



ETUDE STRUCTURE-FONCTION DE GLYCOCONJUGUES ET DE LECTINES BACTERIENNES ET FONGIQUES

Gianluca Cioci

► To cite this version:

Gianluca Cioci. ETUDE STRUCTURE-FONCTION DE GLYCOCONJUGUES ET DE LECTINES BACTERIENNES ET FONGIQUES. Biochimie [q-bio.BM]. Université Joseph-Fourier - Grenoble I, 2006. Français. NNT: . tel-00081084

HAL Id: tel-00081084

<https://theses.hal.science/tel-00081084>

Submitted on 22 Jun 2006

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

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

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

THESE

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

Discipline : Biologie structurale et fonctionnelle

présentée et soutenue publiquement par

Gianluca CIOCI

Le 31 janvier 2006

**ETUDE STRUCTURE-FONCTION DE GLYCOCONJUGUES ET
DE LECTINES BACTERIENNES ET FONGIQUES**

Directeurs de thèse : Dr. Serge Pérez et Dr. Edward Mitchell

JURY

Prof. Anna Bernardi, Rapporteur

Dr. Richard Haser, Rapporteur

Prof. Sine Larsen, Examinateur

Dr. Andrea Dessen, Examinateur

Dr. Edward Mitchell, Examinateur

Dr. Anne Imberty, Examinateur

Dr. Serge Pérez, Examinateur

Avant propos et organisation de la thèse

Les sujets abordés dans cette thèse se situent dans le domaine de la glycobiologie structurale, qui est la branche de la biologie qui étudie les structures et les interactions des glycomolécules. Pendant les trois années passées au CERMAV j'ai été confronté à des thématiques différentes mais complémentaires telles que l'analyse structurale et la modélisation des glycannes et surtout la caractérisation des interactions lectine-sucre. En conséquence, cette thèse montre un évident caractère multidisciplinaire. On peut diviser la thèse en deux parties : une partie qui est dédiée à la structure des glucides et une deuxième partie qui analyse les interactions de ces molécules avec des récepteurs, les lectines. Dans la première partie on a voulu montrer deux techniques différentes, la cristallographie et la modélisation moléculaire en essayant d'explorer les limites de l'une et de l'autre. Dans la deuxième partie de la thèse, sûrement plus riche en quantité de travail, on a étudié les structures et les interactions des trois lectines différentes.

Cette thèse comprend trois chapitres de présentation générale. Le premier chapitre est une introduction à la glycobiologie, à la structure chimique et tridimensionnelle des glucides. Une attention particulière a été portée aux méthodes permettant l'étude structurale et l'analyse conformationnelle des glycannes. Le deuxième chapitre est dédié aux interactions lectines-sucres. On présentera les lectines de façon générale ainsi que les thèmes de recherche qui ont été étudiés. Le troisième chapitre est une introduction aux méthodes expérimentales utilisées, notamment la cristallographie, la calorimétrie et la modélisation moléculaire.

Les six publications (dont quatre déjà publiées) préparées au cours de cette thèse ont été illustrées dans cinq chapitres différents et une annexe. Chaque publication est précédée d'une brève introduction et d'une présentation des résultats. Le quatrième chapitre est consacré à la structure tridimensionnelle de la Tricolorine A, qui est un intéressant glycolipide extrait d'une plante originaire du Mexique. Dans le cinquième chapitre on a abordé la modélisation moléculaire d'un *O*-glycanne complexe. Le sixième chapitre décrit la structure de la lectine bactérienne PA-IL de *Pseudomonas aeruginosa* qui joue un rôle dans les infections causées par cette bactérie opportuniste. Dans le septième chapitre on s'intéressera à la lectine bactérienne CV-IIL de *Chromobacterium violaceum*. Le huitième chapitre décrit le nouveau repliement protéique de la lectine fongique PVL de *Psathyrella velutina*. Quelques conclusions et perspectives terminent la présentation de ces travaux.

Chapitre 1 : Introduction à la structure des glycoconjugués

1.1	Les glucides : généralités	01
1.2	Relations structure-fonction	02
1.3	Glycobiologie : une science récente	03
1.4	Les glycoconjugués	04
1.4.1	Les glycoprotéines	
1.4.1.1	La <i>N</i> -glycosylation	
1.4.1.2	La <i>O</i> -glycosylation	
1.4.2	Les glycolipides	
1.4.3	Importance biologique des glycoconjugués	
1.5	Méthodes expérimentales pour l'analyse structurale	08
1.5.1	Séquençage de sucres	
1.5.2	Résonance Magnétique Nucléaire	
1.5.3	Diffraction de monocrystal	
1.5.4	Diffraction de fibres	
1.5.5	Co-cristallisation avec des protéines	
1.6	Analyse conformationnelle des glucides : facteurs déterminant la structure tridimensionnelle	10
1.6.1	Monosaccharides	
1.6.2	Disaccharides	
1.6.3	Oligosaccharides	

Chapitre 2 : Lectines et interactions protéines-sucres

2.1	Lectines : généralités	15
2.2	Spécificité et affinité	18
2.3	Structures tridimensionnelles	20
2.4	Sites de reconnaissance	23
2.5	Les lectines bactériennes	24
2.5.1	Les lectines fimbriales	
2.5.2	Les toxines	
2.5.3	Les lectines solubles	
2.6	Pseudomonas aeruginosa	27
2.6.1	PA-IL et PA-IIL	
2.6.2	Rôles de PA-IL et PA-IIL dans le processus d'infection	
2.6.3	Homologie de séquence et de structure chez d'autres microorganismes	
2.7	Les lectines de champignons	34
2.7.1	Caractéristiques	
2.7.2	Rôles	
2.7.3	Lectines connues	

Chapitre 3 : Méthodes expérimentales

3.1	Cristallographie	39
3.1.1	Cristallisation	
3.1.2	Collection des données	
3.1.3	Le phasage	
3.1.3.1	La méthode de Patterson	
3.1.3.2	Les méthodes directes	
3.1.3.3	Le remplacement isomorphe	

3.1.3.4	La diffusion anomale	
3.1.3.5	L'utilisation de sucres modifiés dans le phasage de lectines	
3.1.3.6	Le remplacement moléculaire	
3.1.4	Affinement et validation	
3.2	Analyse conformationnelle des oligosaccharides	48
3.2.1	Champs de force pour les glucides	
3.2.2	Calcul de cartes adiabatiques avec MM3	
3.2.3	Exploration de la surface multidimensionnelle : la méthode SCD	
3.3	Thermodynamique des interactions : microcalorimétrie de titration (ITC)	51
3.3.1	Le microcalorimètre et la technique	
Chapitre 4 : Structure cristalline de la Tricolorine A		55
4.1	Introduction	
4.2	Résultats	
4.3	Crystal structure of Tricolorin A: molecular rationale for the biological properties of resin glycosides found in some mexican herbal remedies	
Chapitre 5 : Analyse conformationnelle des O-glycannes : la méthode CICADA		57
5.1	Introduction	
5.2	Résultats	
5.3	Conformational analysis of complex oligosaccharides: the CICADA approach to the uromodulin O-glycans	
Chapitre 6 : La lectine bactérienne PA-IL de <i>Pseudomonas aeruginosa</i>		59
6.1	Introduction	
6.2	Résultats	
6.3	Structural basis of calcium and galactose recognition by the lectin PA-IL of <i>Pseudomonas aeruginosa</i>	
Chapitre 7 : La lectine bactérienne CV-IIL de <i>Chromobacterium violaceum</i>		61
7.1	Introduction	
7.2	Résultats	
7.3	Molecular basis for unusual entropy driven affinity of <i>Chromobacterium violaceum</i> lectin CV-IIL	
Chapitre 8 : La lectine fongique PVL de <i>Psathyrella velutina</i>		63
8.1	Introduction	
8.2	Résultats	
8.3	β-Propeller crystal structure of <i>Psathyrella velutina</i> lectin: an integrin-like fungal protein interacting with monosaccharides and calcium	
Chapitre 9 : Conclusions générales et perspectives		
9.1	Rôles des lectines dans les infections causées par <i>Pseudomonas aeruginosa</i>	65
9.2	Thermodynamique de la reconnaissance lectine-sucré	66
9.3	Evolution de la multivalence : <i>multimères</i> contre <i>β-propellers</i>	67
Références		71
Annexe I		81

Chapitre 1

Introduction à la structure des glycoconjugués

La glycobiologie est la science qui étudie les structures, les propriétés et les fonctions des glucides dans la matière vivante (Taylor 2003). Ces études ne concernent pas seulement les glucides mais également toutes les autres classes de molécules biologiques (protéines, lipides, nucléotides) qui interagissent avec eux. Le but de la glycobiologie structurale est bien évidemment de déterminer la *structure 3D* et d'établir des modèles pour les *interactions* impliquant ces molécules. Ce premier chapitre est dédié à l'étude structurale des glucides et plus particulièrement des oligosaccharides et glycoconjugués.

1.1 Les glucides : généralités

Les glucides représentent la plus importante source de biomasse sur la planète Terre et leur complexité, surtout en terme de variations structurales, n'a pas d'égale dans les autres classes de polymères biologiques. Du point de vue chimique, les sucres sont des polyhydroxy-cétones ou polyhydroxy-aldéhydes. Les *monosaccharides* ont une formule générale du type $C_xH_{2x}O_x$ avec un nombre d'atomes de carbone de la chaîne principale qui peut varier de trois à dix unités. Une caractéristique générale des sucres est qu'un des atomes de carbone porte une fonction de type céto- ou aldéhyde et pour les structures à cinq (*pentoses*) et six atomes de carbone (*hexoses*), la condensation intramoléculaire est à l'origine de la structure cyclique (*furanose* ou *pyranose*) qui est la forme prédominante en solution. La cyclisation donne origine à un nouveau centre chiral dans la molécule, le carbone anomère C1, qui peut avoir deux configurations appelées α ou β . Ce carbone anomère peut condenser avec un des hydroxyles d'un deuxième monosaccharide pour donner un *disaccharide*. Ainsi, l'addition de monomères résulte en la formation de structures dont la complexité augmente considérablement avec le nombre de résidus. Les structures les plus simples, linéaires ou branchées, sont formées par l'addition de deux à cinq monomères et sont nommées en utilisant les préfixes di-, tri-, tetra- et pentasaccharides. Le terme *oligosaccharides*, concerne des molécules qui contiennent jusqu'à une vingtaine de monosaccharides. Quand la taille et la complexité de la structure deviennent très importantes (plus de 20 résidus) on parle d'une manière générale des *polysaccharides*.

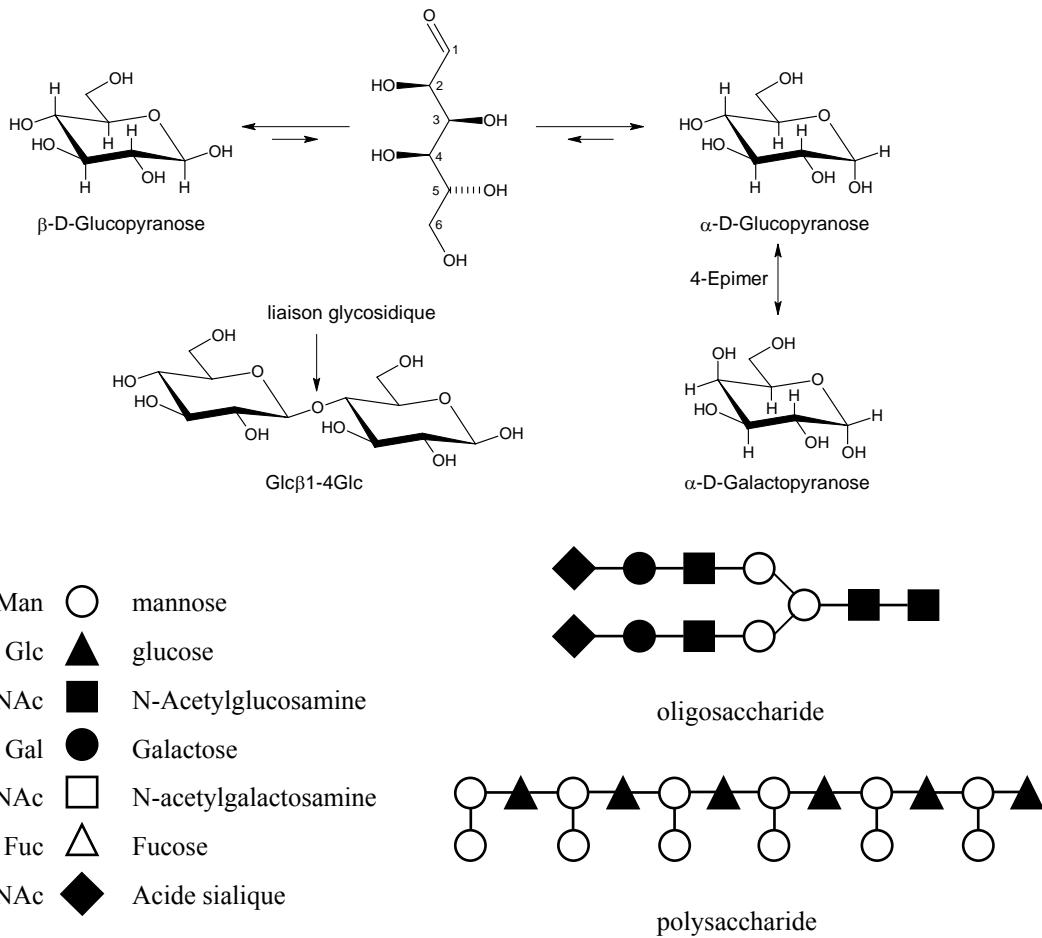


Figure 1.1 Quelques concepts centraux de la structure de glucides

1.2 Relations structure-fonction

L'énorme potentiel biologique codé par les glucides résulte de leurs capacités à créer des structures de grandes complexités, incluant des points de ramification, à partir des simples monomères (Laine). Cette caractéristique des glucides est unique parmi tous les polymères biologiques. Par exemple, deux acides aminés différents peuvent se lier seulement de deux façons différentes. Par contre, un simple monosaccharide comme le glucose peut se lier par son carbone anomère (de façon α ou β) avec les cinq hydroxyles d'un deuxième glucose, donnant naissance à onze isomères structuraux différents. On peut donc comprendre que le nombre de variations structurales assurées par un tel type de molécules est énorme. La structure des glycannes (oligosaccharides, glycoconjugués, polysaccharides) dépend donc fortement de leur composition et de plus, la présence des liaisons glycosidiques introduit une flexibilité structurale et conformationnelle considérable. La conséquence directe de cette variabilité structurale est une grande diversité dans le rôle et dans la localisation des glucides dans les organismes vivants, tels que représentés dans le tableau suivant.

- Renforcement structural (polysaccharides)
- Accumulation d'énergie (polysaccharides)
- Reconnaissance moléculaire (oligosaccharides/glycoconjugués)
- Croissance et développement (oligosaccharides/glycoconjugués)
- Défense (oligosaccharides/glycoconjugués)
- Ancrage (ancre GPI)
- Ciblage (oligosaccharides/glycoconjugués)

Une description exhaustive de la localisation et des fonctions des glucides se situe bien au delà des sujets abordés dans cette thèse et dans mon introduction je me limiterai à une caractérisation générale des glycoconjugués, en portant une attention particulière à leur complexité et aux méthodes utilisées pour les études structurales.

1.3 Glycobiologie : une science récente

C'est seulement à partir de la moitié du vingtième siècle que les glucides ont été reconnus comme des molécules biologiques extrêmement importantes. Il a longtemps été admis que toute l'information biologique était portée par les deux grandes classes de molécules que sont les protéines et les acides nucléiques. Dans les années 1900, on s'intéressait surtout aux polysaccharides des plantes car ils étaient largement utilisés dans de nombreuses applications industrielles. Les connaissances dans la biologie des glucides étaient donc limitées aux rôles structuraux (cellulose, chitine, etc...) et à l'accumulation et au transport d'énergie (amidon, glucose). La prise de conscience que les glycannes, libres ou sous la forme de glycoconjugués, pouvaient jouer des rôles extrêmement divers et de première importance fut la motivation principale qui a donné naissance à la glycobiologie moderne. Des découvertes majeures ont marqué l'élucidation du rôle des oligosaccharides. (i) La structure glucidique des déterminantes du système ABH sanguin a été établie ; (ii) l'interaction entre une hémagglutinine et l'acide sialique a été démontrée lors de l'agglutination des erythrocytes par le virus de la grippe; (iii) il a été observée que la défucosylation des lymphocytes empêche leur migration dans les tissus. En 1968, Ashwell et Morell identifièrent la première lectine animale, une protéine impliquée dans la clairance des glycoprotéines circulant dans le sang (Ashwell 1974). Cette découverte fût à l'origine des nombreuses recherches qui ont depuis été consacrées au rôle joué par les glucides dans les processus de communication et de reconnaissance moléculaire.

1.4 Les glycoconjugués

Les glycoconjugués sont des molécules constituées d'une partie glucidique (mono- ou oligosaccharide) liée de manière covalente à une partie non glucidique (aglycone). Ce sont des composants essentiels des membranes cellulaires, de la matrice extracellulaire et des liquides biologiques et ils sont particulièrement importants dans la vie des organismes pluricellulaires. Selon la nature de la partie aglycone on peut mettre en évidence trois classes différentes de glycoconjugués: les glycoprotéines, les glycolipides et les protéglycans. Je limiterai ma présentation à une brève introduction des deux premiers, en soulignant leur importance biologique.

1.4.1 Les glycoprotéines

La glycosylation des protéines est l'une des modifications post-traductionnelles les plus courantes chez les eucaryotes puisque l'on estime que près de 50% des protéines synthétisées par la cellule seraient glycosylées. Sur la base de la liaison entre la protéine et le glycanne on distingue deux grands types de glycosylation. Dans la *N*-glycosylation une asparagine et un GlcNAc sont invariablement impliqués dans la liaison. Dans la *O*-glycosylation le premier monomère de la chaîne glucidique est lié à un résidu hydroxylé, qui est le plus souvent une sérine ou une thréonine.

1.4.1.1 La *N*-glycosylation

Les *N*-glycannes sont synthétisés dans le réticulum endoplasmique (RE) et ensuite transférés d'un bloc sur un résidu asparagine d'une protéine qui est en train d'être synthétisée. La *N*-glycosylation est donc une modification co-traductionnelle (Pan 1990; Abeijon 1992). La maturation des *N*-glycannes se fait dans l'appareil de Golgi par l'action concertée de glycosyltransférases et de glycosidases. Chez les mammifères, on distingue trois grandes classes des structures différentes : les *N*-glycannes de type oligomannose, les *N*-glycannes de type hybride et les *N*-glycannes de type complexe (Fig. 1.2).

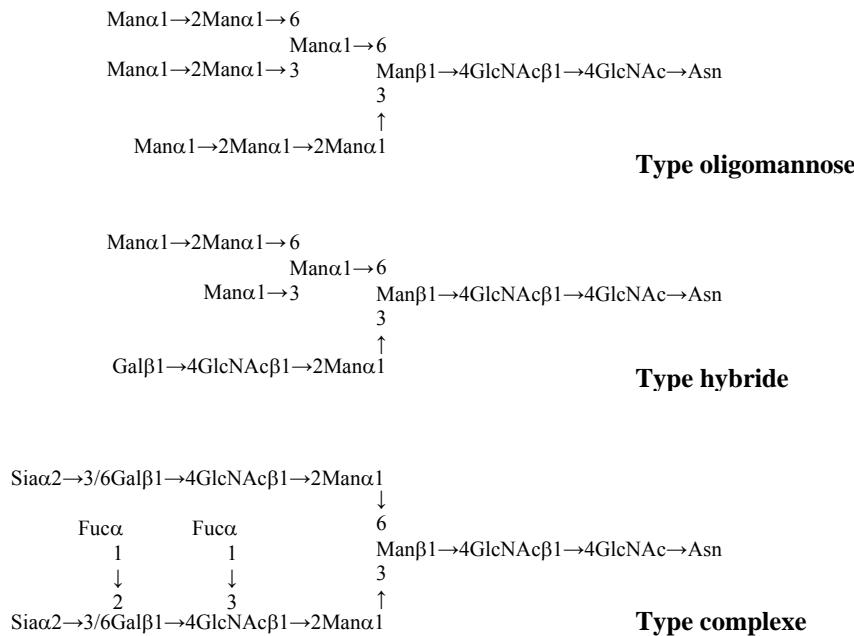


Figure 1.2 Les trois familles de *N*-glycannes identifiés chez le mammifères

1.4.1.2 La *O*-glycosylation

La *O*-glycosylation est une modification post-traductionnelle qui est effectuée dans le Golgi. Contrairement à la *N*-glycosylation, la chaîne *O*-glycanne n'est pas pré-assemblée mais chaque monomère est rajouté sur la chaîne en formation de manière individuelle et séquentielle (Brockhausen 1995). Les structures des *O*-glycannes sont beaucoup plus variables que les *N*-glycannes par la présence de séquences internes (*cores*) différentes. Par exemple, huit séquences différentes ont été identifiées dans les O-glycannes de type mucine. Par l'action de nombreuses glycosyltransférases, chaque séquence est ensuite décorée avec des motifs de glycosylation très spécifiques.

1.4.2 Les glycolipides

Les glycolipides sont formés par une ou plusieurs unités monosaccharidiques, liées de façon covalente à une partie hydrophobe comme un céramide, un sphingoside, un acylglycérol ou un prényl phosphate. Dans les deux premiers cas, ces glycolipides sont des glycosphingolipides et peuvent être de type neutre ou acide. Quand la partie lipidique possède un groupement inositol, ils sont appelés glycophosphatidylinositol. Quelques exemples sont montrés dans la Fig. 1.3. De nombreux glycolipides différents ont été identifiés

et leur structure montre souvent des caractéristiques intéressantes. Dans le quatrième chapitre on présentera la structure d'une glycorésine végétale qui présente une forte activité biologique : la Tricolorine A.

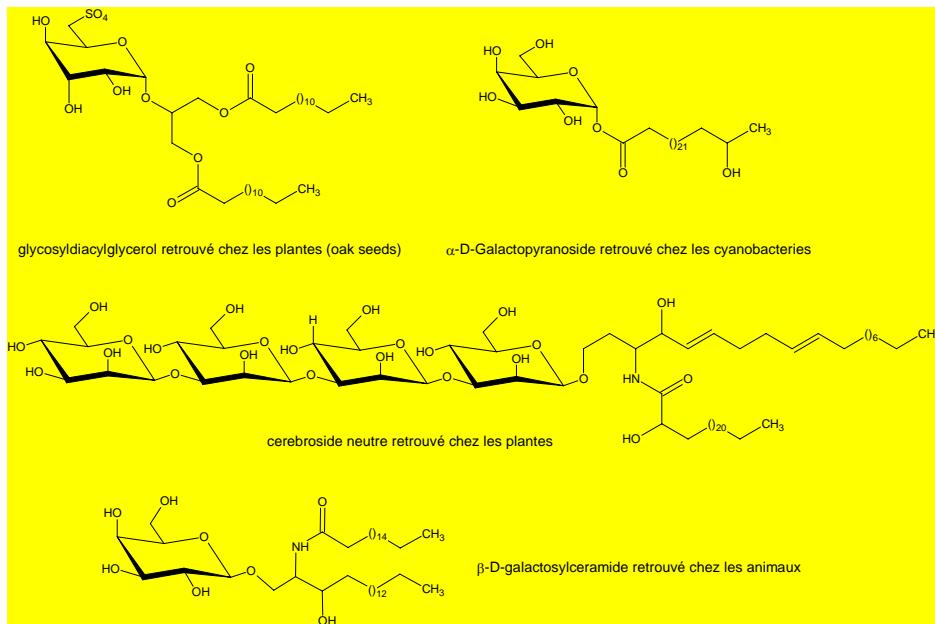


Figure 1.3 Quelques exemples de glycolipides présents chez les procaryotes, les plantes et les animaux

1.4.3 Importance biologique des glycoconjugués

Les fonctions biologiques associées à la partie glucidique des glycoconjugués sont aussi nombreuses que leur diversité structurale. Cependant deux grandes familles des fonctions peuvent être mise en évidence : la fonction structurale et celle de reconnaissance moléculaire.

- **La fonction structurale et de modulation des propriétés physiques**

Ces molécules possèdent souvent une charge électrostatique élevée, ainsi qu'une surface importante et une hydratation importante. Ces propriétés physiques leurs confèrent un rôle de stabilisation structurale, de solubilisation et de protection des glycoprotéines contre la protéolyse. Les glycoconjugués font aussi partie de la couche de haute densité moléculaire, le *glycocalix*, qui recouvre la surface des cellules épithéliales chez les eucaryotes. Cette zone participe à la cohésion cellulaire et à la protection de la cellule contre les chocs physiques et contre l'attaque de microorganismes.

- **Les rôles de reconnaissance moléculaire**

Les oligosaccharides participent à de nombreux phénomènes de reconnaissance moléculaire de type récepteur-ligand impliquant des lectines et de nombreuses études confirment leur rôles dans la vie sociale des cellules (Varki 1993). Comme on le verra dans le prochain chapitre, les lectines reconnaissent généralement les monomères terminaux des oligosaccharides. Au niveau des structures périphériques des glycoconjugués (surtout *O*-glycannes et glycolipides), on trouve souvent des épitopes antigéniques comme les déterminants des groupes sanguins ABO (antigènes A, B, H) et Lewis (Lewis a, b, X et Y) ainsi que leurs dérivés sialylés. Par ailleurs, il a été montré que des anomalies de glycosylation accompagnent fréquemment les états physiologiques anormaux ou pathologiques, comme le cancer. En conséquence, ce sujet reçoit maintenant une grande attention et de nombreuses études ont été effectuées dans le but d'élucider les structures complexes et les rôles de ces glycannes.

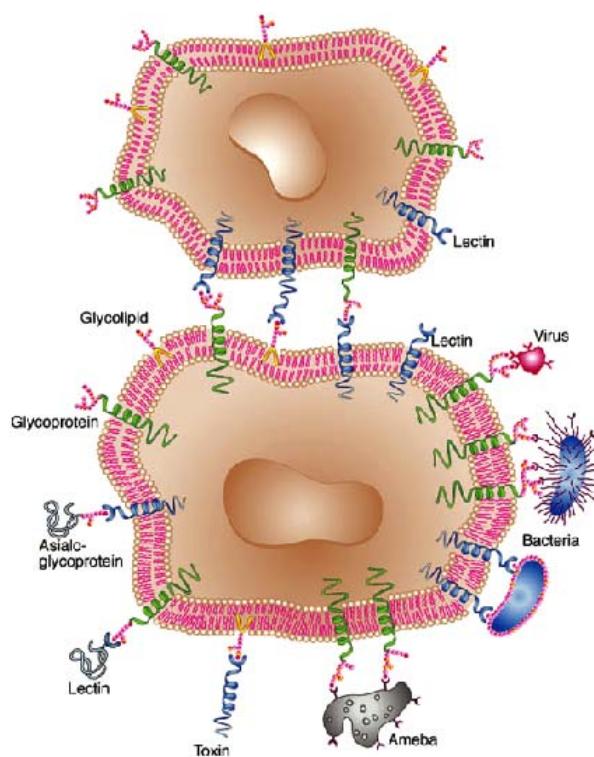


Figure 1.4 Exemple de rôles de glycoconjugués situés sur la surface cellulaire

1.5 Méthodes expérimentales pour l'analyse structurale

1.5.1 Séquençage de sucres

Par analogie avec le séquençage des protéines, la purification et la détermination de la structure primaire d'un oligosaccharide sont les premières étapes dans la connaissance de sa structure. Malgré les efforts et les progrès récents dans le domaine de la spectroscopie, le séquençage des glucides complexes reste un travail long et difficile. Les nouvelles techniques de spectroscopie de masse, qui combinent une haute sensibilité avec l'analyse automatisée des résultats se sont révélées assez puissantes et fiables. Très prometteuse aussi est la combinaison des techniques SM et RMN qui fournissent des informations complémentaires pour la détermination de la séquence primaire. La connaissance préalable des structures semblables mais également des voies possibles de biosynthèse sont d'une grande utilité dans la détermination d'une nouvelle structure (Pérez 2005).

1.5.2 Résonance Magnétique Nucléaire

Les deux seules techniques dont on dispose pour la détermination expérimentale de la structure 3D des glucides sont la RMN et la diffraction des rayons X. En solution, la méthode privilégiée est bien évidemment la RMN (1D, 2D et 3D). Les techniques les plus utilisées sont la mesure des déplacements chimiques, des temps de relaxation, des constantes de couplage et des effets Overhauser (NOEs, ROEs ...). L'inconvénient majeur est du à la présence des liaisons glycosidiques flexibles qui, surtout dans les structures complexes, sont à l'origine de plusieurs conformations alternatives en solution, résultant dans une conformation *moyennée en fonction du temps*. Cependant, l'existence de plusieurs conformations ainsi que leur interconversion peuvent être mises en évidence, ce qui peut donner des informations importantes sur la *dynamique* du système.

1.5.3 Diffraction de monocristal

Une très grande partie des données structurales vient de la cristallographie aux rayons X. Plus de 4000 structures de glucides sont présentes dans la Cambridge Structural Database (CSDB), dont presque la totalité est représentée par des monosaccharides ou des dérivés. Seulement 17 trisaccharides et 5 tétrasaccharides sont présents dans la base de données (Imberty and Pérez 2000). Cette disparité est très révélatrice de la principale difficulté rencontrée dans l'analyse cristallographique des glucides : réussir la cristallisation d'une molécule plus grande qu'un disaccharide est en fait assez rare. La présence d'un équilibre conformationnel au niveau du carbone anomére ($\alpha \leftrightarrow \beta$) et l'inhérente flexibilité de la liaison glycosidique sont les facteurs

principaux qui limitent la formation d'un réseau cristallin ordonné. La difficulté à cristalliser les glucides est encore plus accentuée pour les structures ayant un poids moléculaire situé entre 1000 et 5000 Da, à l'exception des cyclodextrines. Les raisons peuvent être recherchées dans la faible quantité d'échantillon généralement disponible pour les essais de cristallisation mais aussi dans le retard dont souffre la cristallographie des « petites » molécules qui n'a pas suivi les énormes développements techniques et méthodologiques réalisés dans le domaine des macromolécules biologiques. C'est dans ce contexte qu'on présentera dans le quatrième chapitre la structure de la Tricolorine A, qui est l'un des plus grands oligosaccharides dont la structure a été résolue par diffraction X.

1.5.4 Diffraction de fibres

La diffraction de fibres est la méthode la plus utilisée pour l'étude structurale des polysaccharides, mais contrairement à la diffraction de monocristaux, les données de diffraction de fibre ne sont généralement pas suffisantes pour la détermination de tous les paramètres structuraux. Les techniques de modélisation doivent être utilisées pour la construction d'un modèle qui va être comparé avec les données de diffraction obtenues. Le nombre de structures de polysaccharides résolues est autour d'une centaine (Chandrasekaran 1997).

1.5.5 Co-cristallisation avec des protéines

Dans les dernières années nous avons assisté à une augmentation considérable du nombre de structures de protéines résolues. Comme conséquence, de nombreuses structures de glycoprotéines et de lectines complexées avec des glucides ont été résolues et leurs coordonnées atomiques sont disponibles dans la Protein Data Bank (PDB). Dans le cas des glycoprotéines, la flexibilité des N- ou O-glycannes est un facteur qui peut limiter ou empêcher la cristallisation. Généralement dans les structures de glycoprotéines obtenues par diffraction on peut observer seulement les premiers deux ou trois monosaccharides de la chaîne glycosidique (Imberty and Pérez 1995). Par contre, plus de 500 structures de lectines ont été résolues (<http://www.cermav.cnrs.fr/lectines>), souvent en complexe avec des ligands oligosaccharidiques pour lesquels aucune information structurale n'était disponible auparavant.

1.6 Analyse conformationnelle des glucides : facteurs déterminant la structure tridimensionnelle

Les méthodes de la chimie computationnelle sont devenues extrêmement importantes dans l'analyse conformationnelle des molécules organiques et notamment des glucides complexes. L'augmentation constante de la puissance de calcul des ordinateurs a permis de relever de nouveaux défis, notamment l'exploration des surfaces multidimensionnelles, qui étaient techniquement irréalisables auparavant.

Je voudrais ici attirer l'attention sur les facteurs qui déterminent la structure tridimensionnelle de glucides mais aussi sur les difficultés rencontrées dans l'analyse conformationnelle des structures complexes. Les méthodes de calcul seront décrites dans le troisième chapitre.

1.6.1 Monosaccharides

Dans l'analyse conformationnelle des oligosaccharides, la chaîne principale est généralement considérée comme formée de monomères rigides même si les pyranoses sont dotés d'une certaine flexibilité au niveau du cycle. Pour la plupart des monosaccharides, la conformation de plus basse énergie est la forme *chaise* et plus précisément 4C_1 et 1C_4 pour les glucides de série D ou L respectivement (Rao 1988). Certains monosaccharides tels que les cycles idoses peuvent de plus adopter des conformations de type sofa (2S_0). Le groupement hydroxyméthyle exocyclique peut tourner autour de la liaison C5-C6 mais l'analyse cristallographique a montré qu'il adopte trois conformations de basse énergie, nommées *gt*, *gg* et *tg*, pour limiter les interactions avec le O4 et le O5 (Fig. 1.5) (Marchessault and Pérez 1979; Pérez 2000). Par contre, les hydroxydes secondaires sont libres de tourner presque librement même si deux orientations relatives (*sens des aiguilles d'une montre ou sens inverse*), montrées dans la Fig. 1.5, favorisent la formation des réseaux de liaisons hydrogène intramoléculaires (Brady 1993). L'ensemble de ces paramètres doit être pris en compte dans l'analyse structurale des structures complexes.

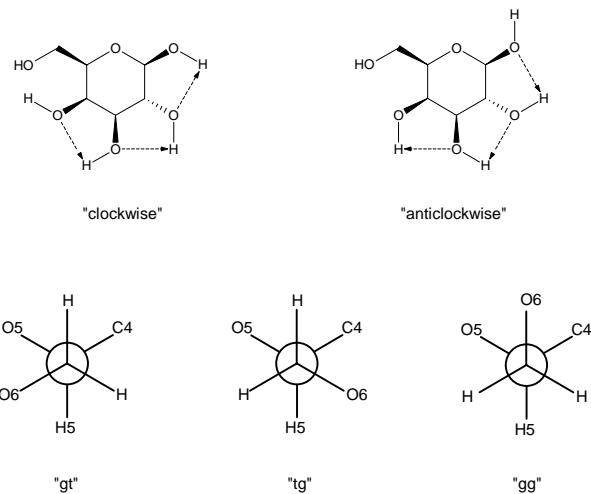


Figure 1.5 Réseaux des liaisons hydrogène pour le galactose et conformation du O6 en projection de Newman

1.6.2 Disaccharides

Nous avons indiqué plus haut que deux pyranoses peuvent se combiner pour donner onze disaccharides différents. La nature de la liaison glycosidique (α ou β , 1 \rightarrow 1, 1 \rightarrow 2, 1 \rightarrow 3, 1 \rightarrow 4 ou 1 \rightarrow 6) est le paramètre clé qui détermine la structure tridimensionnelle d'un disaccharide. En fait, la molécule présente une certaine flexibilité autour des deux angles de torsion (Φ et Ψ) de la liaison glycosidique, mais il a été montré que toutes les conformations ne peuvent pas être atteintes avec la même probabilité. Les deux angles de torsion glycosidiques qui définissent la géométrie de la liaison sont illustrés dans la Fig. 1.6.

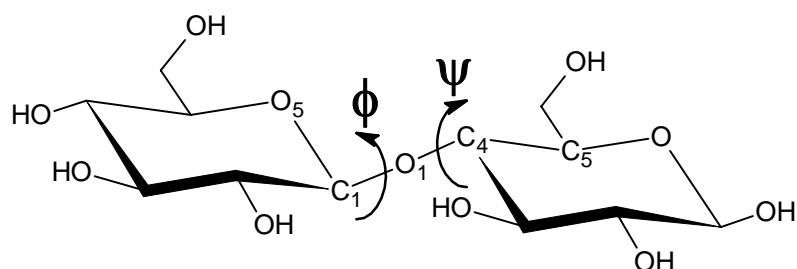


Figure 1.6 : Les angles de torsion glycosidique pour Glc β 1 \rightarrow 4Glc

L'angle ϕ (O5—C1—O1—Cx) est influencé en premier lieu par l'effet *exoanomère*, qui est un effet stéréo-électrique résultant des interactions entre les orbitales des atomes O5,C1,O1 et Cx, impliqués dans la liaison glycosidique (Lemieux 1979). L'angle ψ (C1—O1—Cx—C(x+1)) montre une variabilité plus grande et est plutôt influencé par les interactions non-liées (liaisons hydrogène, interactions de van der Waals) entre les deux monosaccharides. La structure globale d'un disaccharide est donc déterminée par la nature de la liaison et par l'ensemble des ces interactions.

La probabilité d'existence pour une conformation donnée d'un disaccharide peut être représentée sur un diagramme bidimensionnel (Fig. 1.7) qui représente la valeur de l'énergie potentielle de la molécule pour chaque combinaison des valeurs de ϕ et ψ . La méthode de calcul utilisée pour la génération de ces cartes sera décrite dans le troisième chapitre.

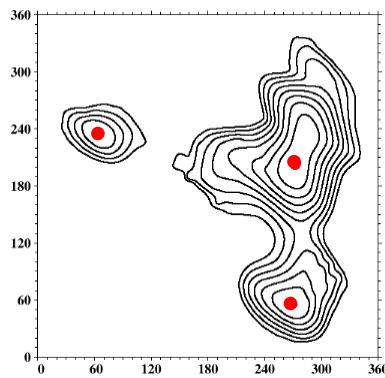


Figure 1.7 Carte adiabatique d'énergie potentielle pour le disaccharide Glc β 1→4Glc. Chaque isocontour représente un niveau d'énergie de 1kcal. Les conformations de plus basse énergie sont indiquées par des point rouges.

On peut voir dans la figure que les conformations de plus basse énergie pour la liaison Glc β 1→4Glc sont confinées dans des régions relativement limitées de l'espace conformationnel. La comparaison des valeurs ϕ et ψ obtenues par cristallographie et RMN a validé cette méthode (Imbert 1997). Des centaines de cartes ont déjà été calculées et un grand nombre sont disponibles dans la base de données de disaccharides du CERMAV (www.cermav.cnrs.fr/disaccharides).

1.6.3 Oligosaccharides

Les oligosaccharides sont des structures généralement flexibles qui peuvent adopter différentes conformations (ou plus souvent familles de conformations) de basse énergie. La liberté conformationnelle peut être limitée par la présence d'interactions (liaisons hydrogène, interactions de van der Waals) entre deux monomères qui ne sont pas forcément liés par une liaison glycosidique. Par exemple, dans le cas particulier des oligosaccharides du type Lewis X, la forte interaction hydrophobe de *stacking* entre le Gal et le Fuc rigidifie la structure de l'oligosaccharide qui n'adopte en fait qu'une seule conformation en solution (Imbert and Pérez 2000).

Dans l'analyse conformationnelle l'oligosaccharide peut être considéré comme constitué d'un ensemble de disaccharides. Les valeurs de ϕ et de ψ pour chaque disaccharide seront donc, en

première approximation, considérées comme indépendantes des résidus voisins. Ceci est une simplification importante mais qui permet de dire que la structure globale d'un oligosaccharide dépend en premier lieu de la nature de chaque liaison entre ses disaccharides constituants. Malgré ça, l'analyse conformationnelle des oligosaccharides reste un travail long et difficile car le nombre des paramètres mis en jeu est généralement très important. Pour un simple trisaccharide, comme celui montré en Fig. 1.8, on doit prendre en compte six conformations différentes pour chaque monomère (trois pour le C6 et deux pour les hydroxyles) et 24^4 combinaisons pour les angles de torsion $\phi, \psi, \phi_1, \psi_1$ pour un totale de 71663616 conformations possibles !

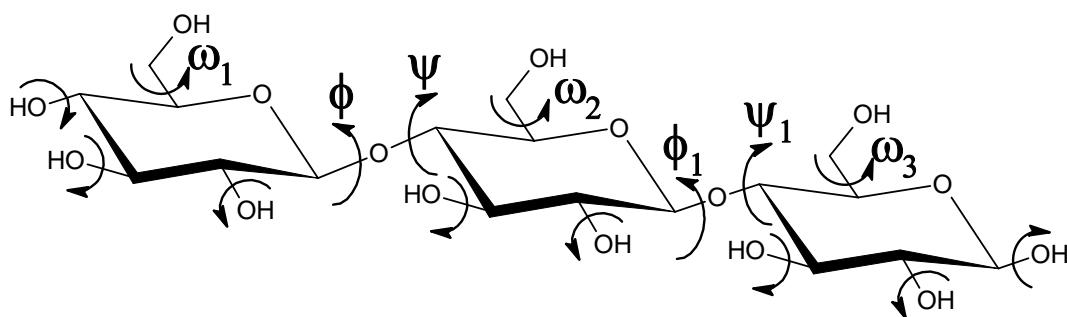


Figure 1.8 : Espace conformationnel pour un trisaccharide $\text{Glc}\beta 1\rightarrow 4\text{Glc}\beta 1\rightarrow 4\text{Glc}$

Bien qu'il s'agisse d'un simple trisaccharide, un tel nombre de minimisations d'énergie nécessite déjà une puissance de calcul non négligeable. On peut donc comprendre que pour l'analyse conformationnelle de glycannes encore plus complexes (plus de 10 monomères) on a forcément besoin d'une stratégie d'exploration de l'espace conformationnel très puissante et intelligente. Dans le cinquième chapitre on présentera l'étude conformationnelle d'une structure d'un *O*-glycanne qui contient 366 atomes et quatorze liaisons glycosidiques. Cette étude porte sur un des plus grands oligosaccharides jamais soumis à une exploration complète de l'espace conformationnel.

Chapitre 2

Lectines et interactions protéine-sucre

2.1 Lectines : généralités

Le mot *lectine* dérive du verbe latin *legere* qui veut dire « sélectionner » ou « choisir », un nom bien approprié pour cette très importante classe de protéines. Les lectines ont été définies comme des protéines d'origine non-immune qui se lient spécifiquement et de façon réversible aux sucres et ne montrent aucune activité enzymatique pour ces substrats (Kocourek 1981). Les lectines sont des molécules ubiquitaires, car elles se retrouvent dans toutes les classes d'organismes, chez les microorganismes (virus, bactéries), chez les plantes, chez les insectes et les animaux. Elles sont aussi appelées agglutinines car elles sont capables d'agglutiner les cellules (comme les érythrocytes) et les glycoconjugués. Cette caractéristique très importante des lectines est due au fait que ces protéines sont généralement multivalentes. Les méthodes anciennement utilisées pour leur identification consistent à mélanger l'extrait à examiner (dérivé par exemple d'un tissu ou d'une plante) avec des érythrocytes en solution. L'agglutination ou la précipitation des cellules indique que la solution analysée contient une (ou parfois plusieurs) molécules agglutinantes. Les premières lectines furent identifiées chez les plantes au début du vingtième siècle mais la communauté scientifique ne commença à s'intéresser à cette classe de protéines qu'à partir des années soixante, en concomitance avec la naissance de la glycobiologie. L'intérêt était motivé surtout par l'utilisation des lectines dans la détection, l'isolation et la caractérisation d'oligosaccharides, tels que les déterminants des groupes sanguins et de glycoconjugués, surtout des glycoprotéines. La découverte majeure que certains états physiologiques et pathologiques étaient associés à un changement de l'état de glycosylation des cellules fut possible grâce à l'utilisation des lectines. Le nombre de travaux publiés sur les lectines a vu une grande croissance principalement grâce à l'abondance des lectines dans tous les organismes vivants, accompagné d'une certaine facilité de purification. Dans les derniers temps, on a aussi commencé à considérer les lectines comme des molécules bioactives et on s'est de plus en plus intéressé aux rôles biologiques de ces molécules. On s'est aperçu tout de suite de l'extrême variabilité des lectines, d'abord en termes de structure primaire, et à la suite des premières structures résolues par diffraction des rayons X, en termes de structure tridimensionnelle.

L'intérêt majeur qui pousse aujourd'hui la recherche sur les lectines est lié sans doute à leur capacité unique de « lire » l'information biologique qui est codifiée dans la structure tridimensionnelle des sucres. Les lectines sont en fait les récepteurs spécifiques pour les interactions protéine-sucre qui jouent des rôles clés dans une multitude de processus de reconnaissance moléculaire et de signalisation cellulaire. Par exemple, l'hémagglutinine du virus de la grippe reconnaît et se lie aux oligosaccharides terminés par un acide sialique qui sont situés sur la surface des cellules épithéliales des voies aériennes supérieures (Wiley 1987). De la même façon, les lectines situées sur la surface des bactéries, des virus ou des parasites intestinaux reconnaissent les glycoconjugués présents sur la surface des cellules épithéliales et donc facilitent les processus de colonisation et d'infection. Les rôles joués par les lectines des plantes restent toujours un mystère. Une hypothèse très probable est le rôle de défense du végétal contre les phytopathogènes (ou contre les animaux qui peuvent se nourrir de la plante), mais d'autres possibilités ont été considérées (Rudiger 2001). Par contre, les rôles des lectines animales ont été étudiés en détail comme l'indique les exemples cités ci-dessous. L'adhésion des lymphocytes sur la paroi interne des vaisseaux sanguins est un processus important, qui précède leur extravasation vers les organes lymphoïdes ou vers les zones infectées. Des lectines appelées *selectines* situées sur la surface des vaisseaux sanguins reconnaissent les glycoconjugués présents sur la surface des lymphocytes (Bevilacqua 1991). Les lectines animales ont des rôles particulièrement importants dans la croissance et dans le développement des organismes supérieurs. La toute première interaction qui s'effectue pendant les processus de fertilisation est la reconnaissance entre une lectine du spermatozoïde (*spermadhesine*) et un glycoconjugué présent sur la surface des ovocytes (Topfer-Petersen 1998). Les exemples qu'on pourrait citer sont très nombreux et cette grande variation de fonction pour chaque classe de lectines est bien évidente dans le tableau 2.1.

Tableau 2.1 Classification et rôles des lectines dans les organismes vivants

Lectine	Rôles
<i>Bactéries</i>	
Lectines fimbriales	Adhésion, infection
Lectines solubles	Adhésion, infection, formation de biofilm
Toxines	Adhésion , infection
<i>Virus</i>	
Influenza haemagglutinine	Adhésion, infection
<i>Amoeba</i>	
Lectines de surface	Adhésion
<i>Plantes</i>	
Légumineuses	Défense, symbiose avec les bactéries fixant l'azote
Autres	Défense
<i>Animaux</i>	
Calnexine	Contrôle de la biosynthèse des glycoprotéines
M-type	Dégradation des protéines dans le RE
L-type	Contrôle de la biosynthèse des glycoprotéines
P-type	Régulation de la croissance cellulaire et de l'apoptose; régulation du cycle cellulaire; modulation des interactions cellule-cellule et cellule-substrat
C-type	Adhésion cellulaire (selectines), Clairance des glycoprotéines, Réponse immunitaire (collectines)
Galectines	Reconnaissance des glycans dans la matrice extracellulaire
I-type	Adhésion cellulaire dans le CNS (siglecs)
R-type	Ciblage des enzymes, régulation du turnover des hormones

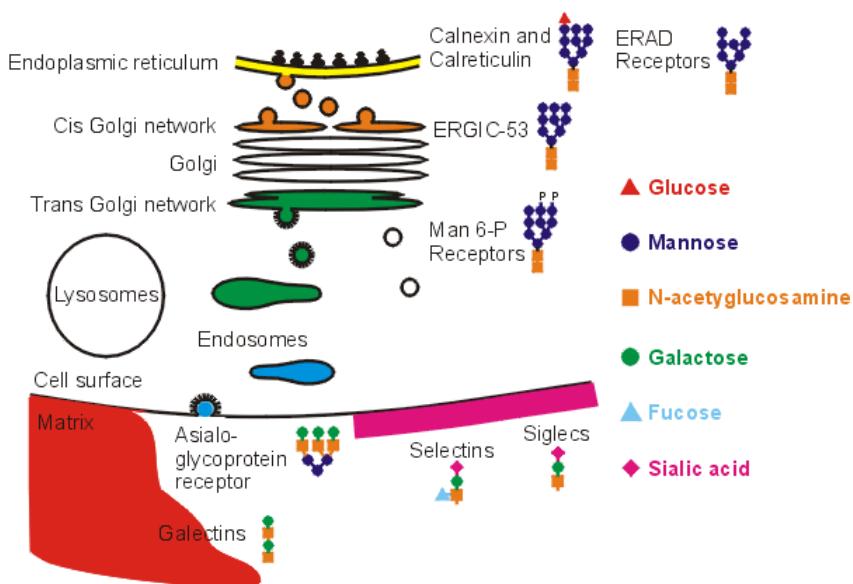


Image 2.1 Localisation cellulaire et spécificité des lectines animales

(<http://www.imperial.ac.uk/research/animallectins/>)

2.2 Spécificité et affinité

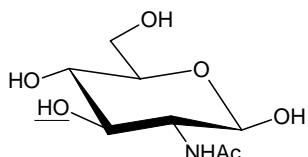
Il est intéressant de noter que la plupart des lectines sont spécifiques pour un petit nombre de sucres et que, dans la majorité des cas, ces sucres sont présents dans et sur la surface des cellules, surtout sous la forme de glycoconjugués. On peut identifier deux classes de lectines par rapport à leur spécificité : celles qui reconnaissent un monosaccharide spécifique et celles qui reconnaissent exclusivement des oligosaccharides (Sharon 2003). Parmi les monosaccharides le plus souvent reconnus par les lectines on retrouve le mannose, le fucose, le galactose/GalNAc et la N-acetylglucosamine. Très peu de lectines reconnaissent l'acide sialique sous la forme de monosaccharide. (Table 2.2)

Tableau 2.2 Lectines spécifiques pour les monosaccharides (Sharon 2003)

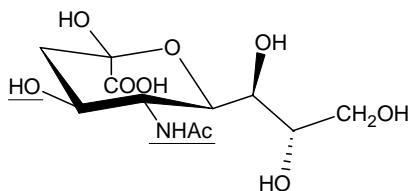
<i>Monosaccharide</i>	<i>Lectine</i>
Man	<i>Allium sativum</i> ; <i>Canavalia ensiformis</i> ; <i>Crocus sativus</i> ; <i>Dioclea grandiflora</i> ; <i>E.Coli</i> type 1 fimbriae ; ERGIC-53 ; <i>Galanthus nivalis</i> ; MBLs of animals ; <i>Pisum sativum</i> ; <i>Vicia faba</i>
Fuc	<i>Aleuria aurantia</i> ; <i>Anguilla anguilla</i> ; <i>Lotus tetragonolobus</i> ; <i>Pseudomonas aeruginosa</i> lectin II ; <i>Ulex europaeus</i> lectin I ; <i>Ulva lactuca</i> ; <i>Chromobacterium violaceum</i> lectin
Gal/GalNAc	<i>Arachis hypogaea</i> ; <i>Coprinus cinereus</i> ; <i>Entamoeba histolytica</i> ; <i>Erythrina corallodendron</i> ; <i>Dolichos biflorus</i> ; <i>Glycine max</i> ; <i>Griffonia simplicifolia</i> lectin I ; <i>helix pomatia</i> ; <i>Hygrophorus hypothejus</i> ; <i>Phaseolus limensis</i> ; <i>Moluccella laevis</i> ; <i>Polyandrocarpa misakiensis</i> ; <i>Ptilota filicina</i> ; <i>Ricinus communis</i> ; <i>Pseudomonas aeruginosa</i> lectin I
GlcNAc	<i>Conglutinin</i> ; <i>Griffonia simplicifolia</i> lectin II ; <i>Tachylectin-2</i> ; <i>Triticum aestivum</i> ; <i>Ulex europaeus</i> lectin II ; <i>Psathyrella velutina</i>
NeuNAc	<i>Achatina fulica</i> ; <i>Cancer antennarius</i> ; <i>Hericium arinaceum</i> ; <i>Homarus americanus</i> lectin I ; <i>Limax flavus</i> ; <i>Scylla serrata</i> ; <i>Triticum aestivum</i> ; <i>Psathyrella velutina</i>

A part quelques exceptions, ces lectines n'ont pas une très grande affinité pour leurs ligands monosaccharidiques, avec une constante d'affinité K_a qui est de l'ordre de 10^3 - 10^4 M $^{-1}$. En général la spécificité est forte pour un monosaccharide et peu de substitutions sont autorisées. L'affinité est généralement modulée par la présence d'un substituant sur le carbone anomérique du monosaccharide et donc ces lectines reconnaissent aussi des di- ou trisaccharides. Certains monosaccharides montrent des similarités structurales importantes, comme on peut le voir dans la figure 2.2. Dans le GlcNAc et dans l'acide sialique le groupement NHAc et le groupement OH adjacent ont la même orientation spatiale (*trans*). Cette situation est encore plus évidente dans les cas du mannose et du fucose, qui montrent trois hydroxyles (2,3,4 chez le mannose et 4,3,2 chez le fucose) dans la même configuration

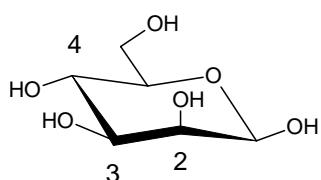
spatiale. De nombreuses lectines qui reconnaissent le mannose peuvent ainsi se fixer au fucose (ou *vice-versa*), bien qu'avec une affinité plus ou moins importante. Deux exemples d'interactions non-spécifiques sont la lectine bactérienne de *Chromobacterium violaceum* (spécificité: fucose/mannose) et la lectine de champignon *Psathyrella velutina* (spécificité: GlcNAc/NeuNAc), qui ont été étudiées au cours de cette thèse.



