



**HAL**  
open science

# Etudes structure-fonction de lectines (DiscI et DiscII) de *Dictyostelium discoideum*

Karoline Saboia Aragão

► **To cite this version:**

Karoline Saboia Aragão. Etudes structure-fonction de lectines (DiscI et DiscII) de *Dictyostelium discoideum*. Biochimie [q-bio.BM]. Université Joseph-Fourier - Grenoble I, 2008. Français. NNT : . tel-00385711

**HAL Id: tel-00385711**

**<https://theses.hal.science/tel-00385711>**

Submitted on 20 May 2009

**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.

**UNIVERSITE GRENOBLE I – JOSEPH FOURIER  
ECOLE DOCTORALE CHIMIE ET SCIENCES DU VIVANT**

**THESE**

Pour l'obtention du Diplôme de  
DOCTEUR DE L'UNIVERSITE JOSEPH FOURIER  
**Discipline : Biologie structurale et nanobiologie**  
présentée par

**Karoline SABÓIA ARAGÃO**

**Le 05 décembre 2008**

**Etudes structure-fonction de lectines (DiscI et DiscII)  
de *Dictyostelium discoideum***

Directeurs de thèse : Dr. Anne Imberty et Dr. Annabelle Varrot

**JURY**

Dr. Juliette Smith-Ravin, Rapporteur  
Dr. Jacques Le Pendu, Rapporteur  
Dr. Benildo Sousa Cavada, Examineur  
Dr. Michel Satre, Examineur  
Dr. Anne Imberty, Directeur de thèse  
Dr. Annabelle Varrot, Directeur de thèse



*Remerciements,*

Je tiens à remercier Madame Annabelle VARROT et Madame Anne IMBERTY pour leur bienveillante attention qui m'a constamment accompagnée tout au long de cette thèse. Leur encadrement et leurs conseils m'ont été très précieux. Je leur prie d'être assurée de ma sincère reconnaissance.

Je remercie Madame Juliette SMITH-RAVIN et Monsieur Jacques LE PENDU d'avoir accepté de juger ce travail. Qu'ils trouvent ici le témoignage de ma reconnaissance. Je remercie également Monsieur Michel Satre pour ses conseils, sa disponibilité et pour avoir accepté la présidence de ce jury.

Je remercie Monsieur Serge PEREZ de m'avoir accueillie à mon arrivée au CERMAV.

Ma reconnaissance va également à toute l'équipe de Glycobiologie Moléculaire dirigée par Madame Christelle BRETON. Un grand merci à Madame Catherine GAUTIER et Madame Valérie CHAZALET pour leurs conseils constructifs et pour leur disponibilité à chaque fois que j'ai eu à solliciter leurs compétences.

Je remercie également mes professeurs de l'Université Fédéral do Ceara : Benildo CAVADA, Nylane ALENCAR, Tatiane GADELHA et Vincente PINTO qui ont su ouvrir mon appétit pour la recherche et ont rendu possible mon arrivée à Grenoble.

Je remercie toutes les personnes du CERMAV qui m'ont soutenue et encouragé tout au long de ces trois années de thèse : Lina, Michael, Alessandra, Mialy, Bertrand, Anna, Charles, Aurélie, Clothilde, Ondrej, Magali, Emilie, Sara, Sophie, Olivier, Malin, Camelia, Issey, Nasir, Elif, Jimmy qui ont pris une part active aux opérations et qui ont fait régner une agréable ambiance de travail. Tous les moments que nous avons passés ensemble au travail et en dehors resteront à jamais gravés dans mes pensées. Je remercie également toutes les personnes qui, au laboratoire, ont contribué à rendre ces années agréables notamment le groupe des doctorants et tout le personnel administratif.

Un grand merci également à tous mes amies brésiliens et grenoblois qui ont partagé avec moi des moments agréables comme les moments difficiles: Manuelle, Angelo, Claudia, Rodrigo, Ledilege, Raquel, Cristiana, Fabi, Lucas, Poly, Rafa, Claudio, Shirley, Isaac, Céline, Marius, Carlos, Davide, Marito, Johana, Andrea, Matheus, Jana, Marina, Leo, Edson, Pedro, Andréia, Luciano.

Une pensée particulière à ma famille. Malgré la distance qui nous sépare, vous avez su être toujours près de moi. Merci à mes parents, mes frères, ma belle famille et mes amies au Brésil d'avoir toujours cru en moi.

En dernier, je tiens à remercier tout particulièrement mon mari, Windson, qui a partagé les joies et les peines inhérentes à cette thèse. Merci d'avoir été toujours là pour moi, de m'avoir motivé quand il le fallait.

Muito obrigada a todos,

Karoline

# Sommaire

---

1	Introduction.....	17
1.1	Les lectines et les glycoconjugués .....	17
1.1.1	Définition des lectines .....	17
1.1.2	Historique .....	17
1.1.3	La spécificité et l'affinité des lectines.....	19
1.1.4	Intérêt des lectines pour l'homme .....	20
1.2	Distribution des lectines dans le monde vivant.....	22
1.2.1	Lectines de vertébrés .....	23
1.2.2	Lectines d'invertébrés .....	25
1.2.3	Lectines de plantes .....	26
1.2.4	Lectines de microorganismes .....	27
1.3	Les discoidines .....	29
1.3.1	L'amibe <i>Dictyostelium discoideum</i> .....	29
1.3.2	Le cycle de vie de l'amibe <i>Dictyostelium discoideum</i> .....	30
1.3.3	Historique des lectines dans <i>Dictyostelium discoideum</i> .....	32
1.3.4	La discoidine I.....	34
1.3.5	La discoidine II.....	36
1.3.6	Le domaine DS.....	37
1.3.7	Le domaine lectine C-terminal et la famille de type H .....	39
1.4	Objectifs de l'étude .....	43
2	Les outils de l'étude de l'interaction lectine-sucre.....	49
2.1	Test d'hémagglutination.....	49
2.2	La calorimétrie isotherme de titration (ITC).....	50

2.2.1	Principe de la microcalorimétrie isotherme de titration.....	50
2.3	Les puces à sucres .....	53
2.3.1	Glycan array .....	54
2.3.2	Printed array .....	54
2.4	DLS : Diffusion dynamique de la lumière .....	55
2.5	La cristallographie aux rayons X .....	56
2.5.1	La cristallogénèse .....	56
2.5.2	La diffraction des rayons X par les cristaux.....	59
2.5.3	Indexation, intégration et mise à l'échelle du jeu de données.....	62
2.5.4	Facteur de structure et densité électronique .....	63
2.5.5	Problème des phases.....	64
2.5.6	Affinement et vraisemblance du modèle.....	65
3	Résultats .....	69
3.1	Production de DiscI et DiscII .....	69
3.2	Analyse par diffusion de la lumière .....	70
3.3	Interaction avec les ligands glucidiques.....	71
3.4	Études structurales de DiscI et DiscII .....	73
4	Conclusions générales et perspectives .....	107
5	Bibliographie.....	113
6	Annexe .....	129

# Figures

---

Figure 1 – Rôle de glycoconjugués situés sur la surface cellulaire. Représentation schématique d'exemples d'interaction lectines-glucides. ....	18
Figure 2 – Représentation graphique de la structure de la E-selectine humaine en complexe avec le Sialyl Lewis <sup>X</sup> (PDB 1G1T) (Somers, <i>et al.</i> 2000). Le calcium est représenté par une sphère bleue et le ligand par des bâtonnets. ....	24
Figure 3 – Tétramère de la protéine ConM de <i>Canavalia maritima</i> complexée avec le tréhalose (code PDB 2CY6) (Delatorre, <i>et al.</i> 2006). Les cations sont représentés par des sphères (Calcium en jaune et Manganèse en vert) et les ligands par des bâtonnets. ....	26
Figure 4 – Repliement général de la lectine du micronème TgMIC1 chez <i>Toxoplasma gondii</i> (gauche, PDB 2JH1). Zoom dans le site de fixation du complexe TGMIC1-6-siallylactosamine (droite, PDB 2JH7) (Blumenschein, <i>et al.</i> 2007). ....	29
Figure 5 – Corps fructifère de <i>Dictyostelium discoideum</i> (issue de M.J. Grimson et de R.L. Blanton, Université de Technologie du Texas, Lubbock, TX, U.S.A) .....	30
Figure 6 – Développement du cycle de vie de l'amibe <i>Dictyostelium discoideum</i> .....	31
Figure 7 – Arbre phylogénétique qui place <i>Dictyostelium discoideum</i> après la séparation entre les plantes et les animaux (Eichinger, <i>et al.</i> 2005).....	32
Figure 8 – Alignement de séquences des trois isoformes de DiscI (dscA, dscC et dscD) et des deux protéines DiscI-like, DD7-like et DD7-1. Les séquences de la DiscII (dscE) et une protéine similaire (dsc-like) sont également représentées.....	34
Figure 9 – Schéma représentatif de quelques protéines à domaine DS. Abréviations F-A, domaine des facteurs V et VIII; CBP, carboxypeptidase; EGF, facteur de croissance épidermique; TK, tyrosine kinase; A5, homologue de l'antigène A5; Ig, immunoglobuline; LamG, laminin-G; FIB, fibronectin-like; CUB, complément binding; et MAM, meprin/A5/PTPmu. (Vogel 1999). ....	38
Figure 10 – Repliement général du monomère du facteur V (FVa - C2) (PDB 1CZT) (Macedo-Ribeiro, <i>et al.</i> 1999) (à gauche) et du trimère de la lectine AAA de <i>Anguilla anguilla</i> complexée au fucose (PDB 1K12) (Bianchet, <i>et al.</i> 2002) (à droite). ....	39

Figure 11 – Alignement des séquences de <i>Helix pomatia</i> (structure cristallographique HPA et déposition uniprot Q575S3), de la lectine d’escargot <i>Cepaea hortensis</i> (CHP-II), des trois isoformes de la lectine de corail <i>Sinularia lochmodes</i> (SLL-2a, SLL-2b et SLL-2c). .....	40
Figure 12 – Le corail <i>Sinularia lochmodes</i> . .....	40
Figure 13 – Escargot <i>Helix pomatia</i> (HPA). La lectine HPA est localisée principalement dans le système reproducteur de cet escargot (Camus 1899), figure issue de Nature Photo-CZ. .....	41
Figure 14 – A) Représentation d’un trimère de lectine de <i>Helix pomatia</i> (HPA) en complexe avec N-acétylgalactosamine (GalNAc) B) représentation du site de liaison d’HPA avec le GalNAc. La sphère rouge représente une molécule d’eau et le ligand (GalNAc) est représenté par des bâtonnets jaunes. (Sanchez, <i>et al.</i> 2006). C) Hexamère du complexe HPA-antigène Tn (GalNAc $\alpha$ -O-Ser/Thr).....	43
Figure 15 – Le test d’hémagglutination sur microplaque de titration. ....	50
Figure 16 – Microcalorimètre Microcal VP-ITC (en haut) et schéma de fonctionnement (en bas). ....	51
Figure 17 – Courbe brute et courbe du titrage d’un VP-ITC. (A) Energie enregistrée en fonction du temps. (B) Courbe de la chaleur totale libérée en fonction de la concentration totale en ligand; calculée à partir de la courbe A. ....	52
Figure 18 – Principe du Glycan array. ....	54
Figure 19 – Principe des Printed Arrays (figure issue de (Blixt, <i>et al.</i> 2004). La méthode permet de tester un très grand nombre d’oligosaccharides (dans notre cas 377) et exige seulement 0,1-2 $\mu$ g de lectine pour obtenir un signal optimal. ....	55
Figure 20 – Etapes pour l’obtention de la structure 3-D d’une protéine. ....	56
Figure 21 – Diagramme de phase d’une protéine. ....	57
Figure 22 – Système de cristallisation par diffusion en phase gazeuse par goutte pendante... ..	58
Figure 23 – A : Modèle d’un cristal à trois dimensions. B : motif cristallin. C : maille élémentaire et paramètres à trois dimensions.....	59
Figure 24 – Réseaux de Bravais.....	60
Figure 25 – A : Faisceau incident de rayons X sur un cristal ; B : Principe de la diffraction selon Bragg. C: Construction d’Ewald.....	62
Figure 26 – Analyse électrophorétique sur gel SDS-PAGE 12%, 1) Discoidin I et II sauvages à 5 mg/mL ; 2) vide ; 3) DiscIr avec Histag 6 mg/mL ; 4) DiscIr sans Histag après	

stockage pendant 4 mois 5,5 mg/mL ; 5) DiscIr sans Histag 5mg/mL ; 6) DiscIIr avec Histag 4,6 mg/mL ; 7) DiscIIr sans Histag 6mg/mL ; 8) Marqueur de poids moléculaire. ....	70
Figure 27 – Courbe de l’intensité relative de la lumière diffuse en fonction du rayon hydrodynamique pour une solution de la DiscI (à gauche) et de la DiscII (à droite). ....	70
Figure 28 – Comparaison de l’activité hémagglutinante des discoidines recombinantes et de discoidine sauvage à 1mg/mL (A) DiscIIr : 32 UH.mL <sup>-1</sup> ; (B) DiscIr : 64 UH.mL <sup>-1</sup> (C) Discoidines sauvages (mélange de DiscI et DiscII) : 1024 UH.mL <sup>-1</sup> . Contrôle négatif (sang + tampon).....	71
Figure 29 – Superposition des résultats de puces à sucres obtenus avec DiscI (rouge) et DiscII (vert) à la même concentration (0.1 µg/mL). Les pics communs au deux sont colorés en violet.....	72
Figure 30 – Cristaux de: A) discoidines sauvages (mélange de DiscI et DiscII) obtenus pour la condition expérimentale 20% PEG4K, 10mM CdCl <sub>2</sub> , 0.1M Hepes pH 7.5 ; B) DiscIIr obtenu par co-cristallisation avec 10 mM GalNAc, pour la condition expérimentale 25% PEG 550MME 0,1M HEPES pH7.5; et C) DiscIr, obtenus pour la condition expérimentale 1.5M sulfate de lithium 0.1M NaHepes pH 7.5.....	73
Figure 31 – Comparaison des sites de reconnaissance du sucre dans les complexes cristallins DiscI-GalNAc (à gauche) et DiscII-GalNAc (à droite). ....	74
Figure 32 – Superposition des lectines DiscI (vert sapin, rouge, bleu foncé), DiscII (vert clair, magenta, bleu clair) et HPA (jaune, saumon, cyan).....	108

# Tableaux

---

Tableau 1 – Principaux glycanes associés au cancer (Guillot, <i>et al.</i> 2004).-----	21
Tableau 2 – Lectines utilisées pour la mise en évidence de glycotopes en oncologie. -----	22
Tableau 3 – Récapitulatif des structures tridimensionnelles de lectines. -----	23
Tableau 4 – Comparaison entre les propriétés biochimiques des DiscI et DiscII. -----	33
Tableau 5 – Le tableau de spécificité de la DiscI 1µg/mL. -----	129
Tableau 6 – Le tableau de spécificité de la DiscII 1µg/mL. -----	133

# Liste des abréviations

---

**AAA** : *Anguilla anguilla* agglutinin

**CCD** : Charge coupled devices

**CEL-I; CEL-III** : *Cucumaria echinata* lectin

**CMF** : Conditioned medium factor

**CRD** : *Carbohydrate Recognition Domain*

**ConA** : Concanavaline A lectin

**ConM** : *Canavalia maritima* lectin

**CPA** : *Cepaea hortensis* lectin

**CupA-J** : Cup pour Calcium up-regulated genes

**3-D** : Trois dimensions

**cDNA** : Acide déoxyribonucléique complémentaire

**DDR** : Récepteurs à domaine discoidine

**DiscI; DiscII** : Discoidine

**DS** : Domaine discoidine

***D. discoideum*** : *Dictyostelium discoideum*

***E. coli*** : *Escherichia coli*

**ELLA** : Enzyme Linked Lectin Assay

**Gal** : Galactose

**GalNAc** : N-Acétyl-D-galactosamine

**GalNAc $\alpha$ -O-Ser/Thr** : Antigène Tn

**$\alpha$ GalNAc1-3GalNAc** : Antigène Forssman

**Glc** : Glucose

**GlcNAc** : N-Acétyleglucosamine

**GM1** : Ganglioside GM1- Gal( $\beta$ 1-3)GalNAc( $\beta$ 1-4)[NeuAc( $\alpha$ 2-3)] Gal( $\beta$ 1-4)Glc( $\beta$ 1-1)ceramide;

**GNA** : *Galanthus nivalis* agglutinin lectin

**HIA** : Hemagglutination inhibition assay

**HPA** : *Helix pomatia* lectin

**IPTG** : Isopropyl- $\beta$ -D-thiogalactopyranoside

**ITC** : Isothermal titration calorimetry

**LacNAc** : N-acétyllactosamine

**$\beta$ MeGal** : Methyl- $\beta$ -D-galactoside

**MAD** : Multiwavelength anomalous diffraction

**MIR** : multiple isomorphous replacement

**MPD** : 4-methyl-2-pentane diol

**NHS** : N58 hydroxysuccimidine

**PA-IIL** : *Pseudomonas aeruginosa* lectin

**PCR** : Réaction en chaîne par polymérase

**PEG** : Polyéthylène glycol

**PDB** : Protein data bank

**PHA-L ; PHA-E** : *Phaseolus vulgaris* lectin

**PSF** : Pre-starvation factor

**RCA** : *Ricinus communis agglutin*

**SAD** : Single anomalous diffraction

**SLL-2a; SLL-2b; SLL-2c** : *Sinularia lochmodes* lectin

**SPR** : Surface Plasmon Resonance

**TEV** : Tobacco Etch Virus

**TgMIC1** : *Toxoplasma gondii* lectin

**WGA** : *Wheat germ* agglutin



# Introduction



# 1 Introduction

## 1.1 Les lectines et les glycoconjugués

### 1.1.1 Définition des lectines

Les lectines sont des protéines d'origine non immunitaire capables de reconnaître des glucides complexes de manière spécifique et réversible. Ces protéines ne montrent aucune activité enzymatique vis-à-vis de leur ligand (Lis and Sharon 1998). L'expression « d'origine non immunitaire » peut être discutée, puisque certaines lectines sont impliquées dans la défense immunitaire innée. Cette expression sert principalement à souligner la différence entre les lectines et les anticorps du système immunitaire acquis. Les lectines sont aussi identifiées chez les bactéries et les êtres vivants primitifs qui ne possèdent pas de système immunitaire.

Les lectines sont présentes dans toutes les branches du règne vivant. L'abondance des ces molécules et leur relative facilité de purification leur ont permis d'être largement caractérisées et d'être utilisées dans différents domaines de la biologie.

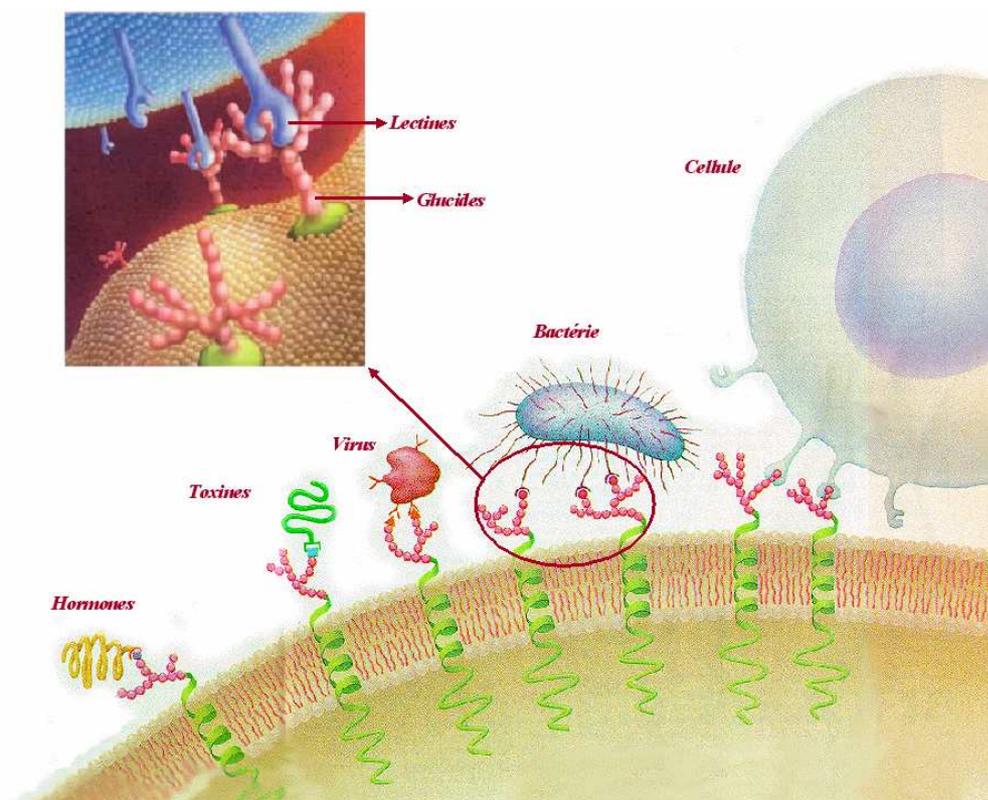
### 1.1.2 Historique

La première lectine a été découverte par Peter Hermann Silltmark en 1888 qui décrit dans sa thèse de doctorat présentée à l'université de Dorpat (maintenant Tartu, Estonie) que des extraits de graines de ricin (*Ricinus communis*) agglutinaient des érythrocytes (Sharon and Lis 2004). A partir de ce moment là, d'autres substances d'origine végétale possédant une activité hémagglutinante ont été découvertes. En 1954, Boyd & Sharpleigh ont démontré la propriété de ces protéines d'agglutiner sélectivement des érythrocytes humains d'un groupe sanguin donné (Boyd and Shapleigh 1954). Cette notion de spécificité est la base de l'étymologie du nom lectine dérivée du mot latin *legere* qui veut dire « sélectionner ».

La plupart des lectines présentes plusieurs sites de liaison pour les glucides. Pour cette raison, l'interaction de lectines avec les glucides présents à la surface des érythrocytes résulte en l'agrégation d'un grand nombre de ces cellules. Cette caractéristique est typique des lectines. Elle est aussi classiquement utilisée pour leur détection (Rüdiger 1993) et leur

caractérisation (Goldstein, *et al.* 1980). Lorsque certains sucres sont ajoutés à ces protéines lors de l'interaction, leur activité hémagglutinante est inhibée ce qui permet de déterminer leur spectre de spécificité (Van Damme, *et al.* 1998).

Les lectines peuvent reconnaître de manière spécifique les glycoconjugués présents sur les surfaces cellulaires. Ces molécules sont constituées d'une partie glucidique (mono ou oligosaccharide) associée de façon covalente à une partie non glucidique (aglycone) de protéines ou de lipides et jouent un rôle primordial dans la vie sociale des cellules (Varki 1993). Les interactions protéine/glucide sont impliquées dans de nombreux phénomènes de reconnaissance et d'adhésion cellulaires. La Figure 1 donne une vue schématique des interactions lectines-glucides.



**Figure 1 – Rôle de glycoconjugués situés sur la surface cellulaire. Représentation schématique d'exemples d'interaction lectines-glucides.**

### 1.1.3 La spécificité et l'affinité des lectines

Les lectines reconnaissent de manière spécifique des mono et oligosaccharides. Les protéines spécifiques pour des monosaccharides sont classifiées en cinq groupes, selon le sucre pour lequel la lectine présente la plus forte affinité: le mannose (Man), le galactose (Gal)/N-acétylgalactosamine (GalNAc), le N-acétylglucosamine (GlcNAc), le fucose (Fuc) et l'acide sialique (NeuAc, acide N-acétylneuraminique) (Lis and Sharon 1998). Cette reconnaissance est souvent désignée comme « la spécificité primaire » des lectines. Ces monosaccharides et leurs dérivés sont ceux qui sont le plus souvent présents sur les épitopes glycaniques des surfaces cellulaires. La plupart des lectines peuvent se lier à des monosaccharides, mais leur affinité sera en général plus forte pour certains oligosaccharides. La constante de dissociation pour les monosaccharides est de l'ordre du millimolaire (mM), en revanche, pour les oligosaccharides elle est de l'ordre du micromolaire ( $\mu\text{M}$ ) (Dam and Brewer 2002). Enfin, certaines lectines telles que PHA-L et PHA-E de *Phaseolus vulgaris* reconnaissent exclusivement des oligosaccharides (Cummings and Kornfeld 1982).

Les similarités structurales entre monosaccharides sont déterminantes pour la spécificité des lectines. Par exemple, la plupart des lectines qui reconnaissent le Gal se lient aussi au GalNAc. Certaines lectines peuvent reconnaître le mannose et le fucose telles que DC-SIGN de mammifère (van Liempt, *et al.* 2006) ou PA-IIL de *Pseudomonas aeruginosa* (Imberty, *et al.* 2004). Ce phénomène est dû à la présence dans ces monosaccharides des trois fonctions hydroxyles qui ont une topologie très similaire. Certaines lectines présentent une spécificité anomérique et peuvent distinguer la configuration en C1 de monosaccharides tels que l' $\alpha$ -methyl-galactoside et le  $\beta$ -methyl-galactoside.

La combinaison de plusieurs techniques expérimentales permet d'élucider la spécificité des lectines (Park, *et al.* 2008). Par exemple, en utilisant les techniques de test ELLA (Enzyme Linked Lectin Assay) et de test « Glycans array » (glycochips avec différents mono et oligosaccharides) la spécificité d'une lectine peut être déterminée. Ces techniques sont simples, rapides et requièrent de quantités réduites de matériel.

La plupart des lectines sont des protéines multivalentes. Elles sont capables de se lier à plusieurs molécules de glucides. La multivalence peut provenir de la répétition de type « tandem » de domaine lectines dans un polypeptide, de l'association de plusieurs monomères ou de la présentation de plusieurs lectines sur une surface cellulaire. Les interactions multiples

entre d'une part les lectines multivalentes et d'autre part les glycoconjugués, sont impliquées dans les processus de reconnaissance (Lee and Lee 1995). Dans la plupart des cas, les systèmes biologiques utilisent des interactions faibles mais multivalente plutôt qu'une seule interaction forte. Ainsi, ils peuvent assurer une haute spécificité et affinité entre ces récepteurs et les surfaces cellulaires des systèmes biologiques.

#### **1.1.4 Intérêt des lectines pour l'homme**

Les lectines peuvent interagir avec des systèmes biologiques et développer une diversité d'événements et fonctions dans ses organismes vivants. Ces interactions ont une grande importance car elles se retrouvent impliquées dans des processus biologiques ainsi que dans des processus pathologiques (Lis and Sharon 1998). Aujourd'hui les lectines sont largement utilisées comme outils dans la recherche et dans le secteur biomédical.

- De par leur spécificité, les lectines immobilisées sur colonne peuvent être utilisées pour l'identification, la purification de glycoconjugués aussi bien que pour leur caractérisation (Hirabayashi 2004).

- Certaines lectines reconnaissent spécifiquement les antigènes des groupes sanguins humains (Boyd and Shapleigh 1954) et sont utilisées pour leur identification dans des banques de sang.

- Certaines lectines purifiées à partir des graines de légumineuses tropicales présentent des propriétés anti-inflammatoires (Alencar, *et al.* 2005).

- Certaines lectines purifiées à partir d'invertébrés terrestres ou marins sont employées comme marqueurs histo-chimiques puisque certaines maladies tel le cancer sont associées à une modification des glycanes présents sur les cellules (Guillot, *et al.* 2004), Tableau 1.

Non de l'antigène	Séquence glycanique
Tn	GalNAc $\alpha$ -Ser/Thr
Sialyl-Tn	SA $\alpha$ 2-6GalNAc $\alpha$ -Ser/Thr
T	Gal $\beta$ 1-3GalNAc $\alpha$ -Ser/Thr
Sialyl-Le $\alpha$	SA $\alpha$ 1-3Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal-   $\alpha$ 1-4 Fuc
Le $^x$	Gal $\beta$ 1-4GlcNAc $\beta$ 1-3Gal-   $\alpha$ 1-3 Fuc
Sialyl- Le $^x$	SA $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-3Gal-   $\alpha$ 1-3 Fuc
Le $^y$	Gal $\beta$ 1-4GlcNAc $\beta$ 1-3Gal   $\alpha$ 1-2   $\alpha$ 1-3 Fuc Fuc
N-acétyllactosamine	Gal $\beta$ 1-4GlcNAc
Séquence $\beta$ 1-6-branchée	GlcNAc $\beta$ 1-6Man $\alpha$ 1-6Man $\beta$ - (oligosaccharides N-liés)

**Tableau 1 – Principaux glycanes associés au cancer (Guillot, *et al.* 2004).**

Le Tableau 2 décrit les principales lectines isolées de plantes ou d'invertébrés qui sont employées pour la mise en évidence de glycotopes en oncologie ou, plus généralement, en histopathologie. Les lectines sont utilisées en fonction des structures glycaniques associées aux états pathologiques (Guillot, *et al.* 2004). De par leur spécificité, les lectines se lient préférentiellement à un monosaccharide soit situé au bout de la chaîne soit placé dans une position non terminale des oligosaccharides complexes.

Lectines	Source	Sucre inhibiteur
LEA	<i>Lycopersicum esculentum</i>	D-GlcNAc
PWL	<i>Phytolacca americana</i>	D-GlcNAc
WGA	<i>Triticum vulgare (wheat germ)</i>	D-GlcNAc
STA	<i>Solanum tuberosum</i>	D-GlcNAc
L-PHA	<i>Phaseolus vulgaris</i>	-GlcNAc $\beta$ 1-6Man $\alpha$ 1-6Man $\beta$ -
GSI-B4	<i>Griffonia (Bandeira) simplicifolia</i>	D-Gal
MPA	<i>Maclura pomifera</i>	D-Gal
DBA	<i>Dolichos biflorus</i>	D-GalNAc
LBL	<i>Phaseolus lunatus (Lima bean)</i>	D-GalNAc
RCA-1	<i>Ricinus communis</i>	D-GalNAc
SBA	<i>Glycine max</i>	D-GalNAc
HAA	<i>Helix aspersa</i>	D-GalNAc
<b>HPA</b>	<b><i>Helix pomatia</i></b>	<b>D-GalNAc</b>
VVA-B4	<i>Vicia villosa</i>	D-GalNAc
WFA	<i>Wistaria floribunda</i>	D-GalNAc
ECA	<i>Erythrina crista-galli</i>	D-Gal $\beta$ 1-4D-GlcNAc
PSL	<i>Laetiporus sulfureus</i>	D-Gal $\beta$ 1-4D-GlcNAc
GSI-A4	<i>Griffonia (Bandeira) simplicifolia</i>	D-Gal $\beta$ 1-4D-GlcNAc
ABL	<i>Agaricus bisporus</i>	D-Gal $\beta$ 1-3GalNAc
PNA	<i>Arachis hypogaea</i>	D-Gal $\beta$ 1-3D-GalNAc
JAC	<i>Artocarpus integrifolia</i>	D-Gal $\beta$ 1-3D-GalNAc
LDL	<i>Lactarius deliciosus</i>	D-Gal $\beta$ 1-3D-GalNAc
VGA	<i>Vicia graminea</i>	D-Gal $\beta$ 1-3D-GalNAc
EEA	<i>Euonymus europaeus</i>	D-Gal $\alpha$ 1-3(L-Fuc $\alpha$ 1-2)D-Gal $\beta$ 1-3/4D-GlcNAc

Tableau 2 – Lectines utilisées pour la mise en évidence de glycotopes en oncologie.

## 1.2 Distribution des lectines dans le monde vivant

Les lectines sont des protéines ubiquitaires, présentes dans tous les organismes vivants. Il existe une grande variété de lectines qui présentent une très grande diversité structurale. Le nombre de structures cristallographiques de lectines est toujours en croissance, et on connaît aujourd'hui la structure tridimensionnelle d'environ 770 lectines. Notre laboratoire a développé une base de données des structures 3-D des lectines qui rassemble les informations structurales de ces protéines (la base de données des lectines est accessible sur:

<http://www.cermav.cnrs.fr/lectines>). La banque de lectines comprend ainsi 48 familles structurales différentes dont 6 d'origine végétale, 10 d'origine bactérienne, 18 d'origine animale, 5 familles d'origine virale, 8 d'origine fongique et 1 famille d'algues, Tableau 3.

Origine	Exemples de Lectines	Native	Complexé	Total
Plantes	ConA Ricine	106	201	307
Bactéries	PA-IL de <i>Pseudomonas</i> Toxine de cholera	37	79	116
Animaux	E-selectin <i>Helix pomatia</i> agglutinin	80	152	232
Virus	Hemagglutinin de virus Capside de rotavirus	43	25	68
Champignons	lectine de mousseron	17	23	40
Algues	Griffithsin	2	7	9
Total		285	487	772

**Tableau 3 – Récapitulatif des structures tridimensionnelles de lectines.**

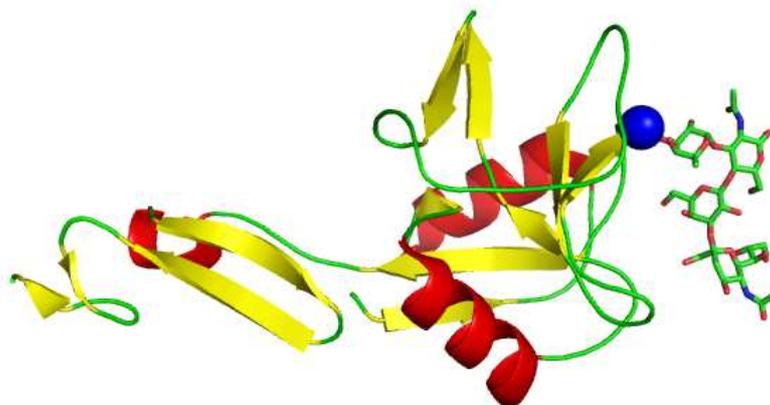
### 1.2.1 Lectines de vertébrés

Les lectines animales sont réparties dans des familles très différentes. Les trois familles les plus étudiées sont les *galectines*, les lectines de *type C* et les *siglecs*. Les lectines d'invertébrés sont traités séparément dans la suite de ce chapitre.

Les structures de *galectines* sont relativement simples. Ces protéines sont spécifiques pour le  $\beta$ -galactose et plus précisément pour le lactose ( $\beta$ Gal1-4Glc) et le N-acetyllactosamine ( $\beta$ Gal1-4GlcNAc) (Leffler, *et al.* 2004). La première structure tridimensionnelle d'une galectine humaine (la galectin-7 (hGal-7)) a été obtenue dans sa

forme native et complexée avec le galactose, le galactosamine, le lactose, et le *N*-acétyllactosamine (Leonidas, *et al.* 1998).

Les lectines du *type C* présentent un CRD (*Carbohydrate Recognition Domain*) bien conservé qui implique un atome de calcium dans l'interaction protéine-sucre (Drickamer 1993). Les lectines de type-C sont soit en circulation dans le plasma, soit attachées aux surfaces cellulaires par la présence d'un segment transmembranaire. Un exemple de la structure 3-D d'une lectine du type C est la E-selectine en complexe avec le Sialyl Lewis<sup>X</sup> (PDB 1G1T) (Somers, *et al.* 2000), Figure 2.



**Figure 2 – Représentation graphique de la structure de la E-selectine humaine en complexe avec le Sialyl Lewis<sup>X</sup> (PDB 1G1T) (Somers, *et al.* 2000). Le calcium est représenté par une sphère bleue et le ligand par des bâtonnets.**

Les *Siglecs*, ou lectines de type I, constituent une famille de lectines qui reconnaissent l'acide sialique (Crocker 2002). Elles sont membres de la superfamille immunoglobuline (Ig) et adoptent un repliement de type immunoglobuline.

Les autres classes de lectines animales comprennent les lectines du *type P* qui sont formées par les lectines qui fixent le mannose-6-phosphate, comme la lectine bovine CD-MPR (Roberts, *et al.* 1998). Les *pentraxines* qui sont une famille des lectines capables de se lier à des ligands de manière  $\text{Ca}^{2+}$  dépendante (Emsley, *et al.* 1994). D'autres types de lectines de vertébrés telles que les *calnexines-calreticuline* comme la lectine calnexine de l'espèce *Canis familiaris* (Schrag, *et al.* 2001) et les ficolines (Garlatti, *et al.* 2007) ont récemment été mises en évidence.

La plupart des lectines de vertébrés ont une localisation extracellulaire et sont capables de détecter les modifications de glycosylation sur les cellules environnantes. Elles jouent donc un rôle dans la vie sociale des cellules et sont impliquées dans des processus tels que la

fécondation, la migration et le développement cellulaire. Par exemple, dans le processus de fécondation, un glycoconjugué de la surface de l'ovule interagit avec une lectine (spermadhésine) du spermatozoïde (Topfer-Petersen, *et al.* 1998). Il a été aussi montré que l'embryon exprime à sa surface cellulaire la L-sélectine six jours après la fécondation. En se liant aux sucres spécifiques de l'endothélium utérin, la L-sélectine détermine ainsi la première étape indispensable à la fixation du blastocyste à la surface de l'endomètre et permet le début de la grossesse (Genbacev, *et al.* 2003).

### 1.2.2 Lectines d'invertébrés

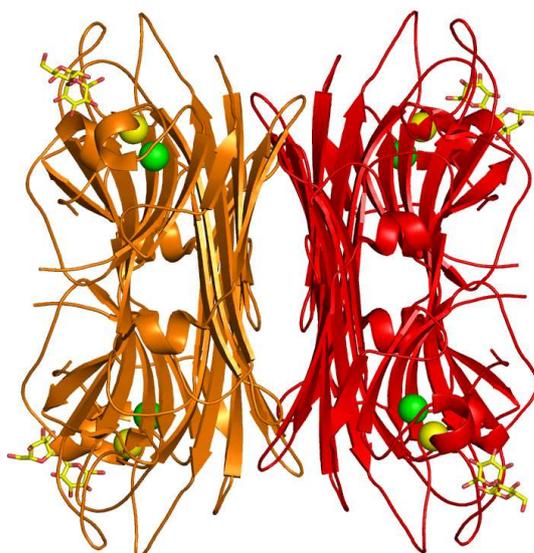
La présence de lectines a été démontrée chez les invertébrés terrestres ou marins. Ces protéines font souvent partie d'un système d'immunité innée et présentent donc des spécificités pour les glucides présents à la surface d'organismes pathogènes (Vasta 1992). Chez les invertébrés, les lectines sont distribuées dans toutes les classes et sous-classes étudiées. Nous pouvons donner en exemple : les gastéropodes avec la lectine d'escargot CHA-II de *Cepaea hortensis* (Gerlach, *et al.* 2002), les échinodermes avec l'hémagglutinine de l'oursin *Hemicentrotus pulcherrimus* (Yamada and Aketa 1982), les crustacés avec la lectine du crabe *Charybdis japonica* (Umetsu, *et al.* 1991), et les insectes avec la lectine du sphinx du tabac *Manduca sexta* (Yu and Kanost 2000). Les lectines présentent des localisations variées : système digestif, appareil de reproduction, hémolymphe ou branchies.

