



**HAL**  
open science

# La lectine PA-IIL de *Pseudomonas aeruginosa*: Structure, affinité et spécificité pour des ligands naturels et glycomimétiques

Charles Sabin

► **To cite this version:**

Charles Sabin. La lectine PA-IIL de *Pseudomonas aeruginosa*: Structure, affinité et spécificité pour des ligands naturels et glycomimétiques. Médicaments. Université Joseph-Fourier - Grenoble I, 2006. Français. NNT: . tel-00144113v1

**HAL Id: tel-00144113**

**<https://theses.hal.science/tel-00144113v1>**

Submitted on 30 Apr 2007 (v1), last revised 30 Apr 2007 (v2)

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

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

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

**THESE**

Pour l'obtention du Diplôme de  
DOCTEUR DE L'UNIVERSITE JOSEPH FOURIER

**Discipline : Biologie structurale et fonctionnelle**

présentée et soutenue publiquement par

**Charles SABIN**

Le 22 novembre 2006

**La lectine PA-III de *Pseudomonas aeruginosa* :  
Structure, affinité et spécificité pour des ligands  
naturels et glycomimétiques**

Directeurs de thèse : Dr. Anne Imberty et Dr. Edward Mitchell

**JURY**

Dr. Solange Morera, Rapporteur  
Dr. Jean-Claude Michalski, Rapporteur  
Dr. Ina Attrée, Examineur  
Dr. Michaela Wimmerova, Examineur  
Dr. Sophie de Bentzmann, Examineur  
Dr. Anne Imberty, Directeur de thèse  
Dr. Edward Mitchell, Directeur de thèse



Cette thèse de recherche a été entièrement financée par  
l'association

« Vaincre La Mucoviscidose »



Vaincre la Mucoviscidose  
181, rue de Tolbiac  
75013 Paris  
Tél : 01 40 78 91 91  
Fax : 01 45 80 86 44



*Remerciements,*

*...alors que je rassemble les dernières affaires sur le bureau que j'ai occupé durant plus de 3 ans au CERMAV, je réalise la chance que j'ai eue d'avoir travaillé dans l'équipe de recherche du Dr. Anne Imberty. Au delà du talent scientifique, ses qualités d'encadrement et ses qualités humaines en font un directeur de thèse tout à fait remarquable! Pour le jeune chercheur que j'étais, ce fut un exemple stimulant, comme le fut tout autant mon co-directeur de thèse, le Dr. Edward Mitchell de l'ESRF, qui fut un formateur véritablement excellent, toujours très disponible et avenant, à qui je n'ai eu de cesse de voler un peu de science autant que de cet esprit anglo-saxon. Leur confiance m'a fortement encouragé.*

*En vidant ma paillasse au labo (finalement seul moyen pour qu'elle soit enfin propre et rangée!), j'imagine difficilement un monde sans Catherine Gautier et Valérie Chazalet, qui par leur support logistique, leurs conseils scientifiques et leur convivialité, m'ont été quotidiennement indispensables et réconfortantes. (Au deuxième étage du labo, « j'ai besoin d'aide » se dit : Catherinnnnne !!! Valouuuuu !!!)*

*Je compte également, parmi mes meilleurs souvenirs, ma rencontre et collaboration avec le Dr. Michaëla Wimmerova et le Prof. Jaroslav Koca qui en m'accueillant gracieusement au sein de leur labo (le NCBR), m'ont permis, de connaître le pays et la culture tchèque, ce qui fut un plaisir et un enrichissement personnel.*

*La Recherche, en effet, est un travail d'équipe et une grande part de mes résultats a pu voir le jour grâce au travail colossal de synthèse chimique réalisé par l'équipe de recherche du Prof. René Roy (UQAM) et ses étudiants : Cathy Préville, Sarah Beha, Karine Marotte et les autres. Je tiens aussi à remercier le Dr. Ina Atrée et toute l'équipe du BBSI/CEA, ainsi que le Dr. Sophie de Bentzmann (CNRS-IBSM) pour m'avoir initié et donné goût à la microbiologie. J'ai toujours pu trouver conseil et aide pour les grandes questions comme les petits doutes auprès du Dr. Annabelle Varrot, indulgente même si j'envahissais sa paillasse régulièrement, ainsi qu'auprès du Prof. Christelle Breton qui, de l'avis de tous, s'est toujours montrée si bon pédagogue que je regrette de ne l'avoir jamais eue comme professeur à l'université.*

*...alors que j'archive 3 ans de données sur DVD, je sais combien je dois aux informaticiens Alain Rivet et Jean-Daniel Dubois, pour m'avoir plus d'une fois sauvé la mise quand l'ordinateur me jouait de sales tours. Quant à mes déboires administratifs... urgents! comme d'habitude!, Isabelle Caldara a toujours su y pallier avec sourire! Les soirées, les fêtes, les barbecues, paellas, repas de Noël, les laser-games et les sorties de ski etc, ont aussi rythmé ce doctorat et je compte aujourd'hui de nombreux collègues et amis parmi tout le personnel du CERMAV, de l'ESRF, et les thésards de ma promotion qu'il m'est difficile de quitter (Martina, Lenka, Jean-Fred, Nikola, Honza, Lina, Gimmy, Émilie, Karoline, Magali, Gianluca, Shirin, Anna, Peter, Caroline, Roberto, Marie-Pierre, Sara... le manque de place m'empêche de les citer tous hélas!).*

*Enfin! enfin..., et d'abord et je dirai même surtout! Je suis reconnaissant au Dr. Serge Perez, directeur du CERMAV, de m'avoir accueilli lorsque j'étais « tout petit » selon ses mots, dans un laboratoire d'excellence, lors d'un stage de licence, pour travailler sur une base de données, puis en DEA chimie-physique me permettant ainsi avec le concours de Anne Imberty de poursuivre en thèse.*

*Je remercie également l'association caritative « Vaincre LA Mucoviscidose » d'avoir intégralement financé et soutenu ce projet de thèse. Qu'elle sache que je me sentirai toujours concerné.*



# Abréviations

**ADP, ATP** : adénosine diphosphate, adénosine triphosphate  
**AHL** : acylhomosérine lactone  
**Ara** : D-arabinose  
**BSA** : bovin sérum albumine  
**CCD** : charge-couple devices  
**CF, CFTR** : cystic fibrosis, cystic fibrosis transmembrane conductance regulator  
**CHO** : chinese hamster ovary  
**CV-III** : *Chromobacterium violaceum* lectin II  
**CUP** : chaperone/usher pathways  
**DDM** : dynamic density modification  
**DMF** : diméthyle formamide  
**E. coli** : *Escherichia coli*  
**ELISA, ELLA** : enzyme linked immunosorbent assay, enzyme linked lectin assay  
**EPS** : exopolysaccharides  
**Fru** : D-fructose  
**Fuc** : L-fucose  
**Gal, GalNAc** : D-galactose, N-acethylgalactosamine  
**Glc, GLCNAC** : D-glucose, N-acétylglucosamine  
**HIV** : human immunodeficiency virus  
**HMOs** : human milk oligosaccharides  
**IC50** : concentration en inhibiteur donnant 50% d'inhibition  
**ITC** : isothermal titration calorimeter  
**Le<sup>a</sup>, Le<sup>b</sup>, Le<sup>x</sup>, Le<sup>a</sup>** : trisaccharide Lewis a, Lewis b, Lewis x, Lewis y  
**LNn,DF-II, LNnFP-V** : lacto-*N*-neodifucohexaose II, lacto-*N*-neofucopentaose V  
**LPS** : lipopolysaccharides  
**MAD** : multiwavelength anomalous dispersion  
**Man** : D-mannose  
**MIR** : multiple isomorphous replacement  
**Me-** : methyl-  
**NBD1** : nucleotide binding domains I  
**NeuAc** : acide sialique  
**NHS** : N-hydroxysuccimidine  
**PAA** : poly[N-(2-hydroxyethyl)acrylamide]  
**PA, PA-IL et PA-IIL** : *Pseudomonas aeruginosa*, *Pseudomonas aeruginosa* lectin I et II  
**PBS** : phosphate buffered saline  
**PEG** : poly(éthylène glycol)  
**PMNs** : poly-morphonuclear neutrophils  
**QS** : quorum sensing  
**RS-L** : *Ralstonia solanacearum* lectin  
**SAD** : single anomalous dispersion  
**SSTT** : système de sécrétion de type III  
**SIR** : single isomorphous replacement  
**TFP** : tetrafluorophenyle  
**UDP** : uridine diphosphate  
**VAP** : ventilator-associated pneumonia





# Sommaire

1. Mucoviscidose et infections bactériennes.....	9
1.1.La maladie du souffle	9
1.2.La bactérie <i>Pseudomonas aeruginosa</i>	14
1.3.Les sucres : éléments stratégiques de l'infection	20
1.4.Pili et flagelle de <i>P. aeruginosa</i>	25
1.5.Lectines solubles de <i>P. Aeruginosa</i>	30
2. Les outils de l'étude de l'interaction lectine-sucré.....	37
2.1.La cristallographie aux rayons X	38
2.2.La microcalorimétrie de titration isotherme (ITC)	49
2.3.Les tests d'interactions	56
3. PA-III: interaction avec des monosaccharides.....	59
3.1.Les lectines bactériennes solubles	59
3.2.Structure de la lectine PA-III	60
3.3.Bases moléculaires et thermodynamiques de la haute affinité (Article n°1)	63
3.4.Spécificité pour d'autres monosaccharides (Article n°2)	63
4. Mutations et spécificité de la lectine PA-III.....	71
4.1.Recherche de séquences similaires chez d'autres bactéries	71
4.2.Productions de mutants de PA-III et données thermodynamiques	73
4.3.Structures cristallines	75
4.4.Discussion et conclusion	78
5. Interaction entre PA-III et les oligosaccharides naturels.....	83
5.1.Recherche de ligands oligosaccharidiques par méthode de criblage	83
5.2.Oligosaccharides fucosylés du lait maternel.	85
5.3.Les applications thérapeutiques des oligosaccharides du lait humain	88
5.4.Interaction entre PA-III et les oligosaccharides du lait humain (Article n° 3)	88
6. Recherche de candidats à l'inhibition de PA-III: les glycomimétiques...	93
6.1.Les médicaments glycomimétiques	93
6.2.Interactions entre PA-III et des séries de glycomimétiques: Approche expérimentale	98
6.3.Interactions entre PA-III et des séries de glycomimétiques: Comparaison des glycomimétiques et affinité	100
6.4.Discussion et conclusion	110
7. Conclusion et perspectives.....	113
7.1.Lectine PA-III et infections bactériennes	113
7.2.Du ligand naturel au médicament	114
Bibliographie et annexes.....	117



# 1

## Mucoviscidose et infections bactériennes

---

### 1.1. La maladie du souffle

#### *1.1.1. Présentation*

La mucoviscidose appelée aussi « fibrose kystique du pancréas », est la maladie génétique héréditaire la plus fréquemment rencontrée dans la population caucasienne. Il s'agit d'une exocrinopathie généralisée et polymorphe qui touche les glandes de sécrétion muqueuse et séreuse et conduit à la production de sécrétions visqueuses insuffisamment hydratées (d'où le nom donné à la maladie, « mucus visqueux »). La maladie touche principalement l'appareil respiratoire, le tube digestif, le pancréas, le foie, les voies biliaires mais aussi les sécrétions sudorales et l'appareil génital.

#### *1.1.2. Génétique*

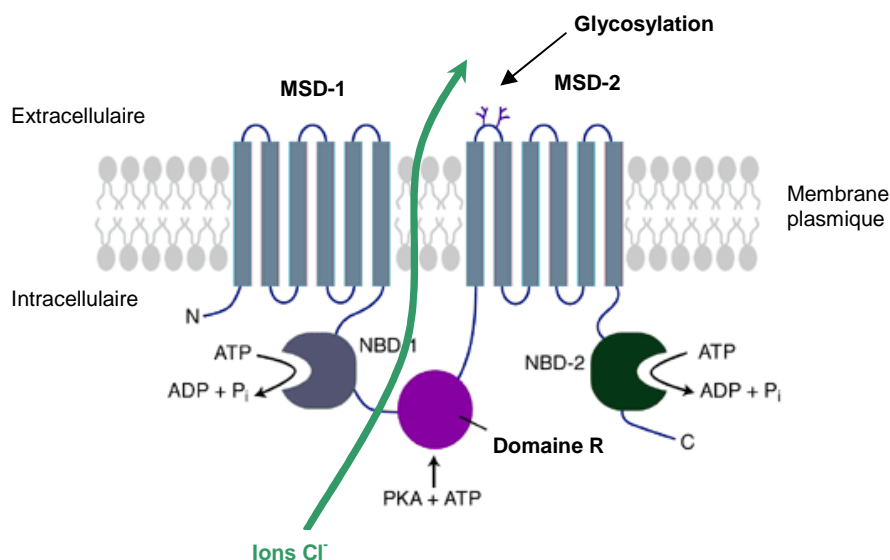
La mucoviscidose est une maladie génétique qui se transmet sur le *mode autosomique récessif*, ce qui signifie que seuls les individus homozygotes (porteurs des deux allèles parentales déficients) sont malades. En Europe, un nouveau-né sur 2000 est atteint, et une personne sur 20 est porteuse du gène responsable de la maladie, soit environ 4 % de la population générale.

La protéine responsable de la maladie est la protéine CFTR (cystic fibrosis transmembrane conductance regulator) composée de 1480 acides aminés dont le gène de 250 kilo-bases appartient au chromosome 7. La protéine est localisée sur les membranes des cellules épithéliales des voies aériennes, intestinales, reproductrices, ainsi que sur les glandes de sécrétion externe telles que les glandes sudorales, pancréatiques et salivaires. Cette protéine est un canal de régulation du passage des ions chlorure à travers la membrane des cellules épithéliales (Quinton, 1983), mais elle agit également comme un régulateur de conductibilité et exerce une influence modulatrice sur de nombreux autres canaux ioniques et protéines de transport tels que le canal épithélial de  $\text{Na}^+$  (Briel *et al.*, 1998), les canaux à eau (Schreiber *et al.*, 1999) et les mécanismes d'exportation de l'ATP (Sugita *et al.*, 1998).

La protéine CFTR appartient à la superfamille des transporteurs qui utilisent l'énergie de l'ATP pour permettre le transport d'une grande variété de substrats à travers les membranes cellulaires (fig. 1.1). Elle est composée de deux motifs répétés, chacun constitué de domaines hydrophobes (MSD, membrane-spanning domains) contenant six hélices transmembranaires et une région hydrophile cytoplasmique qui constitue le domaine de liaison à l'ATP (Riordan *et al.*, 1989). Ces deux motifs sont liés par un domaine cytoplasmique de régulation (R), contenant des résidus chargés et de multiples sites de phosphorylation. La structure aux rayons X de la protéine CFTR humaine reste à déterminer. Néanmoins, la structure cristalline du domaine de liaison à l'ATP de la souris, NBD1, a été résolue. Elle partage 78% de similarité de séquence avec la protéine humaine (Lewis *et al.*, 2004). Plus récemment, des données obtenues par microscopie électronique (résolution de 2 nm) sur la protéine CFTR recombinante humaine intégrale, montrent une forte homologie structurale avec un autre transporteur ABC de cellules eucaryotes, la P-glycoprotéine. Les deux conformations observées suggèrent un état ouvert et un état fermé du canal à chlorure (Rosenberg *et al.*, 2004).

Il existe plus de 1400 mutations affectant le gène CFTR. La mutation la plus répandue est la mutation  $\Delta\text{F508}$  (délétion de la phénylalanine en position 508) et représente environ 70 % des mutations recensées (Kerem *et al.*, 1989). On estime qu'approximativement la moitié des patients atteints de mucoviscidose (CF) sont homozygotes pour la mutation  $\Delta\text{F508}$  associée à une forme grave de la maladie et que plus de 90% de patients CF ont au moins un allèle  $\Delta\text{F508}$ . La mutation conduit à un repliement imparfait de la protéine CFTR ayant pour conséquence de ne plus permettre son insertion correcte dans la membrane plasmique. La protéine mutante maintenue dans le compartiment endoplasmique est alors rapidement dégradée (Ward *et al.*, 1995).

La protéine CFTR peut exister sous trois formes différentes dont les poids moléculaires respectifs de 127, 131, et 160 kDa correspondent à des glycosylations plus ou moins complètes (Cheng *et al.*, 1990). La mutation  $\Delta F508$  donne lieu à une forme de CFTR qui est incomplètement glycosylée. La caractérisation structurale de la glycosylation de CFTR n'a commencé à être étudiée que récemment (O'Riordan *et al.*, 2000). Des polylactosaminoglycanes ont été identifiés sur la forme entièrement glycosylée de CFTR recombinante produite en système CHO (ovaire de hamster chinois), alors que les CFTR partiellement glycosylées issues de cellules d'insecte sf19 présentent des oligomannoses portant une fucosylation. Ces résultats sont controversés quant à l'extrapolation à la forme humaine et des études restent à faire à partir de CFTR endogènes humaines pour comprendre la signification de la glycosylation (Scanlin and Glick, 2001).



**Figure 1.1.** Représentation schématique de la protéine CFTR (*Cystic Fibrosis Transmembrane Regulator*). Schéma adapté de (Davidson and Dorin, 2001).

### 1.1.3. Les symptômes

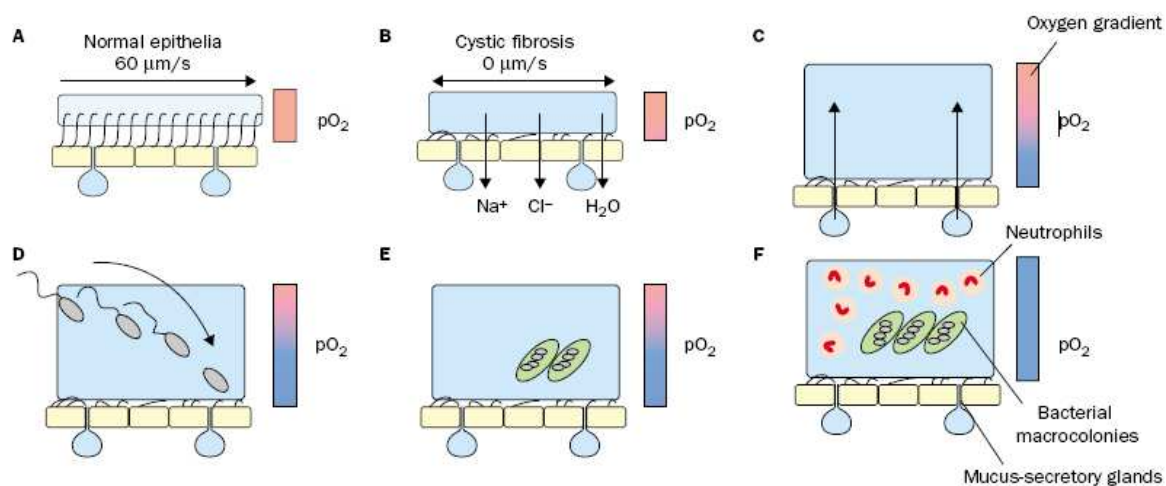
L'absence ou le dysfonctionnement du canal à chlorure CFTR engendre des dérèglements dans les sécrétions au niveau des surfaces épithéliales (Zielenski and Tsui, 1995). Chez environ 90 % des patients, l'obstruction des canaux pancréatiques aboutit à un déficit sécrétoire pancréatique

exocrine responsable d'une fibrose progressive du pancréas (d'où le nom de fibrose kystique du pancréas donné par les anglo-saxons). L'obstruction des canaux biliaires perturbe le fonctionnement du foie et de la digestion. On observe une stérilité chez 95% des hommes due à l'atrésie des canaux déférents (Schulz *et al.*, 2006).

Dans le poumon, le défaut dans le transport des ions chlorure est couplé à une hyper-absorbition du sodium, ainsi qu'à la déshydratation du mucus. L'épaississement du mucus favorise les infections bactériennes. La réaction inflammatoire conduit progressivement à la destruction des parois bronchiques et à une insuffisance respiratoire chronique grave pouvant prendre différentes formes: toux chronique, bronchites infectieuses ou asthmatiformes et récidivantes causant des emphysèmes. Le décès survient suite à une exacerbation de ces signes respiratoires et d'une insuffisance cardiaque dans environ 1/3 des cas. Les manifestations pulmonaires sont responsables de plus de 90% de mortalité. Actuellement, la transplantation de poumon est la seule thérapie efficace. L'espérance de vie atteint aujourd'hui les 38 ans.

#### *1.1.4. Les infections bactériennes pulmonaires et la mucoviscidose*

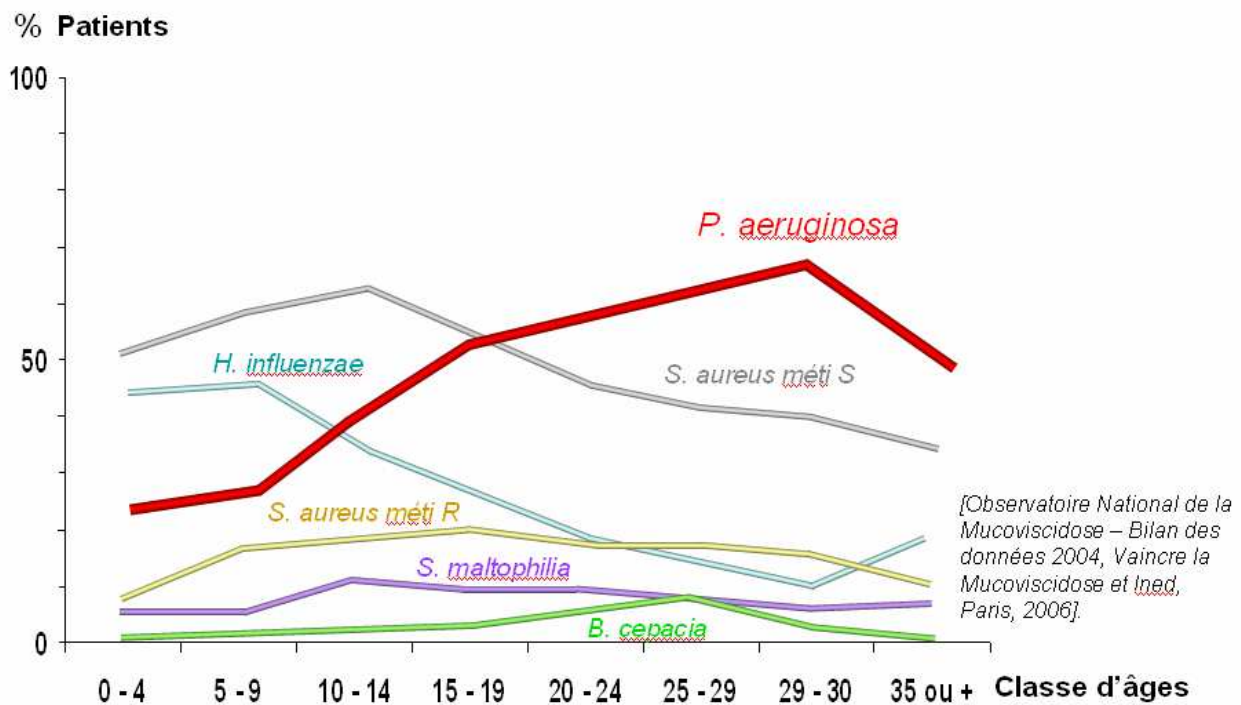
La mucoviscidose présente un environnement particulièrement favorable aux infections bactériennes (fig.1.2). En effet, il existe chez l'individu sain, un mécanisme de défense dans les voies pulmonaires faisant intervenir la clairance mucociliaire d'une part, et l'activité antimicrobienne exercée par le liquide de surface pulmonaire d'autre part. Le mucus est d'ordinaire composé de deux phases, une couche supérieure visqueuse piégeant les particules et les micro-organismes, et une couche inférieure plus fluide dans laquelle battent des cils présents à la surface apicale de l'épithélium respiratoire. Leur mouvement synchrone crée un flux balayant la couche supérieure vers le naso-pharynx afin que celle-ci soit expectorée. Chez un patient CF, le battement des cils est bloqué par la trop grande viscosité du mucus. Les bactéries stagnent puis s'installent sur la paroi bronchique (Lyczak *et al.*, 2002). Par ailleurs, la modification ionique observée à la surface des poumons inhibe l'activité antimicrobienne des lysozymes (Smith *et al.*, 1996), des lactoferrines, des peptides antimicrobiens ( $\beta$ -défensive de type 1) ainsi que des cellules phagocytaires du système immunitaire, telles que les neutrophiles polymorphonucléaires (PMNs) (Tager *et al.*, 1998).



**Figure 1.2.** Processus infectieux dans la mucoviscidose (mucus en bleu, cellules épithéliales en jaune, gradient d'oxygène en bleu et rouge). L'absence de sécrétion en  $\text{Cl}^-$  engendre une hyperabsorption d'eau et d'ions sodium ( $\text{Na}^+$ ). La clearance mucociliaire normale (A) ne fonctionne plus chez le patients CF (B). Le mucus s'épaissit et devient visqueux (C) et propice à l'infection bactérienne (D). Formation de microcolonies et de biofilm (E). Difficulté du système immunitaire à combattre les bactéries (F) (Worlitzsch et al., 2002)

La transmission bactérienne s'effectue le plus souvent soit par contamination directe avec le matériel médical, soit par transmission entre patients. Des études ont montré que la colonisation bactérienne a lieu très tôt dans la vie du patient, et que les espèces bactériennes impliquées varient au cours de la vie (fig.1.3). Au stade précoce de la maladie, les germes le plus souvent retrouvés sont *Staphylococcus aureus* et *Haemophilus influenzae*, mais vers l'âge de 10 ans *Pseudomonas aeruginosa* infecte environ 80 % des patients, et ce, de façon définitive (Aebi et al., 1995). La détérioration préalable de l'épithélium respiratoire par *Staphylococcus aureus* et *Haemophilus influenzae* faciliterait l'adhésion de *Pseudomonas aeruginosa* (Lyczak et al., 2002).





**Figure 1.3.** Prédominance selon l'âge, des bactéries infectant les voies aériennes des patients atteints de mucoviscidose en 2004. [Observatoire National de la Mucoviscidose - Bilan des données 2004, Vaincre la Mucoviscidose et Ined, Paris, 2006] (adaptation).

## 1.2. La bactérie *Pseudomonas aeruginosa*

### 1.2.1. Présentation

*Pseudomonas aeruginosa* est une bactérie Gram-négatif opportuniste vivant normalement à l'état saprophyte dans l'eau, les sols humides et les végétaux, mais qui peut également vivre à l'état commensale sur la peau ou à l'intérieur du système digestif de divers animaux. Le pathogène est inoffensif pour les individus en bonne santé mais devient rapidement très dangereux pour les personnes souffrant de déficiences immunitaires ou présentant des pathologies pouvant initier le développement d'une infection, tels que les grands brûlés. A l'hôpital, certaines souches sont plus dangereuses par leur forte densité et leur résistance due à la sélection exercée par les antibiotiques (Livermore, 2002). *Pseudomonas aeruginosa* est, avec les staphylocoques et les entérobactéries, la bactérie la plus souvent responsable d'infections nosocomiales telles que les septicémies, les infections urinaires, les pancréatites et les dermatites, ainsi que les pneumonies associées aux

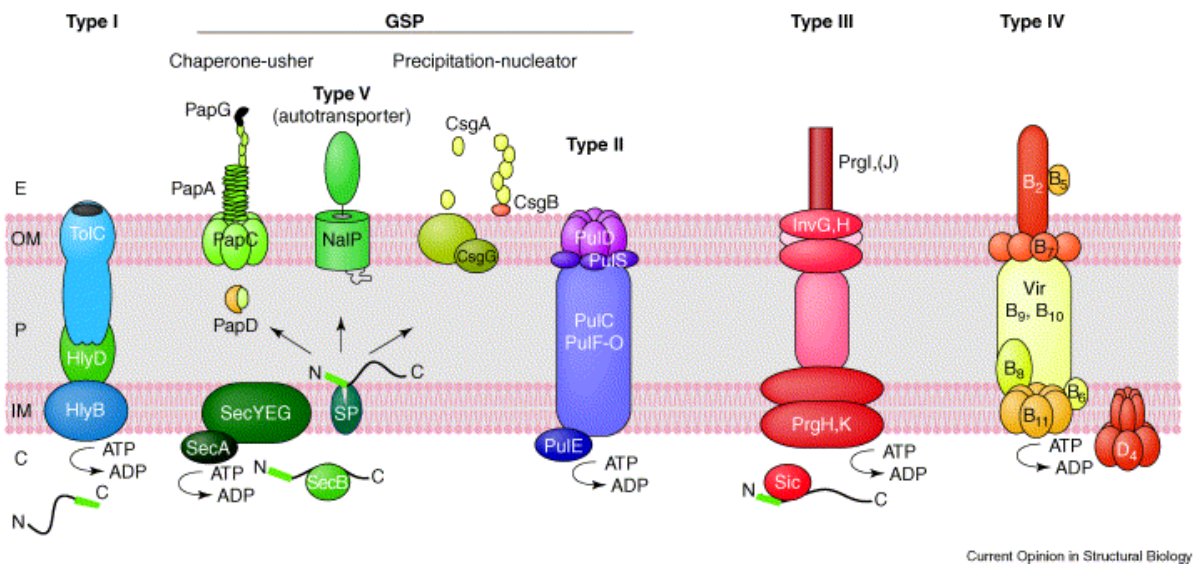
appareils de ventilation (VAP). La colonisation pulmonaire chronique survenant chez les patients atteints de mucoviscidose (CF) s'avère souvent être la cause finale de décès (Odeh and Quinn, 2000).

Le séquençage du génome de la souche de référence PAO1 de *Pseudomonas aeruginosa*, achevé en 2000, a révélé un vaste génome de 6.3 millions de paires de bases codant pour 5570 gènes (Stover *et al.*, 2000). Parmi ces gènes, 8.4 % codent des protéines impliquées dans des fonctions de régulation faisant de cette bactérie le pathogène possédant le plus grand nombre de gènes de régulation. Des travaux comparant différentes souches de *P. aeruginosa* ont montré que son génome était relativement bien conservé mais comportait néanmoins des variations (10%) sur certaines portions codant pour des facteurs de virulence. Des études menées sur les souches de *P. aeruginosa* présentes dans les poumons de patients atteints de mucoviscidose ont révélé que 20% de celles-ci présentent une forte activité de mutation, ce qui pourrait expliquer la grande adaptabilité de *Pseudomonas aeruginosa* aux différents environnements hostiles ainsi que sa grande pathogénicité (Wolfgang *et al.*, 2003).

### 1.2.2. Pathogénicité et toxicité dans les infections respiratoires

*Pseudomonas aeruginosa* utilise plusieurs systèmes de sécrétions distincts pour injecter les protéines toxiques qui jouent un rôle fondamentale lors de l'infection (fig.1.4). Le système de sécrétion de type II secrète des exo-produits (toxines et enzymes de dégradation) dans l'environnement proche des cellules alors que le système de sécrétion de type III délivre les toxines directement dans le cytoplasme de la cellule hôte (Kipnis *et al.*, 2006). Les sécrétions de ces facteurs de virulence sont assujetties à des mécanismes de régulation complexes tel que le « quorum sensing » permettant à la bactérie d'adapter son phénotype au milieu dans lequel elle se trouve en formant notamment un biofilm, souvent rencontré dans les infections chroniques.

Nous aborderons dans les chapitres suivants ces quelques éléments, tout en sachant qu'ils ne constituent en aucun cas une liste exhaustive parmi les facteurs de virulence exprimés par la bactérie *P. aeruginosa*.



**Figure 1.4.** Vue d'ensemble schématique des principaux systèmes de sécrétion de protéines dans les bactéries Gram négatif. Les systèmes de sécrétion de type I, III et IV sécrètent des protéines en une seule étape sans intermédiaire périplasmique. Les systèmes de sécrétion Chaperone-usher, autotransporteur, précipitation-nucleator et de type II utilisent des intermédiaires périplasmiques qui sont transportés par la voie de sécrétion générale. C, espace cytoplasmique ; E, espace extracellulaire ; IM, membrane intérieure ; OM, membrane externe ; P, périplasme (Kipnis, Sawa et al., 2006).

### 1.2.2.1. Système de sécrétion de type II et protéines secrétées

La sécrétion par le système de type II comprend de nombreuses protéines. Parmi celles-ci, citons les ATPases, les protéines chaperones et les peptidases. Les protéines exportées par ce système possèdent une séquence d'acides aminés hydrophobes, dite « séquence signal », qui est nécessaire à leur exportation mais qui sera éliminée par la peptidase dans l'espace périplasmique.

L'élastase (lasB) est une métalloprotéine à zinc sécrétée par *P. aeruginosa* à l'intérieur de l'espace cellulaire via le système de sécrétion de type II. Cette protéine exerce son rôle toxique dans les infections pulmonaires par la rupture de l'élastine, un composant majeur de l'épithélium respiratoire, responsable notamment de la contraction et de l'extension des tissus pulmonaires lors de la respiration. Sa dégradation augmente la perméabilité de l'épithélium et facilite ainsi l'action d'autres protéines toxiques (Azghani et al., 2000). Il a été montré que l'élastase diminuait également la réponse immunitaire en détruisant les protéines surfactantes A et

D (Mariencheck *et al.*, 2003). D'autres études ont montré une activité pro inflammatoire augmentant la concentration en interleukine IL-8 (Kon *et al.*, 1999).

Les phospholipases C sont également sécrétées par le système de sécrétion de type II. Elles ciblent la membrane lipidique des cellules eucaryotes. Elles participent à l'inflammation (Konig *et al.*, 1996) et à l'inactivation des surfactants en s'attaquant aux phosphatidycholines (Holm *et al.*, 1991).

L'exotoxine A (ExoA) est une toxine de type AB contenant un domaine A catalytique et un domaine B interagissant de manière spécifique avec les récepteurs de la cellule hôte. L'exoA tient un rôle majeur dans la virulence de la bactérie (Miyazaki *et al.*, 1995). C'est une ADP-ribosyl transférase qui est sécrétée dans l'espace intercellulaire via le système de sécrétion de type II qui inhibe le facteur d'élongation 2, ce qui entraîne l'arrêt de la synthèse protéique et la mort de la cellule par nécrose (Wick *et al.*, 1990).

La pyocyanine est un pigment bleu sécrété par la bactérie et qui est impliqué dans de nombreux mécanismes pathogènes (Lau *et al.*, 2004). Elle réprime la réponse immunitaire de la cellule hôte, induit l'apoptose des neutrophiles (Allen *et al.*, 2005) et augmente le nombre d'interleukine IL8 (Denning *et al.*, 1998).

#### 1.2.2.2. Le système de sécrétion de type III (SSTT) et toxines

Le système de sécrétion de type III permet à la bactérie d'injecter directement les toxines à l'intérieur de la cellule en utilisant un appendice extracellulaire. La translocation des toxines depuis la bactérie s'effectue en traversant la membrane bactérienne à l'aide d'un appendice ressemblant à une aiguille capable de percer la membrane de la cellule eucaryote. Des travaux ont montré que le système de sécrétion III permettait à des souches de *P. aeruginosa* issues de patients CF de résister à l'action bactéricide des neutrophiles polymorphonucléaires humains (PMNs) et induisait leur mort cellulaire alors qu'un système de sécrétion III rendu non fonctionnel ne pouvait plus induire la mort cellulaire de PMNs (Dacheux *et al.*, 1999) et que les protéines sécrétées par le SSTT conduisaient à l'oncose des macrophages (Dacheux *et al.*, 2001). Plusieurs toxines sont maintenant connues pour être sécrétées via ce système par *P. aeruginosa* : l'exoY, l'exoS, l'exoT et l'exoU.

ExoS et ExoT sont des cytotoxines partageant 76% d'homologie de séquence et comprenant deux domaines actifs : une ADP-ribosyltransférase en C-terminal et un domaine

activateur de la rho GTPase (GAP, GTPase activating protein) en N-terminal. Les protéines Rho sont impliquées dans le maintien du cytosquelette d'actine. L'activité GAP de ExoS et ExoT semble être identique biologiquement bien que l'activité de ExoT apparaisse moindre et plus spécifique. L'activité GAP sur les cellules engendre un dérèglement de l'organisation normale du cytosquelette de la cellule. L'expression du domaine ADP-ribosyltransférase inhibe aussi l'internalisation de *Pseudomonas* par les cellules épithéliales et les macrophages (Shaver and Hauser, 2004, Barbieri, 2004 #126).

ExoY est une adénylate cyclase injectée directement par le système de sécrétion de type III dans le cytosol de la cellule hôte et nécessite l'intervention de cofacteurs eucaryotes. Une étude récente a montré que l'ExoY augmentait le niveau d' AMPc intracellulaire ce qui engendrait la formation de trous intercellulaires et la détérioration des cellules endothéliales pulmonaires (Sayner *et al.*, 2004).

ExoU possède une activité phospholipase/lysophospholipase dégradant la membrane cellulaire une fois transloquée dans le cytosole et activée par des cofacteurs eucaryotes encore inconnus (Sato *et al.*, 2005). L'exoU montre une cytotoxicité 100 fois plus grande que celle de ExoS (Lee *et al.*, 2005).

### 1.2.3. *Le quorum sensing*

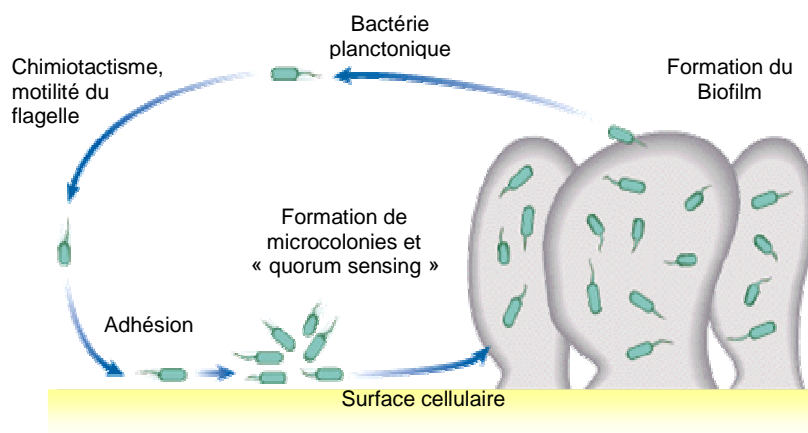
Le « quorum sensing » (QS) est un système de communication entre bactéries dépendant de la densité de population bactérienne et reposant sur la synthèse et la diffusion de petites molécules appelées acylhomosérines lactones (AHL). Ce QS amplifie et coordonne l'expression de gènes de virulence en activant leur transcription. *Pseudomonas aeruginosa* a la particularité de posséder deux systèmes de régulation : le système *las* et le système *rhl*. Le premier comprend les gènes *lasR* et *lasI* codant respectivement pour la protéine régulatrice LasR et une enzyme LasI nécessaire à la synthèse des AHL spécifiques, les 3-oxo-C12-HSL, capables de traverser les membranes facilement. Lorsque la concentration en AHL atteint un seuil critique, témoin d'une forte population bactérienne typiquement entre la phase exponentielle de croissance et la phase stationnaire, les AHL forment un complexe avec deux protéines LasR, ce qui active la transcription de plusieurs gènes de virulence de manière synchrone dans toute la population bactérienne.

Le second système fonctionne de la même façon, en utilisant les gènes *rhlR* et *rhlI* codant pour les protéines RhlR et RhlI nécessaires à la synthèse d'un second type de AHL : les C4-HSL. Il existe une interaction hiérarchique entre les deux systèmes de régulation puisque les AHL sécrétées par le système *las* régule positivement le système *rhl*.

Plusieurs études quantifiant la présence des AHL dans les expectorations et les biopsies des patients CF attestent du rôle que joue le QS dans ce type d'infections chroniques (Davies *et al.*, 1998, Erickson *et al.*, 2002, Geisenberger, 2000 #46. Par ailleurs le QS est impliqué dans les dernières étapes de la formation de biofilm (Davies *et al.*, 1998).

#### 1.2.4. Le biofilm

Parmi les bactéries pathogènes responsables des infections bronchiques, certaines souches se développent au sein d'un biofilm, architecture composée de microcolonies enchâssées dans une matrice d'exopolysaccharides (EPS) et formant des structures en forme de champignons spécifiques de l'infection (fig. 1.5). Cette couche de biofilm protectrice, signe d'une installation durable des bactéries aux structures contaminées, gêne la pénétration des antibiotiques et permet une résistance à la phagocytose et aux anticorps. Trois étapes peuvent être distinguées dans la formation du biofilm : une étape initiale d'attachement à la surface des muqueuses, une étape de prolifération aboutissant à la formation de microcolonies et une étape de structuration du biofilm. Cette architecture ajoute encore à l'obstruction et à la viscosité des poumons CF.



**Figure 1.5.** Formation du biofilm bactérien (Kolter and Losick, 1998).

### 1.3. Les sucres : éléments stratégiques de l'infection

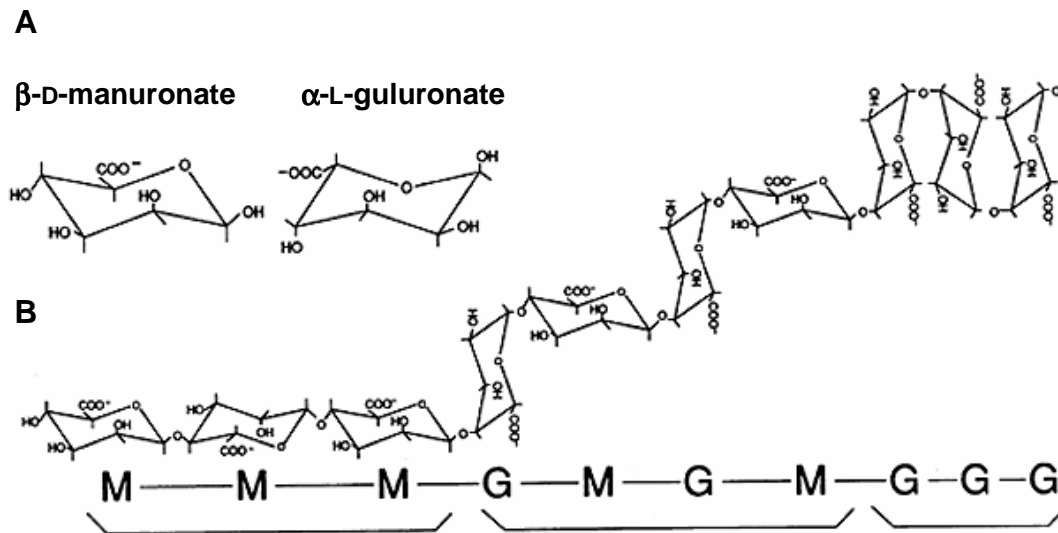
*...home, sweet home*

Les paragraphes qui vont suivre ont pour objet de démontrer que les sucres, qu'il s'agisse de glycoconjugués ou de polysaccharides sécrétés, font partie de l'environnement immédiat de la bactérie et interviennent à divers niveaux dans les mécanismes de l'infection. On les retrouve présents à la surface des membranes cellulaires de la bactérie et de la cellule hôte. Ils assurent aussi une part importante de la structure du biofilm. L'étude de ces glucides ainsi que des molécules susceptibles d'interagir avec ceux-ci présentent donc un intérêt pour la compréhension du processus infectieux.

#### *1.3.1. Les exopolysaccharides (EPS)*

La matrice extracellulaire que constitue le biofilm se compose essentiellement d'exopolysaccharides (EPS) constituant 50 à 90% de la matière organique totale du biofilm, mais aussi de protéines et de sels. (Wingender *et al.*, 1999). La nature biochimique et physico-chimique des EPS dépend largement des microorganismes concernés et des conditions environnementales (pH, salinité, température). Les EPS ont une haute masse molaire et sont très hétérogènes. Ils forment des unités répétitives de carbohydrates plus ou moins longues, pouvant posséder des ramifications, des fonctions alcool, amine, acide, ainsi que des groupements sulfate et phosphate. La forte implication de *Pseudomonas aeruginosa* dans les maladies nosocomiales fait que son biofilm est aujourd'hui l'un des plus étudiés.

Certaines études ont ainsi trouvé dans le biofilm, la présence d'alginate, un polysaccharide linéaire également produit par certaines algues brunes. L'alginate est un polysaccharide chargé négativement et qui se compose d'acides L-guluronique (G) et D-mannuronique (M) liés entre eux par des liaisons  $\beta$ -1,4 (Evans and Linker, 1973) (fig.1.6). Durant son transfert périplasmique, il subit quelques modifications telles qu'une acétylation et une épimérisation (Remminghorst and Rehm, 2006). Le polymère s'organise de trois façons possibles : soit en une succession de blocs d'homopolymères, (MMM) ou (GGG), ou bien en bloc d'hétéropolymères (MGMG).



**Figure 1.6.** (A) Monosaccharides constituant l'alginate. (B) Structure du polymère d'alginate pour différents types d'alternances (MMM), (GMGM) et (GGG) (Chitnis and Ohman, 1990).

Les conditions inflammatoires rencontrées dans les poumons de patients atteints de mucoviscidose engendrent une surproduction d'alginate par la bactérie et sa différenciation en souche dites « mucoïdes » (Mathee *et al.*, 1999, DeVries, 1994 #48).

Bien que l'alginate participe à l'architecture du biofilm (Hentzer *et al.*, 2001), son rôle n'est pas encore bien compris. Des études récentes ont par ailleurs contesté la prédominance ainsi que le rôle des alginates dans le développement de biofilm chez les souches PAO1 et PA14 (Wozniak *et al.*, 2003). D'autre part, de nouveaux gènes *pel* et *psl*, identifiés dans les bactéries de la souche PA14, seraient impliqués dans la production de composants matriciels respectivement riches en glucose et en mannose (Friedman and Kolter, 2004).

### 1.3.2. Les lipopolysaccharides (LPS)

La membrane externe des bactéries Gram négatif est composée principalement de lipopolysaccharides (LPS). Les LPS sont des molécules constituées de trois parties: une partie hydrophobe, le lipide A, qui s'enchaîne dans la bi-couche phospholipidique de la bactérie, un noyau oligosaccharidique et un polysaccharide antigène-O variable, qui constituent la base de



l'identification antigénique du sérotype de la bactérie. Le lipide A, appelé également endotoxine, est responsable de la stimulation excessive du système immunitaire qui peut provoquer des chocs septiques (Ernst *et al.*, 2003). Deux formes de LPS sont synthétisées par *P. aeruginosa*, le type A et le type B, qui diffèrent par la composition oligosaccharidique de l'antigène-O.

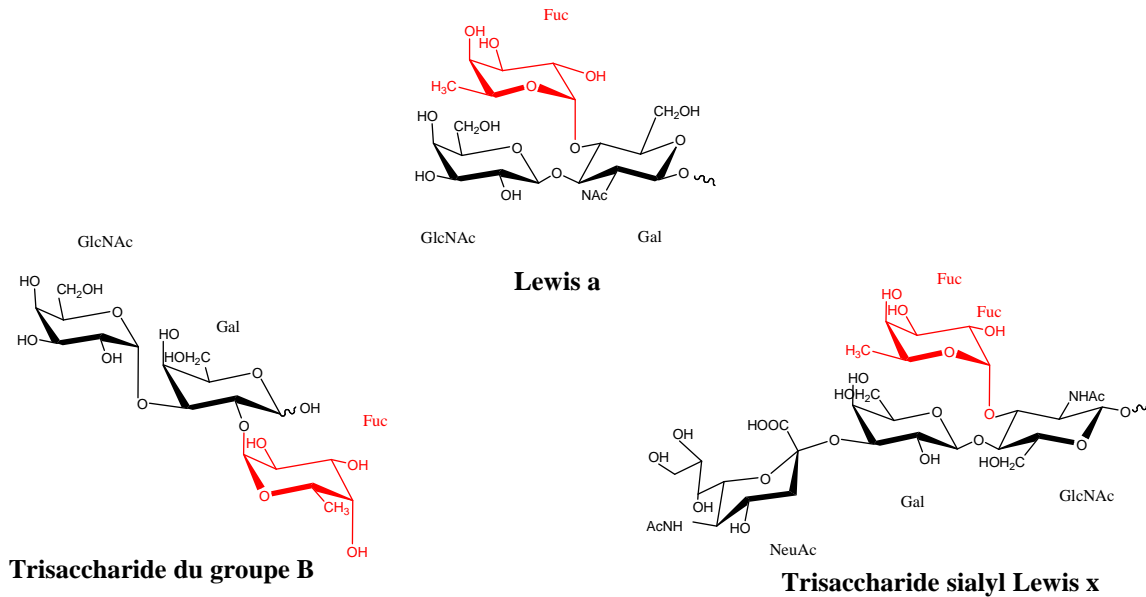
Le noyau LPS de *Pseudomonas aeruginosa* est commun à d'autres bactéries Gram négatif. Il est composé de kétodéoxyoctonate, d'heptoses, d'hexosamines et d'hexoses. Selon la présence ou l'absence de la partie antigène-O, les souches sont dites de phénotype « lisse » ou « rugueux ». Les souches de *P. aeruginosa* causant les infections chez les patients atteints de mucoviscidose sont de phénotype rugueux. La perte de l'antigène-O permettrait à la bactérie d'échapper au système immunitaire qui cible ces antigènes (Goldberg and Pler, 1996) mais le mécanisme de ce changement n'est pas encore connu. Dans le cas de la mucoviscidose, la protéine CFTR interagit avec les LPS de *P. aeruginosa* induisant l'internalisation de la bactérie par les cellules épithéliales pulmonaires. Ce processus est en fait une protection contre les infections car ces cellules sont par la suite éliminées par le mécanisme de clearance mucociliaire. Des travaux ont en effet montré que la protéine CFTR sauvage internalise 50 fois plus de bactéries que la protéine CFTR  $\Delta F508$  (Pier *et al.*, 1997).

### 1.3.3. Les mucines

Les voies aériennes sont recouvertes d'une couche de mucus. Celui-ci est un mélange de différentes molécules sécrétées par les cellules muqueuses et séreuses des glandes bronchiques submucosales de la surface. Il contient entre autres des peptides et des protéines ayant des propriétés protectrices tels que des anticorps, des antiprotéases, des peptides antimicrobiens comme le lysozyme et des défensines capables de détruire les cellules de la paroi bactérienne ainsi que des lactoferrines qui chélatent le fer requis pour le métabolisme bactérien.

Les constituants majeurs du mucus sont les mucines, des glycoprotéines linéaires dont la taille varie de quelques nanomètres à quelques micromètres, et qui contiennent de une à plusieurs centaines de points de glycosylation. Les chaînes de carbohydrates de plusieurs centaines à plusieurs milliers de kDa que l'on retrouve sur les mucines humaines représentent de 70 à 80 % du poids total de la protéine (Lamblin *et al.*, 1991). On compte environ 30 monosaccharides pour 100 acides aminés. L'attachement de ces structures sur la partie peptidique, l'apomucine, s'effectue via une O-glycosylation entre un résidu N-acetylgalactosamine (GalNAc), seul élément

commun à toutes ces chaînes O-glycane, et un acide aminé hydroxylé (sérine ou thréonine). L'apomucine présente en effet une proportion élevée en sérine et en thréonine dont l'ordre et la répétition produisent une grande diversité de forme et influencent vraisemblablement les premières glycosylations. Les apomucines humaines sont codées par au moins sept gènes MUC différents (MUC1, MUC2, MUC4, MUC5B, MUC5AC, MUC7 and MUC8).



**Figure 1.7.** Représentation schématique de quelques oligosaccharides fucosylés (fucose en rouge) présents dans les glycoconjugués et les mucines humaines.

L'élongation de ces chaînes engendre une extrême diversité d'extensions linéaires ou branchées et leur biosynthèse nécessite un processus multi-étapes impliquant de nombreuses glycosyl- ou sulfotransférases. Les mucines humaines contiennent principalement les monosaccharides suivants : fucose, galactose, N-acetylglucosamine, N-acetylgalactosamine et acide sialique (fig.1.7). Elles peuvent également contenir du sulfate et du mannose en petite quantité. On retrouve parmi les extrémités non réductrices des chaînes, des structures portant les oligosaccharides des groupes sanguins A ou B, H et H-sulfaté, ainsi que les épitopes Lewis a, Lewis b, Lewis x ou Lewis y mais aussi les sialyl-, et/ou sulfo- Lewis a et Lewis x. Les mucines interagissent avec des nombreux composants (protéines, peptides, lipides). Leur nature filamenteuse et leur capacité à polymériser confèrent des propriétés visco-élastiques au mucus, essentielles pour l'efficacité de la clearance mucociliaire.

#### 1.3.4. *Modification de la sulfatation et de la glycosylation dans la mucoviscidose*

Diverses anomalies ont été décrites concernant la structure des glycoconjugués observés sur les cellules épithéliales des patients atteints de mucoviscidose. Des études rapportent que la protéine CFTR pourrait, en plus de son rôle sur les mouvements d'ions, influencer directement ou indirectement le fonctionnement des diverses glycosyltransferases (Rhim *et al.*, 2001).

Concernant la glycosylation, une augmentation de la fucosylation a été observée à la fois sur les glycoconjugués des cellules épithéliales et dans les mucines pulmonaires des patients CF. Ce phénomène est réversible lorsque ces mêmes cellules sont transfectées avec la protéine CFTR sauvage (Rhim *et al.*, 2000). Ces études révèlent que ces résidus fucose additionnels sont le plus souvent liés en position 3 ou 4 de la N-acetylglucosamine (Fuc $\alpha$ 1,3/4GlcNAc) correspondant respectivement aux structures Lewis a et Lewis x. Par ailleurs, la sialylation est elle aussi modifiée. Des études ont montré une diminution de la sialylation sur les glycoprotéines et les gangliosides des cellules épithéliales CF (Dosanjh *et al.*, 1994) alors qu'on observe deux fois plus de composants sialylés et sulfatés dans les mucines CF que dans les mucines non-CF (Xia *et al.*, 2005, Lamblin, 2001 #112).

Pour expliquer ces anomalies, certains ont suggéré que l'acidification défectueuse de l'appareil de Golgi, due à un mauvais fonctionnement du canal à chlorure, pourrait modifier l'activité ou la localisation cellulaire de plusieurs sialyltransferases (protéines sensibles au pH) (Barasch and al-Awqati, 1993). Cependant cette hypothèse reste controversée (Seksek *et al.*, 1996) et une hyperacidification des organelles endosomales dans les cellules épithéliales de CF a également été envisagée (Poschet *et al.*, 2001). D'autres études suggèrent que la concentration en donneurs de sulfate (PAPs) serait réglée en partie par CFTR, ce qui influencerait de ce fait le processus de sulfatation (Pasyk and Foskett, 1997).

D'autres travaux ne trouvent aucune corrélation entre l'expression de CFTR et l'activité des glycosyltransferases (Brockhausen *et al.*, 2001) et soulignent que l'état d'inflammation pourrait être responsable des modifications de synthèse et de glycosylation de certaines glycoprotéines. Ainsi, les TNF $\alpha$  pourraient modifier l'activité des glycosyl- et sulfotransferases responsables de la biosynthèse du Lewis x et de son épitope sialylé et/ou sulfaté (De Graaf *et al.*, 1993, Delmotte *et al.*, 2002), en accord avec l'augmentation de ces épitopes dans les mucines pulmonaires des

patients souffrants de mucoviscidose ou de bronchite chronique. Les changements dans la composition du mucus qui surviennent suite à une forte réaction inflammatoire peuvent être un facteur important dans la colonisation des voies respiratoires par *Pseudomonas aeruginosa* (Rhim *et al.*, 2004).

#### 1.4. Pili et flagelle de *P. aeruginosa*

*...Identifier, sonder, adhérer, cibler : le rôle des lectines*

La membrane cellulaire est une structure complexe composée de divers éléments enchâssés dans la bicouche lipidique. Elle présente notamment de nombreux glycoconjugués qui sont autant de marqueurs et de récepteurs conférant une spécificité et une identité à la cellule. Ces sucres exposés ainsi à la surface sont alors la cible de protéines spécifiques capables d'identifier ces structures diverses : les lectines. Avant de s'intéresser plus particulièrement aux lectines bactériennes, rappelons que les lectines sont des protéines ou des glycoprotéines pouvant être d'origine variée : virale, bactérienne, végétale ou animale. Elles sont dépourvues d'activité enzymatique, n'appartiennent pas à un système immunitaire et sont capables de reconnaître spécifiquement des oligosaccharides complexes sans les modifier (Goldstein *et al.*, 1980). Elles sont souvent multivalentes, soit parce qu'elles possèdent plusieurs sites de reconnaissance par molécule, ou bien parce qu'elles s'associent en multimères, ce qui explique le fait qu'elles peuvent précipiter des polysaccharides et agglutiner des bactéries ou des cellules telles que les érythrocytes (hemagglutination).

L'adhésion bactérienne sur les tissus est souvent la première étape du processus d'infection et dans de nombreux cas, des lectines spécifiques sont impliquées dans la reconnaissance et l'adhésion aux récepteurs moléculaires exposés sur les tissus. Certaines adhésines interfèrent avec le système immunitaire, déclenchant des voies signalétiques chez la bactérie et la cellule hôte, assistant la livraison de molécules toxiques dans la cellule hôte et favorisant l'invasion bactérienne. C'est pourquoi ces protéines sont à juste titre considérées comme des facteurs de virulence et font aujourd'hui l'objet d'une attention particulière. Les lectines bactériennes connues sont associées aux fimbriae (pilis et flagelles), aux toxines, ou sont libres en solution.

#### 1.4.1. *Les pilis ou fimbriae*

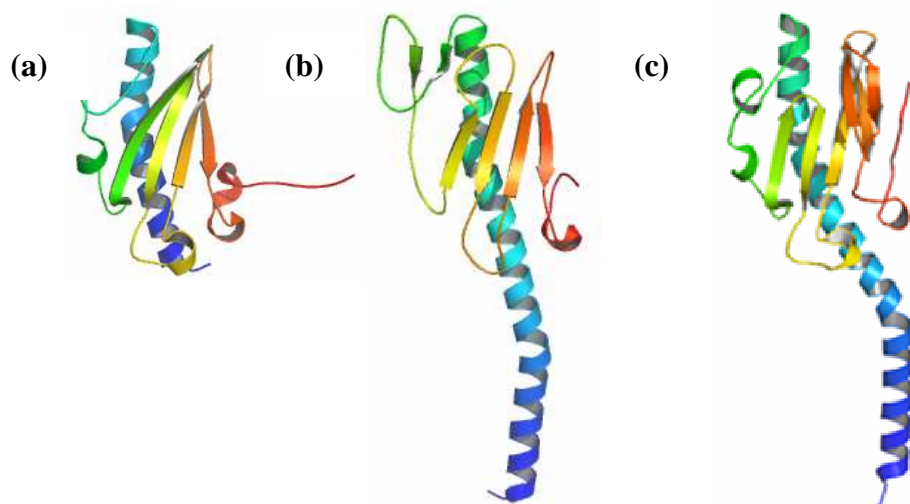
Afin de déterminer le lieu le plus propice à leur colonisation, les bactéries ont développé une sensibilité à leur environnement proche (gradient de pH, de température, de lumière) rassemblé sous le nom de « chimiotactisme ». Pour leur motilité, certaines bactéries se sont dotées d'appendices cellulaires leur permettant sonder la surface de la cellule hôte et se donnant la capacité de se déplacer dans le milieu. Ces organelles de surface sont respectivement les pilis et les flagelles.

##### *Les pilis de type IV*

Les pilis de type IV ont un diamètre de 5 à 7 nm et peuvent s'étendre sur plusieurs µm de longueur. Ils sont composés principalement de piline (protéine PilA), dont les monomères sont assemblés de manière hélicoïdale (5 sous-unités par tour) et qui peuvent être glycosylés et/ou phosphorylés. *P. aeruginosa* possède de nombreux pilis dont les mouvements de « twithing » dus à leurs propriétés rétractiles, permettent à la bactérie de se disperser sur les surfaces humides (Mattick, 2002). Les pilis de type IV sont d'une importance capitale dans la phase d'adhésion et la virulence de la bactérie (Hahn, 1997). Des travaux ont montré que les pilis de type IV étaient responsables de 90 % de la capacité d'adhérence sur les cellules pneumocytaires humaines A549 et responsables de 90% de la virulence sur les cellules de souris (Farinha *et al.*, 1994).

Les pilis de type IV se fixent aux glycosphingolipides asialo-GM1 et asialo-GM2 des membranes cellulaires épithéliales (Gupta *et al.*, 1994). Des tests effectués avec les souches PAO et PAK ont permis d'identifier plus précisément le motif disaccharide GalNAc(β1-4)Gal des glycosphingolipides comme récepteur des pili (Sheth *et al.*, 1994). Des tests d'interaction ont mis en évidence que seule l'extrémité du pilus interagissait avec les asialo-GM1 et qu'une boucle fermée d'un pont disulfure dans la région C-terminale de la sous-unite piline et comptant de 12 à 17 acides aminés semi-conservés, était impliquée dans la reconnaissance. Les données structurales ont par la suite mis en évidence que cette partie lectine est masquée par l'empilement du pilus et n'est exposée qu'à l'extrémité apicale (Hazes *et al.*, 2000). Cette boucle C terminale est conservée chez d'autres types de pilis de type IV comme chez les bactéries *N. gonorrhoeae* et *N. meningitidis*.

La structure tridimensionnelle de la piline MS11 de *N. gonorrhoeae* a été résolue (Parge *et al.*, 1995) et permet de rationaliser les premières données structurales concernant les pili de *P. aeruginosa* (Keizer *et al.*, 2001). Une hélice hydrophobe est fortement conservée et constitue sans doute l'architecture commune à ce type de protéine. L'extérieur de la fibre est constitué par un échafaudage fortement organisé de feuilletts  $\beta$  rassemblés autour de l'hélice ainsi que par une région exposée d'une grande variabilité. Les structures cristallines de pilines tronquées des souches PAK et K122-A de *P. aeruginosa* ont également été résolues (Craig *et al.*, 2003, Audette, 2004 #16) (fig.1.8).

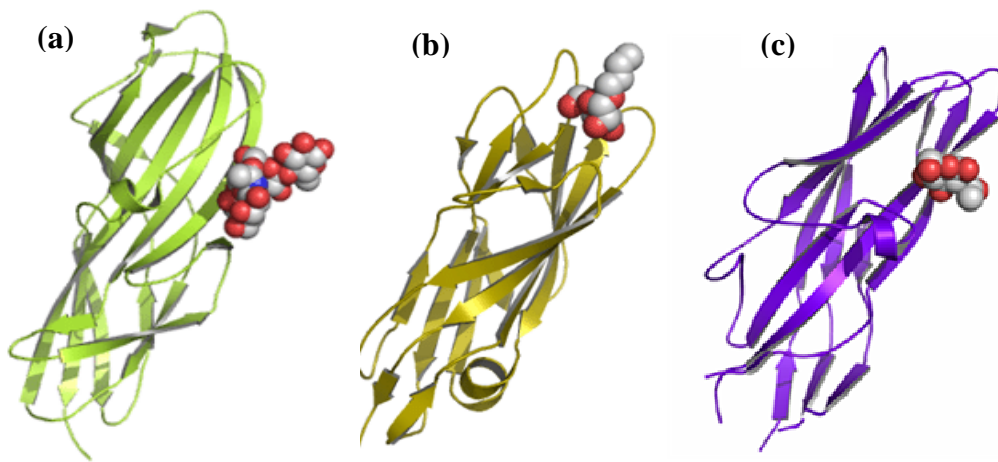


**Figure 1.8.** Structures cristallines de différentes pilines (a) structure tronquée de *P. aeruginosa* K122-4 (1QVE) (Audette *et al.*, 2004); (b) *P. aeruginosa* PAK (1OQW) (Craig, Taylor *et al.*, 2003); (c) piline MS11 de *N. gonorrhoeae* (1AY2)(Parge, Forest *et al.*, 1995).

### *Les pilis de type I*

Des pilis présentant un autre type d'assemblage dépendant d'une protéine chaperone CUP ont été identifiés à partir du séquençage du génome de la souche *P. aeruginosa* PAO1 (Vallet *et al.*,

2001). Les domaines terminaux présentent des homologies avec des adhésines de souches uropathogènes de *E. coli* : FimH qui contient un domaine lectine à mannose (Hung *et al.*, 2002) et PapG qui contient un domaine spécifique de globoside (GalNAc $\beta$ 1-3Gal(1-4Gal $\beta$ 1-4GlcCer) (Dodson *et al.*, 2001). Ces domaines lectiniques adoptent des repliements similaires de type  $\beta$ -sandwich allongé mais ont des sites de reconnaissance différents.