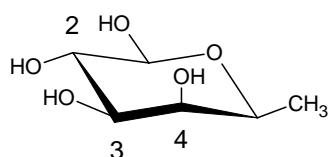
N-Acetylglucosamine



N-Acetylneuraminic acid



Mannose



L-Fucose

Figure 2.2 Similarités structurales entre monosaccharides.

La deuxième classe de lectines est formée par celles qui nécessitent un oligosaccharide spécifique, comme par exemple la cyanovirine-N qui reconnaît des glycannes de type oligomannose tels que la glycoprotéine gp120 du virus VIH (Botos 2002) ou la toxine de choléra qui est spécifique pour le GM1 présent sur la surface de cellules épithéliales (Merritt 1994). L'affinité montrée pour les oligosaccharides est généralement beaucoup plus élevée que pour les monosaccharides par la présence de sites de liaison plus profonds formant plus de contacts avec le ligand. Ces dernières lectines montrent parfois des K_a tout à fait remarquables qui peuvent atteindre le 10^7 - 10^8 M^{-1} .

2.3 Structures tridimensionnelles

La première structure cristalline d'une lectine à être déterminée par diffraction des rayons X fut la concanavaline A (Edelman 1972; Hardman 1972). De 1972 à nos jours, plus de 500 structures cristallines de lectines ont été déterminées, la plupart d'entre elles co-cristallisées avec leurs ligands (<http://www.cermav.cnrs.fr/lectines>) et ce nombre est en augmentation constante.

Un concept important sur la structure moléculaire des lectines est celui du *Carbohydrate Recognition Domain* (CRD), introduit pour la première fois chez les lectines animales (Drickamer 1988). Il s'agit généralement d'une séquence peptidique invariante ou très bien conservée qui forme le domaine structural minimal nécessaire à la reconnaissance du ligand. La multivalence, propriété fondamentale pour l'activité des lectines, peut être réalisée par différentes stratégies (figure 2.3). Quand un seul site de liaison est présent sur chaque chaîne peptidique la structure quaternaire d'une lectine est généralement un multimère formé par l'association non covalente de deux (dimère) à cinq sous-unités (pentamère). Il existe aussi des lectines où le domaine de reconnaissance est répété plusieurs fois dans la structure, comme dans le cas des lectines qui montrent une structure du type *β -propeller*. Une troisième classe de lectines montre la présence de deux CRD distincts reliés par un domaine de jonction (*linker*). Les deux domaines peuvent montrer des spécificités différentes ou bien l'un d'entre eux peut avoir une activité enzymatique (ou autre) comme dans le cas des lectines appelées *chimerolectines*.

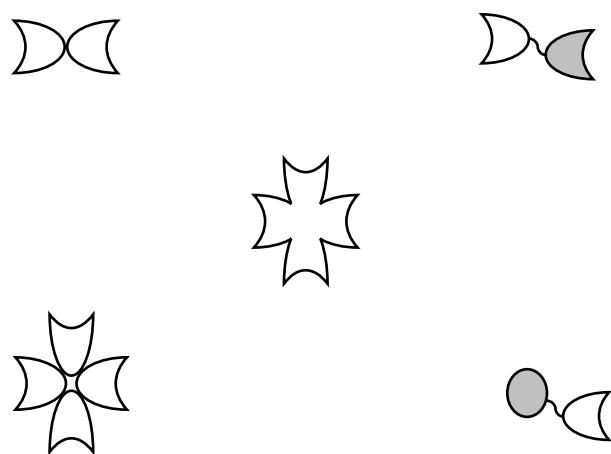


Figure 2.3 Exemple de structure quaternaire de lectines

Les lectines de plantes qui appartiennent à une même famille taxonomique (légumineuses, céréales, etc.) montrent une homologie de séquence/structure remarquable qui a permis de les classifier en sept familles de protéines structurellement similaires (Van Damme 1998). Les lectines de légumineuses adoptent toutes une structure typique appelée « β jelly roll » ou « lectin fold » formée par deux feuillets β superposés. Les lectines de céréales sont très différentes et sont caractérisées par la présence de domaines structuraux conservés très riches en cystéine qui sont appelés domaine hévéine du nom d'une petite protéine de 43 acides amines extraite de l'hévéa. La première structure cristallographique déterminée dans cette famille est la lectine de germe de grain de blé WGA (Fig 2.4), spécifique pour les oligosaccharides à NeuAc ou à GlcNAc (Wright 1980). La ricine, ainsi qu'un grand nombre de toxines végétales, est constituée d'une lectine en trèfle (ricine B) associée à une toxine (ricine A) (Rutenber 1991). La famille de la lectine de perce neige (Hester, Kaku et al. 1995) ainsi que la famille de la jacaline (Sankaranarayanan 1996) adoptent toutes deux des repliements protéiques de type prisme β qui peuvent former une grande variété d'oligomères (Gallego del Sol 2005).

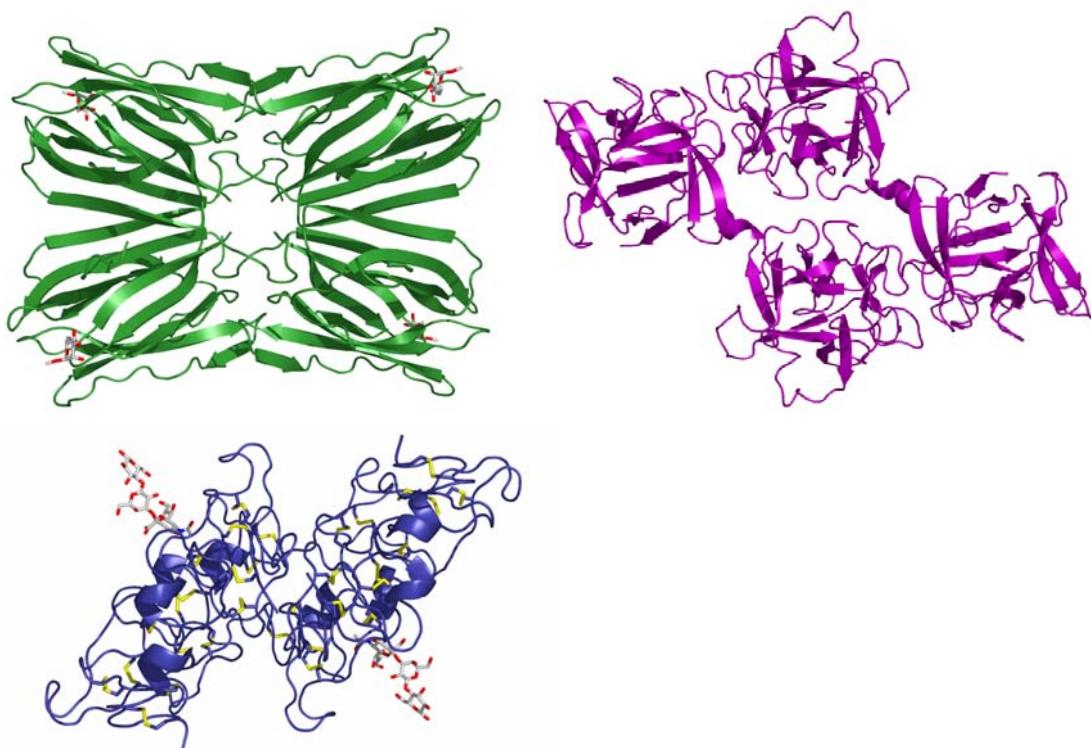


Figure 2.4 Jacaline en complexe avec l' α -Me-mannose (gauche, pdb 1ws5), lectine en trèfle de *Amaranthus caudatus* (droite, pdb 1jly) et WGA en complexe avec le sialyllactose (en bas, pdb 1wgc)

Les lectines animales montrent une grande variabilité structurale et des similarités importantes même entre différentes espèces. Dans l'impossibilité d'adopter une classification du type taxonomique (comme pour les lectines de plantes), elles ont été regroupées dans différentes familles ayant en commun des caractéristiques structurales. Les trois plus grandes familles de lectines animales sont les *galectines*, les lectines de type C et les lectines de type-I (*siglecs*). Les autres lectines animales ont été résumées dans le tableau 2.1. La famille des *galectines* regroupe des lectines solubles qui reconnaissent le β -Gal, et sont caractérisées par la présence d'un domaine très bien conservé appelé *S-type carbohydrate recognition domain* (S-CRD) dont l'intégrité semble indispensable à l'activité de la lectine (Leffler, Carlsson et al. 2004). La grande famille des lectines du type C (*C-type lectins*) est formée par les lectines dont l'interaction sucre-protéine se fait par l'intermédiaire d'un atome de calcium (Drickamer 1999). Les siglecs reconnaissent l'acide sialique et leur CRD adopte un repliement de type immunoglobuline (Crocker 2002). La structure des lectines bactériennes sera le sujet du paragraphe 2.5 et les lectines de champignon seront décrites dans le paragraphe 2.7.

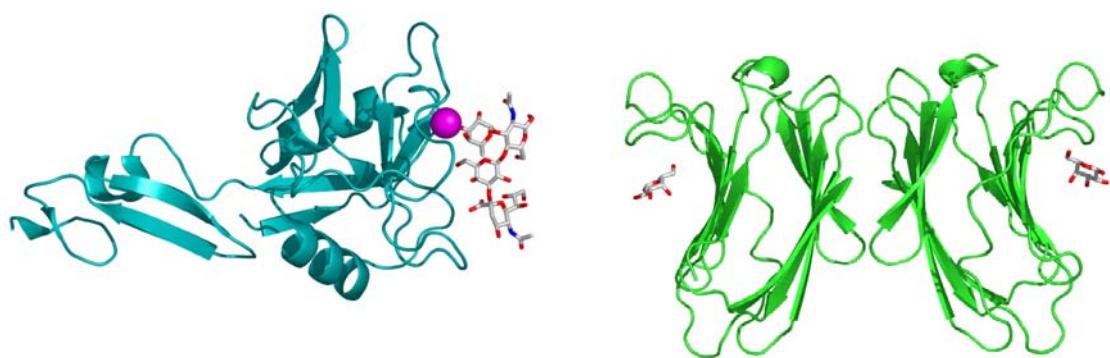


Figure 2.5 E-selectine (lectine de type C) en complexe avec leSialyl Lewis^X (gauche, pdb 1g1t) et galectine-7 humaine en complexe avec le galactose (droite, pdb 2gal)

2.4 Sites de reconnaissance

Toutes les informations disponibles sur les modes de reconnaissance ont été établies sur la base de l'analyse cristallographique des complexes lectine-sucre. Dans les lectines spécifiques pour les monosaccharides, les sites de liaison sont généralement des dépressions peu profondes sur la surface de la protéine. Par contre, dans les lectines spécifiques pour les oligosaccharides les sites de liaison sont plus profonds et montrent une excellente complémentarité pour le ligand qui ressemble à l'interaction protéine-substrat chez les enzymes. Les liaisons hydrogène entre la lectine et le ligand sont des interactions fortes et très directionnelles, ce qui permet d'atteindre une bonne affinité et spécificité. Les chaînes latérales des résidus chargés comme l'acide aspartique, l'asparagine, l'acide glutamique, la glutamine et l'arginine sont souvent impliquées dans les liaisons hydrogène avec les ligands. Les groupements NH ou CO de la chaîne principale de la protéine participent souvent à la reconnaissance. Une importante contribution à la force de l'interaction vient parfois des interactions hydrophobes. Les sucres sont des molécules très polaires mais la disposition des groupements OH peut créer des zones hydrophobes qui peuvent donner lieu à des interactions avec les résidus aromatiques, comme la tyrosine ou le tryptophane. Ce type d'interaction est appelé *stacking interaction* (Vyas 1991).

Dans les lectine de type C un atome de calcium est coordonné par les chaînes latérales des acides aminés dans le site de liaison est par deux des hydroxyles 2-OH, 3-OH et 4-OH du sucre, qui est généralement un fucose, un mannose, ou moins fréquemment un galactose (Drickamer 1996). La lectine PA-IL de *Pseudomonas aeruginosa*, étudiée dans cette thèse, adopte un repliement protéique différent des lectines de type C mais elle reconnaît le galactose par une interaction très similaire. Une nouvelle classe de protéines présentant une grande affinité pour le fucose/mannose a été trouvée chez certaines bactéries opportunistes comme *Pseudomonas aeruginosa*, *Chromobacterium violaceum* et *Ralstonia solanacearum*. Ces lectines montrent la présence exceptionnelle de deux atomes de calcium dans le site de liaison et seront décrites en détail dans les prochains paragraphes.

2.5 Les lectines bactériennes

Les lectines bactériennes sont généralement situées sur la surface de la bactérie ou localisées dans le cytosol et jouent des rôles importants dans la reconnaissance des glycoconjugués présents à la surface des cellules de l'hôte pendant la première étape du processus d'infection, ce qui a attiré bien évidemment beaucoup d'attention dans ces dernières années (Sharon 1996). Les lectines bactériennes connues peuvent être classées en trois familles: les lectines fimbriales, les toxines et les autres lectines solubles qui n'appartiennent pas aux deux premières classes (Imberty 2005).

2.5.1 Les lectines fimbriales

Les bactéries sont dotées d'organelles de surface appelés fimbriae qui servent à différentes fonctions, comme la reconnaissance et l'adhésion sur des surfaces diverses et en particulier sur les cellules des organismes eucaryotes (Low 1996). Des centaines de fimbriae sont attachés à la surface d'une cellule bactérienne. Dans les trois types différents de fimbria qui ont été observés (type 1, type P et type IV) l'organisation structurale est similaire (Soto 1999). La partie principale qui sort de la surface de la bactérie est une longue superstructure hélicoïdale rigide, assemblée à partir des nombreux monomères de piline (16-22 kD). Au sommet des pili, on peut avoir la présence d'un domaine lectine relié au dernier monomère du pilus par une partie flexible (monomères adaptateurs).

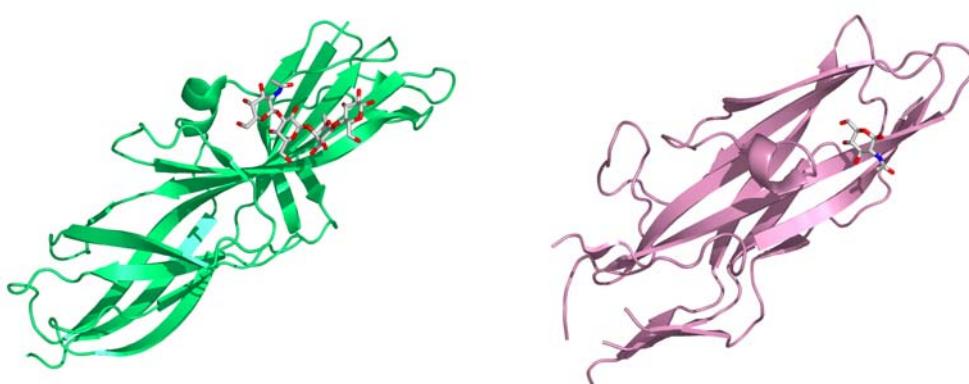


Figure 2.6 Récepteur PAPG en complexe avec le tetrasaccharide GBO4 (gauche, pdb 1j83) et lectine fimbriale F17-AG en complexe avec le GlcNAc (droite, pdb 1o9w)

Les structures cristallines des différentes lectines fimbriales résolues jusqu'à aujourd'hui appartiennent aux pili de type 1 ou de type p et montrent que les domaines lectines adoptent des replis similaires, basés sur une structure allongée de type β -sandwich. Le site de liaison pour le ligand est généralement une dépression peu profonde localisée sur un côté du domaine lectine.

2.5.2 Les toxines

Les toxines sont des protéines secrétées par la bactérie qui montrent une activité毒ique directe dans les cellules cibles. L'endommagement de la cellule peut advenir par la rupture de la paroi cellulaire, par l'inhibition de la synthèse protéique ou par l'activation des métabolismes secondaires. Les toxines du type AB₅ sont secrétées par différents microorganismes tels que *Vibrio cholerae*, certaines souches d'*E.coli* (ETEC), *Shigella dysenteriae* et *Borderella pertussis*. Ces toxines sont formées d'une sous-unité A qui est responsable de l'activité enzymatique et d'une ou plusieurs sous-unités B (Figure 2.7) qui sont spécialement conçues pour reconnaître spécifiquement les sialogangliosides présents sur la surface cellulaire, comme le GM1, GM2 et le Gb3 (Merritt 1995).

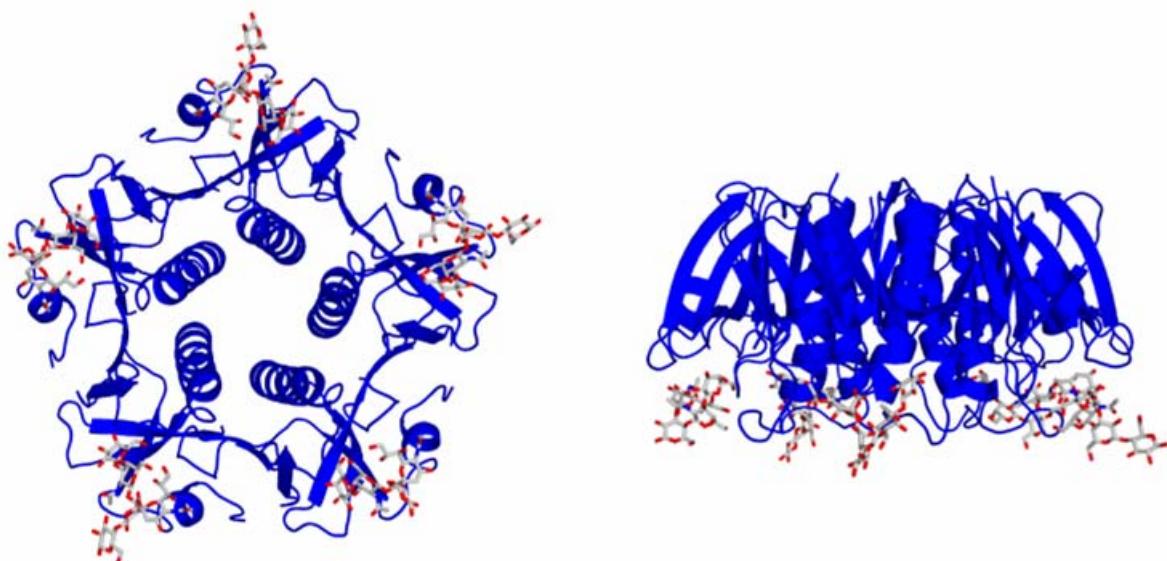


Figure 2.7 Deux orientations pour la sous-unité B de la toxine du Choléra en complexe avec le GM1 (pdb 3chb)

2.5.3 Les lectines solubles

Dans cette famille de lectines bactériennes on regroupe toutes les protéines solubles exprimées par la bactérie ayant une affinité pour des sucres et ne montrant pas d'activité enzymatique. Seul un petit nombre de lectines solubles ont été caractérisées et le domaine d'étude est en pleine expansion. L'équipe de glycobiologie moléculaire du CERMAV s'intéresse depuis quelques années aux relations structure-fonction de ce type de lectines dans les processus d'infection.

Deux lectines, nommées PA-IL et PA-IIL sont produites par le pathogène opportuniste *Pseudomonas aeruginosa*. *Chromobacterium violaceum* est aussi un pathogène pour les humains et produit une lectine, CV-IIL, ayant une forte homologie de séquence avec PA-IIL. Ces trois lectines seront décrites en détail dans le prochain paragraphe ainsi que dans le cinquième et dans le sixième chapitre. Toutes les lectines bactériennes solubles connues sont résumées dans le tableau 2.4.

Tableau 2.4 Lectines bactériennes solubles

Nom	Type	Spécificité	Caractéristiques
<i>Pseudomonas aeruginosa</i> PA-IL	Pathogène humain	galactose	Lectine cytotoxique
<i>Pseudomonas aeruginosa</i> PA-IIL	Pathogène humain	fucose/mannose	Très haute affinité vers le fucose
<i>Cromobacterium violaceum</i> CV-IIL	Pathogène humain	fucose/mannose	Très haute homologie avec PA-IIL
<i>Burkholderia cenocepacia</i> BCLx	Pathogène humain	fucose/mannose	Différents gènes identifiés homologues à PA-IIL
<i>Ralstonia solanacearum</i> RS-IIL	Pathogène végétal	mannose/fucose	Très haute homologie avec PA-IIL
<i>Ralstonia solanacearum</i> RSL	Pathogène végétal	fucose	Similaire à <i>Aleria arantia</i> lectin, une lectine de champignon
<i>Cyanovirin-N</i>	Cyanobactérie	Man ₈₋₉	Puissante activité antivirale

RSL est une lectine produite par la bactérie *Ralstonia solanacearum*, qui attaque les racines des plantes. La séquence montre la présence de deux motifs peptidiques répétés et dans la structure cristalline (Fig 2.8) trois monomères s'associent pour former une structure de type β -propeller à six lames. RSL est la seule protéine qui adopte une structure tertiaire β -propeller dérivée par l'oligomérisation des sous-unités et non par le repliement dans la chaîne (Kostlanova 2005).

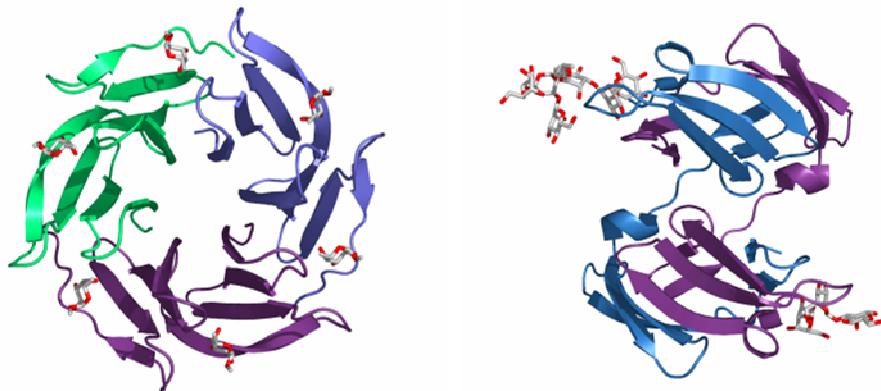


Figure 2.8 RSL en complexe avec le Me-fucoside (gauche, pdb 2bt9) et Cyanovirin-N en complexe avec un hexamannoside (droite, pdb 1m5j)

La Cyanovirine-N, produite par le cyanobactérie *Nostoc ellipsosporum* est sans doute la lectine bactérienne la plus connue et mieux étudiée, car cette protéine montre une puissante activité antivirale et plus précisément elle est capable d'inhiber le virus du SIDA (HIV) (Boyd, Gustafson et al. 1997). Sa structure tridimensionnelle (Fig. 2.8), résolue d'abord par NMR et ensuite par cristallographie a montré un nouveau domaine de reconnaissance (Bewley 2001). La Cyanovirine-N reconnaît avec une très forte affinité les oligosaccharides mannosylées (oligomannose-8 et -9) de la glycoprotéine gp120 qui forme l'enveloppe extérieure du virus.

2.6 Pseudomonas aeruginosa

Pseudomonas aeruginosa est une bactérie gram-négative et opportuniste qu'on retrouve fréquemment dans le sol, dans les eaux et dans la végétation. Elle ne représente pas un danger pour les personnes en bonne santé mais, par contre, elle peut être très dangereuse pour les sujets qui souffrent de déficiences immunitaires ou qui présentent des pathologies qui pourraient être utilisées comme « porte d'entrée » pour l'infection, comme par exemple les grands brûlés. La bactérie est bien connue dans les milieux hospitaliers où elle est responsable de nombreux cas d'infections nosocomiales (Artenstein 1993). Elle peut causer des pathologies de gravité variable comme des septicémies, des infections des voies urinaires, des

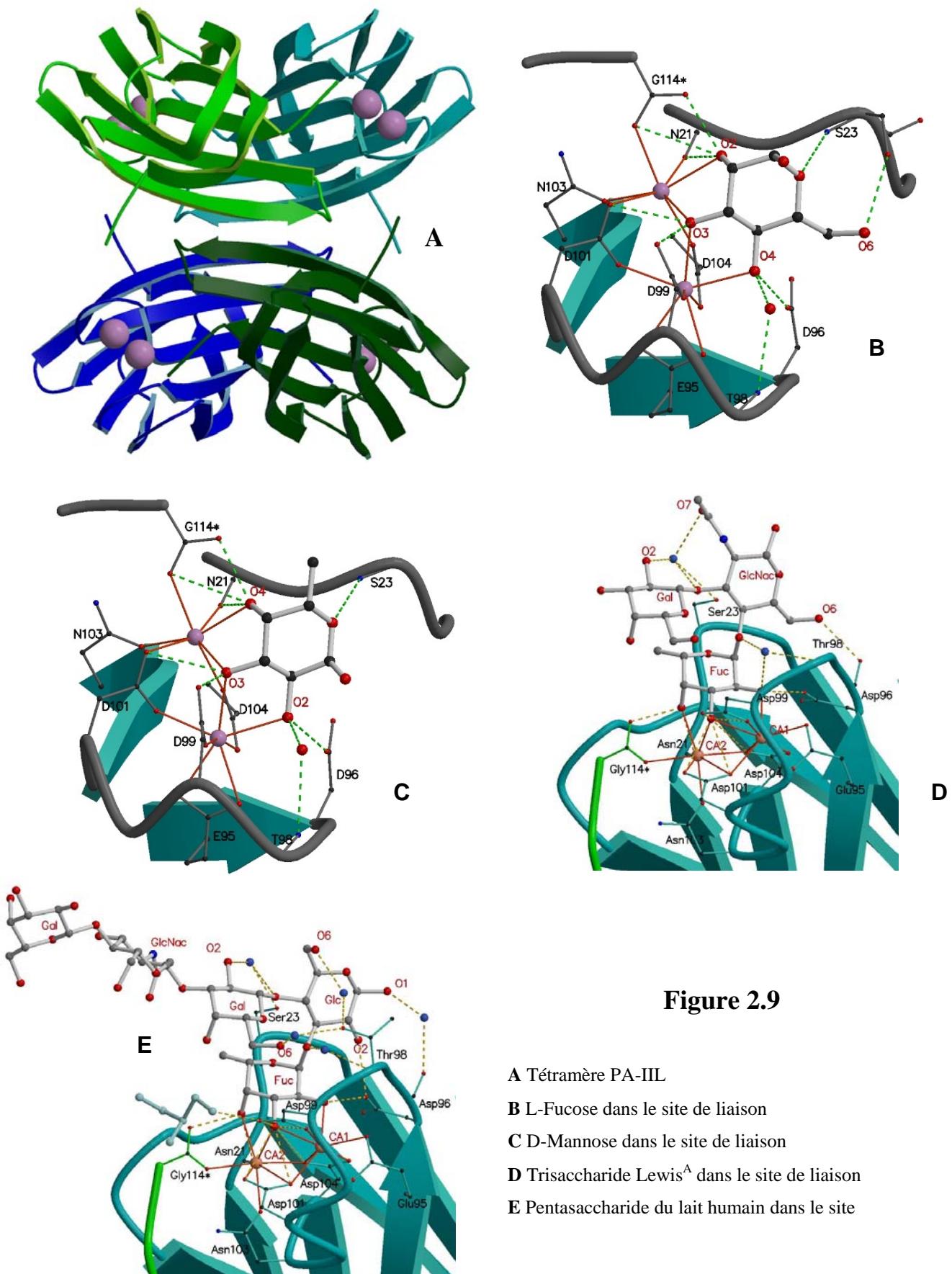
pancréatites et des dermatites. Plus particulièrement, elle est la cause primaire des graves infections pulmonaires chez les patients atteints de mucoviscidose (CF). Elle présente une grande adaptabilité et beaucoup de souches sont résistantes à la plupart des antibiotiques. Dans le processus d'infection de *P. aeruginosa* les carbohydrates jouent des rôles très importants et à différents niveaux. Ils sont les composants principaux des *lipopolysaccharides* (LPS) qui forment l'enveloppe sensée protéger les bactéries gram-negative contre la diffusion des antibiotiques. Dans les infections chroniques, *P. aeruginosa* produit aussi un *exopolysaccharide mucoïde* qui est composé surtout d'acide mannuronique et glucuronique et qui joue un rôle dans la formation du biofilm.

Du point de vue de la reconnaissance entre *P. aeruginosa* et le tissus pulmonaire, la bactérie se sert probablement d'un mécanisme concerté où différentes protéines reconnaissent de manière spécifique les glycoconjugués présents sur les cellules épithéliales et dans les mucines du mucus. Ces protéines sont d'une par les *adhésines* situées au sommet des pili de type IV qui reconnaissent les glycolipides *asialo-GM1* et *-GM2* et qui sont impliquées dans le processus d'adhésion (Hahn 1997). Récemment, un deuxième type de pilus appelé Cup (du nom du système de sécrétion *Chaperone-Usher pathway*) a été identifié chez *P. aeruginosa*. Différents gènes ont été identifiés mais l'implication de ce type d'organelles dans le processus d'adhésion n'a pas encore été démontré chez *P. aeruginosa* (Sauer 2000; Vallet 2001).

2.6.1 PA-IL et PA-IIL

L'arsenal d'adhésion de *P. aeruginosa* est complété par deux lectines, la première (PA-IL) spécifique pour le galactose et la deuxième (PA-IIL) qui montre une très forte affinité pour le fucose (Gilboa-Garber 1986). Les deux protéines sont principalement présentes dans le cytoplasme de la bactérie mais dans certaines conditions elles sont aussi présentes sur la surface de la membrane externe de la bactérie (Tielker 2005). Les deux lectines sont similaires en taille (12.7 kD pour PA-IL et 11.7 kD pour PA-IIL) mais elles ne montrent aucune identité de séquence. Cependant, toutes deux nécessitent des cations divalents pour leur activité.

La lectine PA-IL fut la première à être isolée par chromatographie d'affinité à partir d'une culture de *P. aeruginosa* (Gilboa-Garber, Mizrahi et al. 1972). Sa structure moléculaire en complexe avec du calcium et du galactose sera décrite dans le cinquième chapitre. La structure cristalline de la PA-IIL a été résolue en 2002 (Mitchell 2002). La lectine est un homotétramère et chaque monomère adopte un repliement du type sandwich à 9 brins β qui voit la présence d'un motif structural conservé appelé *clé grecque* (Fig 2.9A). Les quatre chaînes sont fortement associées en deux dimères et plus particulièrement l'extrémité C-terminale de chaque monomère (résidu Gly¹¹⁴) est localisée dans le site de liaison du monomère adjacent et participe à la reconnaissance du sucre, ce qui est une caractéristique unique dans les interactions protéine-sucre. Le site de liaison pour le sucre est une poche qui abrite deux ions calcium coordonnés par les chaînes latérales (deux Asn et quatre Asp) des acides aminés de la *boucle de fixation du calcium* (résidus 95-113) et par l'extrémité C-terminale du monomère adjacent. La structure de la lectine co-cristallisée avec du fucose (Fig 2.9 B) montre que le sucre est en interaction directe avec les deux atomes de calcium par les hydroxyles 2-OH, 3-OH et 4-OH. La « boucle de spécificité » (résidus 22-24) semble pouvoir moduler l'affinité pour les différents ligands. La PA-IIL est la première lectine à montrer un mécanisme de reconnaissance de ce type. L'affinité de la PA-IIL vers le fucose est de l'ordre du micromolaire ($K_a \sim 10^6 M^{-1}$) et peut être expliquée par la forte interaction entre le sucre et les deux atomes de calcium (Mitchell and Imbert 2005). La lectine a également été cristallisée avec le fructose et le mannose (Fig 2.9 C) (Loris, Tielker et al. 2003). Récemment, les études cristallographiques des complexes avec des oligosaccharides, le trisaccharide Lewis^A (Fig 2.9 D) et un pentasaccharide du lait humain (Fig 2.9 E), ont donné les bases moléculaires de la très haute affinité de la lectine pour ces sucres complexes (Perret 2005).



2.6.2 Rôles de PA-IL et PA-IIL dans le processus d'infection

La PA-IL peut être classée dans les facteurs de virulence de la bactérie. Dans *P. aeruginosa* l'expression des gènes de virulence est contrôlée par un système de messagerie biochimique appelé *quorum sensing* qui est utilisé pour la communication entre cellules par de nombreuses bactéries (Whitehead 2001). Le mécanisme de ce système est simple : des molécules appelées *autoinducteurs*, appartenant à la famille des N-acylhomosérine lactones (AHL), sont secrétées par la bactérie dans l'environnement jusqu'à ce qu'une concentration maximale soit atteinte. La détection de cette concentration limite entraîne des modifications dans le mécanisme d'autorégulation de la bactérie qui passe d'une cellule isolée à une colonie. Chez *P. aeruginosa* le mécanisme de *quorum sensing* est impliqué dans la production de facteurs de virulence ainsi que dans la croissance et dans la formation du biofilm, qui est directement impliqué dans la résistance aux antibiotiques (Davies 1998). Dans cette bactérie il existe deux systèmes différents (*las* et *rhl*) qui contrôlent la transcription de gènes en réponse à deux autoinducteurs différents AHL, chacun spécifique pour un type de système. L'expression de la PA-IL dépend directement du système *rhl* et peut être activée par la présence de son autoinducteur AHL spécifique (Winzer, Falconer et al. 2000). PA-IL exerce une activité cytotoxique et est capable d'inhiber la croissance des cellules épithéliales des poumons (Bajolet-Laudinat 1994). De plus, la présence de la lectine en combinaison avec des autres facteurs de virulence (exotoxine A, élastase) a une forte activité cytotoxique *in vivo*, qui peut être inhibée par la présence de N-acetylgalactosamine, un des ligands préférés de PA-IL. Cette dernière observation a permis d'avancer l'hypothèse d'un rôle pour la lectine dans la perméabilisation de la membrane cellulaire qui la rendrait plus vulnérable à l'action des autres facteurs de virulence.

La lectine à fucose PA-IIL est également produite en combinaison avec les facteurs de virulence et, comme PA-IL, elle est sous le contrôle du *quorum sensing* (Winzer, Falconer et al. 2000; Schuster 2003; Wagner 2003). Elle n'est pas cytotoxique mais il a été montré *in vitro* qu'elle peut inhiber le battement ciliaire de cellules pulmonaires (Adam 1997). Des études récentes ont élucidé la localisation extracellulaire de la protéine qui serait fortement associée aux polysaccharides présents sur la surface de la bactérie (Tielker 2005). Cette lectine serait donc impliquée dans la reconnaissance et l'adhésion bactérie-bactérie ou bactérie-hôte et particulièrement dans la formation du biofilm, car il a été démontré qu'une

souche bactérienne privée du gène PA-IIL secrète un biofilm moins épais que celui de la souche normale (Tielker 2005).

2.6.3 Similarité de séquence et de structure chez d'autres microorganismes

A la suite du séquençage de nouveaux génomes bactériens on a pu mettre en évidence la présence de gènes homologues à PA-IL et surtout à PA-IIL dans d'autres bactéries.

Dans le génome du pathogène d'insecte *Photorabdus luminescens* on a identifié une séquence, correspondante à la protéine hypothétique Plu2096, qui montre 36% d'identité et 48% de similarité avec la PA-IL; tous les acides aminés impliqués dans la reconnaissance du sucre sont également bien conservés.

Certaines séquences présentant une haute identité avec PA-IIL ont été identifiées dans les génomes d'autres bactéries opportunistes, comme on peut le voir dans l'alignement montré dans la figure 2.10. *Ralstonia solanacearum* est un redoutable pathogène de plantes qui est responsable de dommages importants chez 200 types de plantations dans le monde entier (Shell 2000). La bactérie est reliée phylogénétiquement à *P. aeruginosa* et produit une lectine RS-IIL (ligne 2 en figure 2.10) qui montre une très forte identité de séquence avec PA-IIL (Mitchell 2002). Sa structure tridimensionnelle est presque identique à PA-IIL et tous les résidus impliqués dans la liaison du calcium sont très bien conservés (Sudakevitz, Kostlanova et al. 2004). Cependant RS-IIL préfère se lier au mannose et non pas au fucose comme PA-IIL et cette préférence semble due à une différence de trois acides aminés dans la boucle de spécificité.

Chlorobacterium violaceum est considéré comme pathogène émergent dans les régions tropicales et subtropicales (Alves de Brito 2004). Sa lectine CV-IIL, spécifique pour le fucose et le mannose, montre une identité de séquence de 60% avec la PA-IIL (Imberty 2004). Très récemment elle a été purifiée et caractérisée du point de vue biochimique (Zinger-Yosovich In press) et dans le sixième chapitre on analysera en détail sa structure tridimensionnelle en complexe avec ses ligands.

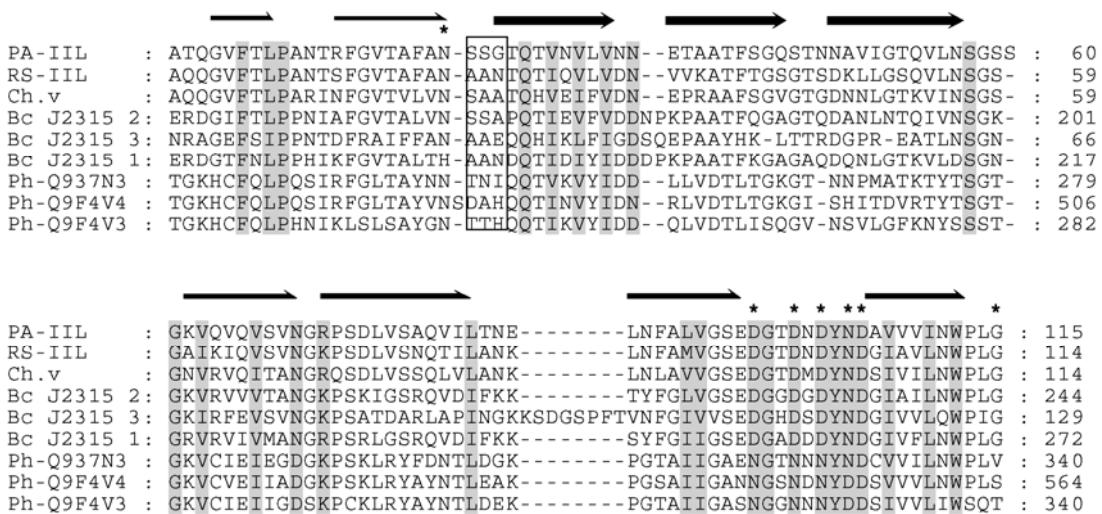


Figure 2.10 Alignement de la famille de type PA-IIL. Les résidus conservés sont représentés sur fond gris et les flèches indiquent l'extension des brins β . Les résidus impliqués dans la reconnaissance du calcium sont marqués avec un astérisque. Les résidus de la boucle de spécificité sont entourés.

Trois copies différentes d'un gène similaire à PA-IIL ont été retrouvées dans le génome de *Burkholderia cenocepacia*. Cette bactérie opportuniste est également responsable d'infections nosocomiales et en analogie à PA-IIL elle peut causer de graves infections pulmonaires chez les patients atteints de mucoviscidose. Pour ces implications médicales, ce sujet d'étude est donc particulièrement intéressant et des efforts considérables sont en train d'être effectués dans l'équipe de glycobiologie moléculaire pour élucider les structures et les rôles de ces trois protéines.

Dans toutes ces séquences, les résidus impliqués dans la reconnaissance du calcium sont très conservés, donc elles peuvent être annotées comme protéines liant le calcium. Par contre, les résidus appartenant à la boucle de spécificité sont variables et comme on a vu chez RS-IIL, ces variations peuvent jouer un rôle important dans la modulation de l'affinité vers les différents ligands.

2.7 Les lectines de champignon

Les champignons, y compris les moisissures et les champignons proprement dits, font partie d'une importante classe d'organismes. De part leurs propriétés nutritionnelles, les champignons ont toujours gardé une place importante dans notre alimentation, dans nos traditions gastronomiques mais également dans la médecine, spécialement dans la médecine traditionnelle orientale. N'étant pas capable d'utiliser la photosynthèse comme les plantes ils doivent extraire du milieu toutes les substances nutritives dont ils se nourrissent et ont donc adopté des modes de vie saprophytes, parasites ou symbiotiques. Ces organismes, qui à première vue peuvent apparaître très simples, ont développé au cours des centaines de millions d'années d'évolution une impressionnante série de gènes et sont donc très riches en métabolites et en protéines qui confèrent à la fois des propriétés bénéfiques ou très toxiques.

Parmi les protéines qui ont été purifiées à partir de champignons on peut trouver des protéines antifongiques, des ribonucléases, des protéines de type *ubiquitine*, des lectines, des cellulases, des xylanases, des laccases, des invertases et des tréhalose phosphorylases (Ng 2004). L'intérêt montré dans les dernières années pour les lectines de champignons est principalement motivé par la découverte que certaines lectines ont des propriétés pharmacologiques intéressantes comme par exemple la stimulation du système immunitaire, contre l'hypertension et contre l'hypercholestérolémie mais aussi antivirales et anticancéreuses (She 1998; Sze 2004). L'abondance de lectines dans les champignons est tout à fait remarquable et un test d'agglutination conduit sur plus de 400 spécimens a permis d'identifier la présence de lectines dans la moitié des champignons analysés (Pemberton 1994).

2.7.1 Caractéristiques

Les lectines de champignon qui ont été purifiées et caractérisées jusqu'à aujourd'hui montrent des caractéristiques très variables, soit en terme de taille (12-190 kD), soit en terme de structure primaire, de glycosylation, de nombre des sous-unités (1-8) et de structure tridimensionnelle. La spécificité pour les sucres est aussi très variable et les ligands reconnus par ces lectines sont soit de simples monosaccharides, soit des structures plus complexes comme des oligosaccharides ou des glycoprotéines.

De plus, les lectines de champignon sont des outils intéressants pour la glycobiologie et ont trouvé des applications dans les études taxonomiques, embryologiques et bactériologiques, dans le fractionnement des glycoconjugués, des cellules, etc. Cependant, malgré l'intérêt montré dans ces dernières années qui a conduit à l'isolation de nombreuses lectines, la connaissance de ces protéines reste encore très limitée.

2.7.2 Rôles

En ce qui concerne les rôles biologiques, différentes hypothèses ont été avancées (Guillot 1997). Elles joueraient probablement un rôle important dans la période de dormance, dans la croissance et la morphogenèse du corps fructifère ou comme protéines de défense immunitaire. Chez les saprophytes, les lectines pourraient avoir un rôle dans la reconnaissance des substrats nutritifs. Par contre, chez les champignons qui adoptent un mode de vie parasitaire ou symbiotique avec d'autres organismes, les lectines sont probablement impliquées dans les processus de reconnaissance de l'hôte ou dans les premières étapes de la mycorrhization (Giollant 1993).

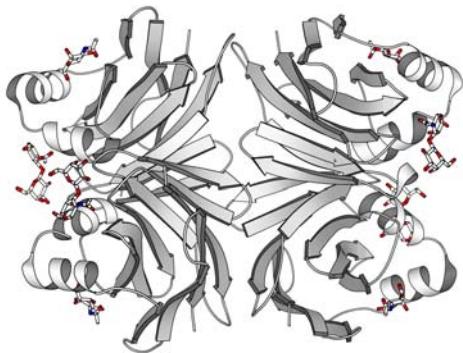
2.7.3 Lectines connues

Les informations disponibles dans la littérature sur la structure tridimensionnelle sont encore limitées, comme on peut le voir dans le tableau 2.5 et dans la Fig. 2.11. En fait, la première structure cristalline d'une lectine de champignon a été résolue seulement en 2003 par l'équipe de glycobiologie moléculaire du CERMAV. La lectine AAL extraite du champignon *Aleuria aurantia* est spécifique pour le fucose et montre une structure du type β -propeller à six lames (Wimmerova 2003). La lectine de *Flammulina velutipes* (Fve) fait partie d'une nouvelle classe de lectines fongiques à activité immunomodulatrice (FIP). Cette protéine montre un intéressant repliement de type fibronectin III mais son site de reconnaissance n'a pas pu être caractérisé (Paaventhamp 2003). La lectine de *Xerocomus chrysenteron* (XCL) et la lectine d'*Agaricus bisporus* (ABL), le champignon de paris, montrent un repliement similaire à celui d'une classe de toxines, le *actinoporines* (Birck 2004; Carrizzo 2005). En outre ABL est bien connue pour sa remarquable activité anticancéreuse (Yu 1993). La résolution de la structure de GCL2, la lectine de *Coprinus cinerea* (Walser 2004), puis la lectine d'*Agrocybe cylindracea* (Ban 2005) a montré pour la première fois que le repliement du type galectine

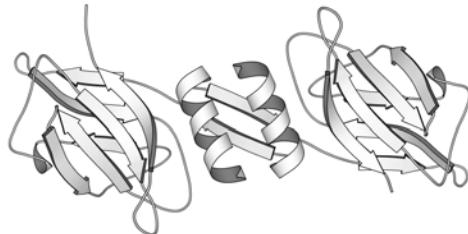
n'est pas une caractéristique exclusive des lectines animales. Un dernier exemple est la lectine toxique du champignon parasite *Laetiporus sulphureus* (LSL) qui montre un CRD du type trèfle similaire à la ricine B, associé à un domaine hydrophobe bien conservé dans toutes les protéines de la famille de l'*aerolysine* (Mancheno 2005). On peut noter que le Protein Data Bank contient d'autres structures de lectines de champignons récemment déposées qui vont être bientôt disponibles et qui sont extraites d'*Agrocybe aegerita*, de *Sclerotum rolfsii* et de *Psathyrella velutina*, qui fait notamment l'objet du huitième chapitre de cette thèse.

Tableau 2.5 Lectines de champignons connues

Mushroom	Year	Specificity	Fold, characteristics
<i>Aleria arantia</i> (AAL)	2003	Fucose	β -propeller à 6 lames
<i>Flammulina velutipes</i> (FVL)	2003		Domaine fibronectin FNIII
<i>Xerocomus chrysenteron</i> (XCL)	2004	Gal/GalNAc	Domaine similaire aux actinoporines
<i>Coprinus cinereus</i> (CGL2)	2004	Gal	Galectine
<i>Agrocybe cylindracea</i> (ACG)	2005	Gal	Galectine
<i>Agaricus bisporus</i> (ABL)	2005	T-antigen	Domaine similaire aux actinoporines
<i>Laetiporus sulphureus</i> (LSL)	2005	LacNAc	Un domaine ricine et un domaine type pore
<i>Psathyrella velutina</i> (PVL)	2006	GlcNAc/NeuNAc	β -propeller à 7 lames



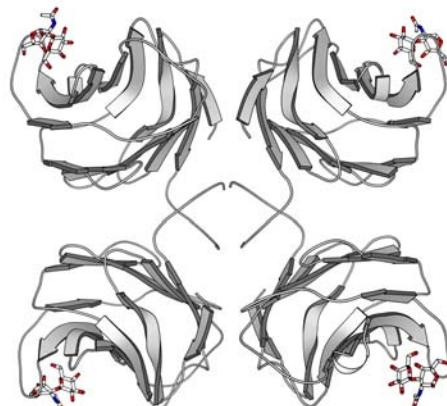
Agaricus bisporus lectin (pdb 1y2w)



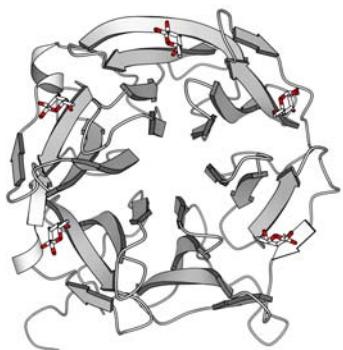
Flammulina velutipes lectin (pdb 1osy)



Laetiporus sulphureus lectin (pdb 1w3g)



Coprinus cinerea galectin (pdb 1ulc)



Aleuria aurantia lectin (pdb 1ofz)



Xerocomus chrysenteron (pdb 1xio)

Figure 2.11 Exemple de structures de lectines de champignons

Chapitre 3

Méthodes expérimentales

3.1 Cristallographie

La cristallographie aux rayons X est sans doute la méthode d'analyse structurale la plus puissante dans l'élucidation de structures tridimensionnelles de molécules et macromolécules biologiques. L'application de la cristallographie à l'étude structurale des macromolécules, et plus particulièrement des protéines, doit sa naissance aux études de Max Perutz et John Kendrew qui ont obtenu le prix Nobel en physique et en biologie pour leur travail sur la détermination de la structure tridimensionnelle de l'hémoglobine et de la myoglobine. De 1959 à nos jours, la biologie structurale a considérablement progressé. L'augmentation du pouvoir de calcul des ordinateurs, la facilité d'accès au rayonnement synchrotron et surtout les découvertes majeures de la biologie moléculaire ont contribué à faire de la cristallographie la technique de pointe pour l'étude structurale des biomolécules. Il suffit de regarder l'augmentation du nombre de structures dans la banque de données PDB pour apprécier ce petit « miracle de la science ».

Cependant, les bases théoriques de cette méthode mais aussi certaines difficultés rencontrées n'ont pas changé, même si aujourd'hui la résolution d'une nouvelle structure pose beaucoup moins de problèmes techniques qu'il y a 30 ans. Plus particulièrement, la cristallisation d'une macromolécule relève encore de méthodes *d'essai-erreur* qui incluent parfois certaines conditions expérimentales très difficiles à contrôler (chance du néophyte, bonne humeur de la protéine ou du matériel utilisé, etc...). Une fois que la protéine a été cristallisée, le *problème du phasage* est la deuxième difficulté à résoudre et on verra plus tard les différentes techniques qui ont été développées.