Un nombre limité de lectines d'invertébrés a été cristallisé. La codakine purifiée à partir des branchies de la palourde blanche *Codakia orbicularis* est une lectine de type C qui a une forte spécificité pour des N-glycannes biantennés (Gourdine, *et al.* in press). D'autres exemples de lectines de type C sont la lectine de tunicier *Polyandrocarpa misakiensis* (Poget, *et al.* 1999) et la lectine I (CEL-I) du concombre de mer *Cucumaria echinata* (Sugawara, *et al.* 2004). Les structure 3D de lectines de type ricine ont aussi été obtenues telles que la lectine III (CEL-III) du même concombre (Uchida, *et al.* 2004) et la lectine du vers de terre *Lumbricus terrestris* (Suzuki, *et al.* 2004), ces deux lectines présentant une spécificité pour le galactose. La limule, *Tachypleus tridentatus*, est un fossile vivant qui présente un grand nombre de lectines impliquées dans l'immunité innée. Les structures de la tachylectine 2, un  $\beta$ -propeller (Beisel, *et al.* 1999) et de la tachylectine 5A, une protéine de type fibrinogène (Kairies, *et al.* 2001), ont été résolues en complexe avec le GlcNAc. La structure de la lectine

extraite de l'escargot *Helix pomatia* (HPA) (Sanchez, *et al.* 2006) a permis de définir une nouvelle famille de lectine, le type H, qui sera décrite dans la Section 1.3.7.

### 1.2.3 Lectines de plantes

Les lectines végétales d'une même famille taxonomique (e.g., les lectines de légumineuse, de céréales, etc.) présentent des homologies de séquence et des similarités structurales. Historiquement, les lectines de légumineuses telle que la concanavaleine A (ConA) ont été les premières à être caractérisées. La première structure cristallographique d'une lectine de légumineuses (la ConA) a été déterminée en 1972 (Edelman, *et al.* 1972, Hardman and Ainsworth 1972). Les lectines de légumineuses adoptent toutes un repliement typique appelé «  $\beta$  jelly roll » formé par deux feuillets  $\beta$  superposés bien conservés dans les nombreuses structures résolues. Les monomères s'associent deux à deux de 7 manières différentes ce qui permet de former une variété d'assemblage macromoléculaires sous la forme de dimères ou de tétramères (Sinha, *et al.* 2007). La Figure 3 représente un tétramère de la protéine ConM de *Canavalia maritima* (PDB 2CY6) complexée avec le tréhalose (Delatorre, *et al.* 2006). La géométrie du site de fixation est fixée par la présence d'un ion manganèse et d'un ion calcium qui sont proches du ligand glucidique mais qui n'interagissent pas directement avec lui. Les variations en acides aminés autour du site de reconnaissance apporte la spécificité pour différents monosaccharides et oligosaccharides (Loris, *et al.* 1998).



**Figure 3 – Tétramère de la protéine ConM de *Canavalia maritima* complexée avec le tréhalose (code PDB 2CY6) (Delatorre, *et al.* 2006). Les cations sont représentés par des sphères (Calcium en jaune et Manganèse en vert) et les ligands par des bâtonnets.**

Une autre grande famille de lectines d'origine végétale qui a pu être caractérisée est la famille des Gramineae (céréales) à laquelle appartient la lectine de germe de grain de blé (WGA). Les lectines de céréales présentent des domaines structuraux conservés riches en cystéine qui sont aussi appelés domaine hévéine. La première structure cristallographique déterminée dans cette famille a été celle de la WGA (Wright, C. S. 1980).

La famille des Monocotylédones présente des lectines spécifiques pour le mannose, la première structure 3-D a été la *Galanthus nivalis* agglutinine (GNA) (Wright, C.S. and Hester 1996). La famille de Moraceae est formée par les lectines qui fixent le galactose telle que la jacaline isolée des graines de *Artocarpus integrifolia* (Sankaranarayanan, *et al.* 1996). La famille Amaranthaceae contient la lectine ACA de *Amaranthus caudatus* qui a été cristallisée avec le Gal $\beta$ 1-3GalNAc (Transue, *et al.* 1997). La famille Euphorbiaceae englobe les lectines qui présentent une grande toxicité, un exemple de lectine de cette famille est la lectine RCA isolée des graines de ricin *Ricinus communis*. La ricine est un hétérodimère où la chaîne B constitue les domaines lectines qui reconnaissent le galactose (Rutenber, *et al.* 1991). La famille Solanaceae englobe les lectines de pommes de terre et des tomates mais aucune structure n'a encore été déterminée.

Les lectines de plantes semblent être impliquées dans la défense contre les phytopathogènes et les prédateurs, certaines ayant des activités insecticides (Chrispeels and Raikhel 1991, Rudiger and Gabius 2001).

#### 1.2.4 Lectines de microorganismes

Les microorganismes pathogènes, virus, bactéries, champignons ou parasites eucaryotes, utilisent fréquemment des lectines pour reconnaître les sucres présents sur la surface des cellules hôte. Ces interactions lectines-sucres jouent également un rôle dans l'adhésion sur les tissus richement glycosylés présents dans les voies respiratoires, digestives ou dans l'appareil urogénital (Imberty and Varrot 2008, Sharon 1996).

L'exemple le plus marquant de lectine de virus est l'hémagglutinine du virus de la grippe (Influenza virus). Cette hémagglutinine interagit avec un récepteur de la surface des cellules hôtes, l'acide 5-N-acétylneuraminique (l'acide sialique). Sa structure 3-D a été résolue en complexe avec son récepteur naturel et avec 12 analogues de ligands (Weis, *et al.* 1990).

Elle est composée de deux polypeptides HA<sub>1</sub> et HA<sub>2</sub>. La sous unité HA<sub>1</sub> correspond à un domaine globulaire qui contient le site de liaison de la lectine.

Les lectines bactériennes connues peuvent être classées en trois familles principales : les lectines *fimbriaes* (pili et flagelles), les *toxines* et les lectines *solubles* (Imberty, *et al.* 2005).

Les lectines *fimbriaes* sont trouvées dans des organelles de surface des bactéries (pili et flagelles) qui présentent diverses fonctions telles que la reconnaissance des glycannes de l'hôte et l'adhésion sur les surfaces de cellules cibles. Il existe trois types de pili : les types IV, les types I et les types P. Par exemple, la PapG est une adhésine du pili de type P trouvée chez les souches uropathogènes de la bactérie *Escherichia coli*. PapG présente une grande importance dans la phase d'adhésion sur les globoside (GbO4) des cellules humaines du rein lors des infections urinaires et la structure en complexe avec la partie tetrasaccharide du GbO4 a été résolue (Dodson, *et al.* 2001). Les lectines de cette famille sur une structure allongée avec un repliement de type  $\beta$ -sandwich.

La famille des toxines est constituée par des protéines secrétées par les bactéries et qui exercent une activité toxique sur les cellules de l'hôte. Les toxines du type AB<sub>5</sub> sont secrétées par différentes bactéries comme *Shigella dysenteriae*, *Borderella pertussis* et *Vibrio cholerae*. Le domaine lectine de la toxine sécrétée par le *Vibrio cholerae* est un  $\beta$ -pentamère et sa structure a été résolue complexée avec le pentasaccharide GM1 qui est un ganglioside présent sur les epithelia (Merritt, *et al.* 1994).

La famille de lectines *solubles* comprend des protéines cytoplasmiques telles que PA-II et PA-III de *Pseudomonas aeruginosa* respectivement spécifiques pour le  $\alpha$ -D-galactose et le  $\alpha$ -L-fucose. Elles présentent des structures tétramériques ou un ou deux ions calciums respectivement sont impliqués dans le site de reconnaissance du sucre (Imberty, *et al.* 2004).

Les protozoaires possèdent aussi des lectines. La surface de l'amibe pathogénique *Entamoeba histolytica* présente deux lectines reconnaissant respectivement le GlcNAc et le Gal/GalNAc (Mirelman, *et al.* 1996). Des lectines ont été isolées à partir du merozoïde de *Plasmodium falciparum*, le parasite responsable de la malaria (von Itzstein, *et al.* 2008) et de *Trypanosoma cruzi* qui cause la maladie de Chagas (Marcipar, *et al.* 2003). Seule la structure tridimensionnelle de la lectine TgMIC1 identifiée dans le micronème du protozoaire *Toxoplasma gondii* a été démontrée complexée avec 6-sialyllactosamine (PDB 2JH7) et 3-sialyllactosamine, Figure 4. Des protéines du micronème (MIC) sont libérées à la surface du

parasite juste avant l'invasion des cellules hôtes et jouent des rôles importants dans l'identification, l'attachement et la pénétration de la cellule hôte (Blumenschein, *et al.* 2007).

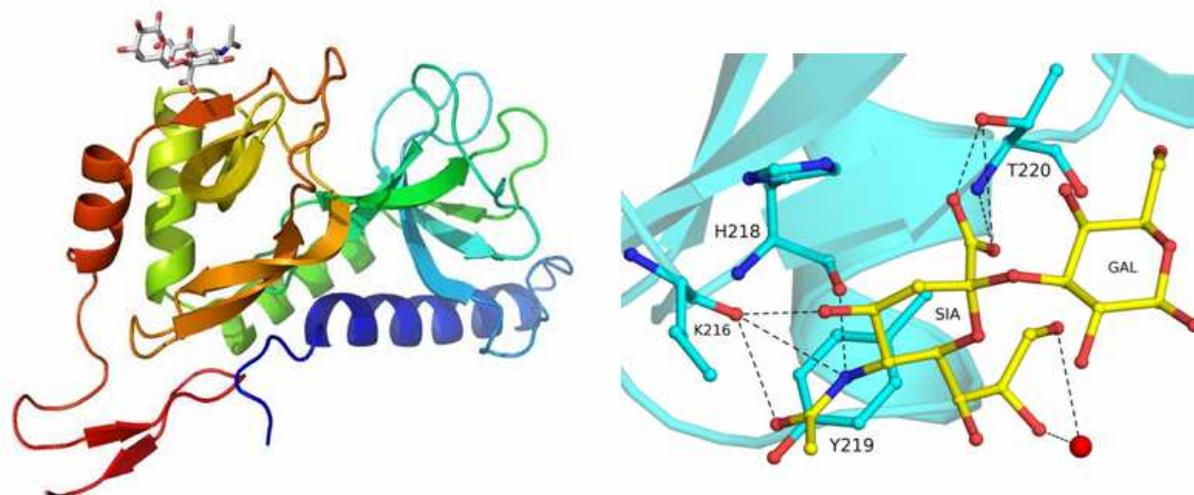


Figure 4 – Repliement général de la lectine du micronème TgMIC1 chez *Toxoplasma gondii* (gauche, PDB 2JH1). Zoom dans le site de fixation du complexe TGMIC1-6-sialyllactosamine (droite, PDB 2JH7) (Blumenschein, *et al.* 2007).

## 1.3 Les discoidines

### 1.3.1 L'amibe *Dictyostelium discoideum*

*Dictyostelium discoideum* est une amibe du sol bien connue des biologistes. Un grand intérêt de cette amibe est le fait qu'elle soit un eucaryote simple haploïde qui présente une grande similarité avec les cellules animales. Le mycologue Oskar Brefeld (Brefeld 1869) fut le premier à observer l'organisme *Dictyostelium mucoroides* qu'il a nommé 'Dicty' qui veut dire 'avec des mailles (comme dans un filet)', 'stelium' qui veut dire 'tour' et 'mucoroides' en raison de son apparente similitude avec des moisissures du sol du genre 'Mucor'.

*D. discoideum* qui est l'espèce la plus étudiée aujourd'hui en laboratoire fut découverte en 1935 par Kenneth Raper dans les forêts de feuillus de la Caroline du Nord aux Etats-Unis, Figure 5 (Raper 1935). Les amibes mesurent environ 20µm de diamètre et vivent en milieu humide, dans l'humus des sols. Au plan taxonomique, *D. discoideum* est un eucaryote qui

appartient au phylum Mycetozoa dans lequel on trouve la classe Dictyostelia qui inclut l'ordre Dictyosteliida et la famille Dictyosteliidae.

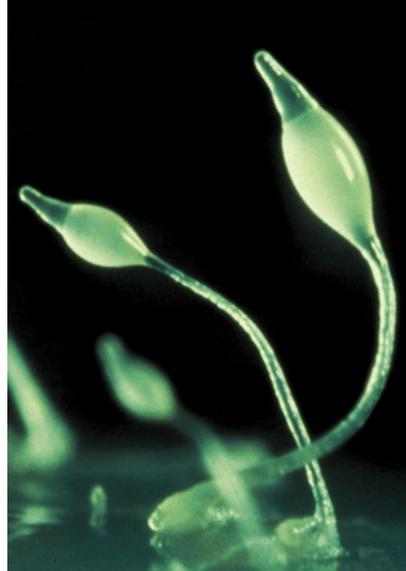
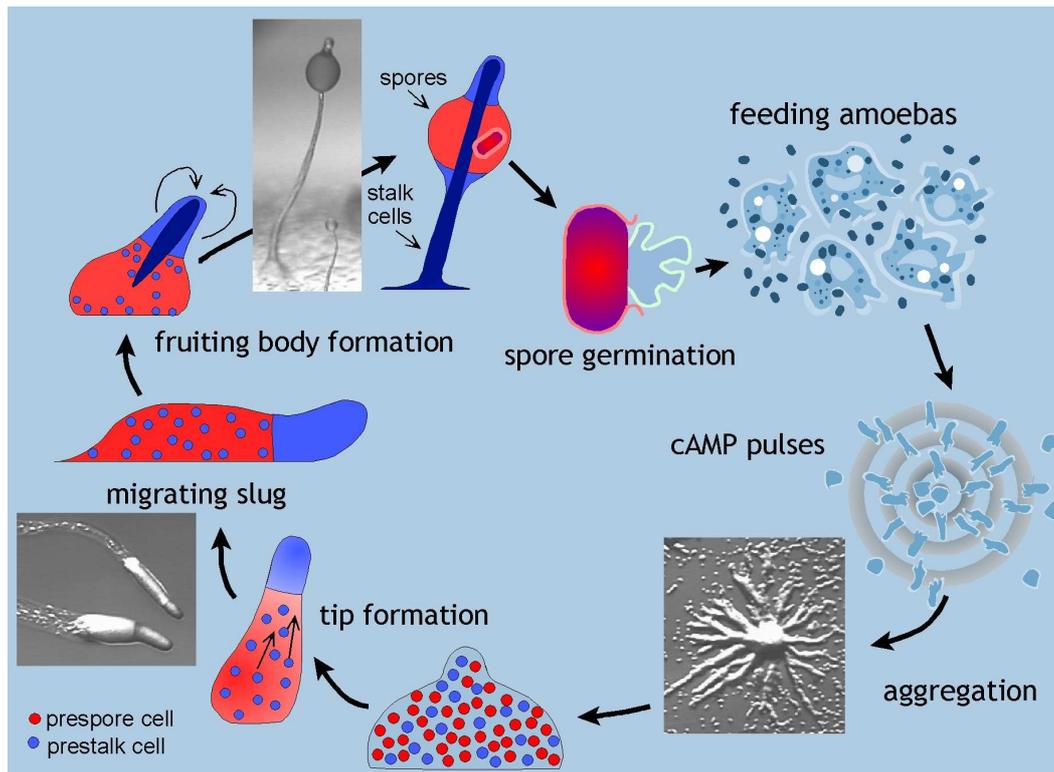


Figure 5 – Corps fructifère de *Dictyostelium discoideum* (issue de M.J. Grimson et de R.L. Blanton, Université de Technologie du Texas, Lubbock, TX, U.S.A)

### 1.3.2 Le cycle de vie de l'amibe *Dictyostelium discoideum*

*D. discoideum* est capable de développer un extraordinaire cycle multicellulaire. Elle présente le pouvoir de vivre à l'état de cellules isolées (état unicellulaire végétatif) ou de colonies multicellulaires (état pluricellulaire). Pendant la phase végétative, dans un environnement favorable en condition alimentaire, l'amibe s'alimente de bactéries par phagocytose. Elle reste isolée de ses consociaux et se multiplie par une simple division en deux cellules filles. Néanmoins, si l'environnement devient difficile, comme dans le cas de carence alimentaire, les cellules s'assemblent et s'engagent dans un cycle de différenciation et de morphogénèse, Figure 6.



**Figure 6 – Développement du cycle de vie de l'amibe *Dictyostelium discoideum*.**

Une phase d'agrégation est induite par chimiotaxie et dirigée par l'impulsion de gradients d'AMPc (Konijn, *et al.* 1967). L'AMPc excrété par une cellule donnée est détecté par les récepteurs membranaires des cellules voisines qui migrent vers la cellule émettrice et produisent et sécrètent de l'AMPc à leur tour. Le signal chimiotactique est ainsi relayé de proche en proche conduisant à la formation d'ondes spirales. Dans les agrégats (environ  $10^5$  cellules), les cellules subissent des processus de différenciation et de morphogenèse afin de former une sorte de petit limaçon multicellulaire qui est capable de réagir à la lumière et à la température. Il se déplace ainsi vers une zone favorable où il se transformera en une fructification d'environ 2mm de hauteur. Ce corps fructifère se compose d'une tige mince formée par un empilement des cellules mortes avec, à son extrémité, une masse de spores rassemblées dans une boule de mucus. Les spores résistent à la dessiccation et aux températures extrêmes. Leur position élevée au sommet de la tige facilite leur dispersion par les petits animaux du sol, insectes ou autres. (Firtel and Meili 2000). Lorsque les conditions environnementales conviennent, les spores germent pour produire à nouveau des amibes qui recommencent le cycle de vie végétatif (Kessin 2001).

Alors que la bactérie *Escherichia coli* est l'organisme procaryote le plus couramment employé dans la génétique moléculaire, l'amibe *D. discoideum* est un des organismes modèles

eucaryotes très utilisé pour étudier les grandes thématiques de la biologie : la transduction du signal, la communication cellulaire, la différenciation, la mort cellulaire, la phagocytose et la motilité cellulaire (Willard and Devreotes 2006). Elle est aussi utilisée comme organisme modèle pour les études des interactions hôte-pathogène (Steinert and Heuner 2005).

Récemment, le génome complet *D. discoideum* souche AX4 a été séquencé (Eichinger, *et al.* 2005) et de nombreuses informations sont disponibles sur le site dictyBase (<http://dictybase.org/>) (Chisholm, *et al.* 2006, Kreppel, *et al.* 2004). Son génome de 34Mb qui se caractérise par une grande richesse en [A+T] (77%) est composé de 6 chromosomes codant environ 12.500 protéines. *Dictyostelium* présente une grande quantité des gènes homologues à ceux d'eucaryotes plus complexes. L'arbre phylogénétique qui s'appuie sur une analyse des données génomiques et protéomiques associées montre que *D. discoideum* diverge de la branche animal-champignon après la séparation entre les plantes et les animaux, Figure 7.

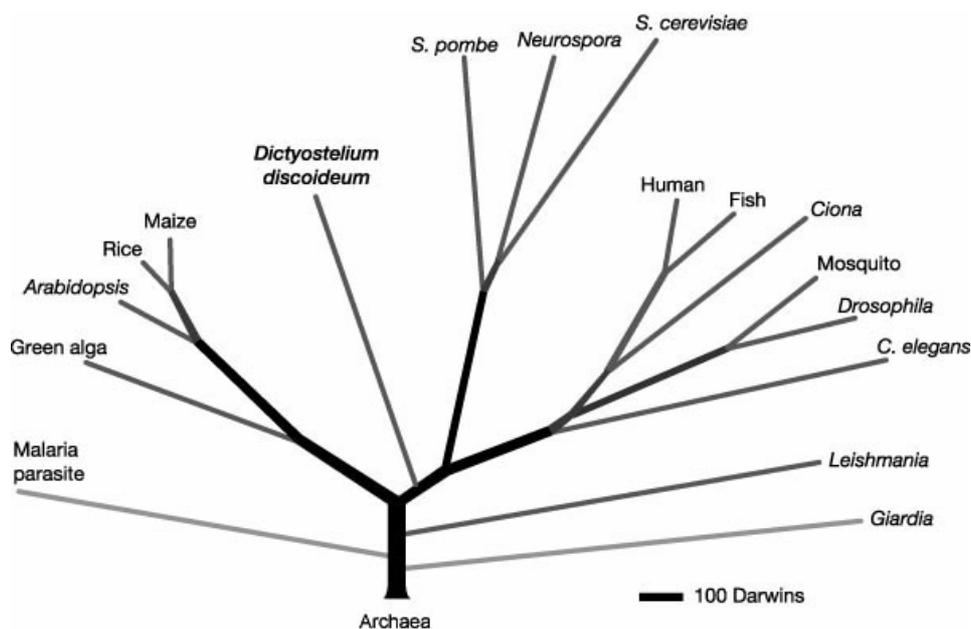


Figure 7 – Arbre phylogénétique qui place *Dictyostelium discoideum* après la séparation entre les plantes et les animaux (Eichinger, *et al.* 2005).

### 1.3.3 Historique des lectines dans *Dictyostelium discoideum*

Un nombre très limité de gènes de *D. discoideum* sont annotés comme codants pour des lectines. Ce sont des séquences codant pour des domaines ricines présents dans 10 protéines

CupA-J (Cup pour Calcium up-regulated genes) ou des CRD de lectines de type C dans 2 protéines (*DDB\_G0269040* et *DDB\_G0289907*). Elles sont toutes très peu caractérisées au plan fonctionnel. Le gène *comA* code pour une protéine de la famille des comitines, des protéines impliquées dans le trafic intracellulaire et présentant un domaine lectine (Jung, *et al.* 1996). La comitine de *D. discoideum* a été purifiée et caractérisée (Stratford and Brown 1985) et un modèle du domaine lectine complexé avec le mannose a été proposé en utilisant la similarité avec la lectine de perce-neige (Barre, *et al.* 1999). Enfin, des lectines à Gal/GalNAc, les discoidines, ont fait l'objet d'un fort intérêt au cours des années 70 et 80.

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. L'activité hémagglutinante spécifique augmente rapidement au cours des 12 h après une période de jeûne (Rosen, *et al.* 1973, Simpson, *et al.* 1974). Le groupe de Rosen et Barondes a démontré ensuite que la protéine lectin-like présente à la surface de *D. discoideum* était constitué de deux lectines, les discoidines I et II (DiscI et DiscII) (Frazier, *et al.* 1975). Ces protéines présentent des similarités mais sont différentes sur de nombreux points tels que 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 4) (Alexander, *et al.* 1983, Barondes, *et al.* 1983, Berger and Armant 1982, Frazier, *et al.* 1975).

	DiscIA	DiscII
Spécificité	GalNAc>Galactose	GalNAc ≈Galactose
Poids moléculaire	28258.5	28574.7
Point isoélectrique	6.20	6.65
Nombre d'acides aminés	253 aa	257 aa
Coefficient d'extinction	1.50	1.59

**Tableau 4 – Comparaison entre les propriétés biochimiques des DiscI et DiscII.**

### 1.3.4 La discoidine I

La discoidine I (DiscI) a d'abord été décrite comme un tétramère de sous-unités de 26 à 29.5 kDa, non-glycosylées et de point isoélectrique 6.1 (Frazier, *et al.* 1975, Ishiguro and Weeks 1978, Simpson, *et al.* 1974). Nos travaux ont démontré que la DiscI forme des trimères. L'analyse génomique a démontré que la discoidine I est en fait codée par une famille de plusieurs gènes qui résultent en trois isoformes pour la protéine (Poole, *et al.* 1981, Rowekamp, *et al.* 1980, Tsang, *et al.* 1981) qui sont appelées chaînes A B/C et D (Figure 8) (P02886, P02887 et P02288 dans Uniprot). Les séquences sont très similaires avec l'isoforme A présentant 96 et 97% d'identité (4 et 3 acides aminés différents) avec les isoformes C et D (seule l'isoforme A sera décrite par la suite). Deux gènes codant pour des protéines similaires (discoidine1-like) ont également été identifiés.

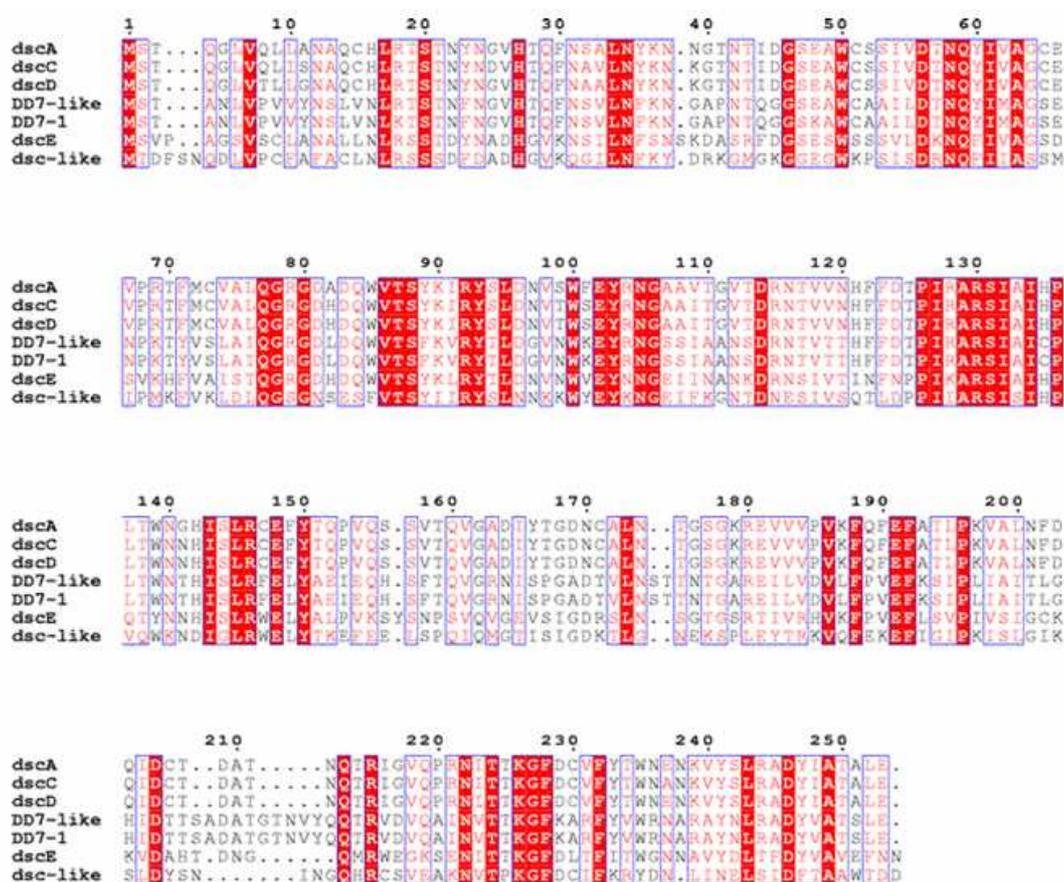


Figure 8 – Aligment de séquences des trois isoformes de DiscI (dscA, dscC et dscD) et des deux protéines DiscI-like, DD7-like et DD7-1. Les séquences de la DiscII (dscE) et une protéine similaire (dsc-like) sont également représentées.

La présence d'une deuxième copie des trois gènes de la discoidine I a été observée dans les souches AX3 et AX4 de *D. discoideum* (dictyBase). La DiscI est constituée de 253 acides aminés avec une masse moléculaire théorique de 28258,5 et un point isoélectrique de 6,2. Deux domaines ont été identifiés : le domaine *N*-terminal (1-152), correspond au domaine discoidine couramment appelé domaine « discoidin-like » (Kiedzierska, *et al.* 2007), suivi d'un domaine lectine *C*-terminal (157-253) appartenant aux lectines de type H (Sanchez, *et al.* 2006). Les caractéristiques de ces deux domaines sont discutées ci-dessous.

La localisation et la régulation de l'expression de la DiscI ont été très étudiées, alors que la fonction biologique n'est pas encore clairement élucidée. Sa production n'apparaît que dans les cellules cohésives, un phénotype induit en stoppant l'alimentation des cellules végétatives (Barondes *et al.*, 1983). DiscI commence à être produite entre 9 et 12 h après l'induction et s'accumule dans les cellules du manteau de la limace et de la tige du corps fructifère (Cooper and Barondes 1984, Fukuzawa and Ochiai 1993). L'expression des gènes codant pour DiscI est régulée de manière positive durant l'agrégation (Devine, *et al.* 1982, Ma and Firtel 1978).

Les études au niveau du transcriptome montre que l'ARN messager de la DiscI est présent avec une fréquence de moins d'une copie par gène durant la phase végétative et augmente de 500-1000 fois pendant les premières heures de développement (Rowekamp, *et al.* 1980). Il est également possible d'induire la production de DiscI durant la phase végétative en traitant les cultures avec des facteurs protéiques sécrétés par *Dictyostelium* comme le CMF (« conditioned medium factor ») (Jain, *et al.* 1992) ou le PSF (« pre-starvation factor ») (Clarke, *et al.* 1987). Un réseau de régulation complexe impliquant une répression par l'AMPc et une induction en deux étapes, d'abord par le PSF, puis par le CMF a été mis en évidence (Blusch, *et al.* 1995). L'expression des gènes de la discoidin I est utilisée comme marqueur du développement précoce chez *Dictyostelium* (Strmecki, *et al.* 2007).

Il a été proposé que DiscI soit impliquée dans le processus de reconnaissance cellule-cellule (Cano and Pestana 1984) ou dans la reconnaissance cellule-substrat (Springer, *et al.* 1984). Les premières études ont proposé que la lectine soit concentrée dans des corps multilamellaires qui serviraient à son exportation (Barondes, *et al.* 1985). La localisation à la surface des cellules cohésives est contreversée (Erdos and Whitaker 1983). Il s'agit d'une protéine cytoplasmique qui est nécessaire au maintien de la morphologie allongée qui correspond au phénotype agrégatif (Alexander, *et al.* 1992). Les études sur des mutants de

délétion n'ont pas montré de défaut d'agrégation mais les cellules perdent leur mobilité sur des surfaces plastiques (Alexander, *et al.* 1983, Springer, *et al.* 1984).

La DiscI agglutine très efficacement les érythrocytes de lapin, et dans une moindre mesure ceux de mouton (Frazier, *et al.* 1975). Cette agglutination est inhibée par des monosaccharides ayant une configuration D-*galacto* tels que le D-galactose, D-fucose, D-GalNAc ainsi que par les disaccharides contenant une unité D-galactose tels que le lactose et le mellibiose. Le GalNAc est le monosaccharide qui se lie le plus fortement à la discoidine I. Le L-fucose et le D-glucose présentent un faible potentiel d'inhibition. L'interaction entre la lectine et les glucides nécessite la présence d'ions bivalents, avec une préférence pour le calcium dans les interactions *in vitro* (Alexander, *et al.* 1983). La recherche de récepteurs endogènes ou exogènes pour la discoidine I n'a pas donné de résultats clairs. Pour le domaine lectine, GalNAc-dépendant, plusieurs glycoprotéines de *D. discoideum* agissent comme récepteurs pour DiscI (Breuer and Siu 1981, Cooper and Barondes 1984, Ray, *et al.* 1979). Une comparaison avec *Polysphondylium pallidum*, un autre dictyostélide, suggère une spécificité d'espèce pour ces récepteurs (Reitherman, *et al.* 1975). Des polysaccharides bactériens extraits de *Escherichia coli* et *Klebsiella aerogenes*, utilisés comme proies pour les cultures d'amibes, ont aussi la capacité d'être reconnus par la lectine (Cooper, *et al.* 1983). Le domaine RGD présent dans le domaine discoidin DS indique également la possibilité d'une liaison indépendante du domaine lectine avec un récepteur spécifique de ce motif (Gabijs, *et al.* 1985).

### 1.3.5 La discoidine II

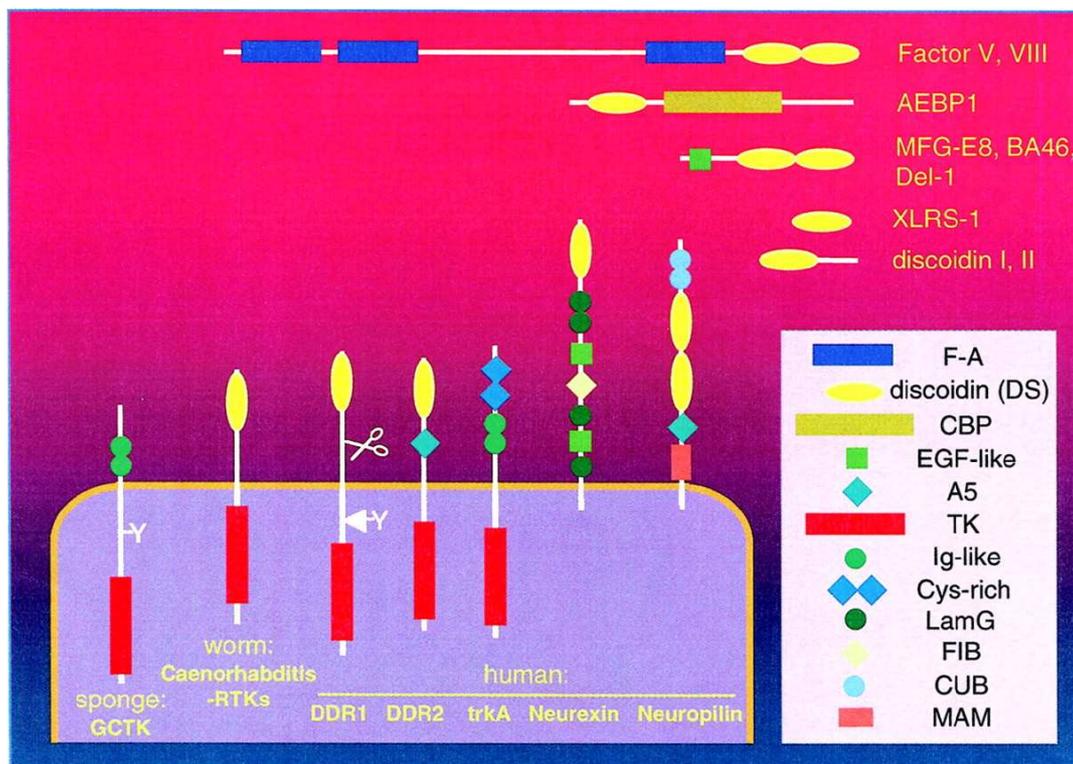
Les premières études sur la discoidine II (DiscII) l'ont décrite comme relativement similaire à DiscI puisque sous la forme d'un tétramère de sous-unités de 23 kDa (Frazier, *et al.* 1975), et présentant des similarités de séquences au niveau des peptides tryptiques (Berger and Armant 1982). Cependant des différences claires ont été observées : la discoidine II reconnaît le galactose avec la même affinité que le GalNAc et cette interaction est indépendante du calcium (Alexander, *et al.* 1983). DiscII est codée par un seul gène qui a été caractérisé en 1996 (Fukuzawa and Ochiai 1996). La séquence peptidique résultant de la traduction comporte 257 acides aminés pour un poids moléculaire de 28575 Da et un point isoélectrique de 6,65. DiscI et DiscII ont seulement 48% d'identité de séquence mais elles

partagent la même organisation en domaine, avec un DS domaine en N-terminal et un domaine H-type lectine en C-terminal.

Les informations sur la distribution spatio-temporelle, la régulation et la fonction biologique de DiscII sont beaucoup plus limitées que celles disponibles pour DiscI. DiscII est également caractéristique des cellules cohésives mais elle exprimée un peu plus tardivement que DiscI et semble s'accumuler dans les vésicules des prés pores (Barondes, *et al.* 1983, Cooper and Barondes 1984, Fukuzawa and Ochiai 1993). Dans ces vésicules, DiscII se fixerait spécifiquement à des polysaccharides qui sont les constituants majeurs de la paroi des spores (Cooper, *et al.* 1983).

### 1.3.6 Le domaine DS

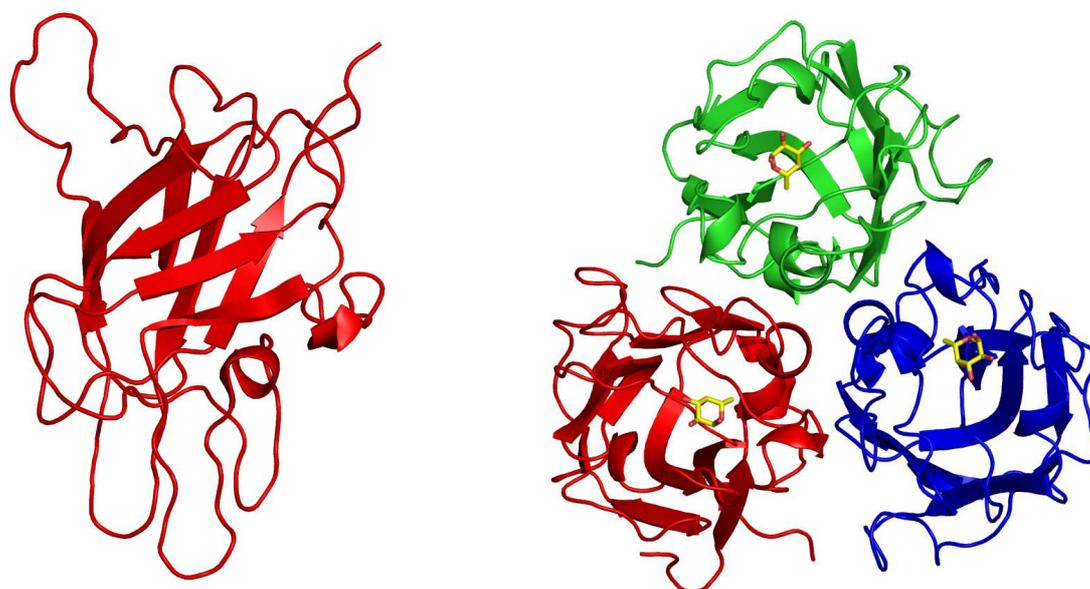
A la suite du séquençage de la DiscI (Poole, *et al.* 1981), le motif formé par les 150 premiers acides aminés a été retrouvé dans un grand nombre de protéines (Kiedzierska, *et al.* 2007). Ce domaine est retrouvé sous les noms de domaine discoidin-like, motif discoidine, FAAA58C ou F5/8C. Il a été détecté seul, ou en répétition tandem, dans de nombreuses protéines telles que des facteurs de coagulation, des enzymes, des récepteurs et des protéines impliquées dans le développement neuronal (Baumgartner, *et al.* 1998, Vogel 1999), Figure 9. Le rôle biologique de ce domaine est basé sur ses capacités d'interactions avec une variété d'autres molécules telles que les facteurs de croissance, les collagènes, les lipides et phospholipides, et certains glucides.



**Figure 9 – Schéma représentatif de quelques protéines à domaine DS.** Abréviations F-A, domaine des facteurs V et VIII; CBP, carboxypeptidase; EGF, facteur de croissance épidermique; TK, tyrosine kinase; A5, homologue de l'antigène A5; Ig, immunoglobuline; LamG, laminin-G; FIB, fibronectin-like; CUB, complément binding; et MAM, meprin/A5/PTPmu. (Vogel 1999).