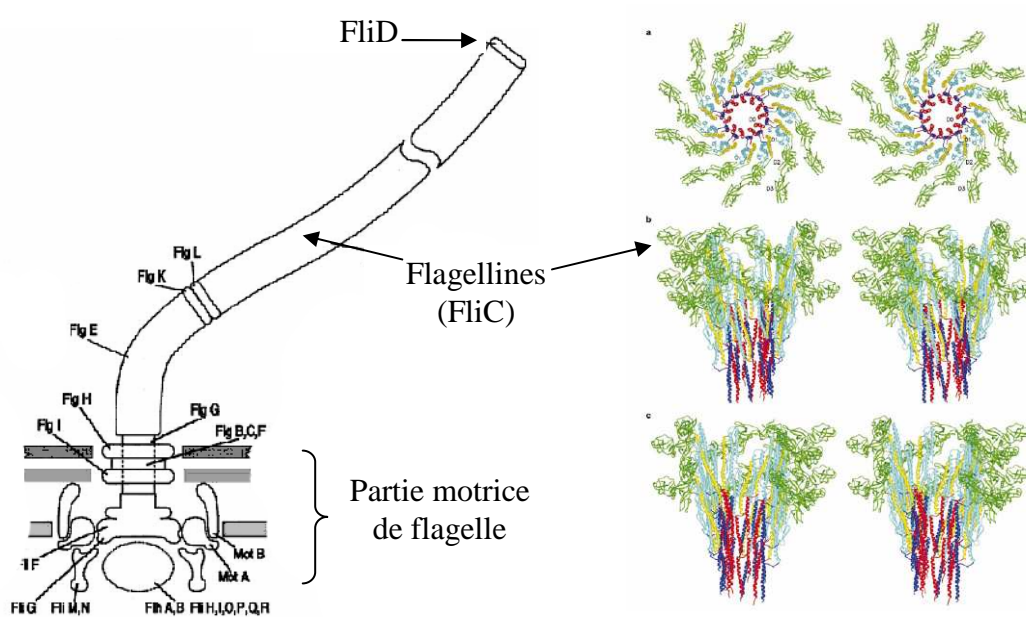


**Figure 1.9.** Structures cristallines de domaines lectines d'adhésines fimbriales : (a) PAPG complexé avec le tétrasaccharide GBO4 (pdb 1J8R). (b) FimH complexé avec butyl-  $\alpha$ -D-mannopyranoside (pdb 1UWF). (c) F17-AG en complexe avec le GlcNAc (pdb 1O9W)

#### 1.4.2. Le flagelle

*P. aeruginosa* possède un unique flagelle responsable de la motilité de la bactérie par un mouvement de type « swimming ». Ce flagelle partage une forte homologie structurale globale avec le système de sécrétion de type III. Il se compose à sa base d'une structure « moteur » enchâssée dans la membrane bactérienne et dont la rotation est assurée par transport de proton, d'une partie supérieure dite « crochet » en raison de sa forme courbe, et d'un long filament flexible constitué de monomères de flagelline (protéine FliC) (Bardy *et al.*, 2003, Berg, 2003) (fig.1.9). Au cours de la mise en place du flagelle, la flagelline est transportée à l'extrémité distale

via l'intérieur du flagelle. L'empilement est assuré par la coiffe flagellaire (protéine FliD) (Ikeda *et al.*, 1996) (Yonekura *et al.*, 2000). Des travaux ont mis en évidence l'implication du flagelle dans la reconnaissance et l'adhérence sur les surfaces épithéliales (Simpson *et al.*, 1992) et son implication dans la formation de biofilm (O'Toole and Kolter, 1998). Des tests d'interactions in vitro ont montré l'adhésion de la flagelline aux glycolipides GM1 et aux asialoGM1 (Feldman *et al.*, 1998). Plus récemment, d'autres travaux ont montré que les flagellines adhéraient aux mucines MUC1 (Lillehoj *et al.*, 2002).



**Figure 1.9.** (a) Représentation du flagelle et localisation des différentes protéines le composant. (b) Vue stéréographique de l'association des flagellines (FliC) qui constitue le filament du flagelle. Données obtenues par microscopie électronique sur le flagelle de *Salmonella typhimurium* (Yonekura *et al.*, 2003).

Les protéines de la coiffe flagellaire (FliD) seraient également impliquées dans la reconnaissance des glycannes de l'hôte. La protéine FliD inhibe l'interaction de *Pseudomonas aeruginosa* sur des mucines (Arora *et al.*, 1998). Deux types seulement de protéines FliD associées respectivement aux deux types de flagelline des souches PAK et PAO1 ont été recensés (Arora *et al.*, 2000, Spangenberg *et al.*, 1996). Des tests d'adhésion indiquent que la protéine FliD de la souche PAO1 (type B) reconnaît les glycoconjugués dérivés trouvés à la périphérie des mucines



respiratoires humaines portant du Lewis x ou sialyl-Lewis x, alors que l'adhésion des flagellines se limiterait au Lewis x. En revanche, la protéine FliD de type A (souche PAK) n'est pas impliquée dans l'attache aux glycotopes Lewis x, sialyl-Le<sup>x</sup>, ou sulfosialyl-Le<sup>x</sup> (Scharfman *et al.*, 2001). Cependant les domaines lectines de la flagelline et de FliD n'ont pas encore été localisés et caractérisés.

## 1.5 Lectines solubles de *P. aeruginosa*

Parmi les facteurs de virulence sécrétés par la bactérie, deux lectines solubles ont été découvertes. La première, PA-IL (ou LecA), d'une taille de 12.7 kDa, montre une spécificité pour le galactose. La lectine PA-IIL (ou LecB) d'une taille de 11.7 kDa montre une très forte affinité pour le fucose (Gilboa-Garber, 1986). Les deux protéines sont principalement présentes dans le cytoplasme de la bactérie mais ont aussi été identifiées sur la surface de la membrane externe de la bactérie (Tielker *et al.*, 2005). Les deux lectines ne montrent aucune identité de séquence mais sont toutes les deux dépendantes de la présence de calcium pour être actives. Les paragraphes suivants feront l'état des données bibliographiques concernant leurs caractéristiques biochimiques, leurs fonctions et leurs structures.

### 1.5.1. La lectine PA-IL

#### *Spécificité et affinité*

La lectine PA-IL fut découverte en 1972 suite à des expériences d'hémagglutination d'érythrocytes traités à la papaïne en présence d'extraits de *P. aeruginosa*. La lectine fut ensuite isolée par chromatographie d'affinité sur une colonne de sépharose à partir d'une culture de *P. aeruginosa* (Gilboa-Garber, 1972). La lectine est spécifique pour le D-galactose avec une constante d'association ( $K_d$ ) de  $3,4 \times 10^4 \text{ M}^{-1}$  (Garber *et al.*, 1992). Elle montre une préférence pour l'anomère  $\alpha$  du galactose et se lie avec une force décroissante aux disaccharides suivants:  $\alpha\text{Gal1-6Glc}$  (mélitiose),  $\alpha\text{Gal1-4Gal}$  (galabiose),  $\alpha\text{Gal1-3Gal}$ . La lectine fixe aussi un grand nombre de glycoprotéines et de glycosphingolipides. Parmi eux, les glycoconjugués contenant les

structures non-substituées  $\alpha$ Gal1-4Gal tels que les antigènes humains du groupe sanguin P1 et Pk, ou encore les disaccharides  $\alpha$ Gal1-3Gal du groupe sanguin B. Cette spécificité en fait un outil intéressant pour tout ce qui concerne les xénotransplantations puisqu'elle peut marquer le xéno-antigène responsable des rejets lors des greffes d'organes de porc sur l'homme.

### *Rôle de PA-IL dans l'infection*

Des études ont montré que l'expression de PA-IL était régulée par le « quorum sensing » et dépendait directement du système RhlR/C4-HSL. En outre sa production dépend de la phase de croissance bactérienne et nécessite le cofacteur RpoS spécifique de la phase stationnaire (Winzer *et al.*, 2000). D'autres études ont identifié un nouveau régulateur MvaT régulant négativement la production de PA-IL (Diggle *et al.*, 2002) ainsi que la présence nécessaire de molécules signal de quinolone (Diggle *et al.*, 2003). Le rôle de la lectine est multiple. Pendant l'infection, les bactéries peuvent se protéger contre les défenses immunitaires en formant des colonies (Gilboa-Garber, 1997), et la lectine PA-IL, de par sa nature multivalente, pourrait jouer un rôle par l'intermédiaire de dérivés de galactose présents sur les lipopolysaccharides (LPS) de la membrane externe (Sadovskaya *et al.*, 1998). Des traces de LPS sont toujours retrouvés lors des purifications de la lectine et il a été également démontré que PA-IL provoque l'agglutination des bactéries *E. coli* (Sudakevitz and Gilboa-Garber, 1982).

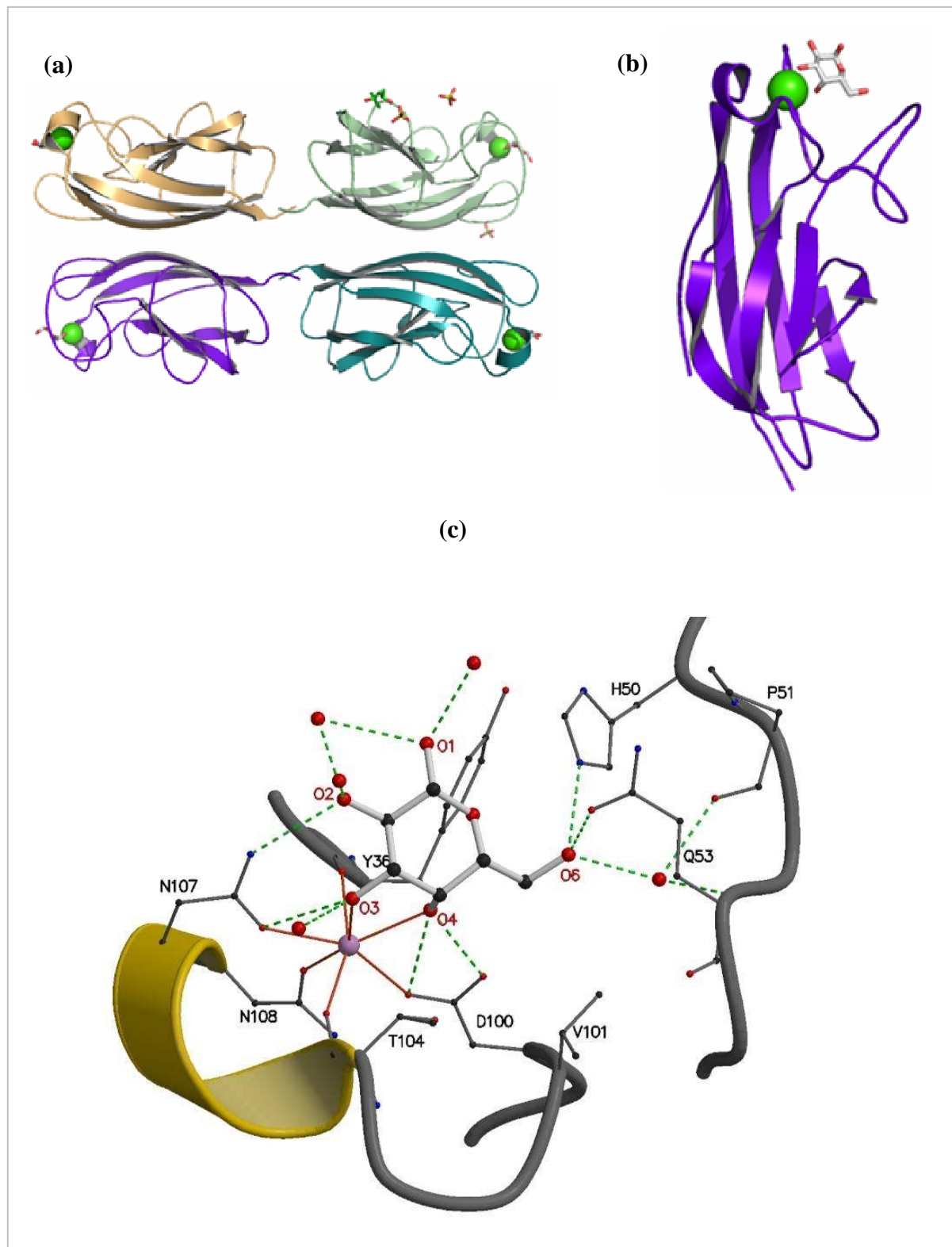
PA-IL pourrait également être impliquée dans la formation et la stabilisation du biofilm. Des analyses statiques d'épaisseur de biofilm sur du polystyrène ont montré une réduction d'épaisseur après déletion du gène *lecA* et une augmentation lorsque celui-ci est surexprimé dans les souches PAO-P47. En outre, un biofilm formé pouvait être dispersé par l'addition d'IPTG (Diggle *et al.*, 2006).

D'autres études ont mis en évidence l'importance de PA-IL dans l'adhérence de *P. aeruginosa* sur les cellules épithéliales intestinales de souris mais aussi dans la perméabilité des cellules permettant ainsi aux exoproduits cytotoxiques, tel que l'exotoxine A, de traverser la barrière épithéliale (Laughlin *et al.*, 2000). D'autres études ont démontré la participation d'interactions de type lectine dans le processus d'adhésion de *P. aeruginosa* NK 125 502 à la fibronectine immobilisée. L'adhésion de *P. aeruginosa* est fortement réduite en présence de l'acide sialique du N-acetylglucosamine (GlcNAc), et du N-acetylgalactosamine (GalNAc) et de façon moindre en présence de glucosamine, de galactosamine, du fucose, du mannose, et du

galactose, confirmant la participation de dérivés de galactose en tant que composants importants pour l'adhérence bactérienne (Rebiere-Huet *et al.*, 2004). Par ailleurs, PA-IL est directement cytotoxique pour les cellules épithéliales respiratoires et inhibe leur croissance (Bajolet-Laudinat *et al.*, 1994).

### *Structure de PA-IL*

La structure quaternaire de PA-IL montre un arrangement tétramérique autour d'un pseudo axe de symétrie C<sub>222</sub> (Cioci *et al.*, 2003) (fig.1.10). Chaque monomère adopte un repliement de brain antiparallèle de type  $\beta$ -sandwich. La structure complexée avec du D-galactose montre que tous les groupements hydroxyle du sucre, à l'exception de O1, interagissent avec la protéine et plus particulièrement O3 et O4, qui participent à la coordination du calcium. Une autre boucle allant des acides aminés 36 à 64 interagit de manière spécifique avec le groupe hydroxyméthyl en O6 afin d'établir des liaisons hydrogène via His50, Gln53. et une molécule d'eau faisant un pont hydrogène. Le calcium en combinaison avec un réseau de liaisons hydrogène forme un site de liaison très spécifique pour le galactose.



**Figure 1.10.** Structure cristallines de la lectine bactérienne PA-IL complexée avec du galactose: (a) tétramère, (b) monomère, (c) site de reconnaissance (pdb 1OKO).

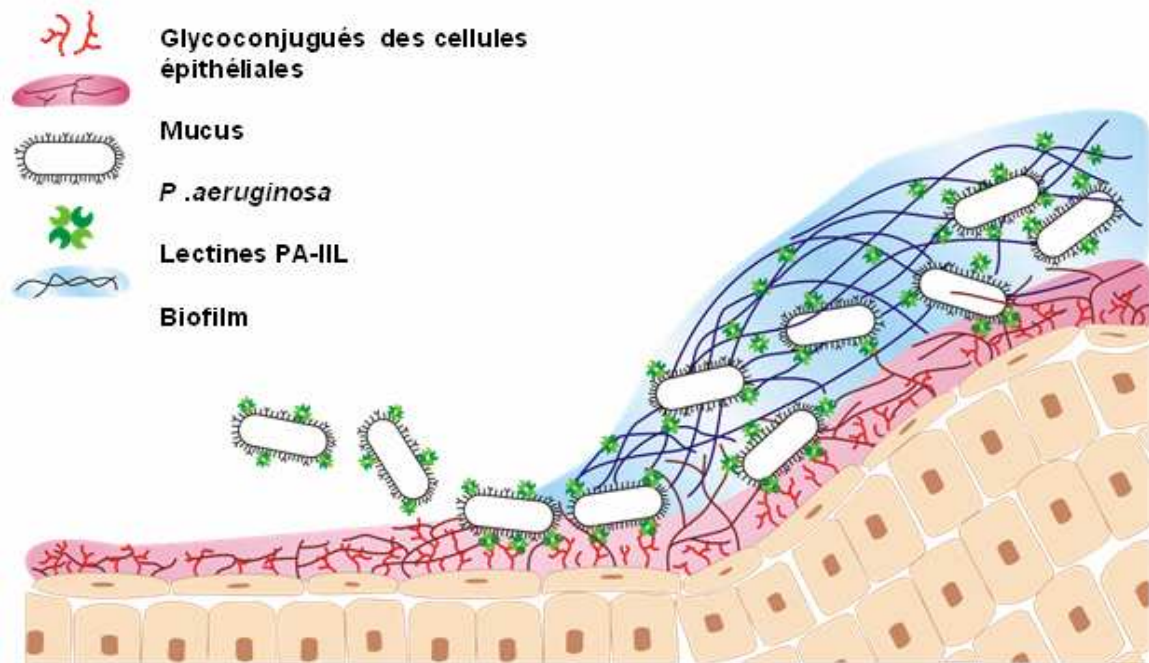
### 1.5.2. La lectine PA-III

#### *Spécificité et affinité de PA-III*

La lectine PA-III, dépendante de la présence d'ions divalents, se caractérise par une forte affinité pour le L-fucose avec une constante d'affinité ( $K_a$  de  $1.6 \times 10^6 \text{ M}^{-1}$ ) rarement rencontrée lors d'une interaction protéine-sucre (Garber *et al.*, 1987). Les tests d'inhibitions d'hémostagglutination ont montré cependant qu'elle était capable de reconnaître avec une affinité décroissante les monosaccharides suivants L-fucose > L-galactose > D-arabinose > D-fructose et enfin le D-mannose, alors que le D-galactose n'est pas reconnu (Imberty *et al.*, 2004). La lectine s'avère aussi reconnaître les oligosaccharides contenant du fucose comme le trisaccharide Lewis a qui est à ce jour son meilleur ligand naturel alors que l'affinité pour le Lewis x est meilleure, indiquant que l'orientation du groupement N-acetylglucosamine joue un rôle important dans la reconnaissance. Nous reviendrons plus en détail sur la structure globale de la lectine ainsi que sur son mode de reconnaissance dans les chapitres suivants.

#### *Rôle de PA-III dans l'infection*

La lectine PA-III est produite sous le contrôle du « *quorum sensing* » (Schuster *et al.*, 2003, Wagner *et al.*, 2003, Winzer, Falconer *et al.*, 2000). Elle a la capacité d'inhiber le battement ciliaire des cellules pulmonaires *in vitro* (Adam *et al.*, 1997). Des travaux récents ont montré que des souches mutantes ne produisant plus la lectine PA-III devenaient inaptes à former du biofilm (Tielker, Hacker *et al.*, 2005). Ces travaux montrent de plus que la lectine se localise sur la paroi extracellulaire de la bactérie en s'associant aux glycoconjugués présents sur la surface de celle-ci (Tielker, Hacker *et al.*, 2005). La lectine pourrait être impliquée dans la reconnaissance et l'adhésion bactérie-hôte, bactérie-bactérie ou encore bactérie-biofilm (fig. 1.11). Mieux comprendre la spécificité et le mode des interactions mises en jeu dans les interactions entre la lectine PA-III et les oligosaccharides présents à la surface des cellules épithéliales des patients atteints de mucoviscidose constitue un axe de recherche nouveau pour combattre les infections causées par la bactérie.



**Figure 1.11.** Les rôles possibles de la lectine PA-IIL durant l'infection (Imberty et al., 2004). La lectine soluble PA-IIL (en vert) présente à la surface de la bactérie pourrait reconnaître les glycoconjugués de la surface épithéliale (en rouge), les mucines (en rose) ou les oligosaccharides présents dans le biofilm (en bleu). Son rôle précis reste à déterminer.

## **Objectif de cette thèse de recherche :**

Mes travaux ont pour objet l'étude de l'affinité et de la spécificité de la lectine PA-IIL pour différents ligands. Ce qui revient d'une part, à identifier les sucres qui sont préférentiellement reconnus par la lectine (notamment à l'aide de tests d'interaction) et d'autre part, à comprendre pourquoi tel sucre est d'avantage reconnu plutôt qu'un autre. L'analyse des structures et des mécanismes d'interaction mis en jeu dans chaque cas est alors nécessaire. Pour ce faire, j'ai utilisé la cristallographie aux rayons X et la microcalorimétrie de titration isotherme (ITC). Dans un premier temps, j'ai étudié la lectine complexée avec différents monosaccharides et oligosaccharides naturels, ainsi que l'impact de mutations ponctuels dans la séquence de PA-IIL sur l'affinité et la spécificité. L'étude

structurale a ensuite été élargie à des complexes avec des glycomimétiques de synthèse afin de chercher des composés pouvant être des ligands encore meilleurs.

### Pourquoi faire tout cela ?

Le rôle exact de PA-III reste à déterminer, mais les éléments trouvés dans la littérature fondent l'hypothèse que la lectine PA-III serait directement ou indirectement impliquée dans l'adhésion de la bactérie *Pseudomonas aeruginosa* à la surface des poumons lors des infections survenant chez les patients atteints de mucoviscidose. Préciser la cible de la lectine revient donc à préciser son rôle biologique, mais surtout, ces données structurales permettront peut-être la conception de composés thérapeutiques dérivés des sucres qui par compétition avec les cibles naturels pourront inhiber l'adhésion de la bactérie sur les cellules épithéliales pulmonaires.

# 2

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

---

### *Introduction*

Devenir soit même tout petit, jusqu'à l'échelle atomique, et voir de ses yeux, et en temps réel une interaction moléculaire comme si on y était, est le rêve de tout biochimiste. Ce n'est hélas encore pas possible. Mais la combinaison de plusieurs techniques expérimentales permet néanmoins de s'en approcher. Parmi celles-ci, la cristallographie des protéines au rayonnement synchrotron constitue l'alternative la plus performante à l'heure actuelle dans l'investigation structurale des molécules biologiques puisqu'elle restitue une image tridimensionnelle de la macromolécule. Si la connaissance structurale permet d'appréhender les topologies de surface, les forces d'interaction et les conflits stériques, autrement dit l'aspect « statique » d'une interaction, elle ne permet pas toujours de comprendre un mécanisme réactionnel. Elle n'offre qu'une vision relativement figée du phénomène tant il est vrai qu'il s'agit d'une « photographie » instantanée de l'interaction et que la protéine sous forme cristallisée est loin des conditions biologiques. C'est pourquoi l'association d'autres techniques comme la microcalorimétrie qui permet d'accéder aux grandeurs thermodynamiques au cours de l'interaction apporte des informations complémentaires nécessaires à la compréhension du mécanisme réactionnel. D'autres techniques comme les tests de compétition entre ligands nous renseignent davantage sur la spécificité de la protéine. C'est donc bien la complémentarité des méthodes qui permet d'élucider le rôle et le fonctionnement d'une protéine et d'envisager la conception de molécules thérapeutiques par exemple.



## 2.1. La cristallographie aux rayons X

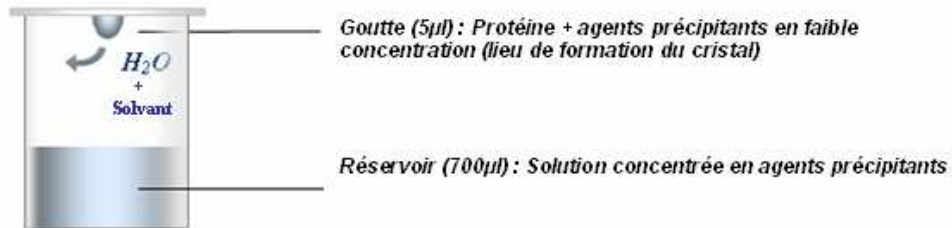
### 2.1.1. La cristallogenèse

Sous certaines conditions, il arrive que des molécules, comme les protéines, s'agglomèrent de manière à former des cristaux. Durant ce processus de cristallisation, les molécules en solution interagissent et passent progressivement d'un état soluble à un état solide ordonné, en s'organisant en un réseau tridimensionnel périodique.

Si la plupart des substances minérales et des petites molécules organiques cristallisent facilement, l'obtention de cristaux de macromolécules biologiques, est souvent plus délicate et constitue en soit le premier obstacle à franchir pour le bio-cristallographe. Les protéines sont en effet des structures biologiques fragiles dont la stabilité dépend fortement de l'environnement immédiat (solvant, pH, lumière, concentration etc), et devient vite précaire en dehors des conditions physiologiques habituelles. Les protéines présentent de surcroît une certaine flexibilité, notamment au niveau des chaînes latérales des acides aminés. Tous ces éléments rendent difficile une cristallogenèse qui impose généralement d'avoir une solution initiale pure et homogène afin de conduire à un solide cristallin et non à un précipité amorphe.

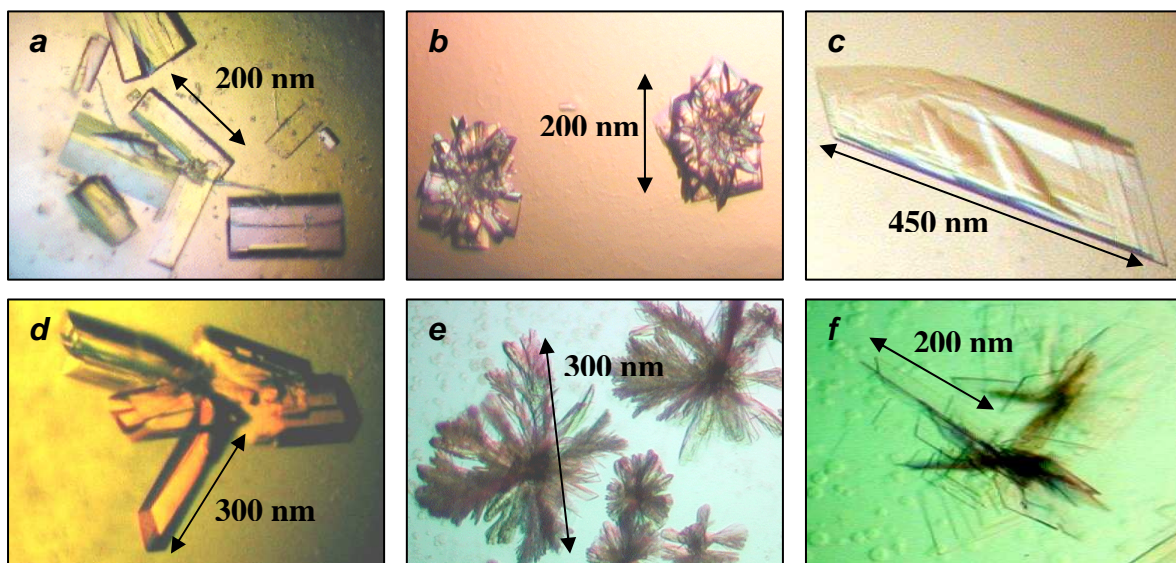
Afin d'obtenir des cristaux de protéine, une des méthodes fréquemment utilisée est « la diffusion en phase gazeuse avec goutte pendante » (fig. 2.1). La technique consiste à mettre en présence dans une enceinte hermétiquement close, une goutte de quelques microlitres d'une solution tamponnée contenant la protéine ainsi que des agents précipitants, avec en contrebas, un réservoir d'un volume largement supérieur (700 µlitres) contenant une solution sans protéine et dont la concentration en agents précipitants est supérieure à celle établie dans la goutte. Avec le temps, les composés volatils (eau ou solvants organiques) vont diffuser dans l'enceinte par voie gazeuse entre les deux liquides jusqu'à ce que les concentrations – ou plutôt les potentiels chimiques – soient équivalents de part et d'autre. Cet équilibre thermodynamique et cinétique engendre du même coup la saturation en protéine, qui alors s'agglomère spontanément initiant ainsi la croissance de microcristaux dans la goutte. Toutefois, la méthode reste empirique et le succès aléatoire tant les paramètres physico-chimiques intervenant sont nombreux. Concentrations des solutés, pH et température, pour ne citer qu'eux, doivent être pris en compte lors de tests successifs tendant à préciser petit à petit l'éventail des conditions expérimentales requises. Des

échantillonnages commerciaux tels que les « Screens Hampton Research » criblant de multiples conditions, constituent souvent le point de départ des essais.



**Figure 2.1.** Principe de la croissance cristalline par la méthode dite de « diffusion en phase gazeuse par goutte pendante ».

Caractériser l'interaction protéine-ligand nécessite l'obtention de cristaux dans lesquels la macromolécule est complexée avec le ligand. Pour ce faire, protéine et ligand sont mélangés dans la goutte afin que tous deux interagissent et cristallisent ensemble : on parle alors de co-cristallisation. Selon les ligands mis en présence, le complexe protéine-ligand ne cristallise pas toujours dans les mêmes conditions expérimentales, ni dans le même groupe d'espace (fig. 2.2). La méthode dite du « trempage », consistant à baigner plus ou moins longtemps un cristal natif (protéine non complexée) dans une solution contenant le ligand afin que celui-ci diffuse dans le cristal, est aussi une alternative, à condition toutefois que l'accessibilité au site de fixation ne soit pas obstruée par les molécules voisines.



**Figure 2.2.** Cristaux de PA-IIL obtenus par co-cristallisation de 10 mg/ml de protéine avec 250 µg/ml de monosaccharide, dans 1,75M de sulfate d'ammonium, 0,1M de Tris-HCl à pH 8,5 et

2mM de  $\text{CaCl}_2$  (Groupe d'espace  $P2_1$ ). **(a)** : Complexe avec le L-fucose **(b)** : Complexe avec le Me- $\alpha$ -L-galactopyranoside **(c)** : Complexe avec le Me- $\beta$ -D-arabinopyranoside. **(d)** : Complexe avec le pentasaccharide LNnFP-V (lacto-N-neofucopentaose-V) (Groupe d'espace  $P1$ ). **(e)** : Complexe avec le trisaccharide Lewis a, dans 0,1M de cacodylate de sodium à pH 6,5, 30% PEG [poly(éthylène glycol)] 8000 et 0.2 M de sulfate d'ammonium. **(f)** : Complexe avec le 2-fucosyl-lactose dans 0,1M de cacodylate de sodium à pH 6,5, 30% PEG 8000 et 0,2 M de sulfate d'ammonium (plaques trop fines et fragiles, les données n'ont pas pu être collectées).

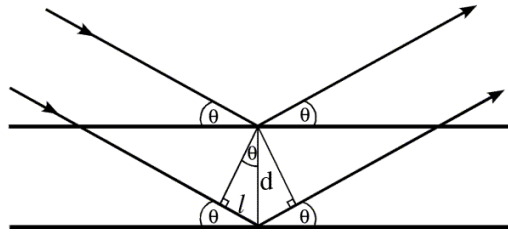
### 2.1.2. Principe de la diffraction aux rayons X

Un cristal étant un réseau tridimensionnel, il est possible de définir une maille élémentaire. La maille élémentaire est le plus petit volume qui, translaté dans les trois directions de l'espace (à distances multiples de a, b et c, qui sont les paramètres de maille), génère un cristal entier. On peut définir aussi une multitude de plans atomiques parallèles équivalents (plans de Bragg) qui coupent le cristal dans toutes les directions. Chaque plan est identifié avec les indices de Miller (h,k,l) définissant la famille de plans. Les vecteurs perpendiculaires à chaque plan de Bragg et ayant une longueur inversement proportionnelle à la distance entre deux plans consécutifs sont nommés vecteurs réciproques. L'ensemble de ces vecteurs prenant leur origine au même endroit, définit le réseau réciproque dans la représentation d'Ewald (figure 2.5).

Notons que les différents jeux de plans parallèles ont des espaces interplanaires différents. W.L. Bragg a démontré que pour un faisceau X incident de longueur d'onde  $\lambda$  traversant un cristal sous un angle  $\theta$ , les plans cristallins parallèles d'indice hkl et d'espace interplanaire  $d_{hkl}$  produisaient un rayon diffracté (que l'on appelle réflexion)(fig. 2.3) seulement si la condition géométrique suivante est rencontrée :

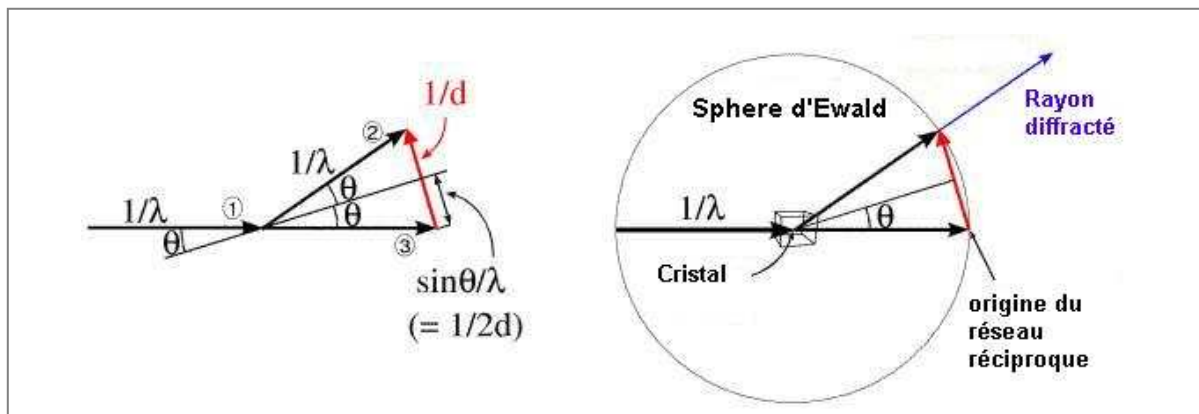
$$2d_{hkl} \sin\theta = n\lambda ,$$

où n est un entier. Dans ce seul cas, le réseau cristallin amplifie la radiation en créant des interférences constructives. Compte tenu du très grand nombre de plans présents dans le cristal, les autres situations finissent toujours par donner lieu à des interférences destructrices.



**Figure 2.3.** Loi de Bragg : La diffraction n'a lieu que si la différence de trajet parcouru par les rayons X correspond à un nombre entier de longueur d'onde

Puisque pour une position fixe donnée, seuls certains plans donneront lieu à des réflexions, il est donc nécessaire pour explorer tous les atomes, de faire tourner le cristal sur lui-même devant le faisceau X incident afin que tour à tour chaque plan atomique puisse engendrer ses réflexions propres. La construction mathématique à l'aide de la sphère d'Ewald et du réseau réciproque est une vue de l'esprit ingénieuse permettant de visualiser et d'indexer chaque réflexion lors de la collecte (fig. 2.4).



**Figure 2.4.** Construction d'Ewald: le cristal placé au centre d'une sphère de rayon  $1/\lambda$  tourne sur lui-même, ce qui équivaut à tourner le réseau réciproque qui a son origine sur l'intersection de la sphère avec le rayon incident. Quand un nœud du réseau réciproque croise la sphère d'Ewald, le rayon est diffracté d'un angle  $2\theta$ , qui peut varier de  $0$  à  $2\pi$

Sur une ligne de lumière d'un synchrotron de troisième génération tel que l'European Synchrotron Radiation Facility (ESRF) à Grenoble, le temps d'acquisition d'une image est de l'ordre de quelques secondes. Chaque image peut être considérée comme une « tranche » de l'espace réciproque. Des centaines de réflexions croisent la sphère d'Ewald et sont mesurées à chaque degré de rotation par un détecteur CCD (Charge-Couple Devices). Celui-ci est recouvert de

phosphore qui émet une lumière visible en réponse aux rayons X. Le dispositif est en fait un compteur de photons qui accumule puis convertit en exacte proportion la quantité de photons reçus en charge électrique, et ce, avec un temps d'acquisition très court (fig. 2.5). Ces données permettent de reconstituer un cliché de diffraction dans lequel chaque tache est l'empreinte d'une unique réflexion d'indices hkl possédant une intensité I propre résultant de la distribution de la densité électronique dans l'unité asymétrique.

### *2.1.3. Paramètres importants et stratégie durant la collecte de données*

Selon l'objectif à atteindre, la stratégie de la collecte ne va pas être la même. En effet, il peut s'agir par exemple d'obtenir les phases à l'aide des méthodes MAD, SAD, auquel cas, on essayera d'avoir le maximum de multiplicité, et l'on travaillera à une longueur d'onde spécifique afin d'obtenir un signal anomal. S'il s'agit de caractériser au mieux un site de reconnaissance, et que l'on possède déjà les phases grâce à une structure précédente, on tentera d'obtenir le maximum de résolution, afin d'obtenir le maximum de détails sur la structure. D'autres paramètres comme le temps d'exposition du cristal dans le faisceau sont aussi importants car le cristal subit des dommages radiatifs, et ceux-ci détruisent petit à petit le cristal durant la collecte.

Les éléments de symétrie du groupe d'espace guident alors la stratégie de la collecte. La complétude définit le pourcentage de réflexions uniques mesurées et le rapport entre le nombre de réflexions totales et de réflexions uniques s'appelle multiplicité. Ces paramètres tabulés, permettent au cristallographe de juger de la qualité des données et donc de la pertinence de la construction moléculaire qui va en découler.

La collecte à l'ESRF se fait sous un flux d'azote afin de minimiser les mouvements atomiques des protéines dans le cristal, et augmenter de ce fait la résolution. Pour éviter la formation de cristaux de glace qui détruisent la symétrie du réseau de la protéine, les cristaux de protéines sont plongés préalablement dans une solution cryoprotectrice (glycerol, PEG, huile minérale etc). Le choix de ce cryoprotectant peut donc être déterminant.

Lorsque l'obtention de données à très haute résolution est envisageable, comme ce fut le cas pour le complexe PA-IIL/fucose à 1,0 Å de résolution, il est alors nécessaire d'effectuer deux collectes sur le même cristal : une à faible résolution et une autre à haute résolution. En effet, la limite de résolution augmente lorsque l'écran du détecteur se rapproche du cristal diffractant. Il

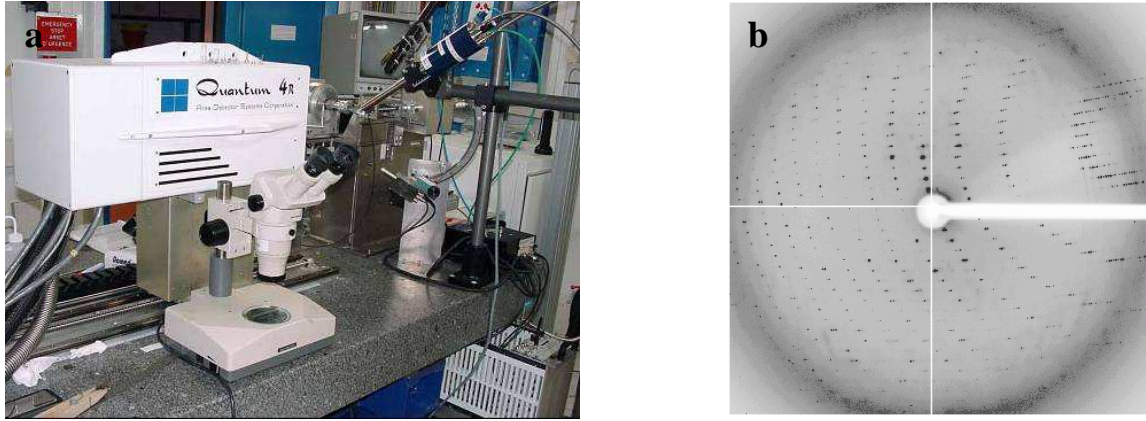
est donc dans notre intérêt de rapprocher le détecteur au maximum. Cependant, si les clichés offrent de meilleures données à haute résolution, ceux-ci sont du même coup, pour des raisons techniques, saturés pour les taches de faible résolution. Il est donc impératif de faire une autre collecte à faible résolution, donc à une distance plus éloignée. Les deux collectes sont ensuite rassemblées et mises à l'échelle.

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

Image par image, les données sont indexées à l'aide de programmes informatiques tels que Mosflm (Leslie, 1992) ou Denzo (Otwinowski and Minor, 1997). La détermination de la maille du cristal est améliorée par l'ajustement de certains paramètres comme la distance cristal-détecteur et la position du faisceau. Un profil de diffraction propre à chaque groupe d'espace est alors prédit; puis ajusté aux taches observées sur le cliché. Chaque tache se voit attribuer les indices de Miller hkl correspondant à la maille déterminée. L'intensité de chaque tache de diffraction est mesurée en prenant soin de retrancher le bruit de fond, puis l'ensemble des images intégrées est mis à l'échelle afin de constituer un jeu de données complet et homogène. En effet, chaque image diffère par la dose de rayons X reçue et par la baisse du pouvoir de diffraction du cristal. Les intensités collectées proviennent parfois de différentes collectes. Afin de garantir la consistance des intensités d'un bloc de données à un autre, une mise à l'échelle est moyennée sur les réflexions équivalentes par symétrie ou collectées plusieurs fois. La qualité de cette mise à l'échelle peut alors être estimée par le paramètre  $R_{sym}$  entre les réflexions équivalentes sur l'ensemble du jeu de données.

$$R_{sym} = \frac{\sum_h \sum_i |I_i(h) - \langle I(h) \rangle|}{\sum_h \langle I(h) \rangle} \quad (\text{Eq.1})$$

où  $I_i(h)$  est la  $i$ -ème mesure et  $\langle I(h) \rangle$  la moyenne pondérée de toutes les mesures  $I_i(h)$  des réflexions équivalentes.



**Figure 2.5.** (a) : détecteur ADSC Quantum 4 CCD de la cabine de collecte ID14-EH2 à l'ESRF de Grenoble. (b) : cliché de diffraction de la protéine PA-IIL, (groupe d'espace  $P2_1$ .)

### 2.1.5. Facteur de structure et densité électronique

Chaque rayon diffracté produisant une réflexion sur le détecteur est une onde qui est la somme des diffractions contributives de tous les atomes de la maille. La fonction mathématique qui la décrit est le facteur de structure. Pour une maille contenant  $n$  atomes, ce facteur de structure  $F_{hkl}$  est la somme de tous les facteurs de structure atomique  $f_{hkl}$  des atomes individuels et s'écrit simplement comme la série de Fourier où chaque terme est la contribution d'un seul atome:

$$F_{hkl} = \sum_j f_{hkl} , \quad f_{hkl} = f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad (\text{Eq. 2})$$

où  $f_i$  est une fonction mathématique assimilant l'atome à une simple sphère de densité électronique. Elle diffère pour chaque atome puisque tous n'ont pas le même nombre d'électrons, ce qui définit donc l'amplitude de leur contribution.  $(x_j, y_j, z_j)$  sont les coordonnées de l'atome  $j$  dans la maille, établissant de ce fait la phase de leur contribution.  $h, k$  et  $l$ , sont les indices de Miller de la réflexion dans l'espace réciproque.

La représentation sphérique et ponctuelle des atomes dans la maille ne rend pas compte de la réalité puisqu'il s'agit en fait de nuages électroniques. Pour plus d'exactitude, le facteur de structure doit être considéré comme la contribution d'éléments volumiques infiniment petits  $dV$ , centré en  $(x,y,z)$  de densité électronique volumique  $\rho(x,y,z)$ . Ce qui équivaut à réécrire la somme discrète  $F_{hkl}$  sous la forme intégrale suivante:

$$F_{hkl} = \int_V \rho(x, y, z) e^{2\pi i(hx+ky+lz)} dV \quad (\text{Eq. 3})$$

En appliquant la transformée de Fourier, on exprime alors la densité électronique comme une fonction des facteurs de structure  $F_{hkl}$

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad (\text{Eq. 4})$$

La densité électronique est la transformée de Fourier des facteurs de structure, ce qui signifie que l'on peut théoriquement convertir les données cristallographiques et faire la reconstitution spatiale du contenu de la maille de cristal. Malheureusement, dans l'équation précédente une information primordiale est manquante pour chaque facteur de structure. Le facteur de structure se caractérise par son amplitude et sa phase. Expérimentalement, la fréquence est connue, l'amplitude est accessible, elle est proportionnelle à  $(I)^{1/2}$  ( racine carré de l'intensité mesurée sur la réflexion), mais la phase est inconnue. Ces phases sont nécessaires pour construire la carte de densité électronique. Pour les obtenir, plusieurs alternatives sont possibles, elles sont développées dans le paragraphe suivant.

#### 2.1.6. *Obtention des phases*

Différentes méthodes sont utilisées pour obtenir les phases. Beaucoup d'entre elles sont des méthodes dites « physiques » basées sur l'utilisation de métaux, telles que les méthodes MIR, SAD, MAD qui sont expliquées plus en détail en annexe 1 de la thèse. Nous développerons ici la méthode « semi-directe » à l'aide du logiciel ACORN, ainsi que le remplacement moléculaire.



La structure à 1 Å de résolution du complexe PA-IIL/fucose a été résolue en utilisant le logiciel ACORN (Yao et al., 2006). Ce logiciel fut développé à l'origine comme un outil de recherche de phases ab initio lorsque des données de très haute résolution étaient disponibles (<1.3 Å). La méthode consiste dans un premier temps à obtenir un ensemble de phases légèrement meilleures que celles obtenues de manière aléatoire. Celles-ci peuvent être calculées à partir d'un fragment de la structure, qu'il s'agisse d'un atome unique de métal, d'un petit fragment (1-8% de la molécule) ou d'un modèle complet de protéine. Comme pour un remplacement moléculaire, des recherches spatiales peuvent être effectuées afin d'orienter un fragment de molécule et fournir les phases initiales. Ces phases permettent de tracer une première carte qui est ensuite itérativement affinée par un procédé dynamique de modification de la densité électronique (DDM). Pour la structure PA-IIL/fucose à 1 Å de résolution, seules les positions de cinq atomes de calcium issus d'une structure préalable à 1,3 Å, ont été utilisées.

La philosophie de la DDM est la suivante : en dehors de la région occupée par le fragment, plus la densité dessinée par la carte est forte plus la probabilité que cette position soit effectivement occupée par la véritable densité électronique de la structure est élevée. Ceci conduit à une modification de la densité électronique  $\rho$  basée sur la déviation standard  $\sigma$  de la carte de densité telle que :

$$\rho' = \begin{cases} 0 & \text{si } \rho \leq 0 \\ \rho \tanh [0,2(\rho/\sigma)^{3/2}] & \text{si } \rho > 0 \\ kn\sigma & \text{si } \rho' > kn\sigma \end{cases}$$

où  $k$  est une constante donnée par l'utilisateur (valeur par défaut 3) et  $n$  est le nombre de cycles d'itération de la DDM (5 au maximum). Une densité négative, (une valeur non physique) est fixée à zéro. Les plus faibles densités, qui ont une moindre probabilité de correspondre à la vraie densité de la protéine, sont atténuées par rapport aux densités plus élevées. Les densités les plus hautes sont tronquées afin que le fragment ne biaise les données. Cette valeur augmente au fur et à mesure que s'améliore la densité, ce qui évite d'enlever l'information structurale nouvellement acquise. L'affinement des phases ne requiert aucune intervention du cristallographe.

Le coefficient de corrélation (CC) rend compte de l'écart entre les facteurs de structure calculés  $F_c$  et ceux observés  $F_o$ .

$$CC = \frac{\langle |FcFo| \rangle - \langle |Fc| \rangle \langle |Fo| \rangle}{\sigma(Ec)\sigma(Eo)}, \quad \text{où} \quad \sigma = \left( \langle |F^2| \rangle - \langle |F| \rangle^2 \right)^{1/2} \quad (\text{Eq. 5})$$

### 2.1.7. Le remplacement moléculaire

Il est parfois possible d'utiliser comme phase de départ, les phases provenant d'une autre molécule. En effet, certaines protéines ou familles de protéines ont des séquences peptidiques similaires. Ces similarités de séquence impliquent le plus souvent une similarité de la structure spatiale. Un minimum de 25 % de similarité de séquence est requis pour avoir une bonne chance de succès. Dans le cas de la lectine PA-IIL, si le seul changement intervient dans le site de reconnaissance de la protéine (mutation ou ligand), le reste de la structure est inchangé. En pratique, le remplacement moléculaire procède de la façon suivante. A l'aide de la fonction de Patterson (cf. annexe1), des logiciels tels que Molrep (Vagin and Teplyakov, 1997) ou Amore (Navaza, 1994) cherchent d'abord l'orientation de la molécule, puis des translations sont effectuées afin de trouver la position de la molécule dans la maille. Compte tenu du nombre de paramètres à considérer (trois pour les rotations autour des axes centrés sur la molécule (x, y, z) et trois encore pour les translations dans la maille), le processus en deux étapes est beaucoup plus rapide qu'une recherche simultanée.

### 2.1.8. Affinement et vraisemblance du modèle

A partir des phases, il est possible de tracer une première carte de densité électronique  $F_{\text{obs}}$  et de la visualiser à l'aide de logiciel graphique tel que O (Jones et al., 1991) et plus récemment Coot (Emsley and Cowtan, 2004) (fig. 2.6). Ces phases initiales ne sont pas parfaites, ce qui engendre une carte de densité partiellement fautive ou incomplète. Le cristallographe peut néanmoins construire un premier modèle moléculaire, générant alors un nouvel ensemble de phases calculées, permettant le traçage d'une nouvelle carte améliorée. Le processus cherche à minimiser les différences entre les facteurs de structure  $F_{\text{calc}}$  calculés à partir de son modèle, et les facteurs de structure observés  $F_{\text{obs}}$ . Des cartes de densité peuvent être calculées : la carte (Fo-Fc) et la carte (2Fo-Fc) d'après l'équation suivante

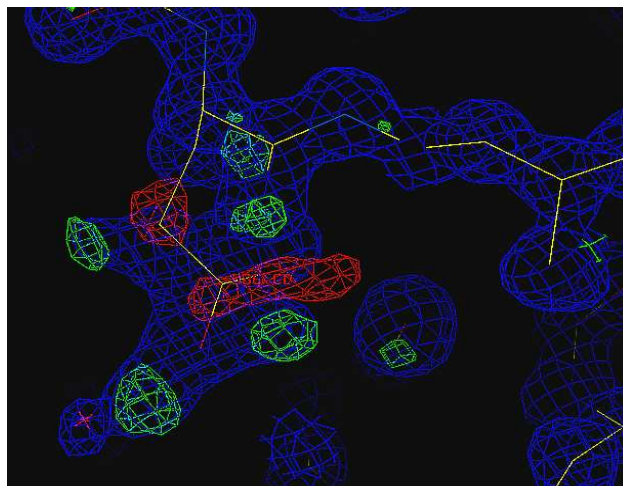
$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} (nF_o - F_C) e^{-2\pi i(hx+ky+lz)}, \quad \text{avec n entier (Eq.6)}$$

Ces cartes permettent de faire des ajustements sur le modèle, et sont le plus souvent effectuées à la main à travers l'interface graphique.

En pratique, le cristallographe alterne modifications manuelles et traitements informatiques à l'aide de logiciels d'affinement comme Refmac (Murshudov et al., 1997), utilisant un algorithme de type statistique nommé le maximum likelihood. Cet algorithme qui maximise la probabilité que le facteur de structure observé  $F_{obs}$  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. Pour juger de la convergence de l'affinement, le facteur d'accord cristallographique R est défini par :

$$R = \frac{\sum_{hkl} \left| |F_{obs}(hkl)| - |F_{calc}(hkl)| \right|}{\sum_{hkl} |F_{obs}(hkl)|} \quad (\text{Eq. 7})$$

La valeur de R indique la progression de l'affinement de la structure. Plus cette valeur est basse, plus le modèle est en accord avec les données. Afin de ne pas suraffiner la structure, un second facteur  $R_{libre}$  se calcule de la même façon que le facteur R conventionnel, mais à partir d'une fraction de réflexions (normalement 500 à 1000 réflexions qui typiquement représentent de 3 à 5% du jeu de données) non prise en compte durant l'affinement. En fin d'affinement, les paramètres structuraux de la protéine sont vérifiés et comparés aux valeurs typiques des structures protéiques avec le logiciel Whatif (Vriend, 1990). La structure peut alors être déposée dans la base de données internationales : RCSB Protein Data Bank.



**Figure 2.6.** Visualisation graphique de la carte de densité électronique. La carte  $2Fo-Fc$  est représentée en bleu, la carte  $Fo-Fc$  négative est en rouge signifiant un excès de densité calculé à cet emplacement. La carte  $Fo-Fc$  positive est en vert et signifie un manque de densité.

## 2.2. La microcalorimétrie de titration isotherme (ITC)

### 2.2.1. Quelques rappels thermodynamiques

L'interaction non covalente entre une macromolécule M et son Ligand L, donnant lieu à la formation d'un complexe ML peut s'écrire comme la réaction réversible suivante :



La propension que cette réaction se fasse dans le sens de la formation du complexe se calcule à l'aide de la constante d'affinité  $K_a$  se définissant comme le rapport des concentrations entre la forme complexée sur les formes libres des différentes molécules. Ainsi  $K_a$  s'écrit :

$$K_a = \frac{[ML]}{[M][L]} \quad (\text{Eq. 8})$$

Comme toute interaction moléculaire, celle-ci donne lieu à des échanges d'énergie qui sont fonction de l'état initial (molécules libres) et final (complexe formé) de la réaction, ainsi que du

chemin réactionnel suivi entre les deux états. D'un point de vue thermodynamique, la constante d'affinité  $K_a$  est fonction de l'énergie libre de l'interaction (énergie de Gibbs) et s'écrit aussi de la manière suivante :

$$K_a = e^{-\Delta G / RT} \quad , \text{ où } T \text{ est la température et } R = 8,31 \text{ J mol}^{-1} \text{ K}^{-1} \quad (\text{Eq. 9})$$

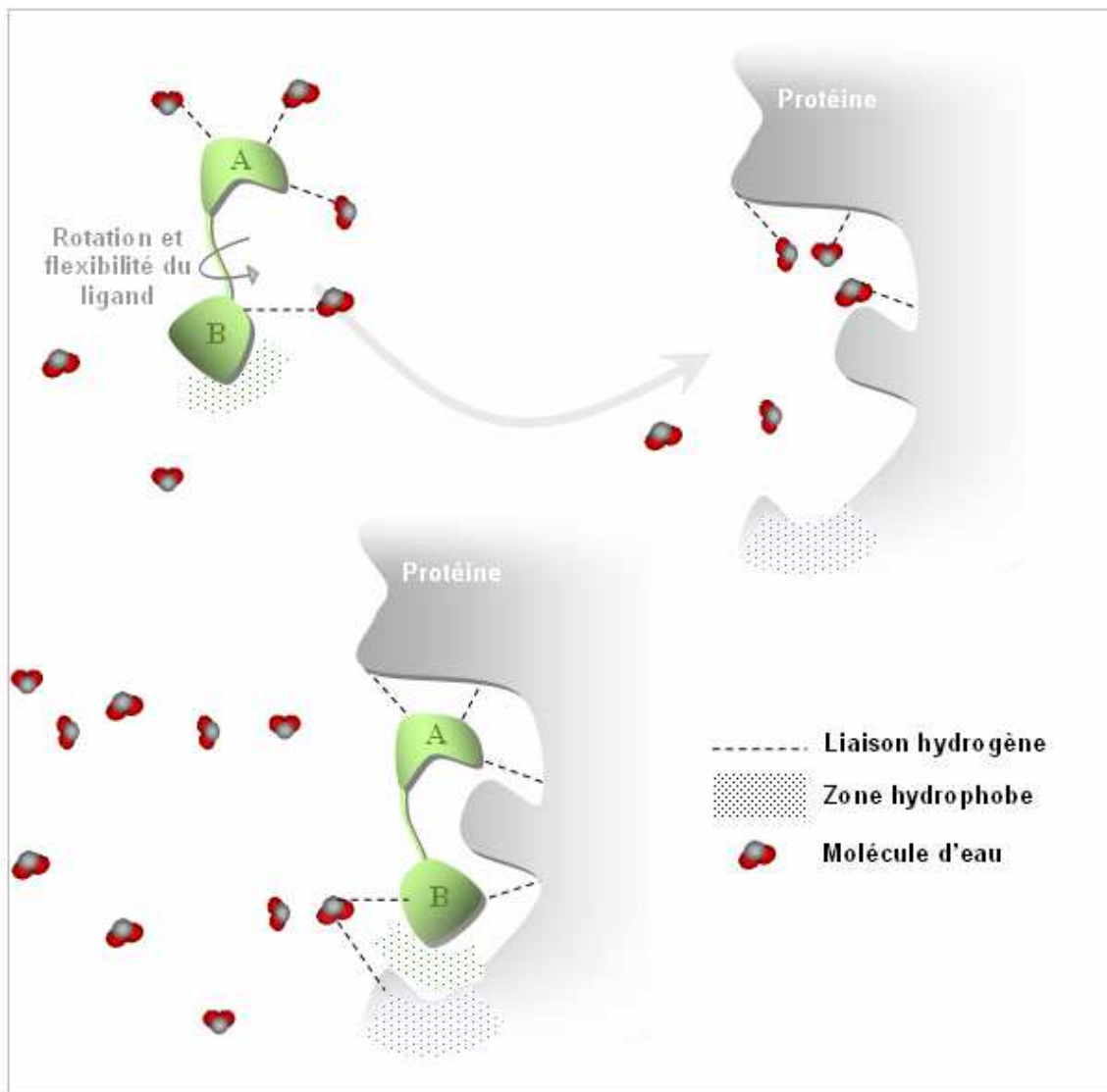
Plus cette constante est grande, plus l'affinité du ligand pour la protéine est forte. La réaction est spontanée lorsque  $\Delta G$  est négatif. Cette énergie est en fait la résultante de deux contributions énergétiques : l'enthalpie  $\Delta H$  et l'entropie  $\Delta S$ ,

$$\Delta G = \Delta H - T\Delta S \quad (\text{Eq.10})$$

D'un point de vue très général, l'enthalpie est une grandeur thermodynamique rendant compte de l'énergie échangée par un système (qu'elle soit sous forme thermique ou mécanique). Tout système ayant tendance à avoir l'énergie la plus basse possible,  $\Delta H$  tend à être négatif. L'entropie fait référence au degré de désordre d'un système. Les mouvements Brownien sont responsables de l'augmentation de l'entropie. Cet effet augmente avec la température.  $T\Delta S$  tend à être positif. Dans le cadre plus restrictif des interactions moléculaires, ces deux grandeurs thermodynamiques correspondent à des mécanismes d'interactions différents.

### 2.2.2. Les contributions enthalpique ( $\Delta H$ ) et entropique ( $\Delta S$ )

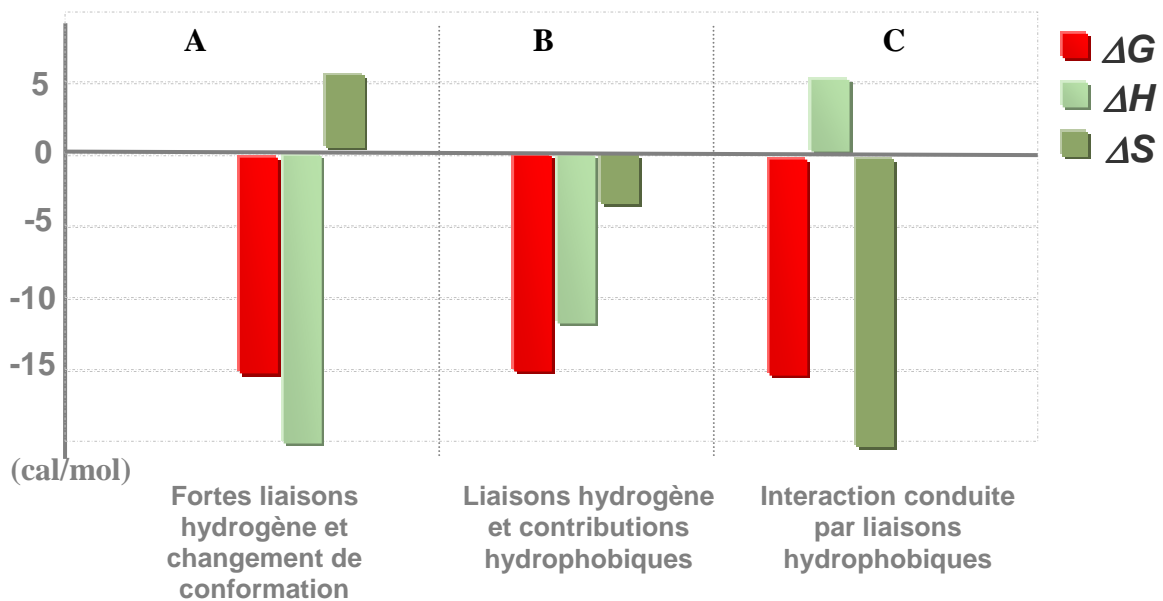
La contribution enthalpique reflète la force des interactions qu'entretient le ligand avec la protéine par rapport à celles entretenues avec le solvant (fig. 2.7) Ces liaisons sont principalement dues à la formation de liaisons hydrogène et aux interactions de Van der Waals. Une enthalpie d'interaction 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.



**Figure 2.7.** Réaménagement du solvant lors d'une interaction. En absence de ligand les molécules d'eau établissent des liaisons hydrogènes dans le site de reconnaissance. Ces molécules sont chassées lorsque vient se positionner la partie A du ligand, ce qui représente un gain entropique et un gain enthalpique. La rigidité occasionnée lorsque la partie B du ligand rentre dans le site, ainsi que la capture d'une molécule d'eau représentent une entropie défavorable.

La contribution entropique peut provenir de deux phénomènes : le premier est dû à un réaménagement du solvant qui s'opère autour de la protéine lors de la reconnaissance. Bien souvent cette entropie de solvation est favorable lorsque les surfaces mises en jeu dans la reconnaissance sont hydrophobiques. En effet, les molécules d'eau présentes dans les premières sphères de solvation de la protéine et/ou du ligand sont en quelque sorte, chassées par le

rapprochement des surfaces hydrophobes et relarguées dans la solution. Le deuxième phénomène est dû à une diminution des degrés de liberté du ligand ou de la protéine au moment de l'interaction. Cette contribution entropique est souvent défavorable. En effet, un ligand en solution peut adopter plusieurs conformations dues aux rotations autour des angles de torsion par exemple. L'interaction avec la protéine contraint le ligand dans une seule conformation, ce qui engendre une perte en entropie (fig. 2.8).



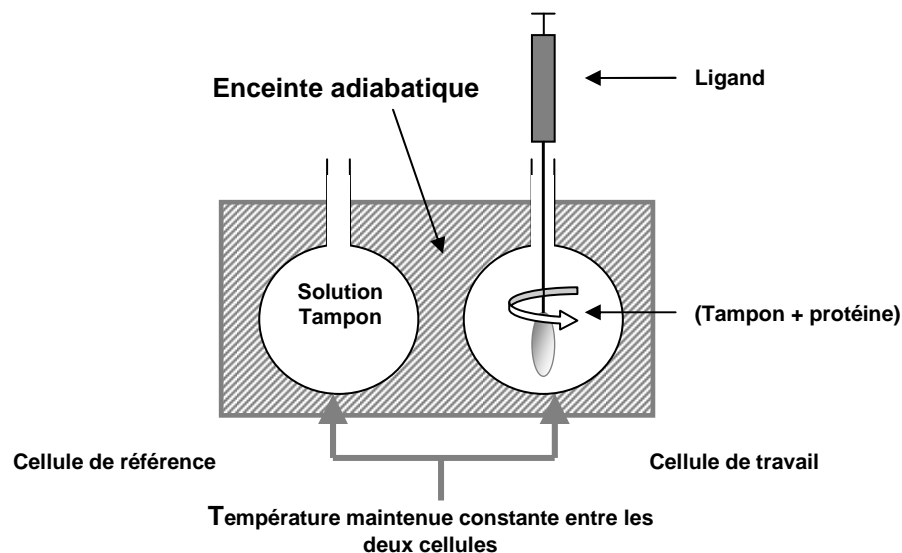
**Figure 2.8.** Trois différents types d'interactions souvent rencontrés. (Adapté d'après *Isothermal Titration Calorimetry and Drug Design, ITC Application Note, Microcal*). Les interactions protéine-sucre sont généralement dans le cas A.

### 2.2.3. Le microcalorimètre de titration isotherme (ITC) : principe

Parmi les différentes méthodes qui permettent d'étudier la thermodynamique de la reconnaissance entre un récepteur et son ligand (protéine-sucre, protéine-protéine ou autre), la microcalorimétrie de titration isotherme est la seule permettant d'obtenir tous les paramètres caractérisant l'interaction en une seule expérience, telle que la constante d'affinité ( $K_a$ ), la stœchiométrie d'interaction ( $n$ ) ainsi que les contributions enthalpiques ( $\Delta H$ ) et entropiques ( $T\Delta S$ ). De plus, contrairement à d'autres méthodes nécessitant l'immobilisation du ligand ou du récepteur,

pouvant parfois gêner l'interprétation des résultats, la microcalorimétrie laisse les molécules libres en solution.