3.1.1 Cristallisation

Un cristal de macromolécule et un cristal de sel sont constitués de la même façon : un arrangement périodique des atomes dans l'espace. Cependant, les cristaux macromoléculaires sont généralement beaucoup moins « denses » car plus riches en solvant et moins ordonnés : les macromolécules ont toutes un certain degré de flexibilité. En définitive, un cristal de protéine (ou d'oligosaccharide) est beaucoup plus difficile à faire croître qu'un cristal de petite molécule. La technique la plus utilisée est appelée *la diffusion de vapeur* et consiste à mélanger la solution de protéine avec une solution précipitante (sels, solvants organiques, polymères, etc.) et de sceller une goutte (4-10 µL) dans un compartiment contenant une grande quantité (0.5-1 mL) de solution précipitante pure (Fig. 3.1). Dans le processus

d'équilibration qui va s'instaurer, l'eau passe de la goutte à la solution concentrée et donc la concentration de protéine dans la goutte augmente jusqu'à ce qu'elle atteigne un certain niveau appelé *sursaturation*. A ce moment, et seulement si les autres conditions dans la goutte sont favorables (pH, concentration en sel, température, présence de molécules stabilisantes, etc...) on peut observer l'apparition des premiers cristaux. Etant donné qu'un cristal est une structure très ordonnée, on peut comprendre que pour que sa croissance soit favorisée, l'homogénéité (taille et forme des macromolécules) de l'échantillon de départ est un paramètre très important.

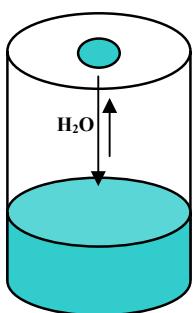
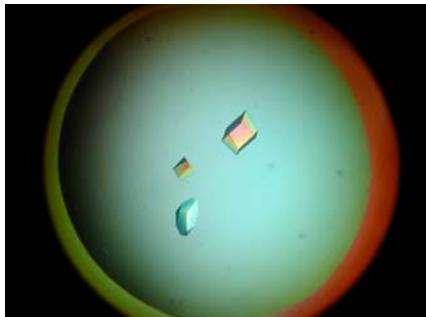


Figure 3.1

A gauche: la technique de *hanging drop* (goutte pendante) est la plus utilisée et se base sur le principe décrit dans le texte.

En bas à gauche: très jolis cristaux de lectine CV-IIIL photographiés dans une goutte.

En bas: le nanorobot de cristallisation installé à l'EMBL-PSB, Grenoble. Cet appareil permet d'essayer 300 conditions différentes de cristallisation avec seulement $100\mu\text{L}$ d'échantillon.



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. L'*unité asymétrique* est la plus petite unité moléculaire (une molécule ou un groupe de molécules) qui suite à l'application d'opérations de symétrie (définis par le *groupe d'espace* du cristal) permet d'obtenir la maille élémentaire. Comme les macromolécules sont des molécules chirales, les éléments de symétrie qui comportent un *centre d'inversion* ou un *miroir* ne peuvent être appliqués.

3.1.2 Collection des données

Quand un cristal est placé dans un faisceau de rayons X, la densité électronique de chaque atome du cristal interagit avec la radiation en la diffusant dans toutes les directions de

l'espace. C'est la présence du réseau cristallin qui amplifie la radiation seulement dans certaines directions et donne naissance au diagramme de diffraction caractéristique. Les bases mathématiques et géométriques du processus de diffraction (Fig. 3.2) ont été formulées par Bragg, Laue et Ewald au début du siècle.

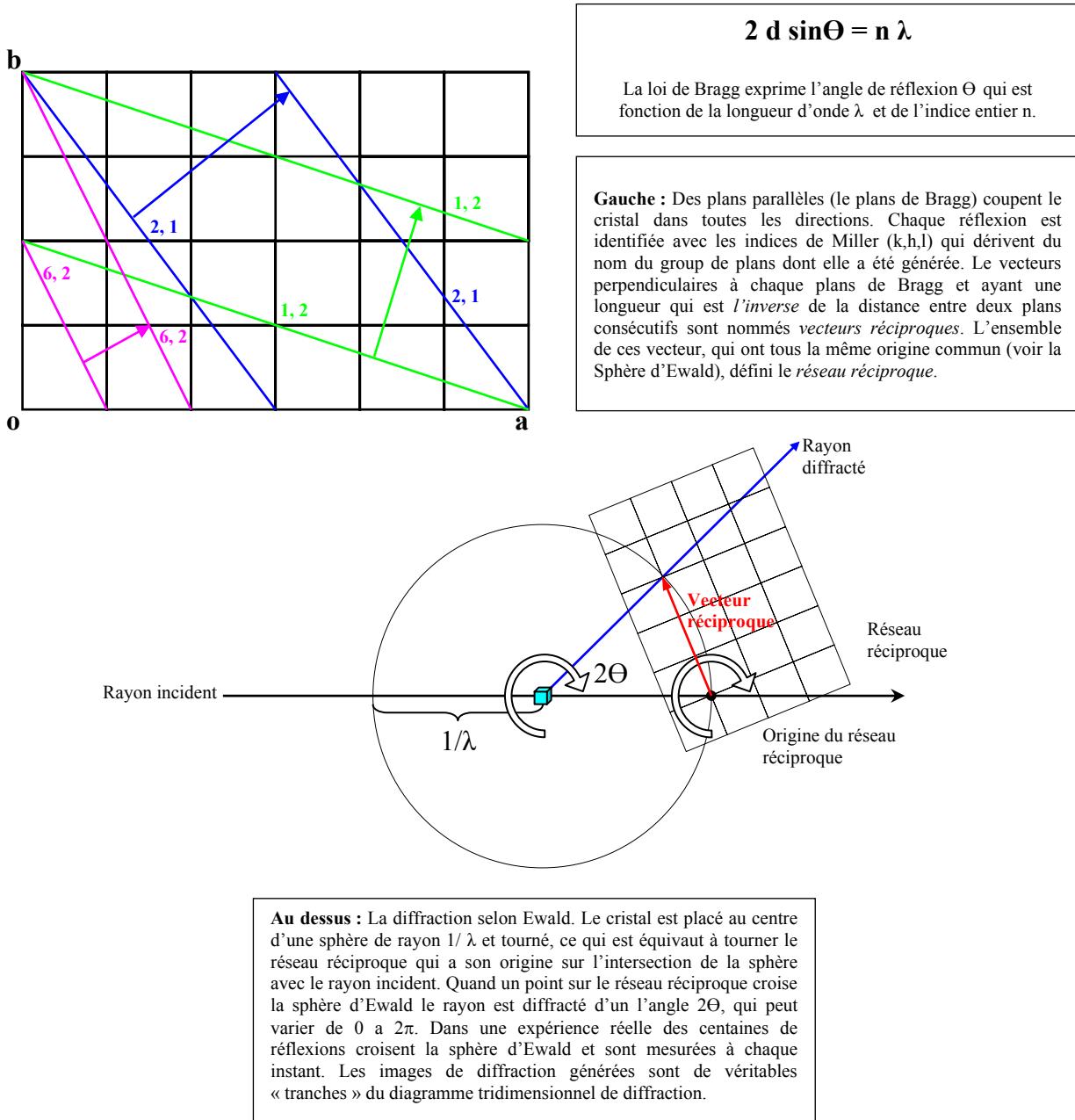


Figure 3.2 Bases géométriques de la diffraction

Le cristal est tourné autour d'un axe généralement perpendiculaire au faisceau de rayons X et les rayons diffractés sont collectés sur un détecteur qui est de nos jours principalement une caméra CCD capable d'enregistrer un grand nombre de taches dans un temps d'acquisition très court. Chaque « image » correspond à une rotation de 0.2-1° degré et peut être considérée

comme une « tranche » de l'espace réciproque. Sur une ligne de lumière d'un synchrotron de troisième génération (comme l'ESRF) le temps d'acquisition pour une image est de l'ordre de quelques secondes. La position d'une réflexion contient les informations sur la géométrie du groupe d'espace et la maille du cristal tandis que l'intensité mesurée I dépend de la distribution de la densité électronique dans l'unité asymétrique. Les images sont analysées à l'aide de programmes informatiques tels que Mosflm (Leslie 1992) ou Denzo (Otwinowski 1997) qui permettent d'attribuer à chaque réflexion ses *indices de Miller* (h,k,l) et son intensité I associée avec une erreur expérimentale σI . Toutes les réflexions sont ensuite intégrées par le logiciel, les intensités I sont mises à la même échelle et converties en *facteurs de structure observés* F .

De par la présence des opérations de symétrie dans le cristal quelques réflexions seront reliées par symétrie et auront donc la même intensité de diffraction. Les informations nécessaires sur la distribution de la densité électronique dans le cristal seront alors contenues dans un plus petit nombre des réflexions appelées *uniques*. La *complétude* est le pourcentage de réflexions uniques mesurées. Le rapport entre les réflexions totales et les réflexions uniques donne la *multiplicité*. Un troisième paramètre est le rapport entre l'intensité d'une réflexion et son erreur, $I/\sigma I$. Ces trois grandeurs, avec la résolution maximale obtenue, sont les paramètres les plus importants pour évaluer la qualité d'un jeu de données expérimentales.

3.1.3 Le phasage

Le concept central de toute la théorie de la diffraction est que le *facteur de structure* est la transformée de Fourier de la distribution de la densité électronique dans le cristal. De par les propriétés d'une transformée de Fourier, le contraire est également vrai, c'est à dire que la densité électronique est la transformée de Fourier du facteur de structure. Plus précisément, pour calculer la densité électronique en un point du cristal xyz on doit sommer la contribution de toutes les ondes diffractées par les plans hkl (dont les amplitudes dépendent du nombre d'électrons dans chaque plan), additionnées en prenant en compte la relation de phase:

$$\rho(xyz) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp[-2\pi i(hx + ky + lz)]$$

Or, pendant une collecte de données de diffraction on peut mesurer l'amplitude de chaque réflexion F_{hkl} mais pas sa phase correspondante α_{hkl} , qui est perdue invariablement pendant l'expérience de diffraction. En fait, au contraire de la lumière visible qui peut être focalisée indéfiniment avec des lentilles, il n'y a aucun moyen (au moins pas encore) de recombiner

l'intensité et la phase d'un rayon X diffracté. Différentes méthodes pour résoudre ce *problème de la phase* ont été développées et seront décrites dans les prochains paragraphes.

3.1.3.1 La méthode de Patterson

Une équation très utile est la *fonction de Patterson*, définie par :

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

Le résultat de cette synthèse de Fourier est une carte de densité électronique un peu spéciale. Les pics qu'on voit dans la carte de Patterson représentent tous les vecteurs inter-atomiques qui peuvent être tracés entre les atomes de la maille réelle. Exemple : si dans la maille réelle il y a deux atomes séparés par un vecteur $r(u,v,w)$, dans la carte de Patterson on aura la présence de deux pics, l'un à (u,v,w) et l'autre à $(-u,-v,-w)$. Pour de petites structures il est possible de « déchiffrer » la carte de Patterson et remonter aux positions atomiques réelles. Pour des structures de protéines, ceci est pratiquement impossible mais on verra plus tard que la méthode de Patterson est d'importance fondamentale dans les différentes méthodes de phasage et dans le remplacement moléculaire.

3.1.3.2 Les méthodes directes

Une énorme quantité de travail a été effectuée pendant les années 50 et 60 pour résoudre *ab initio* le problème de la phase, c'est à dire sans la connaissance d'aucun autre paramètre sauf les intensités expérimentalement mesurées. Ces études ont amené le Prix Nobel à Karl et Hauptmann (mais malheureusement pas à Isabella Karl). Cette méthode, qui a été développé pour la résolution de cristaux moléculaires est applicable à condition que la résolution soit au moins de 1.0 Å. L'application à la cristallographie des macromolécules est donc limitée mais néanmoins quelques structures de protéines ont été résolues avec des logiciels comme Shake-and-Bake (Rapleye 2002), SIR2004 (Burla 2005) ou Acorn (Foadi 2000).

3.1.3.3 Le remplacement isomorphe

Historiquement, la première méthode de phasage *de novo* à être appliquée en cristallographie des protéines fut le *remplacement isomorphe* (Fig. 3.3). Cette méthode, déjà connue dans la cristallographie des petites molécules, consiste à insérer dans le cristal natif un atome « lourd » qui peut modifier les intensités d'une façon détectable. La condition importante à l'application de cette méthode est que le cristal *dérivé* conserve son groupe d'espace et ses dimensions de

maille. Malgré les difficultés techniques de cette méthode qui oblige à tester des dizaines, parfois des centaines d'atomes lourds, le remplacement isomorphe est toujours une méthode de phasage très utilisée.

3.1.3.4 La diffusion anomale

Cette méthode de phasage utilise la propriété de certains atomes qui possèdent une diffusion « anomale » quand ils sont irradiés par un faisceau X avec une longueur d'onde qui est proche de l'énergie du seuil d'absorption de l'atome. Cette diffusion anomale dépend du fait que les électrons des atomes ne sont pas « fixes » mais libres d'osciller et de se déplacer (en énergie), donnant origine aux oscillations typiques qu'on peut voir dans les courbes d'absorption. Comme on peut le voir dans la Fig. 3.3 le facteur atomique f_{anom} de diffusion d'un électron lié se compose de trois parties :

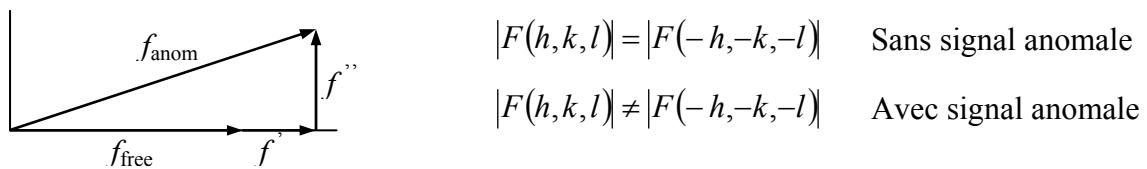
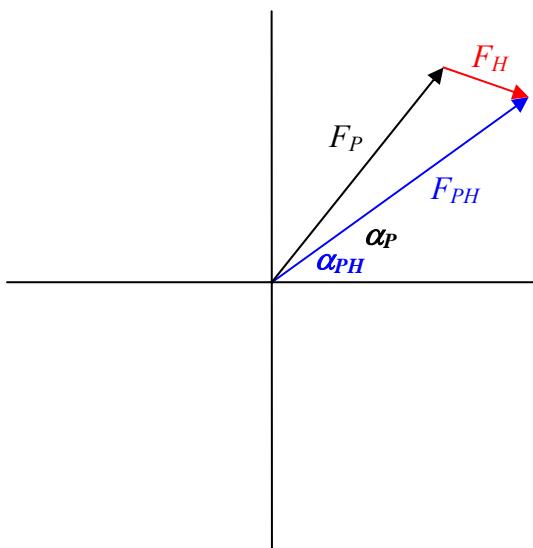


Figure 3.3 Facteur de diffusion pour un électron fixe

Une partie f_{free} qui est le facteur de diffusion pour un électron libre, une partie f' qui est « en phase » avec la f_{free} et une partie f'' dont la phase est toujours retardée de 90° . Dans les méthodes SAD (*Single Anomalous Dispersion*) et MAD (*Multiwavelenght Anomalous Dispersion*) on utilise la propriété du signal anomal de violer la loi de Friedel qui dit que les valeurs absolues des facteurs de structure pour les réflexions reliées par symétrie sont identiques (Fig. 3.4). Les atomes les plus fréquemment utilisés sont le sélénium, incorporé dans une protéine recombinante comme séléniométhionine, le souffre (phasage *long-wavelength*), le brome, le xénon, les métaux de transition et surtout les lanthanides, qui ont un puissant signal anomal. En pratique, presque tous les atomes à partir du souffre peuvent être utilisés, à condition de disposer d'une ligne de lumière capable d'atteindre une longueur d'onde proche du seuil d'absorption de l'atome. Les méthodes décrites ci-dessus peuvent aussi être utilisées ensemble comme pour exemple dans le SIRAS (*Single Isomorphous Replacement with Anomalous Dispersion*).


La méthode SIR

Les amplitudes de la substructure sont calculées avec la :

$$|F_H| \equiv |F_{PH}| - |F_p|$$

Les positions des atomes lourds sont déterminées avec les méthodes directes ou avec une synthèse de Patterson.

Les phases de la protéine sont déterminées avec la formule du cosinus :

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

Cette équation donne deux solutions symétriques pour la α_p (voir en bas dans la *construction de Harker*) on a donc un problème d'ambiguïté de phase.

La méthode MIR

La solution pour résoudre l'ambiguïté de la phase est d'utiliser un deuxième dérivé (vecteur bleu à droite). De cette façon, les probabilités de phase pour chaque dérivé se renforcent pour donner une probabilité de phase finale unimodale.

La méthode SAD

Dans la méthode SAD, le cristal est mesuré à une seule longueur d'onde qui généralement est celle où la composante anomale f'' est maximale. Dans la figure à droite on peut voir que la loi de Friedel pour la F_{PH} (en dessus et en dessous du diagramme) n'est plus respectée, c'est à dire que la $|F_{PH}(+)|$ n'est plus égale à la $|F_{PH}(-)|$. Les différences $\Delta|F|_{ano}$ entre les réflexions liés par symétrie (*paires de Friedel*) sont calculée avec :

$$\Delta|F|_{ano} = [|F_{PH}(+)| - |F_{PH}(-)|] \frac{f'}{2f''}$$

et ensuite utilisées dans un synthèse de Patterson pour déterminer les positions des atomes anomaux.

On peut donc comprendre que la contribution du signal anomal au facteur de structure est essentiel et doit être mesurée avec une grande précision pour que la méthode soit utilisable avec succès. Il est donc nécessaire de mesurer un jeu de données de haute qualité. La *complétude* et la *multiplicité* sont particulièrement importantes et doivent être les plus élevées possibles.

La méthode MAD

Dans la méthode MAD on mesure le cristal à plusieurs longueurs d'onde différentes, ce qui est équivalent à disposer des plusieurs cristaux « dérivés ». La contribution anomale se somme avec la contribution normale pour donner, en analogie avec la méthode MIR, une probabilité de phase encore plus unimodale.

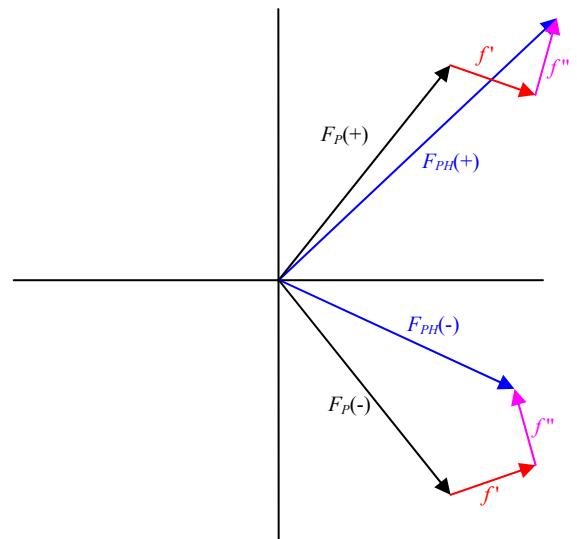
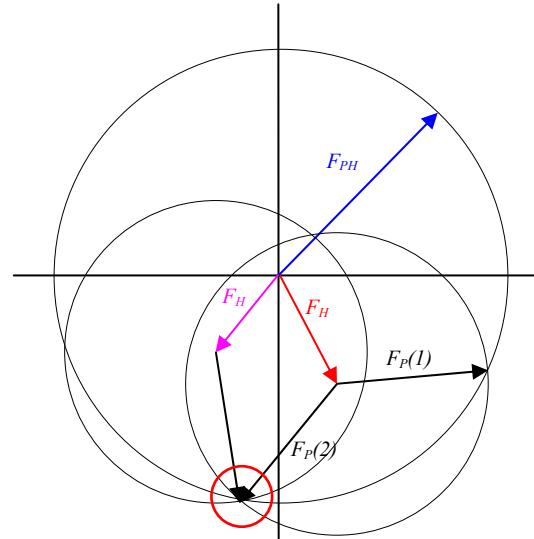


Figure 3.4 Techniques de phasage de macromolécules

Souvent, bien qu'une nouvelle structure ait été phasée, la densité électronique expérimentale n'est pas de qualité suffisante (peu de contraste entre le solvant et la protéine, manque de détail ou zones vides) pour tracer un premier modèle de la chaîne peptidique. Il est donc dans l'intérêt du cristallographe d'améliorer la densité observée par une procédure appelée *density modification*. Dans cette procédure itérative la densité électronique est améliorée par l'utilisation d'algorithmes complexes dont les plus utilisés sont l'aplatissement de solvant (*solvent flattening*) et le *non-crystallographic symmetry averaging*. Ces procédures sont particulièrement efficaces avec la méthode SAD et aussi quand le cristal présente un bon pourcentage de solvant et une symétrie élevée.

3.1.3.5 L'utilisation de sucres modifiés dans le phasage des lectines

Une nouvelle méthode est basée sur l'utilisation de molécules organiques contenant un atome lourd qui peuvent être co-cristallisées ou insérées dans un cristal natif par trempage. Dans le cas particulier des lectines, il est possible de modifier chimiquement *ad-hoc* le sucre naturellement reconnu. Cette technique a été proposée par le groupe de Loris et al., qui a utilisé un sélénio-dérivé, le 2-acétamido-2-déoxy-1-sélénio- β -D-glucopyranoside (β -MeSeGlcNAc), pour phaser en MAD la structure de l'adhésine fimbriale F17-G (Buts 2003). La lectine RSL (paragraphe 2.5.3) a été résolue par la méthode SAD avec un sélénio-derivable du fucose, le méthyl 1-sélénio- α -L-fucoside (Kostlanova 2005) et récemment la structure de la lectine de *Parkia platycephala* a été résolue en MAD avec un dérivé halogéné du mannose, le 5-bromo-4-chloro-3-indolyl- α -D-mannopyranoside (Gallego del Sol 2005). Un ligand sélénier a également été utilisé dans le huitième chapitre pour la résolution de la lectine fongique de *Psathyrella velutina*.

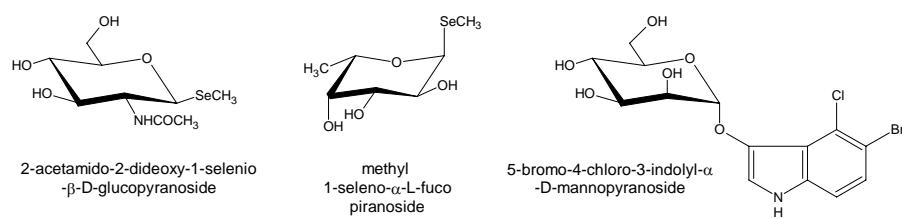


Figure 3.5 Sucres modifiés utilisés dans le phasage de lectines

Cette méthode est très élégante car avec une seule expérience on peut phaser une structure mais aussi obtenir des informations sur le mécanisme de reconnaissance dans le site de liaison. Un deuxième avantage est que les sucres modifiés en position anomère se fixent généralement très bien à la lectine et donc peuvent être co-cristallisés. Dans les cas où la co-cristallisation est impossible on peut essayer le trempage qui ne nécessite qu'une faible

concentration de ligand. L'avantage est bien évidemment que dans la majorité des cas le cristal trempé garde ses propriétés, sa maille et ses limites de diffraction. Il faut cependant noter que les monosaccharides modifiés ne sont pas actuellement disponibles commercialement et que leur synthèse demande l'expertise d'un chimiste des sucres.

3.1.3.6 Le remplacement moléculaire

La dernière technique de phasage est le *remplacement moléculaire* (MR). Cette méthode n'est pas une technique *de novo* mais est conditionnée par la disponibilité de la structure (cristalline, RMN ou modélisée) d'une protéine qui a une forte probabilité d'adopter un repliement similaire à celui de la protéine étudiée. En terme de similarité de séquence entre les deux protéines, l'application du MR a quelques chances de succès si l'identité est au moins de 25-30%, mais il peut y avoir des exceptions. La structure utilisée comme *modèle* est d'abord placée dans la nouvelle maille par une *fonction de rotation* (c'est une méthode de Patterson) et une fois que les orientations « favorables » ont été trouvées, elles sont soumises à une *fonction de translation*, qui essaye de placer la (ou les) molécule(s) correctement orientée(s) dans la maille. Les logiciels disponibles aujourd'hui utilisent des fonctions de recherche automatique de plus en plus puissantes (voir *Phaser*) et ne nécessitent généralement qu'une intervention minime de la part de l'utilisateur.

3.1.4 Affinement et validation

Une fois que la structure a été phasée et qu'un premier modèle est disponible, l'étape suivante est l'affinement des positions atomiques. Le but est de minimiser les différences entre le facteur de structure calculé (F_{calc}) et le facteur de structure observé (F_{obs}) en tenant compte des restrictions structurales et géométriques (longueurs, angles de liaison, torsions, etc.) qui sont introduites par la nature moléculaire du modèle. La fonction à minimiser (*target function*) peut être exprimée comme une somme pondérée de déviations/différences :

$$S = \sum_{hkl} W(hkl) (F_{obs}(hkl) - k|F_{calc}(hkl)|)^2$$

où k et W sont deux facteurs d'échelle. Les logiciels modernes d'affinement, comme *Refmac* (Murshudov 1997), ont adopté un nouveau modèle de type statistique nommé le *maximum likelihood* (ML) qui a remplacé la formule des moindres carrés (LS). Dans le ML, la valeur à maximiser est la *probabilité* que le facteur de structure observé F_{obs} soit en accord avec le modèle tout tenant compte des erreurs qui affectent les intensités mesurées et le modèle. La fonction ML est beaucoup plus puissante dans l'affinement de structures macromoléculaires

car surtout en début d'affinement le modèle n'est pas complet et peut contenir des erreurs macroscopiques. Pour chaque atome dans la structure, les paramètres qui interviennent dans le processus d'affinement sont le type d'atome, ses coordonnées dans l'espace (xyz), son facteur d'agitation thermique B et son occupation. Le *facteur d'accord cristallographique R* indique la progression de l'affinement de la structure. Ce paramètre est défini par l'équation suivante:

$$R = \frac{\sum_{hkl} |F_{obs}(hkl)| - k|F_{calc}(hkl)|}{\sum_{hkl} |F_{obs}(hkl)|}$$

Pour une structure correctement affinée à 2.0 Å, R devrait se situer en dessous de 20%. Des valeurs autour de 15% sont optimales et peuvent être dépassées avec des structures correctement affinées à haute, voire très haute résolution. Pratiquement, les erreurs macroscopiques du modèle sont toujours corrigés « à la main » par le cristallographe qui dispose pour cela de logiciels de modélisation moléculaire comme *Turbo-Frodo*, O (Jones, Zou et al. 1991) et plus récemment *Coot* (Emsley and Cowtan 2004).

La validation des modèles affinés est effectuée avec des logiciels tels que Whatif (Vriend 1990) qui analysent tous les paramètres structuraux comme les longueurs et angles de liaison, les torsions, la géométrie des chaînes latérales et le réseau de liaisons hydrogène. Tous les paramètres recueillis sont confrontés avec les valeurs standards des structures de protéines.

3.2 Analyse conformationnelle des oligosaccharides

On a déjà vu dans le premier chapitre quels sont les facteurs qui déterminent la structure tridimensionnelle des glucides et plus particulièrement des oligosaccharides. L'analyse conformationnelle d'un oligosaccharide (ou glycanne) complexe s'effectue par étapes. Il faut d'abord identifier les liaisons entre les monosaccharides qui composent la structure et analyser la surface d'énergie potentielle de chaque liaison glycosidique. La connaissance des conformations de plus basse énergie pour chaque fragment disaccharidique permet de construire un modèle (ou plusieurs) de la structure globale de la molécule dans une conformation de basse énergie. Ces modèles serviront de structures de départ pour l'analyse de la surface d'énergie potentielle à n -dimensions. La dernière étape est la génération des familles de conformations de basse énergie et l'analyse des résultats.

3.2.1 Champs de force pour les glucides

Les difficultés principales dans l'analyse conformationnelle des glucides sont la taille des structures à modéliser (souvent plusieurs centaines d'atomes), la présence d'effets particuliers (effet exo-anomérique et réseaux de liaisons hydrogène) et l'existence de plusieurs minima énergétiques. Les méthodes quantiques *ab-initio* donnent de bons résultats mais elles sont généralement limitées à l'analyse de petites molécules. Les méthodes semi-empiriques sont de préférence méthodes *ab-initio* très simplifiées. Le désavantage est qu'elles nécessitent un très bon ensemble de paramètres et ne donnent pas de résultat satisfaisant avec les sucres. Les méthodes de mécanique moléculaire (MM) sont les plus simples et permettent d'analyser de très grandes structures. Contrairement à la mécanique quantique (utilisée par les méthodes *ab initio* et semi-empiriques), la MM utilise les lois de la physique classique pour analyser les forces entre les atomes. La fonction d'énergie potentielle à minimiser dans un champ de force MM est définie comme une somme de différentes contributions énergétiques :

$$E_{pot} = \sum E_{bnd} + \sum E_{ang} + \sum E_{tor} + \sum E_{oop} + \sum E_{nb} + \sum E_{el}$$

où E_{bnd} et E_{ang} sont les énergies de déformation pour la longueur et l'angle de liaison, et sont généralement traités par une fonction harmonique. E_{tor} est l'énergie de la déformation des angles de torsion et E_{oop} est l'énergie pour les déformations *out-of-plane*. E_{nb} prend en compte les interactions nonliées et E_{el} est le terme pour l'énergie d'interaction électrostatique (forces de Coulomb). Contrairement aux méthodes *ab-initio*, la mécanique moléculaire est utilisable pour calculer les propriétés moléculaires qui ne dépendent pas des effets électroniques, qui ne sont pas pris en compte par les champs de force MM classiques. Malheureusement, l'effet exo-anomérique est un effet stereo-électrique qui a des conséquences importantes sur la conformation des sucres, notamment sur la géométrie de la liaison glycosidique. Il faut donc soit utiliser un champ de force incluant une correction pour l'effet exo-anomérique qui reproduise bien les conformations des sucres tel que MM3 (Allinger 1989), soit ajouter des paramètres énergétiques optimisés pour les sucres, comme cela a été fait pour les champs de force CHARMM, Amber ou autres (Imbert and Pérez 2000).

3.2.2 Calcul de cartes adiabatiques avec MM3

La surface d'énergie potentielle d'un disaccharide peut être représentée sur une carte d'énergie Φ , Ψ , tel que présentée dans le premier chapitre. Des cartes appelées « adiabatiques » sont calculées prenant en compte toutes les orientations possibles pour les groupements exocycliques de chaque monosaccharide. Pour le groupement hydroxyméthyle,

trois conformations (*gt*, *gg*, *tg*) sont prises en compte et les hydroxyles secondaires sont modélisés dans deux orientation différentes (*clock-* et *anticlock-wise*). Pour une liaison simple de type 1-2, 1-3 ou 1-4 entre deux pyranoses, 36 cartes différentes doivent être calculées. Pour une liaison 1-6, l'angle de torsion ω autour de la liaison C5-C6 doit être aussi pris en compte. Les angles de torsion glycosidique Φ et Ψ sont variés par incrémentation de 10 à 20° et à chaque orientation la structure entière est relaxée (à l'exception de l'orientation des angles de torsion) et son énergie est minimisée. Avec un incrément angulaire de 15°, on peut calculer que le nombre de structures à minimiser s'élève à 207360. A la fin, toutes les cartes calculées sont assemblées dans une carte globale, sur laquelle figure pour chaque conformation Φ et Ψ , l'énergie la plus faible trouvée parmi l'ensemble des conformations calculées. Cette procédure présente l'avantage de surmonter le problème du à l'existence de plusieurs minima d'énergie tout en prenant en considération tous les niveaux de flexibilité conformationnelle autour des liaisons glycosidiques.

3.2.3 Exploration de la surface multidimensionnelle : la méthode SCD

L'application de cette méthode de recherche exhaustive ne peut être envisagée que pour l'étude de structures oligosaccharidiques de complexité moyenne. Dans les cas des oligosaccharides complexes (plus de 15 résidus) le nombre de paramètres structuraux devient trop important, et la méthode précédente ne peut être appliquée, même si l'on dispose de moyens de calcul importants. Dans le cas particulier des oligosaccharides, la liberté conformationnelle est limitée par la présence des liaisons glycosidiques et par les autres interactions qui peuvent s'instaurer entre deux parties différentes de la molécule. Une méthode qui s'est révélée extrêmement puissante dans l'analyse de glucides complexes est la *Single Coordinate Driving* (SCD) implémentée dans le logiciel *CICADA* (Koca 1998). Cette méthode de recherche n'analyse pas toute la surface d'énergie potentielle mais est au contraire optimisée pour la recherche « active » des minima d'énergie. Les avantages de cette méthode sont nombreux: le temps du calcul dépend du nombre de degrés de liberté de façon polynomiale et ne croît pas de manière exponentielle comme dans la méthode *grid search*. Tous les chemins d'interconversion entre les différentes familles de conformations sont explorés. Enfin, l'algorithme de recherche consacre la majeure partie du temps du calcul à l'exploration des zones de basse énergie. (Fig. 3.6) Tout ceci permet une exploration réaliste et efficace de l'espace conformationnel, pour des temps de calculs qui restent raisonnables.

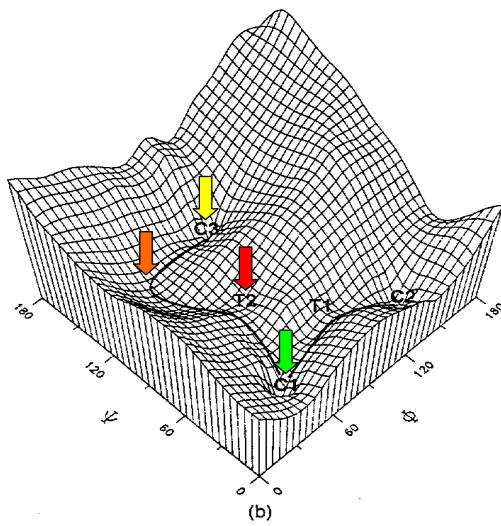


Figure 3.6 Exploration de la surface d'énergie potentielle d'un disaccharide avec la méthode SCD

3.3 Thermodynamique des interactions : microcalorimétrie de titration (ITC)

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), il convient de citer la microcalorimétrie de titration isotherme (*Isothermal Titration Calorimetry*). Cette méthode est la seule qui permette d'obtenir tous les paramètres caractérisant l'interaction en une seule expérience, et plus particulièrement la constante d'affinité (K_a), la stoechiométrie d'interaction (n) ainsi que les contributions enthalpique (ΔH) et entropiques ($T\Delta S$).

3.3.1 Le microcalorimètre et la technique

Le microcalorimètre (Fig. 3.7) contient deux cellules isolées de l'environnement externe par une enceinte adiabatique qui est maintenue à température constante pendant le temps de la titration. La première cellule, dite de référence, contient seulement du tampon. La protéine (ou la molécule à étudier) est placée dans la deuxième cellule, dite de mesure.

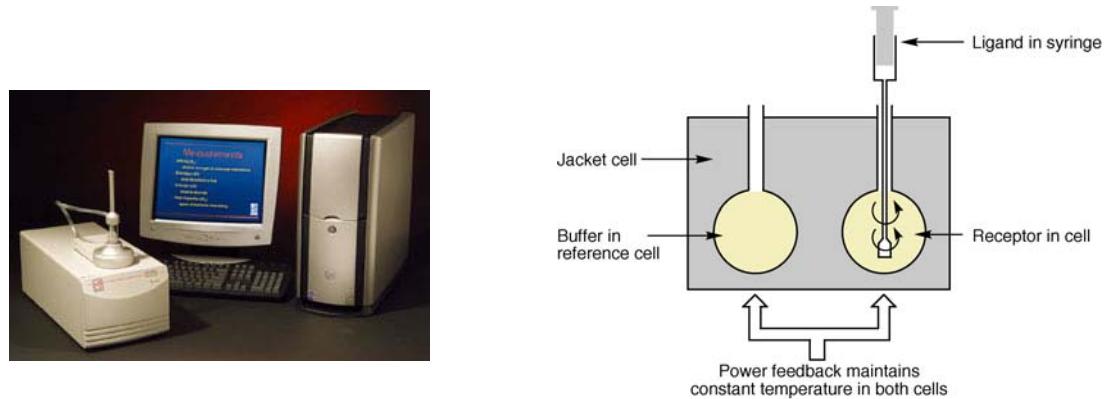


Figure 3.7 Microcalorimètre Microcal VP-ITC (gauche) et chemin de fonctionnement (droite)

Grâce à une seringue, on rajoute à intervalles réguliers des aliquotes de ligand (sucre ou autre) jusqu'à saturation de la solution de protéine. 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 mesure. Dans le but de rétablir la même température entre les deux cellules, l'appareil applique un signal électrique et c'est celui-ci qui est mesuré par le calorimètre. La puissance fournie ($\mu\text{cal/sec}$) est mesurée en fonction du temps, puis intégrée. (Fig. 3.8)

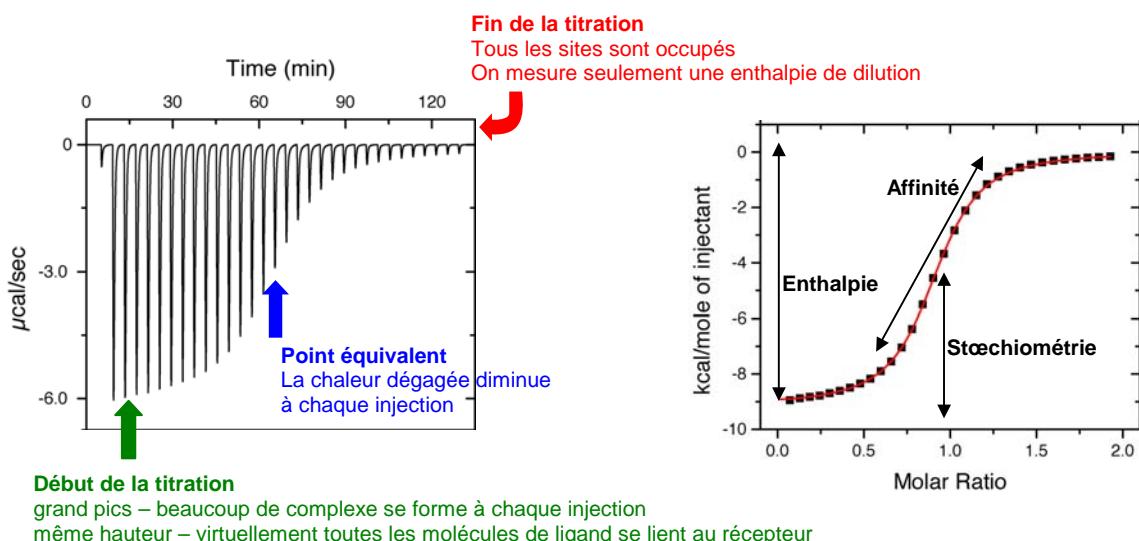


Figure 3.8 Données microcalorimétriques brutes (gauche), intégrées puis « fittées » avec la courbe sigmoïde (droite)

Au début de la titration, des quantités importantes de chaleur sont dégagées, indiquant la formation du complexe protéine-ligand. Par contre, vers la fin de la titration tous les sites de liaison sont désormais occupés et chaque ajout de ligand a comme seule conséquence sa dilution dans la cellule de mesure. La variation d'enthalpie ΔH , la constante d'association K_a et la stœchiométrie n peuvent être déterminées avec un ajustement non-linéaire de la courbe expérimentale utilisant un modèle approprié (Wiseman 1989). Pour une stœchiométrie simple de type 1:1 l'équation utilisée est la suivante :

$$\frac{dQ}{d[X]_t} = \frac{\Delta H^\circ V_o}{2} \left[1 + \left(\frac{1 - [X]_t/[M]_t - n/K_a[M]_t}{\sqrt{(1 + [X]_t/[M]_t + n/K_a[M]_t)^2 - 4[X]_t/[M]_t}} \right) \right]$$

Dans cette équation, $[X]_t$ et $[M]_t$ sont respectivement les concentrations totales du ligand et de la protéine, V_o est le volume effectif de la cellule de mesure et dQ est le différentiel de la quantité de chaleur échangée, calculée à chaque pointe de la courbe de titration.

La variation d'énergie libre ΔG° et la contribution entropique $T\Delta S^\circ$ sont calculées suivant :

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

L'enthalpie ΔH° influence directement l'amplitude de la courbe, tandis que la position du point d'inflexion donne la stœchiométrie n (Fig. 3.8). Par contre, la forme générale de la courbe dépend du paramètre de Wiseman c qui est défini comme :

$$c = n[M]K_a$$

où $n[M]$ représente la « concentration initiale de sites de liaison » en solution. Pour des valeurs de c supérieures à 10, la courbe montre une forme sigmoïdale, qui est nécessaire pour une détermination optimale des paramètres thermodynamiques. En pratique c doit être compris entre 10 et 100 car des valeurs supérieures génèrent des courbes ayant une pente trop élevée, ne permettant pas de mesurer K_a . On peut calculer que pour des valeurs de K_a de l'ordre de 10^4 M^{-1} (typiques des interactions lectine-sucre) une concentration initiale de protéine de 1mM est requise pour que c soit égal à 10. Pour une protéine de taille moyenne (20 kD) ceci équivaut à une solution à 20 mg/ml, concentration limite pour la solubilité d'une grande majorité des protéines. La grande quantité de récepteur nécessaire est donc la principale limitation de cette méthode pour l'étude des interactions lectine-sucre.

Si on dispose d'une quantité suffisante de récepteur pour effectuer plusieurs titrations à différentes températures, on peut envisager d'analyser la variation de la capacité calorifique ΔC_P en fonction de la température. ΔC_P est reliée à l'enthalpie par l'équation :

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

ΔC_P est un paramètre très important qui reflète le comportement du solvant pendant l'association et donne des informations importantes sur les forces qui guident la reconnaissance (Ambrosi 2005). On verra dans le septième chapitre les résultats obtenus avec la lectine CV-IIL de *Chromobacterium violaceum*.

Chapitre 4

Structure cristalline de la Tricolorine A

4.1 Introduction

Une partie d'importance fondamentale dans le développement des nouvelles substances bioactives est consacrée à la recherche, à l'identification et à la caractérisation de principes actifs extraits de plantes. C'est dans cette ligne générale de recherche qui se situe le travail mené sur l'analyse structurale d'un glycolipide, la *Tricolorine A*, présentée dans ce quatrième chapitre.

Ipomea tricolor est le nom d'une belle plante (Fig. 4.1) appartenant à la famille des *convolvulaceae* qui a toujours été utilisée dans l'agriculture traditionnelle des populations indigènes de l'Amérique centrale. Cette plante, connue également sous le nom de *Morning Glory*, a la propriété d'inhiber la croissance de la végétation parasite et donc a été utilisée pendant des siècles comme herbicide naturel. L'activité phytotoxique dérive d'une famille des résines (une dizaine de composés au total) contenues dans la plante et appelées *tricolorines* (Pereda-Miranda 1993; Bah 1996). Chimiquement ces molécules sont des glycolipides contenant un intéressant cycle macrolactonique qui est conservé dans toutes les molécules de cette famille. Ce qui est encore plus intéressant est que la présence de cette chaîne aliphatique semble être indispensable pour l'activité biologique de toutes les tricolorines analysées jusqu'à maintenant (Pereda-Miranda 2003). La Tricolorine A a été la première molécule isolée de la plante à être caractérisée par MS et RMN (Pereda-Miranda 1993; Bah 1996) et ensuite synthétisée par voie chimique (Lu, 1997; Larson 1997; Furstner 1998).

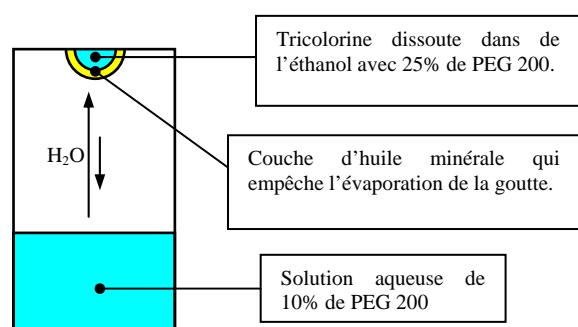


Figure 4.1 *Ipomea tricolor* (gauche) et diffusion de vapeur inverse (droite).

4.2 Résultats

La Tricolorine A a été purifiée à partir d'une extrait de *Ipomea tricolor* par chromatographie en phase liquide. Pour la cristallisation on a utilisé les méthodes développées pour la cristallisation des protéines dans le but de limiter la quantité de produit utilisée. Etant donné que cette molécule est insoluble dans l'eau, on a choisi d'utiliser une méthode de diffusion de vapeur *inverse* (Fig. 4.1) ou l'eau, qui est l'agent précipitant, passe du réservoir (solution moins concentrée) à la goutte (solution plus concentrée). En raison de la petite dimension des cristaux obtenus il a été nécessaire d'utiliser le rayonnement du synchrotron, puissant et très focalisé, pour la collecte de données.

La structure cristalline montre que les quatre molécules dans l'unité asymétrique sont associées de façon à maximiser les interfaces hydrophobes et hydrophiles. Il a été avancé l'hypothèse que l'activité phytotoxique des tricolorines pourrait dépendre du fait qu'elles sont capables de s'insérer dans les membranes cellulaires. L'arrangement observé dans la structure cristalline pourrait être maintenu en solution et dans ce cas l'orientation des surfaces hydrophobiques pourrait être compatible avec un modèle d'insertion parallèlement aux lipides de la membrane cellulaire. Bien sur, il s'agit d'une hypothèse qui nécessite d'autres études pour être validée. En conclusion, ce travail ouvre des voies sur le mécanisme d'action de ces molécules qui pourront amener à d'intéressantes applications dans le *drug design*.

Crystal Structure of Tricolorin A: Molecular Rationale for the Biological Properties of Resin Glycosides Found in Some Mexican Herbal Remedies**

Anna Rencurosi, Edward P. Mitchell, Gianluca Cioci,
Serge Pérez, Rogelio Pereda-Miranda,* and
Anne Imberty*

The traditional uses of several Mexican members of the morning glory family (Convolvulaceae), combined with ecological field observations, have been helpful in the design of an efficient approach for sampling plant materials and the selection of plants for investigation as potential sources of novel biodynamic natural products.^[1] The Mexican variety of the morning glory plant named “heavenly blue” (*Ipomea tricolor* Cav.) has been used for centuries as a cover crop in traditional mesoamerican agriculture because it inhibits the growth of invasive weeds. We now know that the phytotoxins involved are the resin glycosides, collectively called tricolorins.^[2,3] Until recently, the structural complexity of these mixtures seriously hampered the isolation of their individual constituents. The application of recycling high-performance liquid chromatography has allowed no less than 10 lipooligosacharides to be isolated from the aerial parts of “heavenly blue”.^[2-4] Tricolorin A was the first member of the series to be fully characterized through a combination of NMR and MS methods.^[2,3] This compound consists of the tetrasaccharide L-rhamnopyranosyl-(1→3)-O- α -L-rhamnopyranosyl-(1→2)-O- β -D-glucopyranosyl-(1→2)-O- β -D-fucopyranoside linked to jalapinolic acid to form a macrocyclic ester with a 19-membered ring (Scheme 1). Following the elucidation

[*] Dr. R. Pereda-Miranda

Departamento de Farmacia, Facultad de Química
Universidad Nacional Autónoma de México
Ciudad Universitaria, Coyoacán, 04510 D.F. (México)
Fax: (+52) 5-622-5329
E-mail: pereda@servidor.unam.mx

Dr. A. Rencurosi, G. Cioci, Dr. S. Pérez, Dr. A. Imberty
CERMAV-CNRS

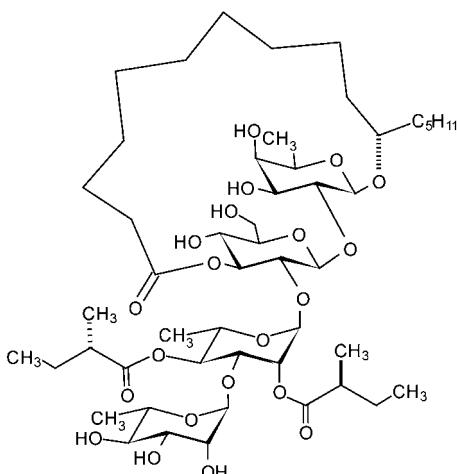
(affiliated with Université Joseph Fourier)
601 rue de la Chimie, BP 53, 38041 Grenoble, Cedex 9 (France)
Fax: (+33) 476-54-7203
E-mail: imberty@cermav.cnrs.fr

Dr. E. P. Mitchell
E.S.R.F. Experiments Division
BP 220, 38043 Grenoble, Cedex 9 (France)

[**] G.C. and A.R. are EEC doctoral and postdoctoral fellows, respectively (HPRN-CT2000-00001). R.P.-M. acknowledges financial support from Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de Mexico (Grant nos. IN200902-2 and IX234504-1), and Consejo Nacional de Ciencia y Tecnología (Grant no. 39951-Q). We thank the European Synchrotron Radiation Facility, Grenoble, for access to synchrotron data collection facilities.



Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.



Scheme 1. Structure of tricolorin A.

tion of its chemical structure, total syntheses of tricolorin A in its natural enantiomeric form were developed by Larson and Heathcock, and by Lu and collaborators. Both teams used a macrolactonization approach.^[5,6] Fürstner and Müller later used a ring-closing-metathesis strategy to form the macro-lactone moiety and succeeded in synthesizing tricolorin A^[7] and several other resin glycosides.^[8,9]

Tricolorin A demonstrates several biological activities of therapeutic interest, such as mammalian cytotoxicity against cultured P-338 and human breast cancer cells,^[2] antibacterial activity against *Staphylococcus aureus* and *Mycobacterium tuberculosis*,^[1] and antifungal potential correlated to its (1→3) β -D-glucan synthase inhibitory activity.^[10] In the cover crop, tricolorin A acts as a nonprotonophoric uncoupler of photophosphorylation and inhibits electron transport in the photosystem II of chloroplasts.^[11] All the bioactivities of this lipopolysaccharide are associated with its macrocyclic structure; the glycosidic acid derived by saponification of the lactone has been shown to be inactive in all resin glycosides biologically tested.^[1]

The difficulty involved in obtaining a useable pure sample of an individual resin glycoside,^[12] in addition to that related to oligosaccharide crystallization,^[13] represented an enormous challenge for the structural investigation reported herein. Protein crystallization techniques were used to avoid wastage of the isolated tricolorin A (20 mg). The compound is insoluble in water, which was therefore selected as the precipitating agent.^[14] The size of the crystal unit cell indicated the presence of four independent tricolorin A molecules per asymmetric unit.^[15] Each unit contains a total of 284 nonhydrogen atoms and is therefore similar in size to a small protein of about 30 amino acids. The size of the asymmetric unit, together with that of the crystal, demanded the use of intense synchrotron radiation to collect the diffraction data. The SIR2002 method was used to solve the structure. Refinement with the SHELX program indicated the presence of 18 water molecules in the asymmetric unit in addition to the four independent tricolorin molecules (Figure 1). All atoms in the structure were clearly visible in the electron density maps, with the exception of two carbon

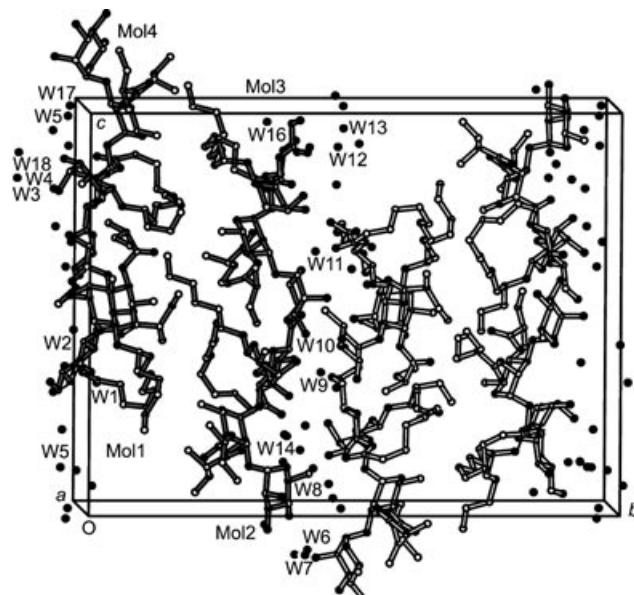


Figure 1. Graphical representation of the unit cell. The contents of the asymmetric unit are shown with gray bonds and the molecules are labeled as Mol1 to Mol4 for tricolorin A and W1 to W18 for water.

atoms in the lipid part of molecule 4. A detailed view of one of the tricolorin A molecules is shown in Figure 2, with ellipsoids representing thermal vibration. The largest temperature

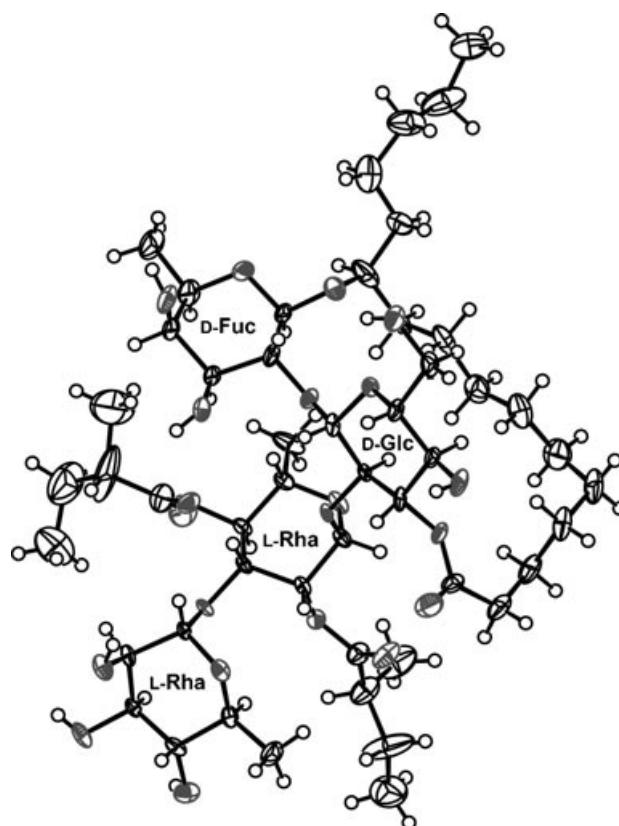


Figure 2. ORTEP representation of one molecule of tricolorin A, drawn with the Platon software.^[28] The ellipsoids of thermal vibration represent a probability of 50%.

factors were measured for the two methylbutyroyl groups that form esters with the internal rhamnose unit. The macrocyclic aglycon core of molecule 4 displayed slightly more disorder than those of the other molecules.