Les domaines DS peuvent appartenir à la région extracellulaire de protéines transmembranaires comme les récepteurs à domaine discoidine (DDRs) qui appartiennent à la famille des récepteurs tyrosine kinases activés par le collagène (Curat, *et al.* 2001). Les neuropilines et la neurexine V sont ancrées dans les membranes et jouent un rôle dans le développement du système nerveux (Peles, *et al.* 1997, Soker, *et al.* 1998). Les domaines DS ont également été identifiés dans des protéines circulantes telles que les facteurs de coagulation V et VIII qui lient les phospholipides à la surface des plaquettes sanguines (Fuentes-Prior, *et al.* 2002). Plus récemment, la structure cristallographique de la lectine d'anguille (AAA) complexée avec le fucose a permis d'identifier des similarités structurales avec les domaines DS (Bianchet, *et al.* 2002). Cette famille, appelée lectine de type-F est présente chez d'autres poissons et autres eucaryotes ainsi chez certains procaryotes (Vasta, *et al.* 2004).

La première structure cristallographique d'un domaine DS est celle de la galactose oxydase (Ito, *et al.* 1991). Depuis un grand nombre de domaines DS a été résolu (facteurs de coagulation, neuropilins...). Le repliement DS consiste en 8 brins  $\beta$  formant un tonneau  $\beta$  compact. Les parties N- et C-terminales sont à une des extrémités du tonneau, alors que l'autre extrémité présente trois boucles variables qui forment la surface de reconnaissance de ligand. Ces domaines apparaissent être monomériques, sauf pour la lectine d'anguille AAA où les trois domaines DS s'associent en trimère et présentent les trois sites à fucose sur une même face (Bianchet, *et al.* 2002) (Figure 10).



**Figure 10 – Repliement général du monomère du facteur V (FVa - C2) (PDB 1CZT) (Macedo-Ribeiro, et al. 1999) (à gauche) et du trimère de la lectine AAA de *Anguilla anguilla* complexée au fucose (PDB 1K12) (Bianchet, et al. 2002) (à droite).**

### 1.3.7 Le domaine lectine C-terminal et la famille de type H

#### 1.3.7.1 Lectines de la famille H

La fonction du domaine C-terminal n'a été identifiée que récemment, lors du clonage de la lectine du corail *Sinularia lochmodes* (Jimbo, *et al.* 2005). La structure cristallographique de la lectine de l'escargot *Helix pomatia* (HPA) a ensuite permis de définir ce nouveau type de lectine à Gal/GalNAc, appelée lectine de type H (Sanchez, *et al.* 2006).

La recherche d'autres séquences similaires aux discoidines peut être réalisée, par exemple avec le programme Blast (Altschul, *et al.* 1990). En supplément des lectines de corail

(SLL-2a, SLL-2b et SLL-2c), d'escargots *Helix pomatia* (HPA) et *Cepaea hortensis* (CPA), on peut identifier des séquences fongiques. Des protéines hypothétiques des champignons *Gibberella zeae* (XP\_388374) et une autre protéine de *Coprinopsis cinerea* (XP\_001838808) présentent 3 domaines lectine de type H répétés (Thèse M.Walti, ETH Zürich 2008). Un alignement des séquences de lectines de type H est représenté dans la Figure 11.

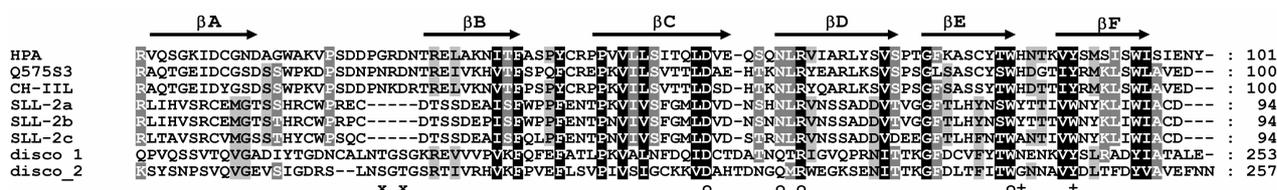


Figure 11 – Alignement des séquences de *Helix pomatia* (structure cristallographique HPA et déposition uniprot Q575S3), de la lectine d'escargot *Cepaea hortensis* (CHP-II), des trois isoformes de la lectine de corail *Sinularia lochmodes* (SLL-2a, SLL-2b et SLL-2c).

### 1.3.7.2 La lectine de Corail *Sinularia lochmodes* (SLL-2)

Des travaux récents ont permis d'établir l'implication de lectines dans la symbiose qui réunit plusieurs espèces de coraux mous avec des algues unicellulaires dinoflagellées (Fenton-Navarro, *et al.* 2003, Koike, *et al.* 2004, Wood-Charlson, *et al.* 2006). Cette association est nécessaire à la survie du corail et la perte de l'algue entraîne le « blanchiment » des récifs et leur mort. Le corail *Sinularia lochmodes* produit une lectine à galactose que l'on retrouve à la surface des cellules de *Symbiodinium* libérées du corail, Figure 12 (Jimbo, *et al.* 2000). Cette lectine présente trois isoformes la SLL-2a, la SLL-2b et la SLL-2c qui peuvent reconnaître spécifiquement les sucres D-galactose et N-acétyl- D-galactosamine (Jimbo, *et al.* 2000).



Figure 12 – Le corail *Sinularia lochmodes*.

L'analyse de la séquence protéique de SLL-2 a démontré une forte similarité avec le domaine C-terminal des discoidines de l'amibe *Dictyostelium discoideum* (33%) (Jimbo, *et al.* 2005). Lorsque la séquence de la lectine HPA a été séquencée, il est apparu qu'elle appartenait à la même famille en termes de similarité de séquence. Une nouvelle famille de lectine, les lectines de type-H, a donc été définie (Sanchez, *et al.* 2006).

### 1.3.7.3 La lectine HPA de *Helix pomatia*

À la fin des années 1960, une lectine a été découverte chez l'escargot de bourgogne *Helix pomatia* de la classe des *Gastropoda phylum Mollusca*, Figure 13 (Prokop, *et al.* 1968) alors que les propriétés agglutinantes d'extraits de la glande à albumen de l'escargot était connue dès la fin du 19<sup>ième</sup> siècle (Camus 1899). Cette lectine représente 8% des protéines de cette glande qui appartient au système de reproduction de l'escargot et produit le fluide perivitellin qui entoure les oeufs. Sa caractéristique principale est d'agglutiner spécifiquement les érythrocytes du groupe sanguin A, d'où vient son nom de *Helix Pomatia* « Agglutinin » (HPA) (Hammarström and Kabat 1971). Le rôle biologique de HPA n'est pas encore très bien connu mais les études précédentes suggèrent que HPA pourraient servir de protection aux œufs, en agglutinant les éventuels microorganismes pathogènes.

Des nombreuses études ont montré la spécificité de HPA pour le résidu N-acétylgalactosamine (GalNAc). Par des tests d'inhibition de la réaction d'agglutination des érythrocytes du groupe A, les chercheurs ont montré que le méthyl- $\alpha$ -D-GalNAc interagit de façon efficace avec HPA (Hammarström and Kabat 1971). Ce résidu est bien présent dans l'épitope glucidique du groupe sanguin A : GalNAc $\alpha$ 1-3[Fuca1-2]Gal $\beta$ 1-R. Le ligand de plus forte affinité est une pentasaccharide, l'antigène de Forssman (GalNAc $\alpha$ 1-3GalNAc $\beta$ 1-3Gal $\alpha$ 1,4Gal $\beta$ 1-4Glc) (Baker, *et al.* 1983).



**Figure 13 – Escargot *Helix pomatia* (HPA). La lectine HPA est localisée principalement dans le système reproducteur de cet escargot (Camus 1899), figure issue de Nature Photo-CZ.**

HPA est une protéine actuellement largement utilisée en histopathologie comme marqueur spécifique de certains cancers. En effet, le processus tumoral entraîne tout un ensemble de modifications dans l'expression des gènes et ces dérèglements peuvent avoir une influence sur les motifs oligosaccharidiques portés par les protéines (Dennis, *et al.* 1999). HPA se lie spécifiquement à la surface des cellules de cancer du sein et du colon à haut pouvoir métastatique et représente donc un intérêt en cancérologie (Brooks 2000, Mitchell and Schumacher 1999). Cette protéine est utilisable pour la détection rapide des cellules métastatiques dans les biopsies, ou les échantillons chirurgicaux ou même comme vecteur de médicament. Cependant, des différentes études réalisées avec cette lectine ont constaté une variabilité dans les résultats selon la méthode histochimique utilisée. Cette variabilité est attribuable en partie à des variations dans les lots de HPA obtenus à partir d'extraits d'escargots, ce qui pose des problèmes de pureté et reproductibilité. L'absence de procédure standardisée a limité l'intérêt de l'utilisation de cette lectine comme outil d'investigation en routine en oncologie (Brooks, *et al.* 1996).

La structure cristallographique de la lectine HPA a été obtenue dans notre laboratoire au CERMAV (Sanchez, *et al.* 2006). Les monomères de 13 kDa sont constitués de deux feuillets  $\beta$  et s'associent dans un arrangement trimérique. Un hexamère fonctionnel est formé de la jonction de deux trimères par de ponts disulfures. La structure cristalline de HPA en complexe avec le GalNAc a également été résolue, Figure 14. Les sites de liaison, à l'interface des monomères, sont localisés sur les deux faces opposées de l'hexamère formées par la lectine.

La structure de HPA a également été résolue en complexe avec deux épitopes présents sur certaines cellules cancéreuses : l'antigène Tn (GalNAc $\alpha$ -O-Ser/Thr) et l'antigène Forssman ( $\alpha$ GalNAc1-3GalNAc). Les structures des différents complexes montrent que la spécificité de cette lectine pour GalNAc est due à un réseau particulier de liaisons hydrogènes (Lescar, *et al.* 2007).

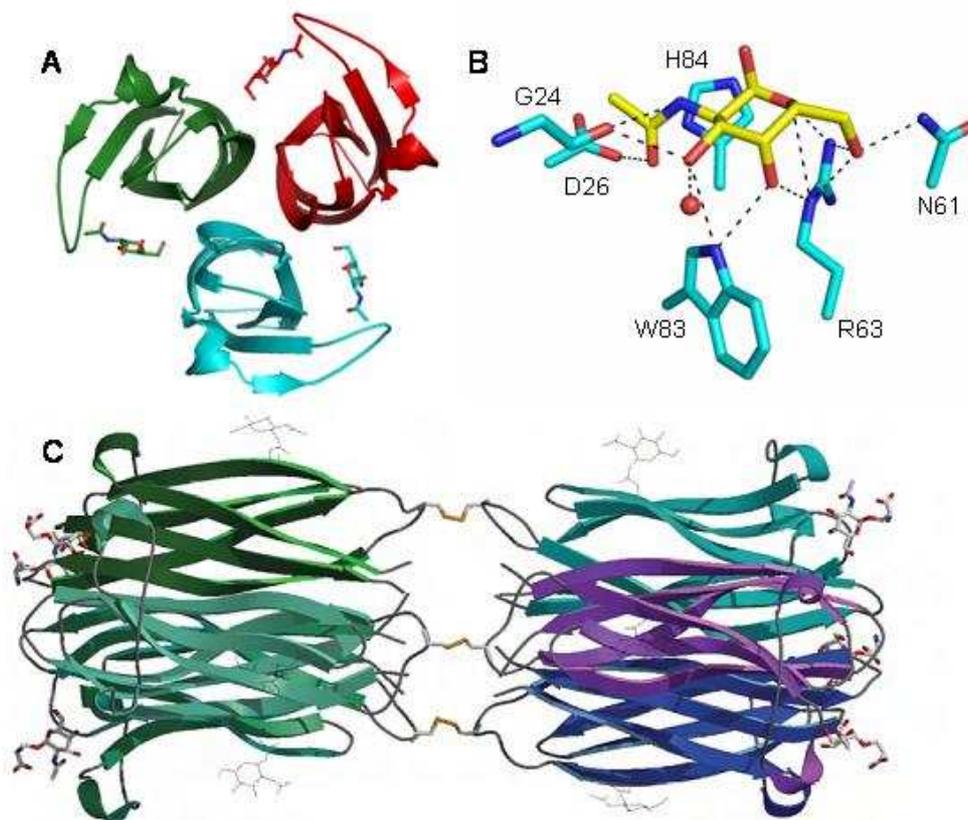


Figure 14 – A) Représentation d'un trimère de lectine de *Helix pomatia* (HPA) en complexe avec N-acétylgalactosamine (GalNAc) B) représentation du site de liaison d'HPA avec le GalNAc. La sphère rouge représente une molécule d'eau et le ligand (GalNAc) est représenté par des bâtonnets jaunes. (Sanchez, *et al.* 2006). C) Hexamère du complexe HPA-antigène Tn (GalNAc $\alpha$ -O-Ser/Thr).

## 1.4 Objectifs de l'étude

Le développement d'outils biotechnologiques pour des applications thérapeutiques devient d'un grand intérêt. Notamment, aujourd'hui, la technique de marquage spécifique en histopathologie peut apporter des progrès importants dans la lutte contre le cancer. Cette approche regroupe différentes stratégies qui servent par exemple à l'identification rapide du caractère métastatique de cellules cancéreuses. Dans ce contexte, l'étude structurale et fonctionnelle de lectines considérées comme d'excellents modèles pour la compréhension de la reconnaissance protéine–sucre offrent d'importantes perspectives pour les applications biotechnologiques comme rechercher, diagnostiquer tôt et soigner les cancers.

La lectine HPA d'escargot *Helix pomatia* représente un marqueur capable de cibler spécifiquement les cellules tumorales 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 en routine en oncologie (Brooks, *et al.* 1996). Le manque de reproductibilité de certains marquages peut être attribué aux variations dans les lots de HPA, qui sont obtenus à partir d'extraits d'escargots. Au laboratoire, des contaminations de 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. 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 HPA sous forme recombinante malgré de nombreux essais dans *Escherichia coli* et *Pichia pastoris*. 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* présentent des similarités suffisantes dans leur région C-terminale pour être de bonnes candidates pour une approche d'ingénierie protéique afin de mimer les propriétés de HPA en outil diagnostic.

Le but principal de cette thèse est l'étude de la structure et de la fonction des lectines à Gal/GalNAc : les discoidines de l'amibe *Dictyostelium discoideum* en utilisant la cristallographie aux rayons X, la microcalorimétrie et les analyses par « Puces à sucres ». Il s'agit également d'évaluer les similarités et les différences entre les discoidines et la lectine HPA. L'étude de discoidines permet aussi une meilleure compréhension de leur rôle biologique dans la vie de l'amibe *Dictyostelium discoideum* étant donné que cette amibe est utilisée comme organisme modèle en diverses expérimentations biologiques.

Cette thèse comprend donc quatre objectifs principaux:

- Mettre au point l'expression et la purification des lectines DiscI et DiscII chez *E. coli* suite au clonage des gènes afin d'obtenir les protéines recombinantes et de les purifier par des techniques de chromatographie classiques.

- Évaluer leur affinité et leur spécificité pour différents mono et oligosaccharides par Microcalorimétrie isotherme de titration et par l'utilisation de « Puces à sucres ».

- Cristalliser et déterminer leur structure 3-D afin d'obtenir les premières structures cristallographiques de lectines de l'amibe *Dictyostelium discoideum*. Ceci permettra d'obtenir de nouvelles structures de lectines de la famille du type H pour lesquelles seule la lectine HPA est connue. Dans un premier temps, il s'agira de cristalliser Disc II sauvage et ensuite, de cristalliser les deux lectines sous forme recombinantes native et complexées avec des monosaccharides.

- Enfin, établir les similarités et les différences au niveau structural, de spécificité et d'affinité entre les discoidines et la lectine HPA.

L'étude de ces deux lectines présentes chez l'amibe (DiscI et DiscII) et leurs caractérisations biochimiques structurales fourniront des informations sur les interactions protéines-sucres nécessaires pour comprendre les mécanismes d'action, de reconnaissance, de spécificité et d'affinité à niveau moléculaire. À plus long terme, ces deux lectines produites sous forme recombinante pourront être testées pour leur capacité à reconnaître les glycannes portés par une tumeur cancérogène. Ainsi les modifications de structure des résidus glycaniques membranaires induites par le développement d'un cancer pourraient être mieux étudiées et possiblement donner naissance à des vaccins anti-cancer.



# **Les outils de l'étude de l'interaction lectine-sucre**



## 2 Les outils de l'étude de l'interaction lectine-sucre

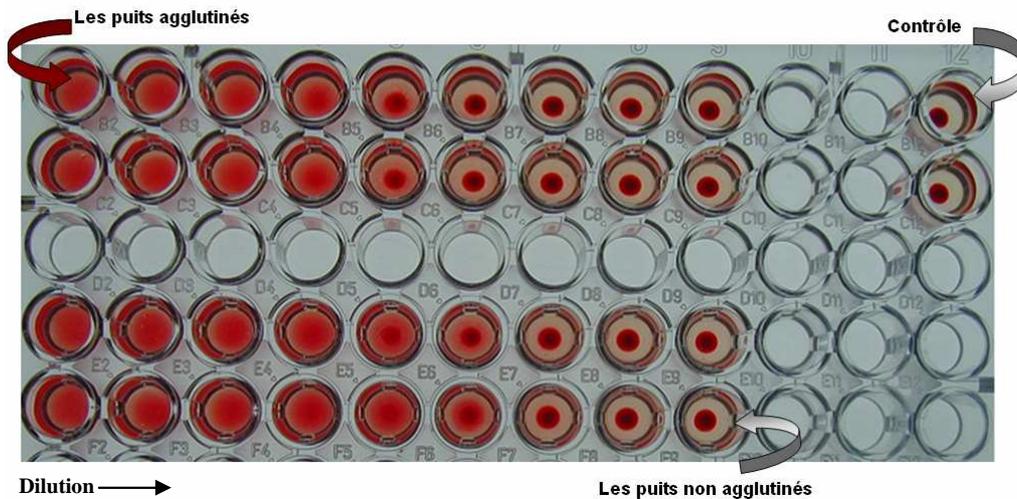
À l'heure actuelle, il existe plusieurs techniques expérimentales qui permettent d'évaluer les interactions lectine-sucre. Certaines mesurent la capacité des lectines à inhiber, à réticuler ou à agréger les sucres. D'autres mesurent la spécificité et/ou l'affinité de ces protéines pour leurs ligands et permettent de comprendre les forces d'interaction mises en jeu.

Parmi les techniques les plus utilisées, nous pouvons citer: le test d'inhibition d'hémagglutination (HIA, Hemagglutination Inhibition Assay), le test par une lectine liée à une enzyme (ELLA, Enzyme-Linked Lectin Assay), la microcalorimétrie par titrage isotherme (ITC, Isothermal Titration Microcalorimetry), la résonance plasmonique de surface (SPR, Surface Plasmon Resonance), les analyses par puce à sucre et la cristallographie des protéines au rayon X qui permet une investigation structurale des lectines. La complémentarité de ces techniques permettra d'élucider le rôle biologique et le fonctionnement des lectines à un niveau moléculaire.

### 2.1 Test d'hémagglutination

La mesure d'activité hémagglutinante est le test le plus simple et le plus largement utilisé pour la détection des lectines et leur caractérisation (Goldstein, *et al.* 1980, Rüdiger 1993). Il est basé sur la propriété de ces protéines de lier des glycoconjugués de la surface des érythrocytes. Ce test repose sur l'observation de l'agglutination (ou agrégation) des érythrocytes par les lectines visible à l'œil nu sous forme d'une phase gélatineuse et sur la détermination du point d'équivalence qui est la concentration minimale de lectine montrant une agglutination évidente, Figure 15. Les érythrocytes de plusieurs mammifères sont parmi les plus utilisées (humain, lapin, mouton, porc, etc...).

L'activité hémagglutinante des lectines DiscI et DiscII sauvages ou recombinantes a été réalisée dans des microplaques de titration. Nous avons fait deux dilutions en série. 25  $\mu$ L de tampon Tris 100 mM pH 7,6, NaCl 150 mM ont été placés dans chaque puits de la microplaque, puis 25  $\mu$ L des lectines (1 mg/mL) ont été ajoutés dans le premier puits. Une double dilution en série a été réalisée jusqu'au 10<sup>ème</sup> puits. Ensuite, 25  $\mu$ L de la suspension d'érythrocytes 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 après 30 minutes d'incubation et après 12 heures à température ambiante.



**Figure 15 – Le test d'hémagglutination sur microplaque de titration.**

En présence d'un ligand soluble, l'hémagglutination peut être entièrement inhibée et la concentration minimale du ligand qui empêche l'hémagglutination est mesurée et donne le potentiel d'inhibition du ligand. Ce dernier est généralement dépendant de variables telles que la source et l'âge des érythrocytes. Le test d'inhibition d'hémagglutination est un test rapide et simple qui permet d'évaluer les ligands solubles de façon comparative, donnant un ordre approximatif d'avidité.

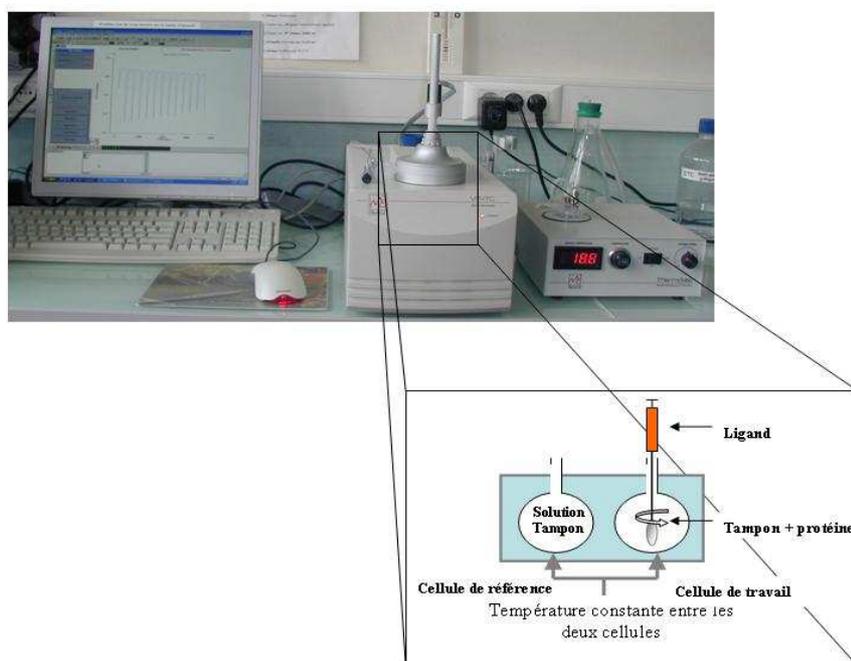
## 2.2 La calorimétrie isotherme de titration (ITC)

Depuis 1977, la microcalorimétrie isotherme de titration (ITC) occupe une place de plus en plus importante en biologie. Elle permet l'étude des interactions entre les protéines et leurs ligands qui peuvent être : d'autres protéines, des glucides, des acides nucléiques, etc... Elle permet d'obtenir tous les paramètres thermodynamiques caractérisant l'interaction en une seule expérience. Il est en effet possible de mesurer et de calculer la constante d'affinité ( $K_a$ ), la stoechiométrie d'interaction ( $n$ ), la variation d'enthalpie libre  $\Delta G$ , l'entropie  $\Delta S$  et l'enthalpie ( $\Delta H$ ).

### 2.2.1 Principe de la microcalorimétrie isotherme de titration

Le microcalorimètre contient deux cellules isolées de l'environnement externe par une enceinte adiabatique qui est maintenue à température constante (25 °C) pendant le temps de la titration. Une cellule de référence contient du tampon ou de l'eau uniquement. La solution

contenant la protéine est placée dans une deuxième cellule, dite cellule de travail ou de mesure. Dans la Figure 16, nous pouvons voir une représentation de l'appareil.



**Figure 16 – Microcalorimètre Microcal VP-ITC (en haut) et schéma de fonctionnement (en bas).**

Cette méthode fonctionne par la mesure d'un dégagement ou d'une absorption de chaleur (réaction exothermique ou endothermique, respectivement) pendant la saturation d'une protéine par un ligand. En général, la lectine est placée dans la cellule de mesure sous agitation continue et le sucre y est ajouté à intervalle régulier à l'aide d'une seringue de 300  $\mu\text{L}$  par une série de 30 injections espacées d'un temps suffisant (3-10 mn). A chaque injection, l'interaction entre les deux molécules s'accompagne d'un dégagement ou d'une absorption de chaleur qui entraîne une variation de la température de la cellule de travail. Afin de rééquilibrer la température entre les deux cellules, l'appareil applique une puissance électrique et c'est ce signal qui est mesuré par le microcalorimètre.

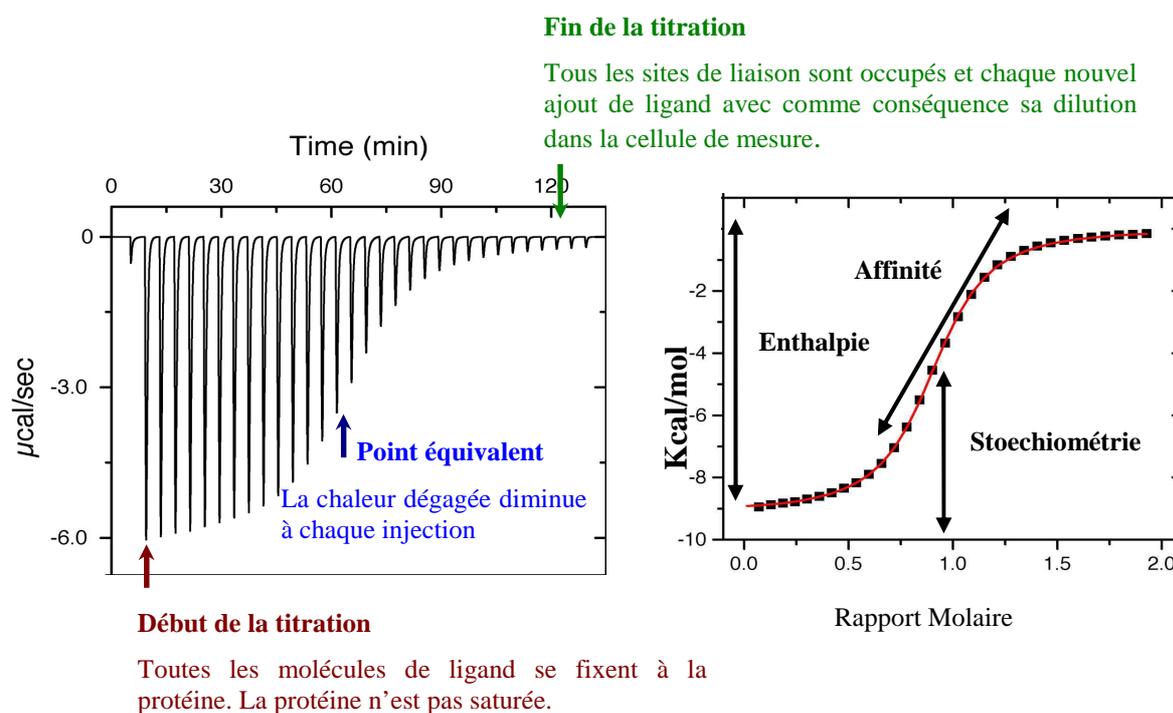
Au début de la titration, la protéine est progressivement saturée par le ligand et, dans le cas d'une réaction exothermique, des quantités mesurables de chaleur sont dégagées indiquant la formation du complexe protéine-ligand. La variation de température du milieu permet d'en déduire la quantité de chaleur échangée. En fin de réaction, tous les sites de liaison sont déjà occupés et chaque nouvel ajout de ligand a comme seule conséquence sa dilution dans la cellule de travail. L'énergie  $\mu\text{cal}/\text{sec}$  est enregistrée en fonction du temps puis intégrée. La

variation d'enthalpie de la liaison,  $\Delta H$  (kcal/mol); la constante d'association,  $K_a$  ( $M^{-1}$ ) et la stoechiométrie d'interaction ( $n$ ) qui indique le nombre de sites de liaison par monomère de la protéine peuvent être déterminées par ajustage sur l'équation d'échange de chaleur (plusieurs modèle mathématiques possibles suivant le nombre de site, la coopérativité...) (Dam and Brewer 2002). Les autres paramètres thermodynamiques, les variations d'entropie ( $\Delta S$ ) et d'énergie libre ( $\Delta G$ ) peuvent être déduits des relations:

$$\Delta G = -RT \ln K_a$$

$$\Delta G = \Delta H - T \Delta S$$

où  $T$  est la température absolue en Kelvin et  $R$  la constante des gaz parfaits,  $8.314 \text{ J mol}^{-1} \cdot \text{K}^{-1}$ . La forme de la courbe intégrée est fonction de la concentration en lectine et en sucre et de la constante d'association du complexe (Figure 17).



**Figure 17 – Courbe brute et courbe du titrage d'un VP-ITC. (A) Energie enregistrée en fonction du temps. (B) Courbe de la chaleur totale libérée en fonction de la concentration totale en ligand; calculée à partir de la courbe A.**

Les systèmes biologiques ont tendance à avoir l'énergie la plus basse possible. L'enthalpie  $\Delta H$  tend à être négative. La réaction est spontanée lorsque  $\Delta G$  est négatif. Cette

énergie est en fait la résultante de deux contributions énergétiques: l'enthalpie  $\Delta H$  et l'entropie  $\Delta S$  (kcal/mol). D'un point de vue thermodynamique, la constante d'affinité  $K_a$  est fonction de l'énergie libre de l'interaction, plus cette constante est grande, plus l'affinité du ligand pour la protéine est forte.

La contribution enthalpique reflète la force des interactions qui maintiennent le ligand avec la protéine et avec le solvant. Ces liaisons sont principalement dues à la formation de liaisons hydrogènes et aux interactions de Van der Waals. Une enthalpie d'interaction favorable requiert un placement correct des groupes donneurs et accepteurs d'électrons dans le site de fixation. Pour la protéine comme pour le ligand, ce sont des groupements polaires qui interviennent dans ce type d'interaction et leur enthalpie de désolvatation est défavorable. Ces deux processus doivent être considérés pour connaître l'enthalpie globale de la réaction.

Les principaux avantages de la méthode microcalorimétrie isotherme de titration sont l'efficacité et la capacité à fournir tous les paramètres thermodynamiques d'une interaction. L'ITC est également devenue une des méthodes favorites pour déterminer la constante d'association d'une interaction ( $K_a$ ). C'est pourquoi elle a vu son utilisation augmentée de façon considérable depuis son introduction commerciale au début des années 90, dans les études d'interactions entre les biomolécules et leur ligand comme les interactions lectine-sucre. L'inconvénient majeur de cette technique est qu'elle nécessite une forte quantité de protéine et ligand, allant jusqu'à plusieurs milligrammes.

## 2.3 Les puces à sucres

Les puces à sucres sont dérivées de la technologie robotique standard. Elles offrent la possibilité de réaliser des analyses en parallèle sur un grand nombre d'oligosaccharides. Les puces à sucres permettent de déterminer la spécificité des protéines qui se lient à des oligosaccharides, tels que les lectines (Blixt, *et al.* 2004). Cette technique, où le grand intérêt est la miniaturisation et la rapidité de l'analyse, utilise peu de protéine.

Les puces à sucres ont été développées par le « Consortium for Functional Glycomics » (CFG ; <http://www.functionalglycomics.org>), fondé par le « National Institute of General Medical Sciences ». Le Consortium a mis au point des outils comme le « Glycan array » et le « Printed array », en français « array » signifie réseau ordonné. Ils se différencient principalement sur la méthode de fixation des sucres. Dans les deux cas, les lectines doivent

être soit marquées par un composé fluorescent comme l'Alexa Fluor® 488-TFP (carboxylic acid, 2,3,5,6 tetrafluorophenyl), soit fournies avec un anticorps.

### 2.3.1 Glycan array

Le « Glycan array » est basé sur le principe ELISA (*Enzyme-linked immunosorbent assay*) qui a été miniaturisé. Cette technique comprend une gamme d'oligosaccharides biotinylés fixés par un bras espaceur sur des plaques de 96 à 384 puits, recouvertes de streptavidine, Figure 18.

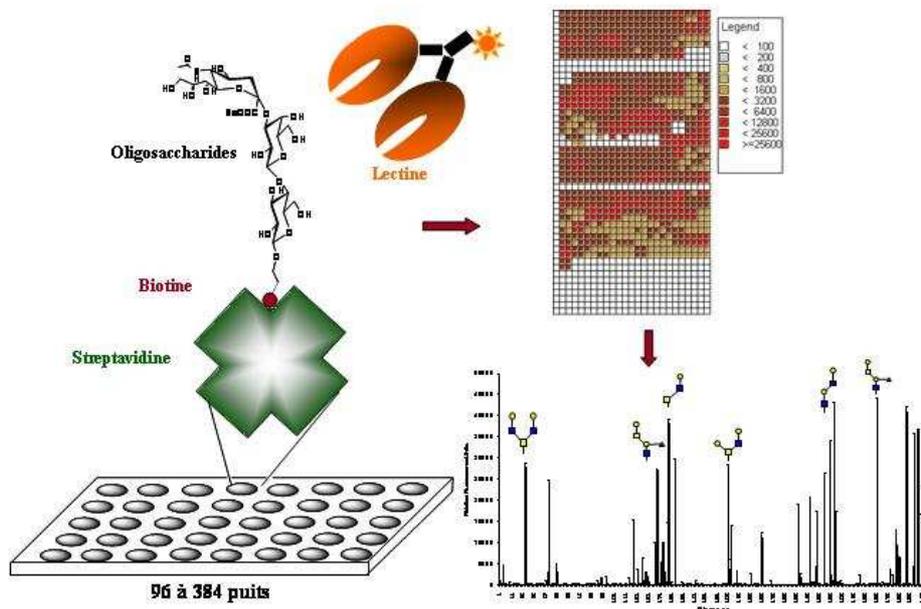


Figure 18 – Principe du Glycan array.

### 2.3.2 Printed array

Le « Printed array » utilise une bibliothèque d'oligosaccharides qui comporte aujourd'hui plus de 40 structures diverses présentes sur les parois cellulaires et représente les structures glycaniques importantes des glycoprotéines et des glycolipides. Des plaques de verre sont fonctionnalisées avec du N58 hydroxysuccimidine (NHS) capable de se lier de manière covalente avec des glycoconjugués préalablement couplés à un groupement amine (Blixt, *et al.* 2004) (Figure 19).

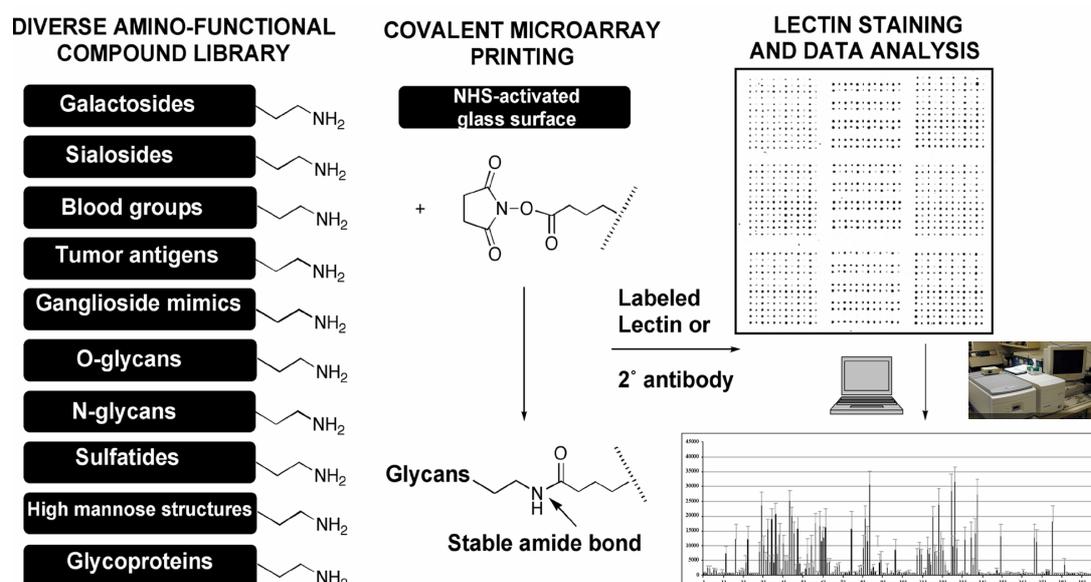


Figure 19 – Principe des Printed Arrays (figure issue de (Blixt, *et al.* 2004). La méthode permet de tester un très grand nombre d'oligosaccharides (dans notre cas 377) et exige seulement 0,1-2 µg de lectine pour obtenir un signal optimal.

## 2.4 DLS : Diffusion dynamique de la lumière

La diffusion de la lumière est une technique très puissante pour caractériser les espèces macromoléculaires. Elle permet la détermination de nombreuses informations sur les particules diffusantes, telles que: la présence ou non d'agrégats, la polydispersité, l'homogénéité, la taille des particules en solution, etc... C'est une technique très efficace, rapide, précise, simple à réaliser où l'échantillon peut être récupéré.

Les particules en solution présentent un mouvement aléatoire (mouvement Brownien) qui est lent pour les grosses particules et rapide pour les petites. La quantification de leur vitesse de diffusion par DLS permet ainsi d'en déduire leur rayon hydrodynamique (grandeur qui dépend notamment de leur masse et de leur forme).

La distribution de la taille est obtenue à partir d'une mesure de la vitesse de fluctuation de la lumière réfléchiée à un angle donné par les particules dans la solution (173° dans notre cas avec « Zetasizer nano series » Malvern instrument). Les résultats sont présentés par une courbe de l'intensité relative de la lumière diffusée en fonction du rayon hydrodynamique des particules. Ce rayon est directement proportionnel à la taille des particules en solution. Dans

le cas des protéines, il permet ainsi d'élucider leur structure quaternaire. La forme des molécules peut être estimée quand la masse moléculaire absolue est connue.

## 2.5 La cristallographie aux rayons X

La cristallographie aux rayons X permet de déterminer spatialement la position de tous les atomes d'un cristal qui peut-être d'une molécule inorganique tel un sel ou d'une macromolécule biologique comme une protéine. La Figure 20 montre les principales étapes du processus de détermination de la structure 3-D d'une protéine par cristallographie.

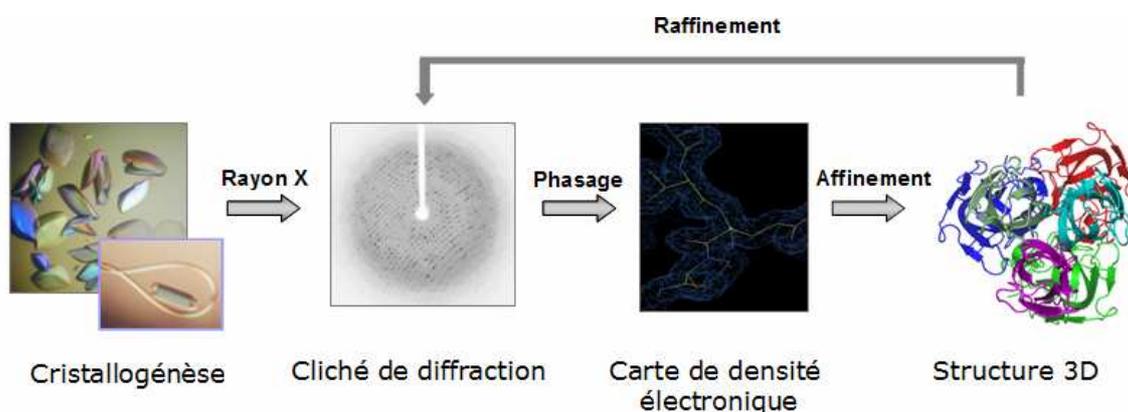


Figure 20 – Etapes pour l'obtention de la structure 3-D d'une protéine.