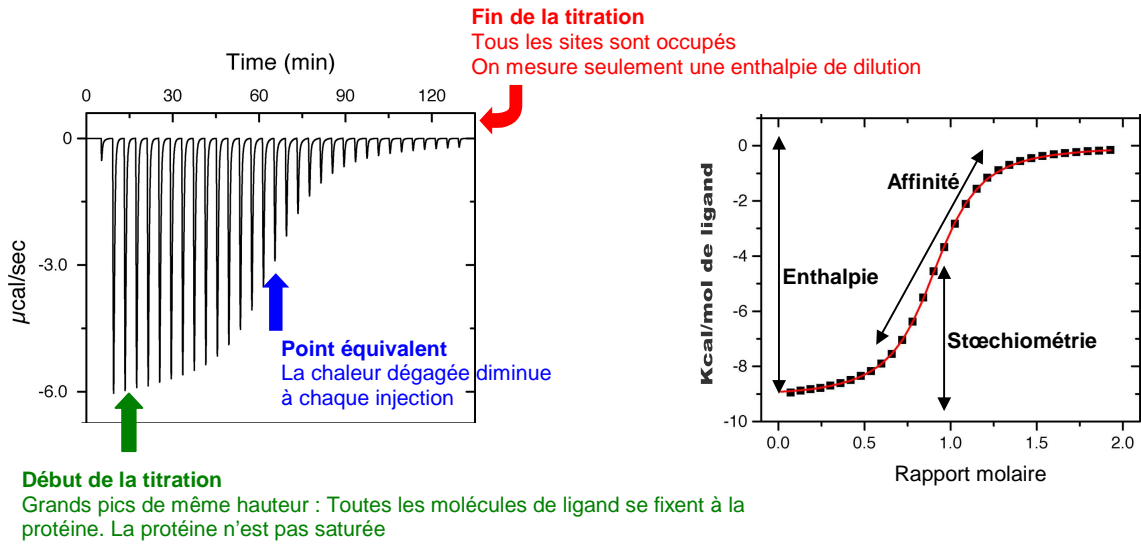
Le microcalorimètre (fig. 2.9) contient deux cellules isolées de l'environnement externe par une enceinte adiabatique qui est maintenue à température constante pendant le temps de la titration. Une cellule de référence contient du tampon uniquement. La même solution contenant la protéine est placée dans une deuxième cellule, dite de travail.



**Figure 2.9.** Microcalorimètre ITC

A intervalle régulier, des aliquotes de ligand sont ajoutés à l'aide d'une seringue. A chaque injection, l'interaction entre les deux molécules s'accompagne d'un dégagement ou d'une absorption de chaleur (réaction exothermique ou endothermique) 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. C'est ce signal qui est mesuré par le calorimètre. La puissance fournie est mesurée en fonction du temps, puis intégrée.





**Figure 2.10.** Données microcalorimétriques brutes (gauche), intégrées puis ajustées à la courbe sigmoïde théorique (droite)

Au début de la titration, des quantités importantes de chaleur sont dégagées, indiquant la formation du complexe protéine-ligand. 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. La variation d'enthalpie  $\Delta H$ , la constante d'association  $K_a$  et la stœchiométrie  $n$  peuvent être déterminées avec un ajustement de la courbe expérimentale sur le modèle théorique approprié (Wiseman *et al.*, 1989) (fig. 2.10). Il est possible d'évaluer la quantité de chaleur  $\Delta Q$  instantanée, qui est proportionnelle à la variation de concentration du complexe formé  $\Delta[ML]$ , le nombre de sites de liaison  $n$ , l'enthalpie de liaison  $\Delta H^0$  et le volume  $V_o$  de la cellule de travail tel que :

$$\Delta Q = n\Delta[ML]\Delta H^0 V_o \quad (\text{Eq. 12})$$

On peut déterminer la quantité de chaleur cumulée au cours de toutes les injections. Pour un modèle d'interaction dans lequel chaque site est indépendant, on obtient :

$$\frac{dQ}{d[L]_t} = \frac{\Delta H^0 V_o}{2} \left[ 1 + \left( \frac{1 - [L]_t/[M]_t - n/K_a[M]_t}{\sqrt{(1 + [L]_t/[M]_t + n/K_a[M]_t)^2 - 4[L]_t/[M]_t}} \right) \right] \quad (\text{Eq. 13})$$

où  $[X]_t$  et  $[M]_t$  sont les concentrations respectives totales du ligand et de la protéine,  $V_o$  est le volume effectif de la cellule de mesure et  $dQ$  est le variation de la quantité de chaleur échangée, calculée à chaque point de la courbe de titration. L'enthalpie  $\Delta H$  et la constante d'affinité  $K_a$  permettent de déduire tous les autres paramètres thermodynamiques. L'enthalpie  $\Delta H^\circ$  influence directement l'amplitude de la courbe, tandis que la position du point d'inflexion donne la stœchiométrie  $n$ . Par contre, la forme générale de la courbe dépend du paramètre de Wiseman  $c$  qui est défini comme

$$c = n[M_o]K_a$$

où  $n[M_o]$  représente la concentration initiale de sites de liaison en solution. La courbe doit se rapprocher de la forme sigmoïdale afin de déterminer au mieux les paramètres thermodynamiques. En pratique,  $c$  doit être compris entre 10 et 100 car des valeurs supérieures génèrent des courbes dont les pentes trop élevées, ne permettent pas une mesure fiable de  $K_a$ . Des titrations à différentes températures sont envisageables afin d'analyser la variation  $\Delta C_p$  (capacité calorifique) en fonction de la température.  $\Delta C_p$  est reliée à l'enthalpie par l'équation :

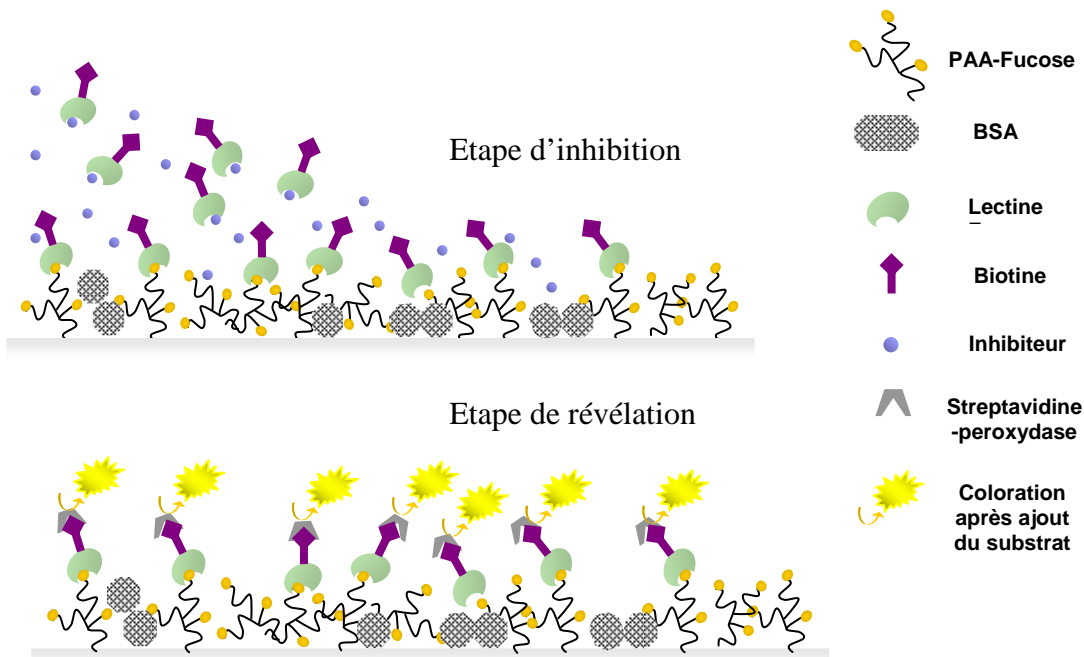
$$\left(\frac{\partial H}{\partial T}\right)_P = \Delta C_p$$

$\Delta C_p$  nous renseigne sur l'évolution des différentes contributions entropiques et enthalpiques en fonction de la température et reflète le comportement du solvant pendant l'association.

## 2.3. Les tests d'interactions

### 2.3.1. Les tests ELLA

Dans la « quête du Graal » que constitue la recherche de nouvelles substances pouvant constituer des molécules thérapeutiques, une des méthodes utilisée dans les reconnaissances lectine-sucre est la méthode ELLA (Enzyme Linked Lectin Assay) (fig. 2.11). Elle consiste à évaluer la faculté que possède une molécule donnée à inhiber l'interaction entre une protéine et une surface recouverte d'un ligand de référence. Dans notre cas, nous avons utilisé une surface de  $\alpha$ -L-fucose puisque il s'agit du ligand monosaccharide de référence dans l'étude de la lectine PA-III. Pour ce faire, des microplaques en polystyrène de 96 puits servent de support à l'adsorption (interactions non spécifiques de type hydrophobe) de poly[N-(2-hydroxyethyl)acrylamide] (PAA) contenant du fucose en ramification (Lectinity Holding, Inc). On obtient ainsi une surface recouverte de  $\alpha$ -L-fucose.



**Figure 2. 11.** Principe du test ELLA: Plus l'inhibiteur (en bleu) est efficace, moins la lectine se fixe sur la surface recouverte du polymère de fucose (en orange). Des lavages éliminent les lectines non fixées sur la surface. La BSA (bovin sérum albumine) empêche les interactions non spécifiques avec la plaque plastique.

Différentes concentrations en inhibiteurs sont alors testées pour inhiber l'interaction de la lectine avec la surface. Les courbes obtenues permettent d'estimer l'IC50 (concentration en inhibiteur donnant 50% d'inhibition). Pour évaluer la quantité de lectine fixée sur la surface, la protéine utilisée a été préalablement biotinylée. La visualisation s'effectue grâce à l'interaction de la streptavidine couplée à une peroxydase qui catalyse la formation d'un composé coloré à partir d'un substrat commercial. Ces tests peuvent aussi s'effectuer en inversant le sens des interactions, c'est-à-dire en recouvrant la surface du puits avec la lectine et en mettant en solution un sucre biotinylé de référence en compétition avec des inhibiteurs.

### 2.3.2. Les « glycans array »

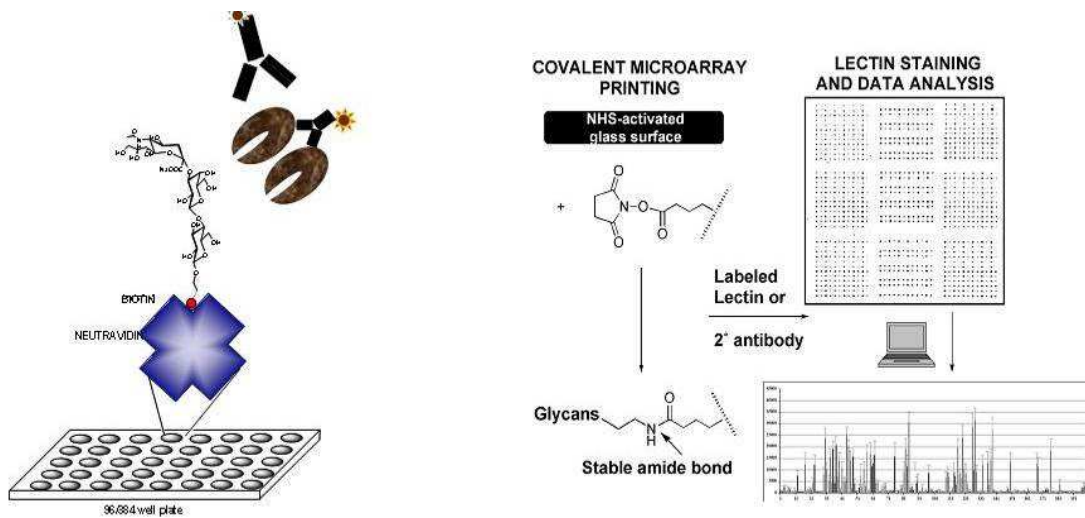
Connaître la ou les cibles d'une lectine est un point clé dans la recherche de sa spécificité et de son rôle biologique. Jusqu'à récemment, la grande complexité des interactions protéine-glycane, l'absence de large bibliothèque de glycane lors des tests effectués et l'hétérogénéité des méthodes, ne permettaient pas une bonne évaluation de la sélectivité des glycoprotéines. Pour palier cette faiblesse, le « Consortium for Functional Glycomics » (CFG), fondé par le « National Institute of General Medical Science » a mis au point de nouveaux outils de routine afin de comprendre le rôle des interactions protéine-carbohydre dans la communication cellule-cellule (voir [http://www.functionalglycomics.org/static/consortium/.](http://www.functionalglycomics.org/static/consortium/))

Les premiers tests proposés sont basés sur le principe ELISA. Ils comprennent une bibliothèque de ligands biotinylés fixés à l'aide de bras espaceurs sur des plaques recouvertes de streptavidine. La miniaturisation des systèmes permet de travailler sur des plaques de 384 puits. Les lectines doivent être soit marquées par un composé fluorescent, soit fournies avec un anticorps (fig. 2.12).

### 2.3.3. Les puces à sucre ou « printed arrays »

Les méthodes plus récentes proposées par Consortium for Functional Glycomics et par d'autres laboratoires dans le monde reposent sur le principe des puces à sucre ou « Printed covalent glycan array ». Elles utilisent des plaques commerciales de verre fonctionnalisées avec du N-

hydroxysuccimidine (NHS) capable de se lier de manière covalente avec des glycanes ou des glycoconjugués, à condition que ceux-ci soient couplés à un groupement amine (Blixt *et al.*, 2004).



**Figure 2. 12.** Principe des test micro-arrays (à gauche) et printed arrays (à droite)

Cette étape étant relativement facile, la bibliothèque de glycanes fonctionnalisés sur la plaque compte aujourd’hui plus de 200 structures diverses présentes sur les parois cellulaires, les glycoprotéines et les glycolipides. La technique nécessite également que la protéine soit marquée à l’aide d’un fluorochrome (dans notre cas le tétrafluorophényle (TFP) Alexa fluor 488) ou soit fournie avec un anticorps.

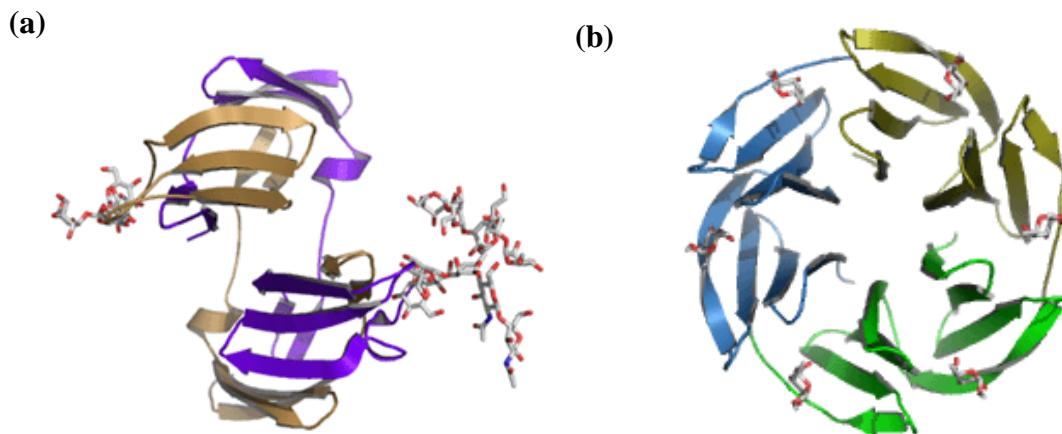
# 3

## PA-IIL : interaction avec des monosaccharides

---

### 3.1. Les lectines bactériennes solubles

Les lectines solubles sont par définition toutes les protéines solubles exprimées par la bactérie, se liant spécifiquement à des glucides et ne montrant pas d'activité enzymatique. Elles peuvent être cytoplasmiques ou sécrétées à l'extérieur de la bactérie. Jusqu'à aujourd'hui, peu de structures de lectines bactériennes solubles ont été caractérisées. Parmi celles-ci, la lectine *Cyanovirin-N* isolée de la cyanobactérie *Nostoc ellipsosporum* (environ 11 kDa) présente un intérêt particulier par son activité anti-HIV car elle possède une affinité nanomolaire pour les oligosaccharides mannosylés Man<sub>8</sub> et Man<sub>9</sub> de la glycoprotéine gp120 qui forme l'enveloppe extérieure du virus (Boyd *et al.*, 1997, Bewley, 2001 #149) (fig. 3.1). Une autre bactérie pathogène, *Ralstonia solanaceanum* qui s'attaque aux racines des plantes, produit une lectine RS-L spécifique du fucose et des xyloglucanes fucosylés présents dans la paroi végétale (fig. 3.1). Sa structure quaternaire est une association de trois monomères. L'assemblage de ces sous-unités forme une structure de type  $\beta$ -propeller à six lames (Kostlanova *et al.*, 2005). Enfin, la bactérie *Pseudomonas aeruginosa* produit deux lectines solubles PA-IL et PA-IIL dont nous avons vu précédemment qu'elles étaient associées aux facteurs de virulence lors des infections survenant chez les patients atteints de mucoviscidose. Des séquences homologues à PA-IIL ont été retrouvées chez *Ralstonia solanaceanu*, *Chromobacterium violaceum* (Zinger-Yosovich *et al.*, 2006) ainsi que chez les bactéries du complexe *Burkholderia cepacia*.

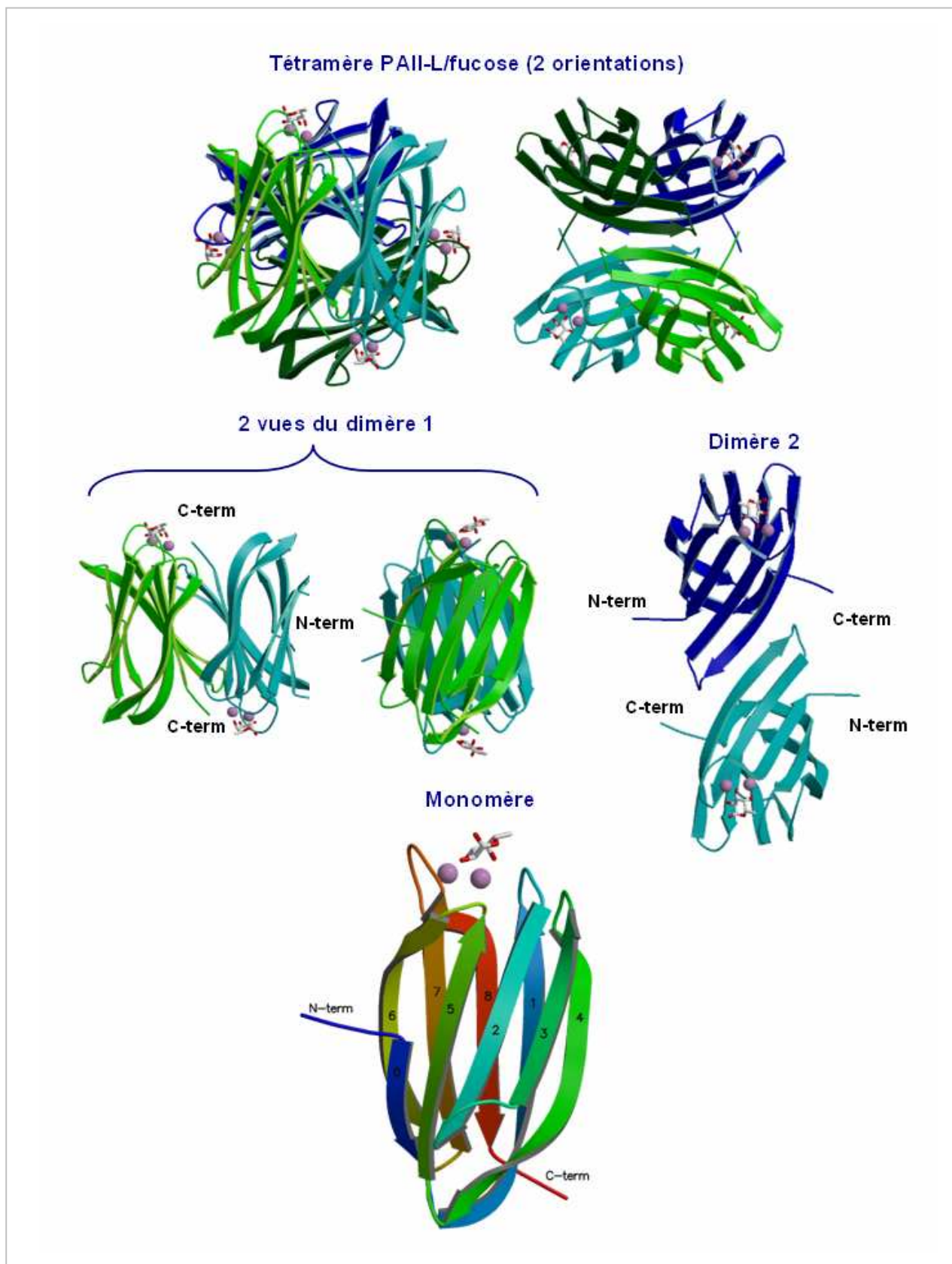


**Figure 3.1.** (a) dimère de la lectine Cyanovirin-N complexée avec du polymannose Man<sub>9</sub> (pdb 1M5M) (b) trimère de la lectine RS-L complexée avec du fucose (pdb 2BS5)

## 3.2. Structure de la lectine PA-III

### 3.2.1. La structure quaternaire

La structure cristalline de la lectine PA-III purifiée de *P. aeruginosa* a été résolue en 2002 (Mitchell *et al.*, 2002) (fig. 3.2). La forme active de la lectine est un homotétramère constitué de quatre sous-unités indépendantes arrangées autour d'un pseudo axe de symétrie C<sub>222</sub>. Chaque monomère adopte un repliement du type sandwich à 9 brins  $\beta$  qui s'agencent en *clé grecque*. La dimérisation résulte d'une association « tête à queue » de deux monomères interagissant par l'intermédiaire de feuillets- $\beta$ . Pour chaque dimère, et de manière insolite dans les interactions protéine-sucre, l'extrémité C-terminale de chaque monomère (résidu Gly114) est localisée dans le site de liaison du monomère voisin et participe à la reconnaissance du sucre. Deux dimères forment un tétramère par l'association antiparallèle de l'un de leurs feuillets avec le feuillet homologue du dimère voisin.

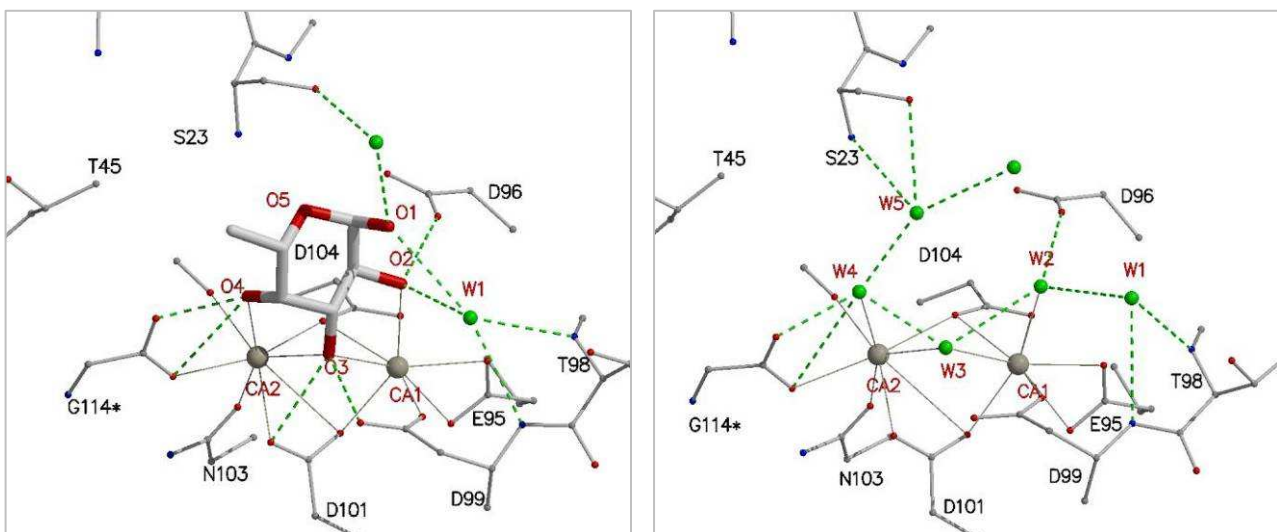


**Figure 3.2.** Structure de la lectine PA-II représentée en ruban. Les atomes de calcium sont représentés par des sphères roses, le fucose par des bâtons rouges et bancs)



### 3.2.2. Le site de reconnaissance

La structure à 1.3 Å de résolution de la lectine co-cristallisée avec du fucose révèle un mode de reconnaissance unique parmi les interactions protéine-sucre (Mitchell, Houles *et al.*, 2002). Le site de reconnaissance est une poche acide constituée principalement de deux asparagines et de quatre acides aspartiques dont les chaînes latérales sont capables de coordonner les deux atomes de calcium distants de 3.8 Å. L'extrémité C-terminale du monomère adjacent participe également à la coordination d'un des deux atomes de calcium. Les deux atomes de calcium ont une coordination classique de huit ligands avec dix atomes d'oxygène de la protéine, provenant principalement des résidus aspartate. Une boucle de coordination (acides aminés 95-103) est essentiellement impliquée dans la fixation des atomes de calcium. Cette boucle est partiellement désordonnée en absence de calcium dans la poche acide (Loris *et al.*, 2003).



**Figure 3.3.** Site de reconnaissance de la lectine PA-IIL. (a) Complexe avec le L-fucose, (b) structure native de PA-IIL. L'extrémité C-terminal du monomère adjacent est marquée d'un astérisque. Les liaisons de coordination sont représentées en traits pleins gris, les liaisons hydrogène sont en vert pointillé, les molécules d'eau (W) sont représentées par des sphères vertes.

Trois des groupes hydroxyle du fucose participent à la coordination des atomes de calcium : Les oxygènes O2 et O4 participent à la coordination d'un atome de calcium chacun, alors que l'oxygène O3 intervient sur les deux atomes de calcium. En plus de ces interactions

électrostatiques, ces trois mêmes groupes hydroxyle participent à plusieurs liaisons hydrogène avec les groupes acides de la boucle de coordination. Le groupement méthyle en position 6 du fucose fait une interaction favorable avec une petite zone hydrophobe de la protéine. En l'absence de fucose, quatre molécules d'eau remplacent les oxygènes O2, O3, O4 des groupes hydroxyle et O5 du cycle pyranose (Loris *et al.*, 2003) (fig. 3.3).

### 3.3. Bases moléculaires et thermodynamiques de la haute affinité

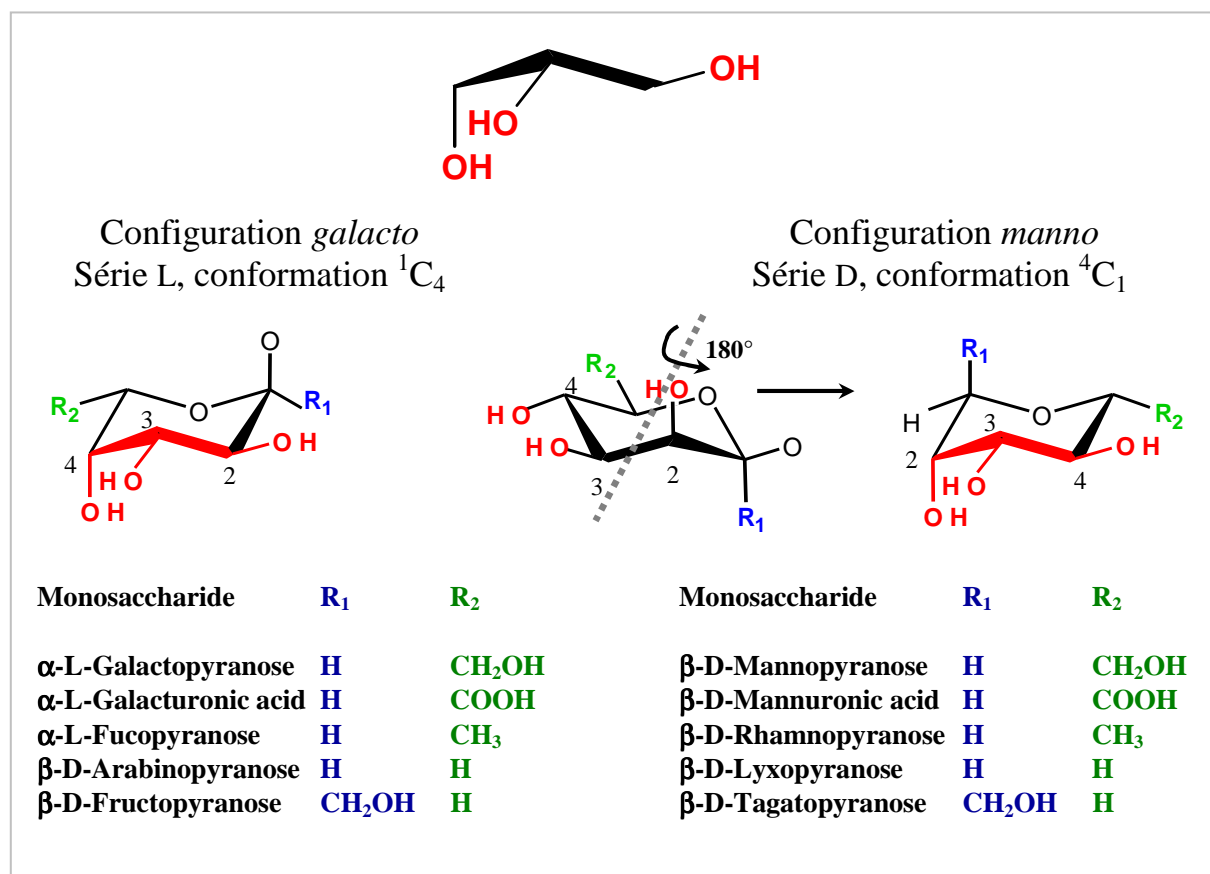
(Article n°1)

Des premiers tests d'inhibition d'hémagglutination et de dialyse à l'équilibre ont révélé une affinité micromolaire de la lectine PA-III pour le L-fucose (Garber *et al.*, 1987). Afin de comprendre les raisons de cette forte affinité pour le fucose, la structure de la lectine PA-III cristallisée avec du fucose et du calcium a été déterminée à la résolution de 1.0 Å (Mitchell *et al.*, 2005). Grâce à ces données de résolution atomique, le positionnement des atomes d'hydrogène dans le site de reconnaissance, et le réseau de liaisons hydrogène ont été modélisés. Ainsi, le rôle stabilisateur d'une molécule d'eau particulière pontant le fucose avec la chaîne principale de la protéine a été identifié. Le calcul des charges partielles, en utilisant des méthodes informatiques *ab initio* a révélé une forte délocalisation des charges entre les atomes de calcium, le fucose et les chaînes latérales aspartate et glutamate. Cette approche a rationalisé le rôle très spécifique des deux ions calcium dans ce mode de reconnaissance inhabituel. Ils fournissent une délocalisation des charges, et sont responsables d'une forte contribution enthalpique d'interaction. De plus, leur désolvatation lorsque le ligand vient se fixer, fournit une contribution entropique favorable, rarement observée lors d'interactions protéine-sucre.

### 3.4. Spécificité pour d'autres monosaccharides (Article n°2)

Les tests d'inhibition ont démontré que plusieurs monosaccharides pouvaient également être reconnus par PA-III avec une affinité décroissante L-fucose (Fuc) > L-galactose (L-Gal) > D-arabinose (Ara) > le D-fructose (Fru) > D-mannose (Man) (Garber, Guempel *et al.*, 1987). Tous

ces monosaccharides présentent soit une configuration *L-galacto*, soit une configuration *D-manno*. Toutes deux présentent deux groupements hydroxyle équatoriaux et un groupement hydroxyle axial lorsque le cycle pyranose est respectivement dans une conformation  ${}^1C_4$  et  ${}^4C_1$ . Cette combinaison d'orientations des groupements hydroxyle s'avère indispensable pour la coordination des deux ions calcium.



**Figure 3.4.** Représentation graphique des différents monosaccharides contenant le motif structural impliqué dans la coordination des atomes de calcium. Les groupes hydroxyle impliqués sont surlignés en rouge. Une rotation de  $180^\circ$  des monosaccharides de la série *D-manno* positionne les groupes OH dans la même orientation que ceux de la série *L-galacto* (Sudakevitz et al., 2004)

Les structures cristallines, déterminées pour des complexes avec du D-mannose et du D-fructose confirment les bases moléculaires de la large spécificité de PA-III (Loris et al., 2003). Nous avons souhaité comparer les modes de liaison des trois monosaccharides ayant l'affinité la plus

élevée pour PA-IIL, à savoir : les L-Fuc, L-Gal et D-Ara, qui diffèrent uniquement par la nature du groupe situé en C5. Deux structures cristallines additionnelles ont été déterminées à haute résolution pour les complexes avec le L-galactose et le D-arabinose. La constante d'affinité ainsi que les grandeurs thermodynamiques associées à leur interaction ont été évaluées puis corrélées avec les données structurales existantes. Le  $\alpha$ -L-Fuc est le meilleur ligand en raison d'une interaction hydrophobe due au groupement methyl en C5 et à un pont hydrogène faisant intervenir l'oxygène O1 du fucose (cf. article 2).



(Article n°1)

**High affinity fucose binding of *Pseudomonas aeruginosa* lectin PA-IIL:  
1.0 Å resolution crystal structure of the complex combined with  
thermodynamics and computational chemistry approaches**

Mitchell EP, **Sabin C**, Snajdrova L, Pokorna M, Perret S, Gautier C,  
Hofr C, Gilboa-Garber N, Koca J, Wimmerova M, Imberty A.

*PROTEINS : Structure, Function, and Bioinformatics* 58(3): 735-46 (2005).



# High Affinity Fucose Binding of *Pseudomonas aeruginosa* lectin PA-IIL: 1.0 Å Resolution Crystal Structure of the Complex Combined with Thermodynamics and Computational Chemistry Approaches

Edward P. Mitchell,<sup>1</sup> Charles Sabin,<sup>1,2</sup> Lenka Šnajdrová,<sup>2,3</sup> Martina Pokorná,<sup>3</sup> Stéphanie Perret,<sup>2</sup> Catherine Gautier,<sup>2</sup> Ctirad Hofr,<sup>4</sup> Nechama Gilboa-Garber,<sup>5</sup> Jaroslav Koča,<sup>3</sup> Michaela Wimmerová,<sup>3,6</sup> and Anne Imberty<sup>2,\*</sup>

<sup>1</sup>E.S.R.F. Experiments Division, Grenoble, France

<sup>2</sup>CERMAV - CNRS (affiliated with Université Joseph Fourier), Grenoble, France

<sup>3</sup>NCBR, Masaryk University, Brno, Czech Republic

<sup>4</sup>Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno, Czech Republic

<sup>5</sup>Bar-Ilan University, Faculty of Life Sciences, Ramat Gan, Israel

<sup>6</sup>Department of Biochemistry, Masaryk University, Brno, Czech Republic

**ABSTRACT** PA-IIL is a fucose-binding lectin from *Pseudomonas aeruginosa* that is closely related to the virulence factors of the bacterium. Previous structural studies have revealed a new carbohydrate-binding mode with direct involvement of two calcium ions (Mitchell E, Houles C, Sudakevitz D, Wimmerova M, Gautier C, Peréz S, Wu AM, Gilboa-Garber N, Imberty A. Structural basis for selective recognition of oligosaccharides from cystic fibrosis patients by the lectin PA-IIL of *Pseudomonas aeruginosa*. *Nat Struct Biol* 2002;9:918–921). A combination of thermodynamic, structural, and computational methods has been used to study the basis of the high affinity for the monosaccharide ligand. A titration microcalorimetry study indicated that the high affinity is enthalpy driven. The crystal structure of the tetrameric PA-IIL in complex with fucose and calcium was refined to 1.0 Å resolution and, in combination with modeling, allowed a proposal to be made for the hydrogen-bond network in the binding site. Calculations of partial charges using *ab initio* computational chemistry methods indicated that extensive delocalization of charges between the calcium ions, the side chains of the protein-binding site and the carbohydrate ligand is responsible for the high enthalpy of binding and therefore for the unusually high affinity observed for this unique mode of carbohydrate recognition. *Proteins* 2005;58:735–746. © 2004 Wiley-Liss, Inc.

**Key words:** *Pseudomonas aeruginosa*; lectin; fucose; crystallography; thermodynamics; quantum chemistry

## INTRODUCTION

*Pseudomonas aeruginosa* is a gram-negative bacterium that is found in various environments including soil, water, and vegetation. It is also an opportunistic pathogen, responsible for numerous nosocomial infections in immunocompromised patients. The bacteria colonize patients suffering from a number of chronic lung diseases, particularly

those under assisted ventilation. Their infections are frequently fatal for cystic fibrosis patients. The airway epithelia glycoconjugates and the abundant mucins of CF patients are known to be modified and serve as attachment points for *P. aeruginosa*.<sup>1–3</sup> For this reason, the different carbohydrate-binding proteins of these bacteria (pilin, flagellin, and non-pili lectins) are of high interest for their role in host recognition and specific adhesion. Under certain culture conditions, *P. aeruginosa* expresses two intracellular lectins, PA-IL and PA-IIL, specific for galactose and fucose, respectively.<sup>4</sup> The two lectins consist of four subunits of 12.7 kDa<sup>5</sup> and 11.7 kDa,<sup>6</sup> respectively. These lectins, which do not display sequence similarity, are mainly associated with the inner membrane of the bacteria,<sup>7</sup> but can also be exposed at the surface.<sup>8</sup>

The present study focuses on PA-IIL, the fucose-binding lectin that recently attracted much attention.<sup>9</sup> In hemagglutination inhibition tests, the strongest monosaccharide derivative inhibition was obtained with *p*-nitrophenyl- $\alpha$ -L-fucose followed by L-fucose > L-galactose > D-arabinose > D-fructose, and D-mannose while D-galactose was not active.<sup>10</sup> PA-IIL is characterized by high affinity for L-fucose<sup>10</sup> with an association constant ( $K_a$ ) of  $1.6 \times 10^6$  M<sup>-1</sup>. Such micromolar affinity range is very unusual among lectin-monosaccharide interactions where the millimolar range is more commonly observed. PA-IIL has been the subject of two crystallographic studies<sup>8,11</sup> that yielded

*Abbreviations:* CF, cystic fibrosis; PA-IL, *Pseudomonas aeruginosa* first (galactose-binding) lectin; PA-IIL, *Pseudomonas aeruginosa* second (fucose-binding) lectin

The Supplementary Materials referred to in this article can be found at <http://www.interscience.wiley.com/jpages/0887-3585/suppmat/index.html>. Included is a table of charges similar to Table V but for other calculation methods (HF and Mulliken).

\*Correspondence to: Anne Imberty, CERMAV-CNRS, 601 rue de la Chimie, BP53, 38042 Grenoble cedex 9, France. E-mail: imberty@cermav.cnrs.fr

Received 4 June 2004; Accepted 13 August 2004

Published online 30 November 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20330



several structures: in addition to the native lectin and the calcium-free one, complexes with three monosaccharides L-fucose, D-mannose, and D-fructopyranose have been described. In the PA-IIL/fucose complex,<sup>11</sup> the fucose residue locks onto a pair of calcium ions with three of its hydroxyl groups participating in the coordination of these cations. Such a protein/carbohydrate binding mode has never been observed before. The different orientation of mannose and fructose in complex with the lectin<sup>8</sup> yielded isosteric arrangements of three sugar hydroxyl groups in direct contact with the calcium ions. More recently, a sequence-related protein purified from the plant pathogenic bacterium *Ralstonia solanacearum*, RS-IIL, has been cocrystallized with mannose and the structure displays the same peculiar role for a pair of calcium ions in the binding site.<sup>12</sup>

The presence of the two calcium ions in the binding site appears to be the clue for the two particularities of PA-IIL: the weak specificity and the high affinity. The weak specificity has been fully characterized by crystallography and the requirement for the particular stereochemistry of two equatorial and one axial hydroxyl groups is now well established<sup>12</sup> and leads to a number of sugars being able to bind. The aim of the present work was to investigate the high affinity of PA-IIL for fucose. The thermodynamic basis of the affinity has been studied by analyzing enthalpy and entropy contributions while the atomic basis is characterized fully by the 1.0 Å resolution crystal structure of the complex. Further rationalization is brought by the study of charge distribution carried out with ab initio calculations. This combination of structural and thermodynamic approaches could serve for further development of anti-adhesion therapeutics against *P. aeruginosa* infections.

## MATERIALS AND METHODS

### Construction of Plasmid for Expression of PA-IIL

The oligonucleotides used as primers were as follows: 5'-GGA GAT ACC ATA TGG CAA CAC AAG GAG-3' (27-mer) and 5'-TTC CAA GCT TCT AGC CGA GCG G-3' (22-mer). The former is designed for the introduction of NdeI and the latter for HindIII restriction sites, respectively. The PCR was performed by using Taq polymerase (Qiagen, GmbH, Germany) and template DNA from *Pseudomonas aeruginosa* ATCC 33347. After digestion with NdeI and HindIII, the amplified fragment was introduced into multiple cloning site of pET-25(b+) vector (Novagen, Madison, WI) resulting in plasmid pET25pa2l.

### Expression and Purification of PA-IIL

*E. coli* BL21(DE3) cells containing the plasmid pET25pa2l were cultured in 1 L of Luria broth (LB) at 37°C. When the culture reached an optical density of 0.5–0.6 at 600 nm, iso-propyl-β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM. Cells were harvested after 3 h incubation at 30°C, washed and resuspended in 10 ml of the buffer (20 mM Tris/HCl, 100 mM NaCl, 100 μM CaCl<sub>2</sub>, pH 7.5). The cells were disrupted by sonication (Soniprep 150, Schoeller instruments, GB). After centrifugation at 10,000 g for 1 h, the

supernatant was further purified on D-Mannose Agarose (Sigma-Aldrich, USA). PA-IIL was allowed to bind to the immobilized mannose in the equilibrating buffer (20 mM Tris/HCl, 100 mM NaCl, 100 μM CaCl<sub>2</sub>, pH 7.5) and then was eluted by the elution buffer (20 mM Tris/HCl, 100 mM NaCl, 100 μM CaCl<sub>2</sub>, 0.1 M mannose, pH 7.5). The purified protein was intensively dialyzed against distilled water for 1 week (for the removal of D-mannose), concentrated by lyophilization and kept at –20°C.

### Hemagglutination Activities and Their Inhibition

Samples (0.05 ml) of the native and recombinant PA-IIL solutions, each at a concentration of 1 mg per ml, were serially twofold diluted with 0.05 ml saline. Then 0.05 ml of 5% suspension of papain-treated type O human erythrocytes<sup>4</sup> was added to each tube. Following a short shaking and 30 min at room temperature the tubes were centrifuged for 30 s (1000×) and the hemagglutination was examined. For the inhibition test L-fucose and D-mannose solutions at 0.15 M concentration were serially twofold diluted in 0.05 ml saline for each lectin preparation analysis. 0.05 ml of each lectin preparation at the dilution giving a titer of 1:32 was added to each tube in the respective series. After 30 min at room temperature 0.05 ml of the erythrocyte suspension was added to each tube and the hemagglutinating activity was examined as described above.

### Microcalorimetry Experiments

Titration calorimetry experiments were performed using two different systems, a Microcal VP-ITC microcalorimeter (Microcal, Northampton, MA) and CSC4200 (Calorimetry Sciences Corp.), respectively. All titrations were performed in 10 mM Tris.HCl buffer containing 0.1 M NaCl and 1–3 mM CaCl<sub>2</sub> at different pH as specified in Table II. Aliquots of 4 or 5 μl of each carbohydrate, dissolved in the same buffer, were added at 4 min intervals into the lectin solution present in the calorimeter cell. In the titrations, the protein concentration in the cell varied from 0.078 to 1 and 0.20 to 0.25 mM for Microcal VP-ITC (cell volume = 1411 μl) and CSC4200 (cell volume = 1300 μl), respectively. The sugar concentration in the 250 μl syringe was 10 to 20 times higher than the protein concentration used in the experiment (typically 1 and 4–5 mM for Microcal VP-ITC and CSC4200, respectively). The temperature of the cell was controlled to 25 ± 0.1°C. Data from a control experiment performed by identical monosaccharide injections into the cell containing only a buffer were subtracted prior to data analysis. Integrated heat effects were analyzed by nonlinear regression using a single site-binding model (Microcal Origin 5.0 and Bindworks 3.0). Fitted data yielded the binding constant ( $K_b$ ) and the enthalpy of binding ( $\Delta H_b$ ). Other thermodynamic parameters, i.e., changes in free energy,  $\Delta G_b$ , and entropy,  $\Delta S_b$ , were calculated from equation

$$\Delta G_b = \Delta H_b - T\Delta S_b = -RT \ln K_b$$

where T is the absolute temperature and R = 8.314 J.mol<sup>-1</sup>.K<sup>-1</sup>. At least two independent titrations were performed for each ligand tested.

TABLE I. Data Collection and Refinement Statistics

Data Collection and Phasing	
Wavelength (Å)	0.934
Resolution (Å)	1.0
Highest resolution shell (Å)	1.00 – 1.04
Completeness (%)	97.7 (94.3)
Space group	P2 <sub>1</sub>
a, b, c (Å)	52.735, 72.827, 54.747
β (°)	94.21
Measured reflections	706877
Unique reflections	216831
Average multiplicity	3.3 (2.9)
Nb of passes (low and high)	2
R <sub>Merge</sub> (%)	0.063 (0.228)
Average I/σ(I)	5.0 (2.8)
Wilson B-factor (Å <sup>2</sup> )	6.6
ACORN phasing statistics	
Correlation coefficient	0.38 (166355 medium E reflections)
R factor	0.44 (166355 medium E reflections)
Refinement	
Resolution range (Å)	30–1.0
Nb of reflections	212499
Protein atoms	3308
Residues modelled with multiple locations	27
Water molecules	682 (59 with occupancy 0.5)
Solvent ions	2 sulfate groups
Metal ions	8 Ca <sup>2+</sup>
Ligands	4 fucose residues
R <sub>crys</sub>	0.1146
R <sub>free</sub>	0.1351
Cruickshank's DPI based on maximum likelihood (Å)	0.023
Average B <sub>iso</sub> (Å)	
All atoms	13.1
Main chain	8.8
Side chain and water	16.7
RMS Deviation from ideality	
Bond (Å)	0.015
Angles (degrees)	1.89
Percent of refl. for R <sub>free</sub>	3% (4332 reflections)
Residues with alternative conformations	
Double	A31,A41,A46,A60,A62, A68,A77,B29,B46, B53,B68,B77, C13,C29, C59,C77,C82,D13, D41,D66,D68,D77.
Triple	A81,B81,C81,D16,D81.

### Crystallization and Data Collection

Lyophilized purified PA-IIL was dissolved in water (10 mg ml<sup>-1</sup>) in the presence of fucose (250 μg ml<sup>-1</sup>), and salts (ZnCl<sub>2</sub>, CaCl<sub>2</sub>, MgCl<sub>2</sub> at 2 mM). Crystallization trials were performed with Hampton crystallization screens I and II (Hampton research, Laguna Nigel, CA) using the hanging drop technique. After optimization of conditions, hanging drops were prepared using 2.5 μl of this preparation and 2.5 μl of the reservoir solution consisting of 1.75 M ammonium sulphate in 0.1 M Tris HCl (pH 8.5), and stored at 20°C. Crystals grew in a rectangular form to maximum dimensions of 0.3 × 0.3 × 0.5 mm<sup>3</sup> after one month. They belong to space group P2<sub>1</sub> with unit cell dimensions of a = 52.735 Å, b = 72.827 Å, c = 54.747 Å, and β = 94.21° at 100K. The asymmetric unit is a tetramer, corresponding to a V<sub>m</sub> 1.75 Å<sup>3</sup> Da<sup>-1</sup> and a solvent content of 29%.

A crystal was cryo-cooled at 100 K, after soaking it for as short a time as possible in glycerol 25% v/v in the precipi-

tant solution. All data images were recorded on an ADSC Q4R CCD detector (Quantum Corp.) on the fixed energy beamline ID14-2 at the ESRF (Grenoble, France). To avoid the loss of the strongest reflections owing to saturation, two sets of data, at low and high resolution, were collected in that order on the same crystal. For the high resolution data the detector was positioned as close as permitted on ID14-2. This did mean that the full resolution available from the crystal was not collected and resulted in a high resolution bin R<sub>merge</sub> of 22.8%. Measurements were made at a single X-ray wavelength of 0.934 Å. Data were processed with the HKL Suite<sup>13</sup> and scaled and converted to structure factors using the CCP4 program suite<sup>14</sup> (see Table I for data collection and diffraction data statistics).

### Structure Determination

Phasing was performed with ACORN,<sup>15</sup> using only the position of five of the calcium ions from the crystal

structure of PA-IIL/fucose complex previously solved at 1.3 Å resolution.<sup>11</sup> No reflections were omitted for the purposes of calculating the normalized structure factors for input to ACORN. Using the ACORN phases, an initial structure was built automatically using ARP/warp<sup>16</sup> to give 440 residues out of a total of 456 for the asymmetric unit cell. The structure was edited using the program O<sup>17</sup> to add missing residues, calcium ions, and fucose ligands that were clearly defined in the electron density maps. This model was then subjected to 20 cycles of refinement in REFMAC<sup>18</sup> yielding  $R_{\text{crys}}$  of 20.10% and  $R_{\text{free}}$  of 21.80%. After several cycles of REFMAC refinement with isotropic *B*-factors, including automatic water molecule placement using ARP/warp, manual rebuilding with O, addition of riding hydrogens in REFMAC, and construction of alternative conformations of side chains,  $R_{\text{crys}}$  and  $R_{\text{free}}$  values respectively decreased to 15.91% and 16.89%. In the next stage, refinement including anisotropic *B*-factors and riding hydrogens on all protein and fucose groups except hydroxyl moieties, was performed with SHELXL.<sup>19</sup> Additional water molecules were placed on the basis of Fo-Fc difference maps. Water molecules were added if the electron density was greater than three sigma and the hydrogen bonding network was sensible. In a number of cases the water molecules, though clearly present, were less than 2.0 Å apart and for these two water molecules, each with an occupancy of 0.5, were positioned in the model.

In order to study the protonated state of the aspartic and glutamic acids in the binding site, the last cycles of refinement were conducted without stereochemical restraints on the carbon-oxygen bond lengths and bond angles of the carboxylate groups. This resulted in a final model of 456 residues with four fucose molecules (one per monomer), eight calcium ions (two per monomer), 682 water molecules of which 59 have a half occupancy, two sulfate ions with  $R_{\text{crys}}$  of 11.46% and  $R_{\text{free}}$  of 13.51% at 1.0 Å resolution (Table I). Additionally, the last six cycles of SHELX refinement were performed in parallel without the hydrogen atoms modeled on the fucose ligands in order to create an omit map and give a gauge for visualizing the binding site hydroxyl hydrogens (which had never been modeled during the refinement). Coordinates for the 1.0 Å resolution X-ray structure of PA-IIL/fucose complex have been deposited in the Protein Data Bank<sup>20</sup> with the code 1UZV.

### Location of Hydroxyl Groups by Molecular Modeling

The hydrogen bond network in the binding site was built using the molecular graphics package Sybyl.<sup>21</sup> Hydrogen atoms on carbons were added geometrically in their riding position. Hydrogens of carbohydrate hydroxyl groups (O2, O3, O4 and O6) and protein (OG or Ser23) were oriented towards the closer acceptor carboxylate group, when present. Several orientations were tested for the water molecules in order to optimize the resulting hydrogen-bond network. Finally, partial charges were added according to the parameterization developed previously for protein-carbohydrate interactions<sup>22</sup> for the Tripos force-

field<sup>23</sup> and the positions of all hydrogen atoms of the binding site were optimized.

### Quantum Chemical Calculations

Quantum-chemical calculations were performed on a selected part of the 1.0 Å crystal structure of PA-IIL/fucose, with hydrogen atoms located as described above. The ligand-site amino-acid residues Asn21, Ser23, Glu95, Asp96, Asp99, Asp101, Asn103, Asp104 from the A-chain and Gly114 (from the B-chain) were included in all the calculations. Two calcium ions, one fucose residue and one water molecule that create a hydrogen-bond bridge between the fucose and the protein were also included.

As the first step, the geometry was optimized to ensure that the X-ray structure would be correctly reproduced by quantum chemical calculations. Because of the size of the system (138 heavy atoms), the initial geometry optimizations were performed by semi-empirical methods. All optimizations were carried out with the program MOPAC2002<sup>24</sup> using eight different combinations of methods and techniques: two semi-empirical methods, AM1<sup>25</sup> and PM3<sup>26,27</sup> were combined with two SCF procedures, conventional or MOZYME, and two optimizers, BFGS or EF. The following optional keywords were used: GEO-OK, MMOK, NOXYZ, and NODIIS. The optimizations were performed with both 1.3 Å and 1.0 Å resolution structures. The data analysis was performed using the program TRITON.<sup>28</sup> The best results, with the lowest RMSD difference between the calculated and X-ray data, were obtained by the combination of PM3/EF with conventional SCF procedure. This procedure was therefore used for optimizing the binding site geometry. In order to render the computation less demanding, all backbone atoms were removed, except for Asn21, replacing the C $\alpha$  atoms by methyl groups. The population analysis on the ab initio level was performed on the binding site with no hydrogen atoms added on the aspartate and glutamate groups, leading to a net charge of -2 for the system. Identical calculations were repeated for different protonation states, adding one or two hydrogen atoms on selected acetate groups. Mulliken population analysis<sup>29</sup> and the population analysis based on the electrostatic potential (ESP) with the subsequent RESP procedure<sup>30,31</sup> were performed on this reduced binding site model. The population analyses were based on two different single-point ab initio calculations, performed at the HF/6-31G\*\* level, and the DFT B3LYP/6-31G\*\* level. All the ab initio calculations were performed using the program Gaussian98.<sup>32</sup>

### Data Mining of Similar Sequences

Blast searches<sup>33</sup> were performed on general databases (www.infobiogen.fr) as well as on microbial databases (http://www.ncbi.nlm.nih.gov/sutils) both at the peptide and DNA sequence levels. When only DNA sequences were available, they were translated starting from the first ATG motif upstream. Peptide sequence alignments were performed using Clustalx<sup>34</sup> and displayed with Genedoc.<sup>35</sup>

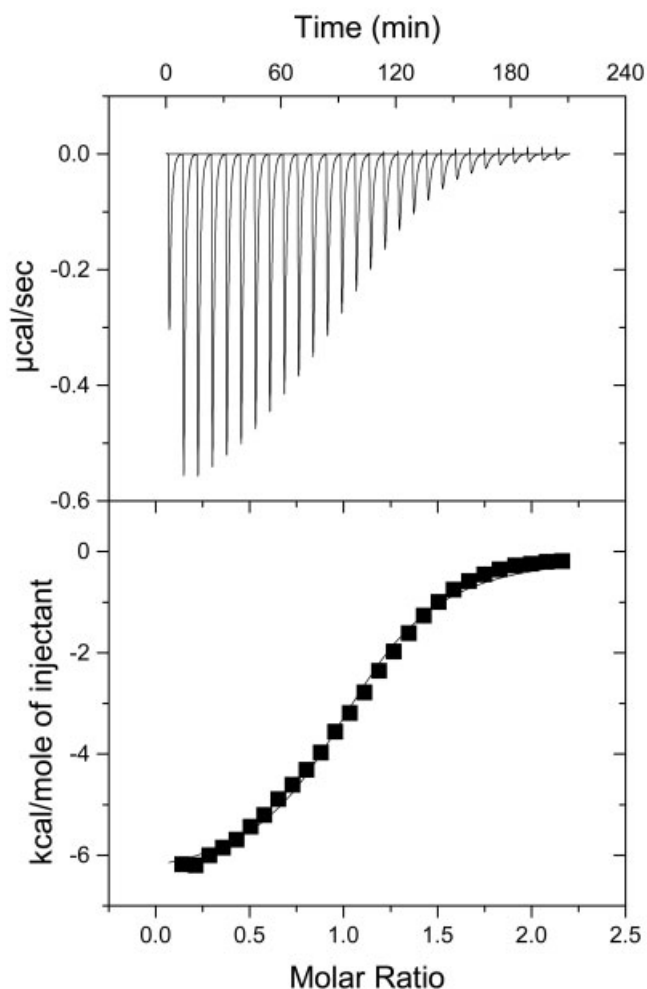


Fig. 1. Titration calorimetry results of L-Fuc (0.78 mM) binding to PA-IIL (0.08 mM) in 10 mM Tris buffer containing 100 mM NaCl and 3 mM  $\text{CaCl}_2$  (pH 7.5), at 25°C. **Top:** data obtained from 28 automatic injections, 10  $\mu\text{l}$  each, of L-Fuc, into PA-IIL in the cell. **Bottom:** plot of the total heat released as a function of total ligand concentration for the titration shown above (squares). The solid line represents the best least-square fit for the obtained data.

## RESULTS

### Comparison of Hemagglutination Activity of Recombinant and Native PA-IIL

The hemagglutinating titers of both the native and recombinant lectin preparations were almost the same, 1:256–1:512. The hemagglutination appeared more rapidly with the recombinant lectin. Both lectin preparations also showed similar carbohydrate affinities: being inhibited by fucose dilution up to 0.147 mM (11 tubes) and by mannose dilution up to 4.7 mM (5–6 tubes).

### Thermodynamic Data

The results of a typical microcalorimetry experiment are displayed in Figure 1, showing a steep decrease in the exothermic heat of binding while saturation is achieved. The values of binding affinity  $K_b$ , binding enthalpy  $\Delta H_b$ , and stoichiometry per monomer  $n$  have been obtained by

fitting a classical equation for single-site binding.<sup>36</sup> The thermodynamic parameters, free energy, and entropy of binding,  $\Delta G_b$ , and  $\Delta S_b$ , were calculated as indicated in the experimental section. The data obtained under different experimental conditions are listed in Table II. Depending on the pH and conditions of measurement, the binding affinity varied between 120 to 180  $10^3 \text{ M}^{-1}$ , corresponding to dissociation constants ranging between 5.6 to 8.3  $10^{-6} \text{ M}$ .

The experiments performed on the CSC and MicroCal calorimeters gave slightly different values for the stoichiometry of the interaction. While an occupancy of almost one fucose per monomer was measured on the Microcal, for the other machine only 75% saturation was obtained. This could be ascribed either to some defects in the protein preparation, with one monomer inactivated, or to problems due to the need of higher protein concentration and/or due to the lower sensitivity of the CSC microcalorimeter. Nevertheless, this does not affect the evaluation of the thermodynamic parameters. In all measurements, a high enthalpy of binding ( $\Delta H_b \approx -27.5 \text{ kJ/mol}$ ) was observed. The entropy contribution is weak but slightly favorable ( $T\Delta S_b \approx 2 \text{ kJ/mol}$ ), contributing to less than 7% of the free energy of binding ( $\Delta G_b \approx -29.5 \text{ kJ/mol}$ ).

### Analysis of the Binding Site in the Very High Resolution Crystal Structure

The crystal structure of the PA-IIL/fucose complex refined at 1.0 Å resolution displayed the same overall features [Fig. 2(A)] as those previously refined at 1.3 Å resolution in the same  $P2_1$  space group<sup>11</sup> and refined at 1.2 Å resolution in  $P2_12_12_1$  space group.<sup>8</sup> In addition to the eight calcium ions, and four fucose ligands, 682 water molecules and two sulfate ions have been refined; 27 amino acids were modeled with two or three conformations.

Each monomer is characterized by a calcium and fucose binding site made up primarily by one long beta-hairpin loop with the participation of a shorter loop and of the C-terminal acidic group at Gly114 of the neighboring monomer. The high resolution allows a detailed analysis of the site. Distances of interest for both the coordination of the two calcium ions, both hepta-coordinated, and the binding of fucose are given in Table III. The two calcium ions are very close to each other, with distances varying from 2.73 to 2.75 Å in the four monomers.

Figure 2(B) illustrates the quality of the electron density map at the ligand site. The omit Fo–Fc difference map allows all of the aliphatic hydrogen atoms of fucose to be located. The difference map contour around the hydrogen at C-1 also indicates that a minority population of fucose adopts the  $\beta$ -configuration, though this was not modeled. The hydroxyl hydrogen atoms are of the greatest interest since they are involved in the hydrogen-bond network between the fucose residue and the protein side chain. Unfortunately only one of three hydroxyl groups can be oriented from the density map. As seen in Figure 2(B), the hydrogen atom of the O-2 hydroxyl group points towards oxygen OD1 of the Asp96 side chain.

Since not all hydrogen atoms can be identified clearly, it is not possible to conclude from the X-ray electron density maps alone the protonation state of the acidic groups

**TABLE II. Thermodynamics—ITC for PA-III with Fucose<sup>†</sup>**

	pH	Calcium (mM)	Kb 10 <sup>5</sup> M <sup>-1</sup>	n	−ΔG (kJ/mol)	−ΔH (kJ/mol)	−ΔS (J/mol. K)	−TΔS (kJ/mol)
Exp 1—fucose <sup>a</sup>	7.4	4	1.2 (± 0.1)	0.78 (± 0.02)	29.0 (± 0.1)	27.7 (± 0.5)	−4.1 (± 1.8)	−1.3 (± 0.5)
Exp 2—fucose <sup>a</sup>	8.5	1–3	1.6 (± 0.2)	0.79 ± (0.06)	29.6 (± 0.3)	27.5 (± 1.4)	−7.1 (± 3.7)	−2.1 (± 0.1)
Exp 3—fucose <sup>b</sup>	7.5	3	1.8 (± 0.2)	1.05 ± (0.02)	29.9 (± 0.2)	27.3 (± 0.3)	−5.9 (± 1.7)	−2.6 (± 0.4)

<sup>†</sup>Experiments 1 to 3 have been conducted using different batches of recombinant protein. Values in parentheses are standard deviations from at least two separate titrations.

<sup>a</sup>Data from CSC ITC microcalorimeter.

<sup>b</sup>Data from a Microcal VP-ITC microcalorimeter.

**TABLE III. Geometrical Details of the Four Binding Sites (Distances in Å) in the 1.0-Å Resolution Crystal Structure of PA-III/Fucose Complex<sup>†</sup>**

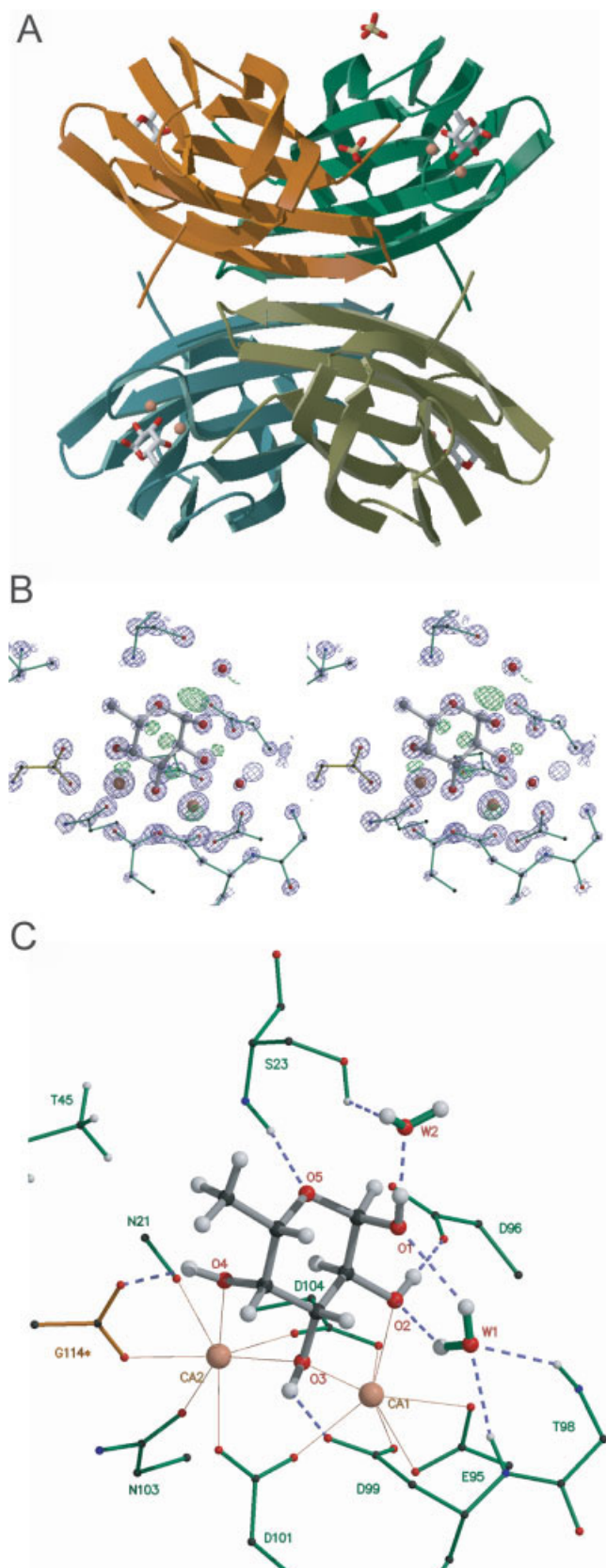
		Chain A	Chain B	Chain C	Chain D
Interactions between the calcium ions and the protein.					
Ca-1	Asp101.OD1	2.37	2.37	2.37	2.36
Ca-1	Glu95.E1	2.46	2.46	2.48	2.45
Ca-1	Glu95.E2	2.41	2.42	2.42	2.41
Ca-1	Asp99.OD1	2.36	2.36	2.37	2.38
Ca-1	Asp104.OD2	2.42	2.41	2.41	2.41
Ca-2	Asn21.O	2.38	2.39	2.37	2.39
Ca-2	Asp101.OD2	2.41	2.41	2.42	2.43
Ca-2	Asn103.OD1	2.34	2.34	2.33	2.33
Ca-2	Asp104.OD1	2.35	2.35	2.36	2.34
Ca-2	Gly114.O <sup>a</sup>	2.42	2.44	2.43	2.41
Interactions between the calcium ions and the fucose residue.					
Ca-1	O2	2.51	2.49	2.51	2.48
Ca-1	O3	2.47	2.48	2.48	2.48
Ca-2	O3	2.48	2.49	2.50	2.46
Ca-2	O4	2.49	2.51	2.48	2.50
Direct interactions between the fucose residue and the protein.					
O2	Asp96.OD1 ***	2.67	2.66	2.62	2.66
O3	Asp99.OD1	2.98	3.00	2.94	3.00
O3	Asp99.OD2 ***	2.55	2.57	2.55	2.58
O3	Asp101.OD1	2.95	2.94	2.96	2.96
O3	Asp101.OD2	2.94	2.94	2.94	2.92
O3	Asp104.OD2	2.99	3.01	3.02	2.99
O4	Asn21.O	3.01	3.11	3.07	3.03
O4	Gly114.OXT <sup>a***</sup>	2.53	2.55	2.58	2.58
O5	Ser23.N ***	2.92	2.96	2.95	2.98
C-Me	Thr45.CG2	4.06	4.27	4.09	4.07
Interactions between the fucose residue and the protein via water molecules.					
O1	Wat_1 ***	3.16	3.27	3.13	3.27
O2	Wat_1 ***	3.00	3.04	3.13	3.04
O1	Wat_2 ***	2.76	2.70	2.75	2.76
Wat_1	Thr98.N ***	2.90	2.97	2.90	2.93
Wat_1	Asp99.N ***	3.32	3.18	3.28	3.16
Wat_2	Ser23.OG ***	2.68	2.71	3.32	2.62

<sup>†</sup>The three stars indicate distances with geometrical feature in agreement with existence of a hydrogen bond.