Figure 3 shows a superposition of the four tricolorin A independent molecules, all of which share the same global shape, albeit with slightly different conformations of the aglycon moiety stacked under the β -D-glucopyranosyl-(1 \rightarrow 2)-O- β -D-fucopyranoside moiety. The glycosidic linkages are superimposed on the corresponding energy maps^[16] (Figure 3c). Although the energy maps of the three disaccharides differ, all three display low-energy regions centered around a Φ -axis *gauche* conformation, as dictated by the *exo*-anomeric effect. A higher level of conformational freedom is apparent along the Ψ axis: the lowest energy region corresponds to a plateau ranging from Ψ =−60 to 180°. While the externally placed L-rhamnopyranosyl-(1 \rightarrow 3)-O- α -L-rhamnopyranoside moiety shows rather different conformations for each of the four molecules in the asymmetric unit, the internal trisaccharide subunit (α -L-rhamnopyranosyl-(1 \rightarrow 2)-O- β -D-glucopyranosyl-(1 \rightarrow 2)-O- β -D-fucopyranoside) has limited conformational freedom. The four molecules each display slightly different sets of torsion angles (see the Supporting Information) but all of these angles yield very similar pseudo-elongated shapes for the macrocyclic aglycon portion of the molecule between the lactone end and the anomeric oxygen atom of the fucose unit. In contrast, the terminal pentyl chain is very flexible. We compared the observed conformations of the tricolorin A molecules to those of the only related molecule that has been crystallized, a synthetic chemical

intermediate of tricolorin A consisting of the β -D-glucopyranosyl-(1 \rightarrow 2)- β -D-fucopyranoside subunit with all its hydroxy groups protected.^[17] The first notable difference is that the lack of the amphipathic properties of the natural sample limits the solubility of the analogue to low-polarity organic solvents. Five independent molecules were refined for the asymmetric unit of the analogue, as opposed to the four molecules found in tricolorin A crystals. This difference resulted in a totally different conformation and molecular packing for the analogue. The lack of water molecules produces a piled parallel arrangement of glycoside residues on one side of the analogue structure, whilst the macrolactone rings stack on the other side with alternating alpha and beta faces. The natural compound structure consists of a succession of hydrophilic and hydrophobic layers.

The most notable feature of tricolorin A in the solid state is the anisotropic repartitioning of the hydrophobic and hydrophilic sections in the crystal packing arrangement (Figure 4). One face of the molecule exhibits an almost flat hydrophobic wall formed by the aglycon unit, the methyl group of the fucose unit, and the three lipophilic inner rhamnose residues (the methyl group and the two esterified methylbutyric acid groups). The other face presents two small hydrophilic areas: one composed of the hydroxy groups of the fucose and glucose residues and the other of those of the external rhamnose unit. The 18 water molecules form a dense network that creates a dividing layer between the hydrophilic faces of the structure (Figure 4). The high water content of the crystal, which is similar to that found in the accepted view of protein crystals, means that the tricolorin A molecular

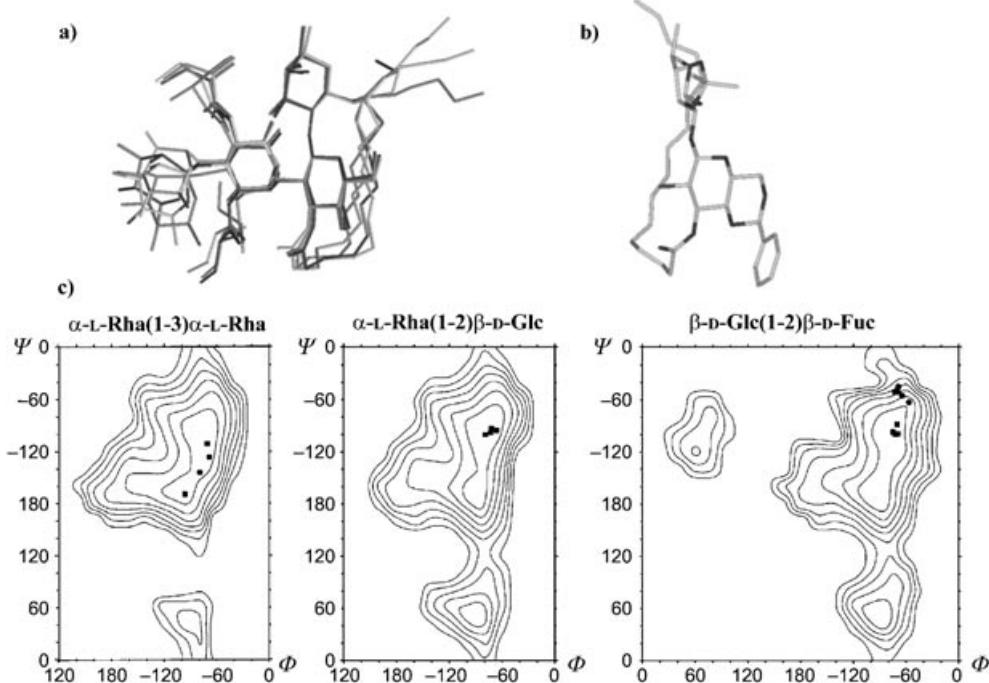


Figure 3. a) Superposition of the four independent molecules of tricolorin A. b) One of the molecules from the crystal structure of the synthetic analogue,^[17] shown with the glucose ring in the same orientation as in (a). c) Glycosidic linkage energy maps for each of the constitutive disaccharide subunits of tricolorin A. The conformations observed in the crystal structure of tricolorin A are indicated by squares and those of the synthetic analogue by circles.

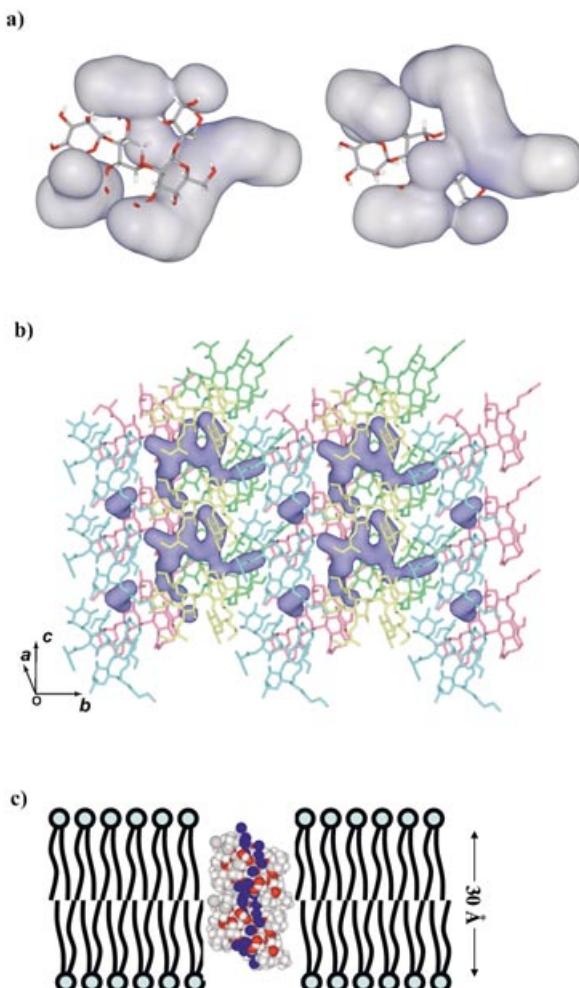


Figure 4. a) Amphiphilic properties of tricolorin A: two orthogonal views of the molecule showing the contours of the accessible surface of the hydrophobic region. b) Graphical representation of packing in the tricolorin A crystal structure: view along the *b* axis with clusters of water molecules represented by their accessible surface (blue). c) Insertion membrane model of the minimal crystal environment made up of four tricolorin A molecules.

conformation in the solid state is not dominated by intermolecular forces and hence might be indicative of the conformation in solution and in supermolecular aggregates.

It has been suggested that the cytotoxic properties of the resin glycosides could be caused by their ability to perturb cell membranes through nonselective pore formation. These compounds have a somewhat peculiar organization in aqueous solution and form micelles or aggregates comparable to those displayed by tricolorin A in the crystalline state. It is therefore of interest to compare the architecture of the water channel formed by the four tricolorin A molecules, which pile up in two pairs along the *c* axis, with the spatial arrangement of a lipid bilayer. The hydrophobic surface exposed externally and elongated along the axis of the water channel is ideally oriented for parallel interaction with the lipids of a biological membrane upon insertion of tricolorin A into the membrane. This hypothesis differs from predictions made by molecular dynamics simulations and NMR studies of micellar solutions

of calonyctin A, another plant-growth-regulating resin glycoside,^[18] which inserts perpendicularly into micelle lipid membranes. The total extension of the channel created by tricolorin A molecules in our study is about 30 Å, which is comparable to the width of a biological membrane such as that of the hydrocarbon core elongation in fluid phospholipid bilayers.^[19] A schematic representation of our insertion model is depicted in Figure 4.

The macrolactone ring is essential to the biological activities of all resin glycosides^[1] since without it the spatial arrangement needed to form aggregates, as well as the above-mentioned channels, probably could not take place. The dried tubers of the medicinal members of the morning glory family yield a purgative remedy of which the main active ingredients are resin glycosides.^[1] All the biological effects displayed by this type of amphipathic oligosaccharide suggest that the activity could be the result of a possible ion flux perturbation in the target cell membrane induced by nonselective pore formation, as illustrated by the insertion model. This model for transmembrane channel formation is based on the crystal structure of tricolorin A and is still speculative in nature. Experimental and theoretical studies are called for to provide substantiation for this hypothesis, as well as to investigate whether other types of architecture could allow better interaction. This first crystallographic analysis of a natural convolvulaceous resin glycoside not only opens avenues for further structural investigations but may also lead to important applications of such compounds in drug design.

Received: April 14, 2004

Keywords: conformation analysis · glycolipids · glycosides · natural products · structure elucidation

- [1] R. Pereda-Miranda, M. Bah, *Curr. Top. Med. Chem.* **2003**, 3, 111.
- [2] R. Pereda-Miranda, R. Mata, A. L. Anaya, D. B. Wickramaratne, J. M. Pezzuto, A. D. Kinghorn, *J. Nat. Prod.* **1993**, 56, 571.
- [3] M. Bah, R. Pereda-Miranda, *Tetrahedron* **1996**, 52, 13063.
- [4] M. Bah, R. Pereda-Miranda, *Tetrahedron* **1997**, 53, 9007.
- [5] D. P. Larson, C. H. Heathcock, *J. Org. Chem.* **1997**, 62, 8406.
- [6] S. F. Lu, Q. O'Yang, Z. W. Guo, B. Yu, Y. Z. Hui, *J. Org. Chem.* **1997**, 62, 8400.
- [7] A. Fürstner, T. Müller, *J. Org. Chem.* **1998**, 63, 424.
- [8] A. Fürstner, T. Müller, *J. Am. Chem. Soc.* **1999**, 121, 7814.
- [9] A. Fürstner, F. Jeanjean, P. Razon, C. Wirtz, R. Mynott, *Chem. Eur. J.* **2003**, 9, 307.
- [10] M. V. Castelli, J. C. Cortes, A. M. Escalante, M. Bah, R. Pereda-Miranda, J. C. Ribas, S. A. Zacchino, *Planta Med.* **2002**, 68, 739.
- [11] L. Achnine, R. Pereda-Miranda, R. Iglesias-Prieto, R. Moreno-Sánchez, B. Lotina-Hennsen, *Physiol. Plant.* **1999**, 106, 246.
- [12] The experimental procedures, including preparative HPLC, handling of the plant material, and extraction of the resin glycosides from the aerial parts of *Ipomoea tricolor*, have been described previously.^[2,3] Preliminary fractionation of the crude resins (100 mg) was achieved by standard column chromatography. The chloroform-soluble pool was subjected to preparative HPLC (Waters column, 150 × 19 mm, μBondapak-amino, 10 μm). This separation was performed to eliminate impurities appearing before and after the selected peak ($t_R = 18$ min). Isocratic elution was applied, with CH₃CN/H₂O (92:8) and a flow rate of 6 mL min⁻¹. The tricolorin A peak was collected by

heart-cutting and independently reinjected (40 mg) into the same column. The HPLC system was operated in the recycle mode to achieve the maximal possible purity of the sample.^[3] This process of purification was monitored by using a refractive index detector. The sensitivity setting of the refractometer was increased from $8\times$ to $64\times$ to facilitate the detection of all minor impurities. Elution was conducted isocratically with $\text{CH}_3\text{CN}/\text{H}_2\text{O}$ (95:5; flow rate = 8 mL min^{-1}) and complete separation of tricolorin A to homogeneity (20 mg) was achieved after twelve consecutive cycles on the same aminopropyl column.

- [13] S. Pérez, C. Gautier, A. Imbert in *Oligosaccharides in Chemistry and Biology: A Comprehensive Handbook* (Eds.: B. Ernst, G. Hart, P. Sinay), Wiley-VCH, Weinheim, **2000**, pp. 969.
- [14] The first microcrystals were obtained by vapour diffusion with a modified version of the hanging drop method. An ethanolic solution (3 μL) of tricolorin A (20 mg mL^{-1}) was mixed with pure poly(ethyleneglycol) 200 (PEG 200; 1 μL , Sigma) and deposited on a glass coverslide. This drop of solution was covered with a layer of mineral oil (Sigma), then the slide was sealed above a reservoir containing a solution of 10% PEG 200 in water. Crystals suitable for x-ray analysis were grown by using the same method and a sample solution (2 μL , 10 mg mL^{-1} in EtOH) mixed with PEG 200 (2 μL) and mineral oil (2 μL). The drop was seeded with the microcrystals obtained previously. The reservoir solution was composed of 75% water, 10% PEG 200, and 15% EtOH.
- [15] A needle-shaped crystal ($0.5 \times 0.01 \times 0.01 \text{ mm}^3$) was soaked in a 60% PEG 6000/water solution for three minutes then cryo-cooled at 100 K. Data were collected from a single crystal on beam line ID29 ($\lambda = 0.8157 \text{ \AA}$) at ESRF (Grenoble) by using an ADSC Q210 CCD detector with a resolution of 0.87 \AA ($\Theta_{\max} = 28.3^\circ$). A total of 38162 reflections were measured, of which 18424 were independent, with $R_{\text{int}} = 0.071$. Data were processed with the MOSFLM package.^[20] The structure was solved by direct methods (SIR-2002).^[21] Refinement was performed with the Shelx-97 program.^[22] The nonhydrogen atoms of the sample (four monomers with the formula $\text{C}_{50}\text{H}_{86}\text{O}_{21}$ and 18 H_2O) were refined with anisotropic displacement parameters, except the water oxygen atoms, for which isotropic refinement parameters were used. A few restraints were included on selected C–C distances. Hydrogen atoms were placed on the model molecules (except the water molecules), which yielded a total of 2770 parameters. All 18424 independent reflections were used in the full matrix least-squares calculations against F^2 . Final refinement cycles yielded factors $R1 = 0.0998$ and $wR = 0.2283$ for 16993 reflections with $I > 2\sigma(I)$. The crystals belong to space group $P2_1$ and have the cell dimensions $a = 14.025(1)$, $b = 33.337(1)$, and $c = 25.512(1) \text{ \AA}$. $\beta = 91.07(1)^\circ$, $V = 11926.1(1) \text{ \AA}^3$, $Z = 8$, $\rho_{\text{calcd}} = 1.211 \text{ g cm}^{-3}$. CCDC 228071 contains the supplementary crystallographic data for this paper. These data can be obtained free of charge via www.ccdc.cam.ac.uk/conts/retrieving.html (or from the Cambridge Crystallographic Data Centre, 12, Union Road, Cambridge CB2 1EZ, UK; fax: (+44) 1223-336-033; or deposit@ccdc.cam.ac.uk).
- [16] Energy maps were calculated for each constituent disaccharide moiety as a function of two glycosidic linkage torsion angles defined as $\Phi = \theta$ (O5–C1–O1–Cx) and $\Psi = \theta$ (C1–O1–Cx–C(x+1)). Each disaccharide was built with the POLYS software^[23] and energies were calculated by using the MM3 program^[24,25] and employing a previously described procedure^[26] involving full optimization of the structure at each point of the (Φ, Ψ) map except for the two driven angles. To allow consideration of the three possible orientations of the hydroxymethylene group of the glucose unit and the clockwise or anticlockwise possibilities for the hydrogen bonding network around each ring, several starting structures, a step of 20° , and a dielectric constant $\epsilon = 80$ were used for these calculations. The single relaxed maps were combined to provide unique adiabatic maps by a procedure designed by our group.^[26] The corresponding plots were generated with the X-Farbe program.^[27]
- [17] C. W. Lehman, A. Fürstner, T. Müller, *Z. Kristallogr.* **2000**, *215*, 114.
- [18] Z. H. Jiang, A. Geyer, R. R. Schmidt, *Angew. Chem.* **1995**, *107*, 2730; *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2520.
- [19] M. C. Wiener, S. H. White, *Biophys. J.* **1992**, *61*, 437.
- [20] A. G. W. Leslie, *Joint CCP4 + ESF-EAMCB Newslett. Protein Cryst.* **1992**, *26*.
- [21] M. C. Burla, M. Camalli, B. Carrozzini, G. L. Cascarano, C. Giacovazzo, G. Polidori, R. Spagna, *J. Appl. Crystallogr.* **2003**, *36*, 1103.
- [22] G. M. Sheldrick, University of Gottingen, **1997**.
- [23] S. Perez, M. Kouwijzer, K. Mazeau, S. B. Engelsen, *J. Mol. Graphics* **1996**, *14*, 307.
- [24] N. L. Allinger, Y. H. Yuh, J.-H. Lii, *J. Am. Chem. Soc.* **1989**, *111*, 8551.
- [25] N. L. Allinger, M. Rahman, J.-H. Lii, *J. Am. Chem. Soc.* **1990**, *112*, 8293.
- [26] A. Imbert, S. Pérez, *Chem. Rev.* **2000**, *100*, 4567.
- [27] A. Preusser, *ACM Trans. Math. Software* **1989**, *15*, 79.
- [28] A. L. Spek, *J. Appl. Crystallogr.* **1988**, *21*, 578.

Chapitre 5

Analyse conformationnelle des *O*-glycannes par la méthode CICADA

5.1 Introduction

La glycoprotéine de Tamm-Horsfall (THP) est la protéine la plus abondante secrétée dans les reins des mammifères, y compris l'homme (Tamm 1950; Kumar 1990). Cette véritable glycoprotéine (94 kDa, dont 28% de glycannes) est attachée dans l'endothélium par une ancre GPI. La THP est aussi secrétée dans l'urine où elle est présente avec une concentration de 50 à 200 mg/L. Les rôles physiologiques de cette glycoprotéine ont été au centre de nombreuses études mais ils n'ont pas encore été démontrés avec certitude. La THP pourrait jouer un rôle comme inhibiteur des infections bactériennes dans l'appareil urinaire. De par sa propriété de s'agréger sous forme de gel, la THP pourrait être très importante dans de nombreux processus physiologiques et pathologiques rénaux.

La THP est fortement glycosylée et montre des structures glycanniques très variées, distribuées sur 7 sites de glycosylation (van Rooijen 2001). Il a été observé récemment que le motif de glycosylation change drastiquement avec la grossesse. La THP normale présente généralement des structures N-glycannes tandis que chez la femme enceinte, cette glycoprotéine (alors appelée *uromoduline*) comporte des O-glycannes complexes. Ce changement de la glycosylation serait responsable de la forte augmentation (x 13) du pouvoir immunomodulateur de la glycoprotéine (Easton 2000). Douze structures différentes de O-glycannes ont été séquencées chez l'uromoduline, et la plus complexe est formée par trois oligosaccharides de type sialyl Lewis^X reliés à un noyau (core) de type II. Par ailleurs, il a été montré que ce type de O-glycannes est une des cibles d'adhésion pour les selectines de type E et P (Mitoma 2003).

Ce cinquième chapitre décrit l'analyse conformationnelle de ce O-glycanne, qui est un des plus grands oligosaccharides jamais soumis à une exploration complète de l'espace conformationnel.

5.2 Résultats

Les surfaces d'énergie potentielle de chacun des 14 disaccharides différents constituant le O-glycanne ont été analysées avec la méthode MM3 décrite dans le troisième chapitre. Les conformations de basse énergie ont été utilisées pour la construction des différentes structures de départ. La surface multidimensionnelle de l'oligosaccharide a été analysée avec la méthode SCD, implémentée dans le logiciel CICADA (Koca 1998). Les résultats ont montré l'existence de quatre grandes familles de conformations qui réunissent 95% des conformations de basse énergie. Les propriétés moléculaires (charge électrostatique, lipophilicité et liaisons hydrogène) de chaque famille ont été étudiées. Ce O-glycanne est un exemple intéressant de structure dendrimérique, et la distribution dans l'espace des trois épitopes sialyl Lewis^X pourrait former une « présentation» idéale pour l'interaction multivalente avec les E- et P-selectines.

Conformational analysis of complex oligosaccharides: the CICADA approach to the uromodulin O-glycans

Gianluca Cioci,^{a,*} Alain Rivet,^a Jaroslav Koča^{a,b} and Serge Pérez^{a,*}

^a*Centre de Recherches sur les Macromolécules Végétales, CNRS and Joseph Fourier University, IFR 2607, BP 53, F-38041 Grenoble, France*

^b*National Centre for Biomolecular Research, Masaryk University, CZ-61137 Brno, Czech Republic*

Received 29 September 2003; accepted 12 December 2003

Abstract—Uromodulin is the pregnancy-associated Tamm–Horsfall glycoprotein, with the enhanced ability to inhibit T-cell proliferation. Pregnancy-associated structural changes mainly occur in the O-glycosylation of this glycoprotein. These include up to 12 glycan structures, made up of an unusual core type 2 sequence terminated with one, two, or three sialyl Lewis^x sequences; this type of O-glycans could serve as E- and P-selectin ligands. The present work focuses on the most complex one; a tetradecamer made up of a type 2 core carrying three sialyl Lewis^x branches. Five different monosaccharides are assembled by 14 glycosidic linkages. The conformational behavior of the constituting disaccharide segments was evaluated using the flexible residue procedure of the MM3 molecular mechanics procedure. For each disaccharide, the adiabatic energy surface, along with the local energy minima were established. All these results were used for the generation, prior to complete optimization of the tetradecamer. This was followed by a complete exploration of conformational hyperspace throughout the use of the single coordinate method as implemented in the CICADA program. Despite the potential flexibility of the tetradecasaccharide, only four conformational families occur, accounting for more than 95% of the total low energy conformations. For each family, the molecular properties (electrostatic, lipophilicity, and hydrogen potential) were studied. The shape of the tetradecasaccharide is best described as a flat ribbon, flanked by three branches having terminal sialyl residues. Two of the branches interact through nonbonded interactions, bringing further energy stabilization, and limiting the conformational flexibility of the sialyl residues. Only one branch maintains the original conformational features of sialyl Lewis^x. This O-glycan can be seen as a fascinating example of ‘dendrimeric’ structure, where the spatial arrangement of three S-Le^x epitopes may favor its complementary ‘presentations’ for the interactions with E- and P-selectins.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Core 2 branched O-glycans; Sialyl Lewis^x

1. Introduction

Tamm–Horsfall (THp) is the major glycoprotein produced by the kidney,^{1,2} where it is expressed via a glycosylphosphatidylinositol (GPI) anchor on the endothelium of the thick ascending limb of the loop of the Henle. This 94 kDa glycoprotein can be released from its membrane anchor through proteases or phospholipase and is excreted in urine. The physiological roles of THp are still under dispute, and many

hypotheses have been formulated. THp could act as an inhibitor of microbial infection of the urinary tract and of antigen-specific T-cell proliferation. Also, the aggregation and gel formation capabilities of THp could play important roles in several physiological and pathological states of the kidney. Pregnancy-associated THp is called uromodulin.³ This is an immunosuppressive molecule whose ability to inhibit T-cell proliferation is increased 13-fold during pregnancy.⁴ THp exhibits extremely large glycosylation microheterogeneities, as over 150 glycan structures that are mainly sialylated and/or sulfated to varying extents have been found on seven glycosylation sites.⁵ The carbohydrate part can account for 28% of the total weight of the THp. Mass

* Corresponding authors. Tel.: +33-4-76037630; fax: +33-4-76037629; e-mail: gianluca.cioci@cermav.cnrs.fr

spectrometric strategies have recently uncovered pregnancy-associated changes in the O-glycosylation of this glycoprotein. THp from nonpregnant females and males expresses primarily core 1 type O-glycans terminated with either sialic acid or fucose, but not with the sialyl Lewis^x epitope. By contrast, the O-glycans linked to uromodulin include unusual core type 2 glycans terminated with one, two, or three sialyl Lewis^x sequences (Fig. 1). These changes in the glycosylation pattern could account for the enhanced immunomodulatory effects of uromodulin.⁶ Meanwhile, different studies have established that sialyl Lewis^x (S-Le^x) in core 2 branched O-glycans serves as an E- and P-selectin ligand while S-Le^x containing type 1 core O-glycans are potential ligands of L-selectin.⁷ The Le^x blood-group antigens have been the subjects of many studies in the past and several NMR and conformational studies converged to the assumption that these molecules exist in just a few, if not only one, conformations in solution.⁸ This does not mean that these oligosaccharides are completely rigid but that their glycosidic torsion angles exhibit small oscillations around the average values.⁹ In the case of S-Le^x, its structure has been idealized as a rigid Le^x core linked to sialic acid via a flexible bond, which allows the co-existence of not more than two or three conformations accessible to the molecule.¹⁰ The S-Le^x has already been extensively studied throughout molecular mechanics calculations.^{8,10} Therefore, from both the computational and biological point of view, it is of interest to see how the interactions with neighboring groups in complex oligosaccharides can modulate or alter its conformational behavior.

There are 12 proposed structures for the uromodulin O-glycans and in this work we focused our attention on the most complex one. This molecule, which we named

Core2(S-Le^x)₃, is a tetradecamer made up of a type 2 core, which is linked to three S-Le^x groups (Fig. 1). Five different monosaccharides are assembled by 14 glycosidic linkages in quite a unique fashion to give a very complex oligosaccharide of 366 atoms.

The treatment of a glycan at this level of complexity can be undertaken using the following route: (a) identification of the monosaccharide constituents; (b) identification of disaccharide segments of Core2(S-Le^x)₃; (c) molecular modeling of each segment, and description of the potential energy surfaces; (d) collection of the 3-D structures of low energy conformers; (e) building realistic possible starting conformations from assembly of the low energy conformers; (f) exploration of potential energy hyperspace; (g) clustering of the low energy conformers into families; (h) analysis and depiction of representative within each low energy family. The present investigation makes use of well-established computational protocols based on the molecular mechanics MM3¹¹ and single coordinate driving method (SCD) as implemented in the CICADA program.^{12,13} There are three main advantages of the method over other conformational search algorithms: (i) the computer time required for a calculation has a polynomial dependence on the degrees of freedom, in contrast to the grid search, which has an exponential one; (ii) all the interconversion pathways between families of conformations are explored but; (iii) the algorithm spends almost its time in the essential highly populated areas, giving a realistic and time-inexpensive description of conformational space. It has been shown that CICADA is an efficient tool to search conformational space, and in the case of small and middle sized oligosaccharides it produces reliable results, which are comparable to a carefully performed grid search, but

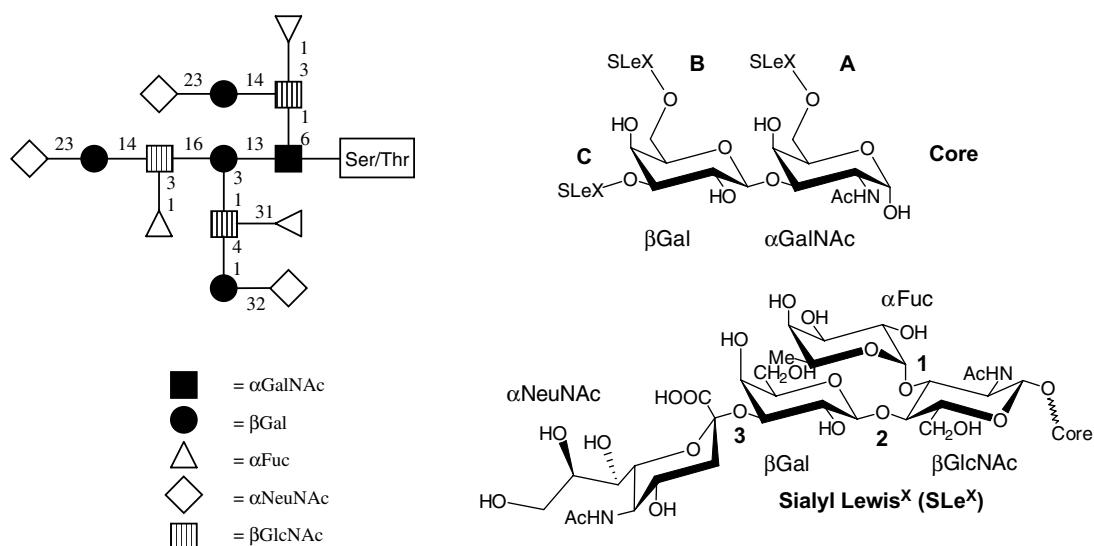


Figure 1. On the left, the structure of the whole molecule. On the right, the representations of the core disaccharide and of the S-Le^x structure. The S-Le^x groups are linked to the core structure via the linkages labeled A, B, and C.

much more time efficient.^{14–16} Despite that, an SCD search with a very complex molecule has never been carried out before. In the present paper, we present the methodology of the study together with information about the three-dimensional structure of the molecule and some biomolecular implications that can be extrapolated from the results of the calculation.

2. Computational procedure

2.1. Nomenclature

The recommendation and symbols proposed by the Commission on Biochemical Nomenclature are used throughout this paper. The torsion angles describing the glycosidic linkages are defined as $\Phi = (\text{O}5\text{--C}1\text{--O}1\text{--C}'\text{X})$, $\Psi = (\text{C}1\text{--O}1\text{--C}'\text{X}\text{--C}'\text{X}+1)$, and $\omega = (\text{O}1\text{--C}6\text{--C}5\text{--O}1)$.

The labeling of the different glycosidic linkages was made according to the following rules. The core disaccharide is $\beta\text{-Gal}(1 \rightarrow 3)\text{-}\alpha\text{-GalNAc}$. With reference to Figure 1, the three branched S-Le^x and the linkages from the core to them are labeled as A, B, and C, respectively. Within each S-Le^x moiety, the $\alpha\text{-Fuc}(1 \rightarrow 3)\text{-}\beta\text{-GlcNAc}$ linkage is referred to as 1, the $\beta\text{-Gal}(1 \rightarrow 4)\text{-}\beta\text{-GlcNAc}$ as 2 whereas the $\alpha\text{-NeuNAc}(2 \rightarrow 3)\text{-}\beta\text{-Gal}$ is referred to as 3. For example, the $\beta\text{-Gal}(1 \rightarrow 4)\text{-}\beta\text{-GlcNAc}$ linkage on the S-Le^x group attached ($1 \rightarrow 3$) to the core disaccharide will be referred to as C2. The nomenclature is summarized in Table 1.

2.2. Building blocks

The starting geometry and suggestion of each of the five different constituting monosaccharides was taken from the MONOBANK data base¹⁷ or constructed. Each residue has been submitted to a complete energy optimization through MM3.

2.3. Relaxed map of disaccharides

The conformational space of each disaccharide segment was explored by stepping the glycosidic linkage Φ , Ψ , and ω torsion angles in 15° increments over the whole angular range. Adiabatic contour maps were drawn for each disaccharide by using different orientations of the

pendent groups. In particular, three staggered positions of the hydroxymethyl groups (*gt*, *gg*, and *gt*) and the two possible networks of hydrogen bonds around the sugar rings (clock- and anticlockwise). As for the sialic acid moiety, two orientations were considered for the acidic group. At each conformational microstate, a geometry optimization was performed by allowing the coordinates of each atom to vary except those defining the Φ and Ψ (and ω) torsion angles. Only the lowest energy conformer at each (Φ , Ψ) point is used for the Ramachandran-like contour plots. This procedure has the advantage of overcoming the well-known multiple-minima problem of the potential energy hypersurface due to the specific orientations of the pendant groups, and also of fully describing the conformational flexibility around the glycosidic linkages. The exact position of the different minima was determined after additional minimizations by removing the constraints on the Φ and Ψ torsion angles.

2.4. Oligomeric structure

Information from the maps computed for the parent disaccharides was used to construct the oligomeric structure. The starting geometry-(ies) of $\text{Core2}(\text{S-Le}^x)_3$ was constructed using the POLYS program.¹⁸

2.5. Exploring the hyperdimensional potential energy surface using CICADA

The potential energy hypersurface of $\text{Core2}(\text{S-Le}^x)_3$ was explored by the CICADA program using the MM3 force field. Several starting structures were used as input for the CICADA analysis with different orientations around the glycosidic torsion angles, based on the low-energy conformations found in the relaxed maps of the corresponding disaccharides. During the CICADA calculations, all glycosidic torsion angles were driven by 15° step as well as those related to the orientation of the COOH and of the primary hydroxyl groups. The remaining torsion angles were only monitored. When CICADA detects a minimum, the conformation is fully optimized. The resulting structure is compared with the previously stored ones and stored if not yet detected. Structures corresponding to energy maxima, that is, the transition states, are also stored. For $\text{Core2}(\text{S-Le}^x)_3$,

Table 1. Nomenclature of all the disaccharides

Core to:	Sialyl Lewis ^x labeling		
	A $\beta\text{-GlcNAc}(1 \rightarrow 6)\text{-}\alpha\text{-GalNAc}$	B $\beta\text{-GlcNAc}(1 \rightarrow 6)\text{-}\beta\text{-Gal}$	C $\beta\text{-GlcNAc}(1 \rightarrow 3)\text{-}\beta\text{-Gal}$
Inside the S-Le^x :			
1	$\alpha\text{-Fuc}(1 \rightarrow 3)\text{-}\beta\text{-GlcNAc}$	$\alpha\text{-Fuc}(1 \rightarrow 3)\text{-}\beta\text{-GlcNAc}$	$\alpha\text{-Fuc}(1 \rightarrow 3)\text{-}\beta\text{-GlcNAc}$
2	$\beta\text{-Gal}(1 \rightarrow 4)\text{-}\beta\text{-GlcNAc}$	$\beta\text{-Gal}(1 \rightarrow 4)\text{-}\beta\text{-GlcNAc}$	$\beta\text{-Gal}(1 \rightarrow 4)\text{-}\beta\text{-GlcNAc}$
3	$\alpha\text{-NeuNAc}(2 \rightarrow 3)\text{-}\beta\text{-Gal}$	$\alpha\text{-NeuNAc}(2 \rightarrow 3)\text{-}\beta\text{-Gal}$	$\alpha\text{-NeuNAc}(2 \rightarrow 3)\text{-}\beta\text{-Gal}$

the dimensionality of the conformational hyperspace explored in the CICADA run was 28. In addition, all the hydroxyl groups were monitored, which means that the orientations were taken into account for detecting new energy minima. The run was considered as complete when no new minima were detected in an energy window of 4 kcal mol⁻¹ above the absolute minimum.

2.6. Force field

Geometry optimization was performed using the molecular mechanics program MM3. This force field contains a correction for the anomeric effect and has been shown to be especially adapted for the study of carbohydrates.¹⁹ The block diagonal minimization method with the default energy convergence criterion of 0.00008*n kcal mol⁻¹ per five iterations, *n* being the number of atoms, was used for grid point optimization. To mimic the hydrated environment of the molecules, a dielectric constant, $\epsilon = 78.5$, was used in all the calculations. All calculations were performed on an SGI Origin 300 (R14000, 600 MHz processor).

2.7. Determination of families of low energy conformers

The raw data from the CICADA calculation were analyzed with the program PANIC,²¹ which extracts the low energy conformations along with the transition states and analyzes the network of interconversions between them. Then the population of the low energy conformations has been analyzed by an ‘in house’ program in order to cluster the conformations into families.²² In this approach, inside a given energy window, a conformer is considered to belong to a family if all its torsion angles differ by no more than a given value from the torsion angles of at least one of the members of the family. In the present study, the value considered for determining the difference has been set to 30°. This means that two conformations are classified into different families if there is at least one glycosidic torsion angle, which differs by more than 30°. The coordinates of the energy minima will be deposited or made available for users on the internet through the CERMAV website.

3. Results and discussion

The exploration of the 28 dimensions of the conformational space throughout CICADA took slightly less than three months of calculation on an R14000 600 MHz processor. From this massive amount of data, 11,471 points in the potential energy hyperspace were stored, of which 2637 were low energy conformations and 5691 were transition states. One should not be surprised at the time employed by the program to perform this calculation, as a classical grid search method

on such a complex molecule would simply not have been feasible. For instance, a 20° step scan on all the 28 glycosidic bonds would require something like 18²⁸ energy minimizations! Even taking into account only the combination of the lowest energy conformations shown by each disaccharide unit, say three each one, we can estimate the number of conformations as 3¹⁴ that equals to 4,782,969.

3.1. Analysis of linkage conformations

Because of the large amount of data produced by the calculations, together with the intrinsic very high multidimensionality of this conformational space, the analysis of the results is not an easy task to accomplish. A general way to proceed is to first take into account the structural information related to the conformations taken at each glycosidic linkage. It is possible to superimpose the conformations found by CICADA on the energy maps of the constituting disaccharides.

The energy map calculated for the α -Fuc(1→3)- β -GlcNAc disaccharide (Fig. 2) displays a main low energy region centered at $\Phi = 270$, $\Psi = 70$ and a second one, slightly higher in energy (+1 kcal mol⁻¹) at $\Phi = 280$, $\Psi = 150$. When looking at conformations found by CICADA, the major part falls over the zone centered at $\Phi = 280$, $\Psi = 150$, while the other minimum, which is the lowest one for the disaccharide, is slightly populated only in the case of the linkage C1.

The same reduction in flexibility takes place at the β -Gal(1→4)- β -GlcNAc linkages. While the map for the disaccharide exhibits three low energy zones (Fig. 3), most of the conformations found for this glycosidic linkage in the oligosaccharide falls around the upper edge of the main energy well, indicating a complete reduction in flexibility for the A2 and B2 linkages. Only the linkage C2 appears to be more flexible, with some conformations occurring within the third low energy zone centered at $\Phi = 280$, $\Psi = 60$.

In the case of an ‘isolated’ Le^x group, these reductions in flexibility are quite a general feature of the molecule. From many NMR and theoretical investigations,^{15,23–25} it has been proposed that important nonbonded interactions occur between the galactose and fucose residues, thereby reducing the conformational flexibility of both linkages. These features have been confirmed by the crystal structure elucidation of this oligosaccharide.²⁶

The orientations of the α -NeuNAc(2→3)- β -Gal linkages in the conformers found by CICADA correspond closely to the low energy zones of the disaccharide map. It can be clearly seen that the conformational flexibility arises mainly from the rotation around the Φ angle. The α -NeuNAc(2→3)- β -Gal is a flexible glycosidic linkage that is not perturbed from the effects shown before. Nevertheless, also in this case, some differences can be observed between the three linkages, as indicated

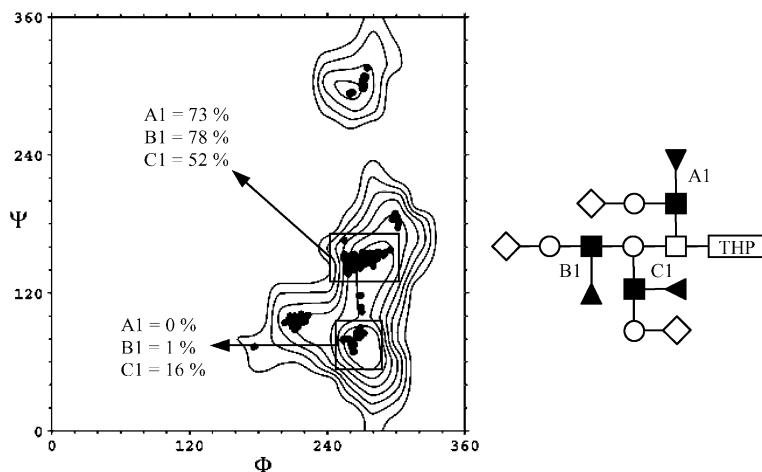


Figure 2. The map for the α -Fuc(1 → 3)- β -GlcNAc disaccharide, with the superimposed energy minima found by CICADA for the A1, B1, and C1 linkages (the black ones on the molecule representation on the right). The isocontours, with 1 kcal mol^{-1} , represent the energy levels of the disaccharide calculated using the grid search method. The map is plotted with the program Xfarbe.²⁰ Only the conformations within an energy window of 10 kcal mol^{-1} are shown. The areas with consistent population have been framed and labeled. The zone centered on $\Phi = 280$, $\Psi = 150$ is the most populated region. The lowest minimum of the disaccharide is populated only in the case of the linkage C1.

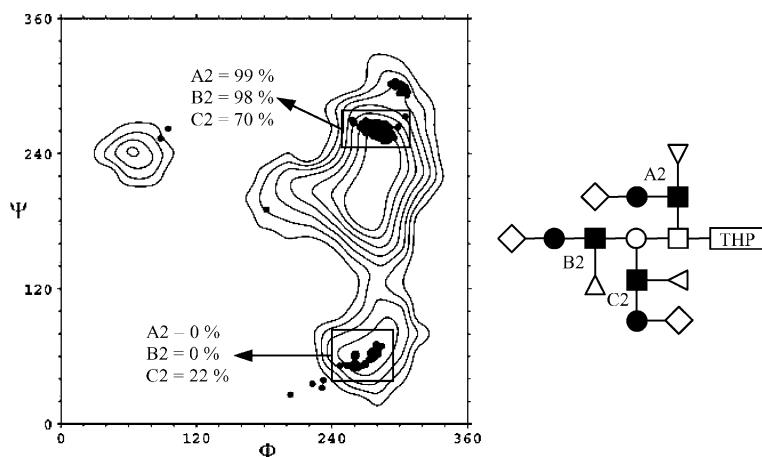


Figure 3. The map for the β -Gal(1 → 4)- β -GlcNAc disaccharide with the superimposed energy minima found by CICADA for the A2, B2, and C2 linkages. The zone centered on $\Phi = 270$, $\Psi = 200$ shows a complete lack of conformer population. Only the linkage C2 shows conformers in the other low energy regions.

by the relative populations of the lowest energy region centered on $\Phi = 60$, $\Psi = 100$ (Fig. 4).

The effects of the reduction in conformational flexibility are well illustrated below in the case of the linkages A and B, corresponding to the β -GlcNAc(1 → 6)- β -GalNAc and β -GlcNAc(1 → 6)- β -Gal disaccharides. The maps for the disaccharides A and B display the same features, both in positions and energy levels. However, the conformations found for these linkages in the oligosaccharide are totally different. In the case of β -GlcNAc(1 → 6)- β -GalNAc (A linkage) the lowest energy region is well populated (Fig. 5a) whereas for β -GlcNAc(1 → 6)- β -Gal (B linkage) the great majority of the low energy conformers are located around the minimum at $\Phi = 270$, $\Psi = 60$.

The ω torsion angles of the (1 → 6) linkages are known to have three possible orientations, *gauche-trans* (*gt*), *trans-gauche* (*tg*), and *gauche-gauche* (*gg*). As for the β -GlcNAc(1 → 6)- β -GalNAc (A linkage), the following percentages of occurrence were calculated: *gt* = 72%, *tg* = 12%, and *gg* = 16%. The *gt* conformation is clearly the preferred one, while the other orientations start to have a consistent percentage only at higher energies (+6 kcal mol^{-1} over the global minima). The β -GlcNAc(1 → 6)- β -Gal (B linkage) is more rigid having shown a *gt* orientation that occurs for more than 98% of the total conformations and neglectable percentages of *tg* and *gg* orientations.

The disaccharide maps for the C and core linkages, which are buried inside the structure of the

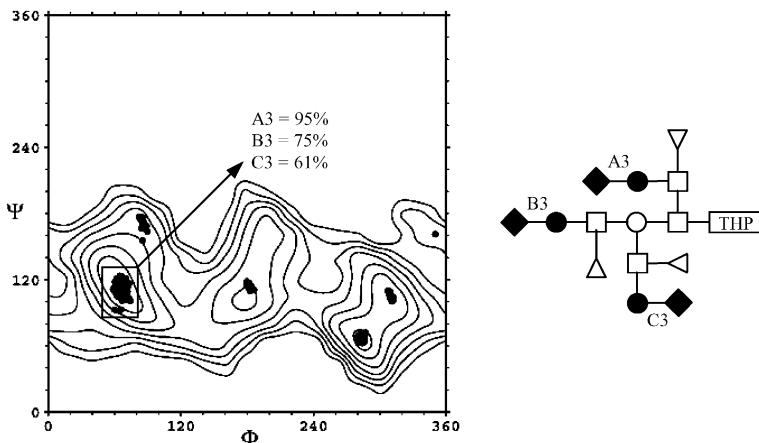


Figure 4. The map for the α -NeuNAc(2 → 3)- β -Gal disaccharide with the superimposed conformations found by CICADA for the A3, B3, and C3 linkages.

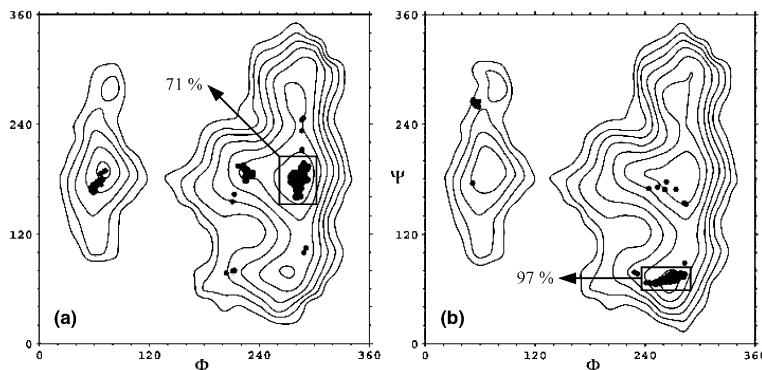


Figure 5. (a) and (b) From left to right, respectively, the maps for the disaccharides A and B, corresponding to the β -GlcNAc(1 → 6)- β -GalNAc and β -GlcNAc(1 → 6)- β -Gal linkages with the superimposed conformations found by CICADA. While the maps look rather the same, the conformations are distributed differently.

oligosaccharide, exhibit the same aspect, with a deep low energy well and two secondary regions. The conformations found for the oligosaccharide are all located in the lowest energy region whereas the secondary minima are not populated at all (Fig. 6).

Moreover, when the energy of the conformers is plotted as a function of the glycosidic torsion angle Ψ (Fig. 7), the occurrence of two distinct families with comparable percentages of occurrence for the β -Gal(1 → 3)- β -GalNAc linkage is found.

The results that have been obtained for the linkages type 1, 2, and 3 groups can be compared with those obtained for an isolated S-Le^x molecule.²⁷ There are no major differences on the overall distribution of the conformations, which means that when linked to such a complex oligosaccharide, the S-Le^x moiety can be considered as a side group whose overall conformational behavior remains unchanged. However, we can see that the distribution of the conformations for the linkages in the C type S-Le^x matches closely the distributions

calculated by CICADA for the isolated S-Le^x molecule. On the other hand, the reductions in flexibility exhibited by the linkages in A and B S-Le^x groups are more significant. This is of course the consequence of the non-equivalency (both in terms of position and neighboring groups) of the three S-Le^x moieties; in other words they ‘feel’ their environments in different ways. For all the linkages, we have shown how the nonbonded interactions that take place between adjacent residues can influence the conformational flexibility of these linkages. These results finally agree with the current idea that any occurrence of interactions between different residues can only result in a reduction of the available conformational space found for the disaccharides.^{28,29}

3.2. Families of conformations and energy minima

To understand how nonbonded interactions can influence the overall shape of Core2(S-Le^x)₃ we first have to cluster all the conformations into families. In Table 2

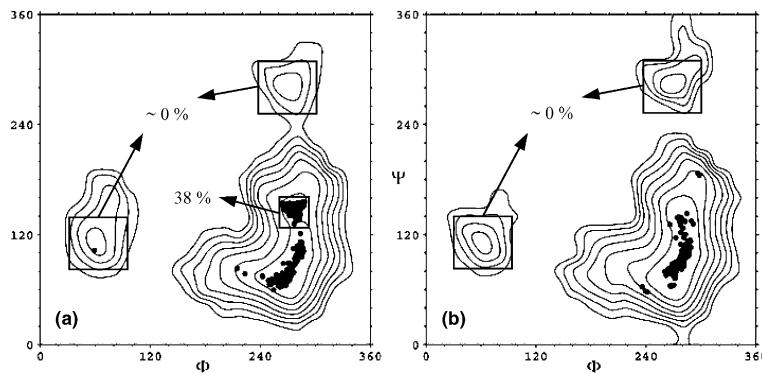


Figure 6. (a) and (b) From left to right, respectively, the maps for the disaccharides C and core, corresponding to the β -GlcNAc(1→3)- β -Gal and β -Gal(1→3)- α -GalNAc linkages, with the superimposed conformations found by CICADA.