En 1934, les deux chercheurs Bernal et Crowfoot Hodgkin ont obtenu pour la première fois un cliché de diffraction à partir de cristaux d'une protéine : la pepsine. Ce fut les débuts de la cristallographie des protéines aux rayons X qui depuis est devenue une technique phare pour la détermination de la structure 3-D des protéines.

### 2.5.1 La cristallogénèse

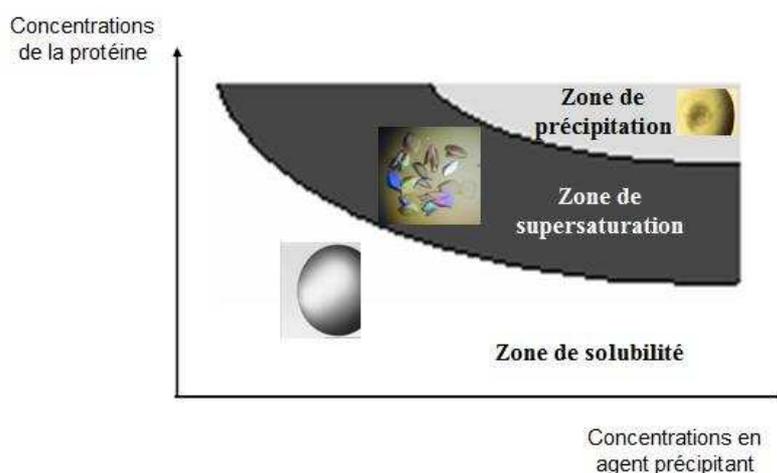
La cristallisation est un processus d'ordination, dans lequel un cristal est formé par la répétition dans l'espace d'un objet : le motif cristallin (par exemple, un atome, un ion ou une molécule) (Janin and Delepierre 1999). Ce phénomène peut se produire de façon naturelle et/ou expérimentale. Durant le processus de cristallisation, les molécules en solution passent d'un état liquide désordonné à un état solide ordonné où elles s'organisent en un réseau cristallin.

Pour avoir un bon cliché de diffraction, en règle générale les cristaux devront être réguliers avec des faces et formes bien définies, ce qui indique une haute pureté et perfection

concernant les molécules qui composent les cristaux (Ducruix and Giégé 1992). Dans beaucoup de cas, l'obtention d'un cristal de qualité est une barrière pour la résolution de la structure de certaines protéines (Geerlof, *et al.* 2006).

La quantité de paramètres engagés rend le processus de cristallisation d'une préparation protéique différent de celui des petites molécules. En effet, les protéines présentent une importante flexibilité de conformation et une grande versatilité. Elles sont des structures biologiques fragiles dont la stabilité dépend fortement de l'environnement immédiat. En conséquence, ces macromolécules sont plus sensibles aux conditions externes et une grande quantité des variables doit être contrôlée afin de rendre possible leur cristallisation. Parmi ces paramètres, nous pouvons citer : la pureté des molécules, la solubilité, la température, la concentration, le pH, la force ionique, l'effet de divers additifs, etc... (Bergfors 1999). Tous ces éléments rendent difficile la cristallogénèse qui reste une méthode empirique au succès aléatoire.

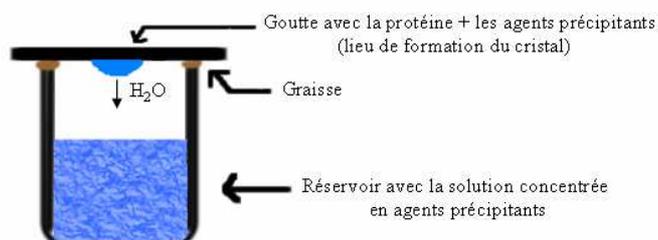
Le processus de cristallisation peut être divisé en trois étapes : nucléation, croissance et cessation de la croissance. Ces étapes sont des compromis entre les facteurs thermodynamiques (solubilité) et cinétiques (nucléation et croissance cristalline). Lors de la cristallisation, la solubilité de la macromolécule diminue jusqu'à atteindre l'état de supersaturation qui est favorable à la formation et à la croissance des cristaux, sans toutefois tomber dans la zone de précipitation, Figure 21.



**Figure 21 – Diagramme de phase d'une protéine.**

Il existe diverses techniques de cristallogénèse et celle utilisée dans notre laboratoire est celle de la « goutte suspendue » basée sur la diffusion de vapeur illustrée sur la Figure 22

(McPherson 1990). Une goutte formée du mélange de la protéine avec une solution contenant des agents précipitants est placée sur une lamelle de verre et scellée au dessus d'un réservoir contenant un grand volume d'agents précipitants. Le volume des gouttes est de 2 $\mu$ l en général lors de nos tests de cristallisation.



**Figure 22 – Système de cristallisation par diffusion en phase gazeuse par goutte pendante.**

L'équilibre s'établit entre la goutte et le réservoir par diffusion des espèces volatiles (eau ou solvants organiques) jusqu'à ce que la tension de vapeur soit la même dans la goutte et dans le réservoir. La goutte avec la protéine est plus diluée que le réservoir, ce qui entraîne une augmentation de la concentration dans la goutte par évaporation d'eau. Cet équilibre thermodynamique et cinétique engendre l'état de sur-saturation. Il s'établit progressivement en augmentant simultanément la concentration de la protéine qui alors s'agglomère spontanément. Si toutes les autres conditions sont favorables, par exemple le pH et la température, des cristaux de protéine se formeront alors dans la goutte.

La recherche des conditions de cristallisation s'effectue souvent par criblage de nombreuses solutions vendues en kits, tel que « Crystal Screens Hampton Research » (Jancarik and Kim 1991). De nos jours, les nombreux kits disponibles et la robotisation permettent de cribler très rapidement un grand nombre de conditions différentes avec peu de matériel, augmentant ainsi les chances d'obtention de cristaux.

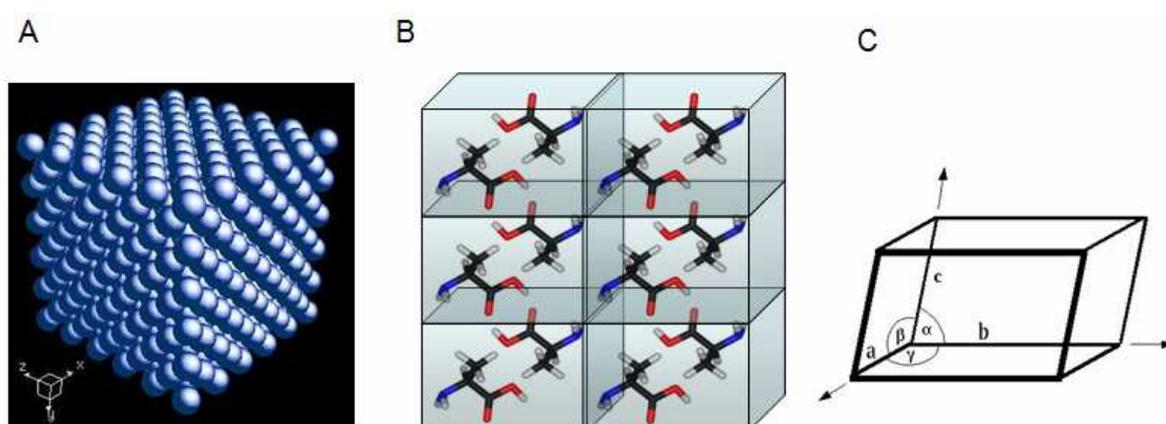
Lors de l'étude de complexe molécule-ligand, il existe deux techniques pour l'introduction des ligands dans les cristaux : le « trempage » et la « co-cristallisation ». La méthode de « trempage » consiste à baigner un cristal natif (protéine non complexée) dans une solution contenant le ligand. Celui-ci doit diffuser dans le cristal pour se fixer à la protéine, en espérant que le site de fixation soit accessible. La méthode de « co-cristallisation » consiste à incuber la protéine avec le ligand avant de réaliser les essais de cristallisations afin qu'ils puissent interagir en solution sans contraintes. Si la formation du complexe

entraîne des changements de conformation dans la protéine qui empêche l'obtention de cristaux dans les conditions natives, de nouvelles conditions de cristallisation pourront être criblées pour le complexe par co-cristallisation.

Avant de soumettre les cristaux de protéines aux rayons X lors de la collecte de données, ils sont congelés dans l'azote liquide afin de limiter les effets dus aux dommages des radiations. Le faisceau de rayons X engendre un échauffement local du cristal et provoque la formation de radicaux libres qui à leur tour endommagent les molécules du cristal. La propagation des radicaux libres dans le cristal est ralentie par sa congélation dans l'azote liquide ce qui prolonge sa durée de vie. Les cristaux sont plongés auparavant dans une solution cryoprotectrice (PEG, glycerol, huile minérale etc...) afin de prévenir la formation de cristaux de glace qui les détruiraient (Garman and Schneider 1997).

## 2.5.2 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 les trois directions de l'espace qui permet de générer un cristal entier est appelée maille élémentaire. Elle est définie par trois vecteurs  $\vec{a}$ ,  $\vec{b}$  et  $\vec{c}$  non colinéaires et trois angles notés  $\alpha, \beta, \gamma$ , Figure 23. Chaque maille élémentaire peut être remplacée par un ou plusieurs points dont l'environnement est égal dans tout le cristal créant ainsi un réseau cristallin auquel un motif structural est attaché (un ou plusieurs atomes).



**Figure 23 – A : Modèle d'un cristal à trois dimensions. B : motif cristallin. C : maille élémentaire et paramètres à trois dimensions.**

Il existe 14 types différents de maille élémentaires pour décrire l'arrangement de points en 3-D qui créent 14 réseaux appelés réseaux de Bravais, Figure 24. Des opérations de symétries (rotation au tour d'un axe, inversion et réflexion) peuvent être appliquées ensuite dans la maille. Elles appartiennent à un des 32 groupes symétriques qui combinés aux 14 réseaux de Bravais forment les 230 groupes d'espace qui représentent toutes les combinaisons d'opérations de symétrie applicables au motif cristallin.

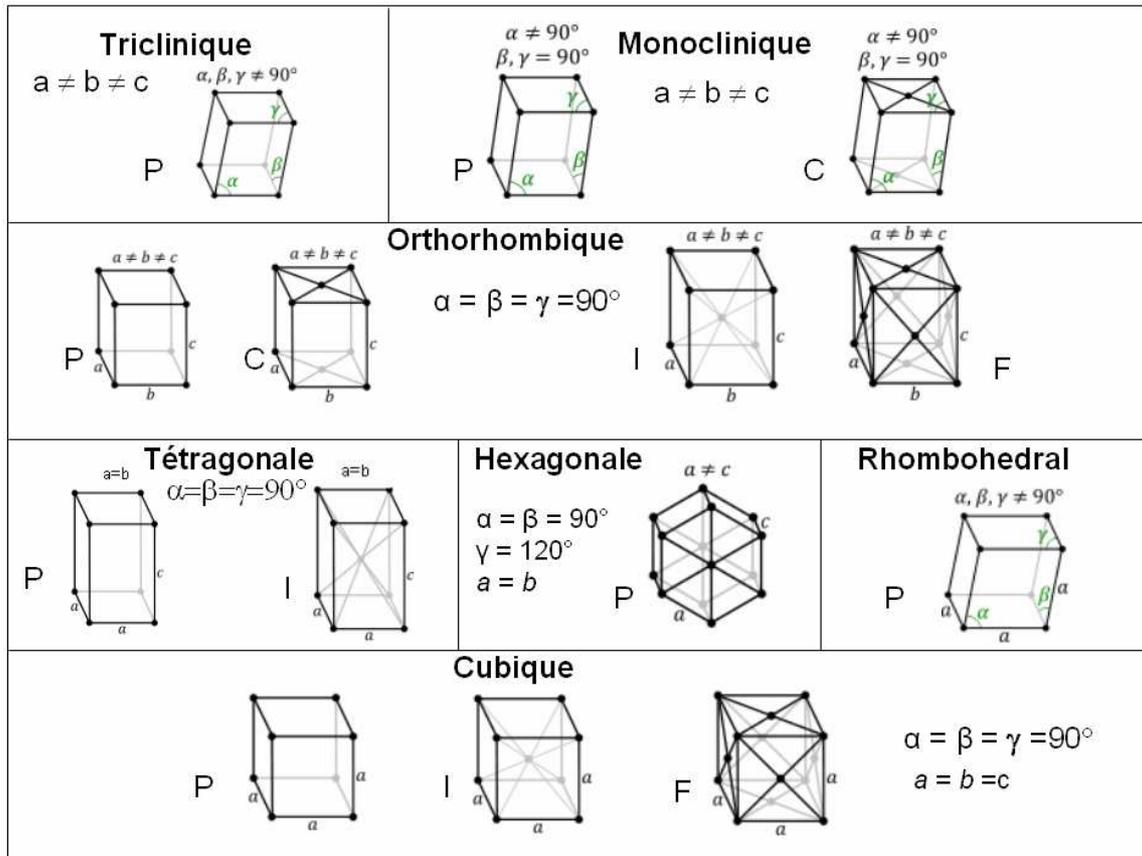


Figure 24 – Réseaux de Bravais.

Vu que les protéines sont des molécules chirales, les opérations de symétrie de type réflexion et inversion ne peuvent exister dans leurs cristaux ce qui réduit à 65 le nombre de groupes d'espace possibles en cristallographie des protéines. On définit ainsi l'unité asymétrique qui est le plus petit volume permettant de reconstituer la maille par symétrie (par exemple dans la Figure 23B l'unité asymétrique comprend une seule molécule d'alanine).

Lorsqu'un faisceau incident de rayons X étroit et parallèle est dirigé sur un cristal, il est dévié par les électrons groupés autour de chaque atome à l'intérieur de la maille du cristal ce qui conduit à l'émission de faisceaux diffractés dans toutes les directions de l'espace, Figure 25A. Les bases mathématiques qui caractérisent le processus de diffraction ont été formulées

principalement par Laue, Bragg et Ewald au début du XX<sup>ème</sup> siècle. Sir Lawrence Bragg a montré que la diffraction par un cristal pouvait être considérée comme la réflexion du faisceau incident par une série de plans parallèles (les plans réticulaires du cristal), comme le ferait une série de miroirs. Chaque ensemble de ces plans est identifié par les indices de Miller hkl. Les rayons X réfléchis par des plans adjacents parcourent des distances différentes et Bragg a montré qu'un faisceau primaire de longueur d'onde ( $\lambda$ ), frappant l'ensemble de plans parallèles d'indice hkl et de distance d'espacement  $d_{hkl}$  avec un angle ( $\theta$ ), sera réfléchi et repartira selon le même angle ( $\theta$ ), Figure 25 B, seulement si la condition géométrique suivante est rencontrée :

$$2d \sin \theta = n\lambda,$$

où (n) est un entier. Si les rayons X frappent les plans parallèles avec un angle qui ne satisfait pas cette équation, les rayons adjacents seront diffractés avec des phases différents donnant lieu à des interférences destructrices. Ainsi, seulement lorsque cette équation est satisfaite, des interférences constructives se produisent et s'additionnent pour être observées sur le cliché de diffraction sous forme de tache de diffraction ou réflexion. La position de chaque tache se rapporte à une famille de plans réticulaires et est définie par ses indices de Miller. Les clichés de diffraction sont une image du cristal et représentent le réseau réciproque, défini par les vecteurs  $a^*$ ,  $b^*$  et  $c^*$  perpendiculaires aux vecteurs  $b$  et  $c$ ,  $a$  et  $c$  et  $a$  et  $b$  respectivement et de longueur  $1/a$ ,  $1/b$  et  $1/c$ . Ils décrivent la position des ondes diffractées tandis que l'espace réel décrit la position des atomes dans le cristal.

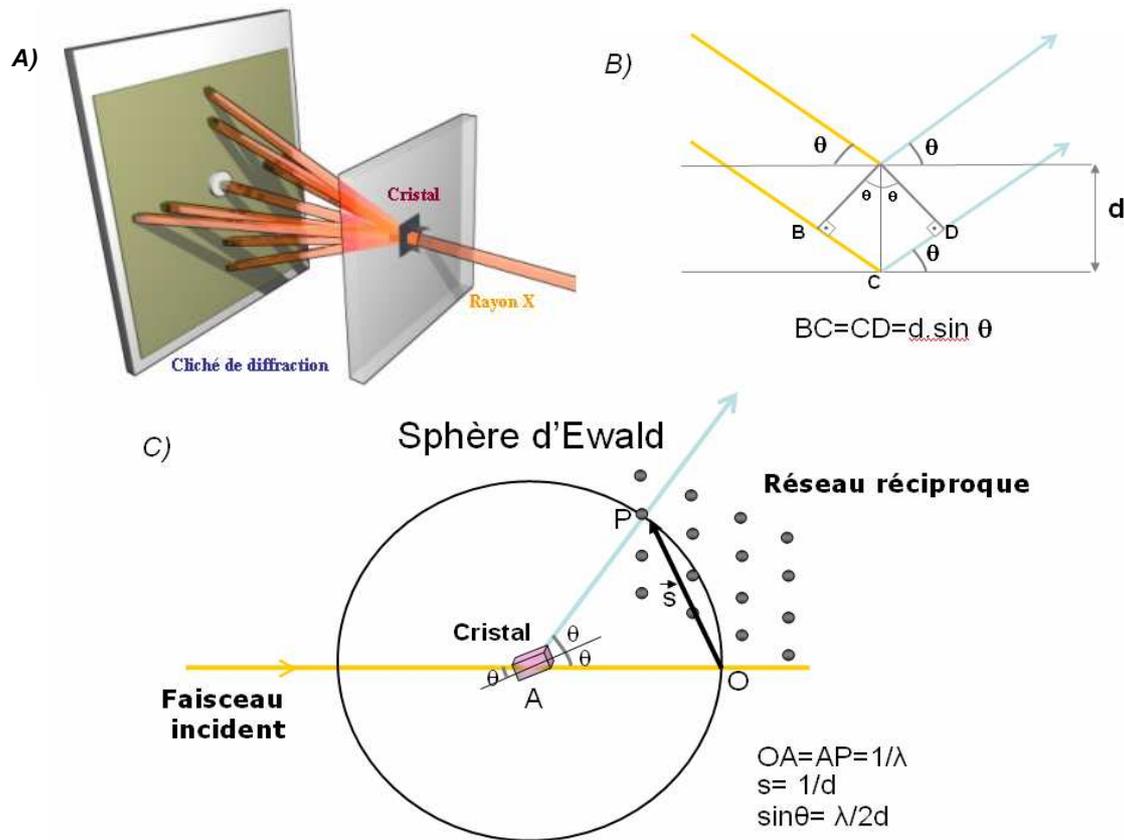


Figure 25 – A : Faisceau incident de rayons X sur un cristal ; B : Principe de la diffraction selon Bragg. C: Construction d'Ewald.

La construction géométrique d'Ewald permet de visualiser la diffraction dans l'espace réciproque plutôt que dans l'espace réel, Figure 25C. Le faisceau incident est placé sur le diamètre d'une sphère de rayon  $1/\lambda$ . Le réseau réciproque représente le cristal et son origine est placée à l'endroit où le faisceau incident émerge de la sphère. Lorsqu'un point du réseau réciproque rencontre la sphère d'Ewald, la loi de Bragg est respectée et il y a diffraction. Pour une orientation du cristal donnée, seuls quelques points du réseau réciproque seront en position de diffraction donc il faut faire tourner le cristal pour que tous les points traversent la sphère d'Ewald.

### 2.5.3 Indexation, intégration et mise à l'échelle du jeu de données

Sur une ligne de lumière d'un synchrotron de troisième génération tel que l'European Synchrotron Radiation Facility (ESRF) à Grenoble, les détecteurs électroniques de type CCD sont capables d'enregistrer un grand nombre d'images dans un temps d'acquisition de l'ordre de quelques minutes.

L'intégration des données des clichés de diffraction est réalisée en utilisant des logiciels informatiques comme Mosflm (Leslie 1992). Les images analysées permettent d'attribuer à chaque réflexion ses indices de Miller  $hkl$  et son intensité ( $I$ ) associée avec une erreur expérimentale ( $\sigma I$ ). Les intensités sont corrigées, indexées et moyennées entre les réflexions équivalentes par symétrie. Les amplitudes des facteurs de structure sont calculées à partir des intensités. L'ensemble des images intégrées est mis à l'échelle afin de constituer un jeu de données complet et homogène à l'aide de programmes tel que Scala (Collaborative Computational Project Number 4 1994). Afin d'évaluer la qualité d'un jeu de données expérimental, il est important d'observer des paramètres importants tels que: la complétude, la multiplicité, le Rmerge et le rapport entre l'intensité d'une réflexion et son erreur ( $\sigma I$ ). De par la présence d'opérations de symétrie dans le cristal, certaines réflexions sont reliées entre elle par symétrie et sont équivalentes. Elles devraient donc avoir la même intensité Et l'accord entre ces réflexions est donné par le Rmerge. Les informations nécessaires sur la distribution de la densité électronique dans le cristal seront contenues dans un nombre des réflexions appelées réflexions uniques. La complétude est le pourcentage de réflexions unique mesurées et doit avoisiner au moins les 95% dans un bon jeu de données. Le rapport entre les réflexions totales et les réflexions uniques donne la multiplicité que l'on cherche à maximiser pour avoir les données les plus précises possible. La résolution maximale d'un jeu de donnée est déterminée par la couche de résolution donnant un ( $I/\sigma I$ ) de 2 et ayant un Rmerge  $< 40\%$ .

#### 2.5.4 Facteur de structure et densité électronique

Chaque tache  $hkl$  du réseau réciproque représente une onde résultante de la somme des diffractions contributives de tous les électrons appartenant à la famille de plans réticulaires qu'elle représente. La fonction mathématique qui décrit cette onde est appelée le Facteur de structure  $F(\vec{s})$ , aussi noté de  $F_{hkl}$ . Cette fonction peut être calculée par la somme infinitésimale de contributions de densité électronique de la maille cristalline par la formule suivante:

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

La densité électronique  $\rho$  pour un point de coordonnées  $(x, y, z)$  dans le volume de la maille  $(V)$  est la transformée de Fourier inverse du facteur de structure  $F_{hkl}$ :

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l} F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

La densité électronique peut ainsi se représenter sous la forme d'une somme de termes exponentiels complexes, décrivant chacun une onde plane se propageant dans la direction d'un nœud du réseau réciproque. Chaque onde est caractérisée par son amplitude  $|F_{hkl}|$  qui est proportionnelle à la racine carrée de l'intensité mesurée et par sa phase  $\phi_{hkl}$ . Malheureusement, les phases sont perdues au cours de l'expérience ce qui représente problème majeur en cristallographie.

### 2.5.5 Problème des phases

Plusieurs méthodes ont été mises au point pour résoudre le problème de la perte des phases. Elles utilisent toutes un outil très important en cristallographie appelée la fonction de Patterson  $P(u,v,w)$ . Celle-ci est indépendante des phases et dépend seulement des intensités déterminées expérimentalement. C'est la transformée de Fourier des intensités.  $P(u,v,w)$  est large quand  $\rho(x,y,z)$  et  $\rho(x+u, y+v, z+w)$  sont tous les deux larges donc quand  $u,v,w$  représente un vecteur interatomique.

$$P(u, v, w) = \frac{1}{V} \sum_h \sum_k \sum_l [F_{hkl}]^2 e^{-2\pi i(hu+kv+lw)}$$

Une carte de Patterson est calculée où les pics correspondent à tous les vecteurs interatomiques qui peuvent être tracés entre les atomes de la maille réelle. Le nombre de pics est proportionnel au carré du nombre d'atomes  $N$ . L'intensité des pics est proportionnelle au produit du nombre atomique des deux atomes à chaque extrémité du vecteur.

Les phases nécessaires à la résolution d'une structure de protéines peuvent être obtenues de manière expérimentale par la méthode du remplacement isomorphe multiple « MIR multiple isomorphous replacement », par l'utilisation de la diffusion anormale par les méthodes SAD « Single anomalous diffraction » et MAD « Multiwavelength anomalous diffraction ». Ces techniques sont basées sur l'utilisation des variations d'intensités suite à l'introduction dans le cristal d'un ou plusieurs atomes dits lourds (MIR) ou diffuseurs anormaux (SAD/MAD). Le calcul des cartes de différence de Patterson permet de déduire la position des atomes

introduits dans la maille cristalline, de calculer leur contribution aux amplitudes et aux phases des rayons diffractés par les autres atomes et enfin de calculer une carte de densité électronique.

Si une protéine de structure inconnue possède un homologue de structure connue, la méthode du remplacement moléculaire est alors la méthode de choix pour résoudre le problème des phases (Rossmann and Blow 1962). Pour avoir une bonne chance de succès de phasage, la similarité de séquence entre les deux protéines doit être au moins de 20 %. Ces similarités de séquence impliquent le plus souvent une similarité structurale. Elle consiste à placer la molécule modèle dans la maille de structure inconnue de façon à reproduire au mieux les facteurs de structures observés et d'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. Celle-ci étant longue, il est plus rapide de faire deux recherches tridimensionnelles. Tout d'abord l'orientation de la molécule est déterminée par la fonction de rotation puis la position par la fonction de translation. De nos jours des programmes de recherche automatique de plus en plus puissants sont disponibles tels que AMoRe; (Navaza 1994) (Navaza, 1994), PHASER (McCoy 2007) qui permettent d'obtenir une solution en quelques minutes.

### **2.5.6 Affinement et vraisemblance du modèle**

Le premier jeu de phases obtenu par chaque technique contient toujours des erreurs donc la première carte de densité est bruitée, partiellement fautive et n'est pas toujours de qualité suffisante pour déterminer la position précise des atomes. Elle permet néanmoins la construction d'un modèle préliminaire permettant le calcul d'un nouvel ensemble de phases et l'obtention d'une nouvelle carte de densité améliorée. On cherche à éliminer la plupart des erreurs au cours de l'affinement du modèle qui a pour but de minimiser les différences entre le facteur de structure calculé à partir du modèle ( $F_c$ ) et le facteur de structure observé expérimentalement ( $F_o$ ). Le modèle est ajusté manuellement entre chaque cycle d'affinement suite à l'inspection des cartes de densités électroniques  $2F_o - F_c$  et  $F_o - F_c$  grâce à des interfaces graphiques telle que Coot (Emsley *et al.*, 2004). La carte  $2F_o - F_c$  donne la densité électronique pour le modèle observé et la carte  $F_o - F_c$  indique les différences de densité entre les cartes observées et calculées. Des logiciels d'affinement comme Refmac (Murshudov, *et al.* 1997) utilisent un algorithme de type statistique nommé l'estimation du maximum de

vraisemblance. Cet algorithme maximise la probabilité que le facteur de structure observé  $F_o$  soit en accord avec le modèle, tout en considérant de nombreux paramètres tels que les coordonnées et le type des atomes, le facteur d'agitation thermique  $B$ , mais aussi des paramètres structuraux comme les longueurs, les angles de liaison et les torsions. La correspondance du modèle aux données expérimentales est mesurée grâce au facteur cristallographique  $R$  qui indique l'accord entre les amplitudes des facteurs de structure calculées et observées et la progression de l'affinement. Il est calculé à partir de la formule suivante :

$$R = \frac{\sum_{hkl} \left| |F_o| - |F_c| \right|}{\sum_{hkl} |F_o|}$$

Un second critère de qualité : le  $R_{free}$  se calcule de la même façon que le facteur  $R$  conventionnel, mais à partir d'une fraction de réflexions (5% à 10% du jeu de données) non prise en compte durant l'affinement. Si la structure n'est pas suraffinée, le facteur  $R$  et le  $R_{free}$  doivent baisser en parallèle.

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

# Résultats



### 3 Résultats

Les résultats obtenus sur les discoidines (DiscI et DiscII) ont fait l'objet de deux publications qui forment l'essentiel de ce chapitre.

Saboia Aragão K., Satre M., Imberty A. & Varrot A. **Structure determination of Discoidin II from *Dictyostelium discoideum* and carbohydrate binding properties of the lectin domain.** *Proteins*. **73**, 43–52, 2008.

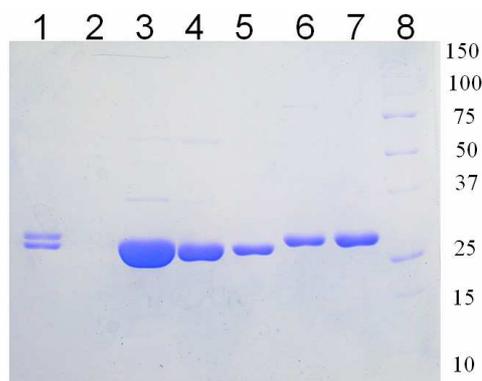
Saboia Aragão K., Imberty A. & Varrot A. **Discoidin I from *Dictyostelium discoideum* : structure and specificity comparison with discoidin II.** *To be submitted*.

Seul un résumé en français, accompagné de quelques figures complémentaires, est donné ci-dessous.

#### 3.1 Production de DiscI et DiscII.

L'amibe *Dictyostelium discoideum* à l'état de cellules cohésives produit des lectines spécifiques du galactose et du GalNAc : les discoidines (DiscI et DiscII). Tout d'abord, nos premières tentatives ont été de purifier les lectines à partir d'un extrait de la souche AX2 de l'amibe *Dictyostelium discoideum* fourni par le groupe de Michel Satre (Laboratoire de Biochimie et Biophysique des Systèmes Intégrés, iRTSV/BBSI, CEA-Grenoble, France). Les deux protéines ont pu être obtenues par chromatographie d'affinité sur colonne Sepharose-4B. Par contre, les tentatives de séparation de ces lectines par chromatographie sur résine d'échangeuse d'ions se sont révélées infructueuses. Le mélange de lectine a cependant permis d'obtenir des cristaux, mais de qualité limitée (faible reproductibilité et résolution limitée à 2.5 Å).

Nous avons modifié notre stratégie, en réalisant le clonage des gènes et l'expression sous forme recombinante chez *E.coli* des deux lectines appelées DiscIr et DiscIIr. Pour atteindre ce but, nous avons ajouté à la partie N-Terminale des protéines une étiquette hexahistidine clivable par la protéase du tabac et du virus (TEV). Les protéines sont bien exprimées dans la bactérie et sont relativement faciles à purifier par chromatographie d'affinité sur résine de Nickel en quantité importante (20 mg/L de culture). La pureté des lectines a été évaluée sur gel électrophorèse SDS-PAGE (Figure 26) Le profil électrophorétique des protéines correspond à un poids moléculaire d'environ 28 kDa.

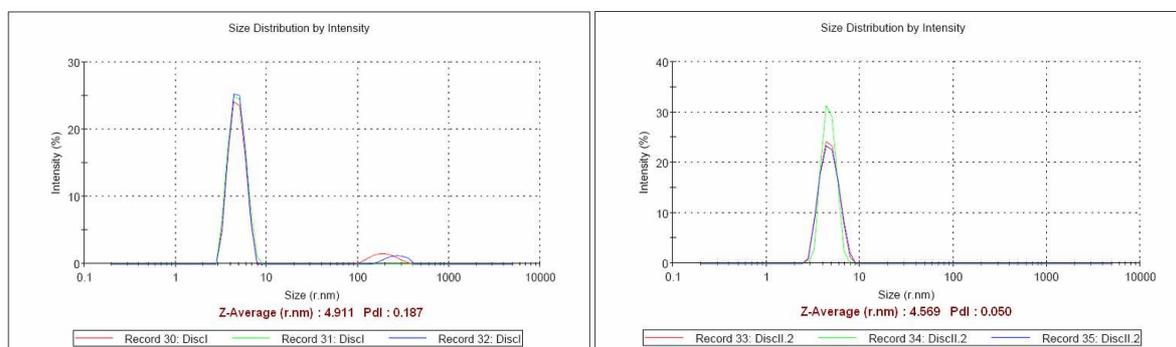


**Figure 26 – Analyse électrophorétique sur gel SDS-PAGE 12%, 1) Discoidin I et II sauvages à 5 mg/mL ; 2) vide ; 3) DiscIr avec Histag 6 mg/mL ; 4) DiscIr sans Histag après stockage pendant 4 mois 5,5 mg/mL ; 5) DiscIr sans Histag 5mg/mL ; 6) DiscIIr avec Histag 4,6 mg/mL ; 7) DiscIIr sans Histag 6mg/mL ; 8) Marqueur de poids moléculaire.**

### 3.2 Analyse par diffusion de la lumière

Par la technique de diffusion de la lumière (DLS), nous avons analysé la présence ou non d'agrégats, la monodispersité et la structure quaternaire de nos solutions de lectines. Toutes les mesures rapportées ici ont été effectuées sur un Zetasizer Nano S à 22°C en utilisant 50µL d'une solution pure de lectines.

La DiscI se retrouve sous forme monodisperse et sans agrégats avec une valeur d'indice de polydispersité égale à 0,187. Le pic majoritaire indique un rayon hydrodynamique de 4,9 nm pour la taille moyenne de particules (Figure 27).

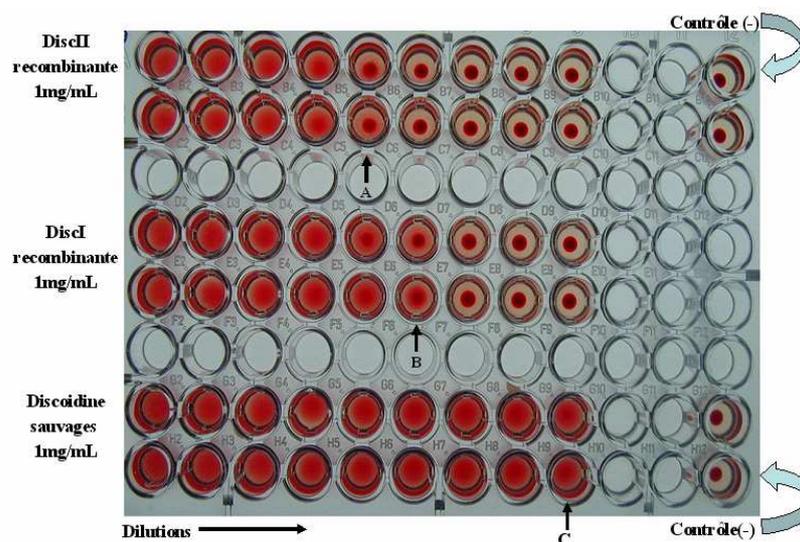


**Figure 27 – Courbe de l'intensité relative de la lumière diffuse en fonction du rayon hydrodynamique pour une solution de la DiscI (à gauche) et de la DiscII (à droite).**

Ces tests, réalisés avec les protéines stockées pendant 4 mois, montrent que ces protéines restent stables dans le temps. Un résultat similaire avec un rayon hydrodynamique moyen pour les particules de 4,6 nm et une valeur de polydispersité très faible (0,050) a été observé pour la DiscII. Les calculs de poids moléculaires indiquent une valeur d'environ 100 KDa ce qui indique que les Discoidines I et II seraient sous la forme de trimères en solution et non de tetramères comme cela avait été proposé lors des premières caractérisations dans les années 70s (Rosen, *et al.* 1973).

### 3.3 Interaction avec les ligands glucidiques

Le potentiel d'hémagglutination des lectines a été étudié en utilisant des érythrocytes natifs de lapin. Les lectines recombinantes ont une très bonne capacité d'hémagglutination, bien qu'elles ne soient pas aussi efficaces que le mélange de discoidines sauvages (Figure 28). Dans cette approche, pour une concentration de départ définie, l'unité d'hémagglutination (UH) est définie comme le facteur de dilution à partir de laquelle l'hémagglutination n'est plus observée. La DiscI présente un potentiel d'hémagglutination plus important que la DiscII (DiscI :  $64 \text{ UH.mL}^{-1}$ ; DiscII :  $32 \text{ UH.mL}^{-1}$ ; Discoidines sauvages :  $1024 \text{ UH.mL}^{-1}$ ). Les tests d'inhibition d'hémagglutination en présence d'un ligand monosaccharidique montrent que la DiscI a une affinité plus forte pour le GalNAc que pour le galactose. La DiscII présente presque la même affinité pour les deux sucres testés.



**Figure 28 – Comparaison de l'activité hémagglutinante des discoidines recombinantes et de discoidine sauvage à 1mg/mL (A) DiscIIr :  $32 \text{ UH.mL}^{-1}$  ; (B) DiscIr :  $64 \text{ UH.mL}^{-1}$  (C) Discoidines sauvages (mélange de DiscI et DiscII) :  $1024 \text{ UH.mL}^{-1}$ . Contrôle négatif (sang + tampon).**

La technique de puces à sucres et la microcalorimétrie isotherme de titration ont permis d'identifier la spécificité et l'affinité des deux lectines pour différents oligosaccharides. DiscI présente une préférence nette pour les motifs GalNac (ou Gal) substitués en position 3 par un monosaccharide d'anométrie  $\beta$  (Gal $\beta$ 1,3GalNac, GlcNac $\beta$ 1,3GalNac, GalNac $\beta$ 1,3Gal...). DiscII a 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). Les tableaux de spécificité des deux lectines sont donnés en annexe.

Un très petit nombre d'oligosaccharides sont reconnus par les deux lectines. Il s'agit soit des gangliosides de la série GD1, GD2... qui contiennent à la fois des acides sialiques et des GalNac, ou de structures très fucosylées (Figure 29).