<sup>a</sup>C-terminal residue from the other monomer.

involved in ligand binding. Nevertheless, due to the high resolution of the data (around 5.6 observation per parameter during anisotropic *B*-factor refinement), it was possible to perform 20 refinement cycles without stereochemical restraints on the bond lengths of the COO groups in the ligand site and to analyze the differences, if any. There are few differences between the four monomers of the asymmetric unit, and therefore only the average bond lengths are listed in Table IV. Larger average bond length differences are observed for Asp99 (1.22 Å for CD-OD1 vs. 1.27 Å for

CD-OD2). This can be compared to values of 1.24 and 1.32 Å observed for COOH observed in the atomic resolution endocellulase crystal structure.<sup>37</sup> The C-O bonds of the fucose site carboxylates, bar Asp99, refine to lengths between 1.25 and 1.27 Å which are compatible with no protonation of those groups (the estimated average coordinate error being 0.023 Å). Further comparison of the molecular model with the protonated structure of the binding site, as treated by computational chemistry approaches are discussed below.



### Modeling of Hydrogen-Bond Network

Where it is assumed, accordant with the X-ray data, that none of the aspartate and glutamate side chains of the binding site are protonated, the modeling study arrives at only one possible network of hydrogen bonds between the sugar and the protein. The most buried hydroxyl groups of fucose, i.e., O3 and O4, donate a hydrogen bond to Asp99.OD2 and Gly114\*.OXT, respectively. The third hydroxyl group, O2, also donates a hydrogen bond to an acidic group, i.e., to OD1 of Asp96, this hydrogen bond being the only one that could be identified from the X-ray data. With the participation of the backbone NH of Ser23, that gives a hydrogen to the fucose ring oxygen O5, this gives a total of four direct hydrogen bonds that can be modeled unambiguously between the protein and the sugar [Fig. 2(C)]. The hydrogen atoms of the conserved water molecule Wat1 could also be modeled: the orientation of the water molecule could be deduced from the facts that it receives two hydrogens from NH of Thr98 and Asp99 and that it donates one hydrogen to the O2 hydroxyl of fucose. The second hydrogen of the same water molecule can then point to O1 of fucose. The special location of this water molecule seems to allow the water to stabilize the tip of the calcium-carbohydrate binding loop made up of 14 amino acids in a beta-hairpin motif. This particular water is conserved in all PA-IIL/fucose complexes (all four monomers, 1.3 and 1.0 Å resolution). Its special role in maintaining the conformation of the loop is illustrated by the fact that it is also conserved in the native PA-IIL structure, but not in the demetallized one, in which the calcium-binding loop is partially disordered.<sup>8</sup>

The second water molecule, Wat2, that stabilizes the alpha configuration of fucose by bridging the anomeric oxygen atom to the side chain of Ser23 is not so well conserved in all monomers and splits in two close sites (modeled at 0.5 occupancy each) in one of them. As a consequence, there are several possibilities for orienting the protons between the Ser23.OG, Wat2 and O1 of fucose and the network shown in Figure 2(C) represents only one possibility.

The proposed network displayed in Figure 2(C) results in the formation of seven direct or indirect hydrogen bonds between one monosaccharide and the protein.

Fig. 2. Structure of the 1.0 Å resolution structure of the PA-IIL/fucose complex. **A:** Ribbon representation of the tetramer consisting of the asymmetric unit with stick representation of sugar and sulfate and cpk representation of the calcium ions. **B:** Final sigma weighted  $2F_o - F_c$  electron density map (blue color contoured at  $4\sigma$ ) around the monosaccharide residue, water molecules and calcium ions and sigma weighted  $F_o - F_c$  omit map (green color contoured at  $2.5\sigma$ ) after six cycles of refinement omitting the hydrogen atoms riding the fucose carbons. **C:** Interactions of PA-IIL with calcium ions and fucose with modeled hydrogen atoms. Coordination contacts are indicated by orange solid lines and hydrogen bonds by green dashed lines. The \* symbol indicates the terminal glycine from the other monomer. All molecular drawings were prepared with MOLSCRIPT<sup>53</sup> and RASTER-3D.<sup>54</sup>

TABLE IV. Carbon–Oxygen Bond Lengths in PA-IIL Binding Site<sup>†</sup>

Residue	C–O bond	Length from X-ray data Å <sup>a</sup>	Length from ab initio calculations, Å			
			Non-protonated	Gly114*-protonated	Asp96-protonated	Both protonated
N21	O	1.25	1.23	1.23	1.23	1.23
N103	OD1	1.23	1.24	1.24	1.24	1.25
G114	O	1.27	1.25	1.25	1.25	1.25
	OXT	1.27	1.26	<b>1.33<sup>b</sup></b>	1.26	<b>1.33<sup>b</sup></b>
E95	OE1	1.26	1.27	1.27	1.27	1.27
	OE2	1.27	1.27	1.27	1.28	1.28
D96	OD1	1.25	1.27	1.26	<b>1.34<sup>b</sup></b>	<b>1.34<sup>b</sup></b>
	OD2	1.26	1.25	1.26	1.22	1.22
D99	OD1	1.22	1.27	1.27	1.28	1.27
	OD2	1.27	1.26	1.26	1.25	1.26
D101	OD1	1.26	1.27	1.27	1.27	1.28
	OD2	1.25	1.27	1.28	1.27	1.28
D104	OD1	1.25	1.28	1.29	1.27	1.28
	OD2	1.27	1.27	1.26	1.27	1.26

<sup>†</sup>Comparison between crystallographic data and models calculated using semi-empirical methods for different protonation states.

<sup>a</sup>Average over the four monomers. Standard deviations less than 0.02 Å.

<sup>b</sup>Protonated oxygen atom.

### Prediction of Protonation State by Ab Initio Calculations

As described above, the PA-IIL binding site is characterized by the presence of six carboxylate groups that are involved in the binding of two calcium ions. Since there are no amino acids that are likely to be positively charged in the close vicinity of the binding site and assuming zero protonation of the carboxyl groups, the overall charge of the site is expected to be  $-2$ . In order to confirm the crystallographic data indicating the absence of protonation, semi-empirical calculations were performed on the binding site in different ionization states. The following residues could potentially be protonated: Glu95, Asp96, Asp99, Asp101, Asp104, and Gly114B. Since both oxygens of Glu95, Asp101, and Asp104 are coordinated by calcium ions, they were excluded. Also Asp99 is unlikely to be protonated because the only position of a hydrogen would be in steric hindrance with O3 hydroxyl group of fucose. Therefore, the most probable candidates for protonation are the C-terminus carboxylate of Gly114B and the Asp96 residue. Four different systems were constructed and optimized: the fully deprotonated structure (i.e., none of the carboxyl groups of the binding site was protonated) with a net charge of  $-2$ , two systems with net charges of  $-1$  (one with Gly114B protonated, the other with Asp96 protonated), and one system with a zero net charge (with both Gly231 and Asp96 protonated). In all the above calculations, the original X-ray structure was correctly reproduced with low RMSD (see Table IV) except for the structural details described below. Analysis of the resulting geometries (Table IV) indicates that the computed C–O bond lengths are indeed dependent upon the protonation state with bond length varying from 1.27 ( $\pm 0.03$ ) Å for a delocalized carboxylate C–O bond to 1.33 ( $\pm 0.01$ ) Å for a carboxylic C–OH bond. These lengths can be compared readily to the experimentally measured values in very high resolution crystal structure of endocellulase that

contains one protonated glutamate with C–O and C–OH distances of 1.24 and 1.32 Å, respectively.<sup>37</sup> Comparison of the X-ray derived distances with the calculated ones indicates that only the calculation performed without any protonation is in agreement with experimental values. The binding site has been therefore considered as nonprotonated for the rest of the study.

### Calculation of Charge Distribution by Ab Initio Calculations

Charge distribution has been calculated on a mini-model of the site as described in the experimental section. Mulliken and RESP charges were calculated using various methods and levels of theory and on different systems: isolated fucose, the binding site of PA-IIL without any ligand, the binding site of PA-IIL with two Ca<sup>2+</sup> ions and, finally, the binding site of PA-IIL with two Ca<sup>2+</sup> ions and either a fucose residue or three water molecules (the water molecules replacing the O2, O3, and O4 atoms of fucose) (Table V). The Mulliken scheme does not separate greatly the above different scenarios of the binding site (data in Supplementary Material). However, when the RESP procedure is applied, the charge distribution is very dependent on the composition of the complex. Results obtained by taking RESP charges from B3LYP/6-31G\*\* calculations and from HF/6-31G\*\* calculations are listed in Table V. The complete results, including all data, are provided in supplementary material.

The atomic charges calculated with B3LYP or with HF/6-31G\*\* calculations are in good agreement. When there is no ligand present in the binding site, the global charge of  $-6$  is distributed mainly over the carboxyl groups. When including two calcium ions in the system, highly negative charges are induced in their neighborhood (i.e., on oxygen atoms of carbonyl and carboxyl groups). The calcium ions themselves keep relatively high positive charges of about  $+1.5$ . On the other hand, isolated fucose

**TABLE V. Charge Distribution in the Binding Site Calculated on Different Systems, Using B3LYP/6-31G\*\* (HF/6-31G\*\* in Italics) Ab Initio Calculations with RESP Charges**

	PA-IIL/Ca2+ Fucose	PA-IIL/Ca2+ Water	PA-IIL/Ca2+	Isolated molecule
Protein				
Asn21.O	-0.36/-0.44	-0.42/-0.50	-0.45/-0.56	-0.32 <sup>a</sup>
Gly114.O	-0.46/-0.56	-0.56/-0.65	-0.67/-0.78	-0.54
Asn103.OD1	-0.52/-0.60	-0.51/-0.63	-0.61/-0.72	-0.56
Asp99.OD1	-0.67/-0.78	-0.74/-0.87	-0.82/-0.93	-0.64
Asp104.OD1	-0.42/-0.52	-0.60/-0.72	-0.77/-0.90	-0.65
Asp101.OD1	-0.50/-0.61	-0.58/-0.70	-0.65/-0.77	-0.51
Asp101.OD2	-0.65/-0.76	-0.69/-0.81	-0.76/-0.87	-0.81
Ca <sup>2+</sup> ions				
Ca1	1.08/1.19	1.33/1.49	1.48/1.62	
Ca2	0.61/0.71	0.99/1.18	1.48/1.64	
Fucose				
Fuc.O2	-0.18/-0.24			-0.52/-0.60
Fuc.O3	0.02/0.01			-0.49/-0.57
Fuc.O4	0.00/-0.03			-0.52/-0.60
Water				
Wat2		-0.73/-0.84		
Wat3		-0.35/-0.46		
Wat4		-0.54/-0.67		

<sup>a</sup>B3LYP/6-31G\*\* calculations on isolated protein binding site did not converge.

is a neutral molecule but displays rather negatively charged hydroxy-oxygens with partial charges of about  $-0.5$ . When two calcium ions and fucose are present in the binding site, the charges on both calcium ions decreases significantly (by about 0.6 to 0.9). A strong charge transfer from three oxygen atoms of fucose and also from the neighboring oxygen atoms of amino acids in the binding site to calcium ions occurs. The situation is somewhere “in between” when there are three water molecules coordinated on calcium ions instead of three oxygens of fucose (see Table V).

In summary, when fucose is not present in the active site, the calcium ions bind by ionic bonds. In this case, the solvent molecules should be very competitive since the free energy of Ca ion solvation is  $-381$  kcal/mol.<sup>38</sup> This would suggest that there is no strong structural role of Ca ions in the unliganded PA-IIL. The fucose binding can be considered as an aqua-substitution reaction for the calcium ion: the fucose hydroxyl groups have a larger mean polarizability than water molecules, resulting in a larger charge redistribution upon binding. In this case, the binding is very strong and this suggests that the Ca ions have a significant structural role in the complex as the charge delocalization is crucial for its stability. Also, the binding site of the protein is stabilized by charge delocalization. It is interesting that when fucose is present, the charge transfer is not as significant on calcium 1 (charge 1.08) as it is on calcium 2 (charge 0.61). This may be caused by the difference in the functional groups that surround the ions and contribute differently to charge delocalization. While calcium 2 is coordinated by oxygens of three carboxyl groups, two carbonyl groups and two oxygens from fucose, calcium 1 is coordinated by oxygens of five carboxyl groups and two oxygens of fucose. So the real difference is the presence of two carbonyl groups which replace two carboxyl groups.

### Search for Similar Sequences in Databases

With the rapid progress of bacterial genomics, genes coding for PA-IIL-like proteins were identified in the newly sequenced genome of other bacteria: *Ralstonia solanacearum*,<sup>39</sup> *Chromobacterium violaceum*<sup>40</sup> and *Burkholderia cenocepacia* (The Sanger Institute, unfinished sequence data were obtained from [http://www.sanger.ac.uk/Projects/B\\_cenocepacia/](http://www.sanger.ac.uk/Projects/B_cenocepacia/)). A similar gene is also present in the newly sequenced genome of another strain of *Pseudomonas aeruginosa*, UCBPP-PA14 (Harvard Partners Center for Genetics and Genomics, <http://www.hpcgg.org:8080/hpcgg/jsp/hpcgg/Sequence/mouse/pseudomonas.jsp>). The gene product, protein accession number ZP\_00136733.1, is apparently shorter at the N-terminus than the protein from *P. aeruginosa* PA01. Detailed analysis of nucleotide sequence prior to the gene sequence revealed that an alternative initiation GTG codon for the automatic gene annotation was taken into account, not considering a presence of the classical initiation ATG codon 15 bp upstream. Moreover, 11 bp further upstream of this new ATG translational start site, a putative Shine-Delgarno sequence can be found, confirming that five amino acids in the annotated gene ZP\_00136733.1 are missing. The corrected N-terminus sequence results in complete homology of both *P. aeruginosa* lectins. The alignment of the PA-IIL like region of all the identified and putative lectins is given in Figure 3. In all cases, the PA-IIL-like domain corresponds to the C-terminus, finishing with a fully conserved WPLG sequence. The calcium and sugar binding loop (Asp96 to Asp104 in PA-IIL) is also highly conserved. Four of the sequences, i.e., the ones from both strains of *P. aeruginosa*, from *R. solanacearum*, and from *C. violaceum*, have very similar lengths and almost similar N-term sequences, while the three *B. cenocepacia* sequences have



			* !!!		
P.a_PA01	:	ATQGVFTLPANTRFGVTAFAANSSGTQTIVNVLVNNETA--	ATFSGQSTNNAVIGTQVLSNSGSS	:	60
P.a_UCBPP	:	ATQGVFTLPANTQFGVTAFAANSAGTQTIVNVQVNNETV--	ATFTGQSTNNAIIGSKVLNSGGG	:	60
R.s.	:	AQQGVFTLPANTSFGVTAFAANAANTQTITQVLVDNVVK--	ATFTGSGTSDKLLGSQLNSGSS-	:	59
Ch.v.	:	AQQGVFTLPARINFGVTVLVNSAATQHVETFEVDNEPR--	AAFGVGTGDNNGIKVINSGS-	:	59
B.c.1	:	ERDCITFNLPPIKFGVTAALTHAANDQITDIYIDDDPKPAATFKGAGAQQNLGIRKVLDSGN-		:	217
B.c.2	:	ERDCITFNLPPIKFGVTAALVNSSAPQITFEVVDNPKPAATFQCAGTQDANLNITQVLSNCK-		:	201
B.c.3	:	NRAGEFSIPNTDFRAIFFANAEEQIHKLFIGDSQEP-AAVHKLITRDGP-REATLNSGN-		:	66
			* * * * *		
P.a_PA01	:	GKVVQVQVSVNGRPSDLVSAQVILITNE-----	LNFALVGSSEDGTDNDYNDAAVVVFNWPLG	:	114
P.a_UCBPP	:	GKVVQLVSVNGRSSDLVSAQVILANE-----	LNFALVGSSEDSTNDYNDAAVVVFNWPLG	:	114
R.s.	:	GATKIQVSVNCKPSDLVSNQITLANK-----	LNFAMVGSSEDGTDNDYNDGIAVFNWPLG	:	113
Ch.v.	:	GNVRVQIILANGRQSDLVSSQLVLANK-----	LNLAVVGSSEDGTDMDYNDSTIVFNWPLG	:	113
B.c.1	:	GRVRVIVMANGRPSRLGSRQVDIFKK-----	SYFGIIGSEDGADDDYNDGIVFNWPLG	:	271
B.c.2	:	GKVRVIVTANGKPSKIGSRQVDIFKK-----	TYFGLVGSSEDGCGDYNDGIAVFNWPLG	:	255
B.c.3	:	GKIRFEVSVNCKPSATDARLAPINGKKS DSGSPFTVNFGLVVSSEDGHDSYNDGIVVFNWPLG		:	128

Fig. 3. Alignment of PA-IIL-like sequences from different bacteria. The sequences are from the following genes, with their NCBI/EMBL deposition number between [brackets]. **P.a\_PA01**, fucose-binding lectin from *Pseudomonas aeruginosa* PA01 [TrEMBL: Q9HYN5]; **P.a\_UCBPP**, ORF Paer104401 from *Pseudomonas aeruginosa* UCBPP-PA14 [NCBI: ZP\_00136733.1GI 32038461]; **R.s.**, mannose-binding lectin from *Ralstonia solanacearum* GM11000 [TrEMBL: Q8XUA5]; **Ch.v.**, ORF CV1744 from *Chromobacterium violaceum* ATCC12472 [NCBI: NP\_901414.1 GI:34495455]; **B.c.1**, **B.c.2** and **B.c.3**, ORFs predicted from unfinished sequence data of *Burkholderia cenocepacia* J2315 [chromosome 2, complement: 215683-216501, 214786-215556 and 216624-217013]; Numberings of amino acids is putative for P.a\_UCBPP, Ch.v. and the three B.c. and correspond to the most probable limits of the ORF. PA-IIL residues involved in calcium binding are marked by an asterisk and the ones responsible for the fine monosaccharide preference are indicated by an exclamation mark.

longer and more variable N-term regions with no similarity with known proteins (data not shown here).

## DISCUSSION

### Thermodynamic Basis of PA-IIL Affinity to Fucose

The present microcalorimetry study confirms the very high affinity of the PA-IIL for fucose, albeit with values one order of magnitude lower than that published previously from equilibrium dialysis study ( $K_a : 1.5 \cdot 10^6 \text{ M}^{-1}$ ).<sup>10</sup> The discrepancies between these values could be due to differences between methodological approaches or between the protein samples. Nevertheless, a review of lectin-protein interactions indicates that the affinity measured here is much higher than the millimolar range usually observed for binding unsubstituted monosaccharides,<sup>41</sup> the highest reported value being a  $K_a$  of  $9 \cdot 10^3 \text{ M}^{-1}$  for the binding to GalNAc of soybean agglutinin.<sup>42</sup>

Analysis of the different thermodynamic contributions demonstrates that the above interaction is characterized by a strong enthalpy of binding, accompanied by a weakly favorable entropy term (Table II). When compared to classical millimolar range lectin-monosaccharide interactions, the unusual affinity is explained not only by the relatively high enthalpy of binding, but also by the absence of an entropy barrier.

The favorable entropy of binding is very unusual in protein-carbohydrate complexes where an entropy barrier is generally observed.<sup>41</sup> The barrier is higher for the binding of flexible oligosaccharides, but unfavorable negative entropy contributions are also observed for monosaccharide binding and are generally attributed to the displacement of water molecules from disordered locations in amphiphilic regions of the carbohydrate binding sites to more ordered networks in bulk water solution.<sup>43</sup> Nevertheless, favorable entropy has already been observed for the

binding of glucose by *Lens culinaris* lectin<sup>44</sup> and the binding of galactose by *Erythrina corallodendron* lectin.<sup>45</sup> In the latter case, structural studies indicated that the favorable entropy may be due to the release of two tightly bound water molecules upon galactose binding.<sup>46</sup> Similar explanation could apply to the favorable entropy observed in PA-IIL/fucose interaction since three water molecules are tightly bound to calcium in the crystal structure of the native lectin<sup>8</sup> and are replaced by fucose hydroxyl groups in the complex.

### Atomic Basis of the PA-IIL Affinity for Fucose

A direct interaction between one calcium atom and a sugar ligand has previously been observed in the carbohydrate recognition domain (CRD) of C-type animal lectins that binds neutral monosaccharides with two equatorial hydroxyl groups such as D-mannose, L-fucose or L-galactose.<sup>47</sup> However, in the C-lectin family, the dissociation constant for monosaccharides is in the millimolar range. Another family of lectins that includes serum amyloid P component (SAP) and C-reactive protein (CRP), contains two close calcium ions involved in binding diverse ligands but it differs from that of PA-IIL since it is only directed to negatively charged carbohydrates.<sup>48</sup>

In the present work, the very particular environment of the two close calcium ions was analyzed. Three of the fucose hydroxyl groups are directly involved in the coordination sphere of the calcium ion. Two of these hydroxyl groups create very strong hydrogen bonds with acetate from protein side chains with an O—O distance of 2.55 (+/−0.02) Å, a value significantly shorter than that currently observed in protein carbohydrate interactions. Furthermore, the acetates that are involved in the hydrogen bonds also coordinate the calcium ions. This yields two peculiar six-membered ring arrangements (—O3—O3H—

D99.OD2—D99.CD—D99.OD1—Ca1— and a similar ring involving O2, Gly114 and Ca2) that could be responsible for the strong charge delocalization shown by the ab initio calculations and could therefore play a major role in the unusually high affinity observed between PA-IIL and the monosaccharide.

### Biological Occurrence of this Sugar Binding Mode

The PA-IIL sequence, first identified in 2000, has been reported recently in other phylogenetically related bacteria, the plant pathogen *R. solanacearum* and *C. violaceum*.<sup>12</sup> This latter gram-negative saprophytic bacterium from soil and water is of high interest for biotechnology since it produces diverse antibiotics. It is normally considered as a nonpathogen, although it is occasionally involved in infections and may cause fatal septicemias in human and animals.<sup>49</sup> Using the homology searches described above, three ORFs are described for the first time which code for three proteins containing C-terminal sequences highly similar to PA-IIL in another emerging pathogen, *Burkholderia cenocepacia* (previously *B. cepacia*). In the 1980s, it was found to cause life threatening pulmonary infections in cystic fibrosis patients with complications caused by *B. cepacia* having a mortality rate of 80%.<sup>50</sup> It has to be noted that all amino acids involved in calcium binding are fully conserved in these different proteins, including the C-terminal glycine. All these proteins can therefore be postulated to bind two calcium ions and also to display the same dimeric structure as PA-IIL. On the other hand, some variations are observed in the three-amino-acid motif that is responsible for sugar specificity<sup>12</sup> and it is difficult to predict what will be the highest affinity monosaccharide ligand for each sequence. The occurrence of the PA-IIL sequence in four different bacteria may give some clues as to the function of this lectin. These bacteria share some particularities: they are soil inhabitants, opportunistic pathogens, and highly dangerous once an infection is established. It has been observed that the ORFs of *C. violaceum* that present a close similarity to ORFs of both *R. solanacearum* (17.4%) and *P. aeruginosa* (9.6%) are mostly coding for proteins involved in the interactions of these bacteria with their environment.<sup>40</sup>

### CONCLUSION

Two paradigms dictate protein–carbohydrate interactions. The first one is that the efficacy of these interactions in biological processes is due to their multivalency that creates an “avidity” or “Velcro” effect that would compensate the low affinity (i.e., millimolar range) generally observed for each binding site.<sup>51</sup> The second paradigm is that protein/sugar binding is mainly due to two types of interaction: the numerous hydrogen bonds, owing to sugar hydroxyl groups, contribute to the affinity, and the van der Waals interactions, mainly due to stacking of the hydrophobic face of sugar to side chains of aromatic amino acids, are responsible of the specificity.<sup>52</sup> The interaction between PA-IIL soluble bacterial lectin and monosaccharides is clearly of a different type: almost no hydrophobic contact is involved and the interaction displays a higher affinity

than other protein–carbohydrate interactions, due to the direct involvement of two cations in the binding.

The very fine present analysis of the interaction gives some clues for the development of high-affinity ligands. Possible routes for a rational design of high-affinity anti-adhesive compounds encompass further stabilization of the charge distribution, enhancement of the hydrophobic interactions that are minimal in the complexes studied to date and rational use of the presence of water molecules. These approaches could yield compounds of therapeutic interest not only against *P. aeruginosa*, but also against several other related dangerous pathogens that use similar lectins for their colonization strategy.

### ACKNOWLEDGEMENTS

We thank the ESRF, Grenoble, for access to synchrotron data collection facilities. The work was supported by Ministry of Education (LŠ, MP, JK, MW contract LN00A016), and by Grant Agency (CH, Grant 202/01/D110) of the Czech Republic, by CNRS, by French Ministry of Research ACI Microbiologie program and by Vaincre la Mucoviscidose. Travels and visits between NCBR and CERMAV are supported by a BARRANDE exchange program.

### REFERENCES

- Scharfman A, Degroote S, Beau J, Lamblin G, Roussel P, Mazurier J. *Pseudomonas aeruginosa* binds to neoglycoconjugates bearing mucin carbohydrate determinants and predominantly to sialyl-Lewis x conjugates. *Glycobiology* 1999;9:757–764.
- Lamblin G, Degroote S, Perini JM, Delmotte P, Scharfman A, Davril M, Lo-Guidice JM, Houdret N, Dumur V, Klein A, Roussel P. Human airway mucin glycosylation: A combinatorial of carbohydrate determinants which vary in cystic fibrosis. *Glycoconj J* 2001;18:661–684.
- Roussel P, Lamblin G. The glycosylation of airway mucins in cystic fibrosis and its relationship with lung infection by *Pseudomonas aeruginosa*. *Adv Exp Med Biol* 2003;535:17–32.
- Gilboa-Garber N. *Pseudomonas aeruginosa* lectins. *Meth Enzymol* 1982;83:378–385.
- Avichezer D, Katcoff DJ, Garber NC, Gilboa-Garber N. Analysis of the amino acid sequence of the *Pseudomonas aeruginosa* galactophilic PA-I lectin. *J Biol Chem* 1992;267:23023–23027.
- Gilboa-Garber N, Katcoff DJ, Garber NC. Identification and characterization of *Pseudomonas aeruginosa* PA-IIL lectin gene and protein compared to PA-IL. *FEMS Immunol Med Microbiol* 2000;29:53–57.
- Glick J, Garber NC. The intracellular localization of *Pseudomonas aeruginosa* lectins. *J Gen Microbiol* 1983;9:3085–3090.
- Loris R, Tielker D, Jaeger K-E, Wyns L. Structural basis of carbohydrate recognition by the lectin LecB from *Pseudomonas aeruginosa*. *J Mol Biol* 2003;331:861–870.
- Imberty A, Wimmerova M, Mitchell EP, Gilboa-Garber N. Structures of the lectins from *Pseudomonas aeruginosa*: Insights into molecular basis for host glycan recognition. *Microb Infect* 2004;6:221–228.
- Garber N, Guempel U, Gilboa-Garber N, Doyle RJ. Specificity of the fucose-binding lectin of *Pseudomonas aeruginosa*. *FEMS Microbiol Lett* 1987;48:331–334.
- Mitchell E, Houles C, Sudakevitz D, Wimmerova M, Gautier C, Pérez S, Wu AM, Gilboa-Garber N, Imberty A. Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nat Struct Biol* 2002;9:918–921.
- Sudakevitz D, Kostlanova N, Blatman-Jan G, Mitchell EP, Lerrer B, Wimmerova M, Katcoff DJ, Imberty A, Gilboa-Garber N. A new *Ralstonia solanacearum* high affinity mannose-binding lectin RS-IIL structurally resembling the *Pseudomonas aeruginosa* fucose-specific lectin PA-IIL. *Mol Microbiol* 2004;52:691–700.

13. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter CWJ, Sweet RM, editors. *Macromolecular Crystallography, Part A*. San Diego: Academic Press; 1997. p 307–326.
14. Collaborative computational project number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 1994;50:760–763.
15. Cowtan KD, Zhang KY. Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol* 1999;72:245–270.
16. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 1999;6:458–463.
17. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 1991;47:110–119.
18. Murshudov GN, A.A.Vagin, E.J.Dodson. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 1997;53:240–255.
19. Sheldrick GM, Schneider TR. Shelxl: High-resolution refinement. *Meth Enzymol* 1997;277:319–343.
20. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242.
21. Tripos Associates, 6.4. 1699 S. Hanley Road, Suite 303, St Louis, MO 63144 USA.
22. Imberty A, Bettler E, Karababa M, Mazeau K, Petrova P, Pérez S. Building sugars: The sweet part of structural biology. In Vijayan M, Yathindra N, Kolaskar AS, editors. *Perspectives in Structural Biology*. Hyderabad: Indian Academy of Sciences and Universities Press; 1999. p 392–409.
23. Clark M, Cramer RDI, van den Opendbosch N. Validation of the general purpose Tripos 5.2 force field. *J Comput Chem* 1989;10:982–1012.
24. MOPAC, 2002. Tokyo: Fujitsu Limited; 1999.
25. Dewar MJS, Zebisch EG, Healy EF, Stewart JJP. AM1: A new general purpose quantum mechanical model. *J Am Chem Soc* 1993;107:3902–3909.
26. Stewart JJP. Optimisation of parameters for semi-empirical methods I. Method. *J Comp Chem* 1989;10:209–220.
27. Stewart JJP. Optimisation of parameters for semi-empirical methods II. Applications. *J Comp Chem* 1989;10:221–264.
28. Damborsky J, Prokop M, Koca J. TRITON: graphic software for rational engineering of enzymes. *Trends Biochem Sci* 2001;26:71–73.
29. Mulliken RS. Electronic population analysis on LCAO-MO molecular wave functions I. *J Chem Phy* 1955;23:1833–1846.
30. Bayly CI, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the RESP model. *J Phys Chem* 1993;97:10269–10280.
31. Cornell WD, Cieplak P, Bayly CI, Kollman PA. Application of RESP charges to calculate conformational energies, hydrogen-bond energies, and free-energies of solvation. *J Am Chem Soc* 1993;115:9620–9631.
32. Gaussian, 98, Revision A. 9. Pittsburgh: Gaussian, Inc.; 1998.
33. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;33:389–3402.
34. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
35. Nicholas KB, Jr. NHB, Deerfield DWI. GeneDoc: Analysis and visualization of genetic variation. *EMBNW News* 1997;4:14.
36. 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.
37. Varrot A, Davies GJ. Direct experimental observation of the hydrogen-bonding network of a glycosidase along its reaction coordinate revealed by atomic resolution analyses of endoglucanase Cel5A. *Acta Crystallogr D Biol Crystallogr* 2003;59:447–452.
38. Magini M, Tanaka N, Ohtaki H, Tamamushi R. Ions and molecules in solution. Amsterdam: Elsevier; 1983. 97 p.
39. Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangelot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, Chandler M, Choise N, Claudel-Renard C, Cunnac S, Demange N, Gaspin C, Lavie M, Moisan A, Robert C, Saurin W, Schiex T, Siguiet P, Thebault P, Whalen M, Wincker P, Levy M, Weissenbach J, Boucher CA. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature* 2002;415:497–502.
40. Brazilian National Genome Project Consortium. The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability. *Proc Natl Acad Sci USA* 2003;100:11660–11665.
41. Dam TK, Brewer CF. Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem Rev* 2002;102:387–429.
42. Gupta D, Cho M, Cummings RD, Brewer CF. Thermodynamics of carbohydrate binding to galectin-1 from Chinese hamster ovary cells and two mutants. A comparison with four galactose-specific plant lectins. *Biochemistry* 1996;35:15236–15243.
43. Lemieux RU, Delbaere LT, Beierbeck H, Spohr U. Involvement of water in host-guest interactions. *Ciba Found Symp* 1991;158:231–245.
44. Schwarz FP, Puri KD, Bhat RG, Surolia A. Thermodynamics of monosaccharide binding to concanavalin A, pea (*Pisum sativum*) lectin, and lentil (*Lens culinaris*) lectin. *J Biol Chem* 1993;268:7668–7677.
45. Surolia A, Sharon N, Schwarz FP. Thermodynamics of monosaccharide and disaccharide binding to *Erythrina corallodendron* lectin. *J Biol Chem* 1996;271:17697–17703.
46. Elgavish S, Shaanan B. Structures of the *Erythrina corallodendron* lectin and of its complexes with mono- and disaccharides. *J Mol Biol* 1998;277:917–932.
47. Drickamer K. Ca(2+)-dependent sugar recognition by animal lectins. *Biochem Soc Trans* 1996;24:146–150.
48. Thompson D, Pepys MB, Tickle I, Wood S. The structures of crystalline complexes of human serum amyloid P component with its carbohydrate ligand, the cyclic pyruvate acetal of galactose. *J Mol Biol* 2002;320:1081–1086.
49. Duran N, Menck CF. *Chromobacterium violaceum*: a review of pharmacological and industrial perspectives. *Crit Rev Microbiol* 2001;27:201–222.
50. Mahenthalingam E, Baldwin A, Vandamme P. *Burkholderia cepacia* complex infection in patients with cystic fibrosis. *J Med Microbiol* 2002;51:533–538.
51. Sacchettini JC, Baum LG, Brewer CF. Multivalent protein-carbohydrate interactions. A new paradigm for supermolecular assembly and signal transduction. *Biochemistry* 2001;40:3009–3015.
52. Vyas N. Atomic features of protein-carbohydrate interactions. *Curr Opin Struct Biol* 1991;1:732–740.
53. Kraulis P. Molscript: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 1991;24:946–950.
54. Merrit EA, Murphy ME. Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr D Biol Crystallogr* 1994;50:869–873.

(Article n°2)

**Binding of Different Monosaccharides by Lectin PA-III from *Pseudomonas aeruginosa*: Thermodynamics data correlated with X-ray structures**

**Sabin C, Mitchell EP, Pokorná M, Gautier C, Utile J.P, Wimmerová M, Imberty A**

*FEBS Letters*, 580: 982-87 (2006)



# Binding of different monosaccharides by lectin PA-IIL from *Pseudomonas aeruginosa*: Thermodynamics data correlated with X-ray structures

Charles Sabin<sup>a,b</sup>, Edward P. Mitchell<sup>b</sup>, Martina Pokorná<sup>c</sup>, Catherine Gautier<sup>a</sup>,  
Jean-Pierre Utille<sup>a</sup>, Michaela Wimmerová<sup>c</sup>, Anne Imberty<sup>a,\*</sup>

<sup>a</sup> CERMAV-CNRS (affiliated with Université Joseph Fourier), 601 rue de la Chimie, Grenoble BP53, F-38041 Grenoble cedex 09, France

<sup>b</sup> E.S.R.F. Experiments Division, BP-220, F-38043 Grenoble cedex 09, France

<sup>c</sup> National Centre for Biomolecular Research and Department of Biochemistry, Masaryk University, Kotlarska 2, 611 37 Brno, Czech Republic

Received 9 December 2005; revised 10 January 2006; accepted 12 January 2006

Available online 19 January 2006

Edited by Irmgard Sinning

**Abstract** The lectin from *Pseudomonas aeruginosa* (PA-IIL) is involved in host recognition and biofilm formation. Lectin not only displays an unusually high affinity for fucose but also binds to L-fucose, L-galactose and D-arabinose that differ only by the group at position 5 of the sugar ring. Isothermal calorimetry experiments provided precise determination of affinity for the three methyl-glycosides and revealed a large enthalpy contribution. The crystal structures of the complexes of PA-IIL with L-galactose and Met-β-D-arabinoside have been determined and compared with the PA-IIL/fucose complex described previously. A combination of the structures and thermodynamics provided clues for the role of the hydrophobic group in affinity.

© 2006 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

**Keywords:** Lectin; Cystic fibrosis; Crystal structure; Thermodynamics; Galactose; Arabinose; *Pseudomonas aeruginosa*

## 1. Introduction

*Pseudomonas aeruginosa* is a gram negative bacterium and an opportunistic pathogen that infects immunocompromised patients and which is often involved in nosocomial infections. The bacterium is also responsible for chronic respiratory tract infections of individuals under mechanical ventilation and patients suffering from cystic fibrosis (CF). Carbohydrates play an important role in infections, since they are a target for bacterial binding through virulence factors such as adhesions, present on pili or flagella, and soluble lectins [1]. In the particular case of CF lungs, an increase of fucosylation is observed, both at the levels of epithelia glycoconjugates [2] and altered mucins [3], and a fucose binding lectin from *P. aeruginosa* (PA-IIL) has been demonstrated to bind specifically to CF air-line culture cells [4]. The tetrameric lectin that was first identi-

fied and characterised from *P. aeruginosa* cytoplasm [5] has been shown to be present in large quantities on the outer membrane of the bacteria and to also be involved in biofilm formation [6].

Through protein crystallography at high resolution, a precise three-dimensional knowledge of lectin has allowed the structural basis of the interaction between PA-IIL and fucose and fucose-containing oligosaccharides to be understood [7,8]. An unusual binding mode, mediated by two calcium ions, is responsible for the micromolar affinity for fucose which is much higher than that classically observed in protein/carbohydrate interactions [9]. Previous inhibition tests have demonstrated that other monosaccharides are also recognised by PA-IIL with affinity (strongest first) L-fucose (Fuc) > L-galactose (L-Gal) > D-arabinose (Ara) > D-fructose (Fru) > D-mannose (Man) [10]. Indeed, these monosaccharides all have either an L-galacto or a D-manno configuration that contains two equatorial and one axial hydroxyl groups when in the <sup>1</sup>C<sub>4</sub> and <sup>4</sup>C<sub>1</sub> pyranose conformation, respectively. This set of groups are needed for the coordination of the two calcium ions (Fig. 1) as was observed in the X-ray crystal structures of the PA-IIL complexes with mannose and fructose [11].

In the present work, we focus on comparing the binding modes of the three monosaccharides with highest affinity for PA-IIL, i.e. Fuc, L-Gal and Ara, that differ only by the nature of the group at C5 (Fig. 1). Precise equilibrium association constants, together with their enthalpic and entropic contributions, were evaluated for PA-IIL interacting with these monosaccharides and two additional crystal structures have been determined at very high resolution for the complexes involving L-Gal and Ara. Such studies can provide the basis for designing higher affinity ligands that could be of therapeutic interest as antibacterial compounds.

## 2. Materials and methods

Recombinant PA-IIL was purified from *Escherichia coli* BL21(DE3) containing the plasmid pET25pa2l as described previously [9]. Fuc, Man, β-methyl-D-arabinopyranoside (Me-β-Ara) and α-methyl-D-mannopyranoside (Me-α-Man) were purchased from Sigma. α-Methyl-L-fucopyranoside (Me-α-Fuc) was purchased from Interchim. L-Gal was either purchased from Sigma or Institute of Chemistry (Bratislava) or prepared from L-galactono-1,4-lactone (Fluka) by preparing 1,2,3,4,6-penta-O-acetyl-L-galactose according to Thiem et al. [12] followed by deacetylation using the conventional

\*Corresponding author. Fax : +33 476 54 72 03.

E-mail address: imberty@cermav.cnrs.fr (A. Imberty).

**Abbreviations:** PA-IIL, lectin II from *Pseudomonas aeruginosa*; CF, cystic fibrosis; L-Gal, L-galactose; Fuc, L-fucose; Man, D-mannose; Ara, D-arabinose; Me-α-L-Gal, α-methyl-L-galactopyranoside; Me-α-Fuc, α-methyl-L-fucopyranoside; Me-α-Man, α-methyl-D-mannopyranoside; Me-β-Ara, β-methyl-D-arabinopyranoside

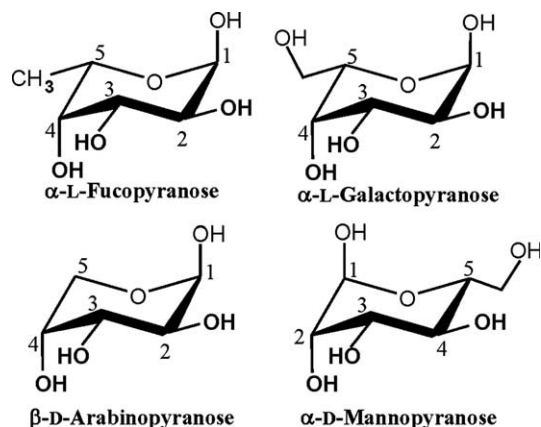


Fig. 1. Schematic representation of the monosaccharides of interest together with numbering of ring carbons. The three hydroxyl groups involved in the coordination of calcium when complexed with PA-IIL are indicated in bold.

Zemplén reaction ( $\text{MeO}^- \text{Na}^+$  1 M in MeOH).  $\alpha$ -Methyl-L-galactopyranoside (Me- $\alpha$ -L-Gal) was synthesised from L-Gal with the following procedure: Me-L-galactoside isomers were prepared by refluxing L-Gal in dry MeOH in the presence of  $\text{BF}_3$  etherate for 15 h. Silica gel chromatography led to the separation of pure Me- $\alpha$ -L-Gal;  $^1\text{H}$  NMR of the acetylated derivative showed a coupling constant  $J_{1,2} = 3.66$  Hz in agreement with an  $\alpha$ -L-configuration [12].  $^{13}\text{C}$  NMR ( $\text{D}_2\text{O}$ , 75 MHz): 100.41 (C1), 71.74, 70.47, 70.26, 69.19 (C2, C3, C4, C5), 62.25 (C6), 56.06 ( $\text{OCH}_3$ ).

Isothermal calorimetry (ITC) experiments were carried out at 25 °C with a VP-ITC calorimeter from Microcal (Northampton, MA). Proteins and sugars were dissolved in the same buffer (0.1 M Tris-HCl with 0.03 mM  $\text{CaCl}_2$  at pH 7.5, filtered through 0.22  $\mu\text{m}$   $\phi$ ). The protein concentration in the microcalorimeter cell (1.4 ml) varied from 0.05 to 0.3 mM in order to keep the  $C$  value between 5 and 100 (where  $C = K_a M$ , where  $M$  is the initial concentration of the macromolecule). A total of 30 injections of 10  $\mu\text{l}$  of sugar solution at concentrations varying from 0.4 to 3 mM were added at intervals of 5 min whilst stirring at 310 rpm. Control experiments yielded insignificant heats of dilution. The thermodynamic parameters  $n$ ,  $K_a$  and  $\Delta H$  were obtained by fitting the classical titration curve [13] to experimental data using Origin7.0 software (Microcal).

The crystals were grown using the hanging drop vapour diffusion method and Hampton Screens I and II (Hampton Research) at 20 °C. The best results were obtained using 2.5  $\mu\text{l}$  of lyophilised purified PA-IIL dissolved in water (10 mg  $\text{ml}^{-1}$ ) in the presence of Me- $\beta$ -Ara or L-Gal (250  $\mu\text{g}$   $\text{ml}^{-1}$ ) and  $\text{CaCl}_2$  salt (2 mM) mixed with 2.5  $\mu\text{l}$  of the reservoir solution consisting of 1.75 M ammonium sulfate in 0.1 M Tris-HCl, pH 8.5. Crystals of both complexes were cryo-cooled to 100 K, after soaking them for as short a time as possible in 25% glycerol v/v in the precipitant solution. All data images were recorded on an ADSC Q4R CCD detector (Quantum corp.) using X-rays of 0.93 Å wavelength on beamline ID14-1 or ID14-4 at the ESRF (Grenoble, France). The data were processed with MOSFLM [14] and scaled and converted to structure factors using the CCP4 programme suite [15] (Table 1).

The phases were first determined by the molecular replacement technique with the MOLREP program of CCP4 using the monomeric PA-IIL/Fuc structure (PDB code 1UZV) [9] stripped of water molecules, calcium ions and sugar ligands as the search probe. Four clear solutions were found for both complexes, resulting in tetramers in the asymmetric units. Initial structures were entirely rebuilt automatically using ARP/warp [16] and clear electron density could be observed for the monosaccharides (Fig. 2). Refinement was performed with REFMAC [17]. Alternating cycles of refinement with manual construction using O [18] of missing residues, alternative conformations, water molecules, calcium ions and sugar ligands resulted in the final structures. Coordinates and structure factors of

Table 1  
Data collection and refinement statistics

	PA-IIL/Me- $\beta$ -Ara	PA-IIL/Me- $\alpha$ -L-Gal
<i>Data collection</i>		
Beamline (ESRF)	ID14-2	ID14-4
Wavelength (Å)	0.933	0.932
Resolution range (Å)	1.8–30.0	1.5–19.4
Highest resolution shell (Å)	1.80–1.90	1.50–1.55
Space group	P2 <sub>1</sub>	P2 <sub>1</sub>
Cell dimensions		
<i>a</i> (Å)	50.64	52.63
<i>b</i> (Å)	80.17	54.53
<i>c</i> (Å)	52.52	94.38
$\beta$ (°)	109.92	94.38
Matthews coef (Å <sup>3</sup> /DA)	1.74	1.90
Solvent content (%)	29	35
Measured reflections	133 138	213 637
Unique reflections	35 769	63 528
Average multiplicity <sup>a</sup>	3.73 (3.45)	3.82 (3.80)
Completeness <sup>a</sup> (%)	97.9 (89.0)	96.8 (95.1)
Average <sup>a</sup> $I/\sigma I$	7.50 (3.64)	10.35 (3.08)
<i>R</i> merge <sup>a,b</sup> (%)	7.4 (19.0)	5.0 (23.0)
Wilson <i>B</i> -factor (Å <sup>2</sup> )	11.4	10.3
<i>Refinement</i>		
<i>R</i> <sub>crys</sub> <sup>c</sup> (no. of observations)	0.137 (34 653)	0.143 (61 594)
<i>R</i> <sub>free</sub> <sup>d</sup> (no. of observations)	0.172 (1116)	0.166 (1934)
<i>RMS deviation from ideality</i>		
Bonds (Å)	0.011	0.012
Angles (°)	1.320	1.398
<i>Total number of atoms:</i>		
Protein	3308	3308
Ligand	44	66
Calcium	8	8
Water	612	430
<i>Average refined B-factors per chains (A/B/C/D)</i>		
Protein	8.6/9.5/9.7/8.3	8.4/10.1/10.0/8.9
Ligand	7.7/10.8/10.6/7.9	6.5/8.8/8.3/8.3
Calcium	6.4/8.5/6.7/8.7	5.9/6.9/6.0/6.0
Water	25.3/26.2/25.1/26.2	22.4/ 22.9/23.5/21.9
Number of side chains modelled with alternative conformations		
	17	16
	Chain A: S23, K62, V81	Chain A: S23, L31, N46, S68, V77, V81
	Chain B: V54, V77, V81	Chain B: S59, V77, V81, I82
	Chain C: S23, T39, Q53, V81, E86	Chain C: N29, S41, S77
	Chain D: S23, T39, N70, V77, V81, E86	Chain D: N46, V77, V81
PDB deposition code	2BOJ	2BP6

<sup>a</sup>Numbers in parentheses refer to the highest resolution shell.

<sup>b</sup> $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$ , where  $I$  = observed intensity.

<sup>c</sup> $R_{\text{cryst}} = \sum |F_o - F_c| / \sum |F_o|$ , where  $|F_o|$  = observed structure factor amplitude and  $|F_c|$  = calculated structure factor amplitude.

<sup>d</sup> $R_{\text{free}}$  is  $R_{\text{cryst}}$  for 3% of reflections excluded from the refinement.

the PA-IIL complexes with L-Gal and Me- $\beta$ -Ara have been deposited in the Protein Data Bank under codes 2BP6 and 2BOJ, respectively.

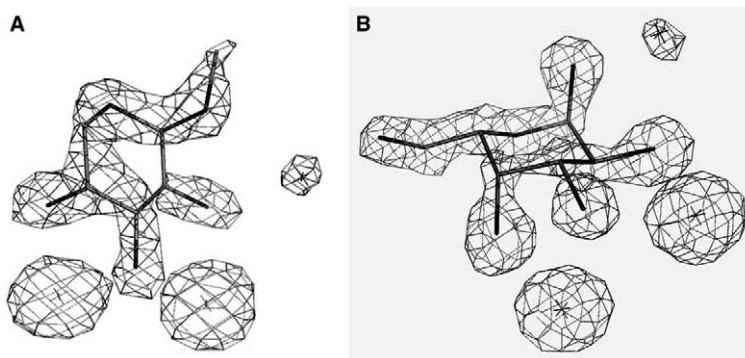


Fig. 2. Initial sigma-weighted  $F_o - F_c$  electron density map. (A) The map has been contoured at  $3\sigma$  around the Met- $\beta$ -D-arabinopyranoside in the PA-IIL-Me- $\beta$ -D-Ara complex calculated after molecular replacement and after the first cycle of refinement without ligand and water molecules. (B) Same representation for the L-Gal in the PA-IIL-L-Gal complex. The final positions of the monosaccharides, calcium and water in the model have been superimposed as stick representation for comparison. Molecular drawings were prepared with Pymol (DeLano Scientific LLC).

### 3. Results

#### 3.1. Thermodynamics of PA-IIL binding to monosaccharides

Thermodynamic data were recorded for the interaction of PA-IIL with methyl-glycosides of the monosaccharides of interest, i.e. Me- $\alpha$ -L-Gal, Me- $\beta$ -Ara and also Me- $\alpha$ -Fuc and Me- $\alpha$ -Man for comparison. The use of methyl derivatives avoids the problems related to anomeric disorder in solution and also gives a better mimic of real biological interactions where the monosaccharides are not free but linked to glycoconjugates.

A typical titration curve of PA-IIL interacting with Me- $\alpha$ -Fuc is shown in Fig. 3. The thermodynamic parameters are reported in Table 2. The stoichiometry of binding is usually close to 1, whilst some protein batches occasionally display lower values (minimum 0.75). Nevertheless, the other binding constants are not affected by the variation in stoichiometry. The affinity is maximal for binding to Me- $\alpha$ -Fuc with a dissociation constant of 430 nM. The affinity for the methyl-fucoside is 10-fold higher than for the unmodified sugar, an effect that is usually observed in protein-carbohydrate interactions. The qualitative order of preference that was observed by inhibition of haemagglutination experiments [10] is confirmed by the present data. Quantitatively, PA-IIL affinity for Me- $\alpha$ -Fuc is threefold stronger than for Me- $\alpha$ -L-Gal and Me- $\beta$ -Ara and 140-fold better than for Me- $\alpha$ -Man.

All complexes formed here are driven by enthalpy. The interaction of the lectin with monosaccharides of the L-galacto configuration gives a higher enthalpy of binding ( $\Delta H < -30$  kJ/mol) than for mannose ( $\Delta H = -20$  kJ/mol). The PA-IIL/Me- $\alpha$ -Fuc complex has the highest enthalpy of binding with  $\Delta H = -40$  kJ/mol. The entropy contribution  $-T\Delta S$  can be slightly unfavourable, as in the case of Me- $\alpha$ -Fuc and Me- $\beta$ -Ara, or slightly favourable, as in the case of Me- $\alpha$ -L-Gal and Me- $\alpha$ -Man, but it is always a low contribution with a maximum of 17% of the free energy of binding.

#### 3.2. Crystal structure of PA-IIL/L-Gal complex

The overall structure of the PA-IIL complex with L-Gal to 1.4 Å resolution is similar to the PA-IIL/Fuc complex determined previously [7,9]. All 114 amino acids can be observed together with the two calcium ions and the monosaccharide of each subunit of the tetramer (Fig. 4). Two sulfate ions and 430 water molecules were incorporated in the model.

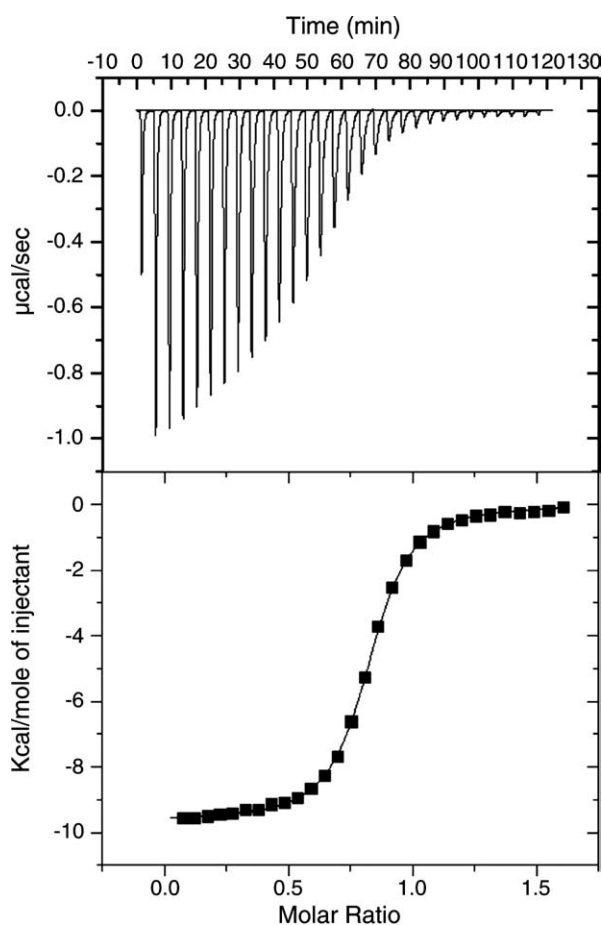


Fig. 3. Titration microcalorimetry results of Me- $\alpha$ -Fuc (0.44 mM) binding to PA-IIL (55.8  $\mu$ M) in 100 mM Tris buffer, pH 7.5, with 0.03 mM  $\text{CaCl}_2$ . Top: data obtained from 30 automatic injections (10  $\mu$ l) of Me- $\alpha$ -Fuc each into the cell containing PA-IIL. Lower: plot of the total heat released as a function of total ligand concentration for the titration shown in the upper panel. The solid line represents the best least-squares fit for the obtained data.

The monomer is a  $\beta$ -sandwich composed of 8 antiparallel strands of which the first five are arranged in Greek key and which form an acidic pocket allowing the binding of two intimate  $\text{Ca}^{2+}$  ions (3.75 Å separation). The amino acids



Table 2  
Thermodynamics–ITC data for different monosaccharides interacting with PA-IIL<sup>a</sup>

	$K_a$ $10^4$ $M^{-1}$	$K_d$ $10^{-6}$ $M$	$n$	$-\Delta G$ (kJ/mol)	$-\Delta H$ (kJ/mol)	$-T\Delta S$ (kJ/mol)
L-Fuc <sup>b</sup>	34 ( $\pm 3$ )	2.9 ( $\pm 0.03$ )	0.96 ( $\pm 0.04$ )	31.5 ( $\pm 0.2$ )	31.2 ( $\pm 0.1$ )	-0.3 ( $\pm 0.1$ )
Me- $\alpha$ -Fuc	235 ( $\pm 8$ )	0.43 ( $\pm 0.01$ )	0.77 ( $\pm 0.03$ )	36.4 ( $\pm 0.1$ )	41.3 ( $\pm 1$ )	4.9 ( $\pm 1$ )
Me- $\alpha$ -L-Gal	85 ( $\pm 5$ )	1.17 ( $\pm 0.07$ )	0.90 ( $\pm 0.04$ )	33.8 ( $\pm 0.1$ )	31.5 ( $\pm 0.6$ )	-2.3 ( $\pm 0.7$ )
Me- $\beta$ -Ara	60 ( $\pm 5$ )	1.7 ( $\pm 0.1$ )	1.10 ( $\pm 0.03$ )	33.0 ( $\pm 0.2$ )	37.9 ( $\pm 0.8$ )	5.4 ( $\pm 1.0$ )
Me- $\alpha$ -Man	1.42 ( $\pm 0.05$ )	71 ( $\pm 3$ )	0.94 ( $\pm 0.01$ )	23.7 ( $\pm 0.1$ )	17.8 ( $\pm 0.4$ )	-5.9 ( $\pm 0.3$ )

<sup>a</sup>Values in parentheses are standard deviations from three separate isothermal titrations.

<sup>b</sup>Taken from Perret et al. [8].

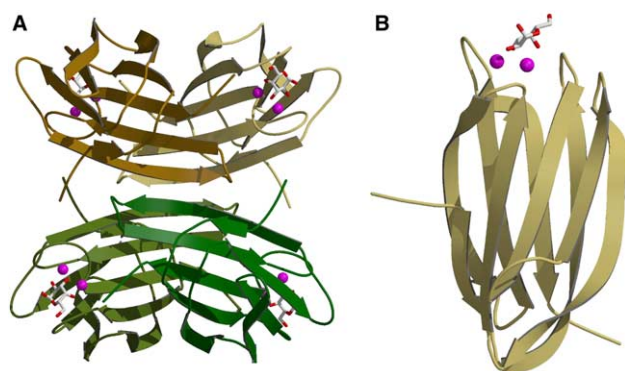


Fig. 4. Ribbon representation of the PA-II lectin/L-Gal complex. (A) Tetramer consisting of the asymmetric unit with stick representation of L-Gal ligand and cpk representation of the calcium ions. (B) Same representation for the monomer. Molecular drawings were prepared with MOLSCRIPT [23] and RASTER-3D [24].

of the pocket, Asn21, Glu95, Asp99, Asp101, Asn103, Asp104, as well as the Gly114\* terminus of the adjacent monomer, complete a large part of the coordination sphere of the two calcium ions (Fig. 5). One calcium ion is coordinated by five oxygen atoms from carboxylate groups and two from the sugar ligand, whereas the other calcium receives three oxygen atoms from carboxylate groups, two from carbonyl groups and two from the sugar ligand. The sugar ring oxygen O5 forms a hydrogen bond with the NH backbone group of Ser23. One water molecule plays a special role by bridging between the L-Gal oxygen atoms O1 and O2 and the main chain NH groups of residues Thr98 and Asp99. Details of distances are listed in Table 3.

The hydroxymethyl group of L-Gal is tucked into a shallow cavity with amphiphilic character between Thr45 and Ser23. The CH2 group at C6 makes hydrophobic contacts with the methyl group of Thr45. Some conformational disorder is observed for the O6 atom of the hydroxymethyl group. In three of the four binding sites, the O6 atom is in the *gauche-trans* (*gt*) conformation (relative to O5 and C4 [19]) and makes a hydrogen bond to a water molecule. This water molecule does not bridge directly to the protein but is part of a complex network with variations between the binding sites. In the fourth binding site (chain A), the hydroxymethyl group adopts another conformation, referred to as *trans-gauche* (*tg*), and makes contact to another water molecule. These two rotamers of the hydroxymethyl group of galactose have been reported to be the two most populated ones, both in the crystalline state [19] and in water solution [20]. Interestingly, this alternative conforma-

tion makes some free space for the hydroxyl group of Ser23 that becomes flexible and adopts two alternative conformations.

### 3.3. Crystal structure of PA-IIL/Me- $\beta$ -Ara complex

The crystal structure of the PA-IIL complex with Me- $\beta$ -Ara is isomorphous to those of PA-IIL/Fuc and PA-IIL/L-Gal and the overall tetrameric structures are similar. In addition to eight calcium ions and four Me- $\beta$ -Ara ligands, one sulfate and 612 water molecules could be located in the electron density map. The orientation of the Me- $\beta$ -Ara ring in the binding site is similar to that of L-Gal described above (Fig. 5 and Table 3). The ring carbon atom C-5 does not carry any substituent and therefore leaves an empty space between Thr45 and Ser23 side chains. As a result, the hydroxyl group of Ser23 adopts two alternative conformations in all binding sites except in chain B.

## 4. Discussion

A comparison of the two crystal structures described here to that reported previously for the PA-IIL/Fuc complex [9] allows the particular role of the moiety at position 5 of the ring to be explored. Indeed, all other contacts between the protein and the monosaccharides are identical in the different complexes and the differences in binding affinity must therefore be correlated with this particular group.

In terms of affinity, the present study highlights the important contribution of hydrophobic contacts. In general, interactions between proteins and carbohydrates present a balance of hydrogen bonds and hydrophobic contacts. PA-IIL is a rather unique case since its interaction with fucose is established mainly through coordination by calcium ions and hydrogen bonds. The only hydrophobic interaction is established between the methyl group at C5 of fucose and the side chain of Thr45. When this interaction is weakened, as in the complex with L-Gal, or absent, as with Me- $\beta$ -Ara, the affinity is decreased by 3-fold.

The relative contributions of the different thermodynamic parameters are more difficult to rationalise. The entropic term is the sum of multiple effects. A favourable entropic contribution is obtained when the increase of solvent entropy coming from water driven out of the site is correlated with minimal loss of conformational degrees of freedom in either ligand or protein chains during binding. This is not usually the case in protein-carbohydrate interactions where a strong entropy barrier is usually observed [21,22].

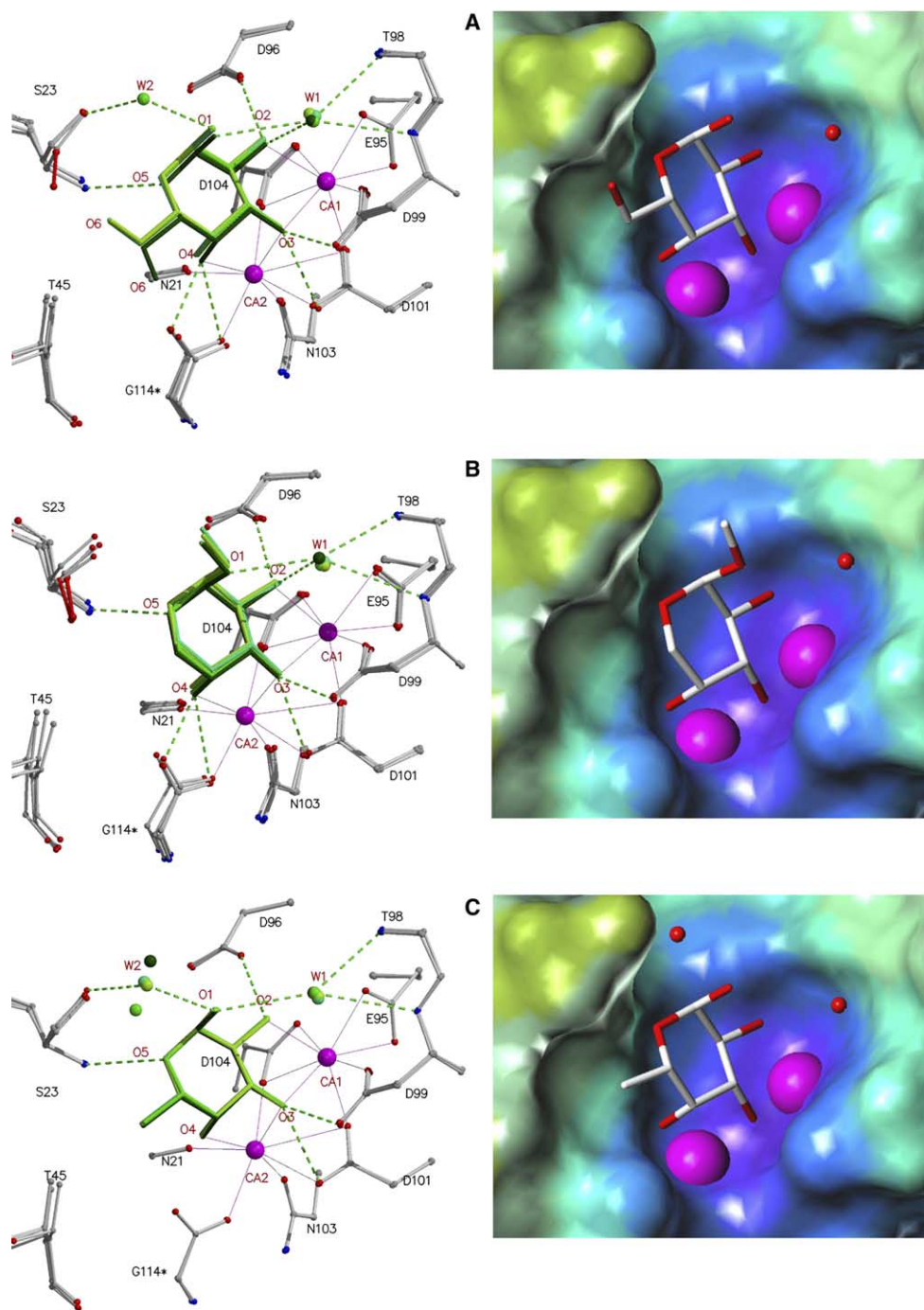


Fig. 5. Binding sites of PA-IIL complexed with monosaccharides. *Left panel*: superposition of the four binding sites of the asymmetric unit. Sugar and water molecules belonging to the same monomer are drawn with the same shade of green. Coordination contacts are indicated by blue solid lines and hydrogen bonds by green dashed lines. The \* symbol indicates the terminal glycine from the second monomer. *Right panel*: Accessible surface of the protein calculated with the MOLCAD program [25] for one binding site. (A) PA-IIL/L-Gal complex, (B) PA-IIL/Me- $\beta$ -Ara complex, (C) PA-IIL/Fuc complex from PDB code 1UZV [9].

