procedure, with a distance-dependent dielectric constant until a gradient deviation of 0.05 kcal mol<sup>-1</sup> Å<sup>-1</sup> had been attained. The validity of the final conformations was checked by comparison with available energy maps (Glyco3D and Imberty *et al.*<sup>60</sup>).

§ <http://www.functionalglycomics.org/>

|| <http://www.cermav.cnrs.fr/glyco3d/>

**PDB accession number**

Coordinates and structure factor amplitudes have been deposited with the PDB with accession codes 2W94 (native crystal), 2W95 (GalNAc complex), 2WN2 (Gal $\beta$ 1–3GalNAc complex), and 2WN3 (GalNAc $\beta$ 1–3Gal complex).

**Acknowledgements**

The glycan microarray analysis was provided by the Consortium for Functional Glycomics funded by National Institute of General Medical Sciences grant GM62116. We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities. We thank Michel Satre and Christopher West for helpful discussions on *D. discoideum*. This work was supported by Centre National de la Recherche Scientifique and the French Ministry of Research. K.S.A. was the recipient of a CNPq grant from the Brazilian Government.

**References**

- Iwanaga, S. & Lee, B. L. (2005). Recent advances in the innate immunity of invertebrate animals. *J. Biochem. Mol. Biol.* **38**, 128–150.
- Vazquez, L., Alpuche, J., Maldonado, G., Agundis, C., Pereyra-Morales, A. & Zenteno, E. (2009). Immunity mechanisms in crustaceans. *Innate Immun.* **15**, 179–188.
- Chaston, J. & Goodrich-Blair, H. (2010). Common trends in mutualism revealed by model associations between invertebrates and bacteria. *FEMS Microbiol. Rev.* **34**, 41–58.
- Muller, W. E., Dorn, A. & Uhlenbruck, G. (1985). The molecular mechanisms of the distinct calcium-dependent aggregation systems in marine sponges and corals. *Acta Histochem. Suppl.* **31**, 37–46.
- Mody, R., Joshi, S. & Chaney, W. (1995). Use of lectins as diagnostic and therapeutic tools for cancer. *J. Pharmacol. Toxicol. Methods*, **33**, 1–10.
- Brooks, S. A. & Wilkinson, D. (2003). Validation of a simple avidin–biotin detection method for *Helix pomatia* lectin (HPA) binding as a prognostic marker in cancer. *Acta Histochem.* **105**, 205–212.
- Queiroz, A. F., Silva, R. A., Moura, R. M., Dreyfuss, J. L., Paredes-Gamero, E. J., Souza, A. C. *et al.* (2009). Growth inhibitory activity of a novel lectin from *Cliona varians* against K562 human erythroleukemia cells. *Cancer Chemother. Pharmacol.* **63**, 1023–1033.
- Kessin, R. H. (2001). *Dictyostelium*—Evolution, Cell Biology, and the Development of Multicellularity. Cambridge University Press, Cambridge, UK.
- Rosen, S. D., Kafka, J. A., Simpson, D. L. & Barondes, S. H. (1973). Developmentally regulated, carbohydrate-binding protein in *Dictyostelium discoideum*. *Proc. Natl Acad. Sci. USA*, **70**, 2554–2557.
- Frazier, W. A., Rosen, S. D., Reitherman, R. W. & Barondes, S. H. (1975). Purification and comparison of two developmentally regulated lectins from *Dictyostelium discoideum*. Discoidin I and II. *J. Biol. Chem.* **250**, 7714–7721.
- Barondes, S. H., Cooper, D. N. & Haywood-Reid, P. L. (1983). Discoidin I and discoidin II are localized differently in developing *Dictyostelium discoideum*. *J. Cell Biol.* **96**, 291–296.