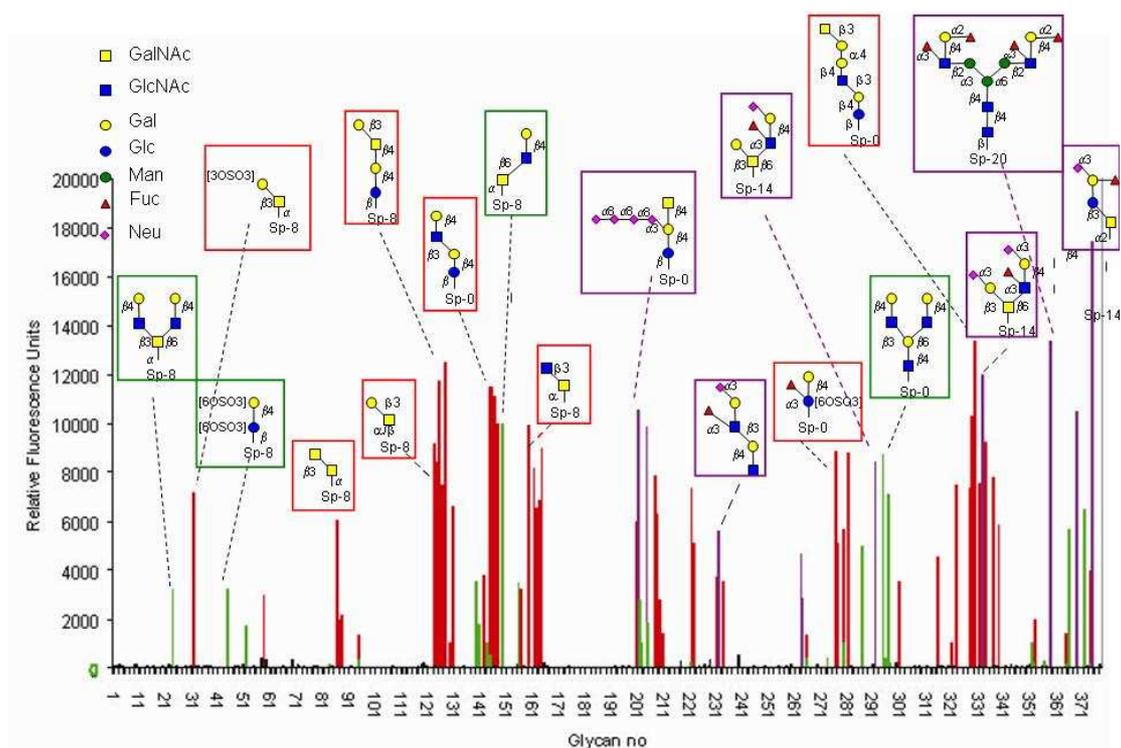


Figure 29 – Superposition des résultats de puces à sucres obtenus avec DiscI (rouge) et DiscII (vert) à la même concentration (0.1 µg/mL). Les pics communs au deux sont colorés en violet.

### 3.4 Études structurales de DiscI et DiscII

L'utilisation de protéines recombinantes a permis d'obtenir des cristaux de très bonne qualité (Figure 30). Les structures cristallines des lectines recombinantes DiscI et DiscII ont été résolues à 1.8Å et à 1.75Å respectivement. La protéine DiscII cristallise dans un groupe d'espace H32 avec un monomère dans l'unité asymétrique et un trimère est organisé autour de l'axe de symétrie d'ordre 3. La DiscI cristallise dans le groupe d'espace C222<sub>1</sub> et le trimère représente le contenu de l'unité asymétrique. La structure de la DiscII recombinante a été résolue par la méthode de phasage SAD (*single anomalous diffraction*) par le Dr. Annabelle Varrot. Les structures de la Disc II sauvage et de la DiscII recombinante complexée avec le GalNAc ou le  $\beta$ -methyl-galactose et de la DiscI recombinante sous forme native ou complexée au GalNAc ont été résolues par remplacement moléculaire à partir de cette première structure.



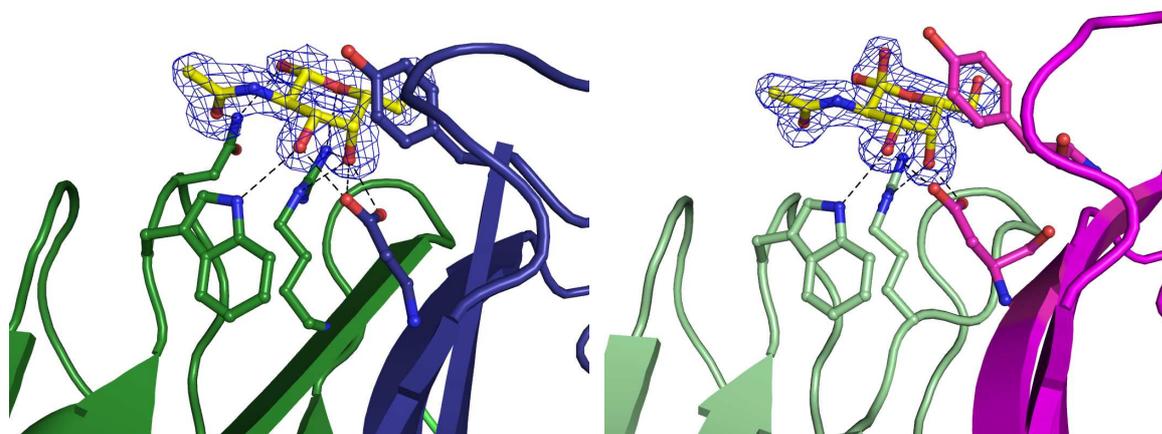
**Figure 30 – Cristaux de: A) discoidines sauvages (mélange de DiscI et DiscII) obtenus pour la condition expérimentale 20% PEG4K, 10mM CdCl<sub>2</sub>, 0.1M Hepes pH 7.5 ; B) DiscIIr obtenu par co-cristallisation avec 10 mM GalNAc, pour la condition expérimentale 25% PEG 550MME 0,1M HEPES pH7.5; et C) DiscIr, obtenus pour la condition expérimentale 1.5M sulfate de lithium 0.1M NaHepes pH 7.5.**

L'oligomérisation des discoidines résulte d'une association de trois monomères interagissant par l'intermédiaire de feuillets- $\beta$ . Ces protéines présentent deux domaines :

- Le domaine C-terminal, typique de lectine type H, adopte un repliement du type sandwich à deux feuillets de trois brins  $\beta$  antiparallèles qui s'agencent en *clé grecque* constituée de six brins antiparallèles. Les résidus du site de la liaison au sucre qui interagissent avec les hydroxyles 3 et 4 de ces sucres sont bien conservés.

- Le domaine N-terminal est composé de huit brins  $\beta$  agencés en deux feuillets  $\beta$  antiparallèles de trois et cinq brins. Ces brins interagissent les uns avec les autres pour constituer une forme globale de tonneau. Ce sandwich de feuillets- $\beta$  est relié par des boucles de surface où il est observé de molécules de PEG ou de MPD pour la DiscII et de MPD pour la DiscI. Ce site est situé à l'opposé du site de liaison au sucre observé sur le domaine C-terminal.

La comparaison des sites de reconnaissance permet de comprendre la plus haute affinité de DiscI pour le GalNAc (Figure 31). Le résidu Asn236 qui n'est présent que dans DiscI établit une liaison hydrogène avec le groupement carbonyle de la chaîne N-acétyl.



**Figure 31 – Comparaison des sites de reconnaissance du sucre dans les complexes cristallins DiscI-GalNAc (à gauche) et DiscII-GalNAc (à droite).**

Une étude complémentaire par modélisation moléculaire (Anne Imberty) a permis d'identifier la poche supplémentaire de la DiscI qui apporte la haute affinité pour les disaccharides de type  $\beta$ 1-3Gal/GalNAc.

# Structure determination of Discoidin II from *Dictyostelium discoideum* and carbohydrate binding properties of the lectin domain

Karoline Saboia Aragão,<sup>1</sup> Michel Satre,<sup>2</sup> Anne Imberty,<sup>1\*</sup> and Annabelle Varrot<sup>1\*</sup>

<sup>1</sup> CERMAV-CNRS (affiliated with University Joseph Fourier), Grenoble, France

<sup>2</sup> Laboratoire de Biochimie et Biophysique des Systèmes Intégrés, iRTSV/BBSI, UMR 5092 CNRS-CEA-UJF, CEA-Grenoble, Grenoble, France

## ABSTRACT

The social amoeba *Dictyostelium discoideum* adopts a cohesive stage upon starvation and then produces Discoidin I and II, two proteins able to bind galactose and *N*-acetyl-galactosamine. The *N*-terminal domain or discoidin domain (DS) is widely distributed in eukaryotes where it plays a role in extracellular matrix binding while the *C*-terminal domain displays sequence similarities to invertebrate lectins. We present the first X-ray structures of the wild-type and recombinant Discoidin II in unliganded state and in complex with monosaccharides. The protein forms a homotrimer which presents two binding surfaces situated on the opposite boundaries of the structure. The binding sites of the *N*-terminal domain contain PEG molecules that could mimics binding of natural ligand. The *C*-terminal lectin domain interactions with *N*-acetyl-D-galactosamine and methyl-β-galactoside are described. The carbohydrate binding sites are located at the interface between monomers. Specificity for galacto configuration can be rationalized since the axial O4 hydroxyl group is involved in several hydrogen bonds with protein side chains. Titration microcalorimetry allowed characterization of affinity and demonstrated the enthalpy-driven character of the interaction. Those results highlight the structural differentiation of the DS domain involved in many cell-adhesion processes from the lectin activity of *Dictyostelium discoidins*.

Proteins 2008; 73:43–52.  
© 2008 Wiley-Liss, Inc.

**Key words:** galactose binding domain; *H*-type lectin; discoidin domain.

## INTRODUCTION

The haploid social soil amoeba *Dictyostelium discoideum* is a simple eukaryotic model, often employed to study signal transduction, chemotaxis, and multicellular development as well as an host model for pathogenesis.<sup>1–4</sup> The life cycle of the slime mold consists of two distinct phases: a unicellular vegetative state where the amoeba grows by mitotic division and feeds by phagocytosis on bacteria and a multicellular state initiated upon starvation where up to 10<sup>5</sup> cells aggregate by chemotaxis in response to AMPc signal before undergoing morphogenesis and cell differentiation to form a fruiting body.<sup>1</sup> Development ends up with the formation of spores that can survive unfavorable environmental conditions.<sup>5</sup>

As the cells become cohesive, two lectins discoidin I and II (DiscI and DiscII) are produced and found at the cell surface.<sup>6,7</sup> DiscII produced in late aggregation, is localized in prespore vesicles and seems to be externalized with the glycoconjugates that form the spore coat.<sup>6,8,9</sup> DiscII have been less studied and its role in *Dictyostelium* life cycle is still unclear. This protein of 257 residues is composed of two domains separated by a small linker. The *N*-terminal domain (1–155) corresponds to the classically named discoidin (DS) domain that is found in numerous unrelated proteins involved in aggregation, cell-surface signaling, and developmental processes and is known to have a β-sandwich structure.<sup>10,11</sup> The *C*-terminal domain (163–257) is responsible for DiscII hemagglutination activity inhibited by *N*-acetyl-D-galactosamine (GalNAc)<sup>6</sup> and could be involved in binding to the glycoconjugates at the cell surface.<sup>8</sup> Recently, sequence similarity was detected with invertebrate lectins from coral<sup>12</sup> and snail.<sup>13</sup> For the *Helix pomatia* lectin, a crystal structure is available demonstrating a trimeric arrangement of pair of β-sandwiches.<sup>13</sup> For DiscI, which displays 48% sequence identity to DiscII, it was shown that the aggregation site was different from the sugar binding

**Abbreviations:** βMeGal, methyl-β-D-galactoside; DDR, discoidin domain receptor; DS, discoidin domain; EDO, ethylene glycol; GalNAc, *N*-acetyl-D-galactosamine; MPD, 4-methyl-2-pentane diol; PEG, polyethylene glycol; TEV, tobacco etch virus.

Grant sponsor: CNPq grant from Brazilian government.

\*Correspondence to: Anne Imberty and Annabelle Varrot, CERMAV-CNRS (affiliated with University Joseph Fourier), BP53, F-38041 Grenoble cedex 09, France. E-mail: imberty@cermav.cnrs.fr or varrot@cermav.cnrs.fr

Received 28 November 2007; Revised 31 January 2008; Accepted 11 February 2008  
Published online 2 April 2008 in Wiley InterScience (www.interscience.wiley.com).  
DOI: 10.1002/prot.22038

site and was located on the DS domain.<sup>14,15</sup> Preliminary structural investigation proposed a tetrameric arrangement for discoidin based on density gradient centrifugation and sedimentation equilibrium experiments.<sup>16</sup>

We report here the 3-D structure resolution of the wild-type and recombinant DiscII in unliganded state and in complex with GalNAc and methyl- $\beta$ -D-galactoside ( $\beta$ MeGal). The quaternary structure of the protein is trimeric and the presence of two different binding surfaces, one on each domain on the opposite faces of the structure, was confirmed. The wild-type structure revealed a possible regulation of DiscII through histidine phosphorylation. The specific structural features of this two domains protein appear to be related to its role in morphogenesis and social life of amoeba.

## METHODS

### Purification of the unliganded protein

Wild-type DiscII was isolated from axenically grown *D. discoideum* AX2 cells collected in their late exponential growth phase. About  $2 \times 10^{10}$  cells were resuspended in 80 mL of the following buffer: 20 mM Tris pH 7.5, 10 mM CaCl<sub>2</sub>, 1 mM KCl, and 150 mM NaCl before sonication. After centrifugation for 20 min at 30,000g, the supernatant was loaded on a 30 mL column of Sepharose 4B (GE healthcare). The lectin was eluted by using 0.3M of D-galactose and concentrated to 5 mg/ml using Microsep<sup>TM</sup> 10 K (Pall).

### Cloning, expression and purification

cDNA for the discII gene was synthesized from *D. discoideum* total mRNAs using rapid amplification of cDNA ends (3'-RACE kit; Invitrogen) with the following primers: Dis2NcoI (5'-GTGCCATGGCCGTTCCAGCTGGTTC-3'), and Dis2XhoI (5'-GACTCGAGTTAATTATAAATTCAACAGCAAC-3'). A fragment of 771 bp was then amplified by PCR using the Taq polymerase (Promega), the same primers and the cDNA as template. After purification, the PCR fragment was cloned in-frame with an N-terminal hexahistidine tag and the tobacco-etch virus (TEV) protease cleavage site of the prokaryotic expression vector pProExHTb (Invitrogen). Sequence analysis was in accordance with the one deposited.

Rosetta2 (DE3) pLysS bacterial cells (Novagen) were used for the expression of the protein. Cells were cultured in LB media at 37°C to an OD<sub>600</sub> of 1.2 before overnight induction at 16°C with 0.2 mM isopropyl  $\beta$ -D-thiogalactoside. Cells were harvested by centrifugation at 4000g for 10 min before resuspension in three volumes of binding buffer (20 mM Tris pH 8.5, 500 mM NaCl, and 10 mM imidazole) and lysis by sonication. After centrifugation at 40,000g for 30 min, the supernatant was applied to a column made of 5 mL of Chelating

Sepharose<sup>TM</sup> Fast Flow resin (GE healthcare) charged with nickel and preequilibrated with the binding buffer. The column was washed in sequence with five volumes of binding buffer, binding buffer supplemented to 1.5M NaCl and binding buffer. Pure protein was eluted with a gradient 10–500 mM imidazole in binding buffer. Fractions were analyzed by SDS-PAGE 15% and then pooled. The Histag was cut by addition of the TEV protease at the ratio 1/100 overnight at 20°C after addition of 2 mM DTT.<sup>17</sup> The buffer was exchanged to 20 mM Tris pH 7.5, 150 mM NaCl by desalting on a PD10 column (GE healthcare) before reloading on the affinity column. The flow-through containing the protein without tag was collected and concentrated to 4.5 mg/mL on Microsep<sup>TM</sup> 10 K (Pall).

### Isothermal titration microcalorimetry measurements

Isothermal Titration Microcalorimetry (ITC) experiments were performed using a VP-ITC isothermal titration calorimeter (Microcal) at 25°C. GalNAc and  $\beta$ MeGal monosaccharides (18 mM) and rDiscII (0.15 mM) solutions were prepared with a 0.02M Tris. HCl, 0.15M NaCl, pH 7.5 buffer solution. Thirty injections of 10  $\mu$ L of sugar solution at 18 mM were added at intervals of 5 min while stirring at 310 rpm in the DiscII containing cell (1.455 mL) at 25°C. Control experiments in which the sugar solution is added to the microcalorimeter cell in the absence of protein yielded insignificant heats of dilution. The experimental data were fitted to a theoretical titration curve using software supplied by Microcal with  $\Delta H$  (enthalpy change),  $K_a$  (association constant), and  $n$  (number of binding sites per monomer) as adjustable parameters.<sup>18</sup> Free energy values and entropy contributions were determined from the standard equation:

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S$$

where  $\Delta G$ ,  $\Delta H$ , and  $\Delta S$  are the change in free energy, enthalpy, and entropy of binding, respectively.  $T$  is the absolute temperature,  $R$  the gas constant, and  $K_a$  is the association constant.

### Crystallization and data collection

All crystals were obtained by the hanging-drop vapor diffusion method using 2  $\mu$ L drops containing 50:50 (v/v) mix of protein and reservoir solution at 20°C. Crystal Screen I and II kits (Hampton Research, Aliso Viejo, CA) and Clear Strategy Screen I and II kits (Molecular Dimensions, Suffolk, UK) were used for screening of crystallization conditions. Crystals of the unliganded protein were obtained using 20% PEG 4K, 10 mM CdCl<sub>2</sub>, 100 mM Hepes pH 7.5. Thin plates appeared after a few days. Preceding data collection, the crystals were transferred in a cryoprotecting were 20% glycerol was added

**Table I**  
Data Collection and Refinement Statistics

Data set	wtDiscII	rDiscII	Hg	$\beta$ MeGal	GalNAc
Spacegroup	C2	H32	H32	H32	H32
a,b,c (Å)	287.9, 56.3, 123.7	81.8, 81.8, 257.8	82.0, 82.0, 259.6	81.1, 81.1, 259.3	82.5, 82.5, 261.6
$\alpha,\beta,\gamma$ (deg)	90.0, 109.1, 90.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution (outer shell), Å	26.8–2.45 (2.58–2.45)	37–1.75 (1.84–1.75)	29.3–1.4 (1.48–1.4)	36.7–1.9 (2.01–1.9)	48.2–1.9 (2.0–1.9)
Meas/Unique reflections	210, 204/66,709	371,819/34,036	2,092,862/66,656	116,327/26,194	166,140/27,554
Multiplicity	3.3 (3.1)	10.9 (11)	31.4 (21.1)	4.4 (4.1)	6.0 (6.1)
Completeness, %	96.8 (91)	100 (100)	100 (100)	99.7 (98.6)	100 (100)
$R_{\text{merge}}$	0.092 (0.372)	0.085 (0.353)	0.084 (0.533)	0.063 (0.312)	0.101 (0.339)
$I/\sigma I$	5.8 (1.9)	7.5 (2.1)	5.6 (1.4)	8.7 (2.3)	6.5 (2.1)
$R_{\text{cryst}}/R_{\text{free}}$	18.1/23.8	15/17.8		17.8/22.7	15.6/19.8
Rmsd bonds, Å/angles, °	0.014/1.55	0.014/1.49		0.014/1.49	0.016/1.41
Protein atoms/Bfac, Å <sup>2</sup>	12,115/26.6	2027/10.2		1996/25.4	2028/14.5
Ligand atoms/Bfac, Å <sup>2</sup>				13/27.6	17/21.5
Water molecules/ Bfac, Å <sup>2</sup>	543/27.9	370/22.9		228/31.8	359/26.2
PDBcode	2VME	2VM9		2VMD	2VMC

prior freezing. Native data were collected at 2.45 Å in C2 spacegroup at ESRF ID23-1.

Crystals of the unliganded recombinant protein were obtained after 2 days using 50% v/v MPD, 200 mM ammonium phosphate monobasic, 100 mM Tris pH 8.5. Derivative crystals were obtained by soaking of the native crystals with 5 mM thimerosal during 5 min prior freezing. Native and derivative data were collected at 1.75 and 1.4 Å in the spacegroup H32 at ESRF ID14-2. Crystals with ligand were obtained by cocrystallization after incubation of the protein at 5 mg/mL with 10 mM of ligand for at least an hour. GalNAc and  $\beta$ MeGal complexes were crystallized using 25% PEG 550MME, 100 mM Hepes pH 7.5, and 25% PEG 550MME, respectively, and data were collected at ESRF ID14-1 and ID29, respectively. All data were obtained from a single crystal directly mounted in a rayon fiber loop and placed in a boiling nitrogen stream at 100 K and collected using an ADSC detector.

Native and derivative data were processed and reduced with MOSFLM and SCALA<sup>19</sup> while XDS was used for processing the complexes data.<sup>20</sup> All further computing was performed using the CCP4 suite,<sup>21</sup> unless otherwise stated. All data statistics are summarized in Table 1.

### Structure determination and refinement

Attempts to use molecular replacement strategy with available structures presenting sequence similarity with lectin and DS domains were unsuccessful. Initial phases and a first model were obtained for the recombinant protein by the SAD method using the derivative data and autoSHARP.<sup>22</sup> 1 Mercury site was found and refined to give an initial figure of merit of 0.325. Solvent flattening with an estimate of 50.5% solvent resulted in excellent and easily interpretable maps with an overall figure of merit of 0.842.

The rDiscII structure and the two complex structures were determined by molecular replacement using AMoRe<sup>23</sup> and respectively the derivative and native coordinates as starting model. The wild-type structure was successfully solved by molecular replacement using PHASER after a search of two copies of the C-terminal domain trimer and six copies of the N-terminal domain.<sup>24</sup> The structures were refined using REFMAC<sup>25</sup> iterated with manual rebuilding in COOT.<sup>26</sup> Details of the models quality are given in the Table I.

All figures were drawn with PyMOL Molecular Graphics System (2002) DeLano Scientific, Palo Alto, CA, unless otherwise stated.

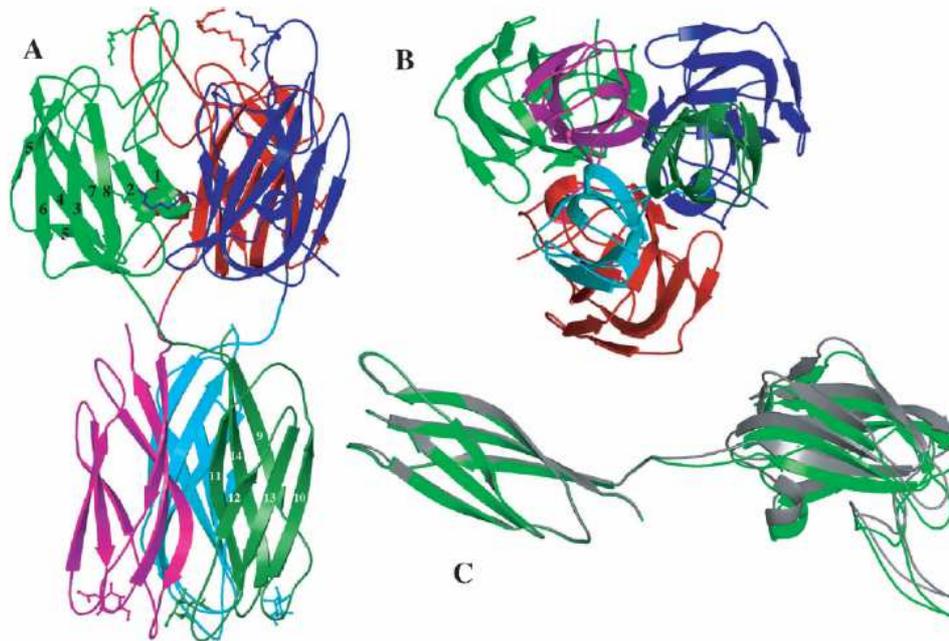
## RESULTS AND DISCUSSION

### DiscII cloning, purification and crystallization

Direct purification from *D. discoideum* AX2 cells on affinity column led to a mix of the two lectins which could not be further purified. Nevertheless, crystallization experiments yielded to plate crystals that diffracted to 2.4 Å in C2 spacegroup. Phasing attempts were unsuccessful since the crystals were fragile, low symmetry, and subjects to rapid radiation damage. The ratio between both lectins as well as the crystallization success was batch dependent. We decided therefore to clone the discII gene for expression in *Escherichia coli*. The recombinant protein was expressed in high yield (20 mg/L of culture) and easily purified thanks to a cleavable N-terminal Histag. Reproducible hexagonal crystals diffracting to high resolution were then obtained.

### DiscII overall structure

Molecular replacement attempts with models of both domains failed despite sequence identity and similarity of

**Figure 1**

Cartoon representation of DiscII trimer perpendicular (A) and along the threefold axis (B). Overlay of the wild-type (grey) and recombinant monomers (green) (C).

around 20 and 40%, respectively. The solutions obtained led only to poor phases and noninterpretable maps. This failure could be attributed to the difficulty to identify clear  $\beta$ -sandwiches in closely packed trimer. Also steric clashes would prevent the trimerization of the DS domain (PDB 1DP7, 1CZT, or 1KEX).

The recombinant DiscII (rDiscII) crystallizes in the H32 spacegroup with one monomer in the asymmetric unit and the trimer is formed around the threefold axis (Fig. 1 and Table I). The structure was solved by SAD using a mercury derivative and residues 6–257 could be modeled in the electron density revealing disorder at the N-terminus. Molecules of precipitant MPD and phosphate for the native or part of polyethyleneglycol (PEG) for the complexes were also found in the electron density. Several large positive peaks were modeled as calcium or chloride based on distance and coordination.

The wild-type structure (wtDiscII) was obtained by molecular replacement using the recombinant coordinates with two trimers in the asymmetric unit. All residues (2–257) are clearly visible in the electron density. The N-terminal methionine has been cleaved and the N-terminus was found formylated. Several molecules of cryoprotectant and precipitant (glycerol, PEG) as well as several ions (calcium, zinc, and chloride) could be modeled in density. Next to His84, we found a large peak of positive density which could be explained by phosphorylation of the histidine [Fig. 3(A)]. This modification is

often found in *Dictyostelium* where histidine kinases have some role in signal transduction and morphogenesis.<sup>27,28</sup> Analysis of the rDiscII electron density indicate that the recombinant protein is not phosphorylated.

The overall structure is similar for both the wild-type and recombinant protein [Fig. 1(A)]. The N-terminal domain is a  $\beta$ -sandwich composed of two antiparallel  $\beta$ -sheets of five ( $\beta_1$  17–20,  $\beta_2$  63–66,  $\beta_3$  71–79,  $\beta_4$  87–96,  $\beta_5$  102–104,  $\beta_5'$  111–114) and three ( $\beta_6$  121–125,  $\beta_7$  129–142,  $\beta_8$  146–153)  $\beta$ -strands packed against each other and a five-turn. The fifth strand is broken by a bend. This  $\beta$ -sandwich core is connected by three short surface loops on the termini face whilst prominent loops protrude like spikes on the opposite side of the barrel and form a ligand binding surface. This topology belongs to the discoidin domain subfamily of the galactose binding domain like superfamily in the SCOP classification.<sup>29</sup> The C-terminal domain presents the six-stranded antiparallel  $\beta$ -sandwich fold ( $\beta_9$  163–170,  $\beta_{10}$  183–190,  $\beta_{11}$  200–210,  $\beta_{12}$  220–224,  $\beta_{13}$  230–237,  $\beta_{14}$  243–254) first observed in the structure of the *Helix pomatia* agglutinin (HPA) which belongs to the H-type lectin family and resembles the immunoglobulin-like fold.<sup>13</sup>

### Oligomerization

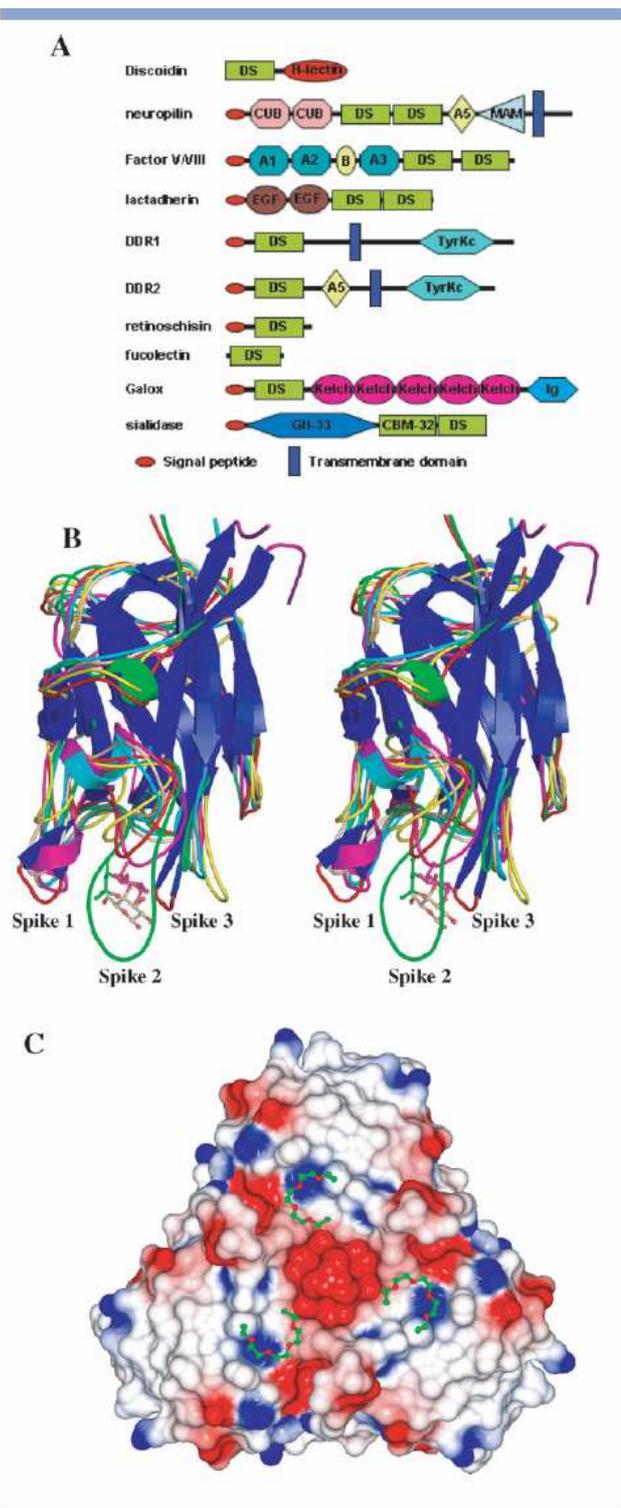
The trimeric association of the protein was confirmed in solution by dynamic light scattering (data not shown).

The trimer is generated by the intertwining of the three monomers with the C-terminal domain of the one molecule facing the N-terminal domain of the next molecule [Fig. 1(A)], which explains the stability of the protein in solution. Both domains are linked by an eight residues linker enclosed by two proline residues Pro155 and Pro162. The peptide angles between those two proline residues define the linker conformation and the orientation of one domain in respect to the other. If we superposed the C-terminal domain between the A molecules of the wild-type and recombinant structures, a rigid movement of up to 4.7 Å is observed for the main chain of the N-terminal domain demonstrating non-negligible flexibility, [Fig. 1(C)]. Moreover, the wild-type protein does not form a perfect trimer and the C-terminal domain overlay of the three monomers in the wtDiscII structure shows similar rigid body movements which extend up to 7.4 Å for the C $\alpha$  of Ser44. This movement permits the optimization of crystal contacts and prevents steric clashes with the lectin domain of the second trimer in the asymmetric unit. The flexibility between the two domains is therefore clearly displayed in wtDiscII structure (C2 spacegroup) but not in the rDiscII structure where the trimer is generated through a crystallographic threefold axis (H32 spacegroup).

### The N-terminal domain, an aggregation module

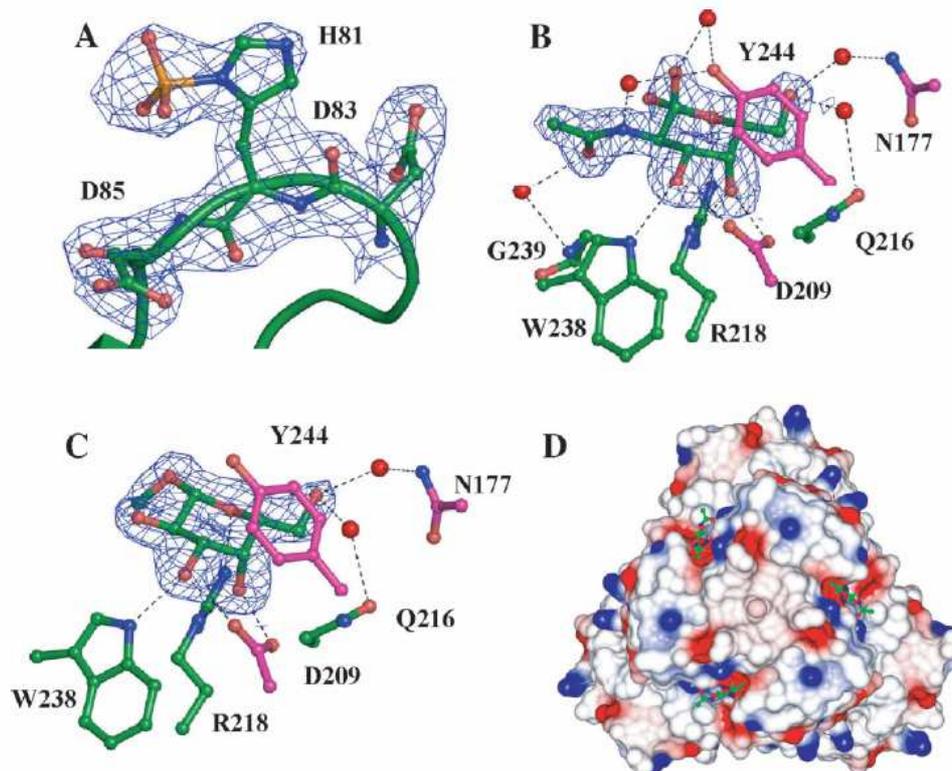
The 155 amino acids long N-terminal domain of discoidins has been identified since the 70s in many proteins both from eukaryotes or prokaryotes and is referred as the DS domain.<sup>6</sup> Despite unrelated functions, all the proteins sharing this domain are involved in cell adhesion and developmental processes [Fig. 2(A)]. There are only five structures of DS domain known to date: the C2 domain of coagulation factor V and VIII and lactadherin, the b1 domain of neuropilin and the DS domain of the discoidin domain receptor 2 (DDR2-DS) where the sequence identities with DiscII are 23, 21, 26, 20, and 19%, respectively.<sup>31–35</sup> DiscII-DS domain present a rmsd of on average 1.5 Å for 136 equivalent C $\alpha$  for the C2 and b1 domains and a rmsd of 2.1 Å for 132 equivalent C $\alpha$  with DDR2-DS structure (calculated with MSDfold).<sup>36</sup> Structural similarities between coagulation factors and DDRs proteins have been previously described.<sup>10,11,37</sup>

Structural similarity search revealed also several hits of proteins belonging to the galactose binding domain like superfamily such as galactose oxidase, sialidase, APC10/Doc1 subunit of the human anaphase promoting complex, *Anguilla anguilla* agglutinin (AAA) and several carbohydrate binding modules.<sup>38–42</sup> The rmsd ranges from 1.84 to 2.5 Å for 120 to 124 aligned C $\alpha$  with Secondary Structure Matching.<sup>36</sup> To our knowledge, only the fucosyltransferase from eel which presents 15% sequence identity of



**Figure 2**

Schematic domain structure of protein containing the DS domain (A). Cartoon representation of the overlay of the DS domain in divergent stereo (B). Loops have been coloured according to protein: rDiscII (green), factor V (red), DDR2 (yellow), galactose oxidase (cyan), sialidase (wheat) and fucoselectin AAA (magenta). Existing ligands are represented in ball and sticks. (C) Electrostatic surface representation of the DS-trimer of the GalNAc-rDiscII structure, PEG molecules are represented in ball and sticks, figure drawn with CCP4MG.<sup>30</sup>

**Figure 3**

Representation of the maximum-likelihood weighted 2mFo-DFc electron density, for the phosphohistidine in wtDiscII (A) and ligand bound in GalNAc-rDiscII (B) and betaMeGal-rDiscII (C) structures. The maps are contoured at 1  $\sigma$  representing 0.27, 0.37, 0.29 eÅ<sup>-3</sup>, respectively. (D) Electrostatic surface representation of the trimer of the lectin domain in the GalNAc-rDiscII structure, ligand is represented in ball and sticks, figure drawn with CCP4MG.<sup>30</sup>

DiscII-DS domain adopts a trimeric organization.<sup>38</sup> The trimeric arrangement of DDR2-DS and AAA are similar with a global rms of 3.26 Å for 356 matched C $\alpha$  but numerous differences in the length and position of  $\beta$ -strands are noticed.

The core of the  $\beta$ -sandwich is well conserved and the main differences are observed on the finger-like loops clustered on the side of the  $\beta$ -barrel opposite to the two termini. Those loops form spikes and a ligand binding surface which is involved in protein-carbohydrate, protein-lipid, or protein-protein interaction and whose topology depends on the spikes length, sequence, and conformation [Fig. 2(B)]. This domain seems to be characterized by the conservation of the overall fold for stability and diversity within the loops to ensure ligand specificity.<sup>10</sup> DiscI-DS domain has been proposed to bind to a glycoprotein and similar ligand could be expected for DiscII but remains to be identified.<sup>15</sup>

A search for possible ligand binding site in the DS trimer was performed by visual analyses of the surface and search for bound organic molecules coming from the crystallizing or cryoprotecting solution. Two putative binding sites could be identified with the secondary

binding site localized on the side of the trimer, where a pocket is created by the five-turn of one molecule (residues 9–14) and the five  $\beta$ -strands antiparallel sheet (residues 17–102) of the other molecule. The main putative binding site is created by three main loops, which comprised residues 21–62, 80–86, and 143–145. The first loop is very long and form two spikes (residues 24–30 and 35–52) where the second is very extended and was found to be stabilized by a calcium ion coordinated by Asn39, Ser40, and Asp47. The second loop forming a third spike contains the GRGDHD motif, found in the cell attachment sequence of fibronectin, which is necessary in DiscI for binding to the substratum and for the formation of organized streams.<sup>14,15</sup> The binding surface presents three narrow clefts formed by charged residues of spike 1 and 3, three depressions formed by basic residues (His28, Arg81, and His144) and stacking interactions of Tyr24 and His84 and a central deep pocket at the trimer interface formed by Asn25, Ala26, and Asp27 [Fig. 2(C)]. We proposed that the ligand will preferentially bind into the three depressions where molecules of PEG, glycerol, or MPD are always observed. It is also noted that His84 would be involved in the interaction

**Table II**  
Interactions Protein–Sugars in DiscII Complexes

Atom	GalNAc complex		BGal complex	
	1st Hbond	2nd Hbond	1st Hbond	2nd Hbond
O1 $\alpha$	HOH (3.0) HOH (2.84)	Tyr244 <sup>a</sup> OH (3.0)		
O1 $\beta$	HOH (2.7)			
N2	HOH (2.74)	HOH (2.73) Tyr244 <sup>a</sup> (3.2)		
O7	HOH (3.0)	Gly239 N (2.90)		
O3	Trp238 NE1 (2.93) Asp209 <sup>a</sup> OD2 (2.67)		Trp238 NE1 (2.77) Asp209 <sup>a</sup> OD2 (2.66)	
O4	Arg218 NE (2.83) Asp209 <sup>a</sup> OD1 (2.61)		Arg218 NE (2.63) Asp209 <sup>a</sup> OD1 (2.66)	
O5	Arg218 NH2 (3.3)		Arg218 NH2 (3.1)	
O6	HOH (2.90) HOH (2.63)	Asn177 <sup>a</sup> ND2 (3.1) Gln216 OE1 (3.3)	HOH (2.75) HOH (3.0)	Asn177 <sup>a</sup> ND2 (3.2) Gln216 OE1 (3.1)
C4–C5		Van der Waals Tyr244 <sup>a</sup>		Van der Waals Tyr244 <sup>a</sup>

Distances (Å) are indicated between parenthesis. Direct (1st Hbond) and water-bridged (2nd Hbond) are given for the ligand.