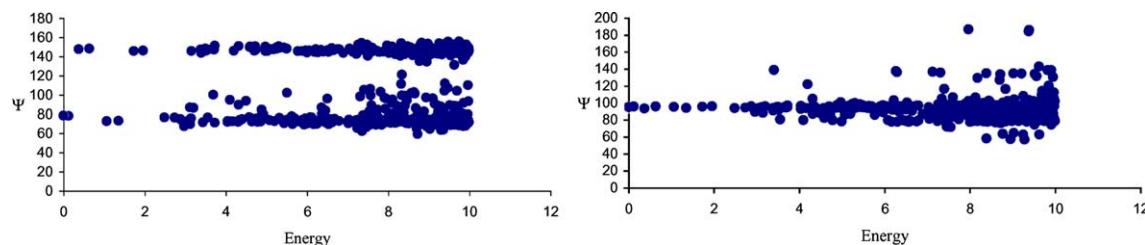


Figure 7. The energy of the conformations is plotted against the Ψ angle for the β -GlcNAc(1→3)- β -Gal linkage (on the left) and for the β -Gal(1→3)- α -GalNAc linkage (on the right). It is clear that for the β -GlcNAc(1→3)- β -Gal there are two main conformational families while the β -Gal(1→3)- α -GalNAc linkage only exhibits oscillations around its average value.

Table 2. The 15 best families found in a window of 10 kcal mol⁻¹ over the global minimum

Family	Energy of the best conformer (kcal mol ⁻¹)	%	Glycosidic torsion angles, which differ by more than 30° from those of the global minimum ($\pm \Delta$ °)	
1	0.00	56.7		Global minimum
2	+0.37	27.6	C Ψ (+69.3)	
3	+1.06	8.4		C3 Ψ (+57.4)
4	+1.72	2.8	C Ψ (+67.4)	C3 Ψ (+57.7)
5	+2.48	0.5		B3 Ψ (+63.9)
6	+2.73	<0.5	C1 Φ (-69.5)	C1 Ψ (-55.9)
7	+2.87	<0.5		C3 Φ (+244.3)
8	+2.96	<0.5	C1 Ψ (-64.0)	C2 Ψ (-200.2)
9	+2.97	<0.5	C1 Ψ (+142.7)	C2 Ψ (-199.0)
10	+3.08	<0.5	C3 Φ (+215.9)	C3 Ψ (-54.1)
11	+3.11	<0.5	B1 Φ (-58.0)	B1 Ψ (-49.8)
12	+3.14	<0.5	C Ψ (+67.4)	B3 Ψ (+63.6)
13	+3.19	<0.5	A1 Φ (-68.9)	A1 Ψ (-56.4)
14	+3.37	<0.5	A1 Φ (-69.2)	A1 Ψ (-56.3)
15	+3.40	<0.5	C Ψ (+65.9)	C Φ (+220.3)
				C Ψ (+69.1)
				B3 Ψ (-46.1)

one can see that despite the great number of glycosidic torsion angles in the molecule, there are only a few families with a significant percentage of occurrence. Indeed, the global distribution of the conformations indicates the occurrence of only a few low energy conformers situated in deep and narrow energy wells. On the other hand, there exist families of conformations, with higher energy, that fill up all the conformational

space. Such a feature is an indication of the occurrence of transition states that the oligosaccharide may follow to go from one low energy conformation to another one.

A careful analysis of the features found in the first 10 families indicates that major variations of glycosidic linkages take place at the C part of the molecule, (i.e., where the S-Le^x is linked 1→3 to the core). This qualitative observation agrees with the results gathered

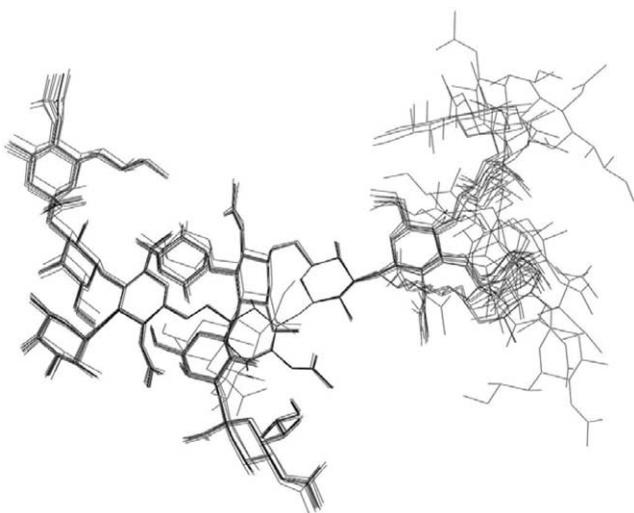


Figure 8. Superimposition of the best conformers from the first 10 families. The structures are superimposed on the core glycosidic linkage and the hydrogen atoms have been removed for clarity. The right part of the molecule is the C S-Le^x. At the extreme left, the A S-Le^x and on the center, the core disaccharide, which is partially covered by the B S-Le^x.

previously on the disaccharide potential energy surfaces. It was noticed that the linkages located at the C part of the oligosaccharide had more conformational flexibility than the linkages in the A and B S-Le^x. This is well illustrated in Figure 8, which shows the superimposition of the best conformers taken from the first 10 families. One can see that the C branch exhibits multiple conformations, particularly for the α -NeuNAc(2 → 3)- β -Gal (C3), the α -Fuc(1 → 3)- β -GlcNAc (C1) and the β -GlcNAc(1 → 3)- β -GalNAc (C) disaccharides. In the rest of the molecule, most of the glycosidic linkages show mainly oscillations around their average values.

When looking at the structure of the molecule, it is clear that the reduction in conformational flexibility of the A and B S-Le^x moieties can only be attributed to the energy stabilization that results from their spatial proximity. Indeed, several hydrogen bonds and stacking interactions take place between these two branches of Core2(S-Le^x)₃. These effects stabilize these two branches of the molecule in their lowest energy conformations, thereby reducing the conformational flexibility of the glycosidic linkages. In contrast, the C S-Le^x group is not involved in any long range interactions with other residues. As a consequence, it is more susceptible to exhibiting higher conformational flexibility.

3.3. Molecular shapes and properties

The whole Core2(S-Le^x)₃ oligosaccharide can be idealized as a backbone, formed by the core disaccharide and the GlcNAc residues of the A, B, and C S-Le^x groups, which is linked to three S-Le^x moieties. Consequently, the overall conformation of the whole molecule can be

changed only by rotations about the glycosidic linkages of the backbone. As shown previously, these linkages are very rigid. Only the β -GlcNAc(1 → 3)- β -GalNAc (C) displays two distinct conformations having comparable percentage of occurrence. Therefore, the most interesting structural feature that is conserved in several families is the shape of the backbone conformation, while the terminal residues are more or less free to explore their conformational spaces. Figure 9 represents the lowest energy conformer of Core2(S-Le^x)₃. The backbone of the molecule is very flat, resulting from four adjacent residues that are in the same plane. The core disaccharide, together with the A and C S-Le^x branches form a curved sheet-like structure, which makes a 180° turn with respect to the plane defined by the backbone of the molecule. The maximum extension of the structure is about 27 Å. The B S-Le^x is placed above the core, forming a third arm that projects from the plane of the molecule. This architecture is highly stabilized by several hydrogen bonds between the A and B S-Le^x residues.

The lowest energy conformers of the first four families are represented in Plate 1, together with their molecular properties, electrostatic, lipophilicity, and hydrogen bonding potential. The second family differs only in the geometry of the β -GlcNAc(1 → 3)- β -GalNAc linkage, showing a twisted backbone conformation. The third family is similar to the first one except for the orientation of the α -NeuNAc(2 → 3)- β -Gal in the C S-Le^x group. The fourth family exhibits both the structural features of the second and the third family.

The Core2(S-Le^x)₃ oligosaccharide should be linked to uromodulin via the anomeric oxygen of the α -GalNAc residue in the core. Unfortunately, the three-dimensional structure of the protein is unknown, as is the shape of the O-glycosylation sites. From the results of the present investigation, one can observe that the anomeric oxygen of the α -GalNAc residue is located in a buried part of the oligosaccharide. As a result, the occurrence of strong interactions between Core2(S-Le^x)₃ and the amino-acid at the surface of the protein can be anticipated. Whether the Core2(S-Le^x)₃ would keep one of its low energy conformations or would induce some conformational changes of the protein in the surrounding of its O-glycosylation sites are still issues to be considered.

With respect to the biological functions of the Core2(S-Le^x)₃ glycan on uromodulin or THP, an important feature has to be noted. If the increased activity of this protein can be related to the presence of these O-glycans at its surface, then the S-Le^x moieties are the most putative candidates for carrying out the molecular interactions that are needed for the activations of the biomolecular processes. For this purpose, the dendrimeric structure of Core2(S-Le^x)₃ seems to be designed ‘ad hoc’ as it provides the maximum capabilities with respect to intermolecular interactions. The

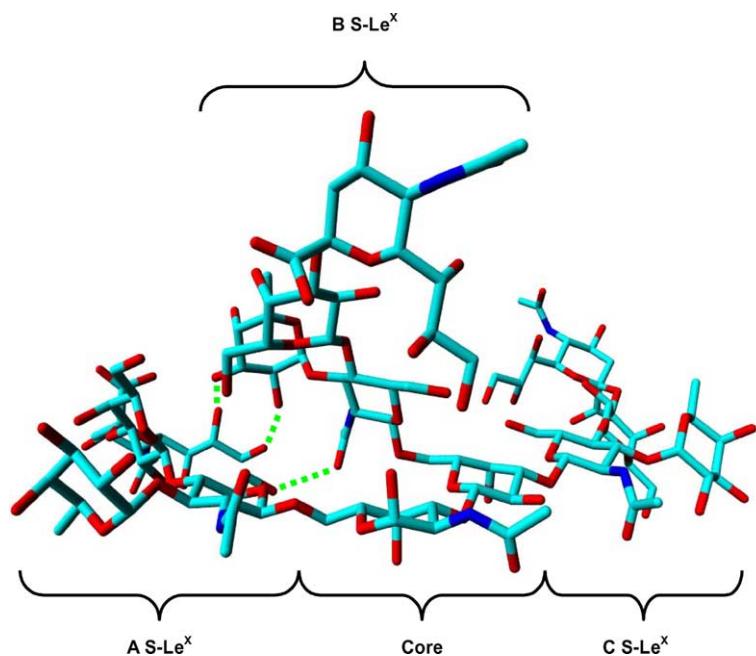


Figure 9. Pictorial three-dimensional representation of the lowest energy conformer for the first family. The hydrogen atoms have been removed and the different parts of the molecule have been framed and labeled. The possible inter-residual hydrogen bond network between A and B S-Le^x groups is suggested with dashed lines.

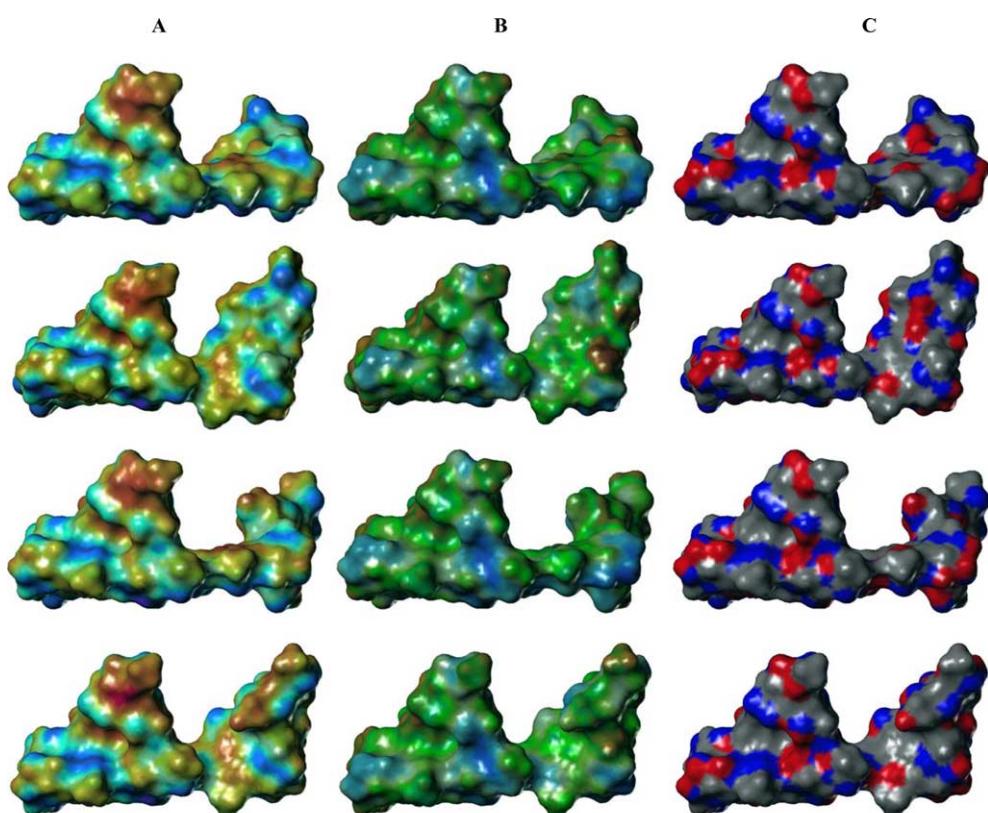


Plate 1. Projection of molecular properties on the Connolly surface for the lowest energy conformer of the four best families (from up to down). (A) Electrostatic potential: colored from blue (negative values) to red (positive values), (B) lipophilicity potential: colored from brown (hydrophobic) to blue (hydrophilic), and (C) hydrogen bonding capacity: blue (acceptor) and red (donor).

conformations of the S-Le^x groups in the lowest energy conformers does not show any differences from those found in previous studies. In particular they are compatible with the binding to E- and P-selectins as derived from crystal structure elucidations.³⁰

4. Conclusions

Along with previous knowledge about the conformations of Le^x and S-Le^x oligosaccharides, the present set of information constitutes the data bank required to characterize the structural features and the dynamics of most of the unusual O-linked glycans that may be related to the enhanced immunomodulatory effects of uromodulin. For each disaccharide constituent, precise information about the low energy domains, along with a description of the features displayed by all the local minima has been provided. The construction of the largest O-linked glycan has been completed; this is a tetradecamer made up of the assembly of three S-Le^x moieties on a type 2 core.

The many possible structures have been explored and submitted to an extensive energy minimization using the molecular mechanics program MM3. This was followed by a complete exploration of conformational hyperspace throughout the use of the single coordinate driving method as implemented in the CICADA program. For the whole oligosaccharide, there exists a substantial reduction of the number of viable conformations. The interactions with neighboring groups in complex oligosaccharides can modulate or alter its conformational behavior. Some generalized pictures have emerged from the present investigation: (i) The occurrence of only four conformational families can account for more than 95% of the total conformations; (ii) The whole glycan can be idealized as a backbone formed by the core disaccharide, and the GlcNAc residue of the three S-Le^x moieties. As such, the overall conformation of the glycan can be changed only by rotations about the glycosidic linkages of this backbone. Three of these linkages are conformationally rigid. It is only the β -GlcNAc(1 → 3)- β -GalNAc portion of the molecules that offers conformational flexibility as it displays two distinct conformations having a comparable percentage of occurrence; (iii) The backbone of the O-glycan is very flat; it is flanked by three branches having terminal sialyl residues. Two of the branches interact through nonbonded interactions that bring further stabilization, and limit the conformational flexibility of the sialyl residues. Only one branch maintains the original conformational feature of S-Le^x epitope; (iv) The Core2(S-Le^x)₃ glycan is a fascinating example of ‘dendrimeric’ structure, with a spatial arrangement of three S-Le^x epitopes that may favor, throughout their complementary ‘presentations’, the interactions with E- and P-selectins.

Acknowledgements

This work was supported in part by the 5th Framework Programme, Research Training Networks under the contract GlycoTrain (HPRN-CT-2000-00001), awarded to G. Cioci. J. Koča was supported by a research associate position from the Centre National de la Recherche Scientifique. The Centre Experimental de Calcul Intensif en Chimie (Grenoble) provided the necessary computational resources.

References

- Tamm, I.; Horsfall, F. L. *Proc. Soc. Exp. Biol. Med.* **1950**, *74*, 108–114.
- Kumar, S.; Muchmore, A. V. *Kidney Int.* **1990**, *37*, 1395–1401.
- Pennica, D.; Kohr, W. J.; Kuang, W.; Glaister, D.; Aggarwal, B. B.; Chen, E. Y.; Goedder, D. V. *Science* **1987**, *236*, 83–88.
- Muchmore, A. V., et al. *Science* **1987**, *237*, 1479–1484.
- van Rooijen, J. J.; Kamerling, J. P.; Vliegenthart, J. F. *Glycoconj. J.* **2001**, *18*, 539–546; Williams, J.; Marshall, R. D.; van Halbeck, H.; Vliegenthart, J. F. *Carbohydr. Res.* **1984**, *134*, 141–155; Afonso, A. M.; Marshall, R. D. *Biochem. Soc. Trans.* **1979**, *7*, 170–173.
- Easton, R. L.; Patankar, M. S.; Clark, G. F.; Morris, H. R.; Dell, A. *J. Biol. Chem.* **2000**, *275*, 21928–21938.
- Mitoma, J.; Petryniak, B.; Hiraoka, N.; Yeh, J.; Lowe, J. B.; Fukuda, M. *J. Biol. Chem.* **2003**, *278*, 9953–9961.
- Imbert, A.; Pérez, S. *Chem. Rev.* **2000**, *100*, 4567–4588.
- Woods, R. J. *Curr. Opin. Struct. Biol.* **1995**, *5*, 591–598.
- Imbert, A. *Curr. Opin. Struct. Biol.* **1997**, *7*, 617–623.
- Allinger, N. L.; Zhu, Z. Q.; Chen, K. J. *J. Am. Chem. Soc.* **1990**, *112*, 6120.
- Koča, J. *Progr. Biophys. Mol. Biol.* **1998**, *70*, 137–173.
- Koča, J. *J. Mol. Struct. (Teochem.)* **1994**, *308*, 13–24.
- Koča, J.; Pérez, S.; Imbert, A. *J. Comput. Chem.* **1995**, *16*, 296–310.
- Imbert, A.; Mikros, E.; Koča, J.; Mollicone, R.; Oriol, R.; Pérez, S. *Glycoconj. J.* **1995**, *12*, 331–349.
- Engelsen, S. B.; Pérez, S.; Braccini, I.; Hervé du Penhoat, C.; Koča, J. *Carbohydr. Res.* **1995**, *276*, 1–29.
- Imbert, A.; Bettler, E.; Karababa, M.; Mazeau, K.; Petrova, P.; Pérez, S. In *Perspectives in Structural Biology*; Vijayan, M., Yathindra, N., Kolashar, A. S., Eds.; Indian Academy of Science (Universities Press), 1999. <http://www.cermav.cnrs.fr/>.
- Engelsen, S. B.; Cros, S.; Mackie, W.; Pérez, S. *Biopolymers* **1996**, *39*, 417–433.
- Pérez, S.; Imbert, A.; Engelsen, S. B.; Gruza, J.; Mazeau, K., et al. *Carbohydr. Res.* **1998**, *314*, 141–155.
- http://www.fhi-berlin.mpg.de/grz/pub/xfarbe_doc.html.
- Koča, J. *J. Mol. Struct.* **1993**, *291*, 255–269.
- Imbert, A.; Pérez, S. *Glycobiology* **1994**, *4*, 351–366.
- Ichikawa, Y.; Lin, Y. C.; Dumas, D. P.; Shen, G. J., et al. *J. Am. Chem. Soc.* **1992**, *114*, 9283–9298.
- Thogersen, H.; Lemieux, R. U.; Bock, K.; Meyer, B. *Can. J. Chem.* **1982**, *60*, 44–57.
- Miller, K. E.; Mukhopadhyay, C.; Casas, P.; Bush, C. A. *Biochemistry* **1992**, *31*, 6703–6709.

26. Pérez, S.; Mouhous-Riou, N.; Nifant'ev, N. E.; Tsvetkov, Y. E.; Bachet, B.; Imberty, A. *Glycobiology* **1996**, *6*, 537–542.
27. Imberty, A.; Pérez, S. In *Carbohydrates Mimics. Concept and Methods*; Chapleur, Y., Ed.; Wiley: Weinheim, 1997; pp 349–364.
28. Imberty, A.; Gerber, S.; Tran, V.; Pérez, S. *Glycoconj. J.* **1990**, *7*, 37–54.
29. Imberty, A.; Delage, M.; Bourne, Y.; Cambillau, C.; Pérez, S. *Glycoconj. J.* **1991**, *8*, 456–483.
30. Somers, W. S.; Tang, J.; Shaw, G. D.; Camphausen, R. T. *Cell* **2000**, *103*, 467–479.

Chapitre 6

La lectine bactérienne PA-IL de *Pseudomonas aeruginosa*

6.1 Introduction

Dans le deuxième chapitre on a introduit l'importance des interactions protéine-sucre pendant les processus d'infections. Ensuite on a vu quelles sont les connaissances sur la bactérie opportuniste *Pseudomonas aeruginosa*, et plus particulièrement sur ses deux lectines solubles.

La PA-IL fut la première lectine à être identifiée dans une culture de *P. aeruginosa* même avant le séquençage complet de son génome (Gilboa-Garber 1972). La spécificité est limitée aux molécules avec un galactose en position terminale, le monosaccharide est fixé avec une affinité moyenne ($K_a = 3.4 \cdot 10^4 \text{ M}^{-1}$) et le GalNAc est aussi reconnu, bien qu'avec une affinité moins importante. L'affinité est plus forte pour les glycoconjugués se terminant par un disaccharide Gal α 1→3Gal ou Gal α 1→4Gal sur l'extrémité non réductrice (Lanne 1994). Les déterminants des groupes sanguins I,B et Pk (qui contiennent le disaccharide Gal α 1→4Gal) sont donc reconnus par la PA-IL, ce qui fait de cette lectine un outil intéressant dans les études menées sur la xénotrasplantation des organes de porc dans les singes ou dans l'homme. Ces oligosaccharides sont en fait les xéno-antigènes responsables du phénomène de rejet hyperaiguë (Kirkeby 2002). Comme on a pu le voir dans le deuxième chapitre, le rôle de la PA-IL dans *P. aeruginosa* ainsi que ces cibles d'adhésion n'ont pas encore été élucidées avec précision mais il a été montré que la lectine est cytotoxique pour les cellules épithéliales des voies respiratoires. La structure cristalline de la PA-IL, résolue en 2002, a mis en évidence une structure quaternaire tétramérique (Liu 2002). L'élucidation des bases moléculaires de la reconnaissance du calcium et du galactose, décrite ici, ajoute un chapitre importante dans l'étude des molécules impliquées dans l'infection de *Pseudomonas aeruginosa*.

6.2 Résultats

PA-IL a été purifiée par chromatographie d'affinité à partir de cultures bactériennes et co-cristallisée en présence de calcium et de galactose. La structure tridimensionnelle, résolue par remplacement moléculaire, montre un arrangement tétramérique où chaque monomère adopte un repliement du type β -sandwich avec le site de liaison qui pointe vers l'extérieur de la molécule. Tous les groupements hydroxyles du sucre, à l'exception de O1, sont en interaction avec la protéine et plus particulièrement O3 et O4 participent aussi à la coordination du calcium. Le mode de reconnaissance qu'on a ici mis en évidence montre des similarités avec

deux lectines animales du type C, bien que le repliement général de la protéine soit différent. Ce travail complète la caractérisation structurale des lectines solubles de *P. aeruginosa*. Dans la PA-IIL, les deux ions calcium sont responsables de la forte affinité et des ligands différents peuvent être reconnus. Au contraire, dans la PA-IL le calcium en combinaison avec un réseau de liaisons hydrogène forme un site de liaison très spécifique pour le galactose qui fait de cette lectine une cible intéressante pour un développement futur d'une nouvelle classe de molécules anti-adhésives.

Structural basis of calcium and galactose recognition by the lectin PA-IL of *Pseudomonas aeruginosa*

Gianluca Cioci^a, Edward P. Mitchell^b, Catherine Gautier^a, Michaela Wimmerová^c, Dvora Sudakevitz^d, Serge Pérez^a, Nechama Gilboa-Garber^d, Anne Imbert^{a,*}

^aCERMAV-CNRS (affiliated with Université Joseph Fourier), P.O. Box 53, F-38041 Grenoble Cedex 09, France

^bE.S.R.F. Experiments Division, P.O. Box 220, F-38043 Grenoble Cedex, France

^cNational Centre for Biomolecular Research and Department of Biochemistry, Masaryk University, Kotlarska 2, 611 37 Brno, Czech Republic

^dBar-Ilan University, Faculty of Life Sciences, Ramat Gan 52900, Israel

Received 23 September 2003; revised 25 September 2003; accepted 21 October 2003

First published online 12 November 2003

Edited by Irmgard Sinning

Abstract The structure of the tetrameric *Pseudomonas aeruginosa* lectin I (PA-IL) in complex with galactose and calcium was determined at 1.6 Å resolution, and the native protein was solved at 2.4 Å resolution. Each monomer adopts a β-sandwich fold with ligand binding site at the apex. All galactose hydroxyl groups, except O1, are involved in a hydrogen bond network with the protein and O3 and O4 also participate in the coordination of the calcium ion. The stereochemistry of calcium galactose binding is reminiscent of that observed in some animal C-type lectins. The structure of the complex provides a framework for future design of anti-bacterial compounds.

© 2003 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

Key words: Lectin; Crystal structure; Galactose; *Pseudomonas aeruginosa*

1. 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 colonise patients with a number of chronic lung diseases, particularly those on assisted ventilation, and especially cystic fibrosis patients. *P. aeruginosa* virulence is based on several properties, most particularly on the ability of this organism to adhere to surfaces, to form biofilms and to secrete hydrolytic enzymes and toxic compounds.

Since host carbohydrates have been known for many years to constitute specific attachment sites for microbial protein receptors [1,2], the lung mucins, and above all those of cystic fibrosis patients, have been analysed thoroughly [3]. On the other hand, some of the carbohydrate binding proteins of *P. aeruginosa* have been studied and their role in recognition and adhesion is far from being fully elucidated. Several types of receptors have been identified. Flagellin and flagellar cap protein FliD recognise mucin oligosaccharides [4], whereas

type IV pilus adhesins have been shown to recognise glycosphingolipids asialo-GM1 and asialo-GM2 [5]. In addition, two soluble lectins PA-IL (gene *lecA*) and PA-IIL (gene *lecB*), specific for D-galactose and L-fucose, respectively, and containing divalent cations have been characterised [6]. These two lectins are produced at high levels by the bacteria in association with the cytotoxic virulence factors [7] and under quorum-sensing control [8]. The galactophilic lectin PA-IL was the first *P. aeruginosa* lectin to be isolated by affinity chromatography [9]. Its 12.75 kDa subunits consist of 121 amino acids and associate as a tetramer [10]. This lectin has a narrow specificity spectrum for D-galactose-containing molecules, including analogues bearing hydrophobic groups [11] and α-D-galactose-containing disaccharides and other glycoconjugates [12,13], but it also binds adenine with high affinity [14]. The lectin has been demonstrated to bind the most common human antigens, I, B and especially Pk [15], and to be cytotoxic to respiratory epithelial cells in primary cultures [16].

We recently solved the structure of PA-IIL, the fucose binding lectin of *P. aeruginosa*, at very high resolution (1.3 Å) [17], revealing a new mode for carbohydrate binding, based on the interaction with two calcium ions. This was proposed to be the basis for the unusually high affinity of PA-IIL for its ligands. The structural basis of PA-IIL's wide specificity spectrum and its interaction modes with different monosaccharides, such as D-mannose and D-fructopyranose, has been recently elucidated [18]. The successful crystallisation of PA-IL has also been reported recently [19]. The crystal structure in the presence of calcium has been solved at 1.5 Å resolution and deposited in the Protein Data Bank (code 1L7L) although not accompanied by a detailed description [20]. The present communication describes the refined crystal structure of the calcium-free form of PA-IL and its complex with calcium and galactose. This structure unveils a new carbohydrate binding mode and may assist in the development of therapeutic drugs against *Pseudomonas* infections.

2. Materials and methods

PA-IL was isolated and purified by affinity chromatography as described previously [6]. Lyophilised protein was dissolved in water (10 mg/ml) either alone or in the presence of galactose (425 µg/ml) and salts (1 mM CaCl₂ and MgCl₂). Initial crystallisation conditions were screened using the complete Hampton Screens I and II from Hampton Research (Laguna Niguel, CA, USA) from which thin, needle-shaped

*Corresponding author. Fax: (33)-4-76 54 72 03.

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

Abbreviations: PA-IL, *Pseudomonas aeruginosa* lectin I; PA-IIL, *Pseudomonas aeruginosa* lectin II; MBPA, mannose binding protein A

crystals appeared after a few days in the presence of ammonium sulphate and isopropanol. These conditions were then optimised and large needle-shaped crystals were obtained from hanging drops of 2 µl of PA-IL solution mixed with 2 µl of reservoir solution (1.5 M ammonium sulphate at pH 4.5, 20% isopropanol for the calcium-free protein and 1.5 M ammonium sulphate at pH 4.7, 5% MPD and 2% glycerol for the complex with galactose). Both crystals belong to space group P2₁2₁2₁ with four molecules per asymmetric unit. The cell dimensions of the calcium-free protein crystal are $a = 49.7 \text{ \AA}$, $b = 51.5 \text{ \AA}$ and $c = 166.8 \text{ \AA}$ and those of the complex are $a = 49.1 \text{ \AA}$, $b = 53.3 \text{ \AA}$ and $c = 160.7 \text{ \AA}$. Crystals were cryo-cooled at 100 K after soaking them for as short a time as possible in either glycerol 30% v/v or MPD 30% v/v in precipitant solution for the calcium-free and complexed form respectively. All data were collected at the ESRF synchrotron (Grenoble, France) at stations ID14-1 and ID14-4 on an ADSC Q4R CCD detector (Quantum) and at station BM30A on a MAR165 CCD detector (Mar Research). Diffraction images were processed using MOSFLM [21] and scaled and converted to structure factors using the CCP4 program suite [22].

The structure of the calcium-free protein was solved by the molecular replacement technique with the MOLREP program [23], using the monomeric structure (PDB code 1L7L) [24] of the calcium-containing PA-IL [20] with calcium and water molecules removed as the search probe. The molecular replacement gave four clear solutions corresponding to the expected tetramer in the P2₁2₁2₁ asymmetric cell. The structure of the complex with galactose was then solved with the same program using one monomer of the refined calcium-free structure, again with calcium and water molecules removed. To retain maximum non-bias, the R_{free} data set of the calcium-free structure was used, extending it to 1.6 Å resolution. After an initial cycle of refinement, the electron density maps of this complex showed clear features corresponding to four galactose molecules and four calcium ions. For the purposes of identification, the ions were modelled as both calcium and magnesium in separate refinements and the B -factors compared after refinement. The B -factors of the magnesium ions were significantly less than those of the surrounding ligating atoms (average of 3.3 Å² against 8.6 Å²) whilst those for calcium were much closer (6.6 Å²), suggesting that the ions were most likely to be calcium. For both structures, automatic placement of water molecules was performed using the ARP/warp program [25]. Crystallographic refinement was carried out with the program REFMAC [26], with an overall B -factor refinement for the lower resolution structure and individual B -factors for the higher resolution structure, and manual model building with O [27]. Co-ordinates have been deposited at the Protein Data Bank [24] under codes 1OKP and 1OKO for the calcium-free and complexed structures, respectively.

3. Results and discussion

The final refined model of calcium-free PA-IL consists of 484 residues, 137 water molecules and two sulphate ions with an R_{crys} of 18.3% and R_{free} of 26.3% to 2.4 Å resolution. The PA-IL/galactose complex contains 484 amino acids, 482 water molecules, three sulphate ions, one MPD molecule, four calcium ions and four galactose residues with an R_{crys} of 15.4% and R_{free} of 18.7% to 1.6 Å resolution (Table 1). In both structures, the asymmetric unit contains four polypeptide chains of 121 amino acids arranged around a pseudo C222 axis (Fig. 1A). The resulting tetramer is very similar to that generated by the I222 space group in the crystal structure of native PA-IL [20]. Each monomer adopts a small jelly-roll type β-sandwich fold, consisting of two curved sheets, each one consisting of four antiparallel β-strands. Tetramerisation occurs by interaction between the largest sheets for one interface and by contacts between C-terminal moieties for the other interface.

The structure obtained from crystals grown in the presence of metal ions and galactose shows each monomer to contain one calcium ion and one galactose ligand in the same binding site (Fig. 1B). There are no major differences between the

Table 1
Data collection and refinement statistics

Crystal	Metal-free	Complex with Ca and galactose
Data collection		
Beamlne	ID14-4	ID14-1
Wavelength (Å)	0.932	0.934
Resolution (Å)	2.4	1.6
Highest resolution shell (Å)	2.40–2.48	1.60–1.66
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions		
a (Å)	49.7	49.1
b (Å)	51.5	53.3
c (Å)	166.8	160.7
α (degrees)	90	90
β (degrees)	90	90
γ (degrees)	90	90
Measured reflections	114 114	241 596
Unique reflections	17 451	53 762
Average multiplicity	6.5 (6.2)	4.5 (4.2)
Completeness (%)	99.8 (99.8)	94.9 (79.1)
Average $I/\sigma(I)$	7.7 (1.9)	6.7 (1.4)
R_{merge} (%)	8.9 (36.0)	7.6 (41.0)
Wilson B -factor (Å ²)	34.0	12.5
Refinement		
Resolution range (Å)	20.37–2.40	18.60–1.60
R_{crys} (observation)	0.183 (16569)	0.154 (51032)
R_{free} (observation)	0.263 (882)	0.187 (2730)
Highest resolution shell		
R_{crys} (observation)	0.200 (1211)	0.197 (3066)
R_{free} (observation)	0.270 (75)	0.256 (144)
Cruickshank's DPI based on maximum likelihood (Å)	0.205	0.062
Average B_{iso} (Å)		
All atoms	30.2	14.0
Protein atoms	30.3	12.5
Solvent atoms	29.2	24.5
RMS deviation from ideality		
Bonds (Å)	0.016	0.017
Angles (degrees)	1.52	1.57
Outliers on Ramachandran plot	0	0
Amino acids	4 × 121	4 × 121
Protein atoms	3604	3604
Sugar atoms	0	4 × 12
Calcium atoms	0	4
Other atoms	2 × 5	3 × 5+8
Water molecules	137	482
Residues with alternative conformations	None	S113 _A , Q120 _A , S121 _A , T39 _B , N71 _B , T95 _B , S113 _B , V15 _C , V85 _C , S113 _C , S121 _C
PDB deposition code	1UOJ	1OKO

Values in parentheses refer to the highest resolution shell.

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

Geometric analyses were performed by PROCHECK [37] and WHATIF [38].

calcium-free structure, the calcium-bound one [20] and the complex with calcium and galactose and comparison between the monomer main chain atoms from the different structures yields rms difference values lower than 0.55 Å.

The ligand binding site is located at one apex of the β-sandwich and is directed towards the solvent. The calcium binding site is made up mainly of the region 100–108 of the amino acid sequence which forms a loop and a short one-turn α-helix. The side chains of Asp100, Asn107 and Asn108 participate in the co-ordination of calcium together with the main chain carbonyl of Thr104 (Table 2). A fifth contact is established by the carbonyl main chain group of Tyr36 which is located on a neighbouring loop. When this calcium-bound site is compared to the calcium-free one, in three out of four

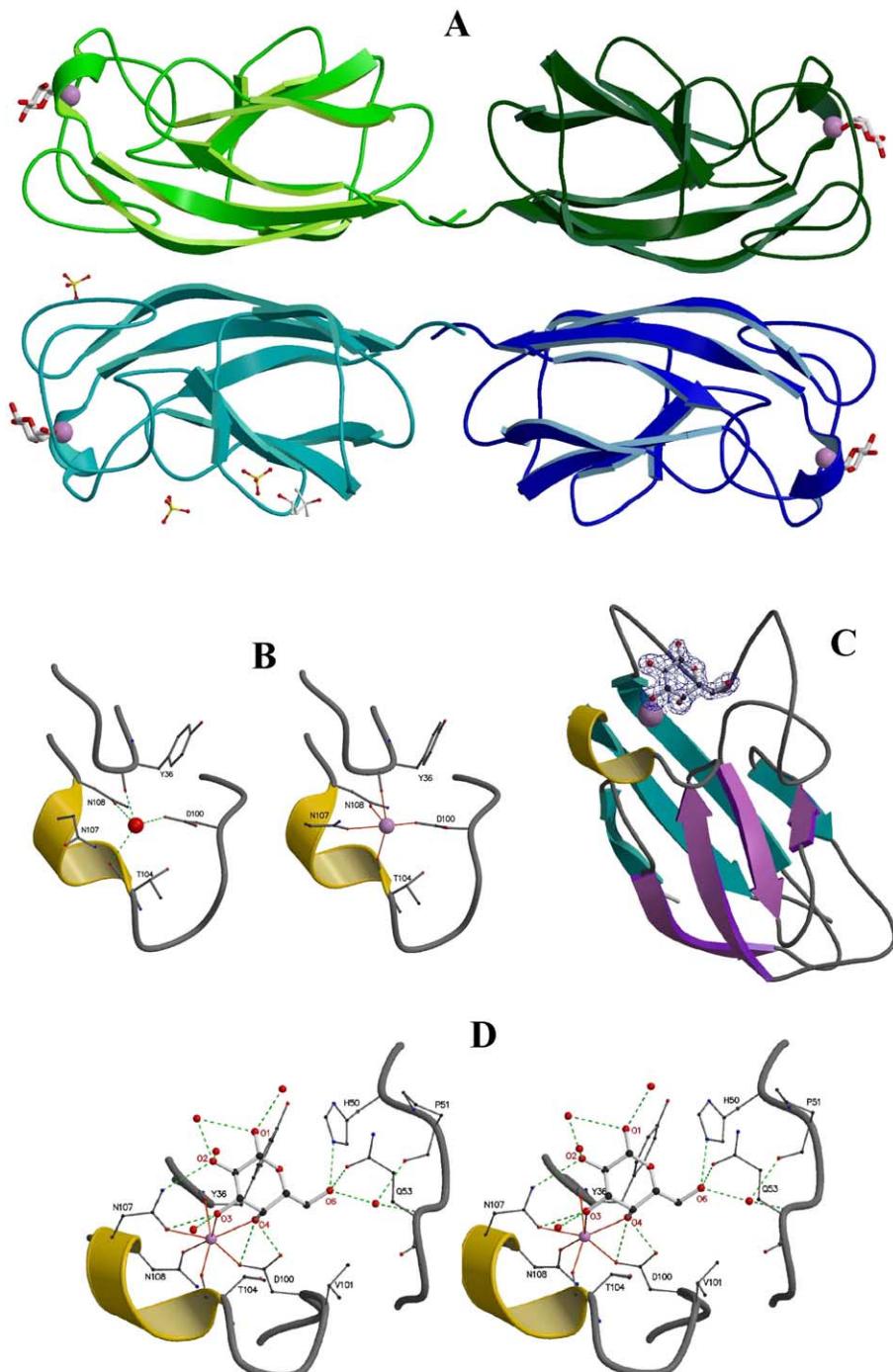


Fig. 1. A: Tetrameric structure of the PA-IL/galactose complex with stick representation of galactose and cpk representation of the calcium ions. Sulphate and MPD molecules are also represented. B: Comparison of the calcium binding site in the calcium-free (left) and in the metal-bound (right) (PDB code 1L7L) structures. Hydrogen bonds are represented as dashed green lines and co-ordination contacts as solid orange lines. Calcium ions and water oxygens are represented by violet and red spheres, respectively. C: Representation of monomer A of the PA-IL/galactose complex showing one sheet in blue, one sheet in pink and the helix in yellow, together with the final sigma weighted $2F_o - F_c$ electron density map (contoured at 1.0σ) around the galactose residue. D: Stereo view of the calcium and galactose binding sites in the crystal structure. Color coding as in B. Molecular drawings in all figures were prepared with MOLSCRIPT [35] and RASTER-3D [36].

monomers of the calcium-free crystal structures a water molecule replaces the calcium ion (Fig. 1B). The oxygen atom of this water is within hydrogen bonding distance of four of the calcium ligands. The only significant difference appears in the Asn107 side chain which is not involved in the hydrogen bond network and adopts a different orientation.

In the structure obtained by co-crystallising PA-IL and ga-

lactose in the presence of calcium and magnesium, electron density is observed at the metal binding site. It has been attributed to calcium, based on the values of *B*-factors, and also on the fact that calcium was observed at high resolution [20]. Electron density corresponding to monosaccharides is clearly identifiable in contact with each calcium ion (Fig. 1C). The galactose ligand adopts the classical 4C_1 pyranose

Table 2
Characteristics of the calcium and galactose binding sites

Atom 1	Atom 2	Distance (Å) ^a
Co-ordination of calcium ion		
Ca	Tyr36.O	2.37 (9)
Ca	Asp100.OD2	2.46 (6)
Ca	Thr104.O	2.29 (3)
Ca	Asn107.OD1	2.45 (4)
Ca	Asn108.OD1	2.41 (8)
Gal.O3	Ca	2.47 (3)
Gal.O4	Ca	2.50 (2)
Hydrogen bonds between PA-IL and Gal		
Gal.O2	Asn107.ND2	3.07 (6)
Gal.O3	Asn107.OD1	2.98 (4)
Gal.O4	Asp100.OD1	2.60 (3)
Gal.O4	Asp100.OD2	2.97 (1)
Gal.O6	His50.NE2	2.76 (7)
Gal.O6	Gln53.OE1	2.70 (6)
Hydrogen bonds with conserved water molecules		
Gal.O1	Wat1	3.1 (1)
Gal.O2	Wat2	2.7 (1)
Gal.O2	Wat2	2.8 (3)
Gal.O3	Wat3	2.7 (1)
Gal.O6	Wat4	2.79 (3)
Bridging water molecules		
Wat4	Pro51.O	2.73 (7)
Wat4	Gln53.N	2.95 (2)
Hydrophobic contacts		
Gal.C1	Tyr36.CD	4.01 (7)
Gal.C2	Tyr36.CD	4.13 (2)
Gal.C4	Thr104.CB	3.71 (3)
Gal.C6	Val101.CG2	3.78 (3)

^aAverage value of the four monomers with standard deviation of the last digit in parentheses.

ring shape. The β configuration is the only one to be observed in monomer A whereas electron density indicates a significant (c. 40%) proportion of α configuration for galactose in the three other monomers. Interactions between galactose and protein include direct hydrogen bonds and hydrophobic contacts but also interactions bridged by either the calcium ion or one water molecule (Table 2). Two oxygen atoms of galactose, i.e. O3 and O4, are involved directly in the co-ordination sphere of the calcium ion (Fig. 1D). Hydrogen bonds to the sugar residues involve three regions of the protein. From the calcium binding loop (100–108), Asp100 establishes contact

with O4 of galactose and Asn107 forms hydrogen bonds with O2 and O3, while Val101 and Thr104 are involved in hydrophobic contacts. Tyr36, also participating in co-ordination of the calcium ion through its carbonyl backbone oxygen, makes hydrophobic contact with C2 of galactose. Another part of the same long loop (36–64) interacts specifically with the O6 hydroxymethyl group that establishes hydrogen bonds with His50, Gln53 and one water molecule making a bridge to the protein backbone. Several other water molecules are in contact with hydroxyl groups O2, O3 and O4. Altogether, four water molecules are conserved in the four binding sites of the asymmetric unit as listed in Table 2.

The PA-IL sequence does not present any significant similarity to any other protein. The monomer fold is related structurally to a β -sandwich family fold described as galactose binding domain-like in the SCOP database [28], although PA-IL represents a truncated form of this fold lacking one of the N-terminal strands. This family contains, among others, several lectins and carbohydrate binding domains, none sharing similarity with those of PA-IL. This fact is not surprising, since it has already been shown that even carbohydrate binding modules from common ancestors may display different locations of their sugar binding sites [29].

Calcium-mediated recognition of galactose by lectins is well characterised for several members of the C-type lectin family. The carbohydrate binding domain of these animal lectins shows specific calcium-dependent binding of various monosaccharides with dissociation constants in the mM range [30]. The majority of C-type lectins can be divided into D-mannose binding (also recognising D-glucose, L-fucose, etc.) and D-galactose binding ones. The first complex with galactose was obtained with the rat mannose binding protein A (MBPA) that, after appropriate mutations, yielded the galactose binding mutant QPDWG [31]. Recently, TC14, a tunicate galactose-specific C-type lectin from *Polyandrocarpa misakiensis*, has been co-crystallised with D-galactose [32]. Despite the lack of similarity at the sequence or at the fold level, the galactose binding mode of PA-IL shows a striking resemblance to each of these C-type lectins (Fig. 2). In all of them, the galactose O3 and O4 hydroxyl groups participate in the calcium ion co-ordination and O4 establishes a strong contact with one Asp residue that participates in the co-ordination.

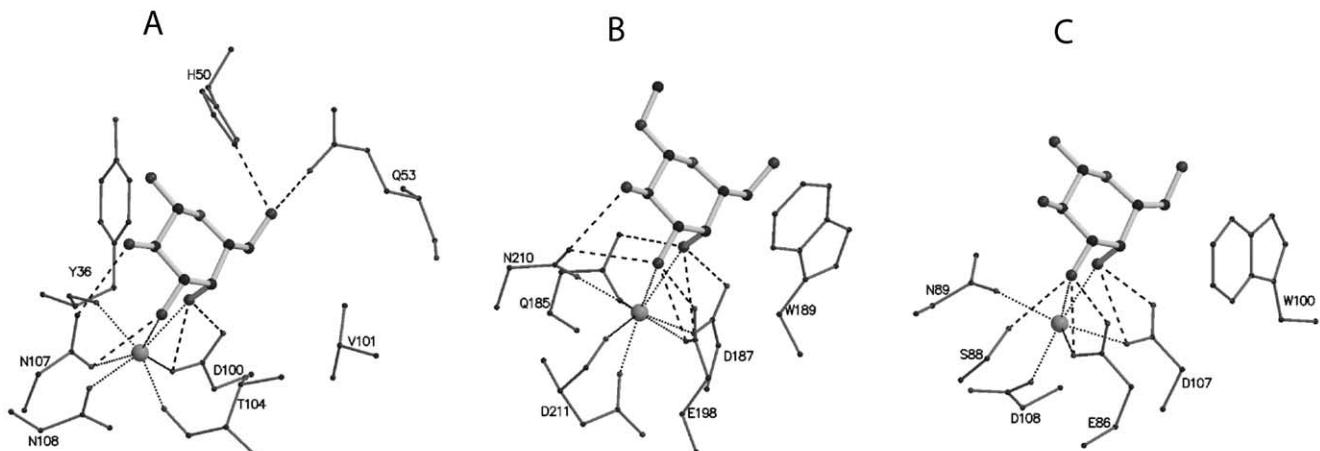


Fig. 2. Comparison of the calcium and galactose binding sites in three different lectin crystal structures. A: PA-IL/βGal complex (this work). B: MBPA (QPDWG)/βGal-O-Me complex (PDB code 1AFA) [31]. C: TC14/βGal complex (PDB code 1TLG) [32]. The galactose ring is shown in the same orientation in the three binding sites.

Since recognition of the axial orientation of O4 is an important point for galactose specificity, this aspartate residue plays a key role in the binding site.

PA-IL displays medium range affinity for D-galactose, with an association constant (K_a) of $3.4 \times 10^4 \text{ M}^{-1}$ as reported from an equilibrium dialysis study [11]. This 10-fold increase in galactose affinity compared to that of C-type lectins could be attributed to the additional strong interaction with O6 of the sugar. PA-IL's limited specificity for galactose does not allow other monosaccharide binding, with the exception of N-acetyl-D-galactosamine that can bind, albeit with a lower affinity than galactose. The presence of a hydrophobic group on the sugar anomeric position, in either α or β configuration, enhances the affinity with the tightest binding obtained for phenyl- β -thiogalactoside [11]. This result is in agreement with the observed galactose binding mode, where only hydroxyl O1 and, to a lesser extent, O2 are exposed to the solvent and are therefore free to carry substitution. PA-IL binds to a large number of glycoproteins [13] and glycosphingolipids [12,15]. Amongst them, the glycoconjugates containing terminal non-substituted α Gal1-4Gal disaccharide, i.e. the human Pk blood group, present on either red blood cell glycosphingolipids or pigeon egg white glycoproteins [33], and α Gal1-3Gal disaccharide, the xeno-antigen present on non-human tissues, are bound most strongly by the lectin. These two disaccharides together with melibiose, α Gal1-6Glc, inhibit the binding of PA-IL to glycoconjugates. Further studies are needed to determine if other galactose-containing ligands such as asialo-GM1 and GM2 and Lewis oligosaccharides that were reported to be present in the lungs of cystic fibrosis patients [4,34] could act as high affinity receptors for PA-IL.

The determination of the crystal structures of the two soluble lectins of *P. aeruginosa* in complex with their monosaccharide ligands gives hope for the future design of anti-adhesive carbohydrate-based therapeutics. The two lectins do not seem to be evolutionarily related even though they both adopt a jelly-roll fold. More interestingly, they both use calcium for carbohydrate binding. In PA-IIL the two calcium ions are associated with high affinity (micromolar range) and a wide range of different monosaccharides that are able to bind. In contrast, in PA-IL, calcium together with a unique network of hydrogen bonds generates a binding site endowed with high specificity for galactose. PA-IL's narrow specificity, together with its unique sequence, makes this protein an attractive target for drug design.

Acknowledgements: We thank the ESRF, Grenoble, for access to synchrotron data collection facilities. Financial support is granted through the French Research Ministry Program ACI Microbiology and Vaincre la Mucoviscidose. Travel and visits between NCBR and CERMAV are supported by a BARRANDE exchange program and partial financial support is acknowledged from the Ministry of Education of the Czech Republic, Grant MSM 143100005. G.C. is an EEC doctoral fellow (HPRN-CT2000-00001).