- Barondes, S. H., Haywood-Reid, P. L. & Cooper, D. N. (1985). Discoidin I, an endogenous lectin, is externalized from *Dictyostelium discoideum* in multilamellar bodies. *J. Cell Biol.* **100**, 1825–1833.
- Poole, S., Firtel, R. A., Lamar, E. & Rowekamp, W. (1981). Sequence and expression of the discoidin I gene family in *Dictyostelium discoideum*. *J. Mol. Biol.* **153**, 273–289.
- Tsang, A. S., Devine, J. M. & Williams, J. G. (1981). The multiple subunits of discoidin I are encoded by different genes. *Dev. Biol.* **84**, 212–217.
- Kessin, R. H. (1988). Genetics of early *Dictyostelium discoideum* development. *Microbiol. Rev.* **52**, 29–49.
- Strmecki, L., Bloomfield, G., Araki, T., Dalton, E., Skelton, J., Schilde, C. *et al.* (2007). Proteomic and microarray analyses of the *Dictyostelium* Zak1–GSK-3 signaling pathway reveal a role in early development. *Eukaryot. Cell*, **6**, 245–252.
- Alexander, S., Shinnick, T. M. & Lerner, R. A. (1983). Mutants of *Dictyostelium discoideum* blocked in expression of all members of the developmentally regulated discoidin multigene family. *Cell*, **34**, 467–475.
- Burdine, V. & Clarke, M. (1995). Genetic and physiologic modulation of the prestarvation response in *Dictyostelium discoideum*. *Mol. Biol. Cell*, **6**, 311–325.
- Blusch, J., Alexander, S. & Nellen, W. (1995). Multiple signal transduction pathways regulate discoidin I gene expression in *Dictyostelium discoideum*. *Differentiation*, **58**, 253–260.
- Crowley, T. E., Nellen, W., Gomer, R. H. & Firtel, R. A. (1985). Phenocopy of discoidin I-minus mutants by antisense transformation in *Dictyostelium*. *Cell*, **43**, 633–641.
- Springer, W. R., Cooper, D. N. & Barondes, S. H. (1984). Discoidin I is implicated in cell–substratum attachment and ordered cell migration of *Dictyostelium discoideum* and resembles fibronectin. *Cell*, **39**, 557–564.
- Ruoslahti, E. & Pierschbacher, M. D. (1987). New perspectives in cell adhesion: RGD and integrins. *Science*, **238**, 491–497.
- Ruoslahti, E. (2003). The RGD story: a personal account. *Matrix Biol.* **22**, 459–465.
- Gabius, H. J., Springer, W. R. & Barondes, S. H. (1985). Receptor for the cell binding site of discoidin I. *Cell*, **42**, 449–456.
- Cooper, D. N. & Barondes, S. H. (1984). Colocalization of discoidin-binding ligands with discoidin in developing *Dictyostelium discoideum*. *Dev. Biol.* **105**, 59–70.
- Cooper, D. N., Lee, S. C. & Barondes, S. H. (1983). Discoidin-binding polysaccharide from *Dictyostelium discoideum*. *J. Biol. Chem.* **258**, 8745–8750.
- Cooper, D. N. W., Haywood-Reid, P. L., Springer, W. R. & Barondes, S. H. (1986). Bacterial glycoconjugates are natural ligands for the carbohydrate binding site of discoidin I and influence its cellular compartmentalization. *Dev. Biol.* **114**, 416–425.
- Alexander, S., Cibulsky, A. M. & Lerner, R. A. (1983). Ion dependence of the discoidin I lectin from *Dictyostelium discoideum*. *Differentiation*, **24**, 209–212.
- Saboia Aragão, K., Satre, M., Imberty, A. & Varrot, A. (2008). Structure determination of discoidin II from *Dictyostelium discoideum* and carbohydrate binding properties of the lectin domain. *Proteins*, **73**, 43–52.