<sup>a</sup>Residue from the neighboring monomer.

with the ligand and its phosphorylation could therefore regulate ligand binding and DiscII activity.

### Sugar binding molecular basis and thermodynamics

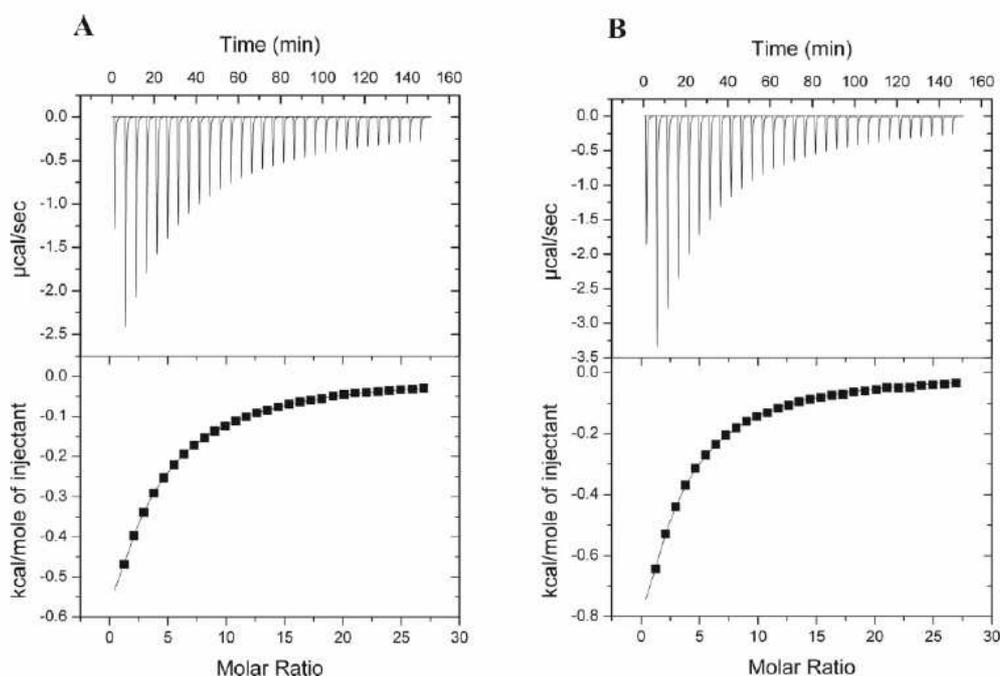
Examination of the difference Fourier map reveals unambiguously the binding of one monosaccharide per monomer in both GalNAc and  $\beta$ MeGal complexes [Fig. 3 (B,C)]. The sugar binding site is located in a shallow groove at the interface between two monomers on the C-terminal domain [Fig. 3(D)]. In both complexes, direct hydrogen bonding interactions are observed with the protein (Table II): O3 hydroxyl with Asp209 OD2 and Trp238 NE1 (neighboring monomer), O4 hydroxyl with Asp209 OD1 and Arg218 NE (neighboring monomer) and O5 hydroxyl with Arg218 NH1 (neighboring monomer). The O6 hydroxyl makes hydrogen bond with Asn177 ND2 (neighboring monomer) and Gln216 OE1 via a water molecule as well as the N2 nitrogen of GalNAc with Tyr244 OH (neighboring monomer). No interactions are observed for the O2 and O1 hydroxyl in the complex with  $\beta$ MeGal. The protein cannot discriminate between the two anomers and both are observed in the GalNAc complex with a preference for the  $\alpha$ -anomer (70% occupancy based on B factor). The equatorial and axial orientations of the O-3 and O-4 hydroxyl, respectively, are decisive for binding since on one hand, any other stereochemistry would lead to steric hindrance with Tyr244 side chain which stacks against the sugar, and on the other hand, they are involved in crucial hydrogen bonds with Asp209, Arg218, and Trp238. These three amino acids are strictly conserved between the members of the H-type lectin family and these interac-

tions explain DiscII specificity for monosaccharides with *D-galacto* configuration.<sup>6</sup>

To measure the affinity between DiscII and its ligands (GalNAc and  $\beta$ MeGal) the thermodynamic binding parameters and the  $K_a$  values were determined using titration microcalorimetry, a method that is well suited for the characterization of protein–carbohydrate interactions.<sup>43</sup> The experiments were only performed with the recombinant protein since the wild-type protein was copurified with DiscI. The thermograms show a large decrease in the exothermic heat of binding while saturation is achieved (see Fig. 4). Using a one site model, a stoichiometry of one ligand per monomer is observed in both cases together with dissociation constants of 1.15 and 0.95 mM for GalNAc and  $\beta$ MeGal, respectively. The interaction was found to be strongly driven by enthalpy with  $\Delta H$  values of  $-20.3$  and  $-24$  kJ/mol, respectively. However, as often observed for the binding of carbohydrates to lectins,<sup>43</sup> the change in enthalpy is strongly counterbalanced by a large, negative change in entropy  $T\Delta S$  of  $-3.5$  and  $-6.7$  kJ/mol, respectively.

### C-terminal lectin domain and related lectins

A search in the structural database by either DALI or MSDfold<sup>13,44</sup> reveals similarities with many proteins of unrelated function sharing the widespread immunoglobulin-like fold.<sup>35,45</sup> As expected from sequence alignment, the closer structural neighbor of DiscII C-terminal domain is the H-type lectin HPA with a rmsd of 1.4 Å for 84 out of 98 equivalent C $\alpha$ . The next two closest structures with MSDfold are the capsid protein of the Brome mosaic virus and the copper resistance protein C.<sup>46,47</sup>



**Figure 4**

Isothermal titration microcalorimetry results for GalNAc (A) and  $\beta$ MeGal (B) monosaccharides (18 mM) binding to DiscII (0.15 mM) in a 0.02M Tris.HCl, 0.15M NaCl, pH 7.5 buffer solution at 25°C. Top, 30 injections of 10  $\mu$ L of sugar solution at 18 mM were added at intervals of 5 min while stirring at 310 rpm in the DiscII containing cell (1.455 mL). 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-square fit for the obtained data.

Many viral coat proteins and carbohydrate binding domains are also found but they present one additional strand and very different loop structures.

DiscII C-terminal domain presents the same trimeric association and sugar binding site location than the HPA lectin [Fig. 5(A)]. Contrary to HPA, it is not stabilized by disulfide bridges and cannot form hexamer because of the presence of the DS domain. The main structural differences observed between the two structures are shorter first loop and fourth  $\beta$ -strand in DiscII. There is indeed an eight amino acid deletion in the first loop of the C-terminal domain (residues 171–182) which is involved in sugar binding, [Fig. 5(B)]. In HPA, this loop and particularly Gly24, Arg25, and Asp26 are involved in several hydrogen bonds which stabilize the *N*-acetyl and O3 hydroxyl of GalNAc. The length and sequence of this loop seem to have a role in the sugar specificity and affinity.

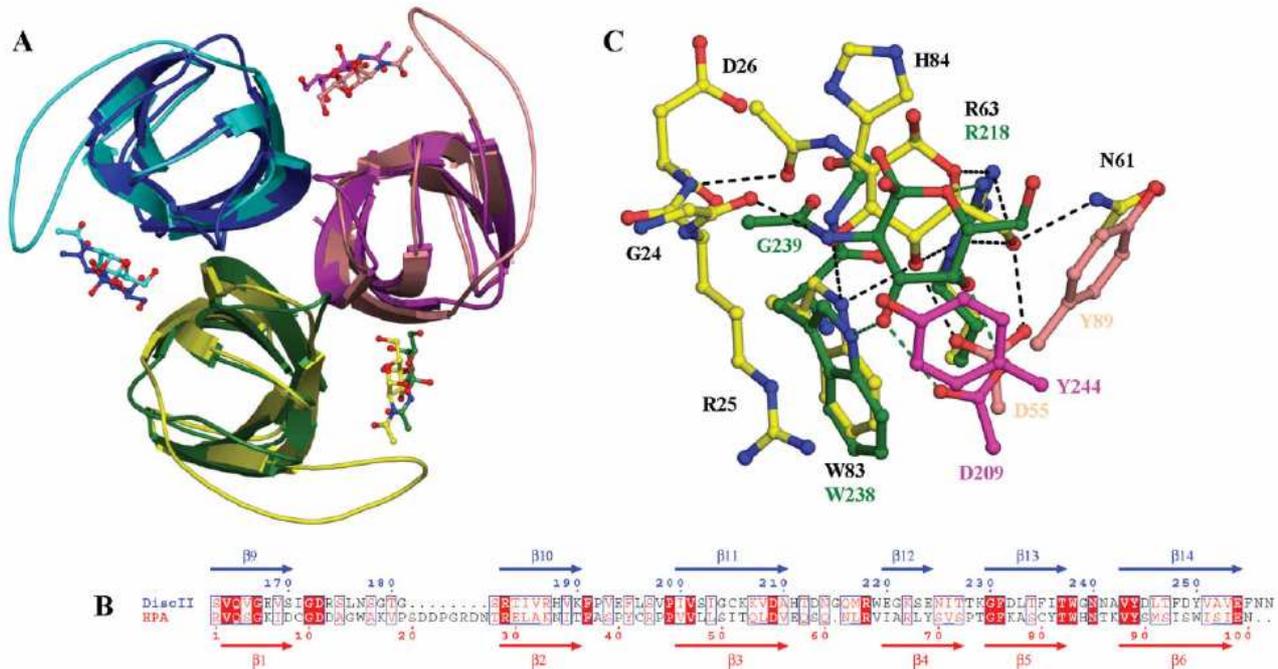
In HPA, all hydroxyls group establish direct hydrogen bonds with the protein. In DiscII, only interactions with the O3 and O4 hydroxyls are direct whereas the others are mediated by water molecules [Fig. 5(C)]. DiscII binding site (shallow groove) is more open and less constrained than HPA's (shallow pocket) which could explain why the GalNAc affinity for DiscII is 10 times lower than

for HPA. HPA is specific for  $\alpha$ -D-GalNAc due to a steric hindrance of His84 with the beta anomer while the equivalent residue, Gly239 in DiscII allows the binding of both anomers.

## CONCLUSIONS

Discoidin II is composed of two distinct domains which possess a definite role for its activity during the morphogenesis of *D. discoideum*. A compact and stable trimer is produced through the intertwining of those two domains that do not required disulfide bridges for stability as observed in structurally related domains.<sup>13,34,38</sup> Oligomerization is essential to DiscII function since carbohydrate binding sites are found at the interface between two monomers. It is necessary in many DS domain containing protein such as retinochisin (octamer), DDRs (dimer)<sup>11</sup> and is observed in many lectins such as HPA.<sup>13</sup> This creates multivalency and enhanced binding strength in order to compensate for the low affinity which is often observed in protein–carbohydrate interactions.<sup>49</sup>

The C-terminal domain of DiscII mediates the recognition of a polysaccharide produced in late aggregation

**Figure 5**

Superposition of the lectin domain of DiscII (green, blue, magenta) and HPA (yellow, cyan, salmon) (A). Sequence alignment based on structural superposition using Indonesia (<http://xray.bmc.uu.se/dennis/>) (B), figure drawn with Esprit.<sup>48</sup> Overlay of the active site of DiscII and HPA in presence of GalNAc (C).

and during fruiting body formation like DiscII. It is mainly composed of galactose and GalNAc.<sup>8</sup> The association of DiscII with glycoconjugates in the prespore vesicles before externalization with them is proposed.<sup>9</sup> Once at the cell surface, DiscII could be involved in cell-adhesion events via its DS domain where a major ligand-binding site has been identified. This binding site was previously described in homologous structures and it was shown that the sequence and orientation of several surface loops have a decisive role in binding site topology and protein specificity. A unique feature in the DS domain of DiscII is the possibility to regulate binding through kinase activity since a phosphorylated histidine has been identified on the border of the binding surface. Further studies are required to determine the natural ligands, definite role and regulation of DiscII to allow a better understanding of *D. discoideum* morphogenesis.

## ACKNOWLEDGMENTS

The authors would like to thank the European Synchrotron Radiation Facility (ESRF), for access to data collection facilities. We thank Gerard Klein (CEA, Grenoble) for providing us with *Dictyostelium discoideum* mRNA and Edward Mitchell for his help during data collection on ID14-2.

## REFERENCES

1. Firtel RA, Meili R. Dictyostelium: a model for regulated cell movement during morphogenesis. *Curr Opin Genet Dev* 2000;10:421–427.
2. Steinert M, Heuner K. Dictyostelium as host model for pathogenesis. *Cell Microbiol* 2005;7:307–314.
3. Williams RS, Boeckeler K, Graf R, Muller-Taubenberger A, Li Z, Isberg RR, Wessels D, Soll DR, Alexander H, Alexander S. Towards a molecular understanding of human diseases using *Dictyostelium discoideum*. *Trends Mol Med* 2006;12:415–424.
4. Willard SS, Devreotes PN. Signaling pathways mediating chemotaxis in the social amoeba, *Dictyostelium discoideum*. *Eur J Cell Biol* 2006; 85:897–904.
5. Bonner JT. Aggregation and differentiation in the cellular slime molds. *Annu Rev Microbiol* 1971;25:75–92.
6. Frazier WA, Rosen SD, Reitherman RW, Barondes SH. Purification and comparison of two developmentally regulated lectins from *Dictyostelium discoideum*. Discoidin I and II. *J Biol Chem* 1975;250: 7714–7721.
7. Berger EA, Armant DR. Discoidins I and II: common and unique regions on two lectins implicated in cell–cell cohesion in *Dictyostelium discoideum*. *Proc Natl Acad Sci USA* 1982;79:2162–2166.
8. Cooper DN, Lee SC, Barondes SH. Discoidin-binding polysaccharide from *Dictyostelium discoideum*. *J Biol Chem* 1983;258:8745–8750.
9. Barondes SH, Cooper DN, Haywood-Reid PL. Discoidin I and discoidin II are localized differently in developing *Dictyostelium discoideum*. *J Cell Biol* 1983;96:291–296.
10. Baumgartner S, Hofmann K, Chiquet-Ehrismann R, Bucher P. The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. *Protein Sci* 1998;7:1626–1631.

11. Kiedzińska A, Smietana K, Czepczynska H, Otlewski J. Structural similarities and functional diversity of eukaryotic discoidin-like domains. *Biochim Biophys Acta* 2007;1774:1069–1078.
12. Jimbo M, Koike K, Sakai R, Muramoto K, Kamiya H. Cloning and characterization of a lectin from the octocoral *Simularia lochmodes*. *Biochem Biophys Res Commun* 2005;330:157–162.
13. Sanchez JF, Lescar J, Chazalet V, Audfray A, Gagnon J, Alvarez R, Breton C, Imberty A, Mitchell EP. Biochemical and structural analysis of *Helix pomatia* agglutinin. A hexameric lectin with a novel fold. *J Biol Chem* 2006;281:20171–20180.
14. Springer WR, Cooper DN, Barondes SH. Discoidin I is implicated in cell-substratum attachment and ordered cell migration of *Dictyostelium discoideum* and resembles fibronectin. *Cell* 1984;39:557–564.
15. Gabius HJ, Springer WR, Barondes SH. Receptor for the cell binding site of discoidin I. *Cell* 1985;42:449–456.
16. Simpson DL, Rosen SD, Barondes SH. Discoidin, a developmentally regulated carbohydrate-binding protein from *Dictyostelium discoideum*. Purification and characterization. *Biochemistry* 1974;13:3487–3493.
17. Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* 2001;14:993–1000.
18. Wiseman T, Williston S, Brandts JF, Lin LN. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* 1989;179:131–137.
19. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. *Jnt CCP4/ESF-EACMB Newsletter on Protein Crystallography*, 1992. p 26.
20. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 1993;26:795–800.
21. Collaborative computational project number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 1994;50:760–763.
22. Vonrhein C, Blanc E, Roversi P, Bricogne G. Automated structure solution with autoSHARP. *Methods Mol Biol* 2006;364:215–230.
23. Navaza J, Saludjian P. AMoRe: an automated molecular replacement program package. In: C W Carter J, Sweet RM, editors. *Methods Enzymol*. Volume 276. London: Academic Press; 1997. pp 581–594.
24. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 2005;61:458–464.
25. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–255.
26. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 2004;60:2126–2132.
27. Loomis WF, Shaulsky G, Wang N. Histidine kinases in signal transduction pathways of eukaryotes. *J Cell Sci* 1997;110:1141–1145.
28. Thomason PA, Sawai S, Stock JB, Cox EC. The histidine kinase homologue DhkK/Sombrero controls morphogenesis in *Dictyostelium*. *Dev Biol* 2006;292:358–370.
29. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995;24:536–540.
30. Potterton E, Briggs P, Turkenburg M, Dodson E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr D Biol Crystallogr* 2003;59:1131–1137.
31. Ichikawa O, Osawa M, Nishida N, Goshima N, Nomura N, Shimada I. Structural basis of the collagen-binding mode of discoidin domain receptor 2. *EMBO J* 2007;26:4168–4176.
32. Lee CC, Kreuzsch A, McMullan D, Ng K, Spraggon G. Crystal structure of the human neuropilin-1 b1 domain. *Structure* 2003;11:99–108.
33. Lin L, Huai Q, Huang M, Furie B, Furie BC. Crystal structure of the bovine lactadherin C2 domain, a membrane binding motif, shows similarity to the C2 domains of factor V and factor VIII. *J Mol Biol* 2007;371:717–724.
34. Macedo-Ribeiro S, Bode W, Huber R, Quinn-Allen MA, Kim SW, Ortel TL, Bourenkov GP, Bartunik HD, Stubbs MT, Kane WH, Fuentes-Prior P. Crystal structures of the membrane-binding C2 domain of human coagulation factor V. *Nature* 1999;402:434–439.
35. Pratt KP, Shen BW, Takeshima K, Davie EW, Fujikawa K, Stoddard BL. Structure of the C2 domain of human factor VIII at 1.5 Å resolution. *Nature* 1999;402:439–442.
36. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 2004;60:2256–2268.
37. Fuentes-Prior P, Fujikawa K, Pratt KP. New insights into binding interfaces of coagulation factors V and VIII and their homologues lessons from high resolution crystal structures. *Curr Protein Pept Sci* 2002;3:313–339.
38. Bianchet MA, Odom EW, Vasta GR, Amzel LM. A novel fucose recognition fold involved in innate immunity. *Nature Struct Biol* 2002;9:628–634.
39. Boraston AB, Wang D, Burke RD. Blood group antigen recognition by a *Streptococcus pneumoniae* virulence factor. *J Biol Chem* 2006;281:35263–35271.
40. Gaskell A, Crennell S, Taylor G. The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* 1995;3:1197–1205.
41. Ito N, Phillips SE, Stevens C, Ogel ZB, McPherson MJ, Keen JN, Yadav KD, Knowles PF. Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature* 1991;350:87–90.
42. Wendt KS, Vodermaier HC, Jacob U, Gieffers C, Gmachl M, Peters JM, Huber R, Sondermann P. Crystal structure of the APC10/DOC1 subunit of the human anaphase-promoting complex. *Nature Struct Biol* 2001;8:784–788.
43. Dam TK, Brewer CF. Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem Rev* 2002;102:387–429.
44. Lescar J, Sanchez JF, Audfray A, Coll JL, Breton C, Mitchell EP, Imberty A. Structural basis for recognition of breast and colon cancer epitopes Tn antigen and Forssman disaccharide by *Helix pomatia* lectin. *Glycobiology* 2007;17:1077–1083.
45. Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993;233:123–138.
46. Lucas RW, Larson SB, McPherson A. The crystallographic structure of brome mosaic virus. *J Mol Biol* 2002;317:95–108.
47. Zhang L, Koay M, Maher MJ, Xiao Z, Wedd AG. Intermolecular transfer of copper ions from the CopC protein of *Pseudomonas syringae*. Crystal structures of fully loaded Cu(I)Cu(II) forms. *J Am Chem Soc* 2006;128:5834–5850.
48. Gouet P, Courcelle E, Stuart DI, Metz F. ESPript: multiple sequence alignments in postscript. *Bioinformatics* 1999;15:305–308.
49. Sharon N. Protein-carbohydrate interactions—at the art of biochemistry. *The Biochemist* 2006; pp 1–5.

# **Discoidin I from *Dictyostelium discoideum*: structure and specificity comparison with discoidin II**

Karoline Saboia Aragão, Anne Imberty and Annabelle Varrot\*

CERMAV-CNRS (affiliated with University Joseph Fourier), BP53, F-38041 Grenoble cedex 09, France.

\* Corresponding author: Telephone 33-476037634, Fax: 33-476547203, E-mail: [varrot@cermav.cnrs.fr](mailto:varrot@cermav.cnrs.fr)

## **Abstract**

Discoidin I (DiscI), a galactose-binding protein is a marker of early development in *Dictyostelium discoideum*. It was cloned and expressed in high yield in a recombinant form in *Escherichia coli* bacteria. Purified recombinant protein gave crystals of good quality and its 3-D structure was solved by X-ray crystallography. We investigated its interactions with oligosaccharides using isothermal calorimetry experiments that provided precise determination of affinity for the ligands and revealed an enthalpy contribution. GalNAc was introduced into the crystals of DiscI by soaking and the complex structure was determined. Thus a combination of structures and thermodynamics permits the identification and analysis of the binding site at a molecular level.

## INTRODUCTION

Lectins are defined as ubiquitous proteins of non-immune origin that interact reversibly and specifically with carbohydrates without modifying them <sup>1</sup>. 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 either establishment of symbiosis or precipitation of pathogens <sup>2;3</sup>. In other cases, they are involved in self-recognition and in cell-cell aggregation that is required for the building of multicellular organisms such as corals and sponges <sup>4</sup>.

The slime mould *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 <sup>5</sup>. The aggregation of individual *Dictyostelium* amoebae occurs by chemotaxis to periodic cAMP signals (review in <sup>6</sup>). Two galactose-binding proteins Discoidin I and II (DiscI and II) were characterized in the slime mold. They were extensively studied in the seventies or eighties but a lot of questions remain about their role and ligand. They are produced upon starvation in cohesive cells and their expression is developmentally-regulated as a result of differential gene expression <sup>7; 8</sup>. The production of DiscI is prominent during aggregation where it comprises about 1% of the cellular proteins whilst DiscII is prominent during fruiting body formation and is localized in the prespore vesicles <sup>9</sup>. DiscI is polymorphic with at least three isoforms per strain <sup>10;11</sup>. The three genes *dscA*, *dscC* and *dscD* are found on chromosome 2 and are duplicated on this chromosome in some strain (AX3 or AX4). Two copies of a homologue are also found on chromosome I but their role is unknown. DiscI is used as a marker of early development in *D. discoideum* <sup>12</sup> and its lectin activity seems dependant on calcium ion <sup>13; 14</sup>. DiscI was shown to be involved in cell substratum attachment and ordered cell migration during aggregation with the cell adhesion site different from the sugar binding site <sup>15</sup>. DiscI is proposed to function in association with glycoconjugates in the slime around aggregates <sup>16</sup> and was found to bind a cell surface glycoprotein of 67 kDa which has not yet been identified <sup>17</sup> via its GRGD motif known to be involved in cell attachment in other proteins review in <sup>18</sup>.

DiscI is a two domains protein of 253 amino acids <sup>10</sup> with higher affinity for N-acetylgalactosamine (GalNAc) than for Galactose (Gal) <sup>8</sup>. It displays 48% sequence identity

with DiscII whose trimeric structure has been recently solved in unliganded state and in complex with Gal and GalNAc<sup>19</sup>. Both DiscI and DiscII are two domains proteins. The *N*-terminal domain referred as the discoidin domain (DS domain) is a structural and functional motif found in various proteins, both eukaryotic and prokaryotic review in<sup>20</sup> It shows a considerable functional diversity via interactions with numerous molecules but is mainly involved in cell adhesion, migration or proliferation, cell-surface signaling and developmental processes<sup>21</sup>. The *C*-terminal domain belongs to the H-type lectin family that has been identified in invertebrates such as snails and corals<sup>22</sup>.

We present here the crystal structure of recombinant DiscI unliganded and in complex with GalNAc. Specificity and affinity for variety of carbohydrate ligands are analysed. DiscI and DiscII present interesting differences in carbohydrate specificity, that may be correlated to their differences in localisation in the slime mold. They are compared in term of specificity and affinity towards oligosaccharides and comparison of binding site structures is discussed.

## **MATERIALS AND METHODS**

### ***Cloning, Expression and Purification of recombinant DiscI***

The discoidin I gene of isoform A (P02886) was cloned in-frame with a *N*-terminal hexahistidine tag and the tobacco-etch virus (TEV) protease cleavage site of the prokaryotic expression vector pProEXHTb (Invitrogen) using polymerase chain reaction and standard molecular biology techniques. The primers used were the following: 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 in Luria broth medium (LB). Cells were cultured at 37°C up to an optical density of 1.2–1.3 at 600nm before induction with 0.2 mM IPTG overnight at 16°C and harvested by centrifugation. Each gram of cell was resuspended in 3ml of buffer A: 20 mM Tris/HCl, 500 mM NaCl and 10mM imidazol, pH 8.5 before using a cell disrupter (Constant system Ltd). After centrifugation 30 min at 40,000g, the supernatant was loaded on 5 mL of Sepharose Fast Flow resin (GE healthcare) charged with nickel. After a wash step of five volumes with buffer A, DiscI was eluted with a gradient 10-500mM imidazol. The Histag was cut by addition of the TEV protease at the ratio 1/100 overnight at 20°C in the presence of 2 mM

DTT. Imidazol was removed using a PD10 Column (GE healthcare) equilibrated with Tris-HCl 20mM, NaCl 150mM pH 7.5 before reloading on the affinity column. The pure protein without tag is eluted in the flow-through and concentrated to 5.8 mg/mL using Microsep™ 10 K (Pall). Each purification step was analyzed by SDS-PAGE 12%.

### ***ITC (Isothermal Titration Microcalorimetry) analysis***

ITC experiments were performed using a Microcal VP-ITC microcalorimeter. All titrations were carried out in 20 mM Tris-HCl and 150 mM NaCl, pH 7.5 at 25 °C. Sugars and proteins were dissolved in the same buffer. A total of 30 injections of 10 µL of sugar solution at concentrations varying from 2 to 40 mM were added at intervals of 5 min 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 solution yielded insignificant heats of dilution. Integrated heat effects were analyzed by non-linear regression using a single-site binding model (Origin 7.0). The experimental data was fitted and data yielded the association constant (K<sub>a</sub>) and the enthalpy of binding (ΔH). 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 (ΔG), entropy (ΔS), and number of binding sites per monomer (n) were calculated from the equation:

$$\Delta G = -RT \ln K = \Delta H - T \Delta S$$

where T is the absolute temperature, R=8.32 J mol<sup>-1</sup>K<sup>-1</sup> and K is the association constant. At least two independent titrations were performed for each ligand tested.

### ***The Glycan Microarray Resource***

Recombinant Discoidin I and II were labelled with Alexa Fluor® 488-TFP (carboxylic acid, 2,3,5,6 tetrafluorophenyl) (Invitrogen) according to the manufacturer's instructions and purified on a D-Salt™ polyacrylamide 6000 desalting column (Pierce). These proteins were used to analyze the printed glycan arrays following the standard procedure of the Consortium for Functional Glycomics (<http://www.functionalglycomics.org/>).

## ***Crystallization and structure determination***

All crystals were obtained by the hanging-drop vapor diffusion method using 2  $\mu$ l drops containing a 50:50 (v/v) mix of the protein and the reservoir solution at 20 °C. Rod crystals were obtained from the solution 16 of the Hampton crystallization screen I (Hampton research) containing 1.5 M lithium sulfate and 100 mM NaHepes pH 7.5 after almost one month. GalNAc was introduced by soaking the crystals for a short time in the mother liquor supplemented with 10 mM GalNAc. Before data collection, the crystals were transferred in a cryoprotecting solution (25% MPD were added to the mother liquor), mounted in a rayon fiber loop and frozen at 100K. Native and complex data were collected at 1.8 and 1.75 Å resolution respectively in the C222<sub>1</sub> spacegroup on beamline ID14-2 at ESRF (Grenoble, France) using an ADSC Q4 CCD detector (Quantum Corp.). The data were processed and reduced using MOSFLM and SCALA<sup>23</sup> All further computing was performed using the CCP4 suite, unless otherwise stated<sup>24</sup>.

## ***Structure Determinations***

The native structure of Discoidin I was solved by molecular replacement using PHASER<sup>25</sup> and the coordinates of the Discoidin II structure (PDB code 2VM9)<sup>19</sup>. The DiscI-GalNAc complex structure was solved by molecular replacement using PHASER and the coordinates of the native DiscI. For the refinement of each structure, 5% of the observations were immediately set aside for cross validation analysis<sup>26</sup> and were used to monitor various refinement strategies such as the weighting of geometrical and temperature factor restraints.

Manual corrections of the model using Coot<sup>27</sup> were interspersed with cycles of maximum-likelihood refinement<sup>28</sup>. Water molecules were added in Coot. The quality of the model was assessed in SFCHECK and PROCHECK<sup>29; 30</sup> before its deposition in the Brookhaven Protein Data Bank<sup>31</sup>. Details of the data and model quality are given in the table 1. All figures were drawn with PyMOL Molecular Graphics System (2002) DeLano Scientific, Palo Alto, CA, unless otherwise stated.

Table 1. Data collection and refinement statistics

	<b>DiscI</b>	<b>DiscI/GalNAc</b>
Spacegroup	C222 <sub>1</sub>	C222 <sub>1</sub>
Cell dimensions <i>a,b,c</i> (Å)	72.5, 274.3, 85.9	72.4, 274.0, 86.1
$\alpha, \beta, \gamma$ (deg.)	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)
Measured/ Unique reflections	299983/75469	390844/86561
Average multiplicity	4.0(3.8)	4.5(4.1)
Completeness (%)	94.8(94.8)	99.9(99.9)
Average $I / \sigma I$	5.9(2.0)	8.1(1.8)
$R_{cryst} / R_{free}$	16.6/20.7	15.5/19.0
$R_{msd}$ bonds, Å	0.014	0.017
$R_{msd}$ angles, °	1.6	1.5
Protein atoms/Bfac, Å <sup>2</sup>	1995/17.7	1990/15.3
Water molecules/ Bfac, Å <sup>2</sup>	974/30.7	1059/29.5
Ligand atoms/ Bfac, Å <sup>2</sup>		15/23.7

### ***Molecular modelling***

Disaccharide  $\beta$ Gal1-3 $\beta$ GalNAc and trisaccharide  $\beta$ Gal1-4 $\beta$ GlcNAc1-6 $\beta$ GalNAc have been modeled in the binding site of DiscI and DiscII, respectively, using the following procedure. For both proteins, the crystal structure of the complex with GalNAc was used as starting point (this work for DiscI and PDB 2VMC for DiscII). The protein moiety was edited with the use of the Sybyl software (St Louis, US). Two adjacent lectin domains were extracted. All the hydrogen atoms were added, the partial charges were included with the Pullman library and the position of hydrogen atoms were optimized with the use of the Tripos force-field<sup>32</sup>. The disaccharide and trisaccharide were built using the monosaccharide library available at Glyco3D (<http://www.cermav.cnrs.fr/glyco3d/>) using the PIM parameters and partial charges for carbohydrates. The reducing GalNAc moieties were placed in the binding site by fitting on the GalNAc in the crystal structures. Possible orientations of the remaining residues were calculated by systematic search around the glycosidic linkages. The lowest energy conformations were further optimized by energy minimization of the full oligosaccharide and the protein hydrogen atoms. The validity of the final conformations was checked by comparison with available energy maps (Glyco3D).

## RESULTS

### ***Specificity of D. discoideum lectins with Glycan Microarray Resource***

Recombinant DiscI was obtained with a yield of 20 mg/liter of culture and the MW checked by mass spectrometry corresponded to the expected value. Recombinant DiscII was produced as previously described<sup>19</sup>. Both lectins were assayed on Glycan Microarray available at the Consortium for Functional Glycomics. Concentrations from 0.01 to 200 µg/mL were assayed. DiscI gave excellent data at very low concentration and the analysis displayed in Figure 1A has been obtained for 0.04 µg/mL. Strong binding to glycochips was observed also for DiscII and Figure 1B has been obtained with 0.5 µg/mL. Quite surprisingly, there is almost no superimposition in the binding specificity of the two discoidins. When looking at disaccharides, Discoidin I binds to several β1,3 linked disaccharides with a *galacto* residue at the reducing position such as GlcNAcβ1,3GalNAc (n° 159), Galβ1,3GalNAc (n° 123) and less strongly to GalNAcβ1,3Gal (n° 300). Elongation by additional residues or by sulphate groups are well tolerated on position 3 or 4 of the non-reducing sugar while the reducing sugar can be only elongated at position 1 and no branching is accepted. Polylactosamine oligosaccharides containing repetitive Galβ1,4GlcNAcβ1,3Gal sequences are high affinity ligands. In few cases, such as in gangliosides of the GM2 to GT2 series (n° 201 to 209), terminal GalNAc are also recognized.

On the opposite DiscII does not recognize efficiently any disaccharides. N-acetyllactosamine (Galβ1,4GlcNAc) is the preferred motif but is recognized only when disulphated or part of larger oligosaccharides. The trisaccharide with the highest affinity is Galβ1,4GlcNAcβ1,6GalNAc (n° 149). Most of the high affinity oligosaccharides also contain a Gal/GalNAc residue with lactosamine on position 6.

The printed glycan array data indicate therefore that DiscI has higher affinity for GalNAc-containing oligosaccharide elongated on position 3, while DiscII prefers Gal/GalNAc-containing motif substituted on position 6. It could be noted that both lectins bind to large fucosylated and/or sialylated glycans such as n° 371 but this may be related to binding to the non-lectin domain as discussed below.

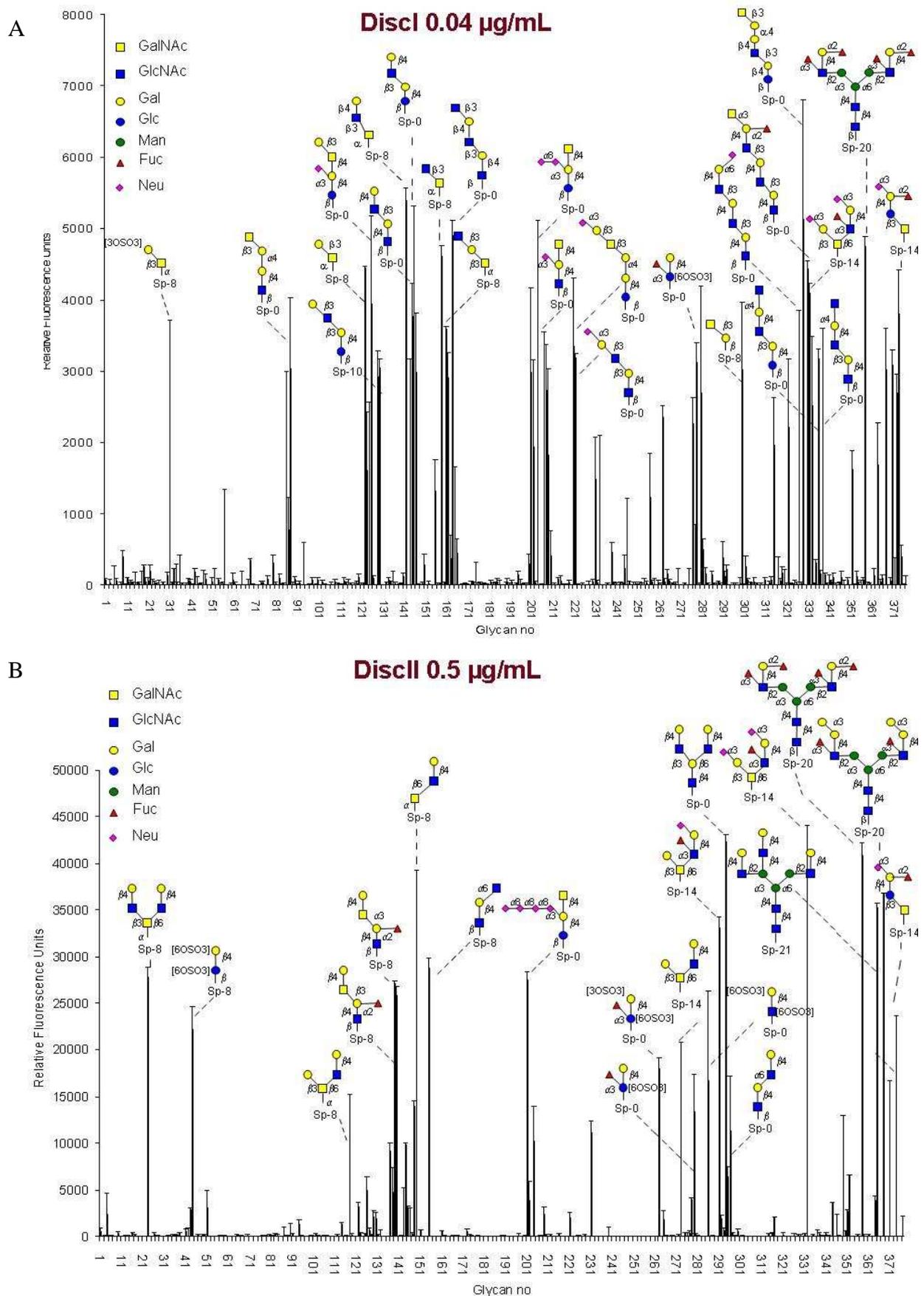


Figure 1. Printed glycan array analysis of A) DisclI (0.04  $\mu\text{g/mL}$ ) and B) DisclII (0.5  $\mu\text{g/mL}$ ) as measured by fluorescence intensity. Complete results including comprehensive oligosaccharide list (Plate Array) are available from the Consortium of Functional Glycomics (<http://www.functionalglycomics.org/>).

## Affinity studies by ITC microcalorimetry

The affinity of DiscII towards GalNAc and  $\beta$ MeGal was previously determined using titration microcalorimetry <sup>19</sup>, leading to dissociation constants of 1.15 and 0.95 mM, respectively. When DiscI is assayed with the same approach, higher affinity is observed towards GalNAc ( $K_d = 0.3$  mM) while galactose and  $\beta$ MeGal have lower affinity ( $K_d$  in the range of 3 to 4 mM) (Table 2). DiscI affinities for disaccharides confirm the preference for  $\beta$ 1,3Gal/GalNAc linked disaccharides. Among tested disaccharides Gal $\beta$ 1,3GalNAc is the best ligand with  $K_d$  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 approx 50  $\mu$ M, while N-acetylglucosamine (Gal $\beta$ 1,4GlcNAc) is a poor ligand with  $K_d$  in the millimolar range.