However, in the sugar free structure of PA-IIL [11] three water molecules strongly coordinated to the calcium ions replace the three hydroxyl groups O2, O3 and O4 of the fucose complex. The release of these water molecules that are more strongly bound than in a classical carbohydrate binding site can explain the absence of a strong entropy bar-

rier for the PA-IIL/monosaccharide interaction. In our study, galactoside and mannoside are the two glycosides with favourable entropy of binding and this observation may be correlated to the conformational flexibility of the O6 hydroxyl group, as observed in the X-ray structures of both complexes.

Table 3  
Comparison of contact distances in Å the PA-IIL binding site<sup>a</sup>

		PA-IIL/ Fuc <sup>b</sup>	PA-IIL/ Me-β-Ara	PA-IIL/ L-Gal
<i>Calcium coordination</i>				
Ca-1	O2	2.50	2.53	2.51
Ca-1	O3	2.48	2.50	2.49
Ca-2	O3	2.48	2.46	2.46
Ca-2	O4	2.49	2.52	2.50
<i>Hydrogen bonds</i>				
Asp96.OD1	O2	2.65	2.63	2.62
Asp 99.OD1	O3	2.98	2.99	3.00
Asp 99.OD2	O3	2.56	2.54	2.57
Asp 101.OD1	O3	2.95	2.97	2.94
Asp 101.OD2	O3	2.93	2.92	2.92
Asp 104.OD2	O3	3.00	3.03	3.05
Asn 21.O	O4	3.07	3.01	3.02
Gly 114*.OXT <sup>c</sup>	O4	2.56	2.47	2.51
Ser23. N	O5	2.95	2.93	2.95
Wat_1 <sup>d</sup>	O1	3.2	3.03	3.17
Wat_1	O2	3.05	3.02	3.05
Wat_2 <sup>e</sup>	O1	3.05	–	2.84
<i>Hydrophobic contacts</i>				
Thr45.CG2	C6 <sub>Me</sub>	4.12	–	4.08

<sup>a</sup>Distances given are a mean from the distances observed in the four independent molecules in the asymmetric unit.

<sup>b</sup>Values taken from PDB structure 1UZV [9].

<sup>c</sup>C-terminal residue from neighbouring monomer.

<sup>d</sup>Water bridging to Thr98.NH and Asp99.NH.

<sup>e</sup>Water bridging to Ser23.OH.

**Acknowledgements:** The help of Jan Adam is acknowledged for microcalorimetry experiments. We thank the ESRF, Grenoble, for access to synchrotron data collection facilities. The work was supported by Ministry of Education of The Czech Republic (Contract MSM0021622413), by CNRS, by French Ministry of Research ACI Microbiologie program, by the Mizutani Foundation for Glycosciences and by French Association Vaincre la Mucoviscidose.

## References

[1] 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.

[2] Rhim, A.D., Stoykova, L.I., Trindade, A.J., Glick, M.C. and Scanlin, T.F. (2004) Altered terminal glycosylation and the pathophysiology of CF lung disease. *J. Cyst. Fibros.* 3, 95–96.

[3] Roussel, P. and Lamblin, G. (2003) The glycosylation of airway mucins in cystic fibrosis and its relationship with lung infection by *Pseudomonas aeruginosa*. *Adv. Exp. Med. Biol.* 535, 17–32.

[4] Stoykova, L.I., Ellway, J., Rhim, A.D., Kim, D.J., Hlick, M.C. and Scanlin, T.F. (2005) Binding of *Pseudomonas aeruginosa* lectin LecB to cystic fibrosis airway cells is inhibited by fucosylated compounds: implication for therapy, in: Proceedings of the 2005 Annual Conference of the Society for Glycobiology, no. 169. *Glycobiology* 15, 1228.

[5] Gilboa-Garber, N. (1982) *Pseudomonas aeruginosa* lectins. *Meth. Enzymol.* 83, 378–385.

[6] Tielker, D., Hacker, S., Loris, R., Strathmann, M., Wingender, J., Wilhelm, S., Rosenau, F. and Jaeger, K.-E. (2005) *Pseudomonas*

*aeruginosa* lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology* 151, 1313–1323.

[7] Mitchell, E. et al. (2002) Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nat. Struct. Biol.* 9, 918–921.

[8] Perret, S. et al. (2005) Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of *Pseudomonas aeruginosa*. *Biochem. J.* 389, 325–332.

[9] Mitchell, E.P. et al. (2005) High affinity fucose binding of *Pseudomonas aeruginosa* lectin PA-IIL: 1.0 Å resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches. *Proteins: Struct. Funct. Bioinfo.* 58, 735–748.

[10] Garber, N., Guempel, U., Gilboa-Garber, N. and Doyle, R.J. (1987) Specificity of the fucose-binding lectin of *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* 48, 331–334.

[11] Loris, R., Tielker, D., Jaeger, K.-E. and Wyns, L. (2003) Structural basis of carbohydrate recognition by the lectin LecB from *Pseudomonas aeruginosa*. *J. Mol. Biol.* 331, 861–870.

[12] Binch, H., Stangier, K. and Thiem, J. (1998) Chemical synthesis of GDP-L-galactose and analogues. *Carbohydr. Res.* 306, 409–419.

[13] Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* 179, 131–137.

[14] Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* 26.

[15] Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* 50, 760–763.

[16] Perrakis, A., Morris, R. and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6, 458–463.

[17] Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* 53, 240–255.

[18] Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* 47, 110–119.

[19] Marchessault, R.H. and Pérez, S. (1979) Conformations of the hydroxymethyl group in crystalline aldohexopyranoses. *Biopolymers* 18, 2369–2374.

[20] Nishida, Y., Ohru, H. and Meguro, H. (1984) Proton NMR studies of (6R)- and (6S)-deuterated D-hexoses: assignment of the preferred rotamers about C5–C6 bond of D-glucose and D-galactose derivatives in solutions. *Tetrahedron Lett.* 25, 1575–1578.

[21] Carver, J.P. (1993) Oligosaccharides: how flexible molecules can act as signals. *Pure Appl. Chem.* 65, 763–770.

[22] Dam, T.K. and Brewer, C.F. (2002) Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* 102, 387–429.

[23] Kraulis, P. (1991) Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* 24, 946–950.

[24] Merritt, E.A. and Murphy, M.E. (1994) Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* 50, 869–873.

[25] Waldherr-Teschner, M., Goetze, T., Heiden, W., Knoblauch, M., Vollhardt, H. and Brickmann, J. (1992) MOLCAD – computer aided visualization and manipulation of models in molecular science in: *Advances in Scientific Visualization* (Post, F.H. and Hin, A.J.S., Eds.), pp. 58–67, Springer, Heidelberg.

# 4

## Mutations et spécificité de la lectine PA-IIL

---

La lectine PA-IIL a un site de reconnaissance peu discriminatoire puisqu'elle reconnaît le L-fucose mais aussi le D-mannose, bien qu'avec une affinité plus faible. Il serait intéressant de pouvoir ingénierer cette lectine afin de lui apporter une spécificité plus stricte. En effet, les lectines sont couramment utilisées en biochimie ou en biotechnologie, et de nouvelles lectines à haute affinité seraient des outils appréciés par les glyco biologistes.

### 4.1. Recherche de séquences similaires chez d'autres bactéries

Des séquences homologues à PA-IIL ont été identifiées dans le génome de deux autres bactéries opportunistes *Chromobacterium violaceum* et de *Ralstonia solanacearum* (Zinger-Yosovich *et al.*, 2006) (fig. 4.1). Les trois séquences ont 60% d'acides aminés en commun. La structure de RS-IIL (*Ralstonia solanacearum*) a été résolue en 2004 (Sudakevitz *et al.*, 2004). La structure tertiaire de la protéine est très similaire à celle de PA-IIL mais ne montre pas la même spécificité envers les monosaccharides. En effet, alors que PA-IIL se lie préférentiellement au L-fucose, RS-IIL montre une affinité élevée pour le D-mannose (Pokorna *et al.*, 2006). La lectine CV-IIL quant à elle, a un comportement intermédiaire et reconnaît le fucose et le mannose de façon similaire (Pokorna, Cioci *et al.*, 2006).

```

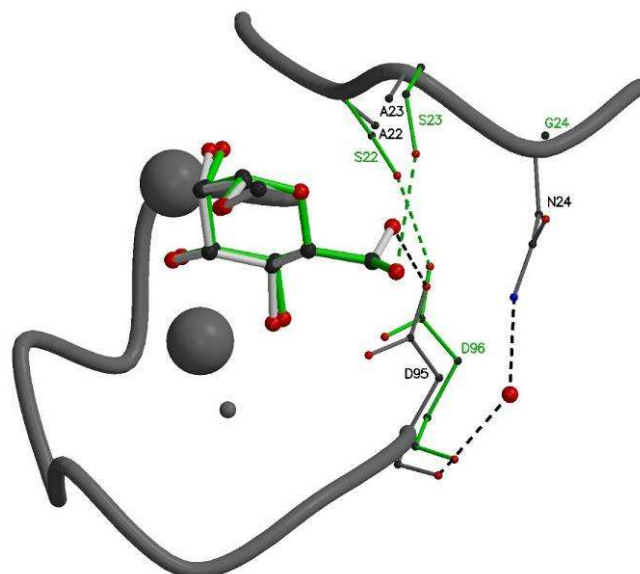
*
PA-IIL: ATQGVFTLPANTRFGVTAFAN-SSG*QTVNVLVNN--ETAATFSGQSTNNNAVIGTQVLNSGSS : 60
RS-IIL: AQQGVFTLPANTSFGVTAFAN-AANTQTIQVLVDN--VVKATFTGSGTSDKLLGSQVLNSGS- : 59
Ch.v : AQQGVFTLPARINFGVTVLVN-SAA*QHVEIFVDN--EPRAAFSGVGTGDNNLGTKVINSGS- : 59

* * * * *
PA-IIL: GKVQVQVSVNGRPSDLVSAQVILTNE-----LNFALVGSSEGDGTDNDYNDVAVVINWPLG : 115
RS-IIL: GAIKIQVSVNGKPSDLVSNQITILANK-----LNFAMVGSSEGDGTDNDYNDGIAVLNWPLG : 114
Ch.v : GNVRVQITANGRQSDLVSSQLVLANK-----LNLAVVGSSEGDGTDMDYNDISIVILNWPLG : 114

```

**Figure 4.1.** Comparaison des séquences des lectines PA-IIL, RS-IIL et CV-IIL. Les acides aminés différents dans les trois structures sont surlignés en gris, les acides aminés impliqués dans la coordination des ions calcium sont marqués par une étoile. Les acides aminés en position 22, 23 et 24 sont encadrés en rouge (« boucle de spécificité »)

L'étude structurale comparative du site de liaison révèle que la différence de spécificité entre PA-IIL et RS-IIL est principalement due à la nature des acides aminés constituant la « boucle de spécificité » (fig. 4.2). La séquence Ser22-Ser23-Gly24 chez PA-IIL est remplacée par la séquence Ala22-Ala23-Asn24 chez RS-IIL. La comparaison des structures cristallines du complexe avec le mannose de PA-IIL (Loris *et al.*, 2003) et de RS-IIL (Sudakevitz *et al.*, 2004) permet de comprendre le rôle d'un acide aminé particulier dans la différence d'affinité. Le groupement hydroxyle de la Ser22 chez PA-IIL établit une liaison hydrogène avec l'Asp96. L'oxygène O6 du D-mannose n'a pas la place nécessaire pour se positionner au fond du site et pointe par conséquent vers l'extérieur du site. Dans le cas de la lectine RS-IIL, l'Asp96 est disponible pour effectuer une liaison d'hydrogène avec l'oxygène O6 du D-mannose. Cette interaction supplémentaire a pour conséquence d'augmenter l'affinité pour le D-mannose.



**Figure 4.2.** Superposition des sites de liaison de RS-IIL complexée avec du Me- $\alpha$ -D-mannoside (en gris et blanc) et de PA-IIL complexé avec du  $\alpha$ -D-mannose (en vert) (Sudakevitz et al., 2004)

Afin de confirmer le rôle de cette boucle particulière dans la spécificité mannose/fucose, nous avons entrepris de produire des mutants, des protéines "hybrides" ou chacun des trois acides aminés de PA-IIL est remplacé par son équivalent dans RS-IIL.

## 4.2. Productions de mutants de PA-IIL et données thermodynamiques

*Cette partie du travail a été effectuée par Martina Pokorna (Brno) dans le cadre de sa thèse soutenue le 12 octobre 2006. Une publication commune est en cours de rédaction.*

Les trois lectines PA-IIL mutantes S22A, S23A et G24N ont été réalisées par le Dr Martina Pokorna du NCBR à l'université Masaryk de Brno, en République Tchèque selon le procédé suivant : la construction pET25pa2l, constitué du plasmide pET-25(b+) (Novagen) et du gène codant l'intégralité de la lectine PA-IIL issue de la souche sauvage de *P. aeruginosa*. La mutagenèse dirigée a été réalisée avec le kit QuikChange<sup>TM</sup> Site-Directed Mutagenesis Kit (Stratagene) d'après les instructions du fournisseur. Les oligonucléotides utilisés pour les mutations ont été conçus comme suit: Pour le mutant PA-IIL/S22A, la paire d'amorces PA-IIL/S22Asens/antisens a été (5` - CCG CCT TCG CCA ACG CGT CCG GAA CCC AGA CGG - 3`/5` - CCG TCT GGG TTC CGG ACG CGT TGG CGA AGG CGG - 3`). Pour le mutant PA-IILS23A, la paire d'amorces PA-IILS23Asens/antisens a été (5` - CCG CCT TCG CCA ACT CGG CCG GAA CCC AGA CGG - 3`/5` - CCG TCT GGG TTC CGG CCG AGT TGG CGA AGG CGG - 3`). Pour le mutant PA-IILG24N, la paire d'amorces PA-IILG24Nsens/antisens a été (5` - CCG CCT TCG CCA ACT CGT CCA ACA CCC AGA CGG TGA ACG - 3`/5` - CGT TCA CCG TCT GGG TGT TGG ACG AGT TGG CGA AGG - 3`). Le succès des mutations a été confirmé par séquençage de l'ADN, et les nouveaux plasmides comportants les mutations désirées ont été transformés dans les souches de *E. coli* : Tuner(DE3) et BL21(DE3) (Novagen). L'expression des mutants a ensuite été réalisée dans les conditions décrites pour la protéine PA-IIL native (Mitchell et al., 2005). Les protéines ont été purifiées par chromatographie d'affinité sur colonne mannose-agarose, puis lyophilisées et stockées à -20C.

Les affinités des lectines mutantes ont été évaluées par microcalorimétrie ITC pour les ligands Me- $\alpha$ -L-fucoside et Me- $\alpha$ -D-mannoside et comparées aux données précédemment obtenues pour PA-IIL sauvage, RS-IIL et CV-IIL (tab. 4.1).

	Ka (10 <sup>4</sup> M <sup>-1</sup> )	Kd ( $\mu$ M)	$\Delta$ G (kJ/mol)	$\Delta$ H (kJ/mol)	T $\Delta$ S (kJ/mol)
<b>PA-IIL G24N</b>					
Me- $\alpha$ -L-Fuc	524.3 ( $\pm$ 36.73)	0.2 ( $\pm$ 0.01)	-38.4 ( $\pm$ 0.18)	-40.1 ( $\pm$ 0.74)	-1.8 ( $\pm$ 0.68)
Me- $\alpha$ -D-Man	2.3 ( $\pm$ 0.17)	42.9 ( $\pm$ 3.12)	-24.9 ( $\pm$ 0.18)	-27.1 ( $\pm$ 0.77)	-2.1 ( $\pm$ 0.88)
<b>PA-IIL S23A</b>					
Me- $\alpha$ -L-Fuc	418.0 ( $\pm$ 72.83)	0.25 ( $\pm$ 0.24)	-37.8 ( $\pm$ 0.41)	-33.9 ( $\pm$ 1.24)	3.8 ( $\pm$ 0.87)
Me- $\alpha$ -D-Man	1.9 ( $\pm$ 0.33)	53.9 ( $\pm$ 9.32)	-24.4 ( $\pm$ 0.43)	-16.7 ( $\pm$ 2.27)	7.7 ( $\pm$ 2.7)
<b>PA-IIL S22A</b>					
Me- $\alpha$ -L-Fuc	26.4 ( $\pm$ 4.9)	3.96 ( $\pm$ 0.91)	-30.9 ( $\pm$ 0.52)	-37.1 ( $\pm$ 0.58)	-6.2 ( $\pm$ 0.95)
Me- $\alpha$ -D-Man	36.0 ( $\pm$ 1.8)	2.78 ( $\pm$ 0.14)	-31.7 ( $\pm$ 0.12)	-24.3 ( $\pm$ 0.60)	7.4 ( $\pm$ 0.70)
<b>PA-IIL native</b>					
Me- $\alpha$ -L-Fuc <sup>b</sup>	235.0 ( $\pm$ 8)	0.4 ( $\pm$ 0.01)	-36.4 ( $\pm$ 0.10)	-41.3 ( $\pm$ 1.00)	-4.9 ( $\pm$ 1.00)
Me- $\alpha$ -D-Man <sup>b</sup>	1.4 ( $\pm$ 0.05)	71.0 ( $\pm$ 3.00)	-23.7 ( $\pm$ 0.10)	-17.8 ( $\pm$ 0.40)	5.9 ( $\pm$ 0.30)
<b>CV-IIL</b>					
Me- $\alpha$ -L-Fuc <sup>a</sup>	59.4 ( $\pm$ 0.2)	1.7 ( $\pm$ 0.01)	-32.9 ( $\pm$ 0.01)	-21.8 ( $\pm$ 0.2)	11.1 ( $\pm$ 0.10)
Me- $\alpha$ -D-Man <sup>a</sup>	5.4 ( $\pm$ 0.6)	19.0 ( $\pm$ 2.00)	-27.0 ( $\pm$ 0.30)	-17.8 ( $\pm$ 1.3)	9.2 ( $\pm$ 1.60)
<b>RS-IIL</b>					
Me- $\alpha$ -L-Fuc <sup>a</sup>	n d*				
Me- $\alpha$ -D-Man <sup>a</sup>	438.0 ( $\pm$ 2)	0.2 ( $\pm$ 0.01)	37.9 ( $\pm$ 0.10)	24.9 ( $\pm$ 1.30)	13.0 ( $\pm$ 0.23)

<sup>a</sup> (Pokorna, Cioci *et al.*, 2006), <sup>b</sup> (Sabin *et al.*, 2006), \* non détectable, valeur < 10<sup>4</sup> M<sup>-1</sup>

**Tableau 4.1.** Données thermodynamiques obtenues par ITC à 25°C, des interactions du Me- $\alpha$ -L-Fuc et du Me- $\alpha$ -D-Man avec les lectines CV-IIL, PA-IIL et RS-IIL (les déviations standards ont été calculées à partir de trois expériences indépendantes)

Les mutants PA-IIL S23A et G24N ne présentent que peu de différence par rapport à la lectine native. Il est cependant intéressant de noter que ces deux mutations augmentent légèrement l'affinité (facteur 2), à la fois pour le mannose et pour le fucose. Par contre, la mutation S22A a un effet très net puisqu'elle multiplie par 10 l'affinité pour le mannose et diminue par 10 l'affinité pour le fucose. La lectine PA-IIL S22A a donc la même affinité pour les deux sucres. Une légère préférence pour le mannose est observée, ce qui permet de dire que la spécificité de la lectine a été fortement modifiée par cette expérience d'ingénierie.

## 4.3. Structures cristallines

### 4.3.1. Préparations des cristaux

Les mutants de la lectine PA-IIL ont d'abord été cristallisés en présence de fucose, car ce complexe cristallin est le plus facile à obtenir. Les meilleurs cristaux obtenus dans les conditions de cristallisations définies par les « screens Hampton I et II » ont ensuite été trempés 24 heures dans des solutions des Me- $\alpha$ -D-Man ou de Me- $\alpha$ -L-Fuc à 10 mM

Les cristaux complexés avec le fucose et utilisés pour le trempage ont été obtenus dans les conditions suivantes:

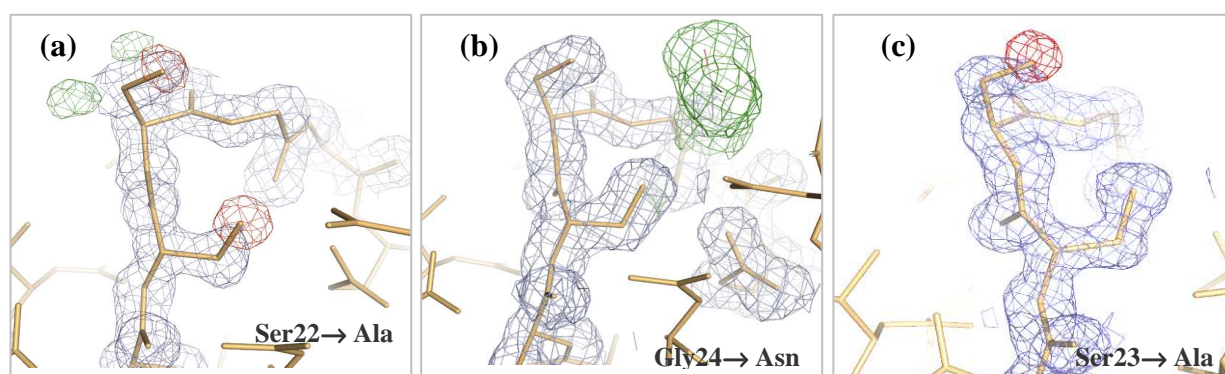
- G24N/Me- $\alpha$ -L-Fuc: 0,2M de sulfate d'ammonium et 30% de PEG 4000
- G24N/Me- $\alpha$ -D-Man: 0,1M d'HEPES (pH 7,5), 10 % d'iso-propanol et de 20% PEG 4000
- S22A/ Me- $\alpha$ -D-Man: 0,2M sulfate de lithium monohydrate, 0,1M de Tris hydrochloride (pH 8,5) et 30% de PEG 4000
- S22A/ Me- $\alpha$ -L-Fuc: 0,2M de sulfate de lithium monohydrate, 0,1M de Tris hydrochloride (pH 8,5) et 30% de PEG 4000
- S23A/ Me- $\alpha$ -L-Fuc: 0,2M de chlorure de calcium dihydraté, 0,1M d'acetate de sodium trihydraté (pH 7.5) et de 20% d'iso-propanol

### 4.3.2. Résultats

La présence des mutations a été vérifiée par contraste de phases lors du remplacement moléculaire de la structure native dans les cartes de densité électronique  $fo$ - $fc$  et  $2fo$ - $fc$  (fig. 4.3). De même, les configurations alternatives de certains acides aminés ont aussi été révélées au cours des affinements successifs en observant les cartes de densité. L'affinement a montré que selon l'empilement cristallin, les monosaccharides d'intérêt n'ont pas toujours pu se substituer au fucose initial dans tous les sites du tétramère. Ainsi, pour la structure PA-IIL S22A/Me- $\alpha$ -D-Man, trois monomères sur quatre contiennent le monosaccharide d'intérêt, et uniquement deux monomères sur quatre ont été substitués pour les structures des complexes PA-IIL G24N/Me- $\alpha$ -L-Fuc et PA-



IIL S22A/Me- $\alpha$ -Fuc. Les données statistiques de l'affinement sont répertoriées dans le tableau 2. La présence de Me- $\alpha$ -D-Man n'a jamais été observée dans les cristaux de PA-IIL S23A/fucose trempés dans le Me- $\alpha$ -D-Man. Les sous-unités du tétramère comportant le sucre d'intérêt ont été superposées afin d'identifier les molécules d'eau structurales participant de manière systématique à l'interaction du monosaccharide. L'analyse des angles, des distances et des facteurs thermiques B des molécules d'eau a aussi été prise en compte dans cette étude. Les données statistiques de la collecte et de l'affinement des données sont détaillées dans le tableau 4.2, les distances interatomiques dans le site de reconnaissance figurent dans le tableau 4.3.



**Figure 4.3.** Cartes de densité électronique  $2fo-fc$  (contour à 1.0 sigma, en bleu) et  $fo-fc$  (contour à 3.0 sigma en vert et -3.0 sigma en rouge), autour de la boucle contenant les acides aminés 22-23-24 lors du remplacement moléculaire avec la structure native. (a) Complexe avec le Me- $\alpha$ -D-Man : mutation de Ser22 en Ala, (et présence de conformations alternatives de Ser23). (b) Complexe avec le Me- $\alpha$ -D-Man : mutation de Gly24 en Asn. (c) Complexe avec le Me- $\alpha$ -L-Fuc : mutation de Ser23 en Ala.

	PA-III G24N		PA-III S22A		PA-III S23A
	<i>Me-<math>\alpha</math>-D-Man</i>	<i>Me-<math>\alpha</math>-L-Fuc</i>	<i>Me-<math>\alpha</math>-D-Man</i>	<i>Me-<math>\alpha</math>-L-Fuc</i>	<i>Me-<math>\alpha</math>-L-Fuc</i>
<b>Collecte de données</b>					
Ligne de lumière (ESRF)	ID29	ID14-EH2	ID29	ID14-EH2	ID14-EH2
Longueur d'onde (Å)	0.987	0.933	0.976	0.933	0.933
Résolution (Å)	1.70-49.45	1.50-28.69	1.30-21.20	1.70-34.69	1.30-87.71
Zone haute résolution (Å)	1.70-1.74	1.50-1.54	1.30-1.33	1.70-1.74	1.30-1.33
Groupe d'espace	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	C2
Maille	a = 50.4 Å; b = 80.1 Å; c = 52.5 Å; $\beta$ = 109.6°	a = 52.6 Å; b = 74.9 Å; c = 53.5 Å $\beta$ = 114.7°	a = 52.5 Å; b = 72.9 Å; c = 54.6 Å; $\beta$ = 94.3°	a = 52.7 Å; b = 73.1 Å; c = 54.9 Å $\beta$ = 94.3°	a = 95.1 Å; b = 45.7 Å; c = 87.9 Å; $\beta$ = 94.2°
Coef de Matthews (Å <sup>3</sup> /Da)	2.13	2.04	2.22	2.25	2.03
Solvant (%)	42.2	39.7	44.7	45.3	39.4
Reflec. mesurées/unique	146877/42206	224211/60074	421970/99982	160284/43407	292549/98919
Multiplicité moyenne <sup>1</sup>	3.5 (3.0)	3.7 (3.7)	4.2 (3.8)	3.7 (3.6)	3.3 (3.2)
Complétude <sup>1</sup> (%)	97.8 (88.7)	99.7 (99.7)	99.2 (96.4)	95.4 (92.4)	95.8 (93.0)
Moyenne <sup>1</sup> I/ $\sigma$ I	20.1 (6.6)	14.8 (3.3)	23.1 (12.4)	14.4 (3.5)	16.2 (2.6)
R merge <sup>1,2</sup> (%)	0.054 (0.20)	0.060 (0.36)	0.042 (0.079)	0.085 (0.32)	0.054 (0.35)
Facteur B de Wilson (Å <sup>2</sup> )	9.68	12.31	10.28	11.29	12.86
<b>Affinement</b>					
R <sub>cryst</sub> <sup>3</sup> (observation)	0.143 (40067)	0.151 (57022)	0.139 (94963)	0.177 (41187)	0.170 (94498)
R <sub>libre</sub> <sup>4</sup> (observation)	0.172 (2117)	0.193 (3024)	0.155 (4993)	0.223 (2200)	0.197 (4421)
Nombre d'atomes:					
Hetero/ Proteine/ Solvant	60/ 3414/ 446	66/ 3525/ 430	60/ 3414/ 446	66/ 3360/ 524	56/ 4232/ 592
Facteur B moyen					
Chaines : A/B/C/D/Eau	7.8/ 9.2/ 81 /7.4/ 22.0	10.2/ 9.7 / 10.2/ 10.2/ 26.4	6.9/ 7.4/ 7.3/ 7.2/ 23.8	9.0/ 9.2/ 9.3/ 10.8/ 22.3	10.8/ 10.9/ 9.8/ 11.3/ 26.2
Deviation RMS/ideal					
Liaisons (Å)	0.013	0.016	0.011	0.011	0.013
Angles (°)	1.26	1.58	1.42	1.29	1.47
Code PDB deposition	non déposé	non déposé	non déposé	non déposé	non déposé

<sup>1</sup>Nombre entre parenthèses font référence à la zone de plus haute résolution. <sup>2</sup> $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle$ , où I = intensité observée. <sup>3</sup> $R_{\text{cryst}} = \sum \|F_o - F_c\| / \sum |F_o|$ , où  $|F_o|$  = amplitude des facteurs de structure observés et  $|F_c|$  = amplitude des facteurs de structure calculés. <sup>4</sup>R<sub>free</sub> correspond au R<sub>cryst</sub> pour 3 % des réflexions exclues de l'affinement.

**Tableau 4.2.** Données statistiques de la collecte et de l'affinement des structures cristallines

	PA-III G24N		PA-III S22A		PA-III S23A
	Me- $\alpha$ -D-Man	Me- $\alpha$ -L-Fuc	Me- $\alpha$ -D-Man	Me- $\alpha$ -L-Fuc	Me- $\alpha$ -L-Fuc
Man (O6) -- Ser23 (OG)	2,80	-	3,50	-	-
Man (O6) -- Asn24 (ND2)	3,31	--	-	-	-
Man (O6) -- Asp96 (OD2)	-	-	2,67	-	-
Man (O6) -- Ser23 (N)	-	-	3,20	-	-
Man (O6) -- Gly24 (N)	-	-	2,94	-	-
O5--Ser22 (OG)	-	3,44	-	-	3,44
O5--Ser23 (OG)	3,20	3,19	3,19	3,35	-
Asp96 (OD2) -- Ser22 (OG)	2,71	2,78	-	-	2,70
Asp96 (OD2) -- H <sub>2</sub> O (2)	-	-	-	2,60	-
H <sub>2</sub> O (2) -- Ser23 (OG)	-	-	-	-	-
Fuc (O1) -- H <sub>2</sub> O (1)	-	3,08	-	3,19	3,03

**Tableau 4.3.** Distances interatomiques des ligands dans le site de reconnaissance de PA-III

## 4.4. Discussion et conclusion

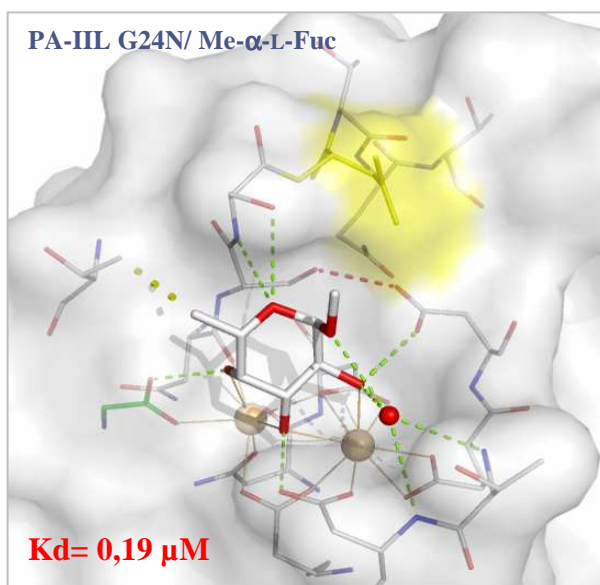
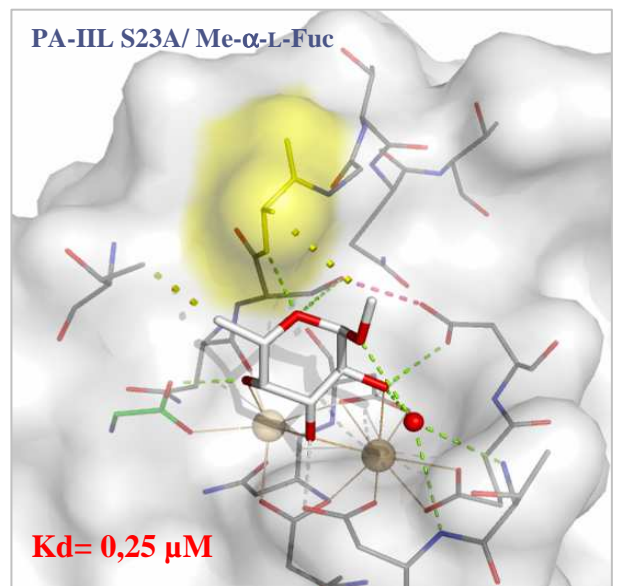
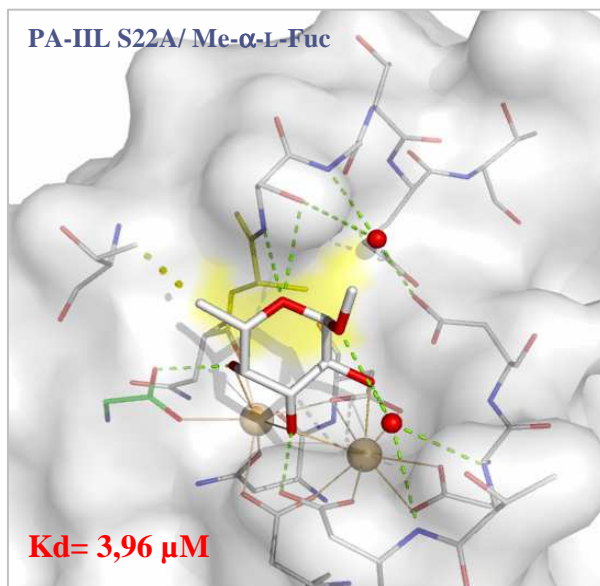
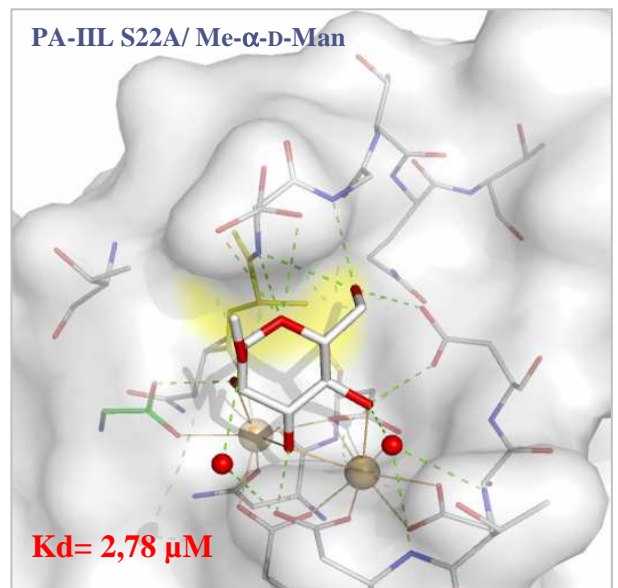
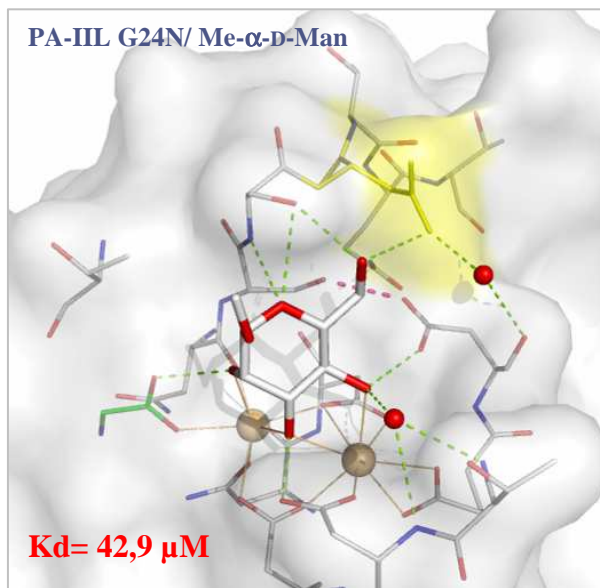
### 4.4.1. Interaction avec le Me- $\alpha$ -L-Fucoside

Lorsque l'on compare l'affinité du Me- $\alpha$ -L-Fucopyranoside pour la lectine PA-III en fonction des différentes mutations (fig. 4.4), on s'aperçoit que l'affinité du monosaccharide est divisée par dix en présence de la mutation S22A. Pour expliquer cette forte baisse d'affinité, il faut regarder la structure native. Le résidu Ser22 fait, via l'oxygène de la chaîne latérale, une liaison hydrogène de 2,7 Å avec l'oxygène de la chaîne latérale de l'acide aminé Asp96 et comble le fond de la poche constituant le site de reconnaissance. La mutation de la Ser22 par une Alanine libère l'acide aminé Asp96 qui peut alors fixer une molécule d'eau, ce qui concorde avec la pénalité entropique observée. Par contre, il est difficile d'expliquer pourquoi l'enthalpie d'interaction de S22A pour le Me- $\alpha$ -L-Fuc est moins bonne que celle de la PA-III native.

A l'inverse, l'affinité du Me- $\alpha$ -L-fucopyranoside pour la lectine ayant subi la mutation S23A est multipliée par deux tout en restant dans un même ordre de grandeur. On note cette fois une contribution entropique favorable consécutive de 3.8 kJ/mol. Comparée avec la contribution défavorable de 4,9 kJ/mol observée dans le cas de la structure native, on observe un gain relatif de 8,7 kJ/mol, ceci au détriment de la contribution enthalpique. En effet, bien que le

résidu Ser23 fasse une liaison hydrogène avec l'oxygène O5 du cycle pyranose des monosaccharides, il est relativement libre et donne fréquemment lieu à plusieurs conformations alternatives, davantage encore en présence de monosaccharides méthylés en O1 (Sabin *et al.*, 2006). La contrainte occasionnée habituellement lors de la complexation avec la forme native et qui engendrait de ce fait une pénalité entropique n'est plus retrouvée dans la forme mutée. De surcroît, le groupe CH<sub>3</sub> de l'alanine de remplacement établit une nouvelle interaction hydrophobe avec le méthyle en O1 du Me- $\alpha$ -L-Fuc.

Il est plus ardu d'expliquer la hausse d'affinité du monosaccharide pour la structure mutée G24N, le résidu ne fait en effet aucune nouvelle interaction directe ou indirecte avec le sucre. La contribution de l'enthalpie quoique diminuée est sensiblement la même, l'entropie quant à elle, est moins défavorable de 3 kJ/mol attestant peut-être d'une meilleure complémentarité de surface entre le sucre et la protéine qui, par la présence de l'Asn24, ferme davantage la poche du site de reconnaissance autour du sucre et ne piège plus d'eau.



**Figure 4.4.** Me- $\alpha$ -L-Fuc et Me- $\alpha$ -D-Man, complexés avec différents mutants de la lectine PA-IIL. Les acides aminés mutés sont représentés en jaune. Eau (sphère rouge), calcium (sphère beige), liaisons hydrogène (pointillés verts), liaisons hydrophobes (gros pointillés jaune). La liaison hydrogène liant la Ser22 et l'Asp96 est représentée par des pointillés violets.

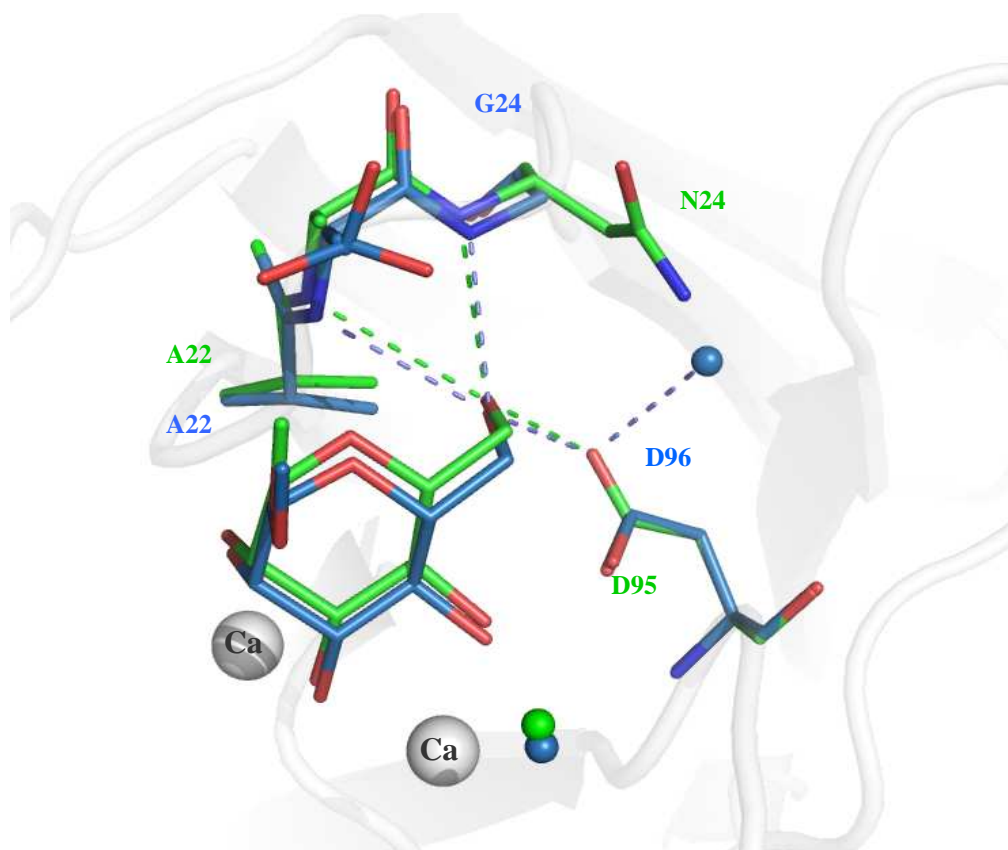
#### 4.4.2. Interaction avec le Me- $\alpha$ -D-mannopyranoside

L'affinité du Me- $\alpha$ -D-Man pour la lectine mutée G24N est sensiblement meilleure mais reste du même ordre que celle obtenue lors du complexe avec la forme native de la protéine. Cependant les contributions des grandeurs thermodynamiques se répartissent différemment. En effet, la contribution de l'enthalpie est significativement plus forte pour la lectine mutée G24N, puisque la valeur de -17.8 kJ pour la forme native est désormais de -27.1 kJ. Par ailleurs, la contribution entropique favorable, qui représentait un tiers de l'énergie libre  $\Delta G$  dans le cas de la lectine native, devient légèrement défavorable et ne représente plus que 8 % du total des contributions pour la lectine mutée G24N. L'analyse de la structure cristalline révèle que l'oxygène O6 du Me- $\beta$ -D-Man pointe vers l'extérieur du site de reconnaissance et participe à deux liaisons hydrogène mettant en jeu l'oxygène de la Ser23 ainsi que l'azote de l'Asn24. Par ailleurs, cet azote fait aussi un pont hydrogène avec l'oxygène de la chaîne principale du résidu Asp96 via la capture d'une molécule d'eau. L'établissement de nouvelles interactions et la capture d'une molécule d'eau structurale supplémentaire expliquent ces variations thermodynamiques

L'affinité ainsi que les grandeurs thermodynamiques obtenues pour la lectine mutée S23A sont sensiblement identiques à celles obtenues pour la forme native. Les données structurales sont absentes ici, mais par analogie avec la structure du Me- $\alpha$ -D-Man complexé avec PA-IIL/G24N, le même raisonnement que celui effectué précédemment pour le Me- $\alpha$ -L-Fuc explique la légère baisse enthalpique et le léger gain entropique.

Le plus grand changement réside dans les données obtenues pour le complexe du Me- $\beta$ -D-Man avec la lectine mutée S22A. La constante d'affinité est environ vingt fois plus forte que celle obtenue lors du complexe avec la lectine native. L'énergie libre  $\Delta G$  a considérablement augmenté due à la contribution favorable et simultanée de l'enthalpie passant -17,8 kJ à -24,3 kJ et de l'entropie passant de +5,9 kJ à +7,4 kJ. Ici encore, la disponibilité retrouvée de l'acide aminé Asp96 du fait de la mutation de la Serine 22 en alanine, a permis la formation d'une nouvelle interaction directe avec le sucre. L'oxygène O6 du Me- $\alpha$ -D-Man pointe cette fois vers l'intérieur du site de liaison et fait trois liaisons hydrogène supplémentaires dont deux avec les azotes des chaînes principales des acides aminés Ser23 et Glu24 et une avec l'oxygène de la chaîne latérale de l'acide aminé Asp96. Il est important de signaler que cette

mutation produit pour la première fois une inversion dans l'affinité de la lectine pour ses ligands. Ainsi, la lectine PA-IIL mutée S22A montre désormais une affinité plus forte pour le mannose que pour le fucose. De plus, le mécanisme d'interaction du Me- $\alpha$ -D-Man avec PA-IIL/S22A est homologue à celui rencontré lors de l'interaction avec la protéine RS-IIL (fig. 4.5).



**Figure 4.5.** Superposition du site de reconnaissance du mutant PA-IIL S22A (en bleu) et de RS-IIL (en vert) en complexe avec le Me- $\alpha$ -D-Man. Les liaisons hydrogène sont représentées par des traits pointillés. Les molécules d'eau sont représentées par des sphères bleues et vertes, les ions calcium sont représentés par les sphères blanches.

# 5

## Interaction entre PA-IIL et les oligosaccharides naturels

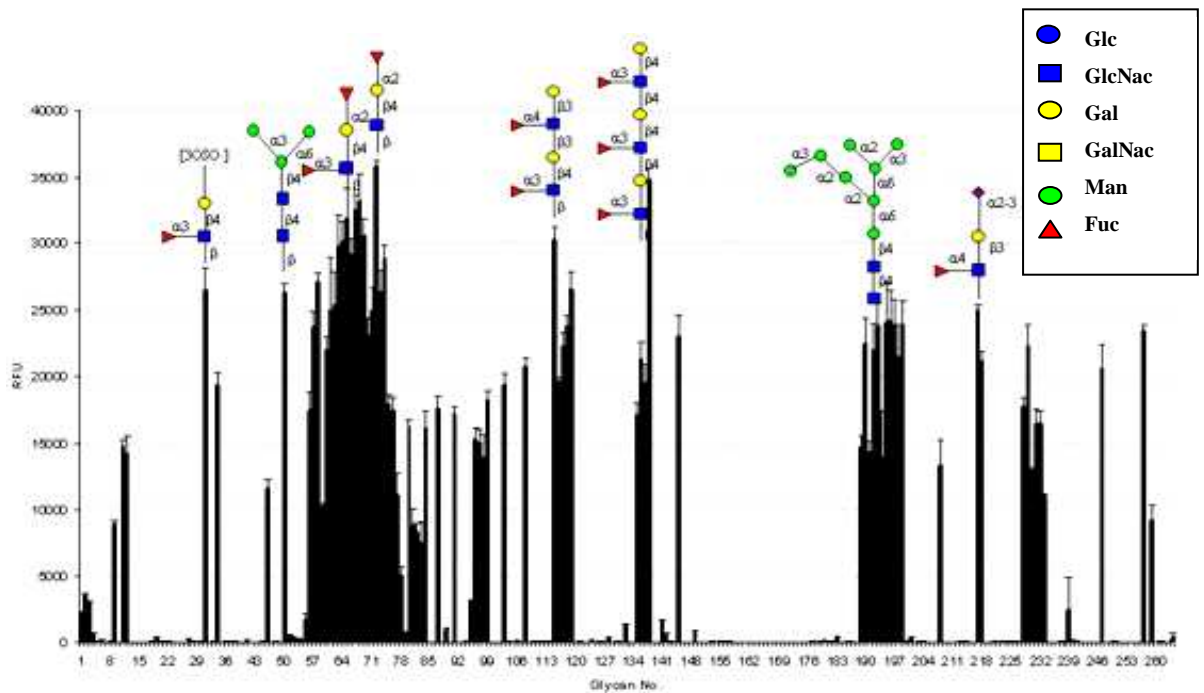
---

La lectine PA-IIL se lie à des oligosaccharides présents sur les voies respiratoires. L'identification des oligosaccharides récepteurs de la lectine est de première importance à la fois pour comprendre le mécanisme d'interaction et pour concevoir des glycomimétiques ayant un fort pouvoir compétiteur avec les ligands naturels.

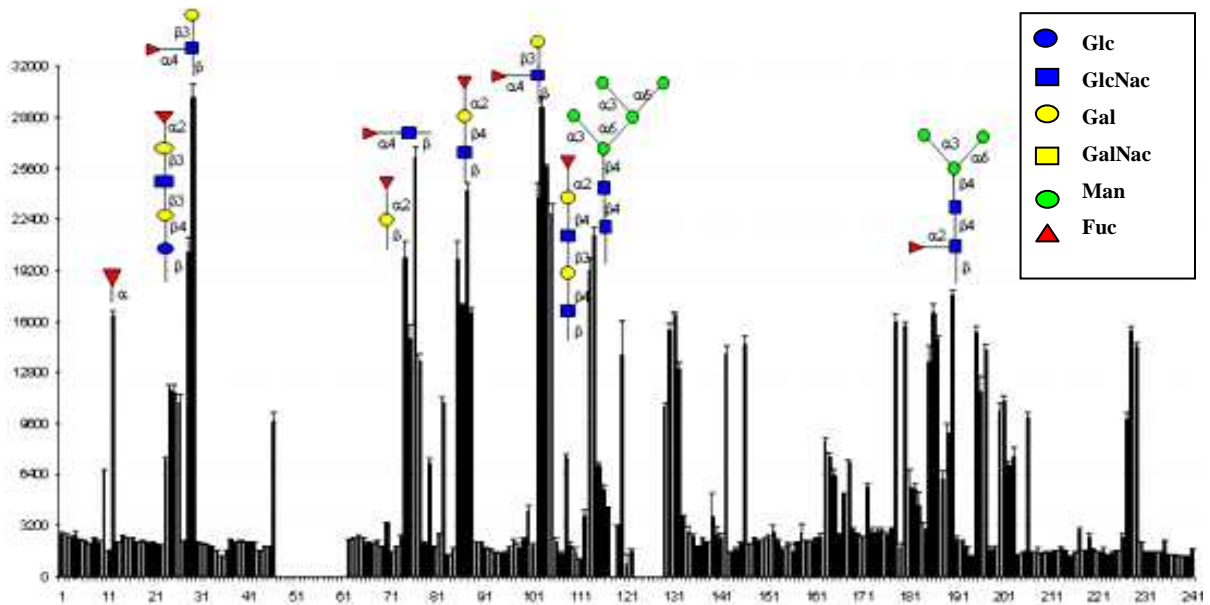
### 5.1. Recherche de ligands oligosaccharidiques par méthode de criblage.

Les méthodes d'analyse de spécificité disponibles dans le "Consortium for Functional Genomics" permettent de tester un très grand nombre d'oligosaccharides. La lectine PA-IIL a été marquée avec le fluorochrome Alexa Fluor 488 et envoyée à Rick Alvarez (U. of Oklahoma). Plus de 250 oligosaccharides naturels ont été testés par les méthodes de « printed array » (fig. 5.2) et « glycan array » (fig. 5.1) Pour des raisons techniques, les informations obtenues par la méthode « printed array » sont essentiellement qualitatives. En effet, la taille minuscule des puits utilisés et le procédé de lecture, discriminent les oligosaccharides fortement reconnus par la lectine (puits éclairés) de ceux qui le sont pas ou peu (puits éteints ou faiblement éclairés), sans fournir de véritables informations quantitatives. La méthode « glycan array », quant à elle, est d'avantage quantitative.





**Figure 5.1.** Interaction de la lectine PA-IIL (intensité de fluorescence) évaluée par la méthode de « printed array ». Seuls les pics intenses ont été étiquetés.



**Figure 5.2.** Interaction de la lectine PA-IIL (intensité de fluorescence) évaluée par la méthode de « glycan array ». Seuls les pics intenses ont été étiquetés.

Quelle que soit la méthode utilisée, on confirme que PA-IIL peut interagir avec des oligosaccharides mannosylés ou fucosylés. La méthode de "Printed array" indique que tous ceux comportant tout ou partie des motifs Lewis a : Fuc( $\alpha$ 1-4)GlcNac(3-1 $\beta$ )Gal ou Lewis x (Fuc( $\alpha$ 1-3)GlcNac(4-1 $\beta$ )Gal), ainsi que les oligosaccharides dans lesquels le L-fucose est lié en position  $\alpha$ 1-2 du  $\beta$ -Galactose sont bien reconnus. On retrouve aussi parmi les meilleurs ligands, des oligosaccharides à D-mannose tel que le motif Man( $\alpha$ 1-6)Man( $\alpha$ 1-6)[Man( $\alpha$ 1-3)] et Man( $\alpha$ 1-2)Man( $\alpha$ 1-6)[Man( $\alpha$ 1-3)].

La méthode "Glycan array", indique que les pics de fluorescence les plus intenses correspondent tous à des oligosaccharides contenant le fucose lié en position 4 sur un GlcNAc. Le Lewis a (Fuc( $\alpha$ 1-4)GlcNac(3-1 $\beta$ )Gal) serait donc le ligand préférentiel de la lectine PA-IIL, comme cela avait été proposé à partir d'études préliminaires sur un petit nombre de déterminants oligosaccharidiques (Mitchell *et al.*, 2002).

## 5.2. Oligosaccharides fucosylés du lait maternel.

Le lait maternel humain est un mélange complexe composé de plusieurs centaines d'oligosaccharides différents (Miller and McVeagh, 1999, Kunz, 2000 #177). Bien que certains composants soient communs à d'autres mammifères, le lait humain comporte un grand nombre d'oligosaccharides spécifiques en quantité significative puisqu'ils viennent en troisième position après le lactose et les lipides, et leur quantité globale est supérieure à celle des protéines (tab. 5.1). La nature des oligosaccharides présents dans le corps d'un individu, y compris ceux du lait humain est déterminé génétiquement. Il n'est donc pas étonnant de retrouver une similitude structurale entre les oligosaccharides du lait et ceux des groupes sanguins. Des expériences ont d'ailleurs montré une corrélation entre les oligosaccharides retrouvés dans le lait maternel humain et le groupe sanguin d'appartenance de la mère (Viverge *et al.*, 1990).

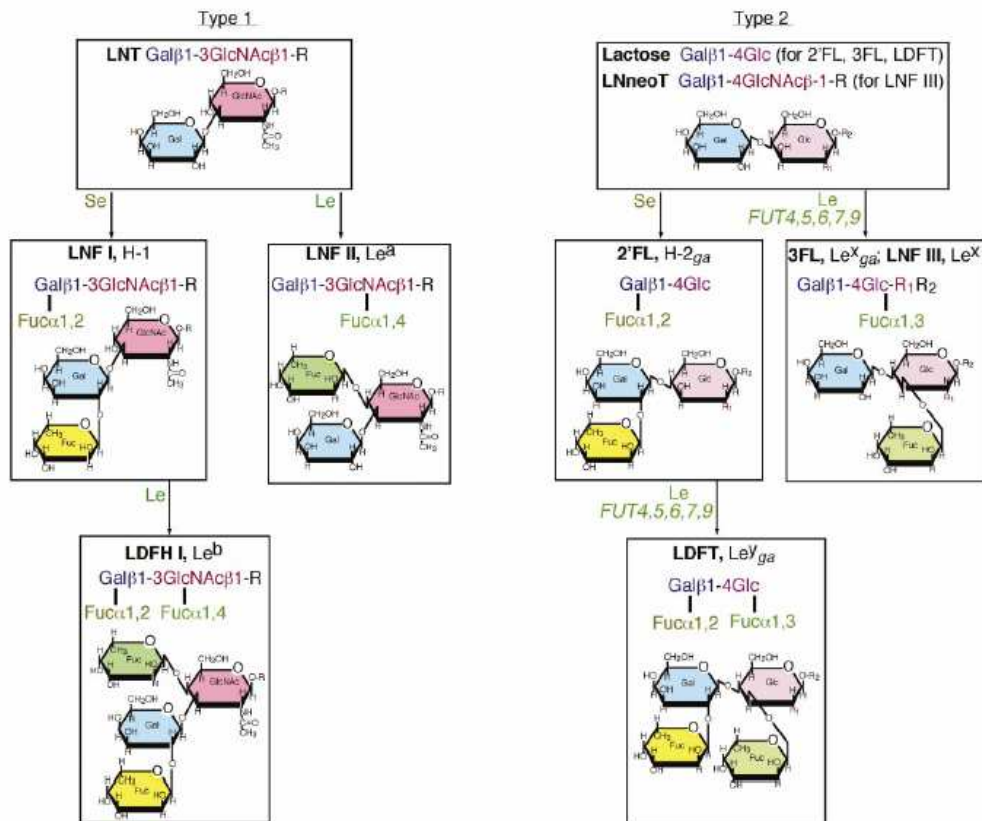
La fonction biologique des oligosaccharides est intimement liée à leurs conformations structurales. Les monosaccharides constituant les oligosaccharides du lait humain (HMOs) sont le D-glucose (Glc), le D-galactose (Gal), le N-acétylglucosamine (GlcNac), le L-fucose et l'acide sialique (NeuAc). Les oligosaccharides isolés jusqu'à présent ont une longueur variant de 3 à 10 monosaccharides et portent tous un lactose ( $\beta$ -D-Gal(1-4)D-Glc) à leur terminaison réductrice. Ces structures sont produites dans l'appareil de Golgi où le disaccharide de base, le lactose, est synthétisé à partir d'UDP-galactose et de glucose sous l'action d'une galactosyltransférase régulée par l' $\alpha$ -lactalbumine. Ce lactose sert ensuite de substrat à d'autres transférases (galactosyl-, fucosyl-, sialyl-, et N-acétylglucosaminyltransférases) qui ajoutent les monosaccharides complémentaires. Il est important de noter que près de la moitié des sucres du lait humain caractérisés sont sialylés.

Composant	Quantité (g/litre)	
	Lait humain	Lait de vache
Lactose	55-70	40-50
Oligosaccharides		
Lacto-N-tetraose	0.5-1.5	Traces
Lacto-N-fucopentaose I	1.2-1.7	—
Lacto-N-fucopentaose II	0.3-1.0	—
Lacto-N-fucopentaose III	0.01-0.2	—
Lacto-N-difucohexaose I	0.1-0.2	—
NeuAc( $\alpha$ 2-6)lactose	0.3-0.5	0.03-0.06
NeuAc( $\alpha$ 2-3)lactose	0.1-0.3	
NeuAc-lacto-N-tetraose a	0.03-0.2	Traces
NeuAc-lacto-N-tetraose c	0.1-0.6	Traces
NeuAc <sub>2</sub> -lacto-N-tetraose	0.2-0.6	Traces
Oligosaccharides (total)	5.0-8.0	Traces

**Tableau 5.1.** Comparaison des oligosaccharides présents dans le lait maternel humain et le lait de vache (Kunz et al., 2000).

La présence du fucose dans le lait est une spécificité du lait maternel humain, il n'est pas présent dans le lait de vache. L'addition du fucose dépend de l'action de trois fucosyltransférases différentes, déterminées génétiquement. L' $\alpha$ -1-2-fucosyltransférase est exprimée chez 77% des caucasiens que l'on classifie comme « sécréteur » (Erney et al., 2000). Le lait humain, chez les individus sécréteurs, se caractérise par la présence du 2-

fucosylactose ( $\text{Fuc}\alpha 1-2\text{Gal}\beta 1-4\text{Glc}$ ), du lacto-N-fucopentaose I ( $\text{Fuc}\alpha 1-2\text{Gal}\beta 1-3\text{GlcNAc}\beta 1-4\text{Glc}$ ) et d'autres oligosaccharides plus complexes comportant tous le motif  $\text{Fuc}\alpha 1-2\text{Gal}\beta 1-3\text{GlcNAc}\beta 1-4\text{Glc}$ . Une autre fucosyltransférase dépendant du gène Lewis attache un résidu fucose en position  $\alpha 1-4$  du résidu subterminal  $\text{GlcNAc}$  des chaînes de type 1. Dans le lait de ces individus (non sécréteurs), l'oligosaccharide fucosylé majoritaire est le lacto-N-fucopentaose II ( $\text{Gal}\beta 1-3[\text{Fuc}\alpha 1-4]\text{GlcNAc}\beta 1-3\text{Gal}\beta 1-4\text{Glc}$ ) correspondant au déterminant du group sanguin Lewis a. On retrouve ce trisaccharide chez 20 % de la population. Si les deux fucosyltransférases sont exprimées par l'individu, l'oligosaccharide du lait humain majoritaire est le lacto-N-difucohexaose ( $\text{Fuc}\alpha 1-2\text{Gal}\beta 1-3[\text{fuc}\alpha 1-4]\text{GlcNAc}\beta 1-3\text{Gal}\beta 1-4\text{Glc}$ ) correspondant au déterminant du groupe sanguin Lewis b (Kunz, Rudloff *et al.*, 2000). Pour les 5% de la population appartenant au group sanguin Lewis a<sup>-</sup> b<sup>-</sup>, on trouve un troisième oligosaccharide portant un fucose lié en  $\alpha 1-3$  du  $\text{GlcNAc}$  dans les chaînes de type 2. L'oligosaccharide qui en résulte est le ( $\text{Gal}\beta 1-4[\text{fuc}\alpha 1-3]\text{GlcNAc}\beta 1-3\text{Gal}\beta 1-4\text{Glc}$ ) correspondant au déterminant du groupe sanguin Lewis x.



**Figure 5.3.** Représentation schématique des oligosaccharides les plus fréquents dans le lait maternel humain (Newburg *et al.*, 2004)

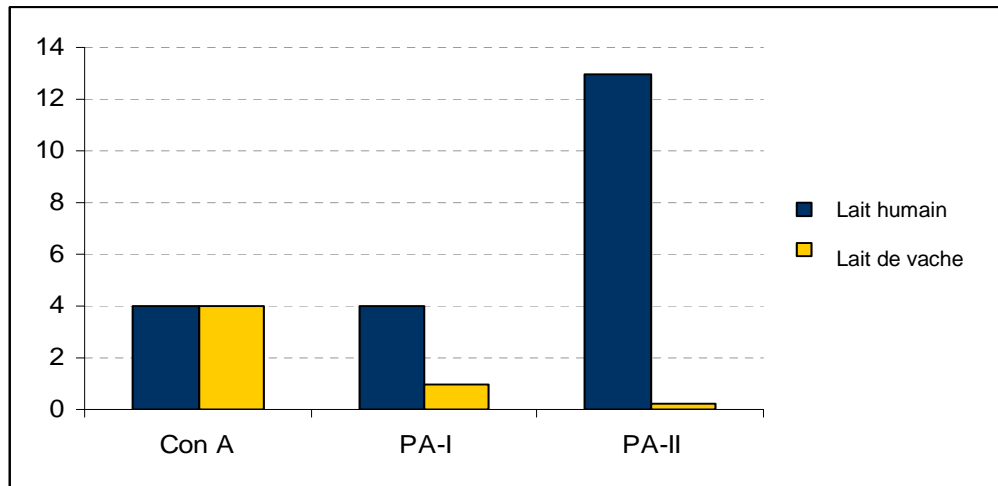
### 5.3. Les applications thérapeutiques des oligosaccharides du lait humain

Pour des raisons évidentes, les études concernant les HMOs ont principalement été menées sur le système digestif des nourrissons. Il a été montré qu'un certain nombre d'oligosaccharides n'étaient pas digérés par l'intestin grêle des nourrissons puisqu'on les retrouve intacts dans le gros intestin et les urines. Ceci suggère d'une part, que ces oligosaccharides auraient un effet prébiotique (Brand-Miller *et al.*, 1998) et d'autre part, qu'ils conservent leur configuration structurale et donc toutes leurs fonctions puisqu'ils ne sont pas dégradés dans l'estomac. En effet, les oligosaccharides de lait humain sont les analogues solubles des glycoconjugués exposés à la surface des cellules épithéliales et peuvent rentrer en compétition avec eux dans l'interaction avec les pathogènes comme les bactéries ou les virus. Chaque fois qu'un nourrisson tète, 100 à 500 mg d'oligosaccharides traversent les voies intestinales (Kunz, Rudloff *et al.*, 2000). 70% des otites survenant chez les nourrissons sont la cause d'infection par *Pneumococci* ou *Haemophilus influenzae*. Il a été démontré que l'adhésion de ces microorganismes à des sucres spécifiques des cellules épithéliales buccales étaient inhibés par le lait humain (Andersson *et al.*, 1986). D'autres études démontrent que les oligosaccharides neutres du lait humain protègent les intestins des nouveaux nés du cholera (Holmgren *et al.*, 1983) ou que les oligosaccharides du lait humain sialylés inhibent l'adhésion des souches *E.coli* responsables de méningite chez le nourrisson. Les oligosaccharides du lait humain pourraient être utilisés comme traitement ou en prévention des infections. Ils constituent des agents thérapeutiques particulièrement intéressants puisqu'ils ne sont ni allergisants ni toxiques pour l'organisme.