References

- [1] Bock, K., Karlsson, K.A., Stromberg, N. and Teneberg, S. (1988) *Adv. Exp. Med. Biol.* 228, 153–186.
- [2] Hooper, L.V. and Gordon, J.I. (2001) *Glycobiology* 11, 1R–10R.
- [3] Lamblin, G., Degroote, S., Perini, J.M., Delmotte, P., Scharfman, A., Davril, M., Lo-Guidice, J.M., Houdret, N., Dumur, V., Klein, A. and Roussel, P. (2001) *Glycoconj. J.* 18, 661–684.
- [4] Scharfman, A., Arora, S.K., Delmotte, P., Van Brussel, E., Mazarier, J., Ramphal, R. and Roussel, P. (2001) *Infect. Immun.* 69, 5243–5248.
- [5] Sheth, H.B., Lee, K.K., Wong, W.Y., Srivastava, G., Hindsgaul, O., Hodges, R.S., Paranchych, W. and Irvin, R.T. (1994) *Mol. Microbiol.* 11, 715–723.
- [6] Gilboa-Garber, N. (1982) *Methods Enzymol.* 83, 378–385.
- [7] Gilboa-Garber, N. (1983) in: *Lectins: Biology, Biochemistry, Clinical Biochemistry*, Vol. 3 (Bog-Hansen, T. and Spengler, G.A., Eds.), pp. 495–502, Walter de Gruyter, Berlin.
- [8] Winzer, K., Falconer, C., Garber, N.C., Diggle, S.P., Camara, M. and Williams, P. (2000) *J. Bacteriol.* 182, 6401–6411.
- [9] Gilboa-Garber, N., Mizrahi, L. and Garber, N. (1972) *FEBS Lett.* 28, 93–95.
- [10] Avichezer, D., Katcoff, D.J., Garber, N.C. and Gilboa-Garber, N. (1992) *J. Biol. Chem.* 267, 23023–23027.
- [11] Garber, N., Guempel, U., Belz, A., Gilboa-Garber, N. and Doyle, R.J. (1992) *Biochim. Biophys. Acta* 1116, 331–333.
- [12] Lanne, B., Ciopraga, J., Bergstrom, J., Motas, C. and Karlsson, K.A. (1994) *Glycoconj. J.* 11, 292–298.
- [13] Chen, C.P., Song, S.C., Gilboa-Garber, N., Chang, K.S. and Wu, A.M. (1998) *Glycobiology* 8, 7–16.
- [14] Stoitsova, S.R., Boteva, R.N. and Doyle, R.J. (2003) *Biochim. Biophys. Acta* 1619, 213–219.
- [15] Gilboa-Garber, N., Sudakevitz, D., Sheffi, M., Sela, R. and Levene, C. (1994) *Glycoconj. J.* 11, 414–417.
- [16] Bajole-Laudinat, O., Girod-de Bentzmann, S., Tournier, J.M., Madoulet, C., Plotkowski, M.C., Chippaux, C. and Puchelle, E. (1994) *Infect. Immun.* 62, 4481–4487.
- [17] Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gauzier, C., Pérez, S., Wu, A.M., Gilboa-Garber, N. and Imbert, A. (2002) *Nat. Struct. Biol.* 9, 918–921.
- [18] Loris, R., Tielker, D., Jaeger, K.-E. and Wyns, L. (2003) *J. Mol. Biol.* 331, 861–870.
- [19] Karaveg, K., Liu, Z.J., Tempel, W., Doyle, R.J., Rose, J.P. and Wang, B.C. (2003) *Acta Crystallogr. D Biol. Crystallogr.* 59, 1241–1242.
- [20] Liu, Z.J., Tempel, W., Lin, D., Karaveg, K., Doyle, R.J., Rose, J.P. and Wang, B.C. (2002) *Am. Crystallogr. Assoc. Abstr. Papers* 29, 98.
- [21] Leslie, A.G.W. (1992) Joint CCP4+ESF-EAMCB Newsletter on Protein Crystallography 26.
- [22] Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. D50*, 760–763.
- [23] Vagin, A. and Teplyakov, A. (1997) *J. Appl. Crystallogr.* 30, 1022–1025.
- [24] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) *Nucleic Acids Res.* 28, 235–242.
- [25] Perrakis, A., Morris, R. and Lamzin, V.S. (1999) *Nat. Struct. Biol.* 6, 458–463.
- [26] Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. and Dodson, E.J. (1999) *Acta Crystallogr. D* 55, 247–255.
- [27] Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard, M. (1991) *Acta Crystallogr. A* 47, 110–119.
- [28] Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) *J. Mol. Biol.* 247, 536–540.
- [29] Czjzek, M. et al. (2001) *J. Biol. Chem.* 276, 48580–48587.
- [30] Drickamer, K. (1993) *Curr. Opin. Struct. Biol.* 3, 393–400.
- [31] Kolatkar, A.R. and Weis, W.I. (1996) *J. Biol. Chem.* 271, 6679–6685.
- [32] Poget, S.F., Legge, G.B., Proctor, M.R., Butler, P.J., Bycroft, M. and Williams, R.L. (1999) *J. Mol. Biol.* 290, 867–879.
- [33] Lerer, B. and Gilboa-Garber, N. (2002) *Electrophoresis* 23, 8–14.
- [34] de Bentzmann, S., Roger, P., Dupuit, F., Bajole-Laudinat, O., Fuchey, C., Plotkowski, M.C. and Puchelle, E. (1996) *Infect. Immun.* 64, 1582–1588.
- [35] Kraulis, P. (1991) *J. Appl. Crystallogr.* 24, 946–950.
- [36] Merrit, E.A. and Murphy, M.E. (1994) *Acta Crystallogr. D* 50, 869–873.
- [37] Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Crystallogr.* 26, 283–291.
- [38] Vriend, G. (1990) *J. Mol. Graph.* 8, 52–56.

Chapitre 7

La lectine bactérienne CV-IIL de *Chromobacterium violaceum*

7.1 Introduction

Dans le deuxième chapitre on a vu que la recherche de séquences similaires à PA-IIL et PA-IIL dans les génomes de nouveaux microorganismes a permis l'identification de certains gènes ayant une forte similarité.

Chromobacterium violaceum est une protobactérie qui habite souvent les eaux, le sol et la végétation des régions tropicales et sub-tropicales. Le séquençage de son génome, complété en 2004 (Alves de Brito 2004), a mis en évidence une grande richesse génomique et une adaptabilité remarquable, qui se traduisent par un potentiel biotechnologique important. Bien que considéré comme un microorganisme saprophyte, *C. violaceum* est aussi un agent pathogène, responsable de cas de septicémie opportuniste dont l'évolution est souvent fatale. De par sa virulence et sa résistance innée aux antibiotiques, *C. violaceum* est considérée comme un pathogène émergent dans les régions tropicales et sub-tropicales de la planète.

L'analyse des ORF a montré la présence d'une gène Q7NX84 codant une protéine ayant une identité de séquence de 60% avec PA-IIL de *P. aeruginosa* et RS-IIL de *R. solanaceum*. Cette protéine, nommée CV-IIL, a été récemment purifiée à partir d'une culture bactérienne. Il s'agit d'une lectine qui montre une haute affinité pour le fucose et le mannose (Zinger-Yosovich 2006). Dans le travail présenté dans ce septième chapitre on a produit la CV-IIL recombinante dans *E. coli* et résolu sa structure tridimensionnelle en complexe avec ses ligands. La thermodynamique de la reconnaissance a été étudiée par microcalorimétrie de titration. Les résultats de ce travail sont une contribution importante dans l'étude de cette importante classe de lectines bactériennes.

7.2 Résultats

La lectine recombinante CV-IIL exprimée dans *E. coli* a été purifiée par chromatographie d'affinité et cristallisée en présence d' α -Me-mannoside et d' α -Me-fucoside. Sa structure tridimensionnelle, résolue à très haute résolution, a montré un mode de reconnaissance identique à celui de PA-IIL et RS-IIL, avec la présence de deux ions calcium en interaction

avec le sucre. Cependant, la force de l'interaction avec ces ligands, étudiée par ITC et test ELLA est un consensus entre la PA-IIL, spécifique pour le fucose, et la RS-IIL, spécifique pour le mannose. De plus, la forte affinité de CV-IIL semble être guidée par une contribution entropique favorable, ce qui est assez rare parmi les interactions lectine-sucre. La mesure de la capacité calorifique a montré une compensation entre enthalpie et entropie en fonction de la température qui conduit à une affinité très stable pour une gamme de température élevée.

Unusual Entropy-Driven Affinity of *Chromobacterium violaceum* Lectin CV-IIL toward Fucose and Mannose^{†,‡}

Martina Pokorná,^{§,||} Gianluca Cioci,^{||,⊥} Stephanie Perret,[⊥] Etienne Rebuffet,[⊥] Nikola Kostlánová,[§] Jan Adam,[§] Nechama Gilboa-Garber,[#] Edward P. Mitchell,[∞] Anne Imbert,^{*,⊥} and Michaela Wimmerová^{*,§}

National Centre for Biomolecular Research and Department of Biochemistry, Faculty of Science, Masaryk University, Kotlárska 2, 611 37 Brno, Czech Republic, Centre de Recherches sur les Macromolécules Végétales, CNRS (affiliated with Joseph Fourier), BP53, 38041 Grenoble cedex 09, France, and Bar-Ilan University, Ramat-Gan 52900, Israel, and ESRF Experiments Division, BP 220, F-38043 Grenoble Cedex, France

Received February 2, 2006; Revised Manuscript Received April 7, 2006

ABSTRACT: The purple pigmented bacterium *Chromobacterium violaceum* is a dominant component of tropical soil microbiota that can cause rare but fatal septicaemia in humans. Its sequenced genome provides insight into the abundant potential of this organism for biotechnological and pharmaceutical applications and allowed an ORF encoding a protein that is 60% identical to the fucose binding lectin (PA-IIL) from *Pseudomonas aeruginosa* and the mannose binding lectin (RS-IIL) from *Ralstonia solanacearum* to be identified. The lectin, CV-IIL, has recently been purified from *C. violaceum* [Zinger-Yosovich, K., Sudakevitz, D., Imbert, A., Garber, N. C., and Gilboa-Garber, N. (2006) *Microbiology* 152, 457–463] and has been confirmed to be a tetramer with subunit size of 11.86 kDa and a binding preference for fucose. We describe here the cloning of CV-IIL and its expression as a recombinant protein. A complete structure–function characterization has been made in an effort to analyze the specificity and affinity of CV-IIL for fucose and mannose. Crystal structures of CV-IIL complexes with monosaccharides have yielded the molecular basis of the specificity. Each monomer contains two close calcium cations that mediate the binding of the monosaccharides, which occurs in different orientations for fucose and mannose. The thermodynamics of binding has been analyzed by titration microcalorimetry, giving dissociation constants of 1.7 and 19 μM for α-methyl fucoside and α-methyl mannoside, respectively. Further analysis demonstrated a strongly favorable entropy term that is unusual in carbohydrate binding. A comparison with both PA-IIL and RS-IIL, which have binding preferences for fucose and mannose, respectively, yielded insights into the monosaccharide specificity of this important class of soluble bacterial lectins.

Chromobacterium violaceum is a free-living bacterium commonly found in the aquatic habitats of tropical and subtropical regions of the world. Its most visible characteristic, noticed at the end of 19th century, is the production of a deep violet pigment named violacein (2). This compound is of great interest because of its antibiotic activity against several tropical pathogens, including *Mycobacterium tuberculosis*, *Trypanosoma cruzi*, and several *Leishmania* species (3). Besides violacein and some related antibiotics, *C. violaceum* produces many other compounds of pharmaceutical or biotechnological interest (4) such as polyhydroxyal-

canoate polymers that could be useful in the production of biodegradable plastics and cyanide that can be used to enhance gold recovery. The bacterium also has potential applications in bioremediation since it can degrade halogenated compounds. Due to the prospective biotechnological applications, the genome of *C. violaceum* strain ATCC 12472 was fully sequenced and annotated (5).

Although *C. violaceum* is considered as a saprophyte, it is also an occasional pathogen of humans and animals and may cause fatal septicaemia from skin lesions with many liver and lung abscesses. Most cases of human infection occur either early in childhood or in immunocompromised individuals (6). The high mortality rate of the infections, once declared, can be related to the unusual resistance of the bacteria to antibiotics (7).

Comparison of the *C. violaceum* ORFs to those of other free-living bacteria from soil reveals that 17.4% are most closely similar to ORFs of the bacterial phytopathogen *Ralstonia solanacearum* (8) and 9.61% to ORFs of *Pseudomonas aeruginosa* (9), which causes opportunistic infections in humans. A high percentage of ORFs (9.75%) also are quite similar to those of *Neisseria meningitidis* serogroup A (10), the causal agent of a serious human disease. Among the genes of *C. violaceum* that present orthologs in both of the

[†] This work was supported by the Ministry of Education (Contract MSM0021622413) and the Grant Agency of Czech Republic (Contract 204/03/H016), by CNRS, by the French Ministry of Research ACI Microbiologie program, and by Mizutani Foundation for Glycosciences.

[‡] The atomic coordinates and structure factors (entries 2BOI and 2BV4) have been deposited in the Protein Data Bank.

^{*} To whom correspondence should be addressed. A.I.: telephone, +33-476037636; fax, +33-476547203; e-mail, imberty@cermav.cnrs.fr. M.W.: telephone, +420-549498166; fax, +420-549492690; e-mail, michaw@chemi.muni.cz.

[§] Masaryk University.

^{||} These authors contributed equally to this work.

[⊥] CNRS (affiliated with Joseph Fourier).

[#] Bar-Ilan University.

[∞] ESRF Experiments Division.

R. solanacearum and *P. aeruginosa* genomes, one encodes the putative protein Q7NX84 that is similar to PA-IIL,¹ the lectin with a preference for fucose in *P. aeruginosa* (11), and to RS-IIL, the lectin with mannose preference in *R. solanacearum* (12). PA-IIL is a calcium-dependent tetrameric lectin with an unusually high affinity for fucose and fucose-containing glycoconjugates (13, 14) and is involved in adhesion and biofilm formation (15). The putative lectin from *C. violaceum*, named CV-IIL, is 113 amino acids long (without the initiating methionine), and its sequence is 60 and 62% identical with the sequences of PA-IIL and RS-IIL, respectively (16).

The production and properties of native CV-IIL have recently been described (1). The lectin is produced as a quorum sensing-driven secondary metabolite, appearing concomitantly with violacein, and its formation is repressed in the *N*-acyl homoserine lactone-deprived CV026 mutant. CV-IIL was purified on a mannose-Sepharose column as a tetramer with a subunit size (~11.86 kDa) similar to those of PA-IIL (11.73 kDa) and RS-IIL (11.60 kDa). CV-IIL is able to agglutinate blood group O erythrocytes and has the following monosaccharide affinities that are very similar to those observed for PA-IIL: L-fucose (Fuc) > D-arabinose (Ara) > D-fructose (Fru) > D-mannose (Man) (while RS-IIL has a preference for mannose).

The high resolution X-ray crystal structure of PA-IIL has helped to clarify the atomic basis for the unusually high affinity between the lectin and the monosaccharide that appears to be related to the presence of two calcium ions in the sugar binding site (17, 18). Further structural characterization of PA-IIL complexed with other monosaccharides demonstrated that D-mannose mimics L-fucose in the binding site, albeit with a different orientation of the sugar ring that presents the hydroxyl groups isosterically in the binding site (19). Comparison with the crystal structure of the RS-IIL- α -Me-mannoside complex demonstrated that a stretch of three amino acids in the binding site may be responsible for the preference for fucose over mannose (20).

In this study, CV-IIL was produced in its recombinant form in *Escherichia coli*. Specificity toward different monosaccharides was investigated by enzyme-linked lectin assay (ELLA) methods. The unusual thermodynamics of binding was investigated using isothermal titration microcalorimetry (ITC), while the molecular basis was investigated by determining two X-ray crystal structures of CV-IIL in complex with methyl α -L-fucopyranoside (Me- α -Fuc) and methyl α -D-mannopyranoside (Me- α -Man). Comparison with the thermodynamic and structural characteristics of related *P. aeruginosa* and *R. solanacearum* lectins is discussed.

EXPERIMENTAL PROCEDURES

Materials. All monosaccharides used in this study were purchased from Sigma-Aldrich (St. Louis, MO) except Me- α -Fuc which was purchased from Interchim (Montluçon, France).

¹ Abbreviations: CV-IIL, lectin II from *C. violaceum*; PA-IIL, lectin II from *P. aeruginosa*; RS-IIL, lectin II from *R. solanacearum*; Ara, D-arabinose; Fuc, L-fucose; Man, D-mannose; Me- α -Fuc, α -methyl L-fucopyranoside; Me- α -Man, α -methyl D-mannopyranoside; ELLA, enzyme-linked lectin assay; ITC, isothermal titration microcalorimetry.

Cloning and Production of Recombinant CV-IIL. The following oligonucleotides were used as primers: 5'-AGG AGT TCA TAT GGC TCA GCA AGG CG-3' (26-mer) and 5'-GCG AAG CTT CAG CCC AGC GG-3' (20-mer). The former was designed for the introduction of NdeI and the latter for HindIII restriction sites (underlined sequences), respectively. PCR was performed using Pfu polymerase (Promega) and genomic DNA from *C. violaceum* CCM 2076 = ATCC 7461 obtained from the Czech Collection of Microorganisms (CCM, Masaryk University) as a template. After digestion with NdeI and HindIII, the amplified fragment was introduced into the multiple cloning site of the pET-25(b+) vector (Novagen, Madison, WI), resulting in plasmid pET25cv2l.

E. coli Tuner (DE3) cells containing the pET25cv2l plasmid were cultured in 1 L of Luria broth at 37 °C. When the culture reached an optical density of 0.5–0.6 at 600 nm, isopropyl β -D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.5 mM. Cells were harvested after incubation for 3 h at 30 °C, washed, and resuspended in 10 mL of buffer [20 mM Tris-HCl, 100 mM NaCl, and 100 μ M CaCl₂ (pH 7.5)]. The cells were disrupted by sonication (Soniprep 150, Schoeller Instruments). After centrifugation at 10000g for 1 h, the supernatant was further purified on a D-mannose-agarose column (Sigma-Aldrich). The CV-IIL lectin was loaded onto the column in the equilibrating buffer [20 mM Tris-HCl, 100 mM NaCl, and 100 μ M CaCl₂ (pH 7.5)] and then was eluted with the elution buffer [20 mM Tris-HCl, 100 mM NaCl, 100 μ M CaCl₂, and 0.1 M D-mannose (pH 7.5)]. The purified protein was dialyzed against distilled water for 1 week (for the removal of D-mannose), concentrated by lyophilization, and stored at –20 °C.

Cloning and Production of Recombinant RS-IIL. The rs2l gene was amplified by PCR using genomic DNA from the *R. solanacearum* ATCC 11696 strain using the following primers: GAG ATT CAC TCC ATA TGG CTC AGC AAG GTG and CAA AGC TTG GGA ACC GAT CAG CCC AGC. The resulting PCR product was digested with NdeI and HindIII (underlined sequences correspond to their recognition sites), and the insert was subcloned into the *E. coli* pRSET A expression vector (Novagen) that was digested similarly. Cells harboring the pRSETrs2l plasmid were grown in LB broth containing 100 μ g/mL ampicillin at 37 °C. Recombinant RS-IIL was further produced and purified as described above for CV-IIL.

Production of Recombinant PA-IIL. Recombinant PA-IIL was purified from *E. coli* BL21(DE3) containing the pET25pa2l plasmid as described previously (18).

ELLA Experiments. Enzyme-linked lectin assays (ELLAs) were performed as described previously (14). Briefly, 96-well microtiter plates (Nunc Maxisorb) were coated with recombinant CV-IIL (30 μ g/mL) in carbonate buffer (pH 9.6). Plates were blocked with 3% BSA in PBS and incubated with L-fucose-biotinylated PAA (Lectinity Holding, Inc.) at 5 μ g/mL in the presence of serial dilution of tested sugars. Streptavidin-peroxidase conjugate (Boehringer-Mannheim, Mannheim, Germany) was then added, and the color was developed using the substrate kit for peroxidase, Sigma Fast OPD (Sigma-Aldrich). The reaction was stopped by adding 30% H₂SO₄, and the absorbance was read at 490 nm in a microtiter plate reader (Bio-Rad, model 680).

ITC Analysis. Titration calorimetry experiments were performed using a Microcal (Northampton, MA) VP-ITC microcalorimeter. All titrations were made in 0.1 M Tris-HCl buffer containing 30 μM CaCl₂ (pH 7.5) at 25 °C. Aliquots of 10 μL of each carbohydrate, dissolved in the same buffer, were added at 5 min intervals to the lectin solution present in the calorimeter cell. In the titrations, the protein concentration in the cell varied from 0.045 to 0.075 mM and from 0.1 to 0.25 mM for Me- α -Fuc and Me- α -Man, respectively. The sugar concentration in the 250 μL syringe was 10–20 times higher than the protein concentration used in the experiment (typically 0.4–0.7 mM for the high-affinity ligands, Me- α -Fuc and Me- α -Man for CV-IIL and RS-IIL, and 1.7–3.5 mM for the low-affinity ligand, Me- α -Man for CV-IIL). The temperature of the cell was kept at 25 ± 0.1 °C. Experiments for the evaluation of heating capacity were performed using similar conditions. Aliquots of 15 and 10 μL of 0.3 mM Me- α -Fuc were added at 5 and 8 min intervals for CV-IIL and PA-IIL, respectively, to the lectin solution present in the calorimeter cell. Data were collected at temperatures varying from 5 to 45 °C. Data from a control experiment performed via identical injections of monosaccharide 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 7). Fitted data yielded the association constant (K_a) and the enthalpy of binding (ΔH). Other thermodynamic parameters, i.e., changes in free energy (ΔG) and entropy (ΔS), were calculated from the equation

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

where T is the absolute temperature and $R = 8.314 \text{ J mol}^{-1} \text{ K}^{-1}$. At least two independent titrations were performed for each ligand that was tested. Linear regression analysis of enthalpy data versus temperature was used for calculation of the heat capacity (ΔC_p).

Crystallization. Lyophilized CV-IIL was dissolved in pure water at a concentration of 7 mg/mL in the presence of salts (CaCl₂ and MgCl₂, 2 mM) and sugars (Me- α -Fuc or Me- α -Man, 5 mM). The protein was crystallized by vapor diffusion using the hanging drop method. Initial crystallization trials were performed with the Jena Bioscience PEG8000 screen, which clearly indicated PEG 8000 in combination with inorganic salts as the starting conditions. Optimization of the initial conditions resulted in well-formed single crystals growing in 1 week from 2.5 μL of the protein sample mixed with 2.5 μL of 10% PEG 8000 and 0.1 M (NH₄)₂SO₄. For crystallization of the Me- α -Man complex, NaCl was used instead of (NH₄)₂SO₄. For data collection, the crystals were soaked in mother liquor containing 30% glycerol for as little time as possible and flash-cooled in liquid nitrogen.

Data Collection and Structure Determination. Atomic-resolution diffraction data were collected at ESRF beamline ID14-2 to 1.1 Å for the Me- α -Fuc complex and to 1.0 Å for the Me- α -Man complex. Diffraction data were processed with MOSFLM (21) and scaled and converted to structure factors using the CCP4 program suite (22). Both crystals belong to space group $P2_12_12$ with two protein molecules in the asymmetric unit. Two successive passes at low and high resolution, using one crystal, were collected and merged together. The phases for the structure were determined with

ACORN (23), using the positions of four calcium ions in the asymmetric unit as a starting point. The positions of the calcium ions were determined by molecular replacement using MOLREP (24) and the PA-IIL (PDB entry 1GZT) monomer as the search model. The two chains were traced, and a first solvent model was made automatically by ARP/Warp (25) using the ACORN experimental phases. Refinement of the model was performed with REFMAC (26) combined with manual rebuilding with O (27). For the final stages of refinement, anisotropic B -factors were used for all atoms. Alternative conformations were inserted where necessary and their occupancies estimated according to the refined B -factors. The data collection and refinement statistics are listed in Table 1.

RESULTS

Expression and Purification of Recombinant Proteins. Using the pET25cv2l expression plasmid, CV-IIL was expressed in *E. coli* Tuner (DE3) with a typical yield of ~30 mg of purified CV-IIL per liter of culture. The resulting protein was purified by affinity chromatography and moved as a sharp band of ~11 kDa. Mass spectrometry analysis confirmed the molecular mass of 11.841 kDa, corresponding to the CV-IIL amino acid sequence lacking the initial methionine residue. The expression level of RS-IIL in *E. coli* was lower than those obtained with the other lectins with a typical yield of ~10 mg of purified protein per liter of culture. However, the protein was found to be less stable than PA-IIL and CV-IIL, and only basic thermodynamics characterization could be carried out.

Monosaccharide Specificity of Recombinant CV-IIL. ELLA experiments conducted on recombinant CV-IIL confirmed that it has the same activity and the same binding preference as the wild-type protein studied by the inhibition of haemagglutination (1). The assays confirm that Fuc is superior to Ara, Man, or Fru for binding but that the differences are small, and therefore, all these free monosaccharides can be described as good ligands for CV-IIL (Table 2). The affinity is stronger for the methyl derivative of fucose, as commonly observed in protein–carbohydrate interactions. A comparison with the data obtained with PA-IIL indicates that both lectins bind preferentially to Fuc but that PA-IIL has a much stronger preference for monosaccharides with the L-galacto configuration (i.e., Fuc and Ara). While the affinity of PA-IIL for the latter is 200-fold higher than for the D-manno configuration (i.e., Man and Fru), the affinity of CV-IIL for them is only 6 times higher.

Crystal Structures of CV-IIL–Monosaccharide Complexes. The crystal structures of CV-IIL in complex with both Me- α -Fuc and Me- α -Man contain two monomers of 113 amino acids per asymmetric unit, together with four Ca²⁺ ions and two sugar residues. The overall topology of one monomer is that of a nine-stranded antiparallel β -sandwich consisting of two sheets, one with four strands and one with five. The dimer is made around a pseudo-2-fold axis of symmetry that brings the hydrophobic face of one β -sheet into close association with its counterpart. Dimers associate to form tetramers by a 2-fold axis of symmetry, creating ball-shaped structures (Figure 1) that are very similar to those described previously for the soluble lectins from *P. aeruginosa* (17) and *R. solanacearum* (20).

Table 1: Statistics for Data Collection and Refinement

	CV-IIL–Me- α -Fuc	CV-IIL–Me- α -Man
PDB entry	2BOI	2BV4
beamline	ID14-2	ID14-2
wavelength (Å)	0.934	0.934
resolution range (Å)	34.400–1.100 (1.129–1.100) ^a	23.200–1.000 (1.026–1.000) ^a
space group	$P2_12_12$	$P2_12_12$
a (Å)	51.09	50.68
b (Å)	89.84	89.50
c (Å)	46.48	46.48
total no. of observations	712652 (44825) ^a	802154 (23973) ^a
no. of unique reflections	87549 (6366) ^a	114226 (7874) ^a
completeness	100.0 (100.0) ^a	99.6 (94.4) ^a
R_{merge} (%)	7.4 (30.0) ^a	6.0 (44.0) ^a
$\langle I/\sigma(I) \rangle$	17.9 (5.6) ^a	17.6 (2.0) ^a
multiplicity	8.1 (7.0) ^a	7.0 (3.0) ^a
ACORN phasing statistics		
correlation coefficient	0.44 (66947) ^{a,b}	0.42 (88443) ^a
R -factor	0.41 (66947) ^a	0.45 (88443) ^a
refinement		
R_{cryst} (%)	10.2 (12.0) ^a	10.7 (26.0) ^a
R_{free} (%)	11.7 (13.7) ^a	12.3 (26.1) ^a
rms deviation		
bonds (Å)	0.014	0.014
angles (deg)	1.634	1.745
solvent content (%)	45	45
estimated coordinate error based on maximum likelihood (Å)	0.012	0.011
average B -factor for main chains/ side chains/others (Å ²)	6.1/8.3/18.3	6.5/8.8/19.4
no. of Ramachandran outliers	0	0
no. of amino acids modeled with alternative conformations	20	26

^a Values in parentheses refer to the highest-resolution shell. ^b Number of medium E reflections.

Table 2: Inhibitory Concentration of Various Carbohydrates Giving 50% Inhibition of Binding of CV-IIL or PA-IIL to L-Fucose-PAA-Biotin in a Microtiter Plate Assay^a

monosaccharide	CV-IIL		PA-IIL	
	IC ₅₀ (μM) ^b	potency ^c	IC ₅₀ (μM) ^b	potency ^c
Me- α -Fuc	48	2.4	2.3	2.6
L-fucose	112	1	6	1
D-arabinose	197	0.55	67	0.09
Me- α -Man	600	0.18	nd ^d	nd ^d
D-mannose	700	0.16	1400	0.0043
D-fructose	860	0.13	1650	0.0036

^a Standard deviations are smaller than 15% except for the PA-IIL–mannose experiments (23%). ^b IC₅₀ values were determined from inhibition curves. Each value represents the mean of at least three independent measurements. ^c Potency is the reciprocal of the inhibiting concentration relative to fucose (taken to be 1.0). ^d Not determined.

In the complexes of CV-IIL with Me- α -Fuc and Me- α -Man, two Ca²⁺ ions and one sugar residue can be clearly observed in each binding site. All amino acids involved in calcium binding are strictly conserved among PA-IIL, RS-IIL, and CV-IIL, and the two close Ca²⁺ ions are bound in a similar manner. This binding involves the “calcium binding loop”, via the side chains of Glu94, Asp98, Asp100, Asn102, and Asp103, a second loop through the side chain of Asn21, and also the acidic group of the C-terminal Gly113 of the other tightly associated monomer (Figure 1).

The electron density of each bound sugar is shown in Figure 2. In the CV-IIL–Me- α -Fuc complex, three of the hydroxyl groups, namely, O₂, O₃, and O₄, participate directly in the coordination of the Ca²⁺ ions, with a special

position for O₃ that participates in the coordination sphere of both. These three hydroxyl groups also establish numerous hydrogen bonds with amino acids (Table 3 and Figure 1B). Ring oxygen O₅ receives a hydrogen bond from the backbone nitrogen of Ala23. One water molecule plays a special role by bridging O₁ and O₂ of Me- α -Fuc with the backbone nitrogen atoms of two amino acids, namely, Thr97 and Asp98. Hydrophobic contacts are limited to the methyl group at C₅ that lies in a shallow hydrophobic pocket limited by the side chains of Ala23 and Thr45.

Although the orientation of Me- α -Man in the other CV-IIL complex is completely different from that of Me- α -Fuc, a very similar network of hydrogen bonds is established (Table 3 and Figure 1C). O₃ and O₅ of Me- α -Man are in the same position as those of Me- α -Fuc, but O₂ takes the place of O₄ (and vice versa) and C₅ the place of C₁. As a result, O₄, O₃, and O₂ are involved in calcium coordination and hydrogen bonds with the protein. The trapped water molecule observed in the Me- α -Fuc complex is conserved and connects the O₄ atom to the backbone of Thr97 and Asp98. O₁ and its methyl group are not involved in any contacts. Interestingly, the interaction of Me- α -Man with CV-IIL does not involve any detectable hydrophobic contacts, which is particularly unusual in protein–carbohydrate interactions. The hydroxymethyl group at C₅ is observed in the three staggered conformations that have been demonstrated to occur in monosaccharides with the D-manno configuration either in crystal structures (28) or in solution (29). In monomer A, electron density indicates the occurrence of both *gt* (*gauche*–*trans* referring to O₅ first and C₄ second)

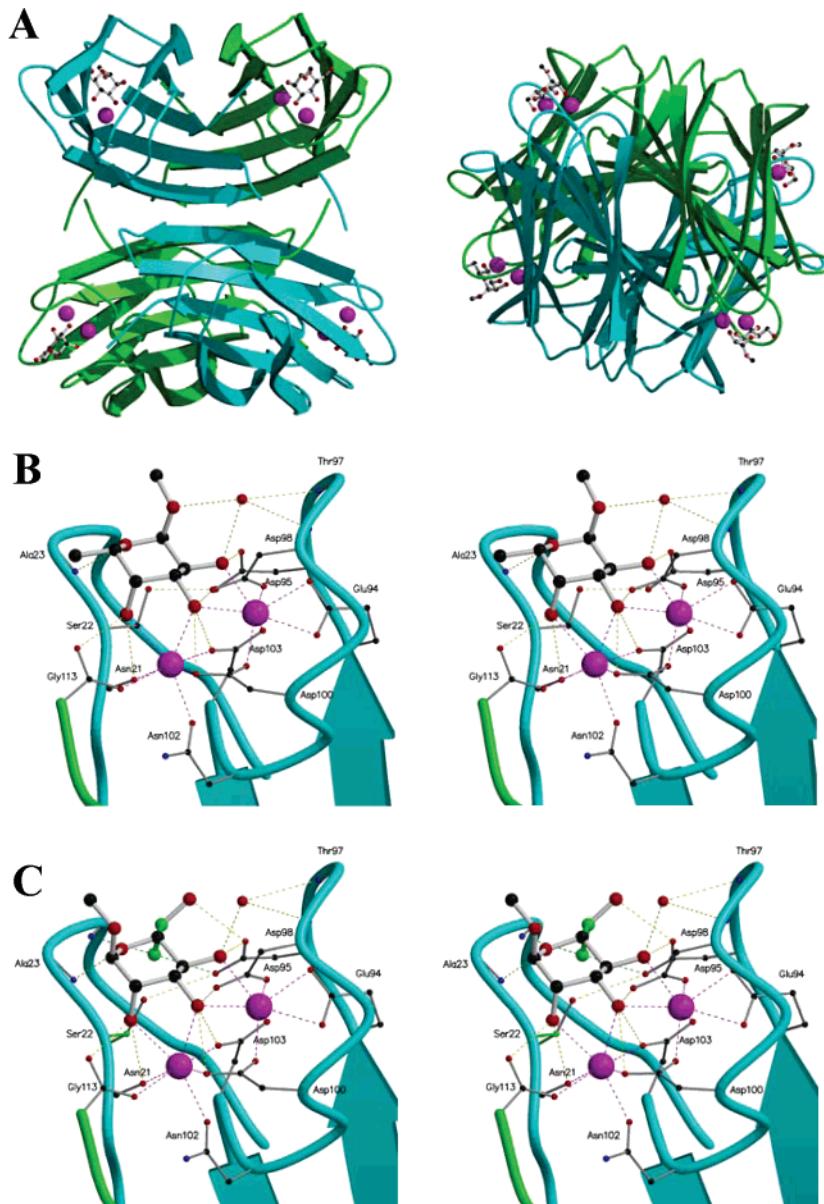


FIGURE 1: Graphical representation of CV-IIL–monosaccharide complexes. (A) Orthographic views of the tetramer of CV-IIL (ribbon) complexed with Me- α -Man (sticks) and calcium ions (CPK). (B) Stereoscopic representation of the binding site in the CV-IIL–Me- α -Fuc complex. Coordination bonds and hydrogen bonds are represented with violet and yellow dashed lines, respectively. (C) Stereo representation of the binding site in the CV-IIL–Me- α -Man complex. The alternative conformation of O6 of mannose and Ser22 are colored green. Molecular drawings were prepared with MOLSCRIPT (39) and RASTER-3D (40).

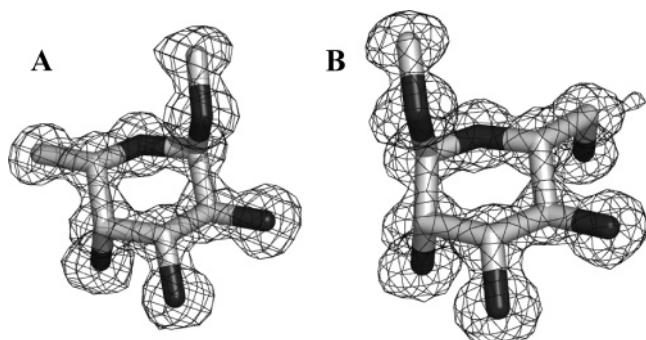


FIGURE 2: σ -weighted $2F_0 - F_c$ electron density map contoured at 1σ around (A) Me- α -Fuc in the CV-IIL–Me- α -Fuc complex and (B) Me- α -Man in the CV-IIL–Me- α -Man complex.

and *gg* conformations, whereas in monomer B, the *gg* and *tg* conformations are observed. These different conformations generate different hydrogen bonding systems. While *gt* and

tg conformations bring O6 out of the binding site, the *gg* orientation directs O6 toward the floor of the site, creating a hydrogen bond with both the backbone of Ala24 and the acidic side chain of Asp95. Interestingly, this *gg* orientation is correlated with an alternative conformation of Ser22 since it brings O6 of Me- α -Man to the position normally occupied by the hydroxyl group of this amino acid.

Thermodynamics of Binding. The result of a typical titration calorimetry measurement is shown in Figure 3. The results exhibited a monotonic decrease in the exothermic heat of the binding until saturation is achieved. The values of binding affinity K_a , binding enthalpy ΔH , and stoichiometry per monomer n have been obtained by fitting a classical equation for single-site binding (30). The thermodynamic parameters, free energy, and entropy of binding (ΔG and ΔS) were calculated as indicated in Experimental Procedures. The affinity constants obtained here for binding of CV-IIL

Table 3: Comparison of Contact Distances (in angstroms) in the CV-IIL Binding Site^a

	Me- α -Fuc	Me- α -Man	
Calcium Coordination			
Ca1	O2	2.51	O4
Ca1	O3	2.46	O3
Ca2	O3	2.48	O3
Ca2	O4	2.51	O2
Hydrogen Bonds			
Asp95 OD1	O2	2.63	O4
Asp98 OD1	O3	2.56	O3
Asp98 OD2	O3	2.97	O3
Asp100 OD1	O3	2.92	O3
Asp100 OD2	O3	2.98	O3
Asp103 OD2	O3	3.01	O3
Asn21 O	O4	3.03	O2
Gly113* OXT ^b	O4	2.56	O2
Ala23 N	O5	3.04	O5
Ala24 N		O6-gg ^e	3.03
Asp95 OD2		O6-gg ^e	2.7
Asp95 OD1		O6-tg ^f	2.72
Wat_1 ^c	O2	3.09	O4
Wat_1	O1	3.04	
Wat_2 ^d			O6A-gt ^g
Hydrophobic Contacts			
Ala23 CB	C6 _{Me}	3.85	C1 _{Me}
Thr45 CG2	C6 _{Me}	3.84	C1 _{Me}
			3.99
			3.85

^a Distances given are means from the distances observed in the two independent molecules in the asymmetric unit. ^b The C-terminal residue from the neighboring monomer. ^c Water bridging Thr97 NH and Asp98 NH. ^d Water bridging Asn69 ND2. ^e O6-gg has 30% occupancy in chain A and 70% in chain B. ^f O6-tg has 30% occupancy in chain B. ^g O6-gt has 70% occupancy in chain A.

to monosaccharides are in the micromolar range, with the highest-affinity ligand being Me- α -Fuc with a dissociation constant of 1.7×10^{-6} M (Table 4), much better than those usually observed for classical lectin–oligosaccharide interactions (31). The 11-fold difference between the affinity for Me- α -Fuc and that for Me- α -Man is in agreement with the relative inhibitory potencies calculated from the ELLA experiments.

PA-IIL was demonstrated before to also have high affinity for fucose (18) and fucose-containing oligosaccharides (14). When compared to PA-IIL, CV-IIL exhibits a slightly lower affinity for Me- α -Fuc, but in contrast, it is better adapted to accommodate Me- α -Man in its binding site (Table 4). The main differences between both protein–sugar interactions are shown in the thermodynamic contributions. The binding of methyl glycosides to CV-IIL is enthalpy-driven, with a ΔH of approximately -30 kJ/mol, but it also includes a strong favorable entropy contribution ($T\Delta S \sim 10$ kJ/mol) that corresponds to 34 and 41% of the free energy of binding for Me- α -Fuc and Me- α -Man, respectively. The interaction of PA-IIL with Me- α -Fuc displays different characteristics with a greater enthalpy contribution and a slightly unfavorable entropy term (32).

Because of the peculiar characteristics of CV-IIL–sugar interactions, further characterization of the thermodynamics was performed to determine the heat capacity change (ΔC_p) of the interaction. ITC measurements performed at temperatures varying from 5 to 45 °C demonstrated linear variations of both enthalpy and entropy contributions with temperature, but with different signs (Figure 4). The variation in enthalpy yielded a negative ΔC_p value for CV-IIL of -335 J mol $^{-1}$ K $^{-1}$ that is in agreement with values usually observed for

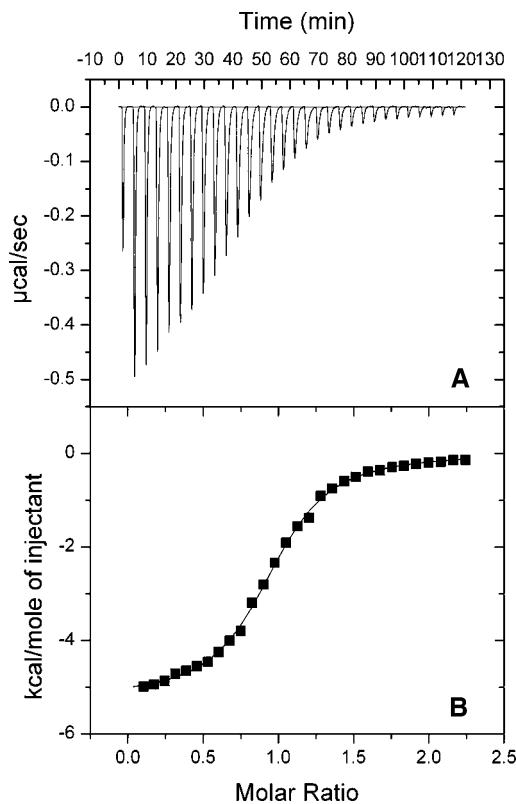


FIGURE 3: Titration microcalorimetry results for binding of Me- α -Fuc (0.4 mM) to CV-IIL (40 μ M) in 0.1 M Tris buffer (pH 7.5) with 30 μ M CaCl₂. (A) Data obtained from 30 automatic injections (10 μ L) of Me- α -Fuc each into the CV-IIL-containing cell. (B) Plot of the total heat released as a function of total ligand concentration for the titration shown in panel A. The solid line represents the best least-squares fit for the obtained data.

lectin–carbohydrate interactions (33). The same ITC measurements performed on PA-IIL showed a similar temperature dependence in ΔH , with binding becoming more exothermic with increasing temperatures and significantly nonlinear above 30 °C. This can be explained by PA-IIL denaturation at the highest assay temperatures with the unfolding process heat adding to the binding process heat. The strong negative decline in enthalpy is compensated by higher unfavorable entropy values, resulting in almost stable free energy values. The resulting heating capacity of PA-IIL calculated from the linear part of the curve (from 5 to 25 °C) gives a value of -258 J mol $^{-1}$ K $^{-1}$ (Figure 4).

The relative contributions of the different thermodynamic parameters are more difficult to rationalize. The entropic term is the sum of multiple effects. A favorable entropic contribution is obtained when the increase in solvent entropy coming from water driven out of the site is correlated with a minimal loss of conformational degrees of freedom in either ligand or protein chains during binding. Chervenak and Toone (34) demonstrated that accurate values of ΔC_p allow determination of the different entropy contributions. Briefly, the total entropy of binding can be considered as a sum of losses in configurational and cratic (rotational and translational) entropy and changes in solvation:

$$\Delta S = \Delta S_{\text{solv}} + \Delta S_{\text{config}} + \Delta S_{\text{rot}} + \Delta S_{\text{trans}}$$

ΔS_{solv} can be expressed as follows:

Table 4: Thermodynamics of Binding of CV-IIL, PA-IIL, and RS-IIL to Me- α -Fuc and Me- α -Man As Determined by ITC at 293 K^a

	$K_A (\times 10^4 \text{ M}^{-1})$	$K_D (\mu\text{M})$	n	$-\Delta G (\text{kJ/mol})$	$-\Delta H (\text{kJ/mol})$	$T\Delta S (\text{kJ/mol})$
CV-IIL						
Me- α -Fuc	59.4 \pm 0.2	1.68 \pm 0.01	0.95 \pm 0.02	32.94 \pm 0.01	21.8 \pm 0.2	11.1 \pm 0.1
Me- α -Man	5.4 \pm 0.6	19 \pm 2	0.89 \pm 0.03	27.0 \pm 0.3	17.8 \pm 1.3	9.2 \pm 1.6
PA-IIL^b						
Me- α -Fuc	235 \pm 8	0.43 \pm 0.01	0.77 \pm 0.03	36.4 \pm 0.1	41.3 \pm 1.0	-4.9 \pm 1
Me- α -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
RS-IIL						
Me- α -Fuc	nd ^c					
Me- α -Man	438 \pm 2	0.23 \pm 0.01	0.83 \pm 0.003	37.9 \pm 0.1	24.9 \pm 1.3	12.97 \pm 0.23

^a Standard deviations calculated from three independent measurements for both ligands. ^b Taken from ref 32. ^c Value of $<10^4 \text{ M}^{-1}$.

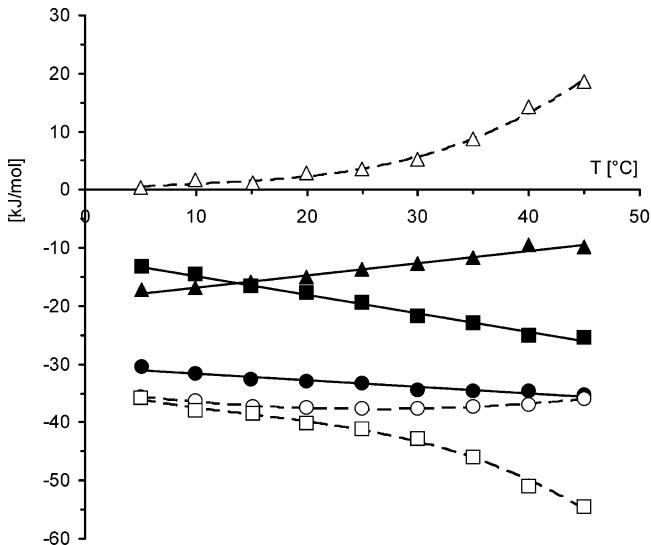


FIGURE 4: Comparison of the heat capacity plot for binding of CV-IIL (filled symbols) and PA-IIL (empty symbols) to Me- α -Fuc. The dependence of enthalpy on temperature is represented by squares together with the calculated variation of the free energy (circles) and entropy (triangles) of binding.

$$\Delta S_{\text{solv}} = \Delta S^*_{\text{solv}} + \Delta C_p \ln(T/T^*_{\text{s}})$$

where T^*_{s} is the convergence temperature with no solvent contributions and is equal to 385.15 K. ΔS^*_{solv} covers protonation and electrostatic interactions and for lectin–sugar interactions is considered to be zero. Murphy and co-workers (35) proposed that the entropic contribution, for the 1:1 complex with 1M as the standard state, is equal to $-33.5 \text{ J mol}^{-1} \text{ K}^{-1}$ (see refs 33 and 34 for details). On the basis of these calculations, ΔS_{solv} and, subsequently, ΔS_{config} were calculated for CV-IIL–Me- α -Fuc and PA-IIL–Me- α -Fuc interactions, respectively (Table 5). From these results, the entropy term due to the freezing of conformational freedom is significantly higher for binding of Me- α -Man to PA-IIL than to CV-IIL. This may be due to slight differences in the sequence of the protein loops involved in the binding site, causing smaller losses of mobility due to shorter or less mobile residues, and might also be correlated to the presence of one glycine residue in the carbohydrate binding loop of PA-IIL (Figure 5).

DISCUSSION

Molecular Basis for Fucose versus Mannose Preference in Soluble Bacterial Lectins. The CV-IIL structure belongs to a new family of soluble bacterial lectins that also includes PA-IIL, a fucose-binding lectin (with a 200-fold lower

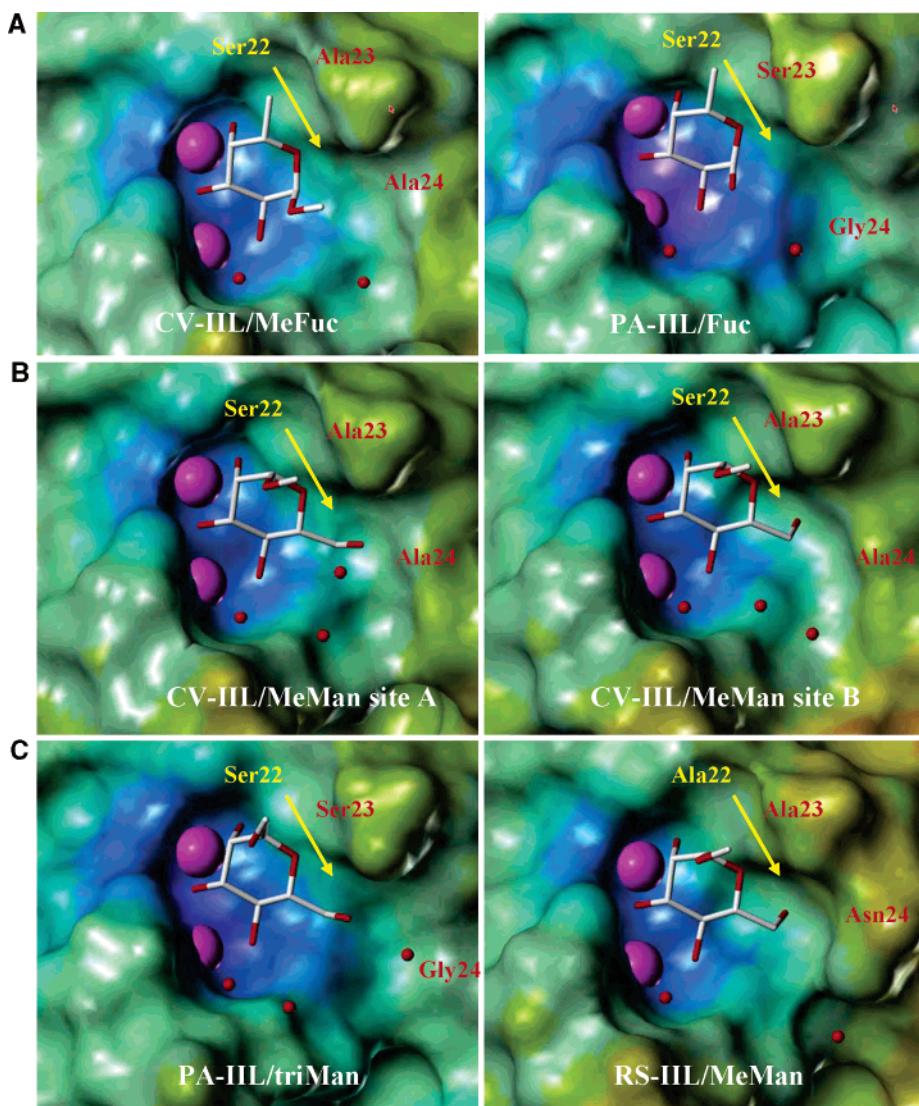
Table 5: Entropic Contribution to the Binding of Me- α -Fuc to CV-IIL and PA-IIL at 293 K

	ΔC_p (J mol $^{-1}$ K $^{-1}$)	ΔS (J mol $^{-1}$ K $^{-1}$)	ΔS_{solv} (J mol $^{-1}$ K $^{-1}$)	ΔS_{config} (J mol $^{-1}$ K $^{-1}$)
CV-IIL	-335	46.3	85.8	-6.0
PA-IIL	-258	-11.5	66.1	-44.0

affinity for mannose), and RS-IIL, a mannose-binding lectin (with a very low affinity for fucose). The preference of CV-IIL for fucose is only 6-fold stronger than that for mannose, and it can therefore be described as a fucose/mannose-binding lectin, i.e., with an intermediate specificity relative to those of PA-IIL and RS-IIL. Alignment of their sequences, as previously published (16), demonstrated that the amino acids involved in calcium binding are strictly conserved, whereas some differences are observed in the so-called “monosaccharide binding loop” in the region of amino acids 20–25 (Figure 5). The amino acids that seem to be crucial for specificity are Ser22 and Ser23 in PA-IIL, Ala22 and Ala23 in RS-IIL, and Ser22 and Ala23 in CV-IIL, which therefore appear also to be between PA-IIL and CV-IIL in terms of the degree of sequence similarity.

The resolution of the RS-IIL–mannose complex structure resulted in the conclusion that the higher affinity for mannose is due to the S22A modification since Asp95 is not involved in hydrogen bonds to Ser22 and therefore is available for bonding to O6 of mannose, yielding a higher affinity (20). However, the crystal structure of the CV-IIL–Me- α -Man complex presented here demonstrates that O6 of mannose is able to displace the OH group of Ser22 and to create the same hydrogen bond as in the RS-IIL–Me- α -Man complex. The two different mannose binding modes observed in the CV-IIL crystal are displayed in Figure 5. In site A, O6 is directed out of the site, because of the presence of the Ser22···Asp95 hydrogen bond, as observed in the crystal structure of the PA-IIL–mannose complex (19). In site B, the Ser22 hydroxyl group points in another direction and O6 can be hydrogen bound to Asp95, as observed in the crystal structure of the RS-IIL–mannose complex (20). Our study confirms that high affinity for mannose can be achieved only by mutation of Ser22 to an amino acid that will not make any hydrogen bonds with Asp95. This opens the route for the design of high-affinity mannose-specific lectins.