Table 2. Microcalorimetry data. All values are averaged over two or three experiments. The reported standard deviations are calculated upon averaging of 2 or 3 experiments, unless otherwise specified.

Protein	Ligand	$K_D(\mu\text{M})$	n	$-\Delta G(\text{kJ/mol})$	$-\Delta H(\text{kJ/mol})$	$T\Delta S(\text{kJ/mol})$
<i>DiscI</i>	GalNAc	302 $\pm$ 27	1 <sup>a</sup>	20.0 $\pm$ 0.2	33.4 $\pm$ 5.8	13.3 $\pm$ 6.0
	Galactose	3770 $\pm$ 200	1 <sup>a</sup>	13.8 $\pm$ 0.1	31 $\pm$ 7.4	17.1 $\pm$ 7.5
	$\alpha$ MeGalactose	8390 $\pm$ 700	1 <sup>a</sup>	11.8 $\pm$ 0.2	42.4 $\pm$ 5.9	30.5 $\pm$ 6.1
	$\beta$ MeGalactose	2960 $\pm$ 270	1 <sup>a</sup>	14.4 $\pm$ 0.2	40.3 $\pm$ 2.4	25.9 $\pm$ 2.2
	LacNAc	2390 $\pm$ 170	1 <sup>a</sup>	15.0 $\pm$ 0.2	38.4 $\pm$ 2.2	23.3 $\pm$ 2.0
	Gal $\beta$ 1,3GalNAc	27.2 $\pm$ 4.6	1.03 $\pm$ 0.01	26.1 $\pm$ 0.4	45.6 $\pm$ 1.1	19.5 $\pm$ 1.5
	GalNAc $\beta$ 1,3Gal	49.4 $\pm$ 1.5	0.81 $\pm$ 0.01	24.5 $\pm$ 0.3	53.2 $\pm$ 2.9	28.81 $\pm$ 2.8
<i>DiscII</i>	GalNAc <sup>b</sup>	1115	1 <sup>a</sup>	16.8	20.3	3.5
	$\beta$ MeGalactose <sup>b</sup>	952	1 <sup>a</sup>	17.4	24.3	6.7
	LacNAc	1221 $\pm$ 23	1 <sup>a</sup>	16.7 $\pm$ 0.5	37.0 $\pm$ 0.3	20.5 $\pm$ 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

<sup>a</sup> fixed stoichiometry for low affinity interaction, <sup>b</sup> taken from <sup>19</sup>, <sup>c</sup> only one experiment due to lack of material.

The affinity of DiscII towards disaccharides was also tested and it was confirmed that none of the disaccharide cited above display high affinity, since all observed dissociation constants were in the millimolar range, as previously measured for DiscII interaction with Gal and GalNAc monosaccharides<sup>19</sup>.

Titration curves of DiscI interacting with GalNAc and Gal $\beta$ 1,3GalNAc are shown in Figure 2. The thermograms display a large decrease in the exothermic heat of binding while saturation is achieved. For all ligands, the interaction was found to be strongly driven by enthalpy with a strong unfavorable entropy contribution as often observed for the binding of carbohydrates to lectins<sup>33</sup>.

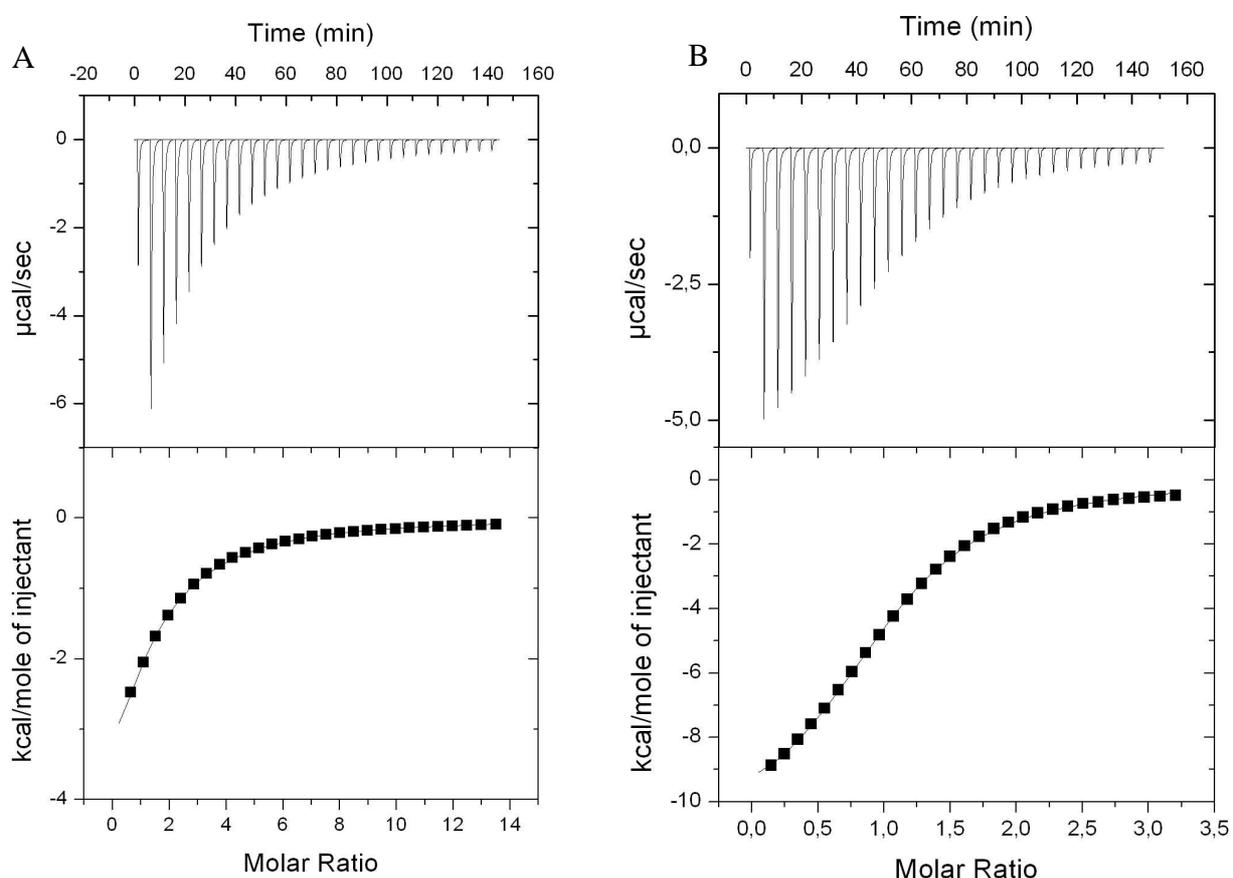


Figure 2. Isothermal titration microcalorimetry results for A) GalNAc monosaccharide (10 mM) and B) Gal $\beta$ 1,3GalNAc disaccharide (2mM) binding to DiscI (0.15 mM) in 20 mM Tris/HCl, 150 mM NaCl, pH 7.5 buffer solution at 25°C. Top: data from 30 automatic injections of 10  $\mu\text{l}$  each of sugar into the DiscI in the cell. Bottom: plot of the total heat released as a function of ligand concentration for the titration shown above (squares). The solid line represents the best least-square fit for the obtained data.

## Crystal structure of *Discl*

*Discl* gives rise to rectangular and reproducible crystals which diffracted to high resolution. They belong to the  $C222_1$  spacegroup with one trimer in the asymmetric unit. The final model of the native structure has a crystallographic Rvalue of 16.6, with a corresponding Rfree of 20.7 for all observed data between 28.6 and 1.8 Å resolution. In this model, all 253 amino acids could be observed in the electron density as well as some MPD molecules, sulphate, calcium and nickel ions (refinement statistics in table 1).

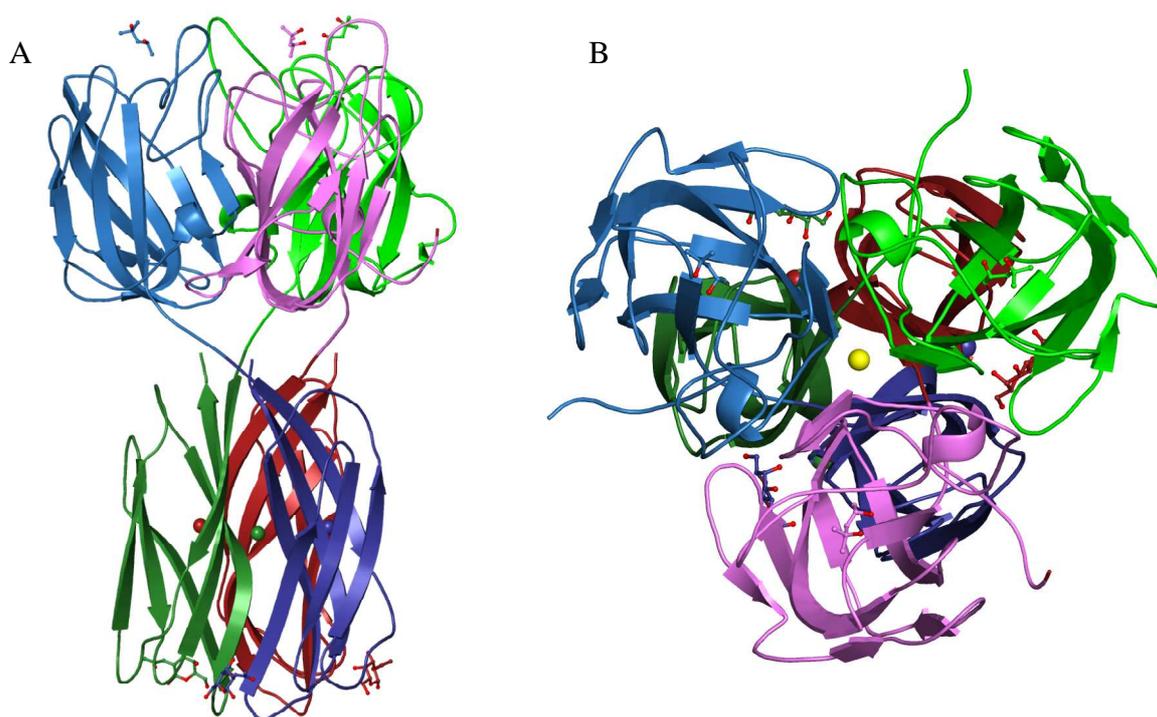


Figure 3. Crystal structure of the *Disc-I* trimer. (A). Representation of the trimer along the pseudo 3-fold axis. (B) Coloration by chain with the N-terminal domain in light colors and C-terminal domain in dark colors. MPD and GalNac are represented in ball and sticks and the calcium and nickel ions as spheres.

The *discl* has an overall fold equivalent to that observed for *DiscII* with the formation of a trimer through the intertwining of the two domains of the protein<sup>19</sup>. The N-terminal domain (1-152) present the classical topology of the discoidin 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, Figure 3<sup>34; 35</sup>. Prominent loops protrude like spikes on the opposite side of the termini creating a ligand binding site where a molecule of MPD is observed. The C-terminal domain presents the  $\beta$ -sandwich fold typical of the H-type lectins consisting of six antiparallel strands<sup>22</sup>. At the level of the pseudo 3-fold axis of the trimer,

some extra density appears between the three His16 side chains. It was modeled as nickel and probably comes from the affinity chromatography on nickel column. Inspection of the electron density map also reveals the presence of three calcium ions according to the distances and coordination<sup>36</sup>. Each calcium is found in the middle of C-terminal domain at the interface between two monomers. It presents classical seven coordinations with five in a plane, one above and one under the plane which comprise bidentate Asp246 ( $\beta$ 15), monodentate Asp164 ( $\beta$ 10) and Asp203 ( $\beta$ 12), Gln219 ( $\beta$ 13) from the neighboring monomer and two water molecules coordinating the main chain oxygens of Leu200 and Val198 ( $\beta$ 15) from the neighboring monomer. The interactions, which involve four  $\beta$ -strands over six, would help to maintain the trimeric arrangement of the lectin domain. That would explain the dependence of galactose binding to the presence of divalent cation and the destabilizing role of EDTA and EGTA observed for DiscI<sup>14</sup>. In DiscII, the position of the calcium is replaced by the NZ of Lys206, the trimerisation is not impaired despite very few interactions at this interface.

### ***Discoidin domain***

The N-terminal domain of DiscI presents a discoidin domain fold found in a variety of extracellular and membrane proteins from eukaryotes and prokaryotes such as coagulation factors, neuropilins, retinoschisin, discoidin domain receptors (DDRs)<sup>20; 37</sup>. Such domains evolved from a common ancestor to bind very diverse ligands (collagen, lipids, carbohydrate or growth factors) via a binding site formed by surface loops that protrude at opening of their core  $\beta$ -barrel structure.

The DS domain has an overall structure that is very similar to the one from DiscII. (rmsd of 0.87 Å for the 150 aligned C $\alpha$  calculated by Secondary Structure Matching<sup>38</sup>). In both DiscI crystal structures discussed here, three molecules of MPD (one per monomer) are observed in a depression at the opening of the  $\beta$ -barrel, Figure 4. In each binding site, the MPD oxygen atoms establishes hydrogen bonds with side chains of Arg79 and Gln84 while the hydrophobic faces makes contact with Tyr23, His27, Cys51 and His142. Analysis of the DiscII structures indicates that the equivalent binding sites host MPD in the native structure of the recombinant protein and PEG in the complexes structures<sup>19</sup>. The fact that MPD molecules are often observed in sugar binding sites, together with the overall similarity of the trimeric DS domain with Anguilla lectin AAA<sup>39</sup>, suggest that the DS domain of DiscI could belong to the F-type lectin family.

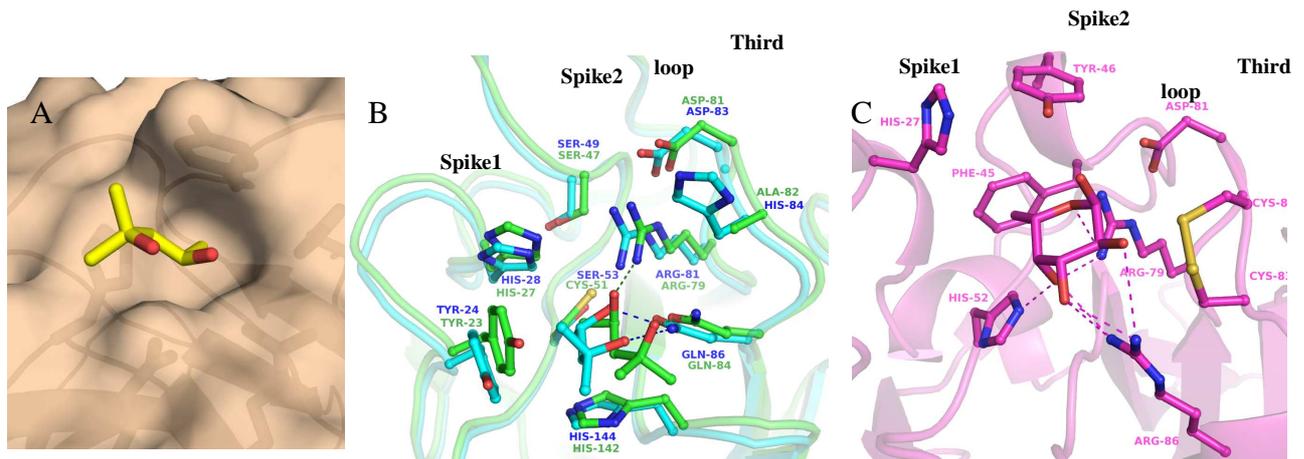


Figure 4. A) Surface representation of the MPD binding site. B) Superposition of the ligand binding site of DiscI (green), DiscII (cyan). C) Blow up of Fucose binding site of AAA (magenta).

The superposition of AAA lectin with the DS domain of DiscI shows that the binding site environment is similar with an rmsd of 2.25 Å for 125 out of 155 aligned C $\alpha$  (calculated with MSDFOLD)<sup>38</sup>. Binding sites of AAA, DiscI and DiscII are formed by the two spikes of the extended first loop (Asp23 to Thr57 in DiscI) and by the third loop Gly78 to Trp85 in DiscI) that contain most residues involved in binding, Figure 4. This third loop has the same conformation in discoidins and AAA and contains the RGD motif which promotes cell attachment to substrate in mammalian cell<sup>18</sup>. The Arg of this motif makes direct contact with fucose O4 and O5 hydroxyl groups in AAA and with the MPD in Discoidins. The conformation of the arginine is stabilized by interactions with the aspartate side chain of the RGD motif. This loop also presents a residue with an amine group (Gln84 in DiscI or Arg86 in AAA) which also make direct hydrogen bond with the ligand. The opposite side of the binding site is formed by the first surface loop with more variations between discoidins and AAA. At the end of spike 1, a histidine residue is found in discoidins (His27 in DiscI) and in AAA and this amino acid also plays a role in substrate binding.

The binding site of the DS domain of discoidins has therefore all required characteristics to bind carbohydrates and we suppose it will bind the N-glycans found in the printed glycan array. Those glycans are fucosylated and sialylated and do not have the prerequisite to bind to the C-terminal lectin domain as discussed below. Recently, the small protein containing only a DS domain: the retinoschisin was shown to bind galactose<sup>40</sup>. It is an octameric protein<sup>41</sup> which plays a crucial role in maintaining the cellular organization and synaptic structure of the vertebrate retina and the formation of the octamer is necessary for the galactose binding.

## *N*-acetyl-galactosamine binding site

The C-terminal domain of DiscI presents the H-type lectins fold first observed for Helix pomatia agglutinin (HPA) and then for DiscII<sup>19; 22</sup>. H-type lectins have trimeric arrangement and show specificity for galactose and GalNAc residues with the sugar binding site located in a shallow groove at the interface between two monomers. The complex structure of DiscI with GalNAc reveals unambiguously one monosaccharide per monomer in the first electron density map (Figure 5A).

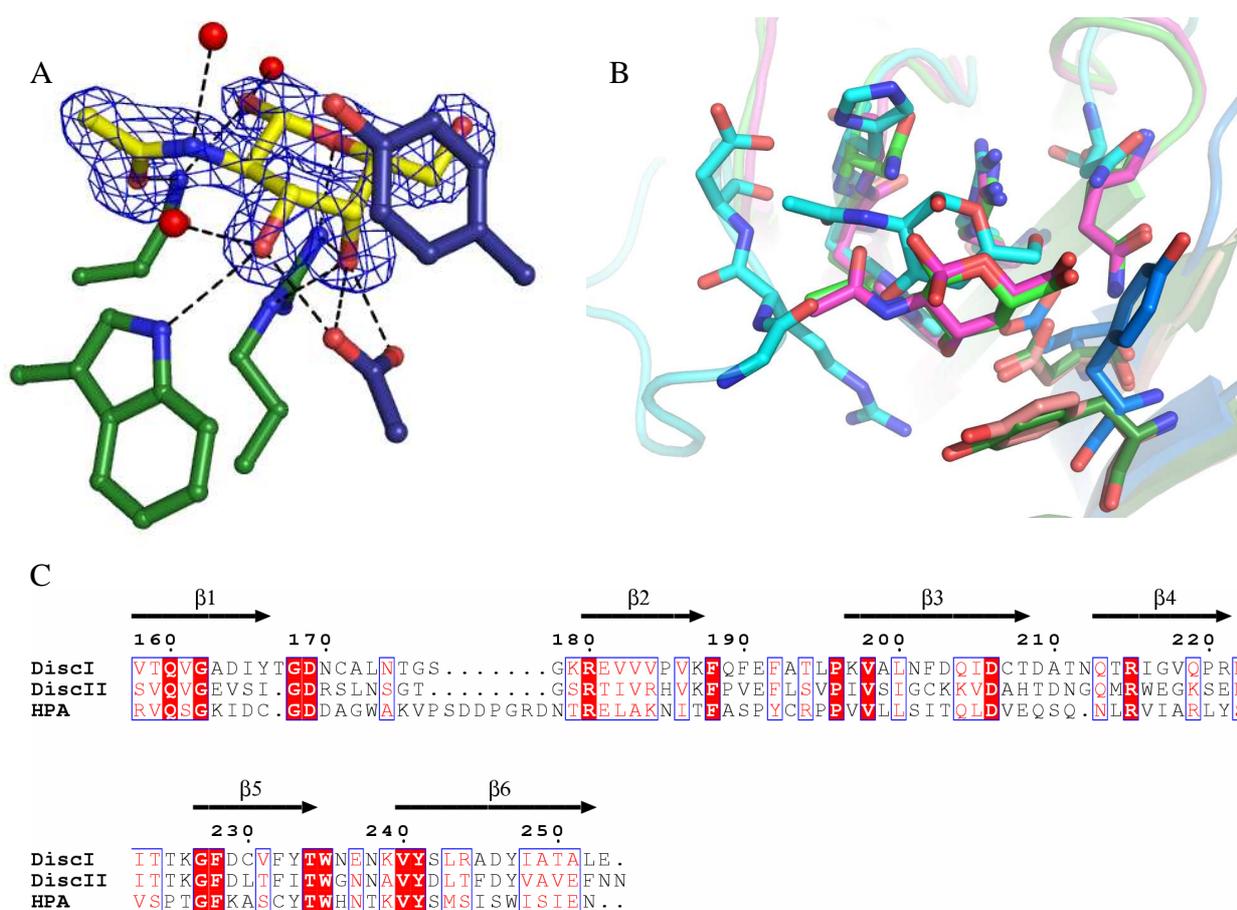


Figure 5. A: Representation of the maximum-likelihood weighted 2Fo-Fc electron density around the GalNAc residue complexed with DiscI. B: Overlay of the GalNAc binding site of DiscI (green), DiscII (magenta) and HPA (cyan). C) Structural based sequence alignment. (figure drawn with ESPrpt<sup>42</sup>).

Only direct hydrogen bonding interactions are observed with the protein. The O3 hydroxyl interacts with Trp235 and Asp206 (neighboring monomer), the O4 hydroxyl with Arg215 and Asp206 (neighboring monomer), the ring oxygen with Arg215, the O7 hydroxyl with Asn236 and the O1 hydroxyl with Asn236 (Details of atoms and distances are listed in

Table 3). No interactions are observed for the O6 hydroxyl in the complex. The sugar is stacked against Tyr241 side chain (neighboring monomer). The residues involved in the interaction with the O4 and O3 hydroxyls are strictly conserved among the members of the H-type lectins and explain the specificity for monosaccharides with *D-galacto* configuration.

Table 3. Contacts between DiscI and GalNAc in the crystal structure of the complex (with standard deviation between parentheses).

GalNAc atom	Protein atom	Distances (Å)
O1	Asn236 ND2	3.23 (0.16)
O7	Asn236 ND2	3.01 (0.09)
O3	Trp235 NE1	2.95 (0.08)
	Asp206 <sup>a</sup> OD2	2.56 (0.10)
O4	Asp206 <sup>a</sup> OD1	2.53 (0.01)
	Arg215 NE	2.79 (0.08)
	Arg215 NH2	3.17 (0.09)
O	Arg215 NH2	3.12 (0.13)
Van der Waals		
C4-C5	Tyr241 <sup>a</sup>	

Inspection of the different binding sites indicates that the anomeric preference of the H-type lectins seems to be dictated by the residue following the conserved Trp in the last surface loop. In HPA, His84 side chain would hinder the accommodation of the  $\beta$ -anomer justifying HPA specificity for  $\alpha$ -GalNAc<sup>22; 43</sup>. In DiscII, Gly239 does not permit any discrimination at the anomeric position and both anomers were found in the complex structure<sup>19</sup>. In DiscI, Asn236 is found interacting with the  $\beta$ -anomer of GalNAc explaining the better affinity and preference observed for this anomer. The N-acetyl group is stabilized through interactions with the extended first loop described as the specificity loop in DiscII

Different stacking interactions are observed with a tyrosine residue at the beginning of the last  $\beta$ -strand. Indeed Tyr89 stacks against the O6 hydroxymethyl whilst Tyr241 (discI) and Tyr 244 (DiscII) stacks against the C4 and C3 imposing their axial and equatorial orientations respectively. The side chain has rotated of about 145° around the C $\alpha$ -C $\beta$  bound

in HPA compared to the discoidins. An aromatic residue is always found in the invertebrates members of the H type lectin whereas it is an alanine in the bacterial member.

The largest difference in architecture of binding site is located in the first surface loop of the H-type lectin. It varies in sequence, length and structure from one member to the other and has a role in sugar specificity and affinity<sup>19</sup>. Each end of the loop participates to one side of the binding site but the C-terminal end is found in one binding site and the N-terminal one in the binding site of the neighboring monomer. The N-terminal end seems involved in the specificity and affinity for oligosaccharides since it is not involved in the binding of the monosaccharides. On the contrary the C-terminal end is implied in the specificity for the monosaccharides. It is extended in HPA and favors the binding of GalNAc to galactose since it stabilized the N-acetyl group via numerous interactions. In discoidins, it is short and does not interact with the monosaccharide.

### ***Molecular modelling of oligosaccharides in discoidins binding site***

The modelling study was performed in order to rationalize the ten fold increase in DiscI affinity for disaccharides Gal $\beta$ 1,3GalNAc (type 3/4) and GalNAc $\beta$ 1,3Gal, when compared to GalNAc monosaccharide. Modelling of the Gal $\beta$ 1,3GalNAc disaccharide in the binding site allows the positioning of the galactose residue in a pocket in between the  $\beta$ -strands of two monomers. This extension of the binding site contains several hydrophilic amino acids (Tyr166, Asp169 and Gln204) that establish hydrogen bonds with O6 of galactose and also an aromatic residue (Trp235 of neighbouring monomer) that makes van der Waals contacts. The modelled galactose is close to a MPD binding site observed in the crystal structure. Since MPD are good mimics of monosaccharides in binding to protein crystals, it is clear that the galactose could be extended on position 3 or 4 to longer oligosaccharides. This hydrogen bond network around the galactose, together with the shape of the binding sites, is due to the conformation of loop Gly167-Ala172. This loop is different in DiscII (figure 6C) explaining the lack of affinity of DiscII type 3/4 disaccharides. DiscII has no strong affinity for disaccharide (expected sulphated ones) but binds efficiently to  $\beta$ Gal1-4 $\beta$ GlcNAc1-6 $\beta$ GalNAc trisaccharide present on core 2 and core 4 mucins. Modeling of this trisaccharide demonstrates an extension towards the surface of the lectin trimer, in the opposite direction as the Gal $\beta$ 1,3GalNAc in DiscI. This area is rather hydrophilic and two hydrogen bonds from O6 of GlcNAc to Asn177 side chain and Ala242 carbonyl group are predicted. Additional

stabilizing interactions are established by van der Waals interaction between the non-reducing galactose of the trisaccharide and the aromatic group of His211. The absence of this residue in DiscI (Thr208) may explain the lack of affinity of this protein for 1-6 branched GalNAc.

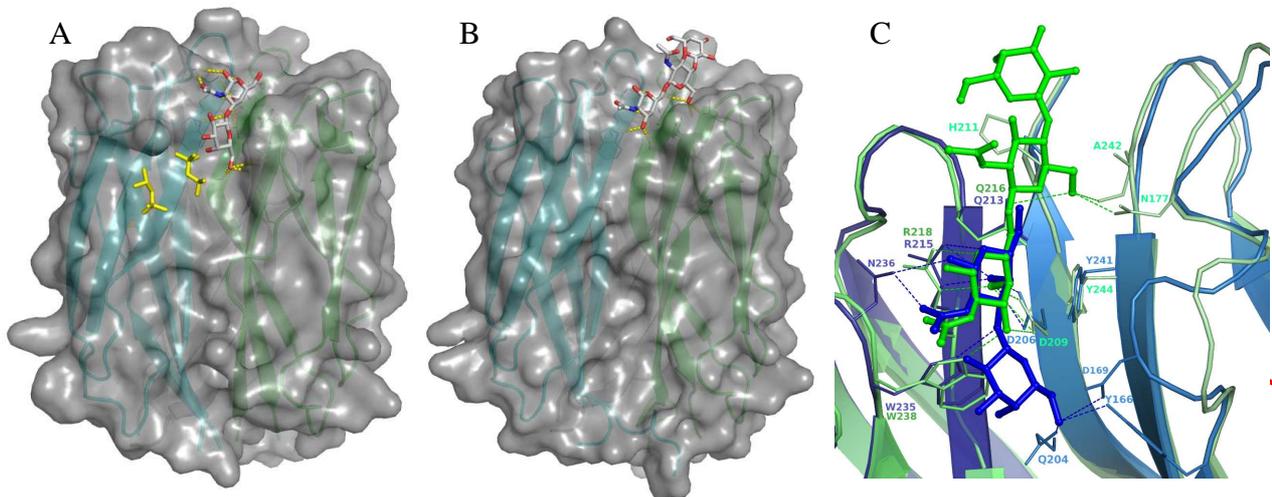


Figure 6. A and B. Models of lectin domain of DiscI interacting with  $\beta$ Gal1-3 $\beta$ GalNAc, and DiscII interacting with  $\beta$ Gal1-4 $\beta$ GlcNAc1-6 $\beta$ GalNAc. The two lectin chains forming the binding site are represented by ribbons colored in blue and green respectively. The oligosaccharides are represented by sticks and hydrogen bonds by dotted lines. MPD molecules observed in crystal structures have been added on the model. C. Superimposition of the DiscI/  $\beta$ Gal1-3 $\beta$ GalNAc and DiscII/  $\beta$ Gal1-4 $\beta$ GlcNAc1-6 $\beta$ GalNAc models colored in blue and green respectively. The variable loop responsible for the high affinity for  $\beta$ Gal/GalNAc1-3GalNAc linkages in DiscI is indicated by a red arrow.

## Conclusion

The present study highlighted the differences in specificity for the two discoidines, proteins that are spatially and temporally regulated during the social life of amoeba. The natural ligands of these proteins are not yet known but from our study, a panel of O-glycans and glycolipids would be recognized with high affinity. This opens a route for defining which possible ligands are related with functional role in the amoeba. The possibility of second carbohydrate binding site in the DS domain is now also opened, and possible ligands will be also tested.

## Acknowledgements

This work was supported by CNRS and French Ministry of Research. KSA is the recipient of a CNPq grant from Brazilian government. Glycan microarray analysis was provided by the Consortium for Functional Glycomics funded by the National Institute of General Medical Sciences grant GM62116. We thank the ESRF, Grenoble, for access to synchrotron data collection facilities.

## References

1. Lis, H. & Sharon, N. (1998). Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637-674.
2. Iwanaga, S. & Lee, B. L. (2005). Recent advances in the innate immunity of invertebrate animals. *J Biochem Mol Biol* **38**, 128-50.
3. Jimbo, M., Koike, K., Sakai, R., Muramoto, K. & Kamiya, H. (2005). Cloning and characterization of a lectin from the octocoral *Sinularia lochmodes*. *Biochem. Biophys. Res. Commun* **330**, 157-62.
4. 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.
5. Raper, K. B. (1935). *Dictyostelium discoideum*, a new species of slime mold from decaying forest leaves. *Journal of Agricultural Research* **50**, 135-147.
6. Kessin, R. H. (2001). *Dictyostelium - Evolution, Cell Biology, and the Development of Multicellularity*, Cambridge University Press, Cambridge, UK.
7. 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**, 2554-7.
8. 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**, 7714-21.
9. 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**, 291-6.
10. 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**, 273-89.
11. Tsang, A. S., Devine, J. M. & Williams, J. G. (1981). The multiple subunits of discoidin I are encoded by different genes. *Dev Biol* **84**, 212-7.
12. Welker, D. L. (1988). The Discoidin I Gene Family of *Dictyostelium discoideum* Is Linked to Genes Regulating Its Expression. *Genetics* **119**, 571-578.
13. Cooper, D. N., Lee, S. C. & Barondes, S. H. (1983). Discoidin-binding polysaccharide from *Dictyostelium discoideum*. *J. Biol. Chem.* **258**, 8745-50.

14. Alexander, S., Cibulsky, A. M. & Lerner, R. A. (1983). Ion dependence of the discoidin I lectin from *Dictyostelium discoideum*. *Differentiation* **24**, 209-12.
15. 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**, 557-64.
16. Cooper, D. N. & Barondes, S. H. (1984). Colocalization of discoidin-binding ligands with discoidin in developing *Dictyostelium discoideum*. *Dev. Biol.* **105**, 59-70.
17. Gabius, H. J., Springer, W. R. & Barondes, S. H. (1985). Receptor for the cell binding site of discoidin I. *Cell* **42**, 449-56.
18. Ruoslahti, E. (2003). The RGD story: a personal account. *Matrix Biol* **22**, 459-65.
19. Aragao, K. S., 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.
20. Kiedziarska, A., Smietana, K., Czepczynska, H. & Otlewski, J. (2007). Structural similarities and functional diversity of eukaryotic discoidin-like domains. *Biochim. Biophys. Acta* **1774**, 1069-78.
21. Vogel, W. F., Abdulhusein, R. & Ford, C. E. (2006). Sensing extracellular matrix: an update on discoidin domain receptor function. *Cell Signal* **18**, 1108-16.
22. Sanchez, J. F., Lescar, J., Chazalet, V., Audfray, A., Gagnon, J., Alvarez, R., Breton, C., Imberty, A. & Mitchell, E. P. (2006). Biochemical and structural analysis of *Helix pomatia* agglutinin (HPA): a hexameric lectin with a novel fold. *J. Biol. Chem.* **281**, 20171-20180.
23. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Jnt CCP4/ESF-EACMB. Newsletter on Protein Crystallography*, **26**.
24. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **D50**, 760-763.
25. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**, 458-64.
26. Brünger, A. T. (1992). Free R-Value - a Novel Statistical Quantity For Assessing the Accuracy of Crystal-Structures. *Nature* **355**, 472-475.
27. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-32.
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255.
29. 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.

30. Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr D Biol Crystallogr* **55**, 191-205.
31. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. T., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol* **112**, 535-542.
32. 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.
33. Dam, T. K. & Brewer, C. F. (2002). Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* **102**, 387-429.
34. Macedo-Ribeiro, S., Bode, W., Huber, R., Quinn-Allen, M. A., Kim, S. W., Ortel, T. L., Bourenkov, G. P., Bartunik, H. D., Stubbs, M. T., Kane, W. H. & Fuentes-Prior, P. (1999). Crystal structures of the membrane-binding C2 domain of human coagulation factor V. *Nature* **402**, 434-9.
35. 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**, 439-42.
36. Harding, M. M. (1999). The geometry of metal-ligand interactions relevant to proteins. *Acta Crystallogr D Biol Crystallogr* **55**, 1432-43.
37. 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**, 1626-31.
38. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-68.
39. Bianchet, M. A., Odom, E. W., Vasta, G. R. & Amzel, L. M. (2002). A novel fucose recognition fold involved in innate immunity. *Nature Struct. Biol.* **9**, 628-34.
40. Dyka, F. M., Wu, W. W., Pfeifer, T. A., Molday, L. L., Grigliatti, T. A. & Molday, R. S. (2008). Characterization and purification of the discoidin domain-containing protein retinoschisin and its interaction with galactose. *Biochemistry* **47**, 9098-106.
41. Wu, W. W., Wong, J. P., Kast, J. & Molday, R. S. (2005). RS1, a discoidin domain-containing retinal cell adhesion protein associated with X-linked retinoschisis, exists as a novel disulfide-linked octamer. *J Biol Chem* **280**, 10721-30.
42. Gouet, P., Courcelle, E., Stuart, D. I. & Metz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.
43. 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.

# **Conclusions générales et perspectives**



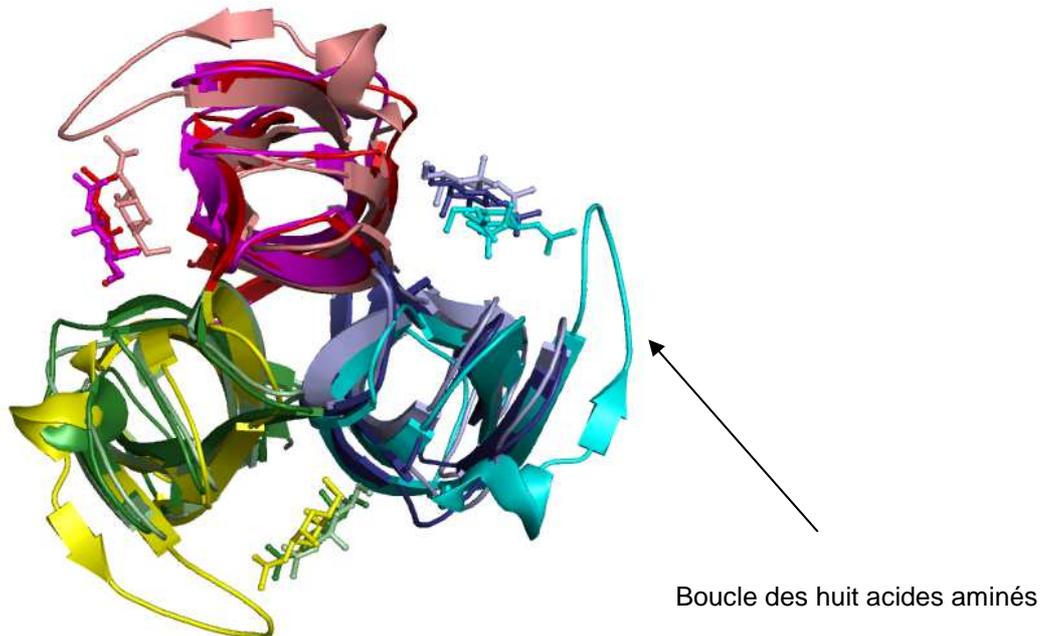
## 4 Conclusions générales et perspectives

Dans le cadre de ces travaux de recherche nous nous sommes intéressés à l'étude structurale et à la caractérisation biochimique des lectines Gal/GalNAc : les discoidines de *Dictyostelium discoideum* (DiscI et DiscII) impliquées dans le cycle de différenciation et de morphogénèse de l'amibe.

Les résultats ont exposés la mise au point des stratégies d'expression et de purification des protéines recombinantes (DiscIr et DiscIIr). On a pu obtenir ces deux protéines recombinantes sous forme suffisamment pure avec un rendement de 20mg de protéine par litre de culture, quantité convenable pour envisager des essais expérimentaux.

La purification de ces lectines a permis d'avoir la qualité suffisante pour les cristalliser et déterminer les structures 3-D avec de bonnes résolutions. Les structures de DiscI et DiscII ont été résolues dans leur forme non complexée (native) et en complexe avec des ligands spécifiques. La résolution de ces structures a donné la possibilité de produire les premières structures cristallographiques des lectines de l'amibe *Dictyostelium discoideum*. Ces protéines représentent une nouvelle famille structurale de protéines possédant un domaine DS (discoidine domain) en N-terminal et un domaine lectine en C-terminal, avec un arrangement en trimère.