### 5.4. Interaction entre PA-IIL et les oligosaccharides du lait humain (Article n° 3)

De nombreux sucres du lait humain sont fucosylés et donc potentiellement reconnaissables par la lectine PA-IIL. Des expériences ont montré que l'hémagglutination par la lectine était

inhibée en présence de lait humain et non en présence de celui de vache (Lesman-Movshovich and Gilboa-Garber, 2003) (fig. 5.4).



**Figure 5.4.** Inhibition d'hémagglutination par du lait humain et bovin des lectines PA-II, PA-I et concanavaline A (ConA). Adapté de (Lesman-Movshovich and Gilboa-Garber, 2003)

Les oligosaccharides complexes comme ceux du lait humain sont cependant difficiles à isoler en grande quantité ou à reproduire chimiquement. Une méthode développée au CERMAV procédant par voie enzymatique dans une souche bactérienne de *E.coli* a permis la fabrication et la purification des pentasaccharides LNnFP-V (lacto-*N*-neofucopentaose V) et LNnDF-II (lacto-*N*-neodifucohexaose II) (Dumon *et al.*, 2001). Les tests ELLA ont pu être menés sur un ensemble d'oligosaccharides fucosylés et les meilleurs ligands ont été retenus pour les études par microcalorimétrie ITC.

La microcalorimétrie confirme que le Le<sup>a</sup> est le meilleur ligand naturel de PA-III avec une affinité environ 10 fois supérieure à celle du fucose ( $K_a = 4.7 \cdot 10^6 \text{M}^{-1}$ ,  $K_d = 2.1 \cdot 10^{-7} \text{M}$ ). La structure du complexe PA-III/Le<sup>a</sup> a été résolue à 1.75 Å de résolution. Il apparaît clairement que, en supplément des contacts précédemment identifiés entre le fucose et la protéine, une stabilisation supplémentaire intervient grâce à un contact entre le O-6 du GlcNAc et la chaîne principale de la protéine. De plus, le galactose établit également un contact avec la protéine par l'intermédiaire d'une molécule d'eau. PA-III a pu également être cristallisé avec le pentasaccharide de lait humain LNnFP-V, fournissant ainsi la première structure démontrant comment le lait humain peut bloquer l'adhésion de pathogène dans l'intestin des nouveaux nés. Cette structure a été résolue à 1.05 Å et le pentasaccharide établit des contacts avec un sucre voisin dans le cristal, ce qui permet de le localiser avec précision.



(Article n°3)

**Structural Basis for the Interaction between Human Milk Oligosaccharides  
and the Bacterial lectin PA-III of *Pseudomonas aeruginosa***

Perret S, **Sabin C**, Dumon C, Pokorná M, Gautier C, Galanina O, Ilia S, Bovin N,  
Nicaise M, Desmadril M, Gilboa-Garber N, Wimmerová M, Mitchell EP and Imberty A

*Biochemical Journal*, 389: 325-32, (2005)





# Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of *Pseudomonas aeruginosa*

Stéphanie PERRET\*<sup>1</sup>, Charles SABIN\*<sup>†1</sup>, Claire DUMON\*, Martina POKORNÁ‡, Catherine GAUTIER\*, Oxana GALANINA§, Shahov ILIA§, Nicolai BOVIN§, Magali NICAISE||, Michel DESMADRIL||, Nechama GILBOA-GARBER¶, Michaela WIMMEROVÁ‡, Edward P. MITCHELL† and Anne IMBERTY\*<sup>2</sup>

\*Centre de Recherches sur les Macromolécules Végétales, CNRS (affiliated with Université Joseph Fourier), BP53, 38041 Grenoble Cedex 09, France, †E.S.R.F. Experiments Division, BP 220, F-38043, Grenoble Cedex, France, ‡National Centre for Biomolecular Research and Department of Biochemistry, Masaryk University, Kotlarska 2, 611 37 Brno, Czech Republic, §Shemyakin & Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences ul Miklukho-Maklaya 16/10, 17871 GSP-7, V-437, Moscow, ||UMR 8619 – IBBMC, Modélisation et Ingénierie des Protéines, Bât. 430 Université Paris XI, 91405 Orsay, France, and ¶Bar-Ilan University, Faculty of Life Sciences, Ramat Gan 52900, Israel

One of the mechanisms contributing to the protection by breast-feeding of the newborn against enteric diseases is related to the ability of human milk oligosaccharides to prevent the attachment of pathogenic bacteria to the duodenal epithelium. Indeed, a variety of fucosylated oligosaccharides, specific to human milk, form part of the innate immune system. In the present study, we demonstrate the specific blocking of PA-IIL, a fucose-binding lectin of the human pathogen *Pseudomonas aeruginosa*, by milk oligosaccharides. Two fucosylated epitopes, Lewis a and 3-fucosyl-lactose (Lewis x glucose analogue) bind to the lectin with dissociation constants of  $2.2 \times 10^{-7}$  M and  $3.6 \times 10^{-7}$  M respectively. Thermodynamic studies indicate that these interactions are dominated by enthalpy. The entropy contribution is slightly favourable when

binding to fucose and to the highest-affinity ligand, Lewis a. The high-resolution X-ray structures of two complexes of PA-IIL with milk oligosaccharides allow the precise determination of the conformation of a trisaccharide and a pentasaccharide. The different types of interaction between the oligosaccharides and the protein involve not only hydrogen bonding, but also calcium- and water-bridged contacts, allowing a rationalization of the thermodynamic data. This study provides important structural information about compounds that could be of general application in new therapeutic strategies against bacterial infections.

**Key words:** crystal structure, fucose, human milk oligosaccharide, lectin, *Pseudomonas aeruginosa*, thermodynamics.

## INTRODUCTION

*Pseudomonas aeruginosa* is an opportunistic pathogen that infects immunosuppressed individuals and compromised tissues, such as burn wounds or the trachea of intubated patients. The bacteria endanger especially patients with cystic fibrosis. For pathogenic organisms, the ability to adhere to host tissues is essential to initiate an infection, and host cell surface glycoconjugates represent a specific target for bacterial receptors [1]. *P. aeruginosa* produces a variety of carbohydrate-binding proteins that could be involved in host recognition and adhesion. Some of them are located at the tip of pili [2] and flagella [3], whereas others are soluble lectins present in the cytoplasm and at the surface of the bacteria [4].

The role of carbohydrate–lectin adhesion in several bacterial infections of the stomach, ears and bladder has been recognized, and oligosaccharide-based treatments have been tried [5,6]. In the intestine of the human newborn, maternal milk oligosaccharides offer natural protection against pathogen infection [7]. Human milk contains a significant amount of structurally diverse oligosaccharides, most of which are fucosylated neutral oligosaccharides carrying lactose at their reducing end [8] (Figure 1). Since human milk oligosaccharides are soluble analogues of epithelial cell surface glycoconjugates, they competitively inhibit the binding of pathogenic bacteria and viruses to epithelial ligands [9]. More particularly, fucosylated human milk oligosaccharides inhibit binding *in vitro* or *in vivo* of several pathogens, such as entero-

pathic *Escherichia coli* [10], *Campylobacter jejuni* [11], Norwalk-like virus [12] and *Helicobacter pylori* [13]. *P. aeruginosa* can also colonize the gastrointestinal tract and cause paediatric diarrhoea. Moreover, undetected endogenous gastrointestinal carriage can lead to severe infection in other parts of the body [14,15].

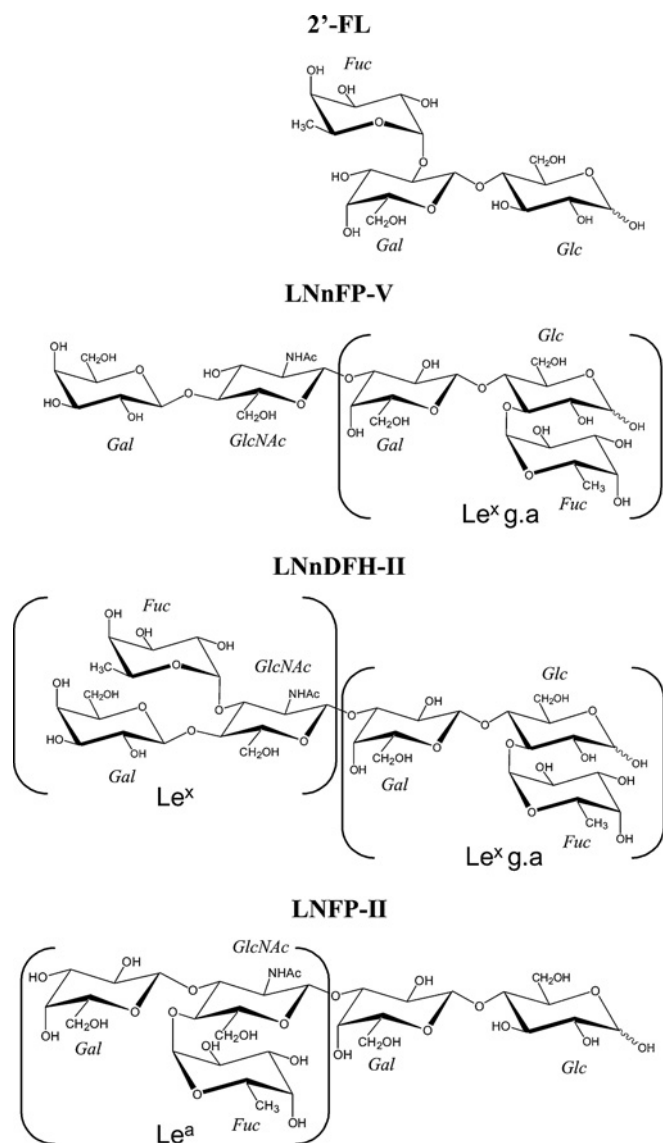
A fucose-binding lectin, PA-IIL, has been isolated from the *P. aeruginosa* cytoplasm [4] and its amino acid sequence characterized [16]. This soluble lectin is also present at the surface of the bacterial cell [17], is regulated by quorum sensing and is associated with virulence factors [18]. It has been demonstrated recently that human milk, but not cow's milk, specifically blocks haemagglutination by PA-IIL [19]. In the last 2 years, our group has deciphered the structural basis of the interaction between PA-IIL and fucose [20,21]. Fucose binding is mediated by two calcium ions, and this very unusual binding mode is responsible for the affinity ( $K_a$   $0.18 \times 10^6$  M<sup>-1</sup>), which is much higher than is classically observed in protein–carbohydrate interactions [22].

Characterization of the interaction between PA-IIL and human milk oligosaccharides may help in understanding the role of PA-IIL in bacterial infection. Furthermore, detailed description of the structural and thermodynamic interactions will help in the design of carbohydrate-based compounds that can be used against *P. aeruginosa* infections such as those that threaten the lives of cystic fibrosis patients. We present here a study of the fine specificity of PA-IIL towards a number of human milk oligosaccharides, and characterization of the thermodynamics of binding

Abbreviations used: ELLA, enzyme-linked lectin assay; FucLac, fucosyl-lactose; ITC, isothermal titration microcalorimetry; Le<sup>a</sup>, Lewis a; Le<sup>x</sup>, Lewis x; Le<sup>x</sup> g.a., Le<sup>x</sup> glucose analogue; LNFP-II, lacto-*N*-fucopentaose-II; LNnDFH-II, lacto-*N*-neodifucohexaose-II; LNnFP-V, lacto-*N*-neofucopentaose-V; Me- $\alpha$ Fuc, methyl- $\alpha$ -L-fucopyranoside; SiaLe<sup>a</sup>, sialyl Lewis a; SiaLe<sup>x</sup>, sialyl Lewis x; T-PBS, PBS containing 0.05% Tween.

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> To whom correspondence should be addressed (email imberty@cermav.cnrs.fr).



**Figure 1** Representation of major human milk oligosaccharides illustrating the great variety of neutral oligosaccharide structures

All core molecules carry lactose at their reducing end. 2'-FL, 2'-FucLac. Lewis structures are shown in brackets.

for the highest-affinity ligand. The crystal structures of PA-IIL in complex with fucosylated tri- and penta-saccharide have been solved to a resolution of 1.75 and 1.05 Å respectively (where 1 Å = 0.1 nm), yielding the first atomic-resolution structures of human milk oligosaccharides and their interaction with bacterial receptors.

## MATERIALS AND METHODS

### Materials

Recombinant PA-IIL was purified from *Escherichia coli* BL21(DE3) containing the plasmid pET25pa2l as described previously [21]. L-Fuc and 2'-FucLac (2'-fucosyl-lactose) were purchased from Sigma; Me- $\alpha$ Fuc (methyl- $\alpha$ -L-fucopyranoside) was purchased from Interchim; LNFP-II (lacto-*N*-fucopentaose II), Le<sup>x</sup> (Lewis x), Le<sup>a</sup> (Lewis a), SiaLe<sup>a</sup> (sialyl Lewis a) and SiaLe<sup>x</sup> (sialyl Lewis x) were purchased from Dextra. 3-FucLac,

LNnFP-V (lacto-*N*-neofucopentaose-V) and LNnDFH-II (lacto-*N*-neodifucohexaose-II) were produced in engineered *E. coli* strains and purified as described previously [23].

### ELLA (enzyme-linked lectin assay) experiments

ELLAs were conducted using 96-well microtitre plates (Nunc Maxisorb) coated with PA-IIL (5  $\mu$ g/ml) diluted in carbonate buffer, pH 9.6 (100  $\mu$ l) for 1 h at 37 °C. After blocking at 37 °C for 1 h with 100  $\mu$ l per well of 3% (w/v) BSA in PBS, plates were incubated at 37 °C for 1 h with 100  $\mu$ l of biotinylated polymeric fucose (Lectinity Holding, Inc.) at 5  $\mu$ g/ml in the presence of serial dilutions of inhibitors. After washing with T-PBS (PBS containing 0.05% Tween), 100  $\mu$ l of streptavidin-peroxidase conjugate (dilution 1:5000; Boehringer-Mannheim) was added and left for 1 h at 37 °C. The colour was developed using 100  $\mu$ l per well of 0.05 M phosphate/citrate buffer containing *O*-phenylenediamine dihydrochloride (0.4 mg/ml) and urea hydrogen peroxide (0.4 mg/ml) (Sigma-Aldrich). The reaction was stopped by the addition of 50  $\mu$ l of 30% H<sub>2</sub>SO<sub>4</sub>. Absorbance was read at 490 nm using a microtitre plate reader (Bio-Rad; model 680).

### ITC (isothermal titration microcalorimetry) analysis

ITC experiments were performed with a VP-ITC isothermal titration calorimeter (Microcal). The experiments were carried out at 25 °C. Sugars and proteins were dissolved in the same buffer, i.e. 0.1 M Tris with 0.03 mM CaCl<sub>2</sub> at pH 7.5. The protein concentration in the microcalorimeter cell (1.4 ml) varied from 23.5 to 50  $\mu$ M. A total of 22 injections of 13  $\mu$ l of sugar solution at concentrations varying from 0.24 to 0.5 mM were added at intervals of 5 min while stirring at 310 rev./min. Control experiments performed by injection of buffer into the protein solution 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, from the relationship:

$$Q_i = \frac{nP_t \Delta H V_o}{2} \left[ 1 + \frac{A_i}{nP_t} + \frac{1}{nK P_t} - \sqrt{\left( 1 + \frac{A_i}{nP_t} + \frac{1}{nK P_t} \right)^2 - 4 \frac{A_i}{nP_t}} \right]$$

where  $P_t$  is the total protein concentration,  $A_i$  is the total concentration of the ligand,  $V_o$  is the volume of the cell and  $Q_i$  is the total heat released for injection  $i$ .  $\Delta G$  values and entropy contributions were determined from the standard equation:

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

where  $\Delta G$ ,  $\Delta H$  and  $\Delta S$  are the changes in free energy, enthalpy and entropy respectively of binding,  $T$  is the absolute temperature,  $R = 8.32 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$  and  $K$  is the association constant. All experiments were performed with  $c$  values  $100 < c < 200$  ( $c = K_a M$ , where  $M$  is the initial concentration of the macromolecule).

### Crystallization

Crystallization conditions were screened using the hanging-drop vapour diffusion method from complete Hampton Screens I and II (Hampton Research) and the best conditions were selected. For the PA-IIL-LNnFP-V complex, hanging drops were prepared using 2.5  $\mu$ l of lyophilized purified PA-IIL dissolved in water (10 mg  $\cdot$  ml<sup>-1</sup>) in the presence of LNnFP-V (1300  $\mu$ g  $\cdot$  ml<sup>-1</sup>) and

CaCl<sub>2</sub> salt (2 mM) and 2.5  $\mu$ l of reservoir solution (1.75 M ammonium sulphate in 0.1 M Tris/HCl, pH 8.5). Crystals, obtained within 3 weeks at 20 °C, belong to space group P1 ( $a = 47.01$  Å,  $b = 48.35$  Å,  $c = 52.47$  Å,  $\alpha = 85.35^\circ$ ,  $\beta = 80.05^\circ$  and  $\gamma = 65.60^\circ$ ), with four molecules per asymmetric unit, corresponding to a  $V_m$  (Matthews coefficient) of  $2.3$  Å<sup>3</sup> · Da<sup>-1</sup> and a solvent content of 46%. For the PA-IIL–Le<sup>a</sup> complex, lyophilized protein was dissolved in water (10 mg · ml<sup>-1</sup>) in the presence of Le<sup>a</sup> (806  $\mu$ g · ml<sup>-1</sup>) and CaCl<sub>2</sub> salt (2 mM). Crystalline plates were obtained from hanging drops of 2.5  $\mu$ l of protein solution mixed with 2.5  $\mu$ l of reservoir solution {0.1 M sodium cacodylate at pH 6.5, 30% PEG [poly(ethylene glycol)] 8000 and 0.2 M ammonium sulphate}. The crystal plates belong to space group P2<sub>1</sub> ( $a = 51.57$  Å,  $b = 75.82$  Å,  $c = 52.74$  Å and  $\beta = 93.48^\circ$ ), with four molecules per asymmetric unit, resulting in a  $V_m$  of  $2.2$  Å<sup>3</sup> · Da<sup>-1</sup> and a solvent content of 44%.

### Data collection and structure solution

Crystals were cryo-cooled at 100 K after soaking them for as short a time as possible in 25% (v/v) glycerol in the precipitant solution. All data images were recorded on an ADSC Q4R CCD detector (Quantum Corp.) at a wavelength of 0.934 Å on beamline ID14-1 at ESRF (Grenoble, France). Data were processed with MOSFLM [24], and scaled and converted to structure factors using the CCP4 programme suite [25]. In the particular case of the 1.05 Å PA-IIL–LNnFP-V complex, to avoid the loss of the strongest reflections owing to saturation, three sets of data (one low- and two high-resolution passes) were collected from the same crystal.

Phasing of the PA-IIL–LNnFP-V complex was performed with ACORN [26], using only the positions of three of the calcium ions from the crystal structure of the PA-IIL–fucose complex solved previously at 1.0 Å resolution [21]. Using the ACORN phases, an initial structure was built automatically using ARP/warp [27]. The resulting electron density maps showed clear features corresponding to the oligosaccharide ligands. Cycles of iterative manual rebuilding and refinement were performed using the programs O [28] and REFMAC [29]. In the binding sites of monomers A and C, only four and three sugar residues respectively could be positioned in the electron density maps, whereas a pentasaccharide could be positioned in monomers B and D.

The structure of the Le<sup>a</sup>-complexed lectin was solved by the molecular replacement technique with the MOLREP program [25], using the monomeric structure of the PA-IIL–fucose complex (PDB code 1GZT) stripped of water molecules and calcium ions as the search probe. Four clear solutions were found, corresponding to a tetramer in the asymmetric cell described previously [20]. The initial  $F_o - F_c$  difference maps showed clearly the presence of four Le<sup>a</sup> ligands, one for each monomer. Cycles of manual model rebuilding with O and refinement with REFMAC gave the final structure. Co-ordinates of the final structures (see refinement statistics in Table 1) have been deposited in the Protein Data Bank under codes 1W8F and 1W8H for the LNnFP-V- and Le<sup>a</sup>-complexed structures respectively.

## RESULTS

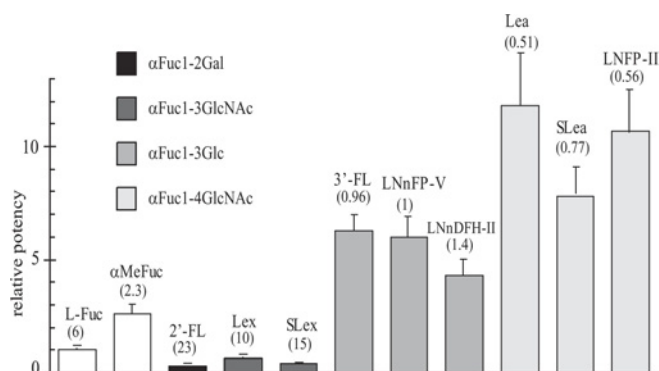
### Specificity of the interaction between PA-IIL and human milk oligosaccharides

Since haemagglutination by PA-IIL has been demonstrated to be blocked specifically by human milk [19], we studied the affinity of different human milk oligosaccharides towards the lectin using ELLAs. The fucosylated oligosaccharides shown in Figure 1 were tested for their inhibitory properties against the binding

**Table 1** Data collection and refinement statistics

Data in parentheses show statistics for the highest-resolution shell. rmsd, root mean square deviation.

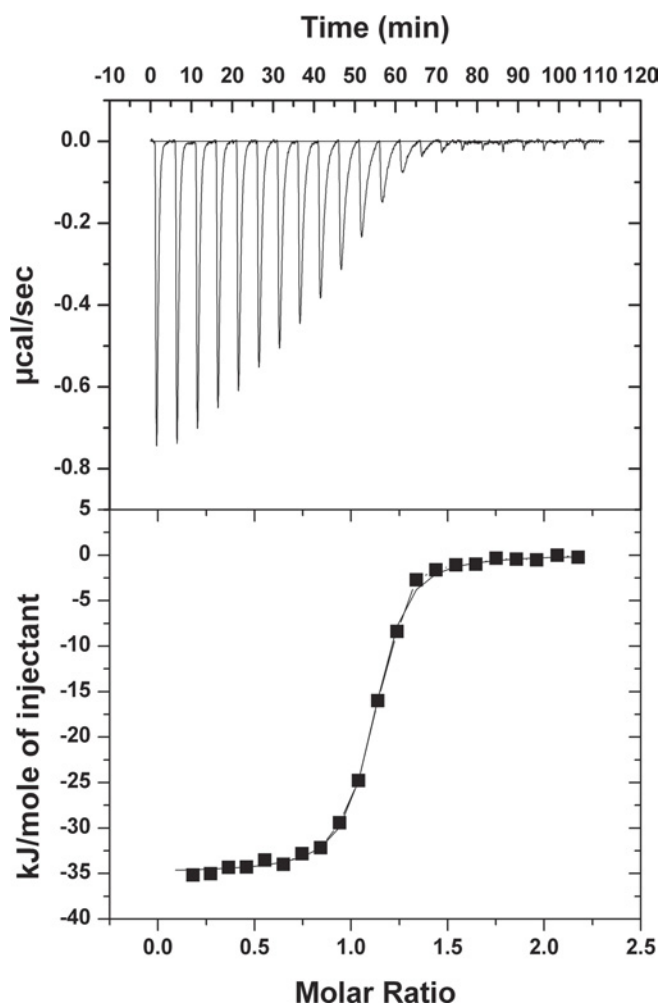
Parameter	PA-IIL–LNnFP-V	PA-IIL–Le <sup>a</sup>
Data collection		
Wavelength (Å)	0.934	0.934
Resolution (Å)	14.51–1.05 (1.08–1.05)	26.82–1.75 (1.81–1.75)
Space group	P1	P2 <sub>1</sub>
Unit cell	$a = 47.015$ Å, $b = 48.352$ Å, $c = 52.472$ Å, $\alpha = 85.35^\circ$ , $\beta = 80.05^\circ$ , $\gamma = 65.60^\circ$	$a = 51.57$ Å, $b = 75.82$ Å, $c = 52.74$ Å, $\beta = 93.48^\circ$
Observed reflections	737 998	141 130
Unique reflections	180 478	34 767
Multiplicity	4.1 (3.5)	4.1 (3.4)
Completeness (%)	93.3 (89.5)	85.7 (41.7)
$I/\sigma I$	7.0 (1.8)	13.2 (3.8)
$R_{\text{merge}}$ (%)	6.6 (32.2)	8.8 (33.0)
Refinement statistics		
$R_{\text{cryst}}$ (%) (no. reflections)	10.5 (171 406)	12.5 (34 037)
$R_{\text{free}}$ (%) (no. reflections)	12.3 (9072)	16.9 (730)
Average B factor (Å <sup>2</sup> )	7.314	9.679
Protein atoms (n)	3564	3362
Ligand/water atoms (n)	197/727	204/438
rmsd bonds (Å)	0.021	0.014
rmsd angles (°)	2.1	1.5



**Figure 2** Potency of different sugars relative to fucose

IC<sub>50</sub> values in  $\mu$ M, as determined from the inhibition curves, are given in parentheses. Each value has been averaged from at least three independent measurements. FL, FucLac; SLe, SiaLe.

of L-fucose-biotinylated polyacrylamide to immobilized PA-IIL. These oligosaccharides are representative of the variety of fucosylated epitopes found in human milk oligosaccharides:  $\alpha$ Fuc1-2Gal,  $\alpha$ Fuc1-4GlcNAc,  $\alpha$ Fuc1-3GlcNAc and  $\alpha$ Fuc1-3Glc. IC<sub>50</sub> values ranged from 23  $\mu$ M for 2'-FucLac to 0.5  $\mu$ M for the best inhibitor, Le<sup>a</sup>. After normalizing values to that of fucose, the inhibitory potency of the oligosaccharides appears to be directly dependent on the linkage between the fucose moiety and the carrying oligosaccharides (Figure 2). Oligosaccharides containing a fucose residue linked to position 2 of galactose (2'-FucLac) are poor inhibitors. Similarly, the Le<sup>x</sup> series, which contain a fucose on position 3 of a GlcNAc residue, bind less efficiently to PA-IIL than does fucose alone. On the other hand, the Le<sup>x</sup> g.a. (Le<sup>x</sup> glucose analogue) series, in which the GlcNAc is replaced by Glc, binds approx. 6-fold better than fucose. The best inhibitors belong to the Le<sup>a</sup> series, which contain fucose linked on position 4 of a GlcNAc residue; these are 8–10 times stronger inhibitors than fucose. These data demonstrate that the presence of a sialyl group does not improve the inhibitory potency of the Le<sup>a</sup> and Le<sup>x</sup>



**Figure 3** ITC analysis of the interaction of PA-IIL with Le<sup>a</sup>

Curves are shown for Le<sup>a</sup> (0.324 mM) binding to PA-IIL (32.7 μM) in 100 mM Tris buffer containing 0.03 mM CaCl<sub>2</sub>, pH 7.5, at 25 °C. Upper panel: data were obtained from 22 automatic injections (13 μl each) of Le<sup>a</sup> into PA-IIL in the cell. Lower panel: plot of the total heat released as a function of total ligand concentration for the titration shown in the upper panel (■). The solid line represents the best least-squares fit for the obtained data. 1 kcal = 4.184 kJ.

oligosaccharides. Moreover, the length of the oligosaccharides containing the Le<sup>x</sup> g.a. or Le<sup>a</sup> epitope does not modify the strength of the interaction. This indicates that only the Lewis epitope part of the oligosaccharide is responsible for the high-affinity binding to the lectin. Together, these data confirm and complete those obtained previously using different binding assays [20].

**Table 2** ITC analysis for binding of PA-IIL to various oligosaccharides

Values are means ± S.D. from at least two separate titrations.

Oligosaccharide	$10^{-6} \times K_a$ (M <sup>-1</sup> )	<i>n</i>	−Δ <i>G</i> (kJ/mol)	−Δ <i>H</i> (kJ/mol)	− <i>T</i> Δ <i>S</i> (kJ/mol)
Fucose	0.34 ± 0.03	0.96 ± 0.04	31.5 ± 0.2	31.2 ± 0.1	−0.3 ± 0.1
Le <sup>a</sup>	4.7 ± 0.2	1.08 ± 0.01	38.06 ± 0.08	34.95 ± 0.07	−3.11 ± 0.01
3-FucLac	2.7 ± 0.2	1.02 ± 0.01	36.7 ± 0.2	40.0 ± 0.5	3.3 ± 0.6
LNnFP-V	1.56 ± 0.04	1.02 ± 0.01	35.33 ± 0.06	35.6 ± 0.2	0.3 ± 0.3
LNFP-II	2.01 ± 0.04	0.94 ± 0.01	35.96 ± 0.05	37.8 ± 0.4	1.8 ± 0.4
Le <sup>x</sup>	0.29 ± 0.02	0.95 ± 0.01	31.1 ± 0.2	22.3 ± 0.1	−8.8 ± 0.3

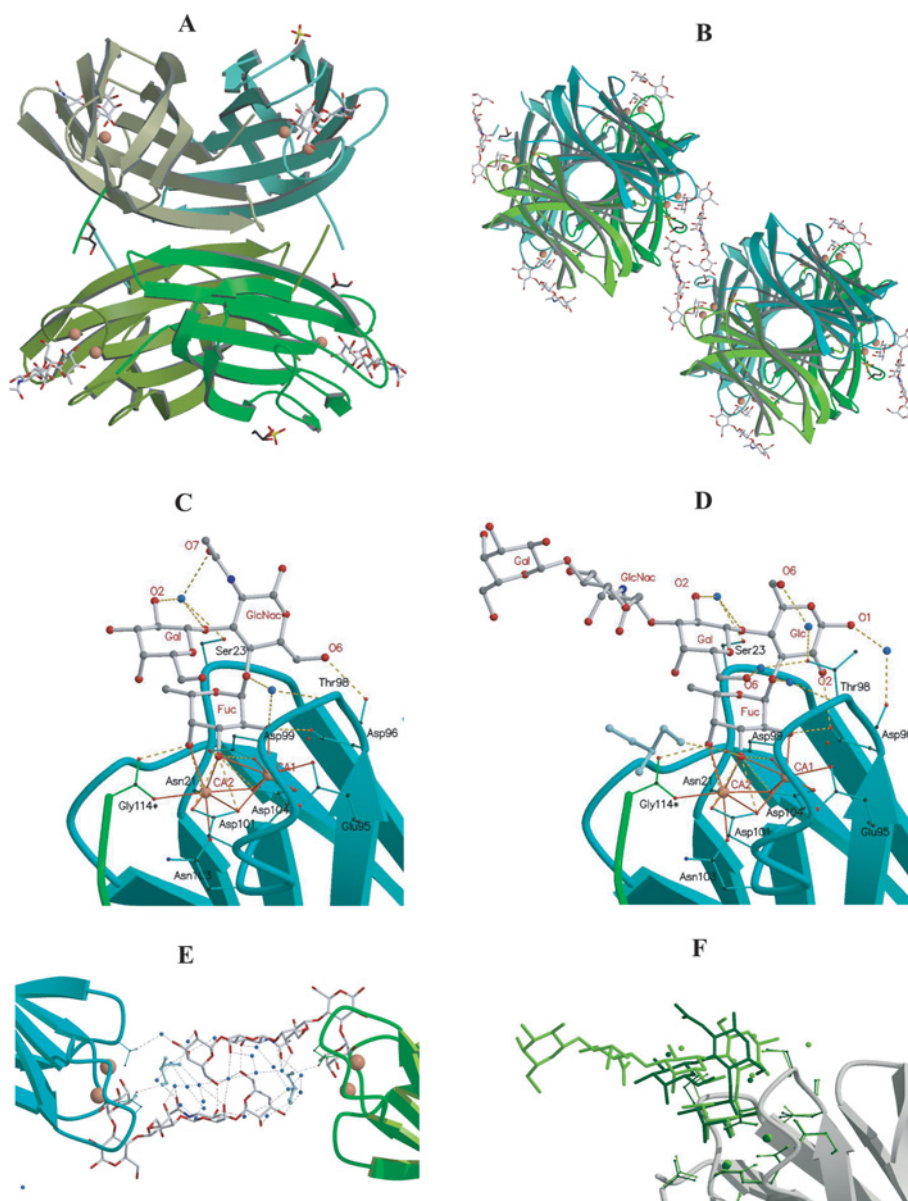
### Thermodynamics of the interactions between PA-IIL and human milk oligosaccharides

In order to characterize further the interactions of the lectin with some of the most potent competitors, *K<sub>a</sub>* values and thermodynamic binding parameters were analysed using ITC. A typical titration curve for the Le<sup>a</sup> trisaccharide, the most potent ligand of PA-IIL found in the ELLA experiments described above, is shown in Figure 3. The association constant of PA-IIL–Le<sup>a</sup> binding was  $4.7 \times 10^6$  M<sup>-1</sup>, some 14 times stronger than for binding of L-fucose (Table 2). The Le<sup>x</sup> g.a. from human milk and the Le<sup>x</sup> g.a.-containing oligosaccharides LNFP-V and LNFP-II presented *K<sub>a</sub>* values 5–8 times higher than that for L-fucose. On the other hand, the affinity of the lectin for the Le<sup>x</sup> trisaccharide was slightly lower than for fucose. These values are in agreement with the relative inhibitory potencies calculated from the ELLA experiments (Figure 2). The affinity constants described here for PA-IIL are much higher than those usually observed for lectin–oligosaccharide interactions [22].

Analyses of the different thermodynamic contributions to ligand binding indicate that the interaction is unambiguously enthalpy-driven. In all cases except one, the entropy term is small, with the *T*Δ*S* value close to 10% of the free-energy term −Δ*G* (Table 2). The strong enthalpy term is not surprising, since the characteristics of fucose binding to the primary binding site are dominated by hydrogen bonding and a very unusual co-ordination to two calcium ions [20,21]. Interestingly, there is no or almost no entropy barrier for the binding of mono- or oligo-saccharides to PA-IIL. Two ligands, L-fucose and Le<sup>a</sup> trisaccharide, present a slightly favourable entropy term, and for one ligand, Le<sup>x</sup>, the entropy contributes almost one-third (28%) of the binding free energy. This is unusual in protein–carbohydrate interactions, where several factors, usually attributed to flexibility of the ligand and displacement of water molecules from amphiphilic surfaces [30], generally yield an unfavourable entropy term [22]. The oligosaccharides of the Lewis series are uncommonly rigid, due to their branched structure [31,32], and this could explain the atypical favourable entropy contribution. Interestingly 3-FucLac, the glucose analogue of Le<sup>x</sup>, binds with unfavourable entropy, which is perhaps due to a decrease in the rigidity of the trisaccharide with the absence of the N-acetyl group at position 2 (Figure 1).

### Crystal structures of complexes of PA-IIL with two human milk oligosaccharides

The crystal structures of the PA-IIL–Le<sup>a</sup> and PA-IIL–LNnFP-V complexes contain the same overall features as the PA-IIL–Fuc complex crystallized in the *P*2<sub>1</sub> space group and previously refined to 1.0 Å resolution [21], and in space group *P*2<sub>1</sub>2<sub>1</sub> at 1.2 Å resolution [17], showing a conserved quaternary arrangement, as well as the calcium- and fucose-binding modes (Figures 4A and 4B). The structural basis for fucose binding was



**Figure 4** Graphical representation of PA-IIL-(milk oligosaccharide) complexes

(A) Ribbon representation of the PA-IIL–Le<sup>a</sup> complex tetramer, consisting of the asymmetric unit with stick representation of sugar and sulphate and cpk representation of the calcium ions. (B) Same representation for the structure of the PA-IIL–LNnFP–V complex, showing the interaction between two tetramers (orientation orthogonal to that of A). (C) Interactions of PA-IIL with calcium ions and Le<sup>a</sup>. Co-ordination contacts are indicated by orange solid lines and hydrogen bonds by green dashed lines. (D) Same representation for the interaction of PA-IIL with calcium ions and LNnFP–V. (E) Interaction between two pentasaccharides in the PA-IIL–LNnFP–V complex. Hydrogen bonds are represented by dotted lines. (F) Superimposition of the binding sites of the PA-IIL–Le<sup>a</sup> and PA-IIL–LNnFP–V complexes.

discussed in detail previously [21]. There are few structural data on fucose-binding lectins at the present time; complexes of fucose with *Ulex europaeus* lectin I [33], eel lectin [34], fungal *Aleuria aurantia* lectin [35] and mutated mannan-binding protein-A [36] interestingly display completely different structural features, and no comparisons can be made.

Of special interest is the analysis of the conformations of the two oligosaccharides that are reported here for the first time in the crystalline state, together with their interactions with the protein receptor. There are no significant differences among the four monomers of PA-IIL–Le<sup>a</sup> present in the asymmetric unit. Analysis of the mode of binding of Le<sup>a</sup> by PA-IIL reveals a large number of contacts between the sugar and the protein (Figure 4C). All three of the monosaccharides interact with the protein surface

either directly (hydrogen bonding and co-ordination of calcium) or indirectly through bridging water molecules (Table 3). In addition to the nine direct hydrogen bonds between the fucose and the protein described previously, the GlcNAc residue establishes a hydrogen bond between its O-6 atom and the main-chain carbonyl group of Asp-96. Two water molecules play an important bridging role in the interaction: one bridges O-1 and O-2 of fucose to the amide nitrogen of Thr-98, whereas the other bridges the Gal residue (O-2 and O-1) to the side chain of Ser-23. Hydrophobic contacts are very limited and only involve the methyl group of fucose with that of Thr-145. Due to its branched structure, Le<sup>a</sup> has been described as a rigid trisaccharide in NMR and modelling studies [31,32]. Indeed, our results show that the conformation adopted by Le<sup>a</sup> in the binding site is similar to that predicted by

**Table 3** Contact of interest in the binding site of PA-IIL-oligosaccharide complexes

Distances (Å) are averaged over the four proteins in the asymmetric unit, and S.D.s on the last digit are given in parentheses. Wat\_1 and Wat\_2 are the water molecules present in the carbohydrate-binding site (coloured blue in Figures 4A and 4B). The asterisk in Gly-114\* OXT indicates that this residue belongs to the neighbouring monomer. Bridging is indicated by superscript numbers as follows: <sup>1</sup> bridged to Thr-98 N, <sup>2</sup> bridged to Thr-98 N and Asp-99 N, <sup>3</sup> bridged to Ser-23 OG, <sup>4</sup> bridged to Asp-96 O, <sup>5</sup> bridged to Thr-98 OG1.

PA-IIL-Le <sup>a</sup> complex			PA-IIL-LNnFP-V complex		
Interacting species		Distance (Å)	Interacting species		Distance (Å)
PA-IIL	Le <sup>a</sup>		PA-IIL	LNnFP-V	
Ca-1	O-2 Fuc	2.50 (3)	Ca-1	O-2 Fuc	2.45 (1)
Ca-1	O-3 Fuc	2.53 (5)	Ca-1	O-3 Fuc	2.47 (1)
Ca-2	O-3 Fuc	2.50 (3)	Ca-2	O-3 Fuc	2.47 (1)
Ca_2	O-4 Fuc	2.47 (2)	Ca_2	O-4 Fuc	2.49 (1)
Wat_1 <sup>1</sup>	O-1 Fuc	2.96 (9)	Wat_1 <sup>2</sup>	O-1 Fuc	2.96 (4)
Wat_1 <sup>1</sup>	O-2 Fuc	3.1 (1)			
Asp-96 OD1	O-2 Fuc	2.64 (2)	Asp-96 OD1	O-2 Fuc	2.7 (1)
Asp-99 OD1	O-3 Fuc	2.98 (7)	Asp-99 OD1	O-3 Fuc	2.97 (2)
Asp-99 OD2	O-3 Fuc	2.48 (1)	Asp-99 OD2	O-3 Fuc	2.55
Asp-101 OD1	O-3 Fuc	2.94 (1)	Asp-101 OD1	O-3 Fuc	2.97 (1)
Asp-101 OD2	O-3 Fuc	2.94 (3)	Asp-101 OD2	O-3 Fuc	2.98 (1)
Asp-104 OD2	O-3 Fuc	3.07 (3)	Asp-104 OD2	O-3 Fuc	3.00 (1)
Asn-21 O	O-4 Fuc	2.98 (4)	Asn-21 O	O-4 Fuc	3.01 (4)
Gly-114* OXT	O-4 Fuc	2.55 (3)	Gly-114* OXT	O-4 Fuc	2.56 (1)
Ser-23 N	O-5 Fuc	3.03 (3)	Ser-23 N	O-5 Fuc	2.96 (1)
Thr-45 CG2	C-6 Fuc	4.10 (5)	Thr-45 CG2	C-6 Fuc	4.1 (1)
Asp-96 O	O-6 GlcNAc	2.7 (1)	Asp-96 OD1	O-2 Glc	2.74 (3)
Wat_2 <sup>3</sup>	O-7 GlcNAc	3.1 (2)	Wat_3 <sup>4</sup>	O-1 Glc	2.4 (1)
			Wat_4 <sup>5</sup>	O-6 Glc	3.0 (1)
Wat_2	O-1 Gal	3.08 (2)	Wat_2	O-2 Gal	2.72 (2)
Wat_2	O-2 Gal	2.72 (3)	Wat_1	O-6 Gal	2.71 (3)
			Wat_5 <sup>4</sup>	O-6 Gal	3.1 (4)

**Table 4** Torsion angles of interest of Le<sup>a</sup> in the binding site of PA-IIL

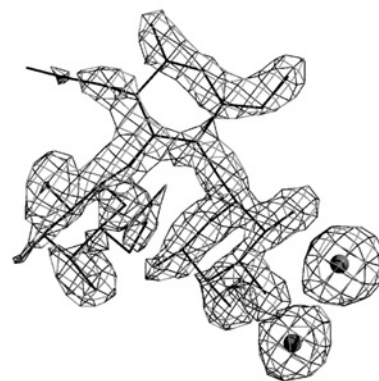
Mol A to Mol D correspond to the four different monomers present in the asymmetric unit.  $\Phi$  is defined as  $\Theta(0-5-C-1-0-1-C-x)$ ;  $\Psi$  is defined as  $\Theta[C-1-0-1-C-x-C(x+1)]$ .

	Mol A	Mol B	Mol C	Mol D
$\alpha$ Fuc1-4 $\beta$ GlcNAc				
$\Phi$	-66.9	-72.8	-75	-77.1
$\Psi$	-105.2	-100.6	-104.9	-105.4
$\beta$ Gal1-3 $\beta$ GlcNAc				
$\Phi$	-65.7	-71	-62.3	-78.3
$\Psi$	134.7	130.5	128.6	144.5

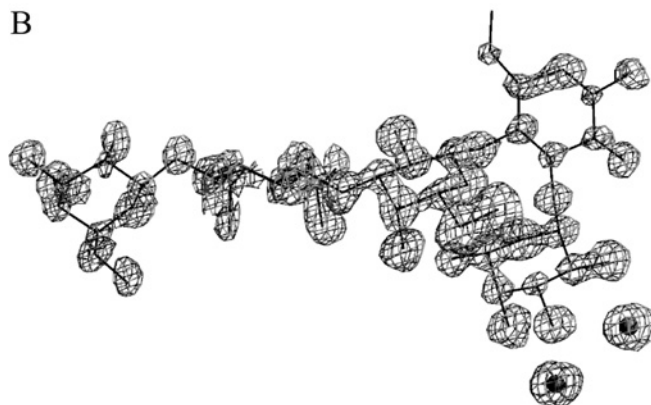
those earlier studies, confirming the rigidity of the trisaccharide. Torsion angles of interest are listed in Table 4.

Among the four independent molecules of the PA-IIL-LNnFP-V complex, clear electron density can be observed for two complete pentasaccharides (Figure 5), whereas only a tetrasaccharide and a trisaccharide can be located for the two remaining sites. The fact that two of the oligosaccharides can be completely visualized is due to the rigidity of their linear part resulting from fortuitous crystal packing that generates extensive contacts between two adjacent sugars centred around a pseudo two-fold axis of symmetry (Figure 4E). It is this interaction that allows the very-high-resolution structure of milk pentasaccharide to be described here for the first time. The protein-binding site, together with the calcium ions and fucose moiety, are identical with those described previously. Additional interactions with the protein surface are

A



B

**Figure 5** Initial sigma-weighted  $F_0-F_c$  electron density map

(A) The map has been contoured at  $1\sigma$  around the trisaccharide in the PA-IIL-Le<sup>a</sup> complex calculated after molecular replacement and first cycle of refinement without ligand and water. (B) Same representation for the pentasaccharide in the PA-IIL-LNnFP-V complex after phasing with Acorn. The final positions of the oligosaccharide and calcium in the model have been superimposed as thin lines for comparison.

established by the glucose and galactose residues of the Le<sup>x</sup> g.a. moiety (Figure 4D and Table 3). The glucose establishes a direct hydrogen bond between its hydroxy group at position 2 and the side chain of Asp-96. In addition, five interactions are mediated by a water molecule, involving O-1 of fucose, O-1 and O-6 of glucose, and O-1, O-2 and O-6 of galactose. Hydrophobic interactions are observed for the methyl group of fucose, but also for C-6 of galactose, which interacts with the carbon of a glycerol molecule located at the protein surface. The other part of the pentasaccharide, i.e. the  $\beta$ Glc1-4 $\beta$ GlcNAc linked to position 3 of galactose, does not interact with the protein surface, but instead is involved in extensive interactions with the neighbouring carbohydrate in the crystal lattice (Figure 4E).

The conformations of the pentasaccharide correspond to a low-energy minimum of each glycosidic linkage, as established previously by conformational analysis studies [32]. The 3-FucLac moiety (Le<sup>x</sup> g.a.) is the most rigid part, and it adopts a conformation with glycosidic linkage orientations similar to those observed in the crystal structure of hydrated Le<sup>x</sup> [37] (Table 5).

## DISCUSSION

The binding of carbohydrate ligands to the fucose-binding lectin of *P. aeruginosa* presents some particular features that are not commonly observed in protein-carbohydrate interactions. Its

**Table 5** Torsion angles of interest of LNNFP-V in the binding site of PA-III.

Mol A to Mol D correspond to the four different monomers present in the asymmetric unit.  $\Phi$  is defined as  $\Theta(0-5-C-1-O-1-C-x)$ ;  $\Psi$  is defined as  $\Theta[C-1-O-1-C-x-C-(x+1)]$ . For  $Le^x$ , torsion angles are given for  $\alpha$ Fuc1-3GlcNAc and  $\beta$ Gal1-4GlcNAc linkages in the two independent molecules of hydrated  $Le^x$  crystal.

	Mol A	Mol B	Mol B	Mol D	$Le^x$
$\alpha$ Fuc1-3 $\beta$ Glc					
$\Phi$	-76.4	-75.5	-78.9	-77.5	-72.5/-76.7
$\Psi$	138.6	141.2	142.3	140.6	139.2/139.0
$\beta$ Gal1-4 $\beta$ Glc					
$\Phi$	-70.3	-69.7	-68.3	-69.3	-80.0/-70.5
$\Psi$	-104.1	-104.2	-101.4	-104.0	-104.6/-107.7
$\beta$ GlcNAc1-3 $\beta$ Gal					
$\Phi$	-91.8	-84.6		-87.7	
$\Psi$	84.2	119.4		120.1	
$\beta$ Gal1-4 $\beta$ GlcNAc					
$\Phi$		-71.3		-76.9	
$\Psi$		-112.2		-111.8	

monosaccharide binding has been characterized previously: the low specificity was rationalized in terms of the structural features required to co-ordinate the two calcium ions [17], whereas the high affinity has also been attributed to the presence of cations with charge delocalization [21]. A special feature is the favourable, albeit very weak, entropy that has been observed for fucose binding that could be generated by the release of three tightly bound water molecules (observed in the crystal structure of the unliganded lectin [17]) that are replaced by fucose hydroxy groups in the complex.

When looking at oligosaccharide binding, and more particularly at human milk oligosaccharides that can play a role in early-stage colonization, the specificity of PA-III is limited to those oligosaccharides with fucose linked to position 4 of GlcNAc, and not to position 3. This is rationalized by the two high-resolution crystal structures presented here. In  $Le^a$ , the Fuc1-4GlcNAc linkage positions O-6 of GlcNAc in a favourable location to create an additional hydrogen bond with the protein backbone (Figure 4C). For the  $Le^x$  trisaccharide, the Fuc1-3GlcNAc linkage would locate the N-acetyl group to the same position, therefore creating steric conflicts; this will result in a deformation of the trisaccharide, weaker binding and, more particularly, as observed in ITC, a large decrease in the enthalpy term. In  $Le^x$  g.a., the N-acetyl group at position 2 is replaced by a hydroxy group that makes a hydrogen bond to the protein (Figure 4D). Superimposition of the two binding sites (Figure 4F) shows that  $Le^x$  g.a. adopts a slightly different orientation than  $Le^a$  in the binding site.

The specificity range established by ELLA experiments gives same order of ligand affinity as the titration study in solution, demonstrating that the lectin is fully active when fixed on the surface. The microcalorimetry study confirmed that the highest-affinity ligand was the  $Le^a$  trisaccharide, and that it binds with very high enthalpy contribution ( $\Delta H = -35$  kJ/mol) and slightly favourable entropy ( $-T\Delta S = -3$  kJ/mol). Interestingly, the  $Le^x$  g.a. presents an unfavourable entropy contribution. This could be due either to the absence of the N-acetyl group that brings higher flexibility to the adjacent linkage, or to the fact that the binding involves three more bridging water molecules, immobilization of which generates an entropy penalty.

The large variety of fucosylated oligosaccharides that are present in human milk (> 100) may represent an adapted and specific response to block a large number of pathogens. For example, *C. jejuni* is specifically blocked by H(O) antigen ( $\alpha$ Fuc1-2 $\beta$ Gal1-4GlcNAc) [11] whereas *H. pylori* is blocked by Lewis b epitope

( $\alpha$ Fuc1-2 $\beta$ Gal1-3[ $\alpha$ Fuc1-4]GlcNAc) [13]. The presence of the three different fucosyl epitopes in milk, i.e.  $\alpha$ Fuc1-2 $\beta$ Gal,  $\alpha$ Fuc1-3Glc/GlcNAc and  $\alpha$ Fuc1-4GlcNAc, is genetically determined primarily by the action of a Lewis gene coding for an  $\alpha$ 1,3/4-fucosyltransferase and a secretor gene coding for  $\alpha$ 1,2-fucosyltransferase [38]. In the European population, as a consequence of polymorphisms yielding inactive or partially active enzymes, approx. 70% of women will have  $\alpha$ 1,2-,  $\alpha$ 1,3- and  $\alpha$ 1,4-linked fucosyl residues in their milk. Fucosyl residues linked  $\alpha$ 1,2 and  $\alpha$ 1,4 will be absent from 20% and 10% of the population respectively. A clear relationship has been established between maternal genotype and the risk of diarrhoea in their infants [39]. The *P. aeruginosa* lectin is blocked by  $\alpha$ 1,4GlcNAc- but also by  $\alpha$ 1,3Glc-linked fucosyl residues that are present in the milk of all women. It is therefore most interesting that these two epitopes are the ones that present the highest affinity found to date for the soluble lectin of *Pseudomonas aeruginosa*.

We thank the ESRF, Grenoble, France, for access to synchrotron data collection facilities. The work was supported by the Ministry of Education of the Czech Republic (M. P. and M. W.; contract MSM 0021622413), by the Russian Academy of Sciences program 'Physico-chemical biology', by the French Ministry of Research ACI Microbiologie program, by the Mizutani Foundation for Glycoscience, and by Vaincre la Mucoviscidose. Travels and visits between NCBR and CERMAV are supported by a BARRANDE exchange program.

## REFERENCES

- 1 Beachey, E. H. (1981) Bacterial adherence: adhesin-receptor interactions mediating the attachment of bacteria to mucosal surface. *J. Infect. Dis.* **143**, 325-345
- 2 Hahn, H. P. (1997) The type-4 pilus is the major virulence-associated adhesin of *Pseudomonas aeruginosa*. *Gene* **192**, 99-108
- 3 Arora, S. K., Ritchings, B. W., Almira, E. C., Lory, S. and Ramphal, R. (1998) The *Pseudomonas aeruginosa* flagellar cap protein, FliD, is responsible for mucin adhesion. *Infect. Immun.* **66**, 1000-1007
- 4 Gilboa-Garber, N. (1982) *Pseudomonas aeruginosa* lectins. *Methods Enzymol.* **83**, 378-385
- 5 Mysore, J. V., Wigginton, T., Simon, P. M., Zopf, D., Heman-Ackah, L. M. and Dubois, A. (1999) Treatment of *Helicobacter pylori* infection in rhesus monkeys using a novel antiadhesion compound. *Gastroenterology* **117**, 1316-1325
- 6 Tong, H. H., McIver, M. A., Fisher, L. M. and DeMaria, T. F. (1999) Effect of lacto-N-neotetraose, asialoganglioside-GM1 and neuraminidase on adherence of otitis media-associated serotypes of *Streptococcus pneumoniae* to chinchilla tracheal epithelium. *Microb. Pathog.* **26**, 111-119
- 7 Sharon, N. and Ofek, I. (2000) Safe as mother's milk: carbohydrates as future anti-adhesion drugs for bacterial diseases. *Glycoconj. J.* **17**, 659-664
- 8 Chaturvedi, P., Warren, C. D., Altaye, M., Morrow, A. L., Ruiz-Palacios, G., Pickering, L. K. and Newburg, D. S. (2001) Fucosylated human milk oligosaccharides vary between individuals and over the course of lactation. *Glycobiology* **11**, 365-372
- 9 Kunz, C., Rudloff, S., Baier, W., Klein, N. and Strobel, S. (2000) Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu. Rev. Nutr.* **20**, 699-722
- 10 Newburg, D. S., Pickering, L. K., McCluer, R. H. and Cleary, T. G. (1990) Fucosylated oligosaccharides of human milk protect suckling mice from heat-stable enterotoxin of *Escherichia coli*. *J. Infect. Dis.* **162**, 1075-1080
- 11 Ruiz-Palacios, G. M., Cervantes, L. E., Ramos, P., Chavez-Munguia, B. and Newburg, D. S. (2003) *Campylobacter jejuni* binds intestinal H(O) antigen (Fuc alpha 1, 2Gal beta 1, 4GlcNAc), and fucosyloligosaccharides of human milk inhibit its binding and infection. *J. Biol. Chem.* **278**, 14112-14120
- 12 Huang, P., Farkas, T., Marionneau, S., Zhong, W., Ruvoen-Clouet, N., Morrow, A. L., Altaye, M., Pickering, L. K., Newburg, D. S., Le Pendu, J. and Jiang, X. (2003) Noroviruses bind to human ABO, Lewis, and secretor histo-blood group antigens: identification of 4 distinct strain-specific patterns. *J. Infect. Dis.* **188**, 19-31
- 13 Xu, H. T., Zhao, Y. F., Lian, Z. X., Fan, B. L., Zhao, Z. H., Yu, S. Y., Dai, Y. P., Wang, L. L., Niu, H. L., Li, N. et al. (2004) Effects of fucosylated milk of goat and mouse on *Helicobacter pylori* binding to Lewis b antigen. *World J. Gastroenterol.* **10**, 2063-2066
- 14 Olson, B., Weinstein, R. A., Nathan, C., Chamberlin, W. and Kabins, S. A. (1984) Epidemiology of endemic *Pseudomonas aeruginosa*: why infection control efforts have failed. *J. Infect. Dis.* **150**, 808-816
- 15 Ohara, T. and Itoh, K. (2003) Significance of *Pseudomonas aeruginosa* colonization of the gastrointestinal tract. *Intern. Med.* **42**, 1072-1076



- 16 Gilboa-Garber, N., Katcoff, D. J. and Garber, N. C. (2000) Identification and characterization of *Pseudomonas aeruginosa* PA-III lectin gene and protein compared to PA-II. *FEMS Immunol. Med. Microbiol.* **29**, 53–57
- 17 Loris, R., Tielker, D., Jaeger, K.-E. and Wyns, L. (2003) Structural basis of carbohydrate recognition by the lectin LecB from *Pseudomonas aeruginosa*. *J. Mol. Biol.* **331**, 861–870
- 18 Winzer, K., Falconer, C., Garber, N. C., Diggle, S. P., Camara, M. and Williams, P. (2000) The *Pseudomonas aeruginosa* lectins PA-II and PA-III are controlled by quorum sensing and by RpoS. *J. Bacteriol.* **182**, 6401–6411
- 19 Lesman-Movshovich, E., Lerrer, B. and Gilboa-Garber, N. (2003) Blocking of *Pseudomonas aeruginosa* lectins by human milk glycans. *Can. J. Microbiol.* **49**, 230–235
- 20 Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gautier, C., Pérez, S., Wu, A. M., Gilboa-Garber, N. and Imberty, A. (2002) Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nat. Struct. Biol.* **9**, 918–921
- 21 Mitchell, E. P., Sabin, S., Šnajdrová, L., Budová, M., Perret, S., Gautier, C., Hofr, C., Gilboa-Garber, N., Koca, J., Wimmerová, M. and Imberty, A. (2005) High affinity fucose binding of *Pseudomonas aeruginosa* lectin PA-III: 1.0 Å resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches. *Proteins: Struct. Funct. Bioinform.* **58**, 735–748
- 22 Dam, T. K. and Brewer, C. F. (2002) Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* **102**, 387–429
- 23 Dumon, C., Priem, B., Martin, S. L., Heyraud, A., Bosso, C. and Samain, E. (2001) *In vivo* fucosylation of lacto-N-neotetraose and lacto-N-neohexaose by heterologous expression of *Helicobacter pylori* alpha-1,3 fucosyltransferase in engineered *Escherichia coli*. *Glycoconj. J.* **18**, 465–474
- 24 Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* **26**
- 25 Computational Collaborative Project Number4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
- 26 Cowtan, K. D. and Zhang, K. Y. (1999) Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72**, 245–270
- 27 Perrakis, A., Morris, R. and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463
- 28 Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119
- 29 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
- 30 Lemieux, R. U., Delbaere, L. T., Beierbeck, H. and Spohr, U. (1991) Involvement of water in host-guest interactions. *Ciba Found. Symp.* **158**, 231–245
- 31 Lemieux, R. U., Bock, K., Delbaere, L. T. J., Koto, S. and Rao, V. S. R. (1980) The conformations of oligosaccharides related to the ABH and Lewis human blood group determinants. *Can. J. Chem.* **58**, 631–653
- 32 Imberty, A., Mikros, E., Koca, J., Mollicone, R., Oriol, R. and Pérez, S. (1995) Computer simulation of histo-blood group oligosaccharides. Energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens. *Glycoconj. J.* **12**, 331–349
- 33 Audette, G. F., Olson, D. J. H., Ross, A. R. S., Quail, J. W. and Delbaere, L. T. (2002) Examination of the structural basis for O(H) blood group specificity by *Ulex europaeus* lectin I. *Can. J. Chem.* **80**, 1010–1021
- 34 Bianchet, M. A., Odom, E. W., Vasta, G. R. and Amzel, L. M. (2002) A novel fucose recognition fold involved in innate immunity. *Nat. Struct. Biol.* **9**, 628–634
- 35 Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C. and Imberty, A. (2003) Crystal structure of fungal lectin: Six-bladed  $\beta$ -propeller fold and novel recognition mode for *Aleuria aurantia* lectin. *J. Biol. Chem.* **278**, 27059–27067
- 36 Ng, K. K., Drickamer, K. and Weis, W. I. (1996) Structural analysis of monosaccharide recognition by rat liver mannose-binding protein. *J. Biol. Chem.* **271**, 663–674
- 37 Pérez, S., Mouhous-Riou, N., Nifant'ev, N. E., Tsvetkov, Y. E., Bachet, B. and Imberty, A. (1996) Crystal and molecular structure of a histo-blood group antigen involved in cell adhesion: the Lewis x trisaccharide. *Glycobiology* **6**, 537–542
- 38 Henry, S., Oriol, R. and Samuelsson, B. (1995) Lewis histo-blood group system and associated secretory phenotypes. *Vox Sang.* **69**, 166–182
- 39 Newburg, D. S., Ruiz-Palacios, G. M., Altaye, M., Chaturvedi, P., Meinzen-Derr, J., Guerrero, M. L. and Morrow, A. L. (2004) Innate protection conferred by fucosylated oligosaccharides of human milk against diarrhea in breastfed infants. *Glycobiology* **14**, 253–263

Received 12 January 2005/22 March 2005; accepted 24 March 2005

Published as BJ Immediate Publication 24 March 2005, DOI 10.1042/BJ20050079

# 6

## Recherche de candidats à l'inhibition de PA-III: les glycomimétiques

---

### 6.1. Les médicaments glycomimétiques

#### 6.1.1. Quelques exemples

La surface cellulaire est recouverte de structures oligosaccharides qui prennent part à un large éventail de processus de reconnaissance biologique, dont certains jouent un rôle dans des pathologies telles que le cancer ou les infections microbiennes. Tous ces processus sont les cibles potentielles d'interventions thérapeutiques et plusieurs médicaments de type «glycomimétiques» représentent un axe de recherche majeur pour les industries pharmaceutiques et biotechnologiques.

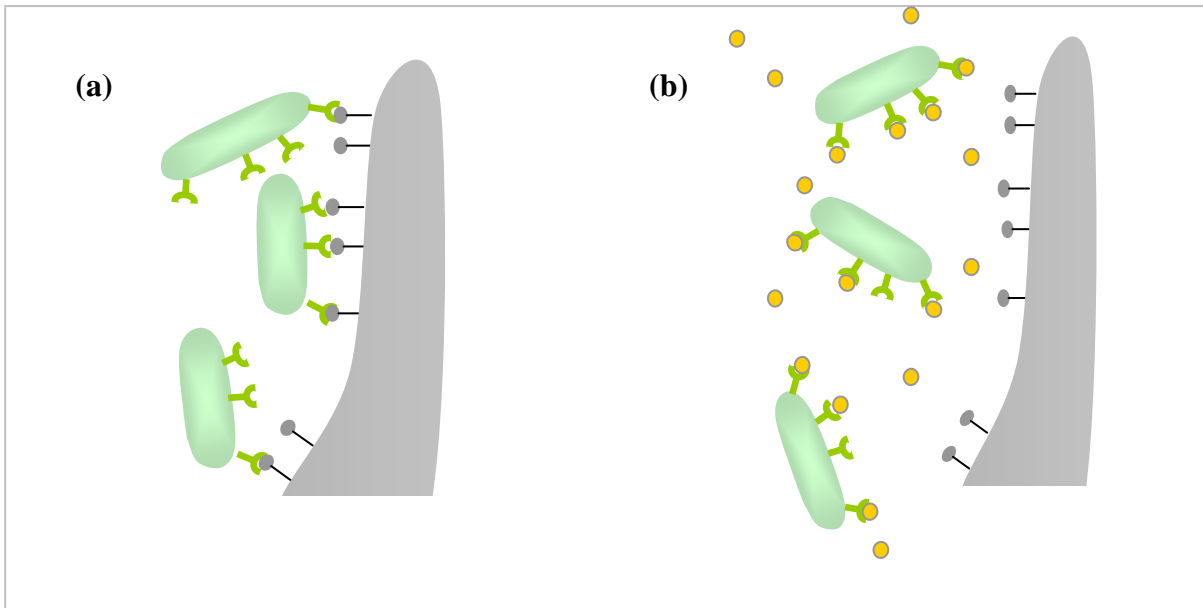
Les applications sont multiples, et incluent par exemple la thérapie contre le cancer. En effet, les cellules tumorales échappent au système immunitaire en présentant des glycolipides et des glycoprotéines spécifiques sur leurs surfaces. Cette découverte a ouvert une voix nouvelle pour la mise au point de vaccins dérivés des motifs glucidiques présents sur ces glycoconjugués (Slovin *et al.*, 1999). Ainsi des tests cliniques de phase I utilisant un vaccin basé sur le globo H, un hexasaccharide complexe, a montré des résultats encourageants contre le cancer du sein. D'autres vaccins développés à partir du tétrasaccharide Lewis Y sont actuellement en cours d'étude contre les cancers des ovaires, du sein et de la prostate (Sabbatini *et al.*, 2000).

Parce qu'ils différencient le soi du non-soi, les glucides jouent un rôle barrière indésirable lors des transfusions sanguines et greffes d'organes. L'utilisation d'organes issus d'animaux tel que le cochon (xénotransplantatino) constituerait un palliatif au don d'organe mais les anticorps humains reconnaissent spécifiquement l'antigène Gal( $\alpha$ 1-3)Gal recouvrant les cellules des endothelia des organes animaux. Une solution consistant à saturer l'anticorps anti  $\alpha$ -Gal avec un pentasaccharide a été explorée (Samuelsson *et al.*, 1994).