Entropy-Driven Binding of Fucosides in CV-IIL. While the binding site of CV-IIL is very similar to that of PA-IIL (Figure 5), the thermodynamics of binding is rather different. The affinity of CV-IIL for Me- α -Fuc is 6-fold lower than that of PA-IIL (Table 4), corresponding to a free energy decrease of $\sim 10\%$ (-33 kJ/mol vs -36 kJ/mol). However,



D	CV-IIL	20 VNSAATQ 25	93 SEDGTDM_DYND_SI 105
	PA-IIL	20 ANSSGTQ 25	94 SEDGT_DN_DYND_AV 106
	RS-IIL	20 ANAANTQ 25	93 SEDGT_DN_DYND_GI 105

FIGURE 5: (A) Comparison of binding of Me- α -Fuc by CV-IIL and fucose by PA-IIL [PDB entry 1UZV (18)]. (B and C) Comparison of binding of Me- α -Man in the two binding sites of CV-IIL, in the binding site of PA-IIL [mannose moiety of the trimannose in PDB entry 1OVS (19)], and in the binding site of RS-IIL [PDB entry 1UQX (20)]. The accessible surfaces of the proteins have been calculated with MOLCAD (41) and color-coded according to the electrostatic potential from blue for negative regions to red for positive ones. (D) Comparison of sequences of the monosaccharide binding loop (residues 20–25) and the calcium binding loop (residues 93–105).

the difference is greater for the enthalpy contribution where binding of Me- α -Fuc to CV-IIL is associated with an enthalpy change of -22 kJ/mol , versus -42 kJ/mol for binding to PA-IIL. Conversely, CV-IIL has a favorable entropy contribution of approximately -11 kJ/mol at 25°C , which corresponds to one-third of the free energy of binding. This behavior is most unusual among lectin–carbohydrate interactions that generally present a strong entropic barrier which can be attributed to either a limitation in the degrees of flexibility in the ligand (36) or solvent rearrangement (37). At present, it is difficult to correlate the thermodynamics and structural differences observed for the binding of Me- α -Fuc by CV-IIL and by PA-IIL, and further studies will be required.

Heat Capacity. The negative heat capacity observed for CV-IIL results in a less favorable enthalpy of binding at low temperatures. Nevertheless, this is offset by a positive slope for the change in the $T\Delta S$ contribution for increasing temperatures. The compensation of both enthalpy and entropy terms results in a free energy of binding that is almost independent of temperature with values varying from -31 kJ/mol at 5°C to -36 kJ/mol at 45°C , i.e., less than 10% variation in the scale of temperature variation that such an opportunistic organism can meet in its environment. This thermodynamic behavior could have evolved and be correlated to the high adaptability of *C. violaceum* to new environments and new hosts (5). PA-IIL, from *P. aeruginosa*, another highly adaptable bacterium, was also previously

demonstrated to have slightly higher agglutination activity at high temperatures (38). Indeed, the microcalorimetry measurement demonstrates that the PA-IIL lectin also exhibits free energy values that are almost independent of temperature, albeit with very different contributions of enthalpy and entropy.

PA-IIL, RS-IIL, and CV-IIL represent three members of a fascinating family of calcium binding lectins. As further bacterial genomes are sequenced, more instances of this family of proteins may be uncovered and lead to an improved understanding of their causative role in infection and disease.

ACKNOWLEDGMENT

The help of Laurianne Rouet is acknowledged for ELLA experiments. We thank the ESRF, Grenoble, for access to synchrotron data collection facilities.

REFERENCES

- 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.
- Lecoq de Boisbaudran, M. (1882) Matière colorante se formant dans la colle de farine, *C.R. Acad. Sci.* **94**, 562.
- Duran, N., and Menck, C. F. (2001) *Chromobacterium violaceum*: A review of pharmacological and industrial perspectives, *Crit. Rev. Microbiol.* **27**, 201–222.
- Stephens, C. (2004) Microbial Genomics: Tropical Treasure? *Curr. Biol.* **14**, R65–R66.
- Brazilian National Genome Project Consortium (2003) The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11660–11665.
- Richard, C. (1993) *Chromobacterium violaceum*, opportunist pathogenic bacteria in tropical and subtropical regions, *Bull. Soc. Pathol. Exot. Ses Fil.* **86**, 169–173.
- Fantinatti-Garbogini, F., Almeida, R., Portillo, V. A., Barbosa, T. A., Trevilato, P. B., Neto, C. E., Coelho, R. D., Silva, D. W., Bartoletti, L. A., Hanna, E. S., Brocchi, M., and Manfio, G. P. (2004) Drug resistance in *Chromobacterium violaceum*, *Genet. Mol. Res.* **31**, 134–147.
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., Billault, A., Brottier, P., Camus, J. C., Cattolico, L., Chandler, M., Choisne, N., Claudel-Renard, C., Cunac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, W., Schiex, T., Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., and Boucher, C. A. (2002) Genome sequence of the plant pathogen *Ralstonia solanacearum*, *Nature* **415**, 497–502.
- Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbroek-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E., Lory, S., and Olson, M. V. (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen, *Nature* **406**, 959–964.
- Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C., Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., Davies, R. M., Davis, P., Devlin, K., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Moule, S., Mungall, K., Quail, M. A., Rajandream, M. A., Rutherford, K. M., Simmonds, M., Skelton, J., Whitehead, S., Spratt, B. G., and Barrell, B. G. (2000) Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491, *Nature* **404**, 502–506.
- Gilboa-Garber, N. (1982) *Pseudomonas aeruginosa* lectins, *Methods Enzymol.* **83**, 378–385.
- Sudakevitz, D., Imberty, A., and Gilboa-Garber, N. (2002) Production, properties and specificity of a new bacterial L-fucose- and D-arabinose-binding lectin of the plant aggressive pathogen *Ralstonia solanacearum* and its comparison to related plant and microbial lectins, *J. Biochem.* **132**, 353–358.
- Imberty, A., Wimmerova, M., Mitchell, E. P., and Gilboa-Garber, N. (2004) Structures of the lectins from *Pseudomonas aeruginosa*: Insights into molecular basis for host glycan recognition, *Microb. Infect.* **6**, 222–229.
- Perret, S., Sabin, C., Dumon, C., Pokorná, M., Gautier, C., Galanina, O., Ilia, S., Bovin, N., Nicaise, M., Desmadril, M., Gilboa-Garber, N., Wimmerova, M., Mitchell, E. P., and Imberty, A. (2005) Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of *Pseudomonas aeruginosa*, *Biochem. J.* **389**, 325–332.
- 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.
- 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.
- 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.
- Mitchell, E. P., Sabin, S., Šnajdrová, L., Pokorná, 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-IIL: 1.0 Å resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches, *Proteins: Struct., Funct., Bioinf.* **58**, 735–748.
- 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.
- Sudakevitz, D., Kostlanova, N., Blatman-Jan, G., Mitchell, E. P., Lerer, B., Wimmerova, M., Katcoff, F. D. J., Imberty, A., and Gilboa-Garber, N. (2004) A new *Ralstonia solanacearum* high affinity mannose-binding lectin RS-IIL structurally resembling the *Pseudomonas aeruginosa* fucose-specific lectin PA-IIL, *Mol. Microbiol.* **52**, 691–700.
- Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data, *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* **26**.
- Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography, *Acta Crystallogr.*, **760–763**.
- Cowtan, K. D., and Zhang, K. Y. (1999) Density modification for macromolecular phase improvement, *Prog. Biophys. Mol. Biol.* **72**, 245–270.
- Vagin, A., and Teplyakov, A. (1997) MOLREP: An automated program for molecular replacement, *J. Appl. Crystallogr.* **30**, 1022–1025.
- Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol.* **6**, 458–463.
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr. D53*, 240–255.
- 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.
- Marchessault, R. H., and Pérez, S. (1979) Conformations of the hydroxymethyl group in crystalline aldohexopyranoses, *Biopolymers* **18**, 2369–2374.
- Nori, H., Nishida, Y., Ohruji, H., and Meguro, H. (1990) Conformational analysis of hydroxymethyl group of D-mannose derivatives using (6S)- and (6R)-(6-²H)-D-mannose, *J. Carbohydr. Chem.* **9**, 601–618.
- 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.
- Dam, T. K., and Brewer, C. F. (2002) Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry, *Chem. Rev.* **102**, 387–429.
- Sabin, C., Mitchell, E. P., Pokorná, M., Gautier, C., Utile, J.-P., Wimmerová, M., and Imberty, A. (2006) Binding of different monosaccharides by lectin PA-IIL from *Pseudomonas aeruginosa*:

- Thermodynamics data correlated with X-ray structures, *FEBS Lett.* **580**, 982–987.
33. Ambrosi, M., Cameron, N. R., and Davis, B. G. (2005) Lectins: Tools for the molecular understanding of the glycocode, *Org. Biomol. Chem.* **3**, 1593–1608.
34. Chervenak, M. C., and Toone, E. J. (1995) Calorimetric analysis of the binding of lectins with overlapping carbohydrate-binding ligand specificities, *Biochemistry* **34**, 5685–5695.
35. Murphy, K. P., Xie, D., Thompson, K. S., Amzel, L. M., and Freire, E. (1994) Entropy in biological binding processes: Estimation of translational entropy loss, *Proteins* **18**, 63–67.
36. Carver, J. P. (1993) Oligosaccharides: How flexible molecules can act as signals, *Pure Appl. Chem.* **65**, 763–770.
37. 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.
38. Gilboa-Garber, N., and Sudakevitz, D. (1999) The hemagglutinating activities of *Pseudomonas aeruginosa* lectins PA-IL and PA-IIL exhibit opposite temperature profiles due to different receptor types, *FEMS Immunol. Med. Microbiol.* **25**, 365–369.
39. Kraulis, P. (1991) Molscript: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr.* **24**, 946–950.
40. Merrit, E. A., and Murphy, M. E. (1994) Raster3D version 2.0. A program for photorealistic molecular graphics, *Acta Crystallogr. D50*, 869–873.
41. Waldherr-Teschner, M., Goetze, T., Heiden, W., Knoblauch, M., Vollhardt, H., and Brickmann, J. (1992) in *Advances in Scientific Visualization* (Post, F. H., and Hin, A. J. S., Eds.) pp 58–67, Springer, Heidelberg, Germany.

BI060214E

Chapitre 8

La lectine fongique PVL de *Psathyrella velutina*

8.1 Introduction

La lectine PVL extraite du champignon sauvage *Psathyrella velutina* (Fig. 8.1) reconnaît spécifiquement les glycannes se terminant par un résidu GlcNAc ou un acide neuraminique (Kochibe and Matta 1989; Ueda 1999). De par cette double spécificité, cette lectine est considérée comme un outil intéressant pour le fractionnement des oligosaccharides et pour la détection des modifications de la glycosylation des glycoprotéines (Tsuchiya 1993; Moore 2005). De plus, il a été montré que PVL peut interagir avec des polysaccharides chargés négativement, tels que les pectines et l'héparine (Ueda 1999).

Les autres caractéristiques intéressantes de la PVL sont le poids moléculaire de la sous-unité qui est relativement important pour une lectine (43 kD) et une séquence qui ne montre aucune similarité avec d'autres protéines de structure ou de fonction connues. Dans ce huitième chapitre on présente la structure tridimensionnelle de la PVL en complexe avec ses ligands et l'étude de l'affinité parmi la calorimétrie de titration.



Figure 8.1 Spécimens de *Psathyrella velutina*

8.2 Résultats

La lectine a été purifiée d'un extrait de champignon par chromatographie d'affinité et co-cristallisée avec un dérivé sélénisé, le β -Me-SeGlcNAc. Sa structure tridimensionnelle, résolue avec la méthode SAD, montre que le monomère de 401 résidus adopte un repliement du type β -*propeller* à sept lames qui n'avait jamais été observé dans une lectine auparavant. Les complexes avec la GlcNAc et le NeuNAc montrent la présence de six sites de liaison, localisés dans l'espace entre deux lames consécutives. En raison des répétitions dans la séquence, les six sites sont très similaires mais pas identiques. Le mode de reconnaissance du ligand rappelle celui de la Tachylectin-2, une lectine animale qui montre aussi un repliement du type β -*propeller*, mais à cinq lames (Beisel 1999). La structure de la PVL montre aussi la présence de domaines acides répétés qui servent de boucles de fixation pour le calcium. Cette caractéristique rappelle une importante classe de β -*propellers* à sept lames qui fixent le calcium, les intégrines (Hynes 2002). La recherche de séquences homologues a mené à l'identification de certaines séquences bactériennes ayant une similarité remarquable et dont la fonction n'est pas connue mais qui sont annotées comme toxines RTX, intégrines ou protéines liant le calcium. Les mesures par microcalorimétrie de titration ont montré une affinité moyenne pour le GlcNAc. En conclusion, la multivalence élevée de cette protéine, sa multispécificité et sa similarité avec les intégrines confortent les rôles importants des β -*propellers* dans les interactions de reconnaissance moléculaire.

β-Propeller Crystal Structure of *Psathyrella velutina* Lectin: An Integrin-like Fungal Protein Interacting with Monosaccharides and Calcium

**Gianluca Cioci¹, Edward P. Mitchell², Valerie Chazalet¹, Henri Debray³
Stefan Oscarson⁴, Martina Lahmann⁴, Catherine Gautier¹
Christelle Breton¹, Serge Perez¹ and Anne Imbert^{1*}**

¹Centre de Recherches sur les Macromolécules Végétales (affiliated with Université Joseph Fourier), CNRS, F-38041 Grenoble, France

²E.S.R.F. Experiments Division F-38043 Grenoble, France

³Unité de Glycobiologie Structurale et Fonctionnelle Université des Sciences et Technologies de Lille, 59655 Villeneuve d'Ascq, France

⁴Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-10691 Stockholm, Sweden

The lectin from the mushroom *Psathyrella velutina* recognises specifically *N*-acetylglucosamine and *N*-acetylneurameric acid containing glycans. The crystal structure of the 401 amino acid residue lectin shows that it adopts a very regular seven-bladed β-propeller fold with the N-terminal region tucked into the central cavity around the pseudo 7-fold axis. In the complex with *N*-acetylglucosamine, six monosaccharides are bound in pockets located between two consecutive propeller blades. Due to the repeats shown by the sequence the binding sites are very similar. Five hydrogen bonds between the protein and the sugar hydroxyl and *N*-acetyl groups stabilize the complex, together with the hydrophobic interactions with a conserved tyrosine and histidine. The complex with *N*-acetylneurameric acid shows molecular mimicry with the same hydrogen bond network, but with different orientations of the carbohydrate ring in the binding site. The β-hairpin loops connecting the two inner β-strands of each blade are metal binding sites and two to three calcium ions were located in the structure. The multispecificity and high multivalency of this mushroom lectin, combined with its similarity to the extracellular domain of an important class of cell adhesion molecules, integrins, are another example of the outstanding success of β-propeller structures as molecular binding machines in nature.

© 2006 Elsevier Ltd. All rights reserved.

*Corresponding author

Keywords: lectin; carbohydrate; β-propeller; crystallography; microcalorimetry

Introduction

Lectins are defined as ubiquitous proteins of non-immune origin that mediate carbohydrate recognition events.¹ The existence of mushroom lectins and their potential link to toxicity have been known for several decades.² Several different roles have been proposed for mushroom lectins with some related to intrinsic mushroom metabolism, such as dormancy, growth and morphogenesis stages. Alternatively, and since mushrooms depend on a symbiotic or parasitic association with a host or

depend on recognition of organic matter (saprophytism), lectins could operate as putative actors in host or substrate recognition and adhesion processes. However, only a few cases have been reported of the direct involvement of a lectin in symbiotic or parasitic recognition.³ Indeed, and by analogy with plant lectins, mushroom lectins probably are able to play different roles in different circumstances, none of which need to be mutually exclusive.²

Agglutination studies of human and rabbit erythrocytes conducted on more than 400 collected mushroom species demonstrated the presence of lectins in about half of them.⁴ A large number of mushroom lectins have now been sequenced and characterised^{2,5} and interest arose when a clear similarity was established between fungal lectins and human galectins and immunoglobulins. In addition, fungal lectins are

Abbreviations used: PVL, *Psathyrella velutina* lectin; AAL, *Aleuria aurantia* lectin; GlcNAc, *N*-acetylglucosamine; Neu5Ac, *N*-acetylneurameric acid; ITC, isothermal titration calorimetry.

E-mail address of the corresponding author:
imbert@cermav.cnrs.fr

valuable tools in glycobiology, providing useful probes for carbohydrate epitopes in taxonomical, embryological and bacteriological studies.^{2,3,5} Despite numerous reports on successful crystallisations, only a limited number of mushroom lectin crystal structures have been described⁶. *Aleuria aurantia* fucose-binding lectin (AAL) was found to be a six-bladed β -propeller;⁷ Fve from *Flammulina velutipes* displays structural similarity to human fibronectin;⁸ XCL from *Xerocomus chrysenteron*⁹ and ABL from *Agaricus bisporus*¹⁰ both resemble actinoporins, a family of pore-forming toxins; CGL2 from *Coprineus cinerea*¹¹ and ACG from *Agrocybe cylindracea*¹² belong to the galectin family and LSL, the lectin from *Laetiporus sulphureus*¹³ adopts a trefoil-based fold observed in many lectins and carbohydrate-binding domains.

A lectin from the wild mushroom *Psathyrella velutina* (PVL) was first purified by Kochibe & Matta as a single polypeptide of 40 kDa and its specificity for *N*-acetylglucosamine (GlcNAc) was characterised.¹⁴ The lectin has been purified from the fruiting bodies of the mushroom and an equivalent lectin has been isolated from the mycelium. Using equilibrium dialysis experiments PVL was estimated to have four binding sites per polypeptide with an affinity constant for the monosaccharide of $6.4 \times 10^3 \text{ M}^{-1}$. PVL was shown to have a distinct binding specificity compared to classical GlcNAc specific lectins since it does not bind preferentially to chitooligosaccharide sequences (GlcNAc β 1-4GlcNAc) but rather to other non-reducing GlcNAc residues such as GlcNAc β 1-2Man present on the antenna of degalactosylated *N*-glycoproteins.¹⁵ Such changes of glycosylation at terminal glycan structures are frequently associated with chronic inflammatory diseases of autoimmune character. For example IgGs of rheumatoid arthritis patients display altered glycosylation with an accumulation of isoforms carrying an N-terminal GlcNAc.¹⁶ PVL has been used to detect such IgGs either in ELISA-based assays¹⁷ or with affinity biosensor techniques.¹⁸ Chronic microbial and viral infections are also associated with deficiency of galactose on IgG *N*-glycans as recently demonstrated in the serum of AIDS patients with use of PVL.¹⁹

Later, it was shown that *N*-acetylneurameric acid (Neu5Ac)²⁰ binds to PVL in the same binding site as GlcNAc.²⁰ The lectin binds to sialylated glycoproteins such as fetuin with very high affinity, about one order of magnitude higher than those measured with *Sambucus nigra* and *Maackia amurensis* lectins.²¹ In addition, Ueda and collaborators have shown that PVL can also bind to negatively charged polysaccharides such as heparin and pectin fragments.²² Interestingly, the polysaccharide binding site appears to be independent from the GlcNAc/Neu5Ac binding site.

In the present study, PVL was purified from the fruiting bodies of *P. velutina* by affinity

chromatography. The X-ray crystallographic structure has been determined by the single anomalous dispersion (SAD) method using a selenium derivative of *N*-acetylglucosamine co-crystallised with the lectin. The crystal structures of the lectin in its native state (to 1.5 Å resolution) and complexed with GlcNAc (to 2.6 Å) and Neu5Ac (to 1.8 Å) are described in detail. The present study elucidates the structural basis for both the dual specificity and the differences in affinity of PVL towards its ligands. Searching for sequence similarity indicates the presence of PVL-like sequences in the newly available genomes of another fungus and of several bacteria. Moreover, the PVL structure shows structural similarity with the β -propeller subunit of integrins with a surprising convergence in their calcium binding properties.

Results

Gene and amino acid sequences of PVL

The *pvl* gene has previously been cloned by a Japanese biotechnology company (Seikagaku Corporation)²³ and the nucleotide sequence of 1256 base-pairs is available under the EMBL accession number E16441. The translated amino acid sequence corresponds to 402 amino acid residues including the initiating methionine residue, with seven tandem repeats of approximately 55 amino acid residues each. Upon solving our first crystal structure of PVL, some clear discrepancies were observed between the Japanese amino acid sequence and the experimental electron density, which could not be unambiguously resolved. Therefore the *pvl* gene sequence was determined for the *P. velutina* mushrooms found locally in the Alps and from which pure lectin for crystallisation and isothermal titration calorimetry (ITC) studies was obtained.

Our cloning experiments resulted in the identification of at least four isoforms of the gene from a single fruiting body that are probably allelic variants. The translated amino acid sequences show a very high identity of 99% between them but a lower identity of 92% to the sequence deposited by the Japanese group. The amino acid sequence shown in Figure 1(a) corresponds to the translated sequence of the cDNA clone that matches best to the electron density of the crystal structure. Since mismatches were observed between the primer sequence (pr-pvl) used for cloning and the crystal structure, the first six amino acids were taken from the crystal structure as they could be determined unambiguously from the electron density map. The nucleotide sequence has been deposited in GenBankTM (accession number DQ232759) as a partial cDNA sequence starting at amino acid 7.

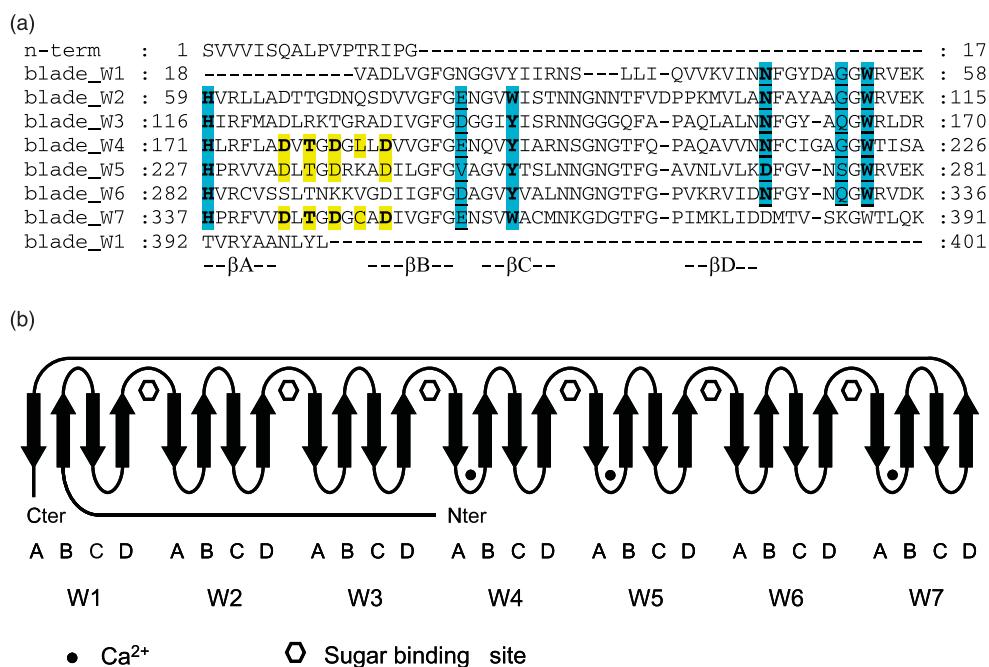


Figure 1. (a) Amino acid sequence of PVL with amino acids involved in sugar and calcium binding highlighted in blue and yellow, respectively. Bold characters indicate binding by side-chain. (b) Schematic description of the PVL fold with labelling of β -strands and blades (W motif).

Purification of PVL

Two different methods have been reported in the literature for the purification of this lectin. Kochibe & Matta first used a multi-step procedure involving four consecutive chitin and ion-exchange columns.¹⁴ An improved method involves a single affinity chromatography step on Sepharose-6B coupled to *N*-acetylchitooligosaccharides (GlcNAc₅₋₆) by a reductive amination.²⁰ In our procedure we used a direct coupling reaction of *N*-acetylglucosamine monomer to epoxy Sepharose-6B. The kinetics for the coupling reaction between the different sugar hydroxyls and epoxy groups is not known but we can expect the primary O6 hydroxyl to react much faster than the sterically hindered O3 and O4 hydroxyl groups. Although the resulting adsorbent must be considered not strictly “monospecific” for the lectin, the remarkable high affinity of PVL for terminal GlcNAc residues bearing a substituent on the C6 position was strong enough to effectively retain the protein. This procedure has the advantages of being both cheaper and faster, resulting in a yield of 30 mg of purified lectin from 50 g of mushrooms.

Affinity studies by ITC microcalorimetry

In order to further characterise the interactions between PVL and monosaccharides, the K_a values and the thermodynamic binding parameters were determined using titration microcalorimetry, a method that is well suited for the characterisation of protein–carbohydrate interactions.²⁴ A typical titration curve for PVL binding to GlcNAc is shown

in Figure 2 and the results from different experiments are listed in Table 1. Using a mathematical model corresponding to a simple binding mode, a stoichiometry of four GlcNAc per PVL monomer is observed together with a modest affinity constant (K_d value of 2×10^{-4} M). In addition to confirming the binding data obtained previously,¹⁴ the present study allows the thermodynamic basis of the affinity between PVL and GlcNAc to be determined. The interaction is strongly driven by enthalpy with a ΔH value of $-55(\pm 3)$ kJ/mol that is counterbalanced by a strong unfavourable entropy contribution ($T\Delta S$ value of $-33(\pm 3)$ kJ/mol). An entropy barrier is often observed for the binding of carbohydrates to lectins²⁴ but is specially strong in the present case.

No significant heat of binding could be detected for Neu5Ac in the same experimental conditions. Since it is not likely that binding of this monosaccharide would be completely entropy-driven, it may be concluded that the affinity of PVL for Neu5Ac is at least a hundred times weaker than the affinity for GlcNAc. PVL was reported to bind sialyl-glycoconjugates and the present results therefore indicate that it is necessary to have more than one Neu5Ac monosaccharide to obtain detectable binding.

Crystal structures

Phasing

The use of a selenium labelled sugar for the phasing of carbohydrate binding proteins was proposed and used for the first time by Loris

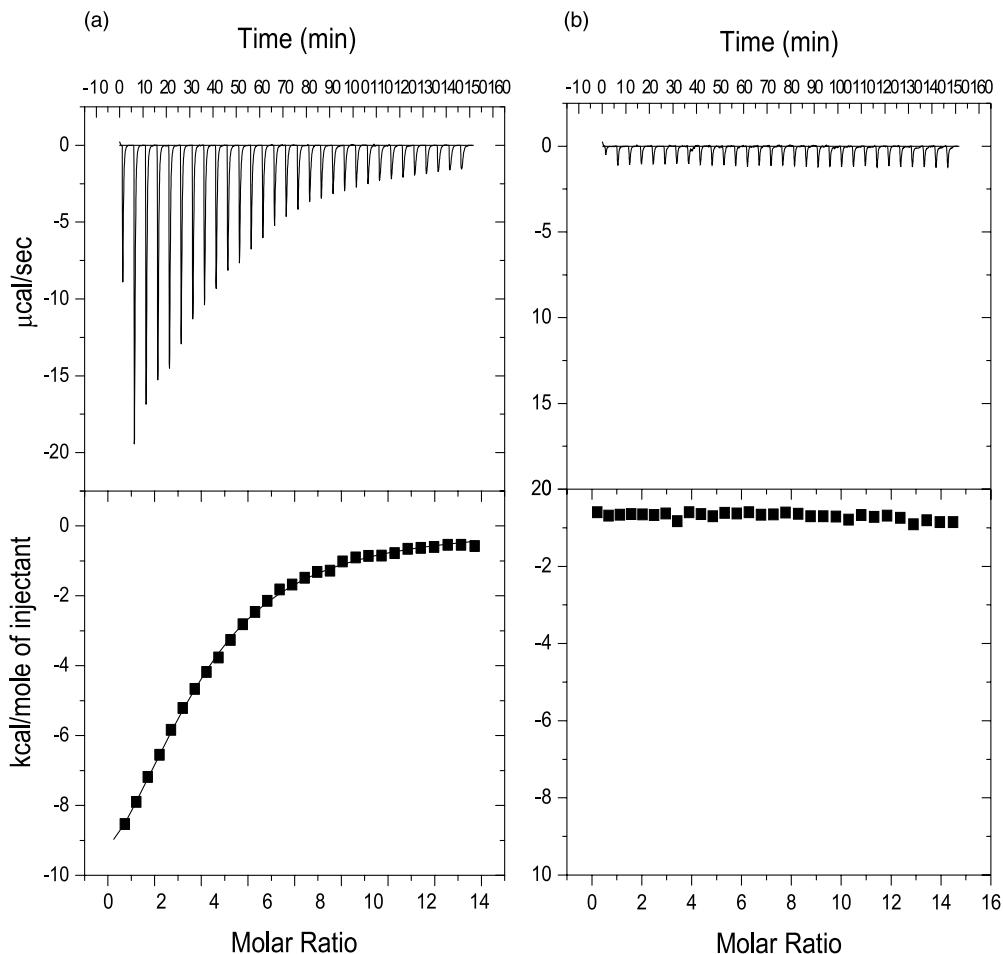


Figure 2. Isothermal titration calorimetry results for GlcNAc (a) and Neu5Ac (b) monosaccharides (7 mM) binding to PVL (110 μ M) in PBS buffer (pH 7.2) at 25 °C. Top, data from 30 automatic injections of 10 μ l sugar each into the PVL-containing cell. Lower, plot of the total heat released as a function of ligand concentration for the titration shown above (squares). The continuous line represents the best least-square fit for the obtained data.

and collaborators for the structure solution of the F17-G fibrial adhesin.²⁵ Depending on the availability of the saccharide substrate, this is the method of choice for phasing lectin structures as generally the seleno-sugar binds very well and relatively low concentrations of sugar (or a short soak) is all that is required to provide sufficient signal for MAD or SAD methods to work admirably. PVL was co-crystallised with methyl-seleno-GlcNAc to give crystals of space group $P2_12_12$. The SAD data, collected at the peak of the selenium K-absorption edge, resulted in a high average figure of merit of 0.471 and phasing power of 1.533 over all data to 1.8 Å resolution with five Se-GlcNAc residues found bound to each PVL monomer. The refinement of the PVL/Se-GlcNAc structure has been performed but the structure will

not be discussed in detail here, since it has less biological relevance than the complexes with natural monosaccharides.

Native structure of PVL

The native protein crystallised in space group $P2_1$ with cell dimensions of $a=51.83$ Å, $b=142.22$ Å, $c=52.16$ Å and $\beta=92.1^\circ$ and two chains of 401 residues arranged around a pseudo 2-fold axis of symmetry in the asymmetric unit. The PVL monomer structure is a single domain showing a very regular seven-bladed β -propeller fold (Figure 3). No disulphide bridges were located in the structure and an inspection of the electron density around putative glycosylation sites showed no indication of carbohydrate units covalently

Table 1. Thermodynamics of binding for PVL with GlcNAc at 298 K

	K_a (10^3 M $^{-1}$)	K_d (10^{-6} M)	n	$-\Delta G$ (kJ/mol)	$-\Delta H$ (kJ/mol)	$T\Delta S$ (kJ/mol)
Exp1	4.7 (± 0.2)	214	4.32 (± 0.06)	20.0	52.5 (± 0.9)	-31.2
Exp2	5.3 (± 0.2)	189	3.72 (± 0.06)	21.2	57.5 (± 1.2)	-36.2

Values between parentheses correspond to uncertainties of the fitting procedure.

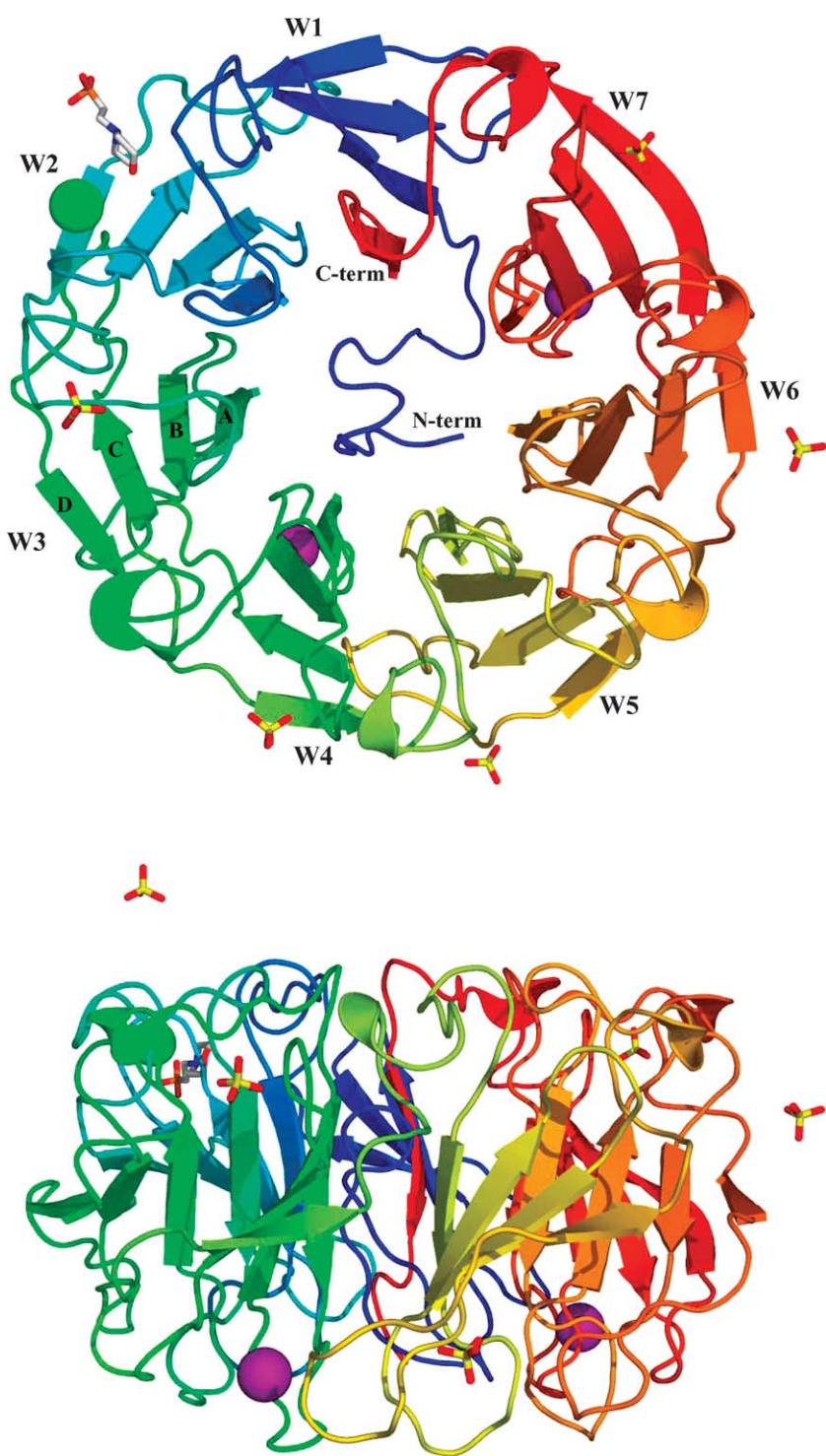


Figure 3. Two orthogonal views of the crystal structure of native PVL. Calcium and chloride ions are represented by magenta and green spheres, respectively. Sulphate ions and 2-N-morpholino-ethane-sulfonic acid (Mes) molecules are represented by capped sticks. Plots are made using Pymol (DeLano Scientific LLC), unless otherwise indicated.

linked to the protein. Each blade (also called W-motif; see Figure 1(b)) is made of four antiparallel β -strands (A-B-C-D) connected by loops. In β -propellers, strand A delimits the central cavity containing the pseudo 7-fold symmetry axis. Strands B and C, the most hydrophobic, are completely buried while strand D, which delimits the edge of the blade, shows amphiphilic properties. The superposition of the seven blades gave a consensus sequence repeat with an overall rmsd of 0.8 Å

(Figure 4). The overall shape of the monomer is a fat torus with a diameter of 52 Å and a height of 31 Å. The closure of the β -propeller fold occurs via a “velcro” interaction between strand B of blade 1 (residues 20–25) and strand A of blade 7 (residues 393–397, close to the C terminus). The N terminus region is located tucked into the central cavity. The contacts between N terminus peptide and amino acid residues nearby are established mainly by hydrophobic contacts and hydrogen bonds

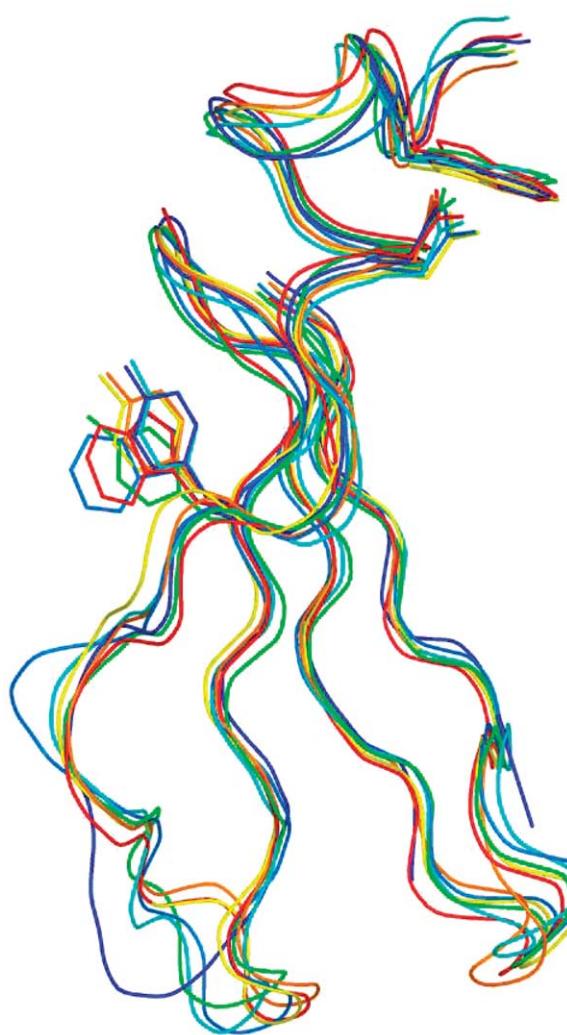


Figure 4. Superimposition of the seven blades of native PVL structure colour-coded as in Figure 3. The side-chains of the aromatic residues interacting with the sugar are shown.

mediated by well ordered solvent molecules. As frequently observed in β -propeller structures,²⁶ the contacts between adjacent blades are mostly achieved by hydrophobic interactions involving phenylalanine residues that are well conserved between the tandem repeats.

Complex of PVL and N-acetylglucosamine

Crystals of the complex of PVL/GlcNAc were obtained in space group $P3_121$ ($a=b=106.8\text{ \AA}$, $c=86.6\text{ \AA}$) with one monomer per asymmetric unit. Electron density for six GlcNAc residues was clearly visible in the 2.6 \AA resolution crystal structure (Figure 5). The sugar binding sites are pockets located in the upper part of the β -propeller, in a space between two consecutive blades, accessible to the solvent (Figure 6). Only the pocket between blade 7 and blade 1 is not occupied. Due to the regular repeats in the protein sequence, the binding sites are very similar, although not identical. The six monosaccharides adopt the same conformation and no major differences were found in the binding mode.

The loop connecting strand D with strand A of two consecutive propeller blades makes the principle interactions with the sugar, in combination with the loop connecting strands B and C. The side-chain of one asparagine (or aspartate) residue forms two hydrogen bonds with the O3 and O4 sugar hydroxyls. The O3 hydroxyl also donates a hydrogen bond to a conserved tryptophan residue. The N-acetyl group of the sugar is involved in two hydrogen bonds with the protein main chain: the nitrogen atom receives from an oxygen while the carbonyl oxygen donates to a nitrogen (Figure 7). One water molecule, conserved amongst most of the sites, takes part in the hydrogen bonding network, bridging the acetyl oxygen to the protein main chain. The hydrogen bonding network for the ligands is summarised in Table 2. The O1 and O6 hydroxyls are not involved in any hydrogen bond with the lectin binding site. In particular, the anomer carbon C1 is solvent-exposed and available for substitution, allowing disaccharides such as the GlcNAc β 1-2Man to bind effectively to the protein. The O6 hydroxyl is also solvent-exposed and is virtually able to carry any substituent, which is in agreement with the previous observations on PVL binding specificity.^{14,15}

Hydrophobic contacts have a strong contribution for the binding of GlcNAc in PVL. The methyl extremity of the N-acetyl group fits in a hydrophobic pocket formed by the conserved tryptophan residue that is involved in hydrogen bonding to O3 and by a histidine that is located

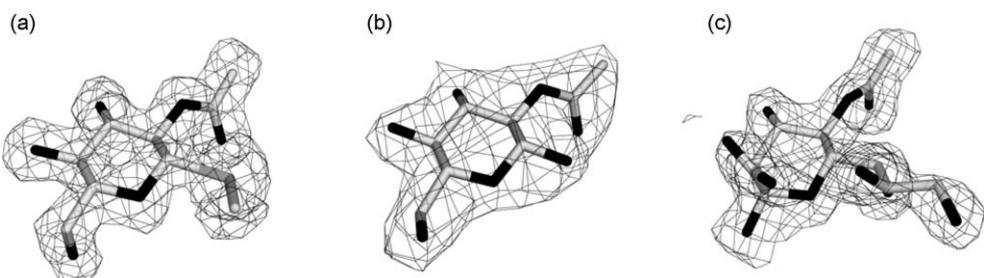


Figure 5. Final sigma $2F_o - F_c$ electron density map (contoured at 1.0σ) around the monosaccharides in crystal structures of PVL complexed with Me-Se-GlcNAc (a), GlcNAc (b) and Neu5Ac (c).

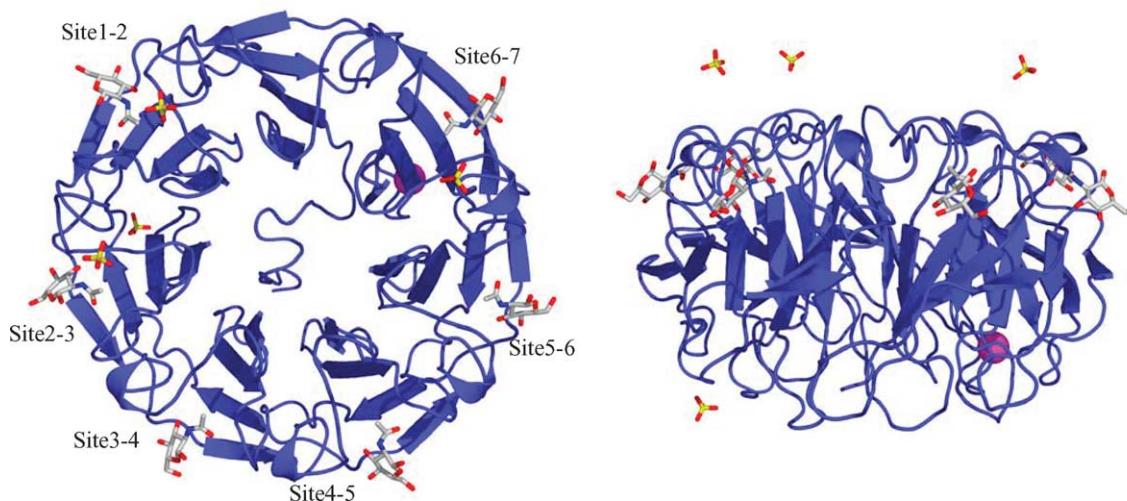


Figure 6. Two orthogonal views of the crystal structure of PVL complexed with GlcNAc. Calcium ions are represented by magenta spheres, sulphate ions and GlcNAc ligands by capped sticks.

on the same connecting loop, at the base of strand A. It is interesting to note that no electron density for the sugar has been located in the binding site 7-1, where this histidine is replaced by a

threonine. Another aromatic residue, a tryptophan or a tyrosine in the middle of strand C, creates a stacking interaction with the hydrophobic face of the GlcNAc residue.

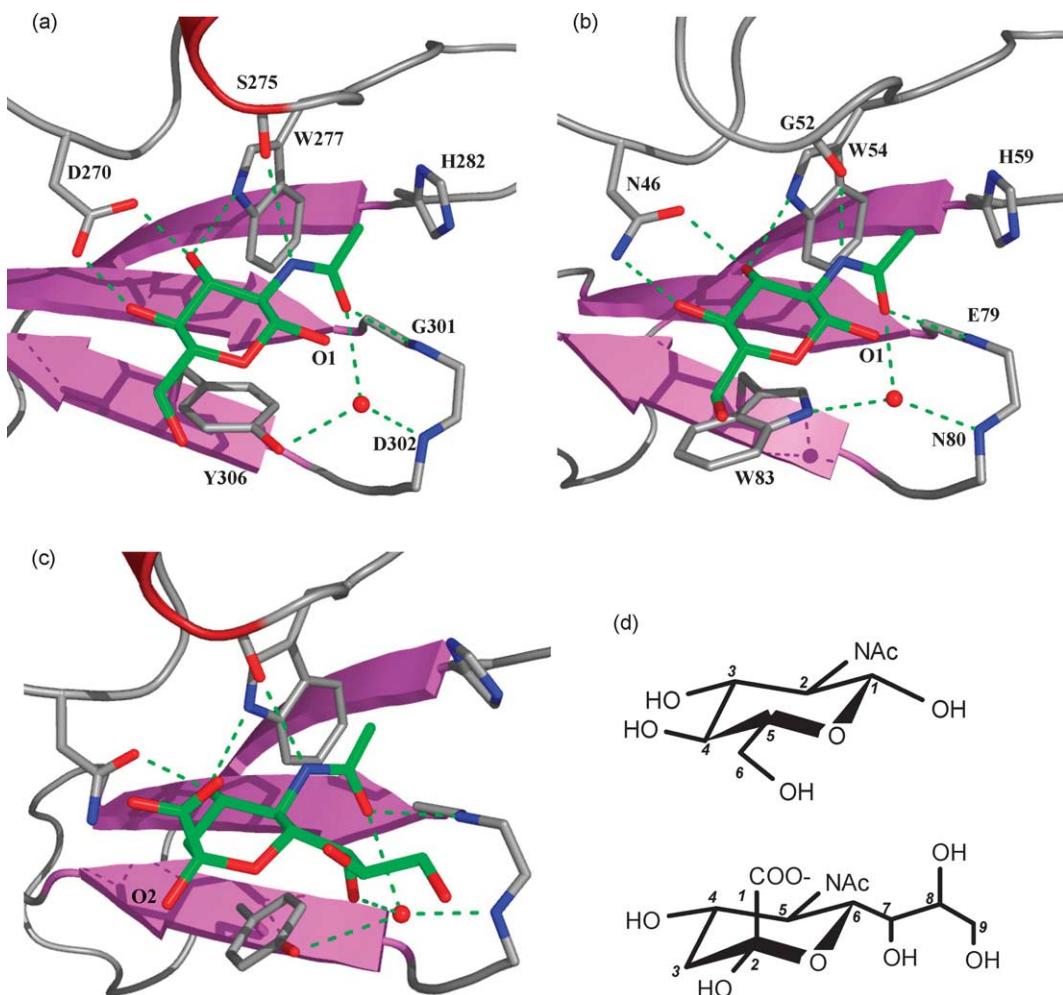


Figure 7. Binding mode for GlcNAc in PVL for site 5-6 (a) and site 1-2 (b) and binding mode for Neu5Ac in PVL for site 3-4 of chain B. (c) Hydrogen bonds are represented as green dotted lines. (d) Schematic representation of GlcNAc and Neu5Ac together with numbering of the carbon atoms.

Table 2. Contacts (distances in Å) between *N*-acetylglucosamine and PVL

	Site 1-2	Site 2-3		Site 3-4		Site 4-5		Site 5-6		Site 6-7		
<i>Hydrogen bonds</i>												
O1	Asp169 O ^{δ2a}	2.6										
N2	Gly52 O	2.7	Gly109 O	2.9	Gln164 O	3.0	Gly220 O	2.8	Ser275 O	3.4	Gln330 O	2.8
O=C	Glu79 N	3.0	Asp136 N	3.0	Glu191 N	2.8	Val247 N	2.8	Asp302 N	3.1	Glu357 N	3.1
	Wat1 ^b	2.6					Wat1 ^b	2.6	Wat1 ^b	3.0		
O3	Asn46 O ^{δ1}	2.6	Asn103 O ^{δ1}	2.7	Asn159 O ^{δ1}	2.6	Asn214 O ^{δ1}	2.4	Asp270 O ^{δ1}	2.3	Asn325 O ^{δ1}	2.7
	Trp54 N ^{ε1}	2.9	Trp111 N ^{ε1}	3.0	Trp166 N ^{ε1}	2.9	Trp222 N ^{ε1}	2.8	Trp277 N ^{ε1}	3.0	Trp332 N ^{ε1}	2.8
O4	Asn46 N ^{δ2}	3.0	Asn103 N ^{δ2}	3.1	Asn159 N ^{δ2}	2.9	Asn214 N ^{δ2}	2.8	Asp270 N ^{δ2}	2.6	Asn325 N ^{δ2}	2.6
O6							Leu266 O ^a	2.6				
							Leu266 N ^a	3.1				
<i>Hydrophobic contacts</i>												
CH3	His59		His116		His171		His227		His282		His337	
H2/H4	Trp83		Tyr140		Tyr195		Tyr251		Tyr306		Tyr306	

^a Belongs to a symmetry-related molecule.^b Bridging to two protein atoms.**Complex of PVL and *N*-acetylneuraminic acid**

Neu5Ac has been co-crystallised with PVL and the resulting complex retains the native *P*2₁ space group with two protein monomers per asymmetric unit. However, since the electron density for the Neu5Ac bound in the co-crystallised samples was ill-defined, one crystal was further soaked in a Neu5Ac solution of higher concentration. Three ligands were found in site 2-3 (between blades 2 and 3), site 3-4 and site 4-5 of the first monomer and two in site 3-4 and site 6-7 of the second monomer. The sialic acid has strong similarity with the binding scheme shown by GlcNAc (**Figure 7**) with identical location of the *N*-acetyl group oriented towards the hydrophobic pockets created by conserved tryptophan and histidine residues. The different orientation of the sugar ring in the binding site makes the carbon atoms C3, C4 and C5 of Neu5Ac superimpose to the carbon atoms C4, C3 and C2 of GlcNAc, providing the structural basis for Neu5Ac recognition. The O4 is involved in two hydrogen bonds with conserved asparagine and tryptophan residues (**Table 3**). The acidic group does not make any contact with the protein surface except in site 4-5 of chain A, where the orientation of the ligand is slightly tilted. The oxygen O2,

carried by the anomeric center C2 is solvent-exposed and could be linked to an additional monomer in either 2-3 or 2-6 geometry. In three sites, the hydroxyl O7 of the three carbon extracyclic chain makes an additional hydrogen bond with the water molecule conserved between the GlcNAc and Neu5Ac structures, which therefore establishes two hydrogen bonds with the protein and two with the sugar. In analogy with the GlcNAc complex, hydrophobic contacts are established by the methyl group of *N*-acetyl and by the hydrophobic face of Neu5Ac.

***Ca*²⁺ binding sites**

In the structure of native PVL refined to 1.5 Å resolution, two metal ions could be located on the opposite face of the β-propeller to the sugar binding sites. The two ions were modelled as calcium according to their coordination geometry, their *B*-factors and by the fact that Ca²⁺ was present in the crystallisation buffer. The calcium ions are located in the β-hairpin loops connecting strands A and B, in blades 4 and 7. Some density could be located in the equivalent site of blade 5 but was not sufficient to justify full calcium ion occupancy.

Table 3. Contacts (distances in Å) between sialic acid and PVL

	Chain A				Chain B					
	Site 2-3		Site 3-4		Site 4-5		Site 3-4		Site 6-7	
<i>Hydrogen bonds</i>										
N5	Gly109 O	3.0	Gln164 O	3.1	Gly220 O	2.9	Gln164 O	3.4	Gln330 O	3.1
O=C	Asp136 N	2.8	Glu191 N	2.9	Val247 N	2.9	Glu191 N	2.8	Glu357 N	2.9
					Wat1 ^a	2.9	Wat1 ^a	3.0	Wat1 ^a	2.8
O4	Asn103 O ^{δ1}	2.8	Asn159 O ^{δ1}	2.8			Asn159 O ^{δ1}	2.5	Asn325 O ^{δ1}	2.7
	Trp111 N ^{ε1}	2.9	Trp166 N ^{ε1}	2.9	Trp222 N ^{ε1}	3.4	Trp166 N ^{ε1}	3.0	Trp332 N ^{ε1}	2.9
O7			Wat1 ^a	2.7			Wat1 ^a	2.5	Wat1 ^a	2.3
O1A					Asn214 N ^{δ2}	2.3			Asn325 N ^{δ2}	3.7
<i>Hydrophobic contacts</i>										
CH3	His116		His171		His227		His227		His337	
H3ax/H5	Tyr140		Tyr195		Tyr251		Tyr251		Tyr306	

^a Bridging to two protein atoms.