Les structures 3D de la DiscI et de la DiscII révèlent les bases structurales de leur spécificité pour les ligands N-Acétylgalactosamine et galactose. Les structures des domaines lectines de ces deux protéines présentent des similarités avec la structure de la lectine HPA qui appartient à la famille de lectine de type H (Sanchez, *et al.* 2006). La résolution des structures 3D de la DiscI et de la DiscII a permis d'ajouter deux nouvelles structures de domaines lectines de la famille du type H. L'alignement des séquences de la lectine HPA avec le domaine C-terminal des discoidines montre que les résidus impliqués au niveau du site de liaison avec le sucre sont bien conservés pour les trois lectines. Cependant, la structure de la HPA présente une première boucle de huit acides aminés qui n'apparaît pas dans la structure du domaine C-terminal de la DiscI et de la DiscII. La séquence de cette boucle et sa taille semble jouer un rôle très important dans la spécificité de HPA pour reconnaître des épitopes tumoraux portant le GalNAc (Figure 32).



**Figure 32 – Superposition des lectines DiscI (vert sapin, rouge, bleu foncé), DiscII (vert clair, magenta, bleu clair) et HPA (jaune, saumon, cyan).**

Nous nous sommes inspirés de la comparaison entre ces lectines pour notre étude dont l'objectif majeur est de caractériser les interactions établies à l'interface entre les ligands et les protéines. La transformation et la croissance tumorale sont associées à des modifications de structure de résidus glycaniques membranaires appartenant aux glycoprotéines, aux glycolipides et aux protéoglycanes présents sur la surface cellulaire. Nos résultats constituent les bases moléculaires pour l'utilisation des DiscI et II dans la recherche de métastase comme marqueurs spécifiques possédant les mêmes propriétés que la lectine HPA, ceci afin de résoudre les problèmes constatés lors de l'étude de HPA.

Les perspectives à court terme sont de définir le rôle et les régulations des ces discoidines pour mieux comprendre le cycle multicellulaire de l'amibe *Dictyostelium discoideum*. A plus long terme, ces deux lectines produites sous forme recombinante pourront être testées pour leur capacité à reconnaître les glycannes portés par les cellules cancéreuses. Récemment, dans notre laboratoire, un projet de thèse a démarré afin d'étudier des mutations

des DiscI et DiscII recombinantes et leurs applications sur les cellules cancéreuses. Ces protéines sont en cour de caractérisation. Les informations devront aider à développer des outils de marquage qui permettront d'identifier rapidement le caractère métastatique de cellules cancéreuses.



# **Bibliographie**



---

## 5 Bibliographie

### A

- Alencar, N.M., Cavalcante, C.F., Vasconcelos, M.P., Leite, K.B., Aragao, K.S., Assreuy, A.M., Nogueira, N.A., Cavada, B.S. and Vale, M.R. (2005) Anti-inflammatory and antimicrobial effect of lectin from *Lonchocarpus sericeus* seeds in an experimental rat model of infectious peritonitis. *J Pharm Pharmacol*, 57, 919-922.
- Alexander, S., Cibulsky, A.M. and Lerner, R.A. (1983) Ion dependence of the discoidin I lectin from *Dictyostelium discoideum*. *Differentiation*, 24, 209-212.
- Alexander, S., Shinnick, T.M. and 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.
- Alexander, S., Sydow, L.M., Wessels, D. and Soll, D.R. (1992) Discoidin proteins of *Dictyostelium* are necessary for normal cytoskeletal organization and cellular morphology during aggregation. *Differentiation*, 51, 149-161.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, 215, 403-410.

### B

- Baker, D.A., Sugii, S., Kabat, E.A., Ratcliffe, R.M., Hermentin, P. and Lemieux, R.U. (1983) Immunochemical studies on the combining sites of Forssman hapten reactive hemagglutinins from *Dolichos biflorus*, *Helix pomatia*, and *Wistaria floribunda*. *Biochemistry*, 22, 2741-2750.
- Barondes, S.H., Cooper, D.N. and 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. and Cooper, D.N. (1985) Discoidin I, an endogenous lectin, is externalized from *Dictyostelium discoideum* in multilamellar bodies. *J Cell Biol*, 100, 1825-1833.
- Barre, A., Van Damme, E.J., Peumans, W.J. and Rouge, P. (1999) Homology modelling of the core domain of the endogenous lectin comitin: structural basis for its mannose-binding specificity. *Plant Mol Biol*, 39, 969-978.
- Baumgartner, S., Hofmann, K., Chiquet-Ehrismann, R. and Bucher, P. (1998) The discoidin domain family revisited: new members from prokaryotes and a homology-based fold prediction. *Protein Sci*, 7, 1626-1631.
- Beisel, H.G., Kawabata, S., Iwanaga, S., Huber, R. and Bode, W. (1999) Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate

- immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. *EMBO Journal*, 18, 2313-2322.
- Berger, E.A. and Armant, D.R. (1982) Discoidins I and II: common and unique regions on two lectins implicated in cell-cell cohesion in *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA*, 79, 2162-2166.
- Bergfors, T. (1999) *Protein Crystallization: Techniques, Strategies, and Tips. A Laboratory Manual*. International University Line, La Jolla USA.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.T., Jr, Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) The protein data bank: a computer-based archival file for macromolecular structures. *J Mol Biol*, 112, 535-542.
- Bianchet, M.A., Odom, E.W., Vasta, G.R. and Amzel, L.M. (2002) A novel fucose recognition fold involved in innate immunity. *Nature Struct. Biol.*, 9, 628-634.
- Blixt, O., et al. (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci U S A*, 101, 17033-17038.
- Blumenschein, T.M., et al. (2007) Atomic resolution insight into host cell recognition by *Toxoplasma gondii*. *EMBO J.*, 26, 2808-2820.
- Blusch, J., Alexander, S. and Nellen, W. (1995) Multiple signal transduction pathways regulate discoidin I gene expression in *Dictyostelium discoideum*. *Differentiation*, 58, 253-260.
- Boyd, W.C. and Shapleigh, E. (1954) Specific precipitation activity of plant agglutinins (lectins). *Science*, 119, 419.
- Brefeld, O. (1869) *Dictyostelium mucoroides*. Ein neuer organismus und der verwandschaft der Myxomyceten. *Abh. Seckenberg. Naturforsch. Ges.*, 7, 85-107.
- Breuer, W. and Siu, C.H. (1981) Identification of endogenous binding proteins for the lectin discoidin-I in *Dictyostelium discoideum*. *Proc Natl Acad Sci U S A*, 78, 2115-2119.
- Brooks, S.A. (2000) The involvement of *Helix pomatia* lectin (HPA) binding N-acetylgalactosamine glycans in cancer progression. *Histol Histopathol*, 15, 143-158.
- Brooks, S.A., Lymboura, M., Schumacher, U. and Leathem, A.J. (1996) Histochemistry to detect *Helix pomatia* lectin binding in breast cancer: methodology makes a difference. *J Histochem Cytochem*, 44, 519-524.

## C

- 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.
- Cano, A. and Pestana, A. (1984) The role of membrane lectins in *Dictyostelium discoideum* aggregation as ascertained by specific univalent antibodies against discoidin I. *J Cell Biochem*, 25, 31-43.

- Chisholm, R.L., Gaudet, P., Just, E.M., Pilcher, K.E., Fey, P., Merchant, S.N. and Kibbe, W.A. (2006) dictyBase, the model organism database for *Dictyostelium discoideum*. *Nucleic Acids Res*, 34, D423-427.
- Chrispeels, M.J. and Raikhel, N.V. (1991) Lectins, lectin genes, and their role in plant defense. *Plant Cell*, 3, 1-9.
- Clarke, M., Kayman, S.C. and Riley, K. (1987) Density-dependent induction of discoidin-I synthesis in exponentially growing cells of *Dictyostelium discoideum*. *Differentiation*, 34, 79-87.
- Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, D50, 760-763.
- Cooper, D.N. and 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. and Barondes, S.H. (1983) Discoidin-binding polysaccharide from *Dictyostelium discoideum*. *J. Biol. Chem.*, 258, 8745-8750.
- Crocker, P.R. (2002) Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. *Curr. Opin. Struct. Biol.*, 12, 609-615.
- Cummings, R.D. and Kornfeld, S. (1982) Characterization of the structural determinants required for the high affinity interaction of asparagine-linked oligosaccharides with immobilized *Phaseolus vulgaris* leucoagglutinating and erythroagglutinating lectins. *J Biol Chem*, 257, 11230-11234.
- Curat, C.A., Eck, M., Dervillez, X. and Vogel, W.F. (2001) Mapping of epitopes in discoidin domain receptor 1 critical for collagen binding. *J Biol Chem*, 276, 45952-45958.

## D

- Dam, T.K. and Brewer, C.F. (2002) Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.*, 102, 387-429.
- Delatorre, P., et al. (2006) Crystal structure of a lectin from *Canavalia maritima* (ConM) in complex with trehalose and maltose reveals relevant mutation in ConA-like lectins. *J Struct Biol*, 154, 280-286.
- Dennis, J.W., Granovsky, M. and Warren, C.E. (1999) Glycoprotein glycosylation and cancer progression. *Biochim. Biophys. Acta*, 1473, 21-34.
- Devine, J.M., Tsang, A.S. and Williams, J.G. (1982) Differential expression of the members of the discoidin I multigene family during growth and development of *Dictyostelium discoideum*. *Cell*, 28, 793-800.
- Dodson, K.W., Pinkner, J.S., Rose, T., Magnusson, G., Hultgren, S.J. and Waksman, G. (2001) Structural basis of the interaction of the pyelonephritic *E. coli* adhesin to its human kidney receptor. *Cell*, 105, 733-743.

- 
- Drickamer, K. (1993) Ca<sup>2+</sup>-dependent carbohydrate-recognition domains in animal proteins. *Curr. Opin. Struct. Biol.*, 3, 393-400.
- Ducruix, A. and Giégé, R. (1992) Crystallization of nucleic acids and proteins, a practical approach. The practicals approach series.

## E

- Edelman, G.M., Cunningham, B.A., Reeke, G.N., Becker, J.W., Waxdal, M.J. and Wang, J.L. (1972) The covalent and three-dimensional structure of concanavalin A. *Proc. Nat. Acad. Sci. USA*, 69, 2580-2584.
- Eichinger, L., et al. (2005) The genome of the social amoeba *Dictyostelium discoideum*. *Nature*, 435, 43-57.
- Emsley, J., White, H.E., O'Hara, B.P., Oliva, G., Srinivasan, N., Tickle, I.J., Blundell, T.L., Pepys, M.B. and Wood, S.P. (1994) Structure of pentameric human serum amyloid P component. *Nature*, 367, 338-345.
- Emsley, P., Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-32.
- Erdos, G.W. and Whitaker, D. (1983) Failure to detect immunocytochemically reactive endogenous lectin on the cell surface of *Dictyostelium discoideum*. *J Cell Biol*, 97, 993-1000.

## F

- Fenton-Navarro, B., Garcia-Hernandez, E., Heimer, E., Aguilar, M.B., Rodriguez, A.C., Arreguin-Espinosa, R. and Falcon, A. (2003) Purification and structural characterization of lectins from the cnidarian *Bunodeopsis antillensis* [corrected]. *Toxicon*, 42, 525-532.
- Firtel, R.A. and Meili, R. (2000) *Dictyostelium*: a model for regulated cell movement during morphogenesis. *Curr Opin Genet Dev*, 10, 421-427.
- Frazier, W.A., Rosen, S.D., Reitherman, R.W. and 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.
- Fuentes-Prior, P., Fujikawa, K. and Pratt, K.P. (2002) New insights into binding interfaces of coagulation factors V and VIII and their homologues lessons from high resolution crystal structures. *Curr Protein Pept Sci*, 3, 313-339.
- Fukuzawa, M. and Ochiai, H. (1993) Spatiotemporal patterning of discoidin I and II during development of *Dictyostelium discoideum*. *Dev. Growth Differ.*, 35, 11-23.

---

Fukuzawa, M. and Ochiai, H. (1996) Molecular cloning and characterization of the cDNA for discoidin II of *Dictyostelium discoideum*. *Plant Cell Physiol*, 37, 505-514.

## G

Gabius, H.J., Springer, W.R. and Baronides, S.H. (1985) Receptor for the cell binding site of discoidin I. *Cell*, 42, 449-456.

Garlatti, V., et al. (2007) Structural insights into the innate immune recognition specificities of L- and H-ficolins. *EMBO J*, 26, 623-633.

Garman, E. and Schneider, T. (1997) Macromolecular cryocrystallography. *J Appl Crystallogr*, 30, 211-237.

Geerlof, A., et al. (2006) The impact of protein characterization in structural proteomics. *Acta Crystallogr D Biol Crystallogr*, 62, 1125-1136.

Genbacev, O.D., et al. (2003) Trophoblast L-selectin-mediated adhesion at the maternal-fetal interface. *Science*, 299, 405-408.

Gerlach, D., Wagner, M., Schlott, B., Zahringer, U. and Schmidt, K.H. (2002) Chemical and physicochemical characterization of the sialic acid-specific lectin from *Cepaea hortensis*. *FEMS Microbiol Lett*, 214, 61-68.

Goldstein, I.J., Hughes, R.C., Monsigny, M., Osawa, T. and Sharon, N. (1980) What should be called a lectin? . *Nature*, 285, 66.

Gourdine, J.-P., Cioci, G., Miguët, L., Unverzagt, C., Varón-Silva, D., Varrot, A., Gautier, C., Smith-Ravin, E.J. and Imberty, A. (in press) High affinity interaction between bivalve C-type lectin and biantennary complex-type N-glycan revealed by crystallography and microcalorimetry *J. Biol. Chem.*

Guillot, J., Guerry, M., Kanska, G., Caldefie-Chezet, F., De Latour, M. and Penault-Llorca, F. (2004) Modification des glycoconjugués au cours du processus de cancérisation : cas des carcinomes mammaires. *Bull Cancer*, 91, 141-158.

## H

Hammarström, S. and Kabat, E.A. (1971) Studies on specificity and binding properties of the blood group A reactive hemagglutinin from *Helix pomatia*. *Biochemistry*, 10, 1684-1692.

Hardman, K.D. and Ainsworth, C.F. (1972) Structure of concanavalin A at 2.4 Å resolution. *Biochemistry*, 11, 4910-4919.

Hirabayashi, J. (2004) Lectin-based structural glycomics: glycoproteomics and glycan profiling. *Glycoconj. J.*, 21, 35-40.

## I

- Imberty, A., Mitchell, E.P. and Wimmerová, M. (2005) Structural basis for high affinity glycan recognition by bacterial and fungal lectins. *Curr. Opin. Struct. Biol.*, 15, 525-534.
- Imberty, A. and Varrot, A. (2008) Microbial recognition of human cell surface glycoconjugates. *Curr. Opin. Struct. Biol.*, 18, 567-576.
- Imberty, A., Wimmerova, M., Mitchell, E.P. and Gilboa-Garber, N. (2004) Structures of the lectins from *Pseudomonas aeruginosa*: Insights into molecular basis for host glycan recognition. *Microb. Infect.*, 6, 222-229.
- Ishiguro, A. and Weeks, G. (1978) Subunit composition and molecular weights of the developmentally regulated lectins from *Dictyostelium discoideum*. *J Biol Chem*, 253, 7585-7587.
- Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D. and Knowles, P.F. (1991) Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase. *Nature*, 350, 87-90.

## J

- Jain, R., Yuen, I.S., Taphouse, C.R. and Gomer, R.H. (1992) A density-sensing factor controls development in *Dictyostelium*. *Genes Dev*, 6, 390-400.
- Jancarik, J. and Kim, S.H. (1991) Sparse matrix sampling: screening method for crystallization of proteins. *J Appl Cryst*, 24, 409-411.
- Janin, J. and Delepierre, M. (1999) *Biologie structurale - Principes et méthodes biophysiques*. Hermann.
- 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, 157-162.
- Jimbo, M., Yanohara, T., Koike, K., Koike, K., Sakai, R., Muramoto, K. and Kamiya, H. (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, 227-236.
- Jung, E., Fucini, P., Stewart, M., Noegel, A.A. and Schleicher, M. (1996) Linking microfilaments to intracellular membranes: the actin-binding and vesicle-associated protein comitin exhibits a mannose-specific lectin activity. *EMBO J*, 15, 1238-1246.

## K

- Kairies, N., Beisel, H.G., Fuentes-Prior, P., Tsuda, R., Muta, T., Iwanaga, S., Bode, W., Huber, R. and Kawabata, S. (2001) The 2.0-Å crystal structure of tachylectin 5A provides evidence for the common origin of the innate immunity and the blood coagulation systems. *Proc Natl Acad Sci U S A*, 98, 13519-13524.
- Kessin, R.H. (2001) *Dictyostelium - Evolution, Cell Biology, and the Development of Multicellularity*. Cambridge University Press, Cambridge, UK.
- Kiedziarska, A., Smietana, K., Czepczynska, H. and Otlewski, J. (2007) Structural similarities and functional diversity of eukaryotic discoidin-like domains. *Biochim. Biophys. Acta*, 1774, 1069-1078.
- Koike, K., Jimbo, M., Sakai, R., Kaeriyama, M., Muramoto, K., Ogata, T., Maruyama, T. and Kamiya, H. (2004) Octocoral chemical signaling selects and controls dinoflagellate symbionts. *Biol Bull*, 207, 80-86.
- Konijn, T.M., Van De Meene, J.G., Bonner, J.T. and Barkley, D.S. (1967) The acrasin activity of adenosine-3',5'-cyclic phosphate. *Proc Natl Acad Sci U S A*, 58, 1152-1154.
- Kreppel, L., Fey, P., Gaudet, P., Just, E., Kibbe, W.A., Chisholm, R.L. and Kimmel, A.R. (2004) dictyBase: a new *Dictyostelium discoideum* genome database. *Nucleic Acids Res*, 32, D332-333.

## L

- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 26, 283-291.
- Lee, Y.C. and Lee, R.T. (1995) Carbohydrate-protein interactions: basis of glycobiology. *Acc. Chem. Res.*, 28, 321-327.
- Leffler, H., Carlsson, S., Hedlund, M., Qian, Y. and Poirier, F. (2004) Introduction to galectins. *Glycoconj. J.*, 19, 433-440.
- Leonidas, D.D., Vatzaki, E.H., Vorum, H., Celis, J.E., Madsen, P. and Acharya, K.R. (1998) Structural basis for the recognition of carbohydrates by human galectin-7. *Biochemistry*, 37, 13930-13940.
- Lescar, J., Sanchez, J.F., Audfray, A., Coll, J.L., Breton, C., Mitchell, E.P. and 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.
- Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Jnt CCP4/ESF-EACMB. Newsletter on Protein Crystallography*, 26.

- Lis, H. and Sharon, N. (1998) Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.*, 98, 637-674.
- Loris, R., Hamelryck, T., Bouckaert, J. and Wyns, L. (1998) Legume lectin structure. *Biochim. Biophys. Acta*, 1383, 9-36.

## M

- Ma, G.C. and Firtel, R.A. (1978) Regulation of the synthesis of two carbohydrate-binding proteins in *Dictyostelium discoideum*. *J Biol Chem*, 253, 3924-3932.
- Macedo-Ribeiro, S., et al. (1999) Crystal structures of the membrane-binding C2 domain of human coagulation factor V. *Nature*, 402, 434-439.
- Marcipar, I.S., Welchen, E., Roodveldt, C., Marcipar, A.J. and Silber, A.M. (2003) Purification of the 67-kDa lectin-like glycoprotein of *Trypanosoma cruzi*, LLGP-67, and its evaluation as a relevant antigen for the diagnosis of human infection. *FEMS Microbiol Lett*, 220, 149-154.
- McCoy, A.J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr*, 63, 32-41.
- McPherson, A. (1990) Current approaches to macromolecular crystallization. *Eur J Biochem*, 189, 1-23.
- Merritt, E.A., Sarfaty, S., Van Den Akker, F., L'hoir, C., Martial, J.A. and Hol, W.G.J. (1994) Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide. *Protein Sci.*, 3, 166-175.
- Mirelman, D., Moody-Haupt, S., Becker, S., Nuchamowitz, Y., Bracha, R. and Alon, R. (1996) Cell surface molecules of pathogenic and nonpathogenic *Entamoeba histolytica* and their relation to virulence. *Adv Exp Med Biol*, 408, 225-227.
- Mitchell, B.S. and Schumacher, U. (1999) The use of the lectin *Helix pomatia* agglutinin (HPA) as a prognostic indicator and as a tool in cancer research. *Histol Histopathol*, 14, 217-226.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, 53, 240-255.

## N

- Navaza, J. (1994) AMoRE - an Automated Package For Molecular Replacement. *Acta Crystallogr A*, 50, 157-163.

## P

- Park, S., Lee, M.R. and Shin, I. (2008) Chemical tools for functional studies of glycans. *Chem Soc Rev.*, 37, 1579-1591.
- Peles, E., Joho, K., Plowman, G.D. and Schlessinger, J. (1997) Close similarity between *Drosophila* neurexin IV and mammalian Caspr protein suggests a conserved mechanism for cellular interactions. *Cell*, 88, 745-746.
- Poget, S.F., Legge, G.B., Proctor, M.R., Butler, P.J., Bycroft, M. and Williams, R.L. (1999) The structure of a tunicate C-type lectin from *Polyandrocarpa misakiensis* complexed with D -galactose. *J. Mol. Biol.*, 290, 867-879.
- Poole, S., Firtel, R.A., Lamar, E. and Rowekamp, W. (1981) Sequence and expression of the discoidin I gene family in *Dictyostelium discoideum*. *J Mol Biol*, 153, 273-289.
- Prokop, O., Uhlenbruck, G. and Kohler, W. (1968) A new source of antibody-like substances having anti-blood group specificity. A discussion on the specificity of *Helix* agglutinins. *Vox Sang.*, 14, 321-333.

## R

- Raper, K.B. (1935) *Dictyostelium discoideum*, a new species of slime. mold from decaying forest leaves. *J. Agr. Res.*, 55, 135-147.
- Ray, J., Shinnick, T. and Lerner, R. (1979) A mutation altering the function of a carbohydrate binding protein blocks cell-cell cohesion in developing *Dictyostelium discoideum*. *Nature*, 279, 215-221.
- Reitherman, R.W., Rosen, S.D., Frasier, W.A. and Barondes, S.H. (1975) Cell surface species-specific high affinity receptors for discoidin: developmental regulation in *Dictyostelium discoideum*. *Proc Natl Acad Sci U S A*, 72, 3541-3545.
- Roberts, D.L., Weix, D.J., Dahms, N.M. and Kim, J.J. (1998) Molecular basis of lysosomal enzyme recognition: three-dimensional structure of the cation-dependent mannose 6-phosphate receptor. *Cell*, 93, 639-648.
- Rosen, S.D., Kafka, J.A., Simpson, D.L. and Barondes, S.H. (1973) Developmentally regulated, carbohydrate-binding protein in *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA*, 70, 2554-2557.
- Rossmann, M.G. and Blow, D.M. (1962) The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst*, 15, 24-31.
- Rowekamp, W., Poole, S. and Firtel, R.A. (1980) Analysis of the multigene family coding the developmentally regulated carbohydrate-binding protein discoidin-I in *D. discoideum*. *Cell*, 20, 495-505.

- 
- Rüdiger, H. (1993) Purification of plant lectins. In Gabius, H.J.G., S. (ed), Lectins and Glycobiology. Springer, Berlin, pp. 31-46.
- Rudiger, H. and Gabius, H.J. (2001) Plant lectins: occurrence, biochemistry, functions and applications. *Glycoconj J*, 18, 589-613.
- Rutenber, E., Katzin, B.J., Ernst, S., Collins, E.J., Mlsna, D., Ready, M.P. and J.D., R. (1991) Crystallographic refinement of ricin to 2.5 Å. *Proteins*, 10, 240-250.

## S

- Sanchez, J.F., Lescar, J., Chazalet, V., Audfray, A., Gagnon, J., Alvarez, R., Breton, C., Imberty, A. and Mitchell, E.P. (2006) Biochemical and structural analysis of *Helix pomatia* agglutinin (HPA): a hexameric lectin with a novel fold. *J. Biol. Chem.*, 281, 20171-20180.
- Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. and Vijayan, M. (1996) A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a  $\beta$ -prism fold. *Nature Struct. Biol.*, 3, 596-603.
- Schrag, J.D., Bergeron, J.J., Li, Y., Borisova, S., Hahn, M., Thomas, D.Y. and Cygler, M. (2001) The Structure of calnexin, an ER chaperone involved in quality control of protein folding. *Mol Cell*, 8, 633-644.
- Sharon, N. (1996) Carbohydrate-lectin interactions in infectious disease. *Adv. Exp. Med. Biol.*, 408, 1-8.
- Sharon, N. and Lis, H. (2004) History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*, 14, 53R-62R.
- Simpson, D.L., Rosen, S.D. and Barondes, S.H. (1974) Discoidin, a developmentally regulated carbohydrate-binding protein from *Dictyostelium discoideum*. Purification and characterization. *Biochemistry*, 13, 3487-3493.
- Sinha, S., Gupta, G., Vijayan, M. and Surolia, A. (2007) Subunit assembly of plant lectins. *Curr Opin Struct Biol*, 17, 498-505.
- Soker, S., Takashima, S., Miao, H.Q., Neufeld, G. and Klagsbrun, M. (1998) Neuropilin-1 is expressed by endothelial and tumor cells as an isoform-specific receptor for vascular endothelial growth factor. *Cell*, 92, 735-745.
- Somers, W.S., Tang, J., Shaw, G.D. and Camphausen, R.T. (2000) Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe<sup>x</sup> and PSGL-1. *Cell*, 103, 467-479.
- Springer, W.R., Cooper, D.N. and 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.
- Steinert, M. and Heuner, K. (2005) *Dictyostelium* as host model for pathogenesis. *Cell Microbiol*, 7, 307-314.

- 
- Stratford, C.A. and Brown, S.S. (1985) Isolation of an actin-binding protein from membranes of *Dictyostelium discoideum*. *J Cell Biol*, 100, 727-735.
- Strmecki, L., et al. (2007) Proteomic and Microarray Analyses of the Dictyostelium Zak1-GSK-3 Signaling Pathway Reveal a Role in Early Development. *Eukaryotic Cell*, 6, 245-252.
- Sugawara, H., Kusunoki, M., Kurisu, G., Fujimoto, T., Aoyagi, H. and Hatakeyama, T. (2004) Characteristic recognition of N-acetylgalactosamine by an invertebrate C-type lectin, CEL-I, revealed by X-ray crystallographic analysis. *J. Biol. Chem.*, 43, 45219-45225.
- Suzuki, R., Fujimoto, Z., Kuno, A., Hirabayashi, J., Kasai, K. and Hasegawa, T. (2004) Crystallization and preliminary X-ray crystallographic studies of the C-terminal domain of galactose-binding lectin EW29 from the earthworm *Lumbricus terrestris*. *Acta Crystallogr D Biol Crystallogr*, 60, 1895-1896.

## T

- Topfer-Petersen, E., Romero, A., Varela, P.F., Ekhlesi-Hundrieser, M., Dostalova, Z., Sanz, L. and Calvete, J.J. (1998) Spermadhesins: a new protein family. Facts, hypotheses and perspectives. *Andrologia*, 30, 217-224.
- Transue, T.R., Smith, A.K., Mo, H., Goldstein, I.J. and Saper, M.A. (1997) Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin. *Nat. Struct. Biol.*, 10, 779-783.
- Tsang, A.S., Devine, J.M. and Williams, J.G. (1981) The multiple subunits of discoidin I are encoded by different genes. *Dev Biol*, 84, 212-217.

## U

- Uchida, T., Yamasaki, T., Eto, S., Sugawara, H., Kurisu, G., Nakagawa, A., Kusunoki, M. and Hatakeyama, T. (2004) Crystal structure of the hemolytic lectin CEL-III isolated from the marine invertebrate *Cucumaria echinata*: implications of domain structure for its membrane pore-formation mechanism. *J. Biol. Chem.*, 279, 37133-37141.
- Umetsu, K., Yamashita, K. and Suzuki, T. (1991) Purification and carbohydrate-binding specificities of a blood type B binding lectin from hemolymph of a crab (*Charybdis japonica*). *J Biochem*, 109, 718-721.

## V

- Van Damme, E.J.M., Peumans, W.J., Barre, A. and Rougé, P. (1998) Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Critical Rev. Plant Sci.*, 17, 575-692.
- van Liempt, E., et al. (2006) Specificity of DC-SIGN for mannose- and fucose-containing glycans. *FEBS Lett*, 580, 6123-6131.
- Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, 3, 97-130.
- Vasta, G.R. (1992) Invertebrate lectins: Distribution, synthesis, molecular biology and function. In Allen, H.J. and Kisailus, E.C. (eds), *Glycoconjugates, Composition, Structure and Function*. Marcel Dekker, New York.
- Vasta, G.R., Ahmed, H. and Odom, E.W. (2004) Structural and functional diversity of lectin repertoires in invertebrates, protochordates and ectothermic vertebrates. *Curr Opin Struct Biol*, 14, 617-630.
- Vogel, W. (1999) Discoidin domain receptors: structural relations and functional implications. *FASEB J*, 13 S77-82.
- von Itzstein, M., Plebanski, M., Cooke, B.M. and Coppel, R.L. (2008) Hot, sweet and sticky: the glycobiology of *Plasmodium falciparum*. *Trends Parasitol*, 24, 210-218.
- Vriend, G. (1990) WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.*, 8, 52-56.

## W

- Weis, W.I., Brunger, A.T., Skehel, J.J. and Wiley, D.C. (1990) Refinement of the influenza virus hemagglutinin by simulated annealing. *J Mol Biol*, 212, 737-761.
- Willard, S.S. and Devreotes, P.N. (2006) Signaling pathways mediating chemotaxis in the social amoeba, *Dictyostelium discoideum*. *Eur J Cell Biol*, 85, 897-904.
- Wood-Charlson, E.M., Hollingsworth, L.L., Krupp, D.A. and Weis, V.M. (2006) Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. *Cell Microbiol*, 8, 1985-1993.
- Wright, C.S. (1980) Crystallographic elucidation of the saccharide binding mode in wheat germ agglutinin and its biological significance. *J Mol Biol*, 141, 267-291.
- Wright, C.S. and Hester, G. (1996) The 2.0 Å structure of a cross-linked complex between snowdrop lectin and a branched mannopentaose: evidence for two unique binding models. *Structure*, 4, 1339-1352.

Y

- Yamada, Y. and Aketa, K. (1982) Purification and partial characterization of hemagglutinins in seminal plasma of the sea urchin, *Hemicentrotus pulcherrimus*. *Biochim Biophys Acta*, 709, 220-226.
- Yu, X.Q. and Kanost, M.R. (2000) Immulectin-2, a lipopolysaccharide-specific lectin from an insect, *Manduca sexta*, is induced in response to gram-negative bacteria. *J Biol Chem*, 275, 37373-37381.



# **Annexe**



## 6 Annexe

Données obtenues par « Printed Glycan » array des DiscI et DiscII auprès du Consortium for Functional Glycomics. Seuls les pics nettement au dessus du bruit de fond parmi les 377 sucres sont listés.

Tableau 5 – Le tableau de spécificité de la DiscI 1µg/mL.

Chart	Oligosaccharide	RFU	STDEV
329	GalNAcb1-3Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	13339	1123
127	Galb1-3GalNAcb1-4Galb1-4Glc-Sp8	12451	557
125	Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0	11717	517
145	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	11459	619
144	Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	11448	863
146	Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	11092	886
201	NeuAca2-8NeuAca2-8NeuAca2-8NeuAca2-3(GalNAcb1-4)Galb1-4Glc-Sp0	10514	312
328	Gala1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	10263	104
147	Galb1-4GlcNAcb1-3Galb1-4Glc-Sp8	10005	523
159	GlcNAcb1-3GalNAca-Sp8	9942	1130
204	Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4Glc-Sp0	9815	641
333	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	9240	643

<b>123</b>	Galb1-3GalNAca-Sp8	9167	1603
<b>164</b>	GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	8945	1239
<b>276</b>	Galb1-3GalNAc-Sp14	8819	717
<b>281</b>	Galb1-4GlcNAcb1-3Galb1-3GlcNAcb-Sp0	8768	516
<b>358</b>	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3[Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20	8489	707
<b>124</b>	Galb1-3GalNAcb-Sp8	8409	958
<b>161</b>	GlcNAcb1-3Galb1-3GalNAca-Sp8	8187	491
<b>207</b>	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp0	7844	495
<b>336</b>	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4Glc-Sp0	7768	101
<b>331</b>	GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	7544	414
<b>322</b>	Galb1-4GlcNAcb1-3GalNAc-Sp14	7473	471
<b>126</b>	Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4Glc-Sp0	7444	4499
<b>327</b>	NeuAca2-6Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	7376	327
<b>221</b>	NeuAca2-3Galb1-3GalNAcb1-3Gala1-4Galb1-4Glc-Sp0	7340	575
<b>31</b>	[3OSO3]Galb1-3GalNAca-Sp8	7171	475
<b>163</b>	GlcNAcb1-3Galb1-4GlcNAcb-Sp8	6833	620
<b>332</b>	(Neu5Aca2-3-Galb1-3)((Neu5Aca2-3-Galb1-4(Fuca1-3))GlcNAcb1-6)GalNAc-Sp14	6827	419
<b>291</b>	Galb1-3(Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-	6721	53

	6)GalNAc–Sp14		
<b>130</b>	Galb1-3GlcNAcb1-3Galb1-4Glc–Sp10	6587	310
<b>162</b>	GlcNAcb1-3Galb1-4GlcNAcb–Sp0	6548	219
<b>374</b>	NeuAca2-3[Fuca1-3]Galb1-4GlcNAcb1-3GalNAc-Sp14	6419	1801
<b>208</b>	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcNAcb-Sp8	6305	4222
<b>86</b>	GalNAcb1-3GalNAca-Sp8	6018	252
<b>200</b>	Neu5Aca2-3Galb1-3GalNAca-Sp8	5982	191
<b>338</b>	GlcNAca1-4Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	5815	2575
<b>279</b>	Galb1-4[Fuca1-3][6OSO3]Glc-Sp0	5646	539
<b>277</b>	Galb1-3GlcNAcb1-3Galb1-3GlcNAcb-Sp0	5110	1086
<b>222</b>	NeuAca2-3Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	5089	2261
<b>368</b>	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3[Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAcb-Sp20	4822	442
<b>263</b>	[3OSO3]Galb1-4(Fuca1-3)(6OSO3)Glc-Sp0	4651	141
<b>315</b>	Neu5Aca2-3Galb1-3GalNAc–Sp14	4507	2958
<b>373</b>	NeuAca2-6Galb1-4GlcNAcb1-3GalNAc-Sp14	3935	323
<b>142</b>	Galb1-4GlcNAcb1-3GalNAca-Sp8	3779	746
<b>231</b>	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAcb-Sp8	3700	341
<b>233</b>	Neu5Aca2-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb–Sp0	3541	326
<b>300</b>	GalNAcb1-3Galb-Sp8	3517	3255

<b>156</b>	GlcNAcb1-2Galb1-3GalNAca-Sp8	3205	243
<b>58</b>	Fuca1-2Galb1-3GalNAcb1-4(Neu5Aca2-3)Galb1-4GlcB-Sp0	2946	1591
<b>209</b>	Neu5Aca2-3(GalNAcb1-4)Galb1-4GlcB-Sp0	2793	4575
<b>88</b>	GalNAcb1-3Gala1-4Galb1-4GlcNAcb-Sp0	2157	2884
<b>352</b>	KDNa2-3Galb1-4(Fuca1-3)GlcNAc-Sp0	1998	1223
<b>87</b>	GalNAcb1-3(Fuca1-2)Galb-Sp8	1946	1283
<b>364</b>	Fuca1-2Galb1-3GlcNAcb1-3[Galb1-4(Fuca1-3)GlcNAcb1-6]Galb1-4Glc-Sp21	1430	870
<b>210</b>	NeuAca2-3(NeuAca2-3Galb1-3GalNAcb1-4)Galb1-4GlcB-Sp0	1387	2386
<b>265</b>	[3OSO3]Galb1-4[Fuca1-3][6OSO3]GlcNAc-Sp8	1327	1260
<b>94</b>	Gala1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	1323	147
<b>129</b>	Galb1-3GlcNAcb1-3Galb1-4GlcNAcb-Sp0	1067	1672
<b>320</b>	Fuca1-2Galb1-3GalNAca-Sp14	1065	569
<b>351</b>	(6SO3)GlcNAcb1-3Gal b1-4GlcNAc-b-Sp0	1027	586

Tableau 6 – Le tableau de spécificité de la DiscII 1µg/mL.

Chart	Oligosaccharide	RFU	STDEV
374	NeuAca2-3[Fuca1-3]Galb1-4GlcNAcb1-3GalNAc-Sp14	14603	2247
358	Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3[Fuca1-2Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAc-Sp20	11167	552
332	(Neu5Aca2-3-Galb1-3)((Neu5Aca2-3-Galb1-4(Fuca1-3))GlcNAcb1-6)GalNAc-Sp14	10033	511
368	Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-3[Gala1-3Galb1-4(Fuca1-3)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAc-Sp20	8764	307
149	Galb1-4GlcNAcb1-6GalNAca-Sp8	8372	413
294	Galb1-4GlcNAcb1-3(Galb1-4GlcNAcb1-6)Galb1-4GlcNAc-Sp0	7310	667
291	Galb1-3(Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-6)GalNAc-Sp14	7053	1072
296	Galb1-4GlcNAca1-6Galb1-4GlcNAcb-Sp0	5908	445
371	Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-3[Fuca1-2Galb1-3(Fuca1-4)GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAc-Sp19	5401	649
365	Galb1-4GlcNAcb1-2(Galb1-4GlcNAcb1-4)Mana1-3[Galb1-4GlcNAcb1-2Mana1-6]Manb1-4GlcNAcb1-4GlcNAc-Sp21	4708	3119
231	Neu5Aca2-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4GlcNAc-Sp8	4668	766
286	[6OSO3]Galb1-4[6OSO3]GlcNAcb-Sp0	4123	210
139	Galb1-4GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	2943	2363

<b>155</b>	GlcNAc1-6Galb1-4GlcNAcb-Sp8	2892	274
<b>23</b>	(Galb1-4GlcNAcb)2-3,6-GalNAc-Sp8	2694	231
<b>44</b>	[6OSO3]Galb1-4[6OSO3]Glc-Sp8	2662	93
<b>263</b>	[3OSO3]Galb1-4(Fuca1-3)(6OSO3)Glc-Sp0	2348	240
<b>201</b>	NeuAca2-8NeuAca2-8NeuAca2-8NeuAca2-3(GalNAcb1-4)Galb1-4Glc-Sp0	2333	769
<b>204</b>	Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4Glc-Sp0	1492	994
<b>140</b>	Galb1-4GalNAcb1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	1477	1304
<b>51</b>	Galb1-4GlcNAcb1-2Mana1-3(Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp13	1416	1388
<b>279</b>	Galb1-4[Fuca1-3][6OSO3]Glc-Sp0	855	820
<b>202</b>	Neu5Aca2-8Neu5Aca2-8Neu5Aca2-3(GalNAcb1-4)Galb1-4Glc-Sp0	843	899
<b>351</b>	(6SO3)GlcNAcb1-3Gal b1-4GlcNAc-b-Sp0	836	100
<b>143</b>	Galb1-4GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb1-3Galb1-4(Fuca1-3)GlcNAcb-Sp0	818	385