30. Baumgartner, S., Hofmann, K., Chiquet-Ehrismann, R. & Bucher, P. (1998). The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. *Protein Sci.* **7**, 1626–1631.
31. Kiedzierska, A., Smietana, K., Czepczynska, H. & Otlewski, J. (2007). Structural similarities and functional diversity of eukaryotic discoidin-like domains. *Biochim. Biophys. Acta*, **1774**, 1069–1078.
32. Vogel, W. F., Abdullhusein, R. & Ford, C. E. (2006). Sensing extracellular matrix: an update on discoidin domain receptor function. *Cell. Signal.* **18**, 1108–1116.
33. Sanchez, J. F., Lescar, J., Chazalet, V., Audfray, A., Gagnon, J., Alvarez, R. *et al.* (2006). Biochemical and structural analysis of *Helix pomatia* agglutinin (HPA): a hexameric lectin with a novel fold. *J. Biol. Chem.* **281**, 20171–20180.
34. Dam, T. K. & Brewer, C. F. (2002). Thermodynamic studies of lectin–carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* **102**, 387–429.
35. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D*, **60**, 2256–2268.
36. Macedo-Ribeiro, S., Bode, W., Huber, R., Quinn-Allen, M. A., Kim, S. W., Ortel, T. L. *et al.* (1999). Crystal structures of the membrane-binding C2 domain of human coagulation factor V. *Nature*, **402**, 434–439.
37. Pratt, K. P., Shen, B. W., Takeshima, K., Davie, E. W., Fujikawa, K. & Stoddard, B. L. (1999). Structure of the C2 domain of human factor VIII at 1.5 Å resolution. *Nature*, **402**, 439–442.
38. Harding, M. M. (1999). The geometry of metal–ligand interactions relevant to proteins. *Acta Crystallogr. Sect. D*, **55**, 1432–1443.
39. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 2780–2781.
40. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. (2009). The Carbohydrate-Active enZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238.
41. Bianchet, M. A., Odom, E. W., Vasta, G. R. & Amzel, L. M. (2002). A novel fucose recognition fold involved in innate immunity. *Nat. Struct. Biol.* **9**, 628–634.
42. Lescar, J., Sanchez, J. F., Audfray, A., Coll, J. L., Breton, C., Mitchell, E. P. & Imberty, A. (2007). Structural basis for recognition of breast and colon cancer epitopes Tn antigen and Forssman disaccharide by *Helix pomatia* lectin. *Glycobiology*, **17**, 1077–1083.
43. Haynes, P. A. (1998). Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology*, **8**, 1–5.
44. West, C. M. (2003). Evolutionary and functional implications of the complex glycosylation of Skp1, a cytoplasmic/nuclear glycoprotein associated with polyubiquitination. *Cell. Mol. Life Sci.* **60**, 229–240.
45. Erdos, G. W. & West, C. M. (1989). Formation and organization of the spore coat of *Dictyostelium discoideum*. *Exp. Mycol.* **13**, 169–182.
46. West, C. M., Nguyen, P., van der Wel, H., Metcalf, T., Sweeney, K. R., Blader, I. J. & Erdos, G. W. (2009). Dependence of stress resistance on a spore coat heteropolysaccharide in *Dictyostelium*. *Eukaryotic Cell*, **8**, 27–36.
47. Sakurai, M. H., Kiyohara, H., Nakahara, Y., Okamoto, K. & Yamada, H. (2002). Galactose-containing polysaccharides from *Dictyostelium mucoroides* as possible acceptor molecules for cell-type specific galactosyl transferase. *Comp. Biochem. Physiol. B*, **132**, 541–549.
48. Kohler, W., Prokop, O. & Kuhnemund, O. (1973). Routine identification of group-C streptococci by means of an agglutinin (protectin) from the albumen gland of the edible snail, *Helix pomatia*. *J. Med. Microbiol.* **6**, 127–130.
49. Gotthardt, D., Blancheteau, V., Bosserhoff, A., Ruppert, T., Delorenzi, M. & Soldati, T. (2006). Proteomics fingerprinting of phagosome maturation and evidence for the role of a Galpha during uptake. *Mol. Cell. Proteomics*, **5**, 2228–2243.
50. Sillo, A., Bloomfield, G., Balest, A., Balbo, A., Pergolizzi, B., Peracino, B. *et al.* (2008). Genome-wide transcriptional changes induced by phagocytosis or growth on bacteria in *Dictyostelium*. *BMC Genomics*, **9**, 291.
51. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data.. *Jt. CCP4/ESF-EACMB. Newsl. Protein Crystallogr.* **26**.
52. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D*, **50**, 760–763.
53. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr. Sect. D*, **61**, 458–464.
54. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D*, **60**, 2126–2132.
55. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D*, **53**, 240–255.
56. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
57. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. T., Jr., Brice, M. D., Rodgers, J. R. *et al.* (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
58. Clark, M., Cramer, R. D. I. & van den Opdenbosch, N. (1989). Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.* **10**, 982–1012.
59. Imberty, A., Monier, C., Bettler, E., Morera, S., Freemont, P., Sippl, M. *et al.* (1999). Fold recognition study of alpha3-galactosyltransferase and molecular modeling of the nucleotide sugar-binding domain. *Glycobiology*, **9**, 713–722.
60. Imberty, A., Delage, M. M., Bourne, Y., Cambillau, C. & Perez, S. (1991). Data bank of three-dimensional structures of disaccharides: Part II. N-acetyllactosaminic type N-glycans. Comparison with the crystal structure of a biantennary octasaccharide. *Glycoconj. J.* **8**, 456–483.