Les glucides pourraient aussi être utilisés comme antiviraux puisqu'une étude a démontré que l'activité du virus de l'hépatite B était inhibée en présence de l'imino-sucres N-butyl-déoxynojirimycine (Block *et al.*, 1994). De même, des analogues solubles du galactosylcéramide, un ligand secondaire du récepteur gp120 du HIV-1 potentiellement impliqué dans la fusion membranaire du virus avec la cellule hôte, s'est montré capable d'inhiber l'intrusion du HIV-1 dans les cellules immunitaires CD4 (Fantini *et al.*, 1997).

Enfin et plus proche des préoccupations abordées dans la thèse, les sucres constituent une ressource prometteuse d'antibactériens. De nombreuses infections sont initiées par la fixation des microorganismes sur les glycoconjugués exposés sur la surface des cellules hôtes. Au cours des dernières décennies, la résistance croissante des bactéries aux divers antibiotiques est devenue un problème de plus en plus préoccupant. Plutôt que de tuer les bactéries comme le fait un antibiotique, l'alternative serait d'interférer au moment de l'adhésion cellulaire. La connaissance des structures oligosaccharides impliqués dans l'adhésion bactérienne peut permettre la conception d'analogues solubles capables d'entrer en compétition avec les glycoconjugués de surface. En inhibant l'adhésion, on s'attaque aux tout premiers stades de l'infection (fig. 6.1).

Malgré la structure complexe des branchements et des liaisons glycosidiques qui rendent difficile l'analyse et la production d'oligosaccharides, ces dernières années ont vu des progrès significatifs dans la conception et la synthèse de molécules thérapeutiques dérivées des sucres.



**Figure 6.1.** Principe imagé de l'adhésion bactérienne en absence (a) et en présence (b) d'analogues solubles des glycoconjugués des surfaces épithéliales

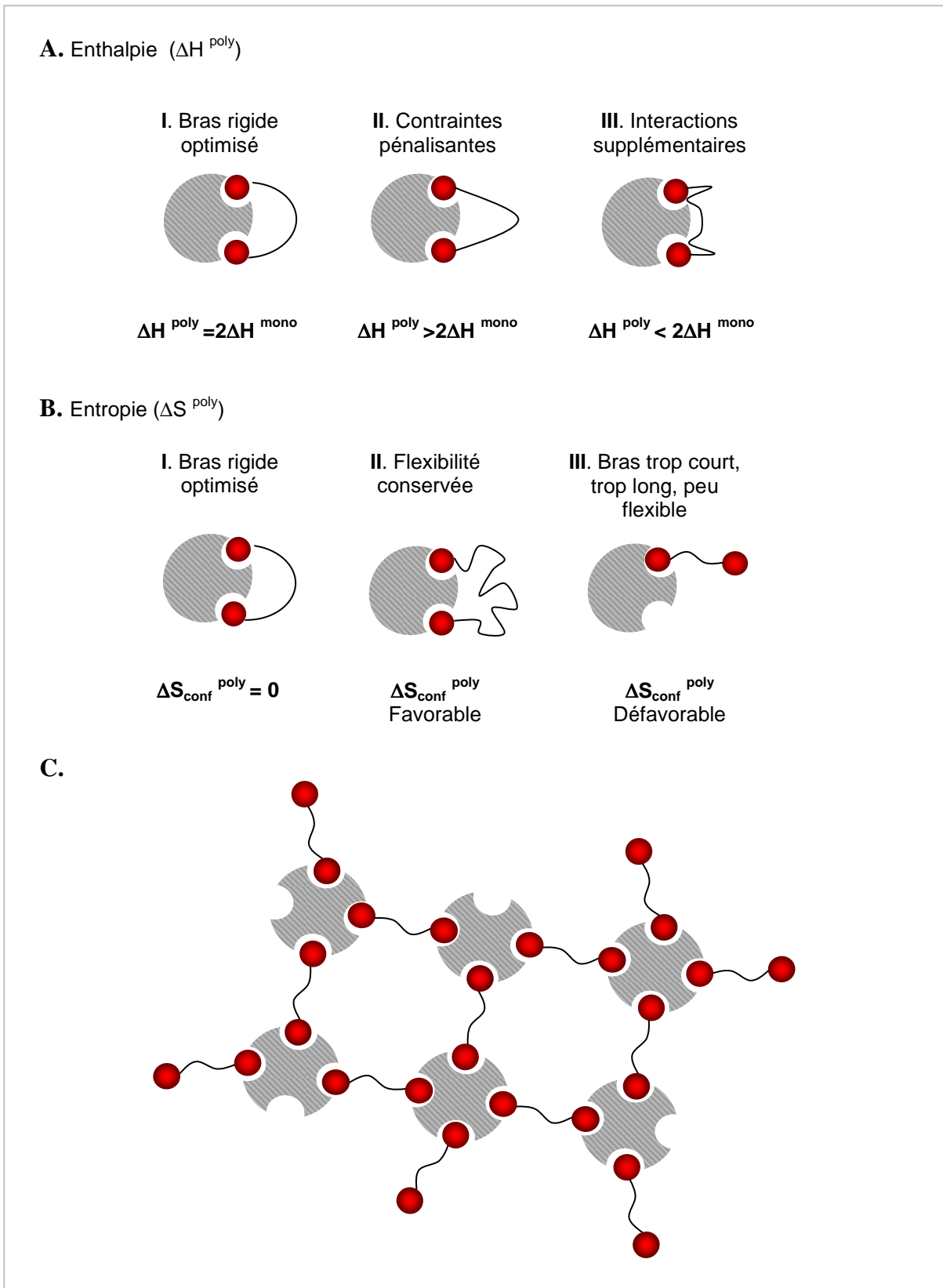
### 6.1.2. L'essor des systèmes multivalents

Une forte affinité est requise pour l'application pratique d'une stratégie de compétition. Or, l'affinité relativement faible (de l'ordre du milimolaire) souvent rencontrée dans les interactions entre monosaccharide et lectines représente l'obstacle majeur qui accompagne la conception de molécules actives basées sur ce type de reconnaissance. Des études ont démontré que la multivalence était une approche efficace pour augmenter de manière significative la force d'interaction des molécules de synthèse.

Différents systèmes modèles ont été développés. Certains ont un faible niveau de valence tels que les glycoclusters qui comportent généralement de 2 à 5 ligands dont l'espacement et les orientations géométriques sont particulièrement étudiés pour s'adapter au mieux à une configuration multimérique de la protéine étudiée. Ainsi, des glycoclusters en forme d'étoile sont des inhibiteurs efficaces de la toxine du cholera (protéine pentamérique), apportant une amélioration significative de l'affinité (Fan *et al.*, 2004). D'autres systèmes de valence plus haute tels que les dendrimères et dendrons (petits dendrimères) sont des polymères arborescents composés d'une structure noyau sur laquelle sont branchés plusieurs groupes fonctionnalisés répétés à la manière d'une grappe de raisin.

Selon la longueur du bras espaceur, deux phénomènes différents peuvent se produire. Soit le ligand s'attache à plusieurs sites de la même protéine (chélation), et on observe un effet coopératif dû à la multivalence ; soit le ligand se lie à plusieurs protéines et on observe la formation d'un réseau tridimensionnel qui précipite en solution. Cette situation est fréquente puisque les lectines sont très souvent multimériques et bon nombre d'entre elles présentent des sites de reconnaissance relativement éloignés dans l'espace.

La chélation est favorisée si l'énergie libre globale des  $N$  interactions ( $\Delta G^{\text{poly}}$ ) est plus favorable que l'énergie libre de ces  $N$  interactions monovalentes ( $N\Delta G^{\text{mono}}$ ). Les principaux systèmes d'interactions possibles sont illustrés dans la figure 6.2 pour un système divalent mais sont extrapolables à un système de  $N$  résidus. L'enthalpie d'interaction reflète la stabilisation du complexe ligand-récepteur. Après la formation du premier contact, l'enthalpie d'interaction du second résidu peut être égale, inférieure ou supérieure au premier. Si l'alignement résidu-récepteur est parfait, son enthalpie est identique ( $\Delta H=2\Delta H^{\text{mono}}$ ). Ce cas de figure est rare puisqu'il est difficile de concevoir des bras espaceurs positionnant de façon optimale les ligands dans les diverses poches réceptrices. Bien souvent, la seconde liaison engendre des contraintes sur les bras qui ajoute une pénalité énergétique à l'enthalpie globale de l'interaction. Parfois des interactions supplémentaires entre le bras espaceur et la protéine viennent ajouter de la stabilité et augmenter l'enthalpie de la seconde interaction. L'entropie reflète le degré de désordre intervenant lors de l'interaction et résulte de diverses contributions : rotation, translation, conformation, hydratation. Etant donné que la plupart des bras possèdent des degrés de liberté, la chélation n'est possible que si ce coût est faible. Dans le cas contraire (bras trop courts, trop longs ou peu flexibles, la chélation ne se fait pas.



**Figure 6.2.** (A) enthalpie, (B) entropie lors de la chélation d'un ligand divalent. (C) Agrégation d'une lectine avec un ligand divalent.

## 6.2. Interactions entre PA-III et des séries de glycomimétiques: Approche expérimentale

### 6.2.1. Synthèse des inhibiteurs

Plus de 60 molécules de synthèse glycomimétiques ont été synthétisées par le Dr René Roy et son équipe (Cathy Préville, Sarah Beha, Karine Marotte) au département de chimie et biochimie de l'Université du Québec à Montréal (UQAM). La liste exhaustive de ces composés est reportée en annexe 2 de la thèse. A l'exception du composé **64**, tous les inhibiteurs comprennent un fucose comme élément de base sur lequel différents groupes substituants ont été ajoutés.

### 6.2.2. Mise au point des tests ELLA

Les tests ELLA ont été réalisés en plusieurs fois, au fur et à mesure que nous parvenions les molécules glycomimétiques. Pour des raisons pratiques les premiers tests ELLA furent réalisés en immobilisant la lectine au fond des puits (interactions non spécifiques) et en mettant en solution pour la compétition les différents inhibiteurs avec un Fuc-PAA biotinylé contenant 20% de fucose. Les résultats obtenus au fur et à mesure des tests ont permis aux chimistes de décider de l'orientation à suivre pour les synthèses ultérieures. Dans les premiers temps, seuls des monomères étaient testés, mais lorsque les chimistes ont entrepris la synthèse de molécules dendrimères, ces tests ELLA sont apparus mal adaptés. Les résultats de ces premiers tests ELLA ne sont pas montrés dans la thèse.

Une nouvelle procédure a été mise au point en inversant le sens des tests ELLA. Pour ce faire, il a fallu fixer le ligand polymérique au fond des puits, et marquer la protéine en la biotinylant. On peut ainsi évaluer le potentiel inhibiteur des glycomimétiques contre l'adhésion de la lectine PA-III à une surface de glycoconjugués. Tous les résultats ELLA donnés dans ce chapitre ont été obtenus par cette méthode.

### 6.2.3. *Protocole des tests ELLA utilisant la lectine PA-III libre en solution*

La biotinylation de la lectine PA-III a été réalisée d'après un protocole de la littérature (Bayer and Wilchek, 1990). 200µM de PA-III en solution dans 0,1M de NaHCO<sub>3</sub> et 0,2M de NaCl sont mélangés à une solution de diméthyle formamide (DMF) contenant 4,5mM de biotine durant 2 heures à température ambiante sous agitation. La solution est ensuite dialysée une première fois contre 0,15 M de NaCl, puis contre de l'eau et enfin lyophilisée

Les tests ELLA ont été réalisés sur des plaques de polystyrène de 96 puits (NuncMaxisorb) recouverte avec 10 µg/ml de PAA-Fuc (poly[N-(2-hydroxyethyl)acrylamide] contenant 20% de fucose en ramification (Lectinity Holding, Inc) pendant 1heure à 37°C. Après rinçage, chaque puits est ensuite rempli avec 100 µl d'une solution de PBS contenant 3% de BSA pendant 1 heure a 37°C. Les plaques sont ensuite incubées 1h à 37 °C en présence de PA-III biotynilée et une série de dilutions en inhibiteurs. Après deux rinçages de 3 minutes au PBS-T (solution de PBS contenant 0.05% de Tween), 100 µl de steptavidine-peroxidase (dilution 1:5000; Boehringer-Mannheim) sont ajoutés dans chaque puits et incubés 1h a 37°C. la couleur est alors développée en utilisant 100µl par puits de peroxyde d'hydrogène (0.4mg/ml) (Sigma-Aldrich). La réaction est stoppée par l'ajout de 50 µl de H<sub>2</sub>SO<sub>4</sub> à 30%. L'absorbance est lue à 490 nm à l'aide d'un lecteur de plaques de microtitration (Bio-Rad; model 680).

### 6.2.4. *Etudes cristallographiques*

Pour quatre inhibiteurs glycomimétiques, il a été possible d'obtenir des cristaux du complexe avec la lectine PA-III. Les complexes ont été obtenus par trempage (24 heures dans une solution de 10mM en inhibiteur) de cristaux préalablement co-cristallisés avec du L-fucose dans les conditions suivantes : 1,75 M de sulfate d'ammonium, 0,1 M Tris-HCl, pH 8,5. Le remplacement moléculaire a été réalisé avec le logiciel Molrep (Vagin and Teplyakov, 1997), et l'analyse visuelle de la structure s'est faite à l'aide du logiciel graphique Coot (Emsley and Cowtan, 2004). Les données de la collecte et de l'affinement figurent dans le tableau 6.1.



inhibiteurs	13	15	50	10
<b>Collecte de données</b>				
Ligne de lumière (ESRF)	ID14-EH2	ID14-EH2	ID14-EH2	ID14-EH2
Longueur d'onde (Å)	0.933	0.933	0.933	0.933
Résolution (Å)	1.10-22.26	1.10-22.30	1.30-27.28	1.70-44.11
Zone haute résolution (Å)	1.10-1.30	1.10-1.30	1.30-1.37	1.70-1.79
Groupe d'espace	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>	P2 <sub>1</sub>
Maille	a= 52.55 Å; b= 72.83 Å; c= 54.37 Å; β=94.28°	a= 52.63 Å; b= 73.09 Å; c= 54.50 Å β= 94.36°	a= 52.55 Å; b= 72.67 Å; c= 54.72 Å; β= 94.41°	a= 53.08 Å; b= 73.18 Å; c= 55.36 Å β= 93.86°
Coef de Matthews (Å <sup>3</sup> /DA)	2.21	2.23	2.22	2.29
Solvant (%)	44.4	44.8	44.6	46.2
Reflexions mesurées	827162	834551	368957	168445
Reflexions uniques	158737	165425	100063	46062
Multiplicité moyenne <sup>1</sup>	5.2 (3.6)	5.0 (3.5)	3.7 (3.6)	3.70 (3.8)
Complétude <sup>1</sup> (%)	96.3 (92.7)	99.4 (97.6)	99.4 (99.5)	99.1 (98.5)
Moyenne <sup>1</sup> I/σI	22.6 (9.4)	18.7 (6.5)	18.6 (9.5)	7.3 (2.2)
R merge <sup>1,2</sup> (%)	0.050 (0.12)	0.054 (0.18)	0.045 (0.11)	0.161(0.49)
Facteur B de Wilson (Å <sup>2</sup> )	9.89	9.61	9.23	14.68
<b>Affinement</b>				
R <sub>cryst</sub> <sup>3</sup> (observation)	0.126 (153917)	0.130 (167060)	0.133 (95040)	0.191 (43714)
R <sub>libre</sub> <sup>4</sup> (observation)	0.145 (4790)	0.150 (9319)	0.152 (4995)	0.210 (2326)
Nombre d'atomes:				
Hetero/ Proteine/ Solvent	117/ 3527/ 588	123/ 3525/ 548	92/ 3451/ 735	89/ 3390/ 608
Facteur B moyen				
Chaines : A/B/C/D/Eau	6.25/ 6.97/ 6.88/ 8.00/ 26.63	6.84/ 7.33/ 7.26/ 8.10/ 22.6	6.30/ 6.40/ 6.37/ 6.90/ 22.67	9.13/ 9.69/ 9.40/ 10.85/ 30.20
Déviation RMS/ideal				
liaisons (Å)	0.016	0.017	0.012	0.015
Angles (°)	1.487	1.540	1.56	1.28

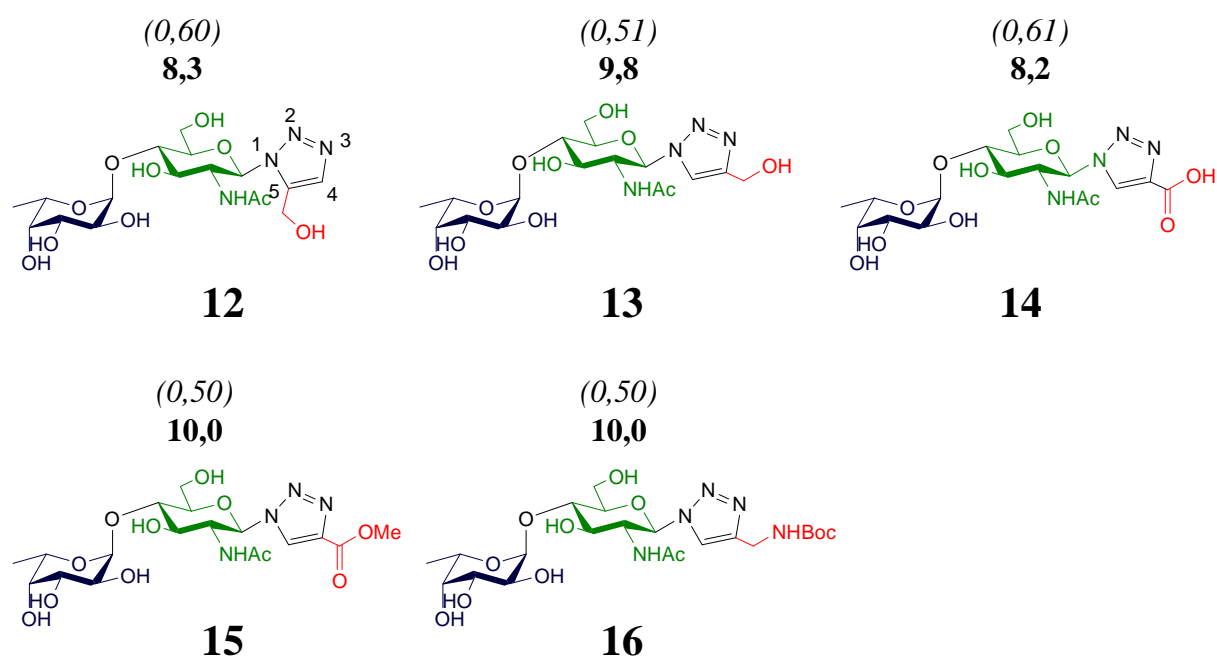
<sup>1</sup>Nombre entre parenthèses font référence à la zone de plus haute résolution. <sup>2</sup> $R_{\text{merge}} = \frac{\sum |I - \langle I \rangle|}{\sum I}$ , où I = intensité observée. <sup>3</sup> $R_{\text{cryst}} = \frac{\sum \|F_o - F_c\|}{\sum |F_o|}$ , où |F<sub>o</sub>| = amplitude des facteurs de structure observés et |F<sub>c</sub>| = amplitude des facteurs de structure calculés. <sup>4</sup>R<sub>libre</sub> correspond au R<sub>cryst</sub> pour 3 % des réflexions exclues de l'affinement.

**Tableau 6.1.** Données statistiques de la collecte et de l'affinement des structures cristallines

### 6.3. Interactions entre PA-IIL et des séries de glycomimétiques : Comparaison des glycomimétiques et affinité

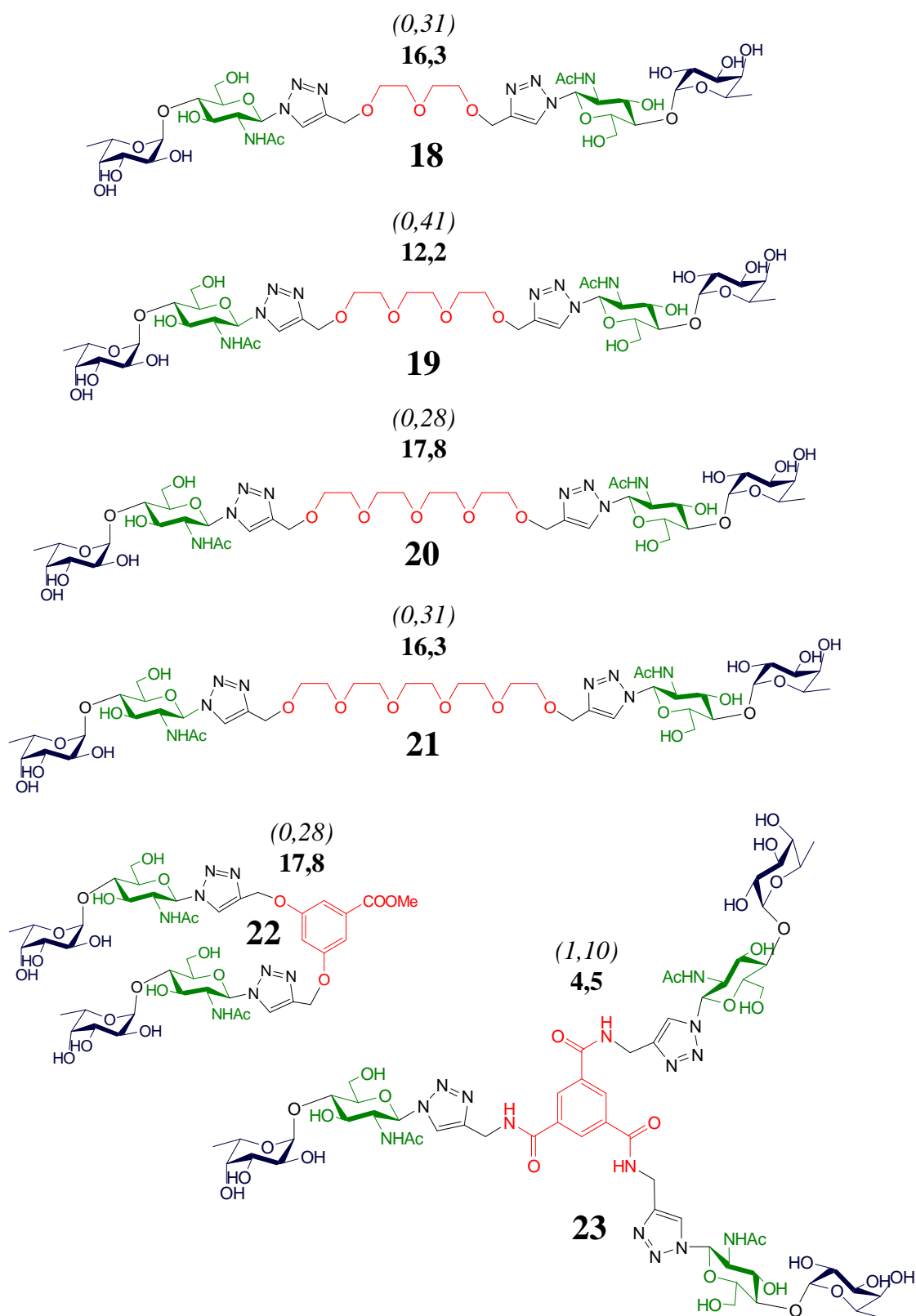
### 6.3.1. Les dérivés du disaccharide L-Fuc( $\alpha$ 1-4)GlcNac

Les tests ELLA effectués antérieurement ont montré que le trisaccharide Lewis a été le meilleur ligand naturel de la lectine PA-IIL (Perret *et al.*, 2005). Les données structurales suggèrent que le disaccharide L-Fuc( $\alpha$ 1-4)GlcNac du Lewis a participé à l'essentiel des interactions avec la protéine et peut constituer le motif commun d'une série d'inhibiteurs à tester en ELLA.



**Figure 6.3.** Molécules inhibitrices dérivées du Fuc( $\alpha$ 1-4)GlcNac. Les données entre parenthèses et en gras font respectivement référence aux valeurs de IC<sub>50</sub> ( $\mu$ M) (concentration nécessaire pour inhiber 50 % de l'interaction) et au potentiel d'inhibition de la molécule par rapport au L-fucose. (Série 1)

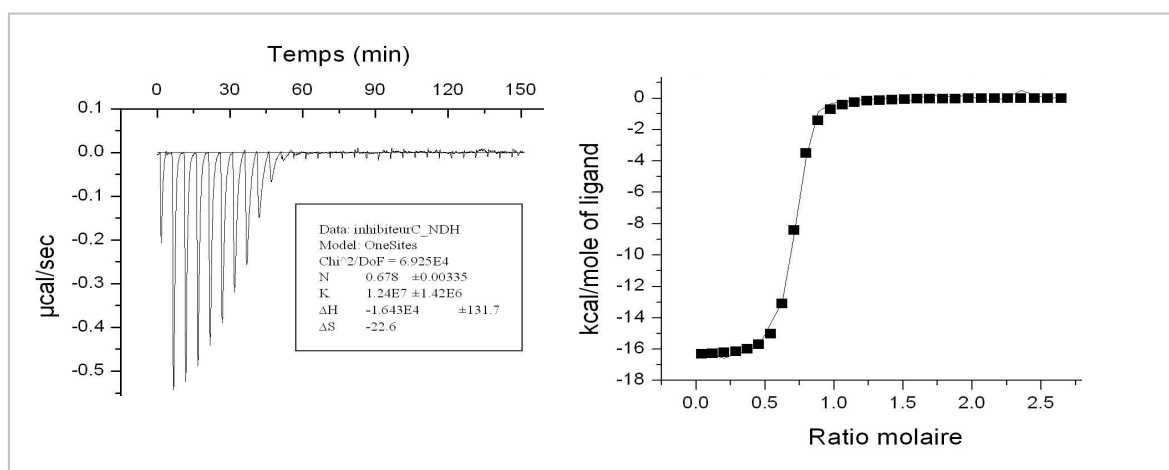
Une première série contient les inhibiteurs possédant comme structure de base commune, le disaccharide L-Fuc( $\alpha$ 1-4)GlcNac, ainsi qu'un triazole lié par l'azote N1 en position C1 du GlcNac (fig. 6.3). Toutes ces molécules ont effectivement montré un IC<sub>50</sub> voisin de celui du Lewis a, et certaines molécules montrent même un potentiel d'inhibition (PI) légèrement supérieur.



**Figure 6.4.** Dendrimères de la molécule 13. Les données entre parenthèses et en gras font respectivement référence aux valeurs de IC<sub>50</sub> (μM) (concentration nécessaire pour inhiber 50 % de l'interaction) et au potentiel d'inhibition de la molécule par rapport au L-fucose durant les tests ELLA. (Série2)

Dans un deuxième lot, plusieurs dendrimères de la molécule **13** ont été développés par les chimistes en variant la nature et la longueur des bras espaceurs (fig. 6.4). Plusieurs dimères liés par une chaîne poly-éthylène-glycol (PEG) de longueur variable ainsi qu'un dimère et un trimère dont les motifs sont liés par un groupement phényle ont été testés par ELLA. Les dimères constituent les meilleurs inhibiteurs testés jusqu'à présent. Tous partagent un potentiel d'inhibition d'un même ordre de grandeur, allant de 12,2 pour le dimère **18** à 17,8 pour les dimères **20** et **22**. Le trimère quant à lui, possède un PI deux fois inférieur à la forme monomérique puisque son IC50 n'est que de 1,10  $\mu\text{M}$  pour une valeur de 0,51  $\mu\text{M}$  pour la molécule **13**.

La microcalorimétrie de titration (ITC) a été utilisée pour mieux caractériser les interactions des molécules les plus intéressantes des deux groupes (monomères, dimères et trimère) avec la PA-IIL. Les molécules multivalentes posent des problèmes particuliers dus à la précipitation de la lectine dans la cellule. Ceci cause des problèmes dans la ligne de base après le point de titration et certaines expériences sont difficiles à interpréter mais il est cependant possible de mesurer l'enthalpie de l'interaction et la constante d'association pour plusieurs composés (fig. 6.5). Les données thermodynamiques des inhibiteurs disaccharidiques figurent dans le tableau 6.2.



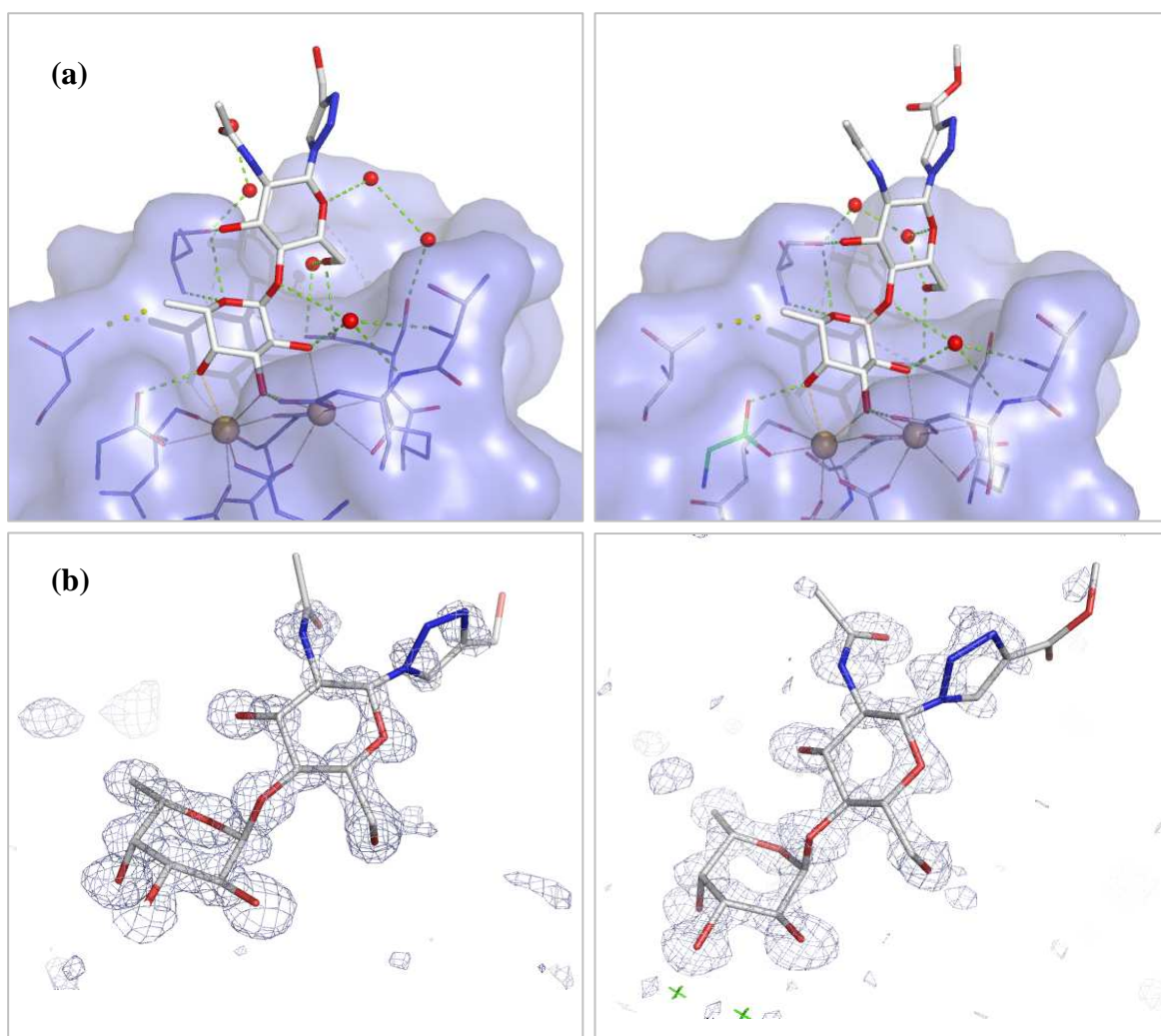
**Figure 6.5.** (a) Courbes de titration ITC obtenue pour le dendrimer **21**. (b) Intégration et ajustement au modèle théorique

	Ka 10 <sup>6</sup> M <sup>-1</sup>	Kd (μM)	n	-ΔG (kJ/mol)	-ΔH (kJ/mol)	-TΔS (kJ/mol)
<b>Monomères</b>						
*Le <sup>a</sup>	4.70 (±0.20)	0.21 (±0.02)	1.08 (±0.01)	38.06 (±0.08)	34.95 (±0.07)	-3.11 (±0.01)
<b>13</b>	3.17 (±0.26)	0.32 (±0.03)	0.98 (±0.01)	37.10 (±0.20)	43.03 (±0.14)	5.92 (±0.16)
# <b>15</b>	3.39 (±0.00)	0.29 (±0.00)	0.78 (±0.00)	37.28 (±0.00)	36.60 (±0.10)	-0.62 (±0.00)
<b>Dimères</b>						
<b>22</b>	5.99 (± 0.32)	0.17 (± 0.01)	0.62 (± 0.01)	38.68 (± 0.13)	73.98 (± 1.93)	35.3 (± 2.06)
<b>21</b>	11.04 (±0.14)	0.09 (± 0.01)	0.66 (±0.02)	40.19 (±0.30)	69.65 (±0.86)	29.46 (±1.16)
<b>Trimère</b>						
<b>23</b>	7.27 (±0.46)	0.14 (± 0.01)	1.13 (±0.12)	39.16 (±0.16)	37.92 (±1.05)	-1.24 (±0.89)

**Tableau 6.2.** Données thermodynamiques obtenues par ITC à 25°C, des interactions de la lectine PA-IIL avec les inhibiteurs basés sur le L-Fuc(α1-4)GlcNac (Les déviations standards ont été calculées à partir de trois expériences indépendantes).\*(Perret, Sabin et al., 2005), #Expérience unique,

L'affinité de PA-IIL pour les inhibiteurs monomériques **13** et **15**, est du même ordre de grandeur que celle obtenue pour le trisaccharide lewis a. L'interaction reste dans les deux cas principalement dirigée par l'enthalpie. On observe dans le cas du composé **13**, une contribution entropique défavorable non négligeable de 5,92 μM comparée aux contributions entropiques faibles ou négligeables observées pour le Le<sup>a</sup> et le composé **15**.

Pour les complexes PA-IIL/dimères, les grandeurs thermodynamiques se répartissent de façon singulière puisqu'on observe dans les deux cas, une grande contribution entropique défavorable de 30 kJ/mol environ, compensée par une très forte contribution favorable de l'enthalpie d'une valeur moyenne de 70 kJ/mol. L'affinité qui en résulte est de deux à trois fois plus forte que celle observée pour le monomère **13**. Le complexe PA-IIL/trimère, ne présente pas une affinité forte puisqu'elle n'est que de 7,27 10<sup>6</sup> M<sup>-1</sup>, soit deux fois plus élevée que celle observée dans le cas du monomère, et la répartition de grandeurs thermodynamiques est sensiblement équivalente. Il est important de noter que la formation d'agrégats dans la cellule de l'appareil ITC a été observée en fin de réaction, attestant des limites de la méthode dans l'étude de ce genre d'interaction.



**Figure 6.6.** Structure cristallines des inhibiteurs **13** (à gauche) et **15** (à droite) dans le site de reconnaissance. **(a)** Site de reconnaissance, les liaisons hydrogène sont en vert pointillé, les liaisons de coordination en orange, les molécules d'eau sont représentées comme des sphères rouges, les ions calcium comme des sphères orange, la surface de la protéine est en Bleu pétrole transparent. **(b)** Cartes de densité électronique  $f_o - f_c$  (contour à 3.0 sigma, en bleu) avant la mise en place du ligand dans le site de reconnaissance.

Les contacts entre la partie disaccharidique et la protéine (fig. 6.6) sont très semblables à ceux observés dans la structure du complexe PA-III/Le<sup>a</sup> (Perret *et al.*, 2005). En particulier, la liaison hydrogène établie par l'hydroxyle secondaire O6 du GlcNAc avec Asp96 est conservée dans les différentes structures. Le groupement triazole n'apporte pas de contact supplémentaire ce qui explique que l'affinité ne soit pas meilleure que celle observée pour le Lewis a.

### 6.3.2. Les dérivés du fucose

Plusieurs autres séries d'inhibiteurs basés sur le monosaccharide fucose modifié en position C1, O1 ou C5 ont aussi été testées (fig. 6.7), (fig. 6.8), (fig. 6.9), (fig. 6.10). Les valeurs signalées font respectivement référence aux valeurs de IC<sub>50</sub> en  $\mu\text{M}$  (entre parenthèses) et au potentiel d'inhibition de la molécule par rapport au L-fucose durant les tests ELLA.

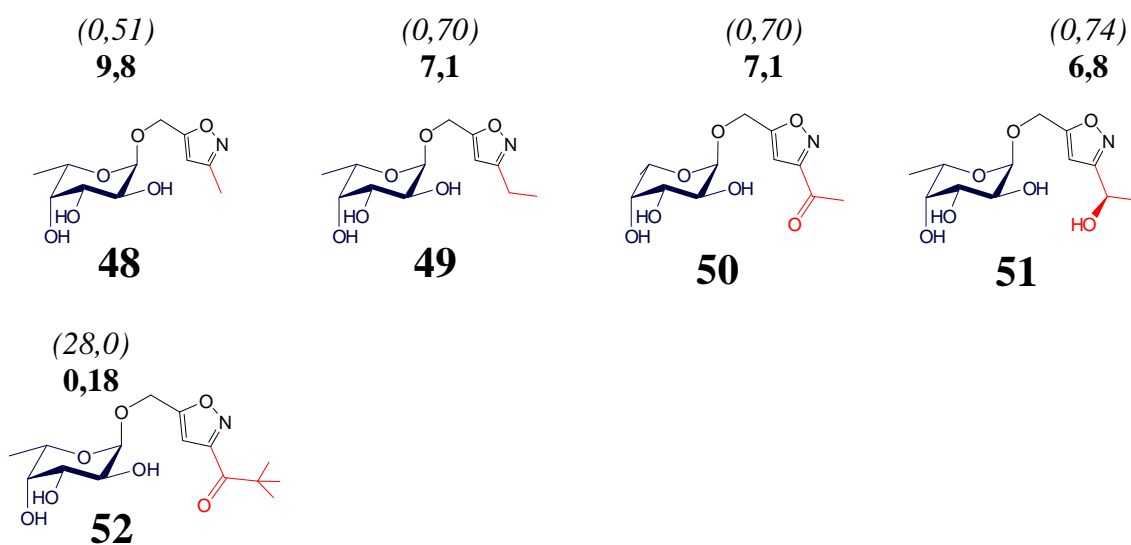


Figure 6.7. Inhibiteurs : groupement isoxazole-(R) en position O1 du fucose (Serie3).

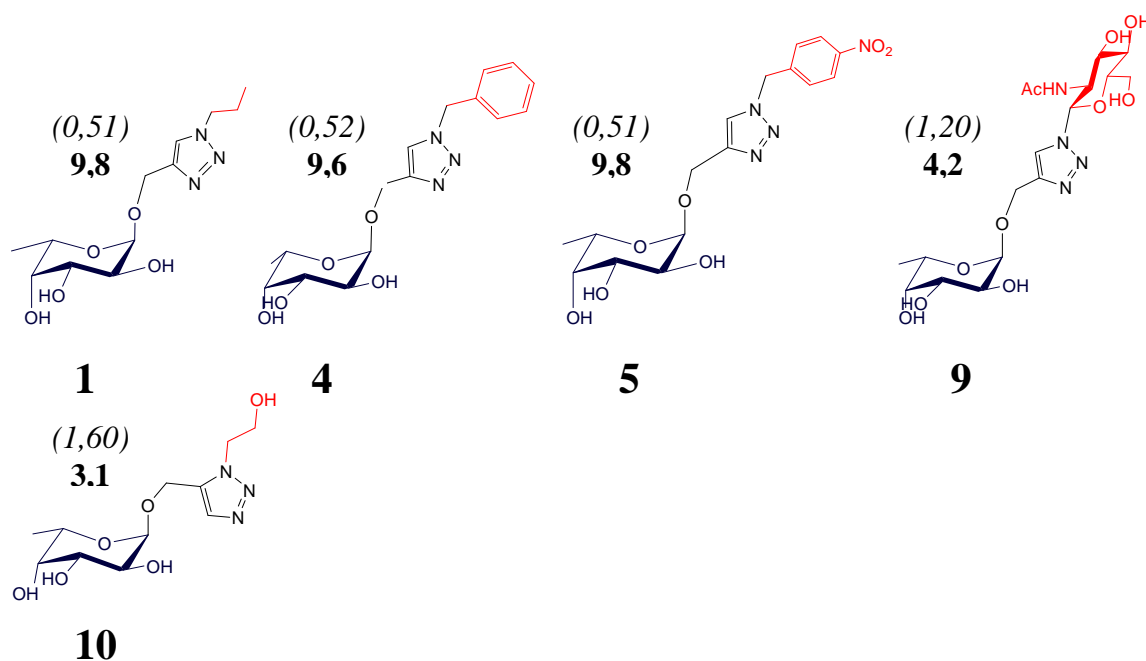
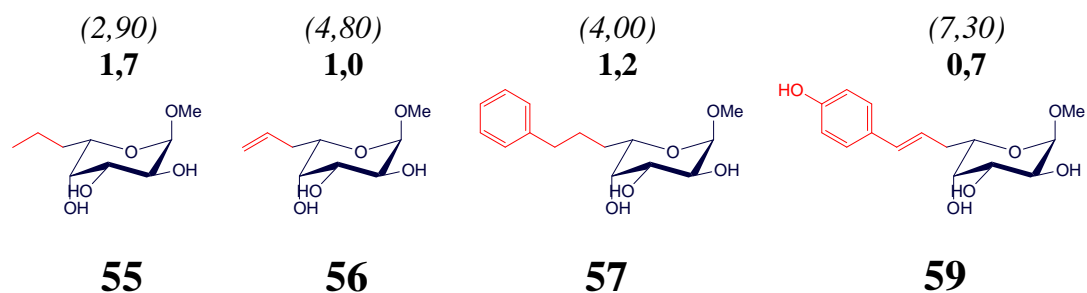
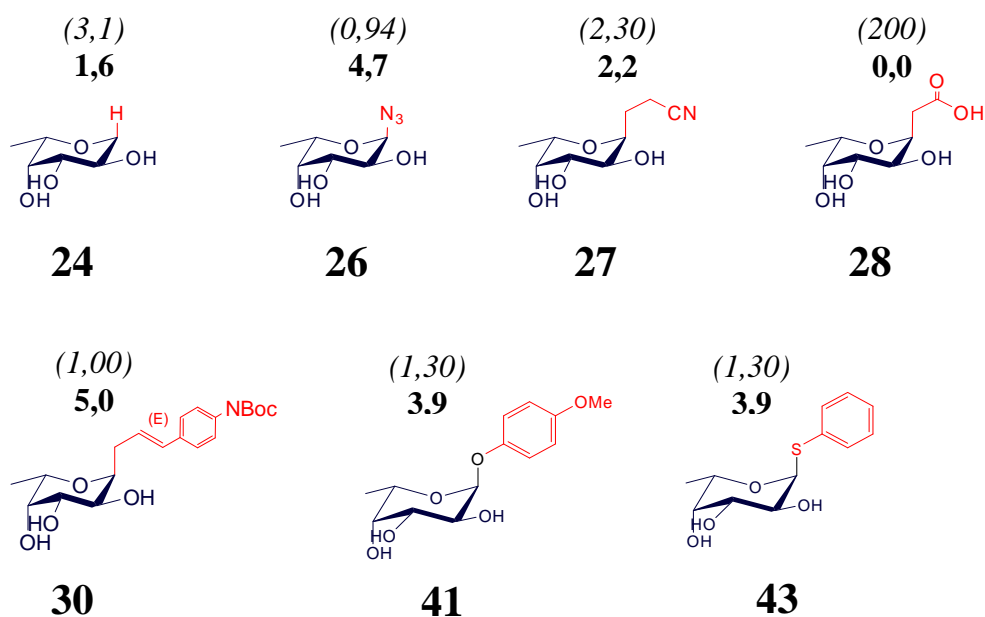


Figure 6.8. Inhibiteurs : groupement triazole-(R) en position O1 du fucose (Série 4).



**Figure 6.9.** Inhibiteurs: groupement hydrophobe en position C5 du fucose (Série 5).



**Figure 6.10.** Inhibiteurs : groupement en position C1 du fucose (Série 6).

Il apparaît clairement que seuls les inhibiteurs portant un hétérocycle en position anomérique du fucose peuvent avoir un PI d'environ 10, c'est-à-dire du même ordre de grandeur que celui du Lewis a. Ces hétérocycles azotés ont un caractère amphiphiles car ils présentent à la fois

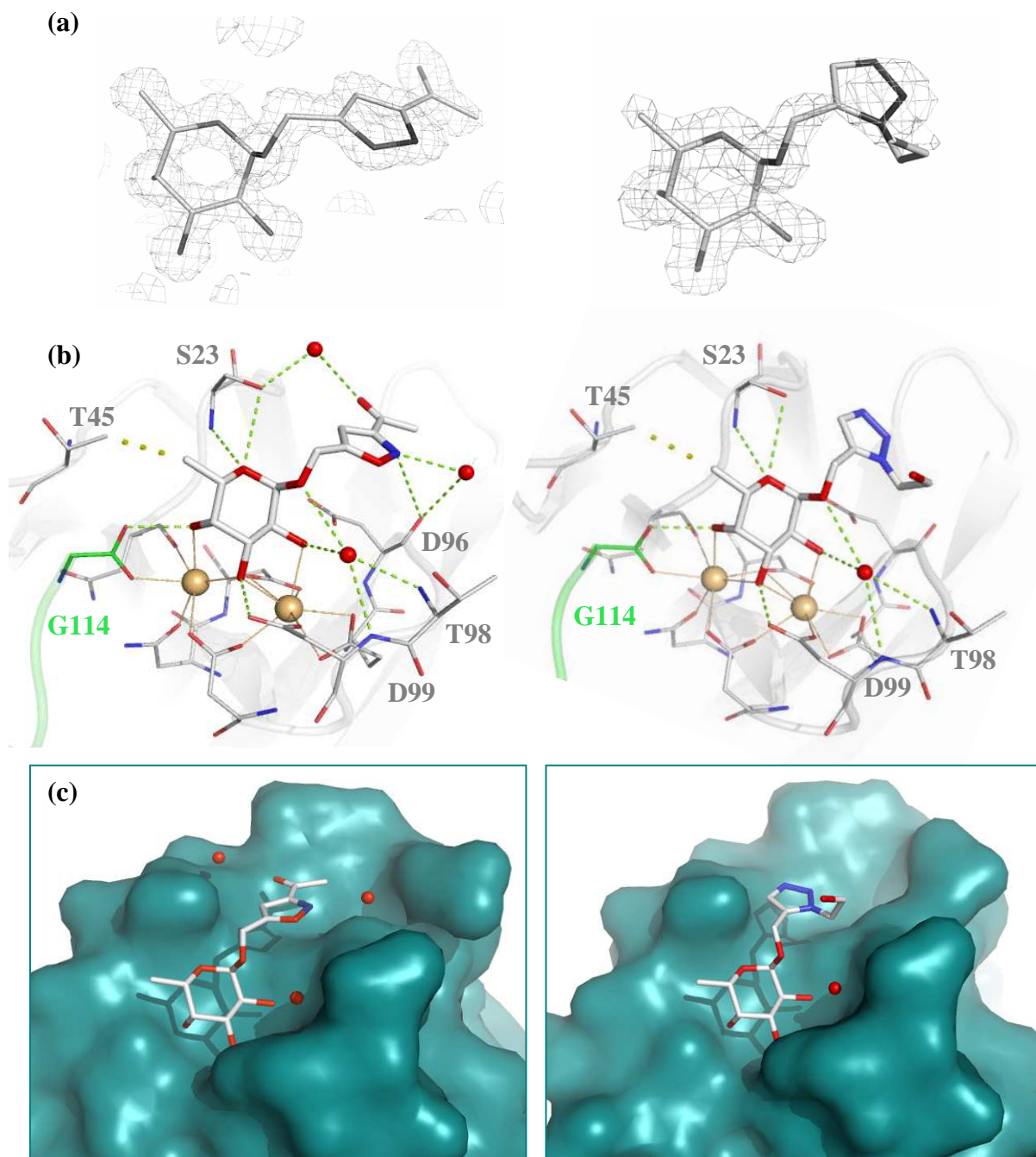


des carbones hydrophobes et des N ou O pouvant accepter des liaisons hydrogène. Ils semblent donc pouvoir mimer le rôle que joue le GlcNAc dans le Lewis a.

	Ka 10 <sup>6</sup> M <sup>-1</sup>	Kd (μM)	n	-ΔG (kJ/mol)	-ΔH (kJ/mol)	-TΔS (kJ/mol)
<b>45</b>	0.24 (±0.29)	4.25 (±0.56)	0.79 (±0.26)	30.70 (±0.30)	24.66 (±0.91)	-6.04 (±1.20)
<b>43</b>	1.85 (±0.17)	0.55 (±0.05)	1.05 (±0.01)	35.76 (±0.22)	39.19 (±5.40)	3.43 (±5.40)
<b>50</b>	1.87 (± 0.52)	0.58 (±0.16)	1.15 (± 0.33)	35.71 (± 0.70)	41.50 (± 0.40)	5.8 (± 1.70)
<b>1</b>	2.41 (±0.03)	0.42 (±0.01)	1.20 (±0.04)	36.42 (±0.03)	35.59 (±0.50)	-0.83 (±0.51)
<b>48</b>	2.97 (±0.28)	0.34 (±0.03)	1.18 (±0.01)	36.94 (±0.24)	36.21 (±0.31)	-0.73 (±0.32)
<b>4</b>	3.05 (±0.27)	0.33 (±0.03)	0.95 (±0.04)	37.00 (±0.06)	44.09 (±0.67)	7.08 (±0.72)
<b>5</b>	3.38 (±0.01)	0.30 (±0.01)	1.03 (±0.01)	32.27 (±0.01)	41.83 (±0.01)	4.57 (±0.01)

**Tableau 6.3.** Données thermodynamiques obtenues par ITC à 25°C, des interactions de la lectine PA-IIL avec les inhibiteurs fucose modifiés (les déviations standards ont été calculées à partir de trois expériences indépendantes).

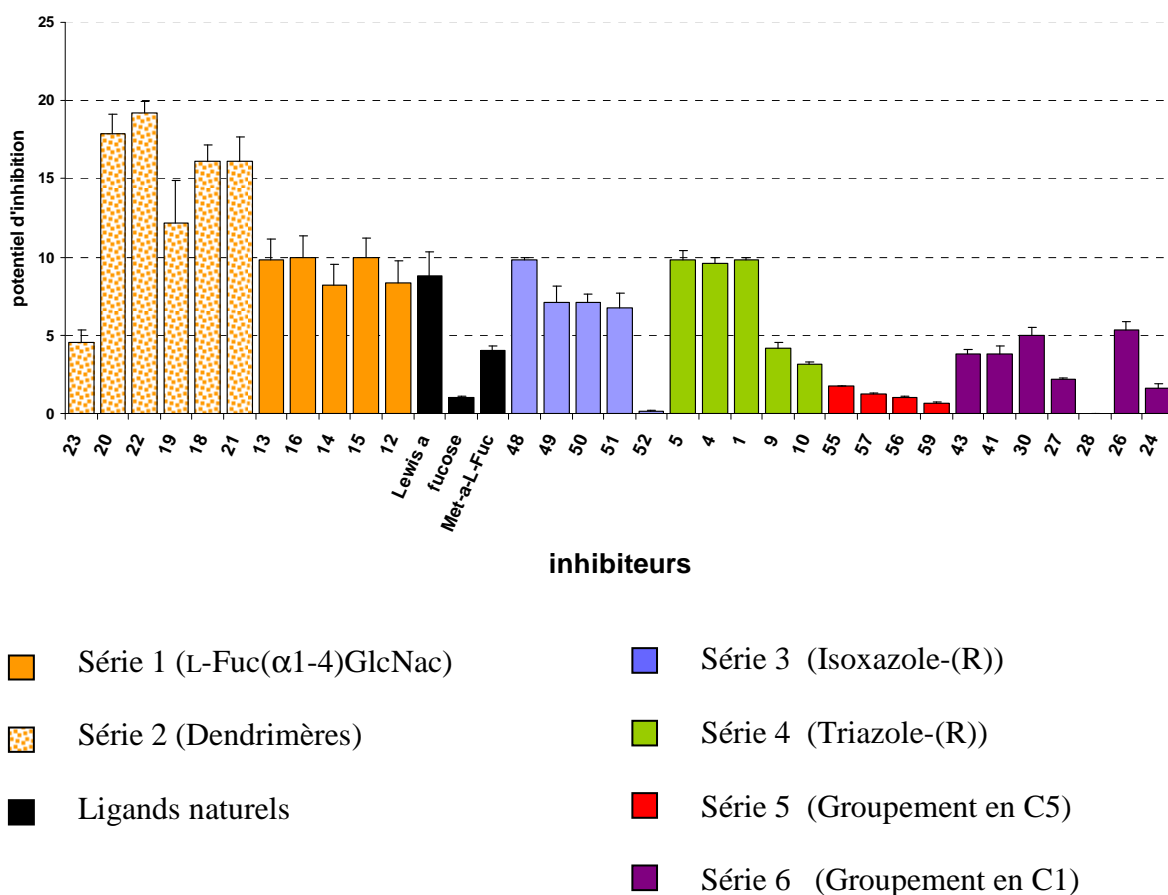
Les expériences ITC (tab. 6.3) confirment que les produits **4**, **5** et **48** présentent une constante d'association pour le PA-IIL d'environ 3 10<sup>6</sup> M<sup>-1</sup>, proche de celle mesurée pour le Lewis a (4 10<sup>6</sup> M<sup>-1</sup>). Les composés **4** et **5** contiennent un groupement aromatique supplémentaire qui semble favoriser l'enthalpie d'interaction. Par contre, les liaisons additionnelles résultent en une entropie légèrement défavorable.



**Figure 6.11.** Inhibiteurs **50** (à gauche) et **10** (à droite) dans le site de reconnaissance. **(a)** Cartes de densité électronique  $2fo-fc$ , (contour à 1.0 sigma). **(b)** Site de reconnaissance, (liaisons hydrogène en vert pointillé, liaisons de coordination en orange, les molécules d'eau sont représentées comme des sphères rouges, les ions calcium comme des sphères oranges, les liaisons hydrophobes en jaune pointillé). **(c)** Mise en évidence de la complémentarité de la surface de la protéine (en Bleu pétrole) avec l'inhibiteur.

Des cristaux de complexes de PA-III<sub>L</sub> avec les composés **50** et **10** ont pu être obtenus et les structures ont été affinées (fig. 6.11). Dans les deux cas, l'hétérocycle vient se placer dans une poche légèrement hydrophobe de la protéine. Grâce à leur taille réduite, les cycles à 5 s'adaptent mieux dans cette poche que ne le fait le GlcNAc du Lewis a. Cependant contrairement au composé **10** dont le groupement triazole ne fait aucune interaction avec la protéine, le composé **50** fait deux nouvelles interactions avec la protéine, la première interaction est indirecte, et met en jeu un pont hydrogène entre le groupe C(O)CH<sub>3</sub> du groupement isoxazole et la Ser23 via une molécule d'eau intermédiaire. La deuxième interaction est directe, et met en jeu l'azote du cycle a 5 de l'isoxazole et l'oxygène de la chaîne principale de l'Asp96.

## 6.4. Discussion et conclusion



**Figure 6.12.** Schéma récapitulatif du potentiel d'inhibition des meilleurs glycomimétiques.

Parmi les inhibiteurs monomériques, ces premiers résultats confirment que le seul motif disaccharide L-Fuc( $\alpha$ 1-4)GlcNac suffit pour égaler l'affinité du Lewis a. Les données structurales confirment ce résultat en révélant un nombre important de liaisons hydrogène effectuées par le disaccharide alors que le groupement triazole ajouté sur le GlcNac ne fait aucune interaction directe ou indirecte avec la protéine.

Parmi les inhibiteurs basés sur un fucose modifié, peu d'amélioration est possible lorsque l'on change de groupement hydrophobe en position C5 du fucose, et ce sans doute à cause de l'encombrement stérique qu'occasionnent ces groupements dans le site de liaison. Les meilleurs inhibiteurs sont les résidus fucose modifiés en position C1. Les séries 3 et 4 comptent parmi les meilleures familles d'inhibiteurs testées. Les composés **48** (serie3) et **5, 4, 1** (série 4) atteignent des potentiels d'inhibition supérieurs au Lewis a durant les tests ELLA (fig. 6.12). Les données thermodynamiques des composés **1, 4, 5, 48** montrent que ces inhibiteurs ont une affinité élevée proche de celle du Lewis a. Tous les composés de ces séries n'ont malheureusement pas pu être cristallisés. Cependant, à en juger par la structure **50** appartenant à la série 3, on peut constater une forte complémentarité de forme entre le composé et la surface de la protéine expliquant peut-être son fort potentiel d'inhibition. Le composé **50** n'est pas le meilleur inhibiteur de sa série (serie3) dans les test ELLA (fig. 6.12). Ceci s'explique peut-être par le pont hydrogène que fait l'oxygène du groupe C(O)CH<sub>3</sub> de l'isoxazole avec la Ser23 et qui fige la Ser 23, alors que le composé **48** ayant à cette même place un groupement méthyle laisse la Ser23 libre en rotation.

En ce qui concerne les dimères de composé **13**, ceux-ci sont effectivement les meilleurs inhibiteurs testés mais leur potentiel d'inhibition n'est que deux fois supérieur à ceux des disaccharides monomères, La forte pénalité entropique observée en microcalorimétrie résulte sans doute de la perte de flexibilité des chaînes PEG espaçant les composés **13**. Compte tenu des longueurs de chaîne courtes et de la topologie de la protéine, la chélation de plusieurs sites de liaison n'est pas possible. Seule la formation d'un réseau de type « *-protéine-ligand-protéine-* » est possible, (comme ce fut confirmé par l'examen visuel des solutions après expériences ITC). Dans un tel cas de figure, l'augmentation de l'affinité s'explique par le fait que la dissociation est considérablement ralentie une fois que l'agrégation est établie puisque les différents constituants sont maintenus en contact avec des partenaires complémentaires.

La forme tridentée (**23**) de la molécule **13** est un composé relativement rigide, dans lequel les ligands sont peu espacés les uns des autres. Un conflit stérique empêchant l'approche d'une troisième molécule de PA-III une fois que les deux premières sont en place peut expliquer la faible affinité observée durant les tests ELLA. L'affinité de 0,14  $\mu\text{M}$  observée en microcalorimétrie confirme que ce trimère a le même comportement qu'un dimère.

# 7

## Conclusion et perspective

---

### 7.1. Lectine PA-III et infections bactériennes

Au cours de cette thèse, je me suis intéressé à l'étude des mécanismes de l'interaction entre la lectine PA-III, sauvage et mutée, et différents ligands: monosaccharides, oligosaccharides naturels et glycomimétiques de synthèse. Les données cristallographiques, thermodynamiques ainsi que les tests d'interaction, ont permis de caractériser l'affinité et la spécificité de la lectine PA-III pour ces différents composés. Cependant, les véritables cibles de l'adhésion ne sont pas clairement identifiées et le rôle physiologique joué par la lectine est encore loin d'être complètement élucidé. Des études menées en parallèle par un autre groupe ont néanmoins montré que PA-III était associée à la membrane externe de la bactérie et qu'elle jouait un rôle important dans la formation du biofilm (Tielker *et al.*, 2005).

Par ailleurs, la production de la lectine est sous le contrôle du « quorum sensing » (Winzer *et al.*, 2000) et cette protéine est capable de bloquer le battement ciliaire des cellules épithéliales des voies respiratoires (Adam *et al.*, 1997). Beaucoup d'indices retrouvés dans la littérature attestent que la lectine PA-III est bien un facteur de virulence, ce que confirme, le fait de retrouver des séquences similaires chez d'autres bactéries opportunistes. En effet, les génomes de *Ralstonia solanacearum*, un pathogène des plantes, et *Chromobacterium violaceum*, un pathogène humain opportuniste, contiennent un gène codant pour une protéine très similaire à PA-III. *Burkholderia cenocepacia*, responsable d'infections particulièrement graves chez les patients atteints de mucoviscidose, contient dans son génome, trois gènes codant pour des protéines contenant toutes un domaine similaire à PA-III (Imberty *et al.*, 2004). Ces bactéries opportunistes sont capables de s'adapter très rapidement à de nouveaux

environnements et sont fortement dangereuses une fois que l'infection est établie. De plus, l'étude structurale des PA-IIL mutées a confirmé que de ponctuelles mutations dans le site de reconnaissance modifiaient l'affinité et la spécificité de la lectine, puisqu'une seule mutation a transformé une lectine à fucose en une lectine spécifique du mannose. On peut voir dans ces résultats, une illustration plausible de l'adaptabilité bactérienne à l'environnement extérieur.

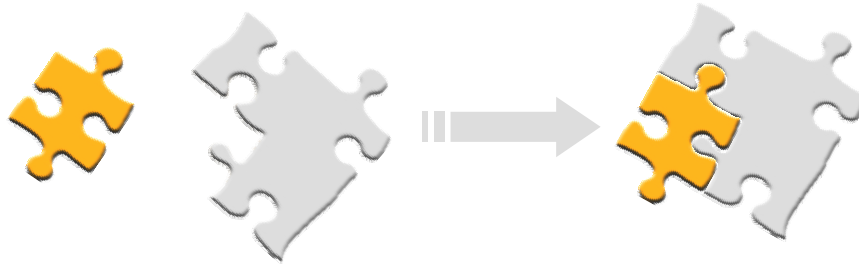
## 7.2. Du ligand naturel au médicament

Développer une molécule active revient à choisir un "lead" puis à augmenter son affinité pour le site de reconnaissance. Dans cette optique, il est possible d'utiliser la thermodynamique en jouant sur les contributeurs de l'énergie de Gibbs, en rendant  $\Delta H$  plus négatif et/ou  $\Delta S$  plus positif. Pour un substrat naturel ayant une affinité initiale de l'ordre du micromolaire, on s'accorde à dire qu'une molécule thérapeutique requiert une affinité de l'ordre du nanomolaire pour avoir un intérêt pharmaceutique.

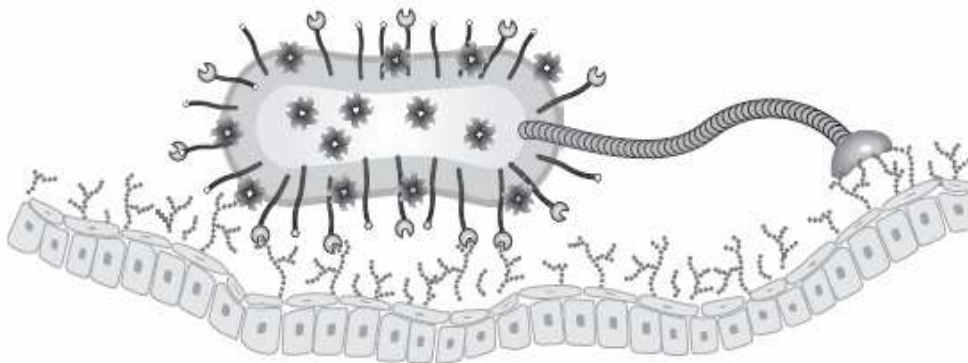
Puisqu'une perte d'entropie conformationnelle est défavorable, une approche consiste à minimiser cet effet indésirable en rigidifiant la molécule afin que celle-ci ne soit plus pénalisée en rentrant dans le site de reconnaissance. L'autre méthode souvent utilisée pour améliorer l'entropie d'interaction est l'ajout judicieux de groupements hydrophobes sur la molécule candidate. L'inconvénient qui peut alors survenir est une perte de solubilité de la molécule active. De plus, si l'astuce augmente certes l'affinité de la molécule, c'est parfois davantage par « désamour » du solvant que par « amour » du site actif, et la molécule peut venir se loger dans n'importe quelle autre poche hydrophobe. On s'aperçoit que l'augmentation de l'affinité ne s'accompagne pas pour autant d'une amélioration de la spécificité.

Comment jouer sur l'enthalpie ? Nous savons qu'il ne suffit pas simplement de multiplier les groupements polaires pouvant interagir car l'énergie de désolvatation en sera d'autant plus coûteuse. (~9kcal/mol pour  $NH_2$ ,  $NH$ ,  $OH$ ). C'est d'avantage la configuration des liaisons hydrogène ou des autres contacts (direction angle et longueur) qui définit la spécificité et la sélectivité. C'est finalement la complémentarité des formes de type clef-serrure, ou pièce de puzzle qui offre les meilleures chances de succès. Une molécule préformée à la géométrie du

site cible garantit simultanément l'affinité et la spécificité car la probabilité de trouver une autre situation équivalente devient faible.



Des travaux sont en cours, concernant les molécules glycomimétiques étudiés durant la thèse, notamment, en évaluant leur impact sur la croissance du biofilm bactérien (collaboration Dr. Sophie de Bentzmann, Marseille et Dr. Ina Attrée à Grenoble). Des tests *in vivo*, sur des souris infectées par *P. aeruginosa* sont aussi prévus dans un futur proche (collaboration Prof. B. Guery, Lille). Cet axe de recherche apparaît prometteur pour prévenir les stades précoces d'infections bactériennes et pourrait représenter un complément ou une alternative aux traitements antibiotiques. Cependant, étant donné que la bactérie possède un vaste arsenal de récepteurs se liant aux cellules hôtes (fig. 7.1), il faut envisager de bloquer plusieurs lectines et adhésines à la fois, peut-être à partir d'un mélange d'oligosaccharides, afin de reproduire dans les poumons de patients atteints de mucoviscidose, l'effet du lait maternel humain dans le système digestif des nourrissons.



**Figure 7.1.** Représentation d'une bactérie *Pseudomonas aeruginosa* avec différents récepteurs interagissant avec les glycoconjugués de la surface de la cellule hôte. (Imberty et al., 2006).





# Bibliographie



# Bibliographie

## A

- Adam, E.C., Mitchell, B.S., Schumacher, D.U., Grant, G. and Schumacher, U. (1997) *Pseudomonas aeruginosa* II lectin stops human ciliary beating: therapeutic implications of fucose. *Am. J. Respir. Crit. Care Med.*, **155**, 2102-2104.
- Aebi, C., Bracher, R., Liechti-Gallati, S., Tschappeler, H., Rudeberg, A. and Kraemer, R. (1995) The age at onset of chronic *Pseudomonas aeruginosa* colonization in cystic fibrosis--prognostic significance. *Eur J Pediatr*, **154**, S69-73.
- Allen, L., Dockrell, D.H., Pattery, T., Lee, D.G., Cornelis, P., Hellewell, P.G. and Whyte, M.K. (2005) Pyocyanin production by *Pseudomonas aeruginosa* induces neutrophil apoptosis and impairs neutrophil-mediated host defenses in vivo. *J Immunol*, **174**, 3643-3649.
- Andersson, B., Porras, O., Hanson, L.A., Lagergard, T. and Svanborg-Eden, C. (1986) Inhibition of attachment of *Streptococcus pneumoniae* and *Haemophilus influenzae* by human milk and receptor oligosaccharides. *J Infect Dis*, **153**, 232-237.
- Arora, S.K., Dasgupta, N., Lory, S. and Ramphal, R. (2000) Identification of two distinct types of flagellar cap proteins, FliD, in *Pseudomonas aeruginosa*. *Infect Immun*, **68**, 1474-1479.
- Arora, S.K., Ritchings, B.W., Almira, E.C., Lory, S. and Ramphal, R. (1998) The *Pseudomonas aeruginosa* flagellar cap protein, FliD, is responsible for mucin adhesion. *Infect Immun*, **66**, 1000-1007.
- Audette, G.F., Irvin, R.T. and Hazes, B. (2004) Crystallographic analysis of the *Pseudomonas aeruginosa* strain K122-4 monomeric pilin reveals a conserved receptor-binding architecture. *Biochemistry*, **43**, 11427-11435.
- Azghani, A.O., Miller, E.J. and Peterson, B.T. (2000) Virulence factors from *Pseudomonas aeruginosa* increase lung epithelial permeability. *Lung*, **178**, 261-269.

## B

- Bajolet-Laudinat, O., Girod-de Bentzmann, S., Tournier, J.M., Madoulet, C., Plotkowski, M.C., Chippaux, C. and Puchelle, E. (1994) Cytotoxicity of *Pseudomonas aeruginosa* internal lectin PA-I to respiratory epithelial cells in primary culture. *Infect Immun*, **62**, 4481-4487.

- Barasch, J. and al-Awqati, Q. (1993) Defective acidification of the biosynthetic pathway in cystic fibrosis. *J Cell Sci Suppl*, **17**, 229-233.
- Bardy, S.L., Ng, S.Y. and Jarrell, K.F. (2003) Prokaryotic motility structures. *Microbiology*, **149**, 295-304.
- Bayer, E.A. and Wilchek, M. (1990) Protein biotinylation. *Methods in Enzymology*, **184**, 138-160.
- Berg, H.C. (2003) The rotary motor of bacterial flagella. *Annu Rev Biochem*, **72**, 19-54.
- Blixt, O., Head, S., *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.
- Block, T.M., Lu, X., Platt, F.M., Foster, G.R., Gerlich, W.H., Blumberg, B.S. and Dwek, R.A. (1994) Secretion of human hepatitis B virus is inhibited by the imino sugar N-butyldeoxyjirimycin. *Proc Natl Acad Sci U S A*, **91**, 2235-2239.
- Boyd, M.R., Gustafson, K.R., *et al.* (1997) Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrob Agents Chemother*, **41**, 1521-1530.
- Brand-Miller, J.C., McVeagh, P., McNeil, Y. and Messer, M. (1998) Digestion of human milk oligosaccharides by healthy infants evaluated by the lactulose hydrogen breath test. *J Pediatr*, **133**, 95-98.
- Briel, M., Greger, R. and Kunzelmann, K. (1998) Cl<sup>-</sup> transport by cystic fibrosis transmembrane conductance regulator (CFTR) contributes to the inhibition of epithelial Na<sup>+</sup> channels (ENaCs) in *Xenopus* oocytes co-expressing CFTR and ENaC. *J Physiol*, **508 ( Pt 3)**, 825-836.
- Brockhausen, I., Vavasseur, F. and Yang, X. (2001) Biosynthesis of mucin type O-glycans: lack of correlation between glycosyltransferase and sulfotransferase activities and CFTR expression. *Glycoconj J*, **18**, 685-697.

## C

- Cheng, S.H., Gregory, R.J., Marshall, J., Paul, S., Souza, D.W., White, G.A., O'Riordan, C.R. and Smith, A.E. (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell*, **63**, 827-834.
- Chitnis, C.E. and Ohman, D.E. (1990) Cloning of *Pseudomonas aeruginosa* algG, which controls alginate structure. *J Bacteriol*, **172**, 2894-2900.
- Cioci, G., Mitchell, E.P., Gautier, C., Wimmerova, M., Sudakevitz, D., Perez, S., Gilboa-Garber, N. and Imberty, A. (2003) Structural basis of calcium and galactose recognition by the lectin PA-IL of *Pseudomonas aeruginosa*. *FEBS Lett*, **555**, 297-301.

Craig, L., Taylor, R.K., *et al.* (2003) Type IV pilin structure and assembly: X-ray and EM analyses of *Vibrio cholerae* toxin-coregulated pilus and *Pseudomonas aeruginosa* PAK pilin. *Mol Cell*, **11**, 1139-1150.

## D

Dacheux, D., Attree, I., Schneider, C. and Toussaint, B. (1999) Cell death of human polymorphonuclear neutrophils induced by a *Pseudomonas aeruginosa* cystic fibrosis isolate requires a functional type III secretion system. *Infect Immun*, **67**, 6164-6167.

Dacheux, D., Goure, J., Chabert, J., Usson, Y. and Attree, I. (2001) Pore-forming activity of type III system-secreted proteins leads to oncosis of *Pseudomonas aeruginosa*-infected macrophages. *Mol Microbiol*, **40**, 76-85.

Davidson, D.J. and Dorin, J.R. (2001) The CF mouse: an important tool for studying cystic fibrosis. *Expert Rev Mol Med*, **2001**, 1-27.

Davies, D.G., Parsek, M.R., Pearson, J.P., Iglewski, B.H., Costerton, J.W. and Greenberg, E.P. (1998) The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science*, **280**, 295-298.

De Graaf, T.W., Van der Stelt, M.E., Anbergen, M.G. and van Dijk, W. (1993) Inflammation-induced expression of sialyl Lewis X-containing glycan structures on alpha 1-acid glycoprotein (orosomucoid) in human sera. *J Exp Med*, **177**, 657-666.

DeLano, W.L. The PyMOL Molecular Graphics System.

Delmotte, P., Degroote, S., Lafitte, J.J., Lamblin, G., Perini, J.M. and Roussel, P. (2002) Tumor necrosis factor alpha increases the expression of glycosyltransferases and sulfotransferases responsible for the biosynthesis of sialylated and/or sulfated Lewis x epitopes in the human bronchial mucosa. *J Biol Chem*, **277**, 424-431.

Denning, G.M., Wollenweber, L.A., Railsback, M.A., Cox, C.D., Stoll, L.L. and Britigan, B.E. (1998) *Pseudomonas* pyocyanin increases interleukin-8 expression by human airway epithelial cells. *Infect Immun*, **66**, 5777-5784.

Diggle, S.P., Stacey, R.E., Dodd, C., Camara, M., Williams, P. and Winzer, K. (2006) The galactophilic lectin, LecA, contributes to biofilm development in *Pseudomonas aeruginosa*. *Environ Microbiol*, **8**, 1095-1104.

Diggle, S.P., Winzer, K., Chhabra, S.R., Worrall, K.E., Camara, M. and Williams, P. (2003) The *Pseudomonas aeruginosa* quinolone signal molecule overcomes the cell density-dependency of the quorum sensing hierarchy, regulates rhl-dependent genes at the onset of stationary phase and can be produced in the absence of LasR. *Mol Microbiol*, **50**, 29-43.

Diggle, S.P., Winzer, K., Lazdunski, A., Williams, P. and Camara, M. (2002) Advancing the quorum in *Pseudomonas aeruginosa*: MvaT and the regulation of N-acylhomoserine lactone production and virulence gene expression. *J Bacteriol*, **184**, 2576-2586.

- 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.
- Dosanjh, A., Lencer, W., Brown, D., Ausiello, D.A. and Stow, J.L. (1994) Heterologous expression of delta F508 CFTR results in decreased sialylation of membrane glycoconjugates. *Am J Physiol*, **266**, C360-366.
- Dumon, C., Priem, B., Martin, S.L., Heyraud, A., Bosso, C. and Samain, E. (2001) In vivo fucosylation of lacto-N-neotetraose and lacto-N-neohexaose by heterologous expression of Helicobacter pylori alpha-1,3 fucosyltransferase in engineered Escherichia coli. *Glycoconj J*, **18**, 465-474.

## E

- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
- Erickson, D.L., Endersby, R., Kirkham, A., Stuber, K., Vollman, D.D., Rabin, H.R., Mitchell, I. and Storey, D.G. (2002) Pseudomonas aeruginosa quorum-sensing systems may control virulence factor expression in the lungs of patients with cystic fibrosis. *Infect Immun*, **70**, 1783-1790.
- Erney, R.M., Malone, W.T., *et al.* (2000) Variability of human milk neutral oligosaccharides in a diverse population. *J Pediatr Gastroenterol Nutr*, **30**, 181-192.
- Ernst, R.K., Hajjar, A.M., Tsai, J.H., Moskowitz, S.M., Wilson, C.B. and Miller, S.I. (2003) Pseudomonas aeruginosa lipid A diversity and its recognition by Toll-like receptor 4. *J Endotoxin Res*, **9**, 395-400.
- Evans, L.R. and Linker, A. (1973) Production and characterization of the slime polysaccharide of Pseudomonas aeruginosa. *J Bacteriol*, **116**, 915-924.

## F

- Fan, E., O'Neal, C.J., *et al.* (2004) Structural biology and structure-based inhibitor design of cholera toxin and heat-labile enterotoxin. *Int J Med Microbiol*, **294**, 217-223.
- Fantini, J., Hammache, D., Delezay, O., Yahi, N., Andre-Barres, C., Rico-Lattes, I. and Lattes, A. (1997) Synthetic soluble analogs of galactosylceramide (GalCer) bind to the V3 domain of HIV-1 gp120 and inhibit HIV-1-induced fusion and entry. *J Biol Chem*, **272**, 7245-7252.
- Farinha, M.A., Conway, B.D., Glasier, L.M., Ellert, N.W., Irvin, R.T., Sherburne, R. and Paranchych, W. (1994) Alteration of the pilin adhesin of Pseudomonas aeruginosa PAO results in normal pilus biogenesis but a loss of adherence to human pneumocyte cells and decreased virulence in mice. *Infect Immun*, **62**, 4118-4123.

Feldman, M., Bryan, R., Rajan, S., Scheffler, L., Brunnert, S., Tang, H. and Prince, A. (1998) Role of flagella in pathogenesis of *Pseudomonas aeruginosa* pulmonary infection. *Infect Immun*, **66**, 43-51.

Friedman, L. and Kolter, R. (2004) Two genetic loci produce distinct carbohydrate-rich structural components of the *Pseudomonas aeruginosa* biofilm matrix. *J Bacteriol*, **186**, 4457-4465.

## G

Garber, N., Guempel, U., Belz, A., Gilboa-Garber, N. and Doyle, R.J. (1992) On the specificity of the D-galactose-binding lectin (PA-I) of *Pseudomonas aeruginosa* and its strong binding to hydrophobic derivatives of D-galactose and thiogalactose. *Biochim. Biophys. Acta*, **1116**, 331-333.

Garber, N., Guempel, U., Gilboa-Garber, N. and Doyle, R.J. (1987) Specificity of the fucose-binding lectin of *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.*, **48**, 331-334.

Gilboa-Garber, N. (1972) Purification and properties of hemagglutinin from *Pseudomonas aeruginosa* and its reaction with human blood cells. *Biochim Biophys Acta*, **273**, 165-173.

Gilboa-Garber, N. (1986) Lectins of *Pseudomonas aeruginosa*: Properties, biological effects and applications. In *Microbial Lectins and Agglutinins: Properties and Biological Activity*, Mirelman, D. (ed), John Wiley & Sons, New York, pp. 255-269.

Gilboa-Garber, N. (1997) Multiple aspects of *Pseudomonas aeruginosa* lectins. *Nova Acta Leopoldina NF 75*, **301**, 153-177.

Goldberg, J.B. and Pler, G.B. (1996) *Pseudomonas aeruginosa* lipopolysaccharides and pathogenesis. *Trends Microbiol*, **4**, 490-494.

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

Gupta, S.K., Berk, R.S., Masinick, S. and Hazlett, L.D. (1994) Pili and lipopolysaccharide of *Pseudomonas aeruginosa* bind to the glycolipid asialo GM1. *Infect Immun*, **62**, 4572-4579.

## H

Hahn, H.P. (1997) The type-4 pilus is the major virulence-associated adhesin of *Pseudomonas aeruginosa*. *Gene*, **192**, 99-108.

Hazes, B., Sastry, P.A., Hayakawa, K., Read, R.J. and Irvin, R.T. (2000) Crystal structure of *Pseudomonas aeruginosa* PAK pilin suggests a main-chain-dominated mode of receptor binding. *J. Mol. Biol.*, **299**, 1005-1017.



Hentzer, M., Teitzel, G.M., Balzer, G.J., Heydorn, A., Molin, S., Givskov, M. and Parsek, M.R. (2001) Alginate overproduction affects *Pseudomonas aeruginosa* biofilm structure and function. *J Bacteriol*, **183**, 5395-5401.

Holm, B.A., Keicher, L., Liu, M.Y., Sokolowski, J. and Enhorning, G. (1991) Inhibition of pulmonary surfactant function by phospholipases. *J Appl Physiol*, **71**, 317-321.

Holmgren, J., Svennerholm, A.M. and Lindblad, M. (1983) Receptor-like glycoconjugates in human milk that inhibit classical and El Tor *Vibrio cholerae* cell adherence (hemagglutination). *Infect Immun*, **39**, 147-154.

Hung, C.S., Bouckaert, J., *et al.* (2002) Structural basis of tropism of *Escherichia coli* to the bladder during urinary tract infection. *Mol Microbiol*, **44**, 903-915.

## I

Ikeda, T., Oosawa, K. and Hotani, H. (1996) Self-assembly of the filament capping protein, FliD, of bacterial flagella into an annular structure. *J Mol Biol*, **259**, 679-686.

Imberty, A., Wimmerova, M., Mitchell, E.P. and Gilboa-Garber, N. (2004) Structures of the lectins from *Pseudomonas aeruginosa*: insight into the molecular basis for host glycan recognition. *Microbes Infect*, **6**, 221-228.

Imberty A., Wimmerova M., Sabin C. & Mitchell E.P. (2006) Structures and roles of *Pseudomonas aeruginosa* lectins. In “*Protein-Carbohydrate Interactions in Infectious Disease*”, C. Bewley (ed.), RSC Biomolecular Sciences Series, Royal Society of Chemistry, pp 30-48.

## J

Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110-119.

## K

Keizer, D.W., Slupsky, C.M., Kalisiak, M., Campbell, A.P., Crump, M.P., Sastry, P.A., Hazes, B., Irvin, R.T. and Sykes, B.D. (2001) Structure of a pilin monomer from *Pseudomonas aeruginosa*: implications for the assembly of pili. *J Biol Chem*, **276**, 24186-24193.

Kerem, B., Rommens, J.M., Buchanan, J.A., Markiewicz, D., Cox, T.K., Chakravarti, A., Buchwald, M. and Tsui, L.C. (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science*, **245**, 1073-1080.

Kipnis, E., Sawa, T. and Wiener-Kronish, J. (2006) Targeting mechanisms of *Pseudomonas aeruginosa* pathogenesis. *Med Mal Infect*, **36**, 78-91.

- Kolter, R. and Losick, R. (1998) One for all and all for one. *Science*, **280**, 226-227.
- Kon, Y., Tsukada, H., Hasegawa, T., Igarashi, K., Wada, K., Suzuki, E., Arakawa, M. and Gejyo, F. (1999) The role of *Pseudomonas aeruginosa* elastase as a potent inflammatory factor in a rat air pouch inflammation model. *FEMS Immunol Med Microbiol*, **25**, 313-321.
- Konig, B., Jaeger, K.E., Sage, A.E., Vasil, M.L. and Konig, W. (1996) Role of *Pseudomonas aeruginosa* lipase in inflammatory mediator release from human inflammatory effector cells (platelets, granulocytes, and monocytes). *Infect Immun*, **64**, 3252-3258.
- Kostlanova, N., Mitchell, E.P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerova, M. and Imberty, A. (2005) The fucose-binding lectin from *Ralstonia solanacearum*. A new type of beta-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan. *J Biol Chem*, **280**, 27839-27849.
- Kunz, C., Rudloff, S., Baier, W., Klein, N. and Strobel, S. (2000) Oligosaccharides in human milk: structural, functional, and metabolic aspects. *Annu Rev Nutr*, **20**, 699-722.

## L

- Lamblin, G., Lhermitte, M., Klein, A., Houdret, N., Scharfman, A., Ramphal, R. and Roussel, P. (1991) The carbohydrate diversity of human respiratory mucins: a protection of the underlying mucosa? *Am Rev Respir Dis*, **144**, S19-24.
- Lau, G.W., Ran, H., Kong, F., Hassett, D.J. and Mavrodi, D. (2004) *Pseudomonas aeruginosa* pyocyanin is critical for lung infection in mice. *Infect Immun*, **72**, 4275-4278.
- Laughlin, R.S., Musch, M.W., Hollbrook, C.J., Rocha, F.M., Chang, E.B. and Alverdy, J.C. (2000) The key role of *Pseudomonas aeruginosa* PA-I lectin on experimental gut-derived sepsis. *Ann Surg*, **232**, 133-142.
- Lee, V.T., Smith, R.S., Tummler, B. and Lory, S. (2005) Activities of *Pseudomonas aeruginosa* effectors secreted by the Type III secretion system in vitro and during infection. *Infect Immun*, **73**, 1695-1705.
- Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, **26**.
- Lesman-Movshovich, E. and Gilboa-Garber, N. (2003) *Pseudomonas aeruginosa* lectin PA-III as a powerful probe for human and bovine milk analysis. *J. Dairy Sci.*, **86**, 2276-2282.
- Lewis, H.A., Buchanan, S.G., *et al.* (2004) Structure of nucleotide-binding domain 1 of the cystic fibrosis transmembrane conductance regulator. *Embo J*, **23**, 282-293.
- Lillehoj, E.P., Kim, B.T. and Kim, K.C. (2002) Identification of *Pseudomonas aeruginosa* flagellin as an adhesin for Muc1 mucin. *Am J Physiol Lung Cell Mol Physiol*, **282**, L751-756.

- Livermore, D.M. (2002) Multiple mechanisms of antimicrobial resistance in *Pseudomonas aeruginosa*: our worst nightmare? *Clin Infect Dis*, **34**, 634-640.
- Loris, R., Tielker, D., Jaeger, K.E. and Wyns, L. (2003) Structural basis of carbohydrate recognition by the lectin LecB from *Pseudomonas aeruginosa*. *J Mol Biol*, **331**, 861-870.
- Lyczak, J.B., Cannon, C.L. and Pier, G.B. (2002) Lung infections associated with cystic fibrosis. *Clin Microbiol Rev*, **15**, 194-222.

## M

- Mariencheck, W.I., Alcorn, J.F., Palmer, S.M. and Wright, J.R. (2003) *Pseudomonas aeruginosa* elastase degrades surfactant proteins A and D. *Am J Respir Cell Mol Biol*, **28**, 528-537.
- Mathee, K., Ciofu, O., *et al.* (1999) Mucoid conversion of *Pseudomonas aeruginosa* by hydrogen peroxide: a mechanism for virulence activation in the cystic fibrosis lung. *Microbiology*, **145 ( Pt 6)**, 1349-1357.
- Mattick, J.S. (2002) Type IV pili and twitching motility. *Annu Rev Microbiol*, **56**, 289-314.
- Miller, J.B. and McVeagh, P. (1999) Human milk oligosaccharides: 130 reasons to breast-feed. *Br J Nutr*, **82**, 333-335.
- Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gautier, C., Perez, S., Wu, A.M., Gilboa-Garber, N. and Imberty, A. (2002) Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nat Struct Biol*, **9**, 918-921.
- Mitchell, E.P., Sabin, C., *et al.* (2005) High affinity fucose binding of *Pseudomonas aeruginosa* lectin PA-IIL: 1.0 Å resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches. *Proteins*, **58**, 735-746.
- Miyazaki, S., Matsumoto, T., Tateda, K., Ohno, A. and Yamaguchi, K. (1995) Role of exotoxin A in inducing severe *Pseudomonas aeruginosa* infections in mice. *J Med Microbiol*, **43**, 169-175.
- Murshudov, G.N., A.A.Vagin and E.J.Dodson (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.*, **D53**, 240-255.

## N

- Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A*, **50**, 157-163.
- Newburg, D.S., Ruiz-Palacios, G.M., Altaye, M., Chaturvedi, P., Meinzen-Derr, J., Guerrero Mde, L. and Morrow, A.L. (2004) Innate protection conferred by fucosylated

oligosaccharides of human milk against diarrhea in breastfed infants. *Glycobiology*, **14**, 253-263.

## O

Odeh, R. and Quinn, J.P. (2000) Problem pulmonary pathogens: *Pseudomonas aeruginosa*. *Semin Respir Crit Care Med*, **21**, 331-339.

O'Riordan, C.R., Lachapelle, A.L., Marshall, J., Higgins, E.A. and Cheng, S.H. (2000) Characterization of the oligosaccharide structures associated with the cystic fibrosis transmembrane conductance regulator. *Glycobiology*, **10**, 1225-1233.

O'Toole, G.A. and Kolter, R. (1998) Flagellar and twitching motility are necessary for *Pseudomonas aeruginosa* biofilm development. *Mol Microbiol*, **30**, 295-304.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Macromolecular Crystallography, part A*, Carter, C.W.J. and Sweet, R.M. (eds), Academic Press, San Diego, pp. 307-326.

## P

Parge, H.E., Forest, K.T., Hickey, M.J., Christensen, D.A., Getzoff, E.D. and Tainer, J.A. (1995) Structure of the fibre-forming protein pilin at 2.6 Å resolution. *Nature*, **378**, 32-38.

Pasyk, E.A. and Foskett, J.K. (1997) Cystic fibrosis transmembrane conductance regulator-associated ATP and adenosine 3'-phosphate 5'-phosphosulfate channels in endoplasmic reticulum and plasma membranes. *J Biol Chem*, **272**, 7746-7751.

Perret, S., Sabin, C., *et al.* (2005) Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of *Pseudomonas aeruginosa*. *Biochem. J.*, **389**, 325-332.

Pier, G.B., Grout, M. and Zaidi, T.S. (1997) Cystic fibrosis transmembrane conductance regulator is an epithelial cell receptor for clearance of *Pseudomonas aeruginosa* from the lung. *Proc Natl Acad Sci U S A*, **94**, 12088-12093.

Pokorna, M., Cioci, G., *et al.* (2006) Unusual entropy-driven affinity of *Chromobacterium violaceum* lectin CV-IIL toward fucose and mannose. *Biochemistry*, **45**, 7501-7510.

Poschet, J.F., Boucher, J.C., Tattersson, L., Skidmore, J., Van Dyke, R.W. and Deretic, V. (2001) Molecular basis for defective glycosylation and *Pseudomonas* pathogenesis in cystic fibrosis lung. *Proc Natl Acad Sci U S A*, **98**, 13972-13977.

## Q

Quinton, P.M. (1983) Chloride impermeability in cystic fibrosis. *Nature*, **301**, 421-422.

## R

- Rebiere-Huet, J., Di Martino, P. and Hulen, C. (2004) Inhibition of *Pseudomonas aeruginosa* adhesion to fibronectin by PA-IL and monosaccharides: involvement of a lectin-like process. *Can J Microbiol*, **50**, 303-312.
- Remminghorst, U. and Rehm, B.H. (2006) Bacterial alginates: from biosynthesis to applications. *Biotechnol Lett*, **28**, 1701-1712.
- Rhim, A.D., Kothari, V.A., Park, P.J., Mulberg, A.E., Glick, M.C. and Scanlin, T.F. (2000) Terminal glycosylation of cystic fibrosis airway epithelial cells. *Glycoconj J*, **17**, 385-391.
- Rhim, A.D., Stoykova, L., Glick, M.C. and Scanlin, T.F. (2001) Terminal glycosylation in cystic fibrosis (CF): a review emphasizing the airway epithelial cell. *Glycoconj J*, **18**, 649-659.
- Rhim, A.D., Stoykova, L.I., Trindade, A.J., Glick, M.C. and Scanlin, T.F. (2004) Altered terminal glycosylation and the pathophysiology of CF lung disease. *J Cyst Fibros*, **3 Suppl 2**, 95-96.
- Riordan, J.R., Rommens, J.M., *et al.* (1989) Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, **245**, 1066-1073.
- Rosenberg, M.F., Kamis, A.B., Aleksandrov, L.A., Ford, R.C. and Riordan, J.R. (2004) Purification and crystallization of the cystic fibrosis transmembrane conductance regulator (CFTR). *J Biol Chem*, **279**, 39051-39057.

## S

- Sabbatini, P.J., Kudryashov, V., *et al.* (2000) Immunization of ovarian cancer patients with a synthetic Lewis(y)-protein conjugate vaccine: a phase 1 trial. *Int J Cancer*, **87**, 79-85.
- Sabin, C., Mitchell, E.P., Pokorna, M., Gautier, C., Utille, J.P., Wimmerova, M. and Imberty, A. (2006) Binding of different monosaccharides by lectin PA-III from *Pseudomonas aeruginosa*: thermodynamics data correlated with X-ray structures. *FEBS Lett*, **580**, 982-987.
- Sadovskaya, I., Brisson, J.R., Lam, J.S., Richards, J.C. and Altman, E. (1998) Structural elucidation of the lipopolysaccharide core regions of the wild-type strain PAO1 and O-chain-deficient mutant strains AK1401 and AK1012 from *Pseudomonas aeruginosa* serotype O5. *Eur J Biochem*, **255**, 673-684.
- Samuelsson, B.E., Rydberg, L., *et al.* (1994) Natural antibodies and human xenotransplantation. *Immunol Rev*, **141**, 151-168.
- Sato, H., Feix, J.B., Hillard, C.J. and Frank, D.W. (2005) Characterization of phospholipase activity of the *Pseudomonas aeruginosa* type III cytotoxin, ExoU. *J Bacteriol*, **187**, 1192-1195.

- Sayner, S.L., Frank, D.W., King, J., Chen, H., VandeWaa, J. and Stevens, T. (2004) Paradoxical cAMP-induced lung endothelial hyperpermeability revealed by *Pseudomonas aeruginosa* ExoY. *Circ Res*, **95**, 196-203.
- Scanlin, T.F. and Glick, M.C. (2001) Glycosylation and the cystic fibrosis transmembrane conductance regulator. *Respir Res*, **2**, 276-279.
- Scharfman, A., Arora, S.K., Delmotte, P., Van Brussel, E., Mazurier, J., Ramphal, R. and Roussel, P. (2001) Recognition of Lewis x derivatives present on mucins by flagellar components of *Pseudomonas aeruginosa*. *Infect Immun*, **69**, 5243-5248.
- Schreiber, R., Nitschke, R., Greger, R. and Kunzelmann, K. (1999) The cystic fibrosis transmembrane conductance regulator activates aquaporin 3 in airway epithelial cells. *J Biol Chem*, **274**, 11811-11816.
- Schulz, S., Jakubiczka, S., Kropf, S., Nickel, I., Muschke, P. and Kleinstein, J. (2006) Increased frequency of cystic fibrosis transmembrane conductance regulator gene mutations in infertile males. *Fertil Steril*, **85**, 135-138.
- Schuster, M., Lostroh, C.P., Ogi, T. and Greenberg, E.P. (2003) Identification, timing, and signal specificity of *Pseudomonas aeruginosa* quorum-controlled genes: a transcriptome analysis. *J. Bacteriol.*, **185**, 2066-2079.
- Seksek, O., Biwersi, J. and Verkman, A.S. (1996) Evidence against defective trans-Golgi acidification in cystic fibrosis. *J Biol Chem*, **271**, 15542-15548.
- Shaver, C.M. and Hauser, A.R. (2004) Relative contributions of *Pseudomonas aeruginosa* ExoU, ExoS, and ExoT to virulence in the lung. *Infect Immun*, **72**, 6969-6977.
- Sheth, H.B., Lee, K.K., Wong, W.Y., Srivastava, G., Hindsgaul, O., Hodges, R.S., Paranchych, W. and Irvin, R.T. (1994) The pili of *Pseudomonas aeruginosa* strains PAK and PAO bind specifically to the carbohydrate sequence beta GalNAc(1-4)beta Gal found in glycosphingolipids asialo-GM1 and asialo-GM2. *Mol. Microbiol.*, **11**, 715-723.
- Simpson, D.A., Ramphal, R. and Lory, S. (1992) Genetic analysis of *Pseudomonas aeruginosa* adherence: distinct genetic loci control attachment to epithelial cells and mucins. *Infect Immun*, **60**, 3771-3779.
- Slovin, S.F., Ragupathi, G., *et al.* (1999) Carbohydrate vaccines in cancer: immunogenicity of a fully synthetic globo H hexasaccharide conjugate in man. *Proc Natl Acad Sci U S A*, **96**, 5710-5715.
- Smith, J.J., Travis, S.M., Greenberg, E.P. and Welsh, M.J. (1996) Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell*, **85**, 229-236.
- Spangenberg, C., Heuer, T., Burger, C. and Tumber, B. (1996) Genetic diversity of flagellins of *Pseudomonas aeruginosa*. *FEBS Lett*, **396**, 213-217.
- Stover, C.K., Pham, X.Q., *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature*, **406**, 959-964.

Sudakevitz, D. and Gilboa-Garber, N. (1982) Effect of *Pseudomonas aeruginosa* lectins on phagocytosis of *Escherichia coli* strains by human polymorphonuclear leucocytes. *Microbios*, **34**, 159-166.

Sudakevitz, D., Kostlanova, N., Blatman-Jan, G., Mitchell, E.P., Lerrer, B., Wimmerova, M., Katcoff, D.J., Imberty, A. and Gilboa-Garber, N. (2004) A new *Ralstonia solanacearum* high-affinity mannose-binding lectin RS-III structurally resembling the *Pseudomonas aeruginosa* fucose-specific lectin PA-III. *Mol Microbiol*, **52**, 691-700.

Sugita, M., Yue, Y. and Foskett, J.K. (1998) CFTR Cl<sup>-</sup> channel and CFTR-associated ATP channel: distinct pores regulated by common gates. *Embo J*, **17**, 898-908.

## T

Tager, A.M., Wu, J. and Vermeulen, M.W. (1998) The effect of chloride concentration on human neutrophil functions: potential relevance to cystic fibrosis. *Am J Respir Cell Mol Biol*, **19**, 643-652.

Tielker, D., Hacker, S., Loris, R., Strathmann, M., Wingender, J., Wilhelm, S., Rosenau, F. and Jaeger, K.E. (2005) *Pseudomonas aeruginosa* lectin LecB is located in the outer membrane and is involved in biofilm formation. *Microbiology*, **151**, 1313-1323.

## V

Vagin, A. and Teplyakov, A. (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.*, **30**, 1022-1025.

Vallet, I., Olson, J.W., Lory, S., Lazdunski, A. and Filloux, A. (2001) The chaperone/usher pathways of *Pseudomonas aeruginosa*: identification of fimbrial gene clusters (cup) and their involvement in biofilm formation. *Proc. Natl. Acad. Sci. USA*, **98**, 6911-6916.

Viverge, D., Grimmonprez, L., Cassanas, G., Bardet, L. and Solere, M. (1990) Discriminant carbohydrate components of human milk according to donor secretor types. *J Pediatr Gastroenterol Nutr*, **11**, 365-370.

Vriend, G. (1990) WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 52-56.

## W

Wagner, V.E., Bushnell, D., Passador, L., Brooks, A.I. and Iglewski, B.H. (2003) Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. *J. Bacteriol.*, **185**, 2080-2095.

Ward, C.L., Omura, S. and Kopito, R.R. (1995) Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell*, **83**, 121-127.

- Wick, M.J., Hamood, A.N. and Iglewski, B.H. (1990) Analysis of the structure-function relationship of *Pseudomonas aeruginosa* exotoxin A. *Mol Microbiol*, **4**, 527-535.
- Wingender, J., Neu, T.R. and Flemming, H.C. (1999). In *Microbial extracellular polymeric substances - Characterization, Structure and Function*, Wingender, J., Neu, T.R., et al. (eds), Springer.
- Winzer, K., Falconer, C., Garber, N.C., Diggle, S.P., Camara, M. and Williams, P. (2000) The *Pseudomonas aeruginosa* lectins PA-IL and PA-III are controlled by quorum sensing and by RpoS. *J Bacteriol*, **182**, 6401-6411.
- Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, **179**, 131-137.
- Wolfgang, M.C., Kulasekara, B.R., Liang, X., Boyd, D., Wu, K., Yang, Q., Miyada, C.G. and Lory, S. (2003) Conservation of genome content and virulence determinants among clinical and environmental isolates of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A*, **100**, 8484-8489.
- Worlitzsch, D., Tarran, R., *et al.* (2002) Effects of reduced mucus oxygen concentration in airway *Pseudomonas* infections of cystic fibrosis patients. *J Clin Invest*, **109**, 317-325.
- Wozniak, D.J., Wyckoff, T.J., Starkey, M., Keyser, R., Azadi, P., O'Toole, G.A. and Parsek, M.R. (2003) Alginate is not a significant component of the extracellular polysaccharide matrix of PA14 and PAO1 *Pseudomonas aeruginosa* biofilms. *Proc Natl Acad Sci U S A*, **100**, 7907-7912.

## X

- Xia, B., Royall, J.A., Damera, G., Sachdev, G.P. and Cummings, R.D. (2005) Altered O-glycosylation and sulfation of airway mucins associated with cystic fibrosis. *Glycobiology*, **15**, 747-775.

## Y

- Yao, J.X., Dodson, E.J., Wilson, K.S. and Woolfson, M.M. (2006) ACORN: a review. *Acta Crystallogr D Biol Crystallogr*, **62**, 901-908.
- Yonekura, K., Maki, S., Morgan, D.G., DeRosier, D.J., Vonderviszt, F., Imada, K. and Namba, K. (2000) The bacterial flagellar cap as the rotary promoter of flagellin self-assembly. *Science*, **290**, 2148-2152.
- Yonekura, K., Maki-Yonekura, S. and Namba, K. (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature*, **424**, 643-650.



## Z

- Zielenski, J. and Tsui, L.C. (1995) Cystic fibrosis: genotypic and phenotypic variations. *Annu. Rev. Genet.*, **29**, 777-807.
- Zinger-Yosovich, K., Sudakevitz, D., Imberty, A., Garber, N.C. and Gilboa-Garber, N. (2006) Production and properties of the native *Chromobacterium violaceum* fucose-binding lectin (CV-IIL) compared to homologous lectins of *Pseudomonas aeruginosa* (PA-IIL) and *Ralstonia solanacearum* (RS-IIL). *Microbiology*, **152**, 457-463.

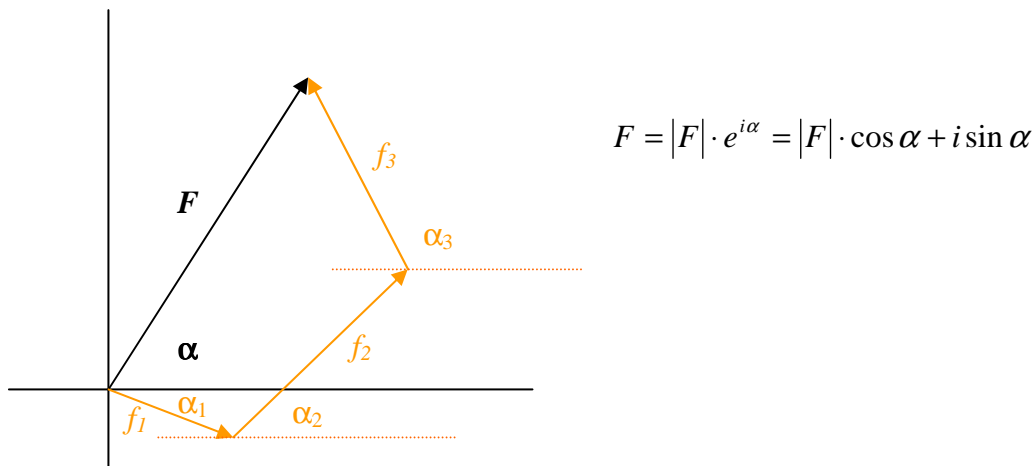
# Annexes



# Annexe 1

## Le problème des phases

Afin de résoudre le problème des phases, il convient d'utiliser la représentation imaginaire du facteur de structure et se souvenir que celui-ci est la somme de facteurs de structure atomiques comme le montre la figure ci-dessous



**Figure 1.** Représentation vectorielle d'un facteur de structure comme la somme de facteurs de structure atomiques.

Les méthodes exposées par la suite afin de résoudre le problème des phases, nécessitent de localiser les atomes lourds. Pour ce faire, la *fonction de Patterson*, définie par :

$$P(uvw) = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 \cos[2\pi (hx + ky + lz)]$$

est une sorte de série de Fourier débarrassée des phases. Les pics de densité visibles sur une carte dite « carte de Patterson » représentent tous les vecteurs inter-atomiques entre les atomes de la maille réelle. Exemple : si dans la maille réelle il y a deux atomes séparés par un vecteur  $r(u,v,w)$ , alors on retrouvera la présence de deux pics dans la carte de Patterson, l'un à  $(u,v,w)$  et l'autre à  $(-u,-v,-w)$ . Il est alors possible de remonter aux positions atomiques. En pratique,

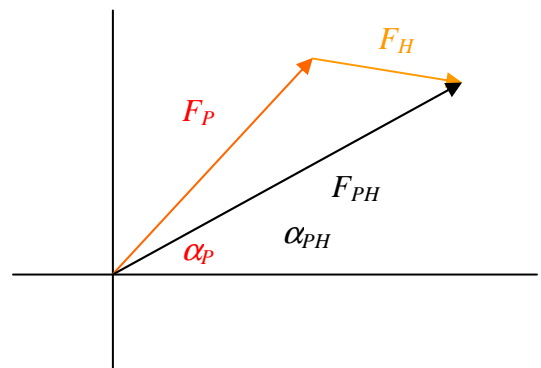
si la longueur d'onde le permet, presque tous les atomes à partir du soufre peuvent être utilisés comme dérivés lourds tels que le brome, le xénon, les métaux de transition et surtout les lanthanides. Le sélénium, incorporé dans une protéine recombinante comme la séléniométhionine présente l'avantage de ne pas poser de problème d'isomorphisme puisque le même cristal sert à la fois pour la forme native et dérivée. Dans le cas de la lectine PA-III, la toute première structure a été résolue par méthode SAD en trempant le cristal natif dans une solution contenant de l'holmium.

### 1.1. Le remplacement isomorphe

Cette méthode consiste à insérer dans le cristal natif un atome lourd, c'est-à-dire, ayant un nombre d'électrons importants capable de modifier les intensités des taches de façon détectable, ce qui équivaut à modifier les amplitudes des facteurs de structures. Pour être utilisable, la méthode impose que le cristal *dérivé* (contenant les atomes lourds) conserve le même groupe d'espace et les mêmes dimensions de maille que le cristal natif, puisque c'est la différence entre les deux jeux de données qui permettra de trouver les phases. Ainsi, les amplitudes et les phases des facteurs de structure, de la structure native  $P$  et de la structure dérivée  $PH$  sont tels que :

$$|F_H| \cong |F_{PH}| - |F_P|$$

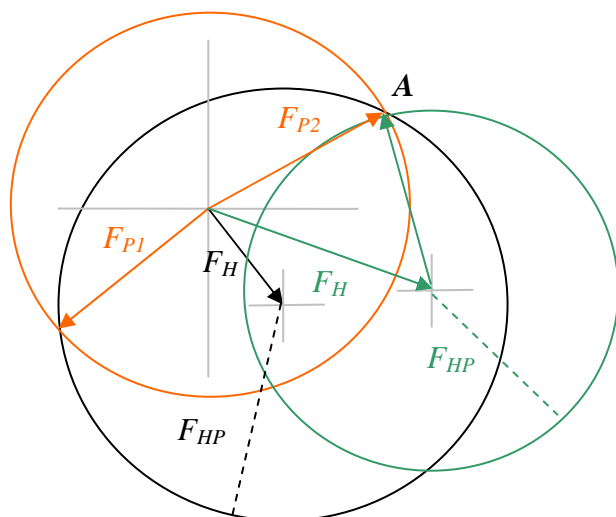
$$\alpha_P = \alpha_H + \cos^{-1} \left[ \frac{(F_{PH}^2 - F_P^2 - F_H^2)}{2 F_P F_H} \right]$$



où  $F_H$  est le facteur de structure de l'atome lourd seul.

Les positions des atomes lourds sont déterminées en utilisant la méthode de Patterson. Cette équation donne deux solutions symétriques. On peut alors essayer les deux possibilités.

Cependant, pour lever cette ambiguïté de phase, une solution est d'utiliser un deuxième ou plusieurs dérivés de façon à discriminer la phase finale. On parle alors de méthode MIR (Multiple Isomorphous Replacement).

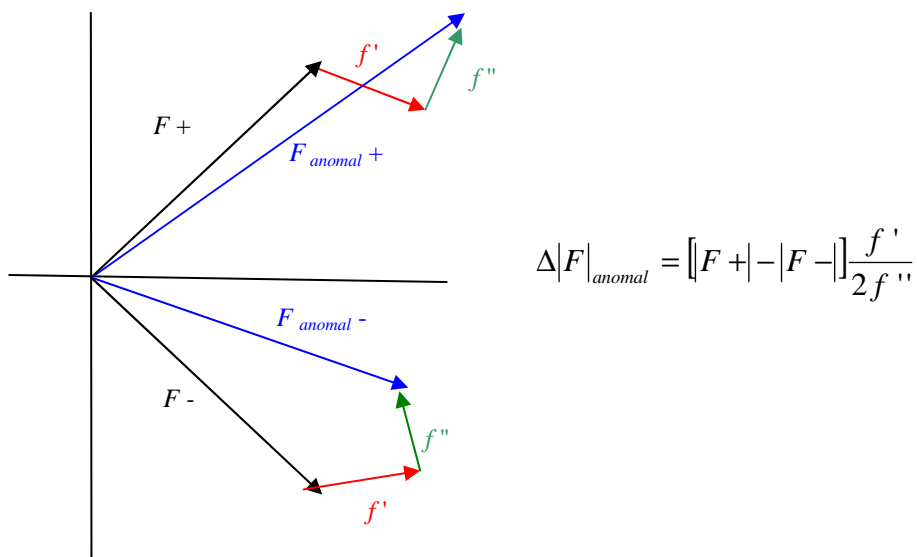


**Figure 2.** Construction vectorielle de la méthode MIR. Un seul dérivé laisse une incertitude entre les phases de  $F_{P1}$  et  $F_{P2}$ , un deuxième dérivé lourd permet de discriminer la solution A comme étant la phase correcte.

### 1.2. La diffusion anormale

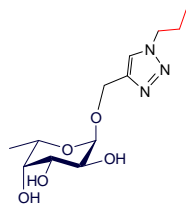
Quand des atomes lourds sont irradiés par un faisceau X d'une longueur d'onde d'énergie proche de celle de leur seuil d'absorption, ceux-ci réémettent une radiation avec une phase altérée appelée diffusion « anormale ». Pour la construction vectorielle, il faut ajouter au facteur de diffusion classique  $F$ , un couple de vecteur, dont l'un  $f'$  est « en phase » avec  $F$  et l'autre  $f''$  dont la phase est toujours retardée de  $90^\circ$ . Cette particularité confère à  $F$  de ne plus respecter la loi de Friedel voulant que le module des facteurs de structure pour les réflexions reliées par symétrie soit identique. La méthode SAD (*Single Anomalous Dispersion*) utilise la différence  $\Delta|F|_{\text{anomal}}$  entre les réflexions liées par symétrie (*paires de Friedel*). La méthode de Patterson détermine ensuite les positions des atomes anormaux pour résoudre la phase. De même, la méthode MAD (*Multiwavelength Anomalous Dispersion*) mesure le cristal à plusieurs longueurs d'onde différentes, ce qui équivaut à disposer des plusieurs cristaux

« dérivés ». Par analogie avec la méthode MIR, chaque contribution anormale se somme afin d'augmenter la probabilité de phase.

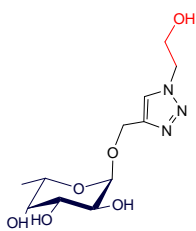


**Figure 3.** Représentation vectorielle de la diffusion anormale dans laquelle  $F_{hkl}$  n'est plus égale à  $F_{-h-k-l}$

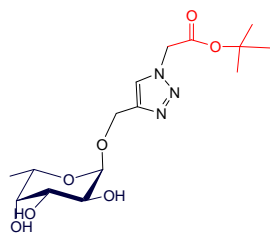
# Annexe 2



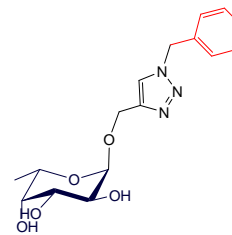
**1**



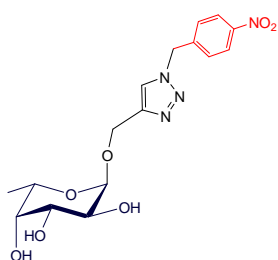
**2**



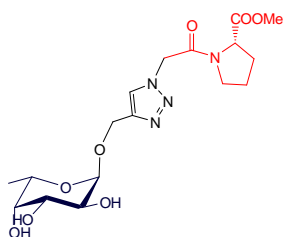
**3**



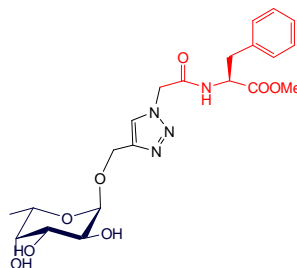
**4**



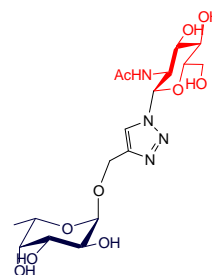
**5**



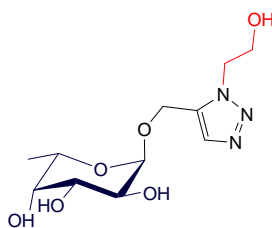
**7**



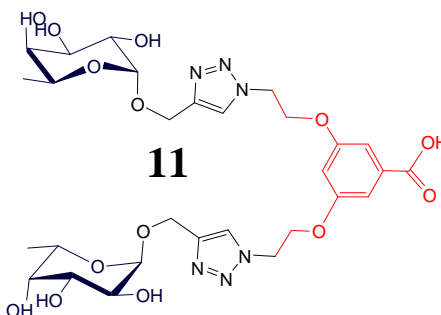
**8**



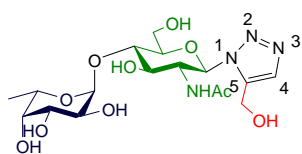
**9**



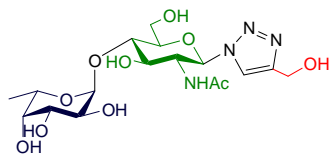
**10**



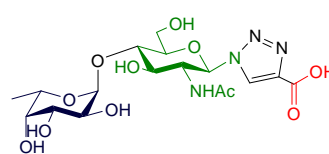
**11**



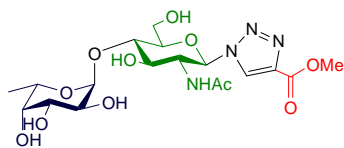
**12**



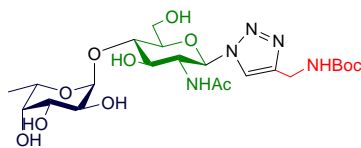
**13**



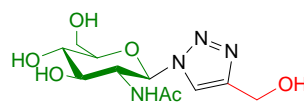
**14**



**15**

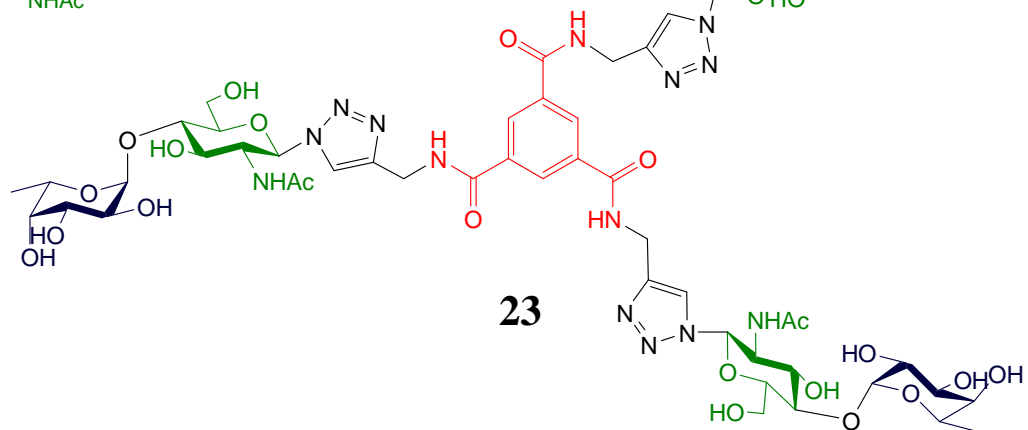
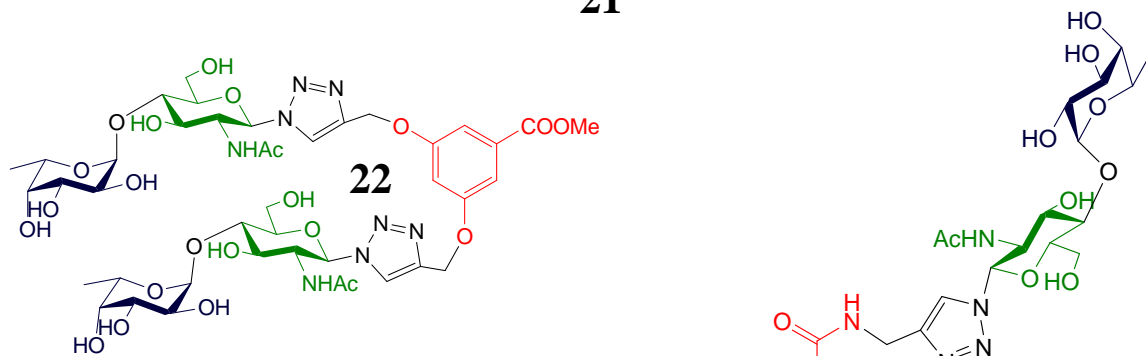
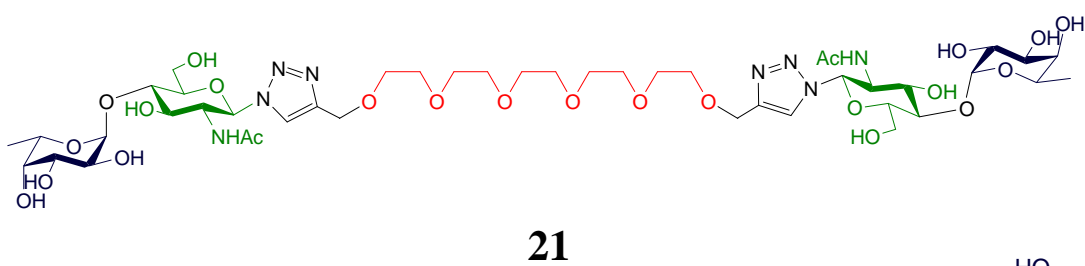
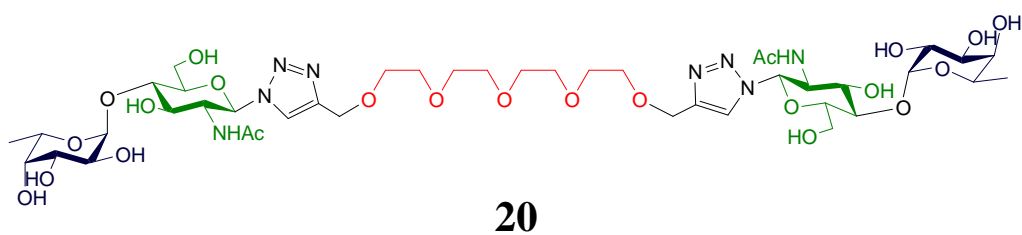
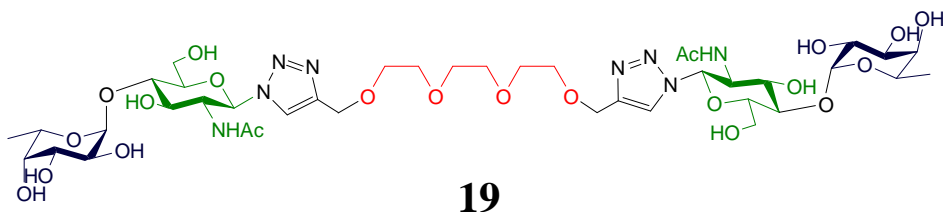
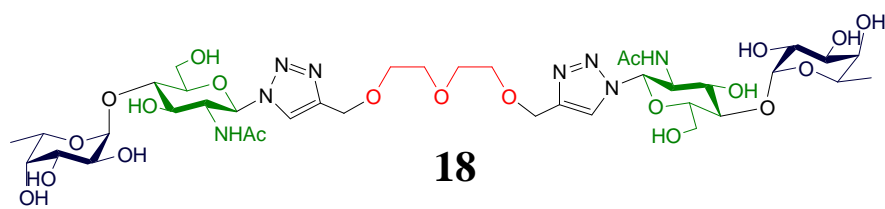


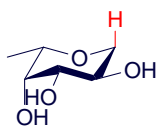
**16**



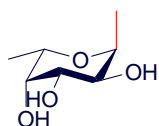
**17**







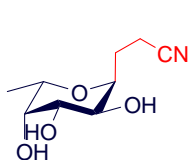
**24**



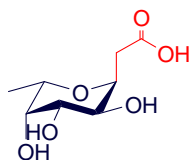
**25**



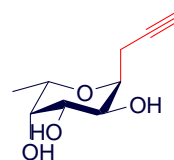
**26**



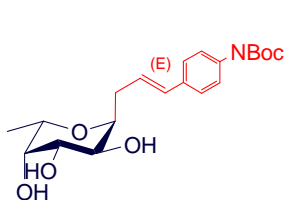
**27**



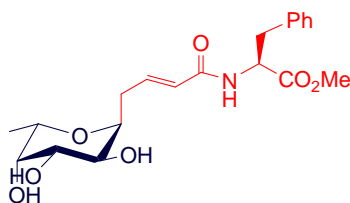
**28**



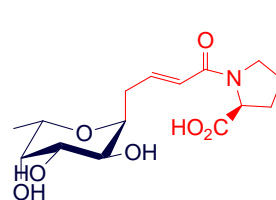
**29**



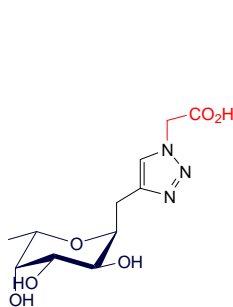
**30**



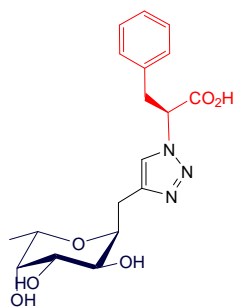
**31**



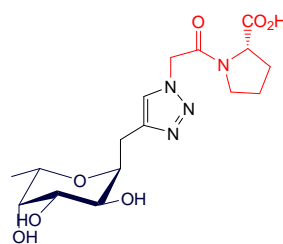
**32**



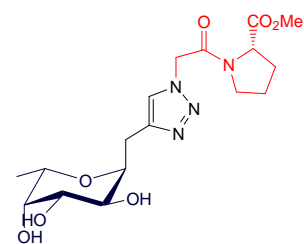
**33**



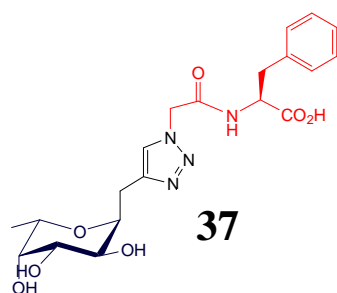
**34**



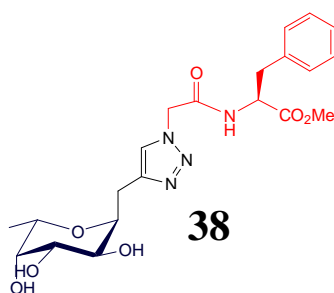
**35**



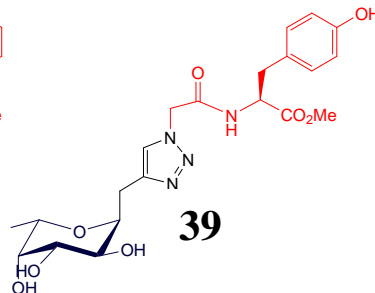
**36**



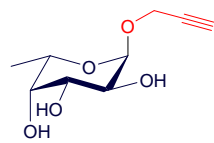
**37**



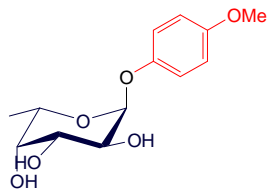
**38**



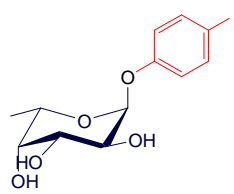
**39**



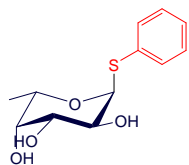
**40**



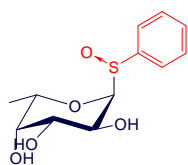
**41**



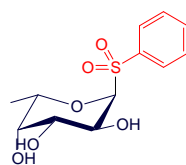
**42**



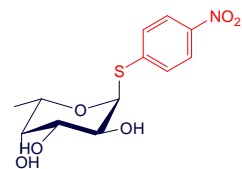
**43**



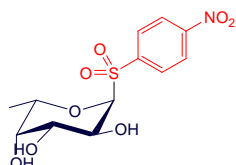
**44**



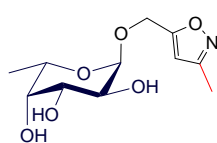
**45**



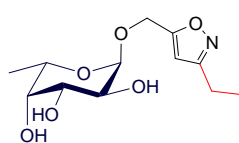
**46**



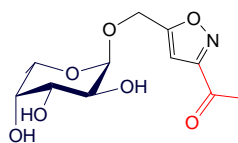
**47**



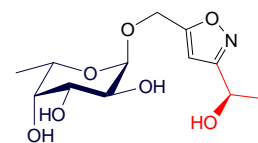
**48**



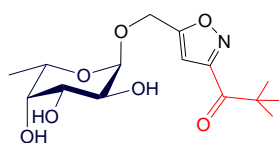
**49**



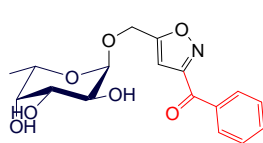
**50**



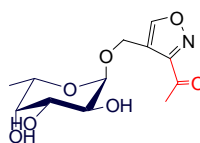
**51**



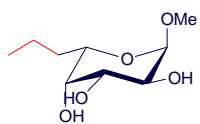
**52**



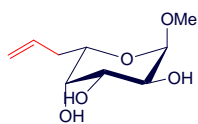
**53**



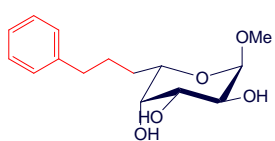
**54**



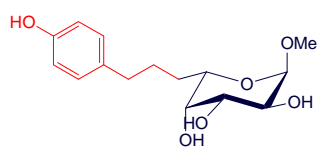
**55**



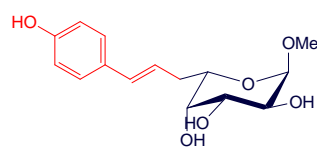
**56**



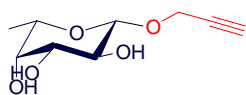
**57**



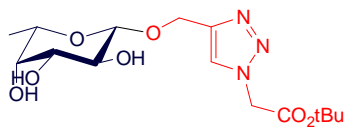
**58**



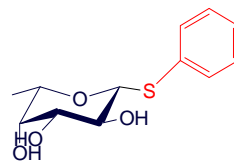
**59**



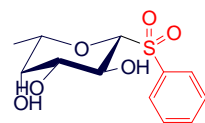
**60**



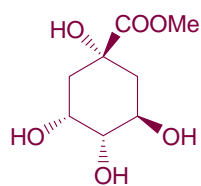
**61**



**62**



**63**



**64**





## RESUME

La colonisation chronique des poumons par *P. aeruginosa* est une cause majeure d'infection et de mortalité chez les patients atteints de mucoviscidose (CF). Pour la bactérie pathogène, la capacité à adhérer aux cellules hôtes est essentielle pour initier l'infection, et les glycoconjugués présents à la surface des cellules épithéliales sont autant de cibles pour les récepteurs bactériens. Chez les patients CF, des modifications surviennent dans les glycoconjugués des surfaces épithéliales ainsi que dans les mucines pulmonaires et salivaires. Une augmentation d'oligosaccharides fucosylés de type Lewis est observée, souvent accompagnée de sialylation et/ou sulfatation. Une lectine soluble, PA-IIL, reconnaissant spécifiquement le fucose, a été identifiée dans le cytoplasme de *P. aeruginosa*. Cette lectine, également présente sur la surface de la bactérie, pourrait être impliquée dans la reconnaissance de l'hôte et/ou dans la formation du biofilm responsable de la résistance de la bactérie contre les antibiotiques. Les travaux développés dans la thèse portent sur l'étude structurale et thermodynamique de l'interaction de PA-IIL avec des ligands naturels et des glycomimétiques de synthèse. Des tests d'interaction ELLA (*Enzyme-Linked Lectin Assays*) ont été mis en place afin d'évaluer les propriétés inhibitrices de ces composés. Ces travaux constituent une première approche dans la conception de composés dérivés des sucres pouvant servir comme alternative aux antibiotiques utilisés dans les infections survenant chez les patients atteints de mucoviscidose. (*Projet financé par l'association: Vaincre la Mucoviscidose*)

## ABSTRACT

The chronic colonization of the lungs by *P. aeruginosa* is a major cause of infection and mortality among cystic fibrosis patients (CF). For the pathogenic bacterium, the ability to adhere to the host cells is essential to initiate the infection, and the glycoconjugates of the surface of the epithelial cells are as many targets for the bacterial receptors. Among CF patients, modifications occur in glycoconjugates of the epithelial cell surface on pulmonary and salivary mucins. An increase in fucosylated oligosaccharides of Lewis type is observed, often with sialylation and/or sulfatation. A soluble lectin, PA-IIL, recognizing the fucose specifically, was identified in the cytoplasm of *P. aeruginosa*. This lectin, also found on the surface of the bacterium, could be implied in the recognition of the host and/or the formation of the biofilm that is responsible for the resistance of the bacterium against antibiotics. The work developed in the thesis concerns the structural and thermodynamic study of the interaction of PA-IIL with natural ligands and glycomimetics. Tests of interaction ELLA (*Enzyme-Linked Lectin Assays*) were set up in order to evaluate the properties of inhibition of these compounds. This work constitutes a beginning in the design of new compounds derived from sugars. These glycomimetics could be use as anti-bacterial in the infections occurring among CF patients. (*Project financed by association: Vaincre la Mucoviscidose*)

**DISCIPLINE : BIOLOGIE STRUCTURALE**

**MOTS CLES: structure, microcalorimétrie, tests ELLA, lectine, oligosaccharides, glycomimétiques, *Pseudomonas aeruginosa*, Mucoviscidose**