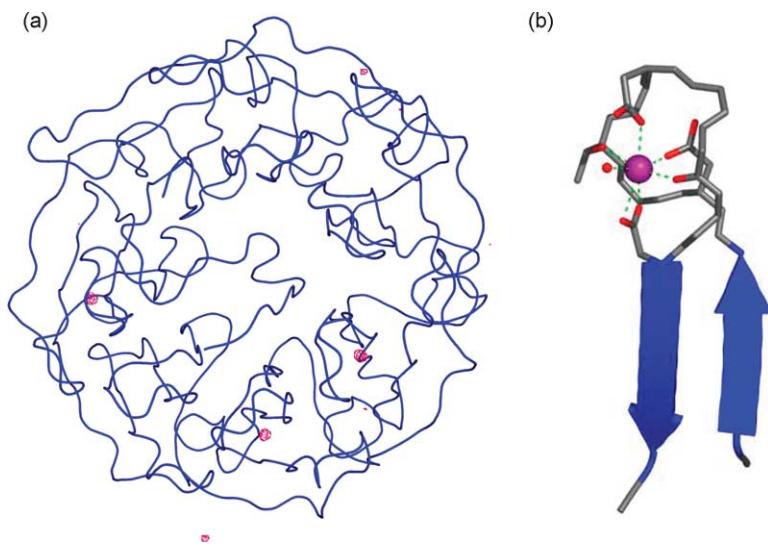


Figure 8. (a) Representation of PVL monomer together with the phased anomalous map contoured at 5σ showing the position of three calcium atoms. (b) Calcium binding site in blade 7.

In order to test for other calcium binding sites, a native crystal was soaked into a CaCl_2 containing solution and diffraction data collected (statistics not shown). Although calcium anomalous signal is very low using a wavelength of 0.934 \AA , six major peaks could be located in the resulting phased anomalous difference map, corresponding to calcium ions bound in the β -hairpin loops in the blade 4, 5 and 7 of each monomer in the asymmetric unit (Figure 8(a)). The binding sites for calcium ions are made up by a nine residue loop with the consensus sequence Asp-h-Thr-Gly-Asp-Gly-h-h-Asp, where h is a hydrophobic residue. The Ca^{2+} ions are coordinated by the side-chain oxygen of consensus sequence residues 1, 3, 5, 9 and the main-chain oxygen of residue 7, the last coordination position is occupied by a water molecule (Figure 8(b)).

Oligomerisation

Although native PVL and the complexes obtained by co-crystallisation resulted in three different space groups, one interface between monomers is conserved between the three systems. This interface is created by a 2-fold crystallographic symmetry axis for the complexes with GlcNAc (space group $P3_121$) and Se-GlcNAc (space group $P2_12_12$), and by the pseudo 2-fold symmetry axis between the two monomers in the asymmetric unit of the native lectin (space group $P2_1$). PVL was previously described as a monomer,¹⁴ and our own assays on exclusion chromatography column confirm that it is retained as a monomer at a concentration of 1 mg/ml using PBS (pH 7.2) (data not shown). However, PVL tends to aggregate irreversibly, forming an insoluble precipitate,^{14,22} and the dimerisation observed in all the crystal structures may therefore only occur in concentrated

solution. This feature may be of importance for the biological role of the lectin.

Discussion

Carbohydrate specificity of PVL

PVL was first described as a GlcNAc binding protein¹⁴ and the present study gives the structural and thermodynamic bases of this specificity. Different experimental approaches agree that, in solution, four GlcNAc residues bind to one PVL monomer, whereas six bound sugars are observed in the crystal structure of PVL/GlcNAc complex. The existence of four high affinity binding sites and two lower affinity ones can be correlated with the role of tyrosine (four sites) versus tryptophan (two sites) for establishing hydrophobic interactions with the GlcNAc ring. Indeed, similar observations were made for fucose binding lectin AAL from *Aleuria aurantia*, where five or three binding sites are occupied depending upon the crystallisation conditions.^{7,27} The special role of a tyrosine residue in the high affinity binding of some of the sites was later demonstrated through site-directed mutagenesis studies.²⁸

Later, it was reported that PVL could also bind sialic acid containing glyconjugates,²⁰ although no binding data have been published for the recognition of the monosaccharide. In our study, the binding of Neu5Ac is very weak (K_d lower than 10^{-3} M). In fact, the soaking of PVL crystals in a solution of Neu5Ac only resulted in a partial binding of the sugar, but it is very likely that extension of the ligand from position 2 would create further interactions with the protein thereby conferring Neu5Ac-containing oligosaccharides a much higher affinity for PVL than the monomer alone.

The dual-specificity of a lectin binding site for both GlcNAc and Neu5Ac has already been observed for the isolectins of wheat germ agglutinin. In particular, WGA2 has been cocrystallised with chitobiose,²⁹ with Neu5Ac³⁰ and with sialoglycopeptide from glycophorin A.³¹ As observed in PVL, the N-acetyl group of both ligands is buried in the same fashion in the protein binding site, while the orientation of the sugar ring is different.

PVL has been shown to bind also to negatively charged polysaccharides such as heparin and pectin.²² Our efforts to co-crystallise PVL with pectin fragments have not been successful to date. However, analysis of the structure demonstrates that the positively charged amino acids located at the end of the strands D and A connecting loop (e.g. Arg55 and Lys58 for blade 1 and corresponding residues on the other blades) do cluster on the top face of the propeller, thereby creating a strongly positive area in the electrostatic potential surface. In most of the crystal structures described here, several sulphate residues from the crystallisation buffer are observed interacting with these amino acid residues. Such an area should be very favourable for interaction with negatively charged polysaccharides and, as described previously,²² this interaction should not affect the GlcNAc binding sites.

The ability of PVL to bind GlcNAc or Neu5Ac and negatively charged polysaccharides has been related to the possible role of mushroom lectins in the saprophytic colonisation of humic soil derived from plant or wood decomposition. As already pointed out by Ueda and collaborators,²² PVL might be able to cross-link substrates such as pectins or polygalacturonic acids, both components of plant cell walls, with GlcNAc or sialic acid containing glycoconjugates present on bacterial cell surfaces. In this case the lectin may act as a sort of receptor able to "guide" the mycelium colonisation towards soil zones rich in organic materials. In addition the massive presence of the lectin in the mushroom could be explained by the need of the organism to mobilise or store the nutrients needed for the development of the fruiting body. It is also possible that multi-adhesive lectins like PVL could be involved in the mobilisation of the defence system, since these proteins have the ability to agglutinate bacteria and are therefore often involved in protection against infection.

β-Propellers and lectins

Since the discovery of the β-propeller fold in the crystal structure of influenza virus neuramidase,³² this fold has been found in a number of subsequent protein structures.²⁶ β-Propeller structures are classified according to the number of blades (also called W-motifs after the sequence of antiparallel strands forming a "W") and propellers showing from four to eight blades have been reported so far. The W-motif forms the structural scaffold of the β-propeller fold, which means that

different β-propellers could show significant structural similarity in the β-strand regions although no sequence similarity is observed.³³ Conversely, the most structurally variable regions, which are very often related to enzymatic or binding activity, are located in the loops between the β-strands, especially those connecting strands D-A and B-C, which in the case of PVL form the ligand binding site.

At the present time, only three lectins other than PVL are known to adopt a β-propeller fold (Figure 9). The lectin from the orange peel mushroom *A. aurantia* (AAL) is the first mushroom lectin whose structure was solved by X-ray crystallography. AAL is specific for fucose and it has been shown to adopt a six-bladed β-propeller fold.⁷ The lectin was co-crystallised giving either three or five bound fucose residues and suggesting that even if all the binding sites are very similar, they do not have the same affinity. A bacterial lectin from the potent plant pathogen *Ralstonia solanacearum* (RSL) is closely related with the AAL lectin and its crystal structure has been solved recently.³⁴ This lectin, specific for fucose, is three times shorter in sequence than AAL and it is the only known example of a β-propeller structure generated by the oligomerisation of the monomer and not by sequence repeats in the single peptide chain. Both the β-propeller structure and the binding sites of AAL and RSL lectins are very similar. Tachylectin-2, isolated from hemocytes of the Japanese horseshoe crab (*Tachypleus tridentatus*), is the first protein to have been shown to adopt a five-bladed β-propeller fold and it is believed to play important roles in the innate immunity defence mechanism of the crab.³⁵ This lectin is specific for terminal GlcNAc and GalNAc monomers and closely resembles PVL in the binding specificity towards N-acetyl-chitooligosaccharides, showing an affinity constant that decreases with the increasing of the length of the sugar chain. Also, and analogous to PVL, tachylectin-2 binds well to GlcNAc residues substituted at the C6 position, like Galβ1-6GlcNAc, but is unable to recognize GlcNAc residues substituted at C3 or C4. Interestingly, the two lectins show a remarkable similarity in their binding sites, although there is no sequence homology and the two propeller structures are profoundly different. In tachylectin-2, the principle interactions of the sugar are with the residues of the strands D-A connecting loop, by analogy to PVL. The sugar O3 hydroxyl binds to the same conserved tryptophan and the N-acetyl group shows a hydrogen bonding scheme that closely matches that of PVL.

These three different lectin families (AAL/RSL, tachylectin-2 and PVL) show both similarities and differences, but it is clear that the β-propeller fold, with its high degree of pseudo symmetry, has converged through evolution in order to meet one of the most important features of lectins: multivalency. The apparent robustness conferred on a protein structure by the β-propeller fold is another interesting topic: PVL retains some of its binding

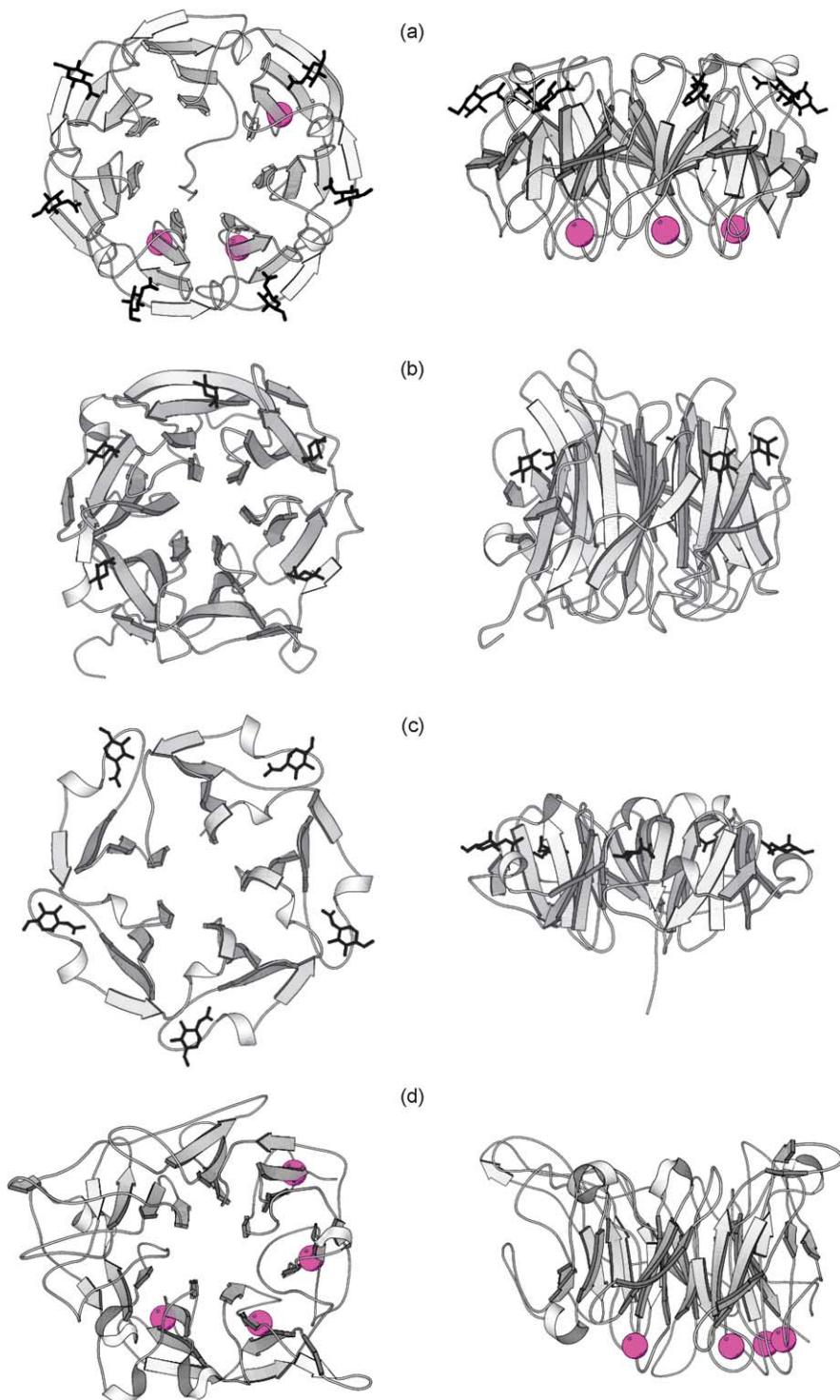


Figure 9. Orthographic representation of different β -propellers with ligand represented as sticks and calcium ions represented as magenta spheres. (a) PVL/GlcNAc complex, including the calcium atoms observed after additional soaking. (b) *Aleuria aurantia* lectin AAL/fucose complex (PDB code 1OFZ). (c) Tachylectin-2/GlcNAc complex (PDB code 1TL2). (d) Extracellular domain of integrin $\alpha V\beta 3$ with four calcium ions (PDB code 1JV2). Plots were made with MOLSCRIPT.⁵⁹

properties even after the strong denaturing conditions used in SDS-PAGE analysis.²² This robustness associated with its specificity could naturally make this lectin useful in biotechnology applications.

As pointed out recently,³⁶ cyclic lectin oligomers appear to be a convergent structural solution for multivalency. In addition to the five, six and seven bladed β -propellers now identified and cited above, ring-shaped oligomers having carbohydrate-binding

sites at their periphery have been observed for the plant lectins from *Galanthia nivalis* bulbs and *Parkia platycephala* seeds,³⁶ for the human serum pentraxin³⁷ and the snake *Crotalus atrox* C-type lectin.³⁸

Similarities with other proteins

PVL has no sequence similarity with any other previously reported lectin. A search in the fungal DNA databases at NCBI using Genomic Blast³⁹ indicated a related gene (50% identity) in the genome of *Coprinopsis cinerea* (or *Coprinus cinereus*) okayama7#130,⁴⁰ that is also a member of *Psathyrellaceae* family (Figure 10). A full homology search in protein databases with PHI-BLAST⁴¹ identified local similarities with a large number of sequences annotated as Ca^{2+} -binding protein, α -integrins or RTX protein. Using PROSITE,⁴² further searching for the conserved peptide motif GWx[VIL]xxHxR observed in the loops connecting two consecutive blades indicated three sequences

with at least six repeats of this motif in the translation of the newly sequenced genome from *Vibrio vulnificus* CMCP6,⁴³ *Thiobacillus denitrificans* ATCC 25259,⁴⁴ and uncultured bacterium soil samples of a fossil nuclear reactor.⁴⁵ The *T. denitrificans* sequence contains two domains with sequence similarity to PVL. For the three bacterial proteins, the alignment shows a strong similarity for a sequence length corresponding to six blades (Figure 10). Identity scores vary between 33% for *V. vulnificus* sequence (Q8D4R5_VIBVU), 35% for the unknown bacterium (Q52UU0_9BACT) to 46% and 48% for the C terminus and N terminus moieties of *T. denitrificans* sequence (Q3SJ62_THIDE). The nonapeptide motif DxxxDxxxD involved in Ca^{2+} binding is also highly conserved but it can also be found in many other classes of bacterial proteins and is not sufficient for identifying the present family. The amino acids involved in GlcNAc binding are conserved in several blades in the three proteins, which therefore can be predicted

PVL	-----VADLVGFNGGGVYI	TRNS---LLI-QVVKVINN-FGYD--AGGWRVEK	58
C.c.	-----SADIVGVFGDVGVH	WAPNS---TFCMQTKQVLSN-FGLD--TGGWKTCSR	92
T.d. Nterm	-----TADIVGVGST	GVWTALGNGDGSF-QPPRVVLA--FGTR--AGGWQFDK	153
T.d. Cterm	-----FSRDDPFARYPDH	TVLMRARDLH--KITER-T-DWDT--PEGWVRDVR	979
Envir.	-----RADVVAFGDPGVF	VAQQQADGTF--AAPVLRVR-DFGT--AQGWRTD	177
V.v.	-----IPDFVGFAKDG	YVAFGEVGGTYSAPOQKLVRGY-VEET--GGWNVAS	262
PVL	EVRILLADTTGDNQ-	SDVVVGFGENGVWISTNNCNNLFVDPPKMVLAN-FAYAA--GGW	115
C.c.	HPRILLCDTTGDN-	GKDLVGFGESGVWISRNNGTFR-PAECILRD-FGAAE	150
T.d. Nterm	HPRRELADLSGDR-	ADIAGFGDAGVWTALGNCDGTF-QPPQFVLAD-LGYE--SGW	208
T.d. Cterm	HPRLLTDLTASGR-	ADVVVGFAGDAGVWTALSKCGGF-ADPRFVLAD-LGYD--AGW	1034
Envir.	HPRLVADVDGDR-	DDLVGFGNAGVIVSYAQPDSF-TAPQLKVRD-FGWN--QGW	232
V.v.	HPRMLSDVNGDGK-	ADIVGFANAGIYVSLSECK-TEKQGT-LWSNA-YCAS--GGW	316
PVL	HIREFLADLIRKTGR-	ADIVVGFAGDGGIYISRNNCGQE-APAQLALNN-FGYA--QGW	170
C.c.	HIREFLADLIRNVGR-	VDIVVGFGERGVFVALNEKG-FEAPR-IAIDG-FGYN--QGW	205
T.d. Nterm	HIREFLVDLTGDR-	ADLVVGFAGDAGVWTALGNCDGTF-QPPQFVLAD-LGYE--AGW	263
T.d. Cterm	HIREFLVDLTGDR-	ADIVVGFAGDAGVWTALGNCDGTF-QPPQFVLAD-LGYE--AGW	1089
Envir.	HIREFLADLSGNGT-	ADIVVGFAGSAGVTVSHGRADRF-E-DGTGKRIDS-YGAD--RGW	287
V.v.	HERMLADVNGDGF-	NDVVVFAHOQVIVSLSK-C-TF-SEPKLWISQ-FGNN--QGW	370
PVL	HIREFLADVTGDL-	LDVVVGFGENQVYIARNSGNGT	226
C.c.	HIREFLADVTGDL-	LDIVVGFDSFVFSRSN	261
T.d. Nterm	HIREFLADVN	GDGR-ADLVVGFAGDAGVWTALGNCDGTF-QPPQFVLAD-LGYE--AGW	318
T.d. Cterm	HIREFLVDLTGDR-	ADIVVGFAGDAGVWTALGNCDGTF-QPPQFVLGD-LGYD--SGW	1144
Envir.	NPRMLADMDGNGT-	ADIVVGFAGDAGTVVSYSFAL	342
V.v.	HIREFLADINGDM-	ADIVVGFAGDAGTVVSYSFAL	425
PVL	HPRVVADLTGDRK-	ADILGFGVAGVYTSLNNGNT	281
C.c.	HPRFLADMTGDT-	VDVVVFKGSGAFVFSFNDCGSE	315
T.d. Nterm	HPRFLADLNGDGR-	PDIIVGVGDAGVWTALGNCDGTF-QPPQFVLAD-LGYE--SGW	373
T.d. Cterm	HPRFLVDLNGDGR-	ADIVVGGAGVWTALGNCDGTF-QPPQFVLAD-LGYE--SGW	1199
Envir.	HPRVVGVNVSGHLP-	PDVVGFGRGALAAYS	397
V.v.	HIREFLADVNGDGL-	ADIIIGFSSHGVKVSLSTCISLS-GE-SLWSDH-FGYA--AGG	480
PVL	EVRCVSSLTNKKV-	GDIIGFGDAGVYVAINNNGNTF-GAVNLVLKD-FGVN--SGW	336
C.c.	HPRFLVPVSRNNAAAD	IIGFGNDGVSVAINNNGDGT	371
T.d. Nterm	HPRFLADLTGDR-	ADLVVAFGDA	428
T.d. Cterm	HPRFLVVDLTGNGP-	TDLVGFAGDAGVWTALGNCDGSE-QPPQFVLAD-LGYD--AGW	1254
Envir.	HPRFLADLNGDGI-	LDVVVGFAGVLYAGDDFLT-GPE-LTSSS-FGWA--AGW	451
V.v.	HPRFLSDVNGDSI-	PDIIGFANDRVLTSISLETCFATRE-EGIRA-YGTR--AGS	535

Figure 10. Alignment of amino acid sequence repeats from PVL compared to translation genome fragment of *Coprinopsis cinerea* okayama7#130 (numbering starting from previous stop codon), the two domains of Q3SJ62_THIDE from *Thiobacillus denitrificans* ATCC 25259, Q52UU0_9BACT from unknown bacterium and Q8D4R5_VIBVU from *Vibrio vulnificus* CMCP6. The background is shaded according to similarity between the sequences. Symbols x and o indicate amino acids eventually involved in calcium and GlcNAc binding, respectively.

as six or seven-bladed β -propellers with calcium-binding and lectin activity. Interestingly, the occurrence of highly similar lectins in fungi and bacteria has already been observed since a fucose-binding lectin with a six-bladed β -propeller fold has been identified in the fungi *A. aurantia* and *Aspergillus oryzae* and in the bacterium *Ralstonia solanacearum*.⁷

A search for structurally related proteins in the protein database using the DALI server⁴⁶ resulted in a large family of β -propeller structures showing an overall rmsd of 4.0 Å. Amongst them, the most interesting ones in terms of structural similarity and Ca^{2+} -binding function are the α -integrins, i.e. the N terminus domain of a very important class of cell adhesion receptors. Integrins are cell surface glycoproteins that transmit conformational changes bidirectionally across the membrane, regulating various processes such as proliferation, differentiation, apoptosis, cell adhesion and migration.⁴⁷ Integrins have been the subject of extensive studies in the past few years and the crystal structure of the extracellular domain of integrin $\alpha V\beta 3$ has been elucidated.⁴⁸ As anticipated from theoretical studies,⁴⁹ the N terminus domain of integrins shows a seven-bladed β -propeller fold with the presence of calcium binding sites. Analogous to PVL, the calcium ions in integrins are located in the β -hairpin loops connecting strands A and B of some blades and the nonapeptide sequence Asp-X-X-X-Asp-X-X-X-Asp is identical. These calcium binding sites in integrins have been proposed to play a role in ligand binding by stabilising the β -propeller domain and modulating the integrin subunit association in an allosteric way.⁴⁸ The comparison between the integrin fold and PVL is shown in Figure 9. The overall shape and the calcium binding sites are very similar but there are no other similarities in the sequence. From the studies presented here, it is not known if the calcium ions in PVL are related to some specific biological or binding activity in the mushroom metabolism. Probably, and like integrins, they could play a role in the stabilization of the most external and therefore flexible regions of the protein, contributing to the stabilization of the entire structure.

Materials and Methods

PVL purification and characterisation

All the procedures were carried out at 4 °C. The extraction and the acidic treatment were adapted from Ueda *et al.*⁵⁰ Fruiting bodies of *P. velutina* collected in France (50 g) were ground in a mortar with liquid nitrogen and homogenised into 500 ml of sodium phosphate buffer (PBS) (pH 7.2) supplemented with 5 mM EDTA, 5 mM GlcNAc and 0.5 mM PMSF. The homogenate was centrifuged at 1000g for 30 min and the supernatant was collected. The solution was then slowly acidified at pH 4 using 20% acetic acid and gentle stirring. After standing for several hours the preparation was centrifuged at 1000g for 30 min. The clear solution was neutralised with concentrated

NaOH, filtered and subsequently dialysed against several changes of PBS (pH 7.2) containing 5% (v/v) glycerol.

Affinity adsorbent was prepared from epoxy-activated Sepharose-6B (Amersham Bioscience) and GlcNAc (Sigma) following the method described by Vretblad.⁵¹ Briefly, Sepharose-6B (5 g) is suspended in 10 ml of 0.1 M NaOH and added to 10 ml of 0.1 M NaOH containing 750 mg of GlcNAc. The mixture was gently mixed at 37 °C for 24 h. The gel was packed into a column (1 cm × 9 cm) and washed with PBS. The dialysate was injected and after extensive washing, PVL was eluted with 0.2 M GlcNAc. The purified lectin showed a single band on SDS-PAGE with apparent molecular mass of 40 kDa (data not shown). For crystallisation assays, the protein was then dialysed against pure water containing 150 mM NaCl and 5% glycerol and concentrated by centrifugation using a PALL Macrosep® centrifugal device. The protein used for the microcalorimetry experiments was extensively dialysed against several changes of PBS (pH 7.2) to ensure the complete removal of the bound GlcNAc. The sample was concentrated by centrifugation and re-dialysed against a small amount of buffer just before the microcalorimetry experiment.

Isothermal titration microcalorimetry (ITC)

Experiments were performed with a VP-ITC isothermal titration calorimeter (Microcal). The experiments were carried out at 25 °C. Protein concentration in the microcalorimeter cell (1.4478 ml volume) was 110 µM. Sugar solutions (prepared in PBS) in 30 injections of 10 µl each at concentrations of 7 mM, were added at intervals of 5 min while stirring at 310 rpm. Control experiments in which sugars are added to the microcalorimeter cell in the absence of protein yielded insignificant heats of dilution. The experimental data were fitted to a theoretical titration curve using software supplied by Microcal with ΔH (enthalpy change), K_a (association constant), and n (number of binding sites per monomer) as adjustable parameters, from the relationship:

$$Q_i = \frac{n P_t \Delta H V_0}{2} \times \left[1 + \frac{A_t}{n P_t} + \frac{1}{n K P_t} - \sqrt{\left(1 + \frac{A_t}{n P_t} + \frac{1}{n K P_t} \right)^2 - 4 \frac{A_t}{n P_t}} \right]$$

where P_t is the total protein concentration, A_t the total concentration of the ligand, V_0 the volume of the cell and Q_i the total released heat for the injection i . ΔG values and entropy contribution were determined from the standard equation:

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

where ΔG , ΔH and ΔS are the change in free energy, enthalpy and entropy of binding, T is the absolute temperature, $R=8.32 \text{ J mol}^{-1} \text{ K}^{-1}$ and K the association constant.

Crystallisation, data collection, structure determination and refinement

PVL was crystallised by vapour diffusion, using the hanging-drop method. Initial crystallisation conditions were screened with Hampton Screen I, II and the Jena Bioscience ammonium sulphate-based screen. Also, a crystallization screen comprising 576 conditions was

performed using the PSB/EMBL crystallisation platform. The results of the trials indicated that $(\text{NH}_4)_2\text{SO}_4$ in combination with Mes buffer represented good starting conditions and optimisation resulted in crystals after two weeks from 2 μl of PVL solution (2.5 mg/ml, 150 mM NaCl, 5% glycerol) mixed with 2 μl of 2.4 M $(\text{NH}_4)_2\text{SO}_4$, 0.1 M Mes buffer (pH 6.5). Seeding trials and the addition of salts (1 mM CaCl_2 and MgCl_2) improved crystal quality and the diffraction limit. Final crystallisation conditions are shown in Table 4. Crystals were cryo-protected with 10–15% glycerol in precipitant for as a short time as possible and flash-cooled in liquid nitrogen. Data were collected at ESRF (Grenoble, France) on beamline ID-29 on an ADSC Q210 CCD detector for the native and β -Me-SeGlcNAc crystals. The GlcNAc complex was collected at beamline ID23-1 on a Marmosaic 225 CCD detector while the Neu5Ac complex was collected at station ID14-1 on an ADSC Q4R CCD detector. Diffraction images for all datasets were integrated with MOSFLM⁵² and scaled and converted to structure factors with the CCP4 program suite.⁵³

The carbohydrate analogue methyl 2-acetamido-2-deoxy-1-seleno- β -D-glucopyranose (β -Me-SeGlcNAc) was synthesised as described,⁵⁴ and co-crystallised with the lectin resulting in well-formed single crystals of space group $P2_12_12$ with one molecule in the asymmetric unit. The structure was solved by the SAD method using data to 1.8 \AA resolution collected at the peak of the selenium K-absorption edge as determined from a fluorescence scan. The anomalous Patterson map clearly showed peaks corresponding to three major selenium sites per asymmetric unit. AutoSHARP⁵⁵ found four selenium sites from which initial phases were derived. Density modification provided an improved electron density map of excellent quality into which arp/WARP⁵⁶ built and

docked side-chains of 381 out of 401 amino acid residues. For this initial model building step only, the PVL sequence accessible in EMBL database (accession number E16441) was used.

Prior to data collection, the crystal obtained by co-crystallisation of the lectin with Neu5Ac was soaked for 1 h in a 2.4 M $(\text{NH}_4)_2\text{SO}_4$, 50 mM Hepes (pH 7), 10% glycerol solution containing 20 mM Neu5Ac. The crystals of native PVL and the complex with Neu5Ac proved to be of space group $P2_1$ containing a dimer in the asymmetric unit, while the GlcNAc complex crystallised in space group $P3_121$ with one monomer per asymmetric unit. For these datasets, the monomeric structure of the β -Se-GlcNAc complex was used as a molecular replacement search probe by the program MOLREP⁵³ followed by automatic full reconstruction using arp/WARP for the native and the sialic acid complex. Refinement and initial solvent model building were performed by Refmac,⁵⁷ combined with manual rebuilding using O.⁵⁸ Double conformations were modelled where necessary and the occupancy estimated according to their B -factors. The statistics for the final refined models are summarised in Table 5. The high-resolution (1.5 \AA) native structure revealed some discrepancies with the deposited PVL sequence (E16441) and this led us to clone the *pvl* gene to avoid ambiguous interpretation in the lower resolution electron density maps. All the other structures were corrected accordingly.

In order to confirm the presence of calcium in the structure, a native crystal was soaked for 1 h in a solution of its mother liquor containing 100 mM CaCl_2 and 10% glycerol. The resulting crystal was flash-cooled and a highly redundant data set of 720° of data was collected at station ID14-2 with a wavelength of 0.934 \AA . Phases for the calcium-soaked crystal structure were obtained by

Table 4. Data collection for PVL and PVL complexes

	Native	β -Me-SeGlcNAc	GlcNAc	NeuNAc
Beamline	ID29	ID29	ID23-1	ID14-1
Space group	$P2_1$	$P2_12_12$	$P3_121$	$P2_1$
Cell dimensions				
<i>a</i> (\AA)	51.83	82.62	106.8	51.73
<i>b</i> (\AA)	142.22	105.29	106.8	144.40
<i>c</i> (\AA)	52.16	51.83	86.6	52.60
α , β , γ (deg.)	90, 92.1, 90	90, 90, 90	90.0, 90, 120	90.0, 93.0, 90
Subunits/asym.	2	1	1	2
Solvent content (%)	44.5	54.7	62.6	45.7
Matthews coeff.	2.2	2.6	3.3	2.3
Wavelength (\AA)	0.976	0.979	0.954	0.934
Resolution (\AA)	21.1–1.50	25.2–1.80	25.6–2.60	26.7–1.80
High resolution shell (\AA)	1.55–1.50	1.85–1.80	2.67–2.60	1.85–1.80
Measured reflections	410,940 (35,125) ^a	964,565(69,580)	199,289(14,594)	279,137(20,377)
Unique reflections	115,491(10,551)	42,217(3042)	17,901(1293)	70,971(5260)
Average multiplicity	3.6 (3.3)	22.8 (22.9)	11.1 (11.3)	3.9 (3.9)
Completeness (%)	96.3 (90.1)	99.0 (98.2)	99.9 (100)	99.8 (99.8)
Anomalous completeness (%)		98.9 (98.2)		
Average $I/\sigma I$	5.3 (2.6)	6.3 (1.7)	5.5 (1.6)	8.3 (1.8)
R_{merge} (%) ^b	6.9 (26.1)	10.2 (44.0)	10.2 (46.7)	4.8 (28.8)
Wilson B -factor (\AA^2)	12.2	11.4	63.3	21.4
Figure of merit (before density modification)		0.471		
Rcullis (anom) ^c		0.686		
Phasing power (anom) ^d		1.533		

^a Values in parentheses refer to the highest resolution shell.

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

^c $R_{\text{cullis(anom)}} = \text{rms}(\text{DANO}_{\text{calc}} - \text{DANO}_{\text{obs}}) / \text{rms}(\text{DANO}_{\text{obs}})$.

^d $\text{Phasing power}_{\text{(anom)}} = \text{rms}(\text{DANO}_{\text{calc}}) / \text{rms}(\text{DANO}_{\text{calc}} - \text{DANO}_{\text{obs}})$.

Table 5. Refinement statistics for PVL and PVL complexes

	Native	β -Me-SeGlcNAc	GlcNAc	NeuAc
Resolution range (Å)	21.15–1.50	25.17–1.80	25.65–2.60	24.68–1.80
R_{cryst} (observations)	0.159 (109,647)	0.151 (40,041)	0.205 (16,965)	0.188 (67,358)
R_{free} (observations)	0.183 (5801)	0.183 (2127)	0.271 (912)	0.230 (3575)
<i>Highest resolution shell</i>				
R_{cryst} (observations)	0.191 (7842)	0.151 (2882)	0.320 (1226)	0.241 (4992)
R_{free} (observations)	0.234 (392)	0.215 (157)	0.363 (63)	0.262 (261)
Cruickshank's DPI based on maximum likelihood (Å)	0.069	0.096	0.297	0.127
Average B_{iso} (Å)	10.4	9.5	43.1	21.4
<i>rms deviations</i>				
Bonds (Å)	0.016	0.016	0.017	0.017
Angles (degrees)	1.655	1.682	1.713	1.614
Ramachandran's outliers	0	0	0	0
<i>Atoms used in refinement</i>				
Protein/sugar/other/solvent	6076/0/72/894	3038/80/18/414	3038/90/21/55	6076/105/22/447
Residues showing alternate conformations	AsnA45, ValA94, IleA323, AspA335, ValA342, IleA352, IleA374, MetA381, SerB200, IleB323	Val13, Met375		IleA323, ValA342, IleA374, IleB323
PDB deposition code	2bwr	2bwm	2c4d	2c25

molecular replacement using the refined native dimer as search probe and an anomalous phased map was calculated using the experimental anomalous differences and the calculated phases (after initial refinement) as coefficients.

Reverse transcription-PCR and cDNA cloning of *Psathyrella velutina* lectin

Total RNA was prepared by grinding *P. velutina* fruiting bodies in liquid nitrogen followed by an extraction using TRI reagent (Sigma Chemical Co.). RNA preparation was used to synthesise first-strand cDNA using the Super-Script™II reverse transcriptase and an oligo(dT) primer from Invitrogen, and the coding region of *pvl* gene was then amplified by PCR. The following primers were designed: pr-pvl (sense 5'-GGAATTCCATATGTGATCCCAGT CATCAG-3') and pr-pvrev (antisense, 5'-CGGGATCCT CAGAGGTAGAGGTTCCG-3'). Sequences in bold characters code for the N-terminal (MSIPVIS) and C-terminal (ANLYL) peptide sequences, respectively, which were taken from a previously published *pvl* nucleotide sequence (EMBL accession number E16441). The primers contained the restriction sites NdeI (sense) and BamHI (antisense) on their 5'-ends for further subcloning. PCR was performed using Taq DNA polymerase (Promega), with the following temperature profile: 2 min at 95 °C, 45 s at 95 °C, 30 s at 48 °C, 3 min at 72 °C, for 35 cycles. These conditions permitted the amplification of a fragment of 1200 bp. The purified PCR fragment was cloned into the pGEM-T Easy vector (Promega). Seven independent clones were sequenced on both strands (Genome Express, Meylan, France). Oligonucleotides were purchased from MWG S.A. Biotech (Courtaboeuf, France).

Protein Data Bank accession codes

The atomic coordinates have been deposited in the RCSB Protein Data Bank with accession codes as shown in Table 5.

Acknowledgements

We thank the ESRF, Grenoble, for access to synchrotron data collection facilities and EMBL, Grenoble Outstation, for access to the crystallization platform. Maria Pechlaner is acknowledged for her help in protein purification. We thank Flora Pérez and the Société mycologique du Dauphiné for their help in picking and identifying mushrooms. G.C. is an EEC doctoral fellow (HPRN-CT2000-00001).

References

- Lis, H. & Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* **98**, 637–674.
- Guillot, J. & Konska, G. (1997). Lectins in higher fungi. *Biochem. System. Ecol.* **25**, 203–230.
- Nordbring-Hertz, B. & Chet, I. (1986). Fungal lectins and agglutinins. In *Microbial Lectins and Agglutinin: Properties and Biological Activity* (Mirelman, D., ed.), pp. 393–408, Wiley, New York.
- Pemberton, R. T. (1994). Agglutinins (lectins) from some British higher fungi. *Mycol. Res.* **98**, 277–290.
- Wang, H., Ng, T. B. & Oooi, V. E. C. (1998). Lectins from mushrooms. *Mycol. Res.* **102**, 897–906.
- Imberty, A., Mitchell, E. P. & Wimmerová, M. (2005). Structural basis for high affinity glycan recognition by bacterial and fungal lectins. *Curr. Opin. Struct. Biol.* **15**, 525–534.
- Wimmerová, M., Mitchell, E., Sanchez, J. F., Gautier, C. & Imberty, A. (2003). Crystal structure of fungal lectin: six-bladed β -propeller fold and novel recognition mode for *Aleuria aurantia* lectin. *J. Biol. Chem.* **278**, 27059–27067.
- Paaventham, P., Joseph, J. S., Seow, S. V., Vaday, S., Robinson, H., Chua, K. Y. & Kolatkar, A. R. (2003). A 1.7 Å structure of Fve, a member of the new fungal immunomodulatory protein family. *J. Mol. Biol.* **332**, 461–470.
- Birck, C., Damian, L., Marty-Detraves, C., Lougarre, A., Schulze-Brieze, C. K. P., Fournier, D. et al. (2004). A new

- lectin family with structure similarity to actinoporins revealed by the crystal structure of *Xerocomus chrysenteron* lectin XCL. *J. Mol. Biol.* **344**, 1409–1420.
10. Carrizo, M. E., Capaldi, S., Perduca, M., Irazoqui, F. J., Nores, G. A. & Monaco, H. L. (2005). The antineoplastic lectin of the common edible mushroom (*Agaricus bisporus*) has two binding sites, each specific for a different configuration at a single epimeric hydroxyl. *J. Biol. Chem.* **280**, 10614–10623.
 11. Walser, P. J., Haebel, P. W., Künzler, M., Sargent, D., Kües, U., Aebi, M. & Ban, U. (2004). Structure and functional analysis of the fungal galectin CGL2. *Structure*, **12**, 689–702.
 12. Ban, M., Yoon, H. J., Demirkan, E., Utsumi, S., Mikami, B. & Yagi, F. (2005). Structural basis of a fungal galectin from *Agrocybe cylindracea* for recognizing sialoconjugate. *J. Mol. Biol.* **351**, 695–706.
 13. Mancheno, J. M., Tateno, H., Goldstein, I. J., Martinez-Ripoll, M. & Hermoso, J. A. (2005). Structural analysis of the *Laetiporus sulphureus* hemolytic pore-forming lectin in complex with sugars. *J. Biol. Chem.* **280**, 17251–17259.
 14. Kochibe, N. & Matta, K. L. (1989). Purification and properties of a *N*-acetylglucosamine-specific lectin from *Psathyrella velutina* mushroom. *J. Biol. Chem.* **264**, 173–177.
 15. Endo, T., Ohbayashi, H., Kanazawa, K., Kochibe, N. & Kobata, A. (1992). Carbohydrate binding specificity of immobilized *Psathyrella velutina* lectin. *J. Biol. Chem.* **267**, 707–713.
 16. Parekh, R. B., Dwek, R. A., Sutton, B. J., Fernandes, D. L., Leung, A., Stanworth, D. *et al.* (1985). Association of rheumatoid arthritis and primary osteoarthritis with changes in the glycosylation pattern of total serum IgG. *Nature*, **316**, 452–457.
 17. Tsuchiya, N., Endo, T., Matsuta, K., Yoshinoya, S., Takeuchi, F., Nagano, Y. *et al.* (1993). Detection of glycosylation abnormality in rheumatoid IgG using *N*-acetylglucosamine-specific *Psathyrella velutina* lectin. *J. Immunol.* **151**, 1137–1146.
 18. Liljeblad, M., Lundblad, A. & Pahlsson, P. (2000). Analysis of agalacto-IgG in rheumatoid arthritis using surface plasmon resonance. *Glycoconj. J.* **17**, 323–329.
 19. Moore, J. S., Wu, X., Kulhavy, R., Tomana, M., Novak, J., Moldoveanu, Z. *et al.* (2005). Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS*, **19**, 381–389.
 20. Ueda, H., Kojima, K., Saitoh, T. & Ogawa, H. (1999). Interaction of a lectin from *Psathyrella velutina* mushroom with *N*-acetylneurameric acid. *FEBS Letters*, **448**, 75–80.
 21. Ueda, H., Matsumoto, H., Takahashi, N. & Ogawa, H. (2002). *Psathyrella velutina* mushroom lectin exhibits high affinity toward sialoglycoproteins possessing terminal *N*-acetylneurameric acid α 2-3-linked to penultimate galactose residues of trisialyl *N*-glycans. *J. Biol. Chem.* **277**, 24916–24925.
 22. Ueda, H., Saitoh, T., Kojima, K. & Ogawa, H. (1999). Multi-specificity of a *Psathyrella velutina* mushroom lectin: heparin/pectin binding occurs at a site different from the *N*-acetylglucosamine/*N*-acetylneurameric acid-specific site. *J. Biochem.* **126**, 530–537.
 23. Asano, Y. & Kamoashi, T. (1998). Polypeptide of lectin and DNA coding the same, Patent number JP1998174588-A/1. SEIKAGAKU KOGYO CO LTD., Japan.
 24. Dam, T. K. & Brewer, C. F. (2002). Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.* **102**, 387–429.
 25. Buts, L., Loris, R., De Genst, E., Oscarson, S., Lahmann, M., Messens, J. *et al.* (2003). Solving the phase problem for carbohydrate-binding proteins using selenium derivatives of their ligands: a case study involving the bacterial F17-G adhesin. *Acta Crystallogr. sect. D*, **59**, 1012–1015.
 26. Jawad, Z. & Paoli, M. (2002). Novel sequences propel familiar folds. *Structure*, **10**, 447–454.
 27. Fujihashi, M., Peapus, D. H., Kamiya, N., Nagata, Y. & Miki, K. (2003). Crystal structure of fucose-specific lectin from *Aleuria aurantia* binding ligands at three of its five sugar recognition sites. *Biochemistry*, **42**, 11093–11099.
 28. Amano, K., Fujihashi, M., Ando, A., Miki, K. & Nagata, Y. (2004). Involvement of tyrosines at fucose-binding sites of *Aleuria aurantia* lectin: non-equal response to site-directed mutagenesis among five sites. *Biosci. Biotechnol. Biochem.* **68**, 841–847.
 29. Wright, C. S. (1984). Structural comparison of the two distinct sugar binding sites in wheat germ agglutinin isolectin II. *J. Mol. Biol.* **178**, 91–104.
 30. Wright, C. S. (1980). Location of the *N*-acetyl-D-neuraminic acid binding site in wheat germ agglutinin. A crystallographic study at 2.8 Å resolution. *J. Mol. Biol.* **139**, 53–60.
 31. Wright, C. S. & Jaeger, J. (1993). Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycoporphin A. *J. Mol. Biol.* **232**, 620–638.
 32. Varghese, J. N., Laver, W. G. & Colman, P. M. (1983). Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution. *Nature*, **303**, 35–40.
 33. Witarto, A. B. & Sode, K. (2001). Increasing the hydrophobic interaction between terminal W-motifs enhances the stability of *Salmonella typhimurium* sialidase. A general strategy for the stabilization of beta-propeller protein fold. *Protein Eng.* **14**, 891–896.
 34. Kostlanová, N., Mitchell, E. P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N. *et al.* (2005). The fucose-binding lectin from *Ralstonia solanacearum*: a new type of β-propeller architecture formed by oligomerisation and interacting with fucoside, fucosyllactose and plant xyloglucan. *J. Biol. Chem.* **280**, 27839–27849.
 35. Beisel, H. G., Kawabata, S., Iwanaga, S., Huber, R. & Bode, W. (1999). Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. *EMBO J.* **18**, 2313–2322.
 36. Gallego del Sol, F., Nagano, C., Cavada, B. S. & Calvete, J. J. (2005). The first crystal structure of a Mimosoideae lectin reveals a novel quaternary arrangement of a widespread domain. *J. Mol. Biol.* **353**, 574–583.
 37. Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J. *et al.* (1994). Structure of pentameric human serum amyloid P component. *Nature*, **367**, 338–345.
 38. Walker, J. R., Nagar, B., Young, N. M., Hirama, T. & Rini, J. M. (2004). X-ray crystal structure of a galactose-specific C-type lectin possessing a novel decameric quaternary structure. *Biochemistry*, **43**, 3783–3792.
 39. Cummings, L., Riley, L., Black, L., Souvorov, A., Resenchuk, S., Dondoshansky, I. & Tatusova, T. (2002). Genomic BLAST: custom-defined virtual databases for complete and unfinished genomes. *FEMS Microbiol. Letters*, **216**, 133–138.

40. Birren, B., Nusbaum, C., Abebe, A., Abouelleil, A., Adekoya, E., Ait-zahra, M., et al. (2003). The genome sequence of *Coprinus cinereus*. Whitehead Institute/MIT Center for Genome Research.
41. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
42. Hulo, N., Sigrist, C. J., Le Sauv, V., Langendijk-Genevaux, P. S., Bordoli, L., Gattiker, A. et al. (2004). Recent improvements to the PROSITE database. *Nucl. Acids Res.* **32**, D134–D137.
43. Kim, Y. R., Lee, S. E., Kim, C. M., Kim, S. Y., Shin, E. K., Shin, D. H. et al. (2003). Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. *Infect. Immun.* **71**, 5461–5471.
44. Beller, H. R., Chain, P. S., Letain, T. E., Chakicherla, A., Larimer, F. W., Richardson, P. M. et al. (2006). The genome sequence of the obligately chemolithoautotrophic facultatively anaerobic bacterium *Thiobacillus denitrificans*. *J. Bacteriol.* **188**, 1473–1488.
45. Chapon, V. & Chanal, A. (2005). A new strategy to exploit environmental genes resources reveals the role of a eukaryotic-type topoisomerase in bacterial radioresistance. EMBL/GenBank/DDBJ databases. In the press.
46. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.
47. Hynes, R. O. (2002). Integrins: bidirectional, allosteric signaling machines. *Cell*, **110**, 673–687.
48. Xiong, J. P., Stehle, T., Diefenbach, B., Zhang, R., Dunker, R., Scott, D. L. et al. (2001). Crystal structure of the extracellular segment of integrin alpha Vbeta3. *Science*, **294**, 339–345.
49. Springer, T. A. (1997). Folding of the N-terminal, ligand-binding region of integrin alpha-subunits into a beta-propeller domain. *Proc. Natl Acad. Sci. USA*, **94**, 65–72.
50. Ueda, H., Takahashi, N. & Ogawa, H. (2003). *Psathyrella velutina* lectin as a specific probe for N-acetylneuraminc acid in glycoconjugates. *Methods Enzymol.* **363**, 77–90.
51. Vretblad, P. (1976). Purification of lectins by biospecific affinity chromatography. *Biochim. Biophys. Acta*, **434**, 169–176.
52. Leslie, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4+ESF-EAMCB News letter on Protein Crystallography*, **26**.
53. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **760–763**.
54. Ogra, Y., Ishiwata, K., Takayama, H., Aimi, N. & Suzuki, K. T. (2002). Identification of a novel selenium metabolite, Se-methyl-N-acetylselenohexosamine, in rat urine by high-performance liquid chromatography-inductively coupled plasma mass spectrometry and electrospray ionization tandem mass spectrometry. *J. Chromatogr. sect. B*, **767**, 301–312.
55. de La Fortelle, E. & Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494.
56. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
57. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. sect. D*, **53**, 240–255.
58. Jones, T. A., Zou, J. Y., Cowan, S. W. & 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. sect. A*, **47**, 110–119.
59. Kraulis, P. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950.

Edited by R. Huber

(Received 24 November 2005; received in revised form 16 January 2006; accepted 17 January 2006)
Available online 6 February 2006

Chapitre 9

Conclusions générales et perspectives

9.1 Rôles des lectines dans les infections causées par *Pseudomonas aeruginosa*

Une partie importante de cette thèse a été consacrée à l'étude structurale et à la caractérisation de l'affinité des lectines bactériennes, notamment celles impliquées dans les processus d'infection causés par certaines bactéries opportunistes. La résolution de la structure de PA-IL en complexe avec son ligand de haute affinité, décrite dans le sixième chapitre, a montré les bases structurales de sa spécificité et a complété l'étude structurale des lectines solubles de *Pseudomonas aeruginosa*. Cependant, de nombreuses questions restent encore sans réponse. Plus précisément, les véritables cibles de l'adhésion et les rôles physiologiques joués par ces deux lectines sont encore loin d'être complètement élucidés, bien que des hypothèses intéressantes et plausibles aient déjà été avancées. En raison de sa haute affinité pour le fucose et de son mode de reconnaissance unique, la lectine PA-IIL a été étudiée beaucoup plus en détail que PA-IL. Il a été démontré que la PA-IIL est associée à la membrane externe de la bactérie et joue un rôle important dans la formation du biofilm, qui protège la bactérie, notamment contre les antibiotiques. La lectine peut également bloquer le battement des cils des cellules épithéliales dans les voies respiratoires. Les autres rôles possibles, tels que l'implication directe dans la reconnaissance des cellules hôtes, l'adhésion ou encore le contact entre bactéries, sont en cours d'étude. La lectine PA-IL a été moins étudiée et il apparaît nécessaire dans le futur d'identifier les oligosaccharides galactosylés qui seraient les ligands de haute affinité et de caractériser les complexes.

L'identification des ligands naturels des lectines de *P. aeruginosa* est nécessaire pour envisager la conception d'inhibiteurs glycomimétiques qui pourront bloquer l'adhésion bactérienne. En raison des stratégies bactériennes qui utilisent plusieurs récepteurs pour se lier aux cellules de l'hôte, il faut envisager de bloquer plusieurs lectines et adhésines. Cet axe de recherche apparaît prometteur pour prévenir ou bloquer les stades précoce d'infection bactérienne et pourrait représenter un complément ou une alternative aux traitements antibiotiques.

9.2 Thermodynamique de la reconnaissance lectine-sucré

Les valeurs de K_a caractéristiques pour les lectines de légumineuses en interaction avec des monosaccharides se situent autour de 10^2 à 10^3 M^{-1} . Par contre, les lectines qui reconnaissent des oligosaccharides montrent souvent des K_a de l'ordre de 10^4 M^{-1} . La microcalorimétrie de titration permet d'évaluer les contributions respectives de l'enthalpie et de l'entropie, et donc de caractériser la thermodynamique de la reconnaissance lectine-sucré (Dam 2002). En générale, ces interactions impliquent la formation de nombreuses liaisons hydrogène qui génèrent des valeurs de ΔH fortes et négatives. Cette contribution enthalpique est contre-balancée par une contribution entropique généralement défavorable qui résulte du réarrangement des molécules de solvant (Lemieux 1991) ou d'une perte de flexibilité du ligand (mais aussi de la protéine) à la suite de la complexation (Carver 1993). Les deux contributions peuvent être représentées dans un graphe du type de celui montré en Fig. 9.1 (Imbert 2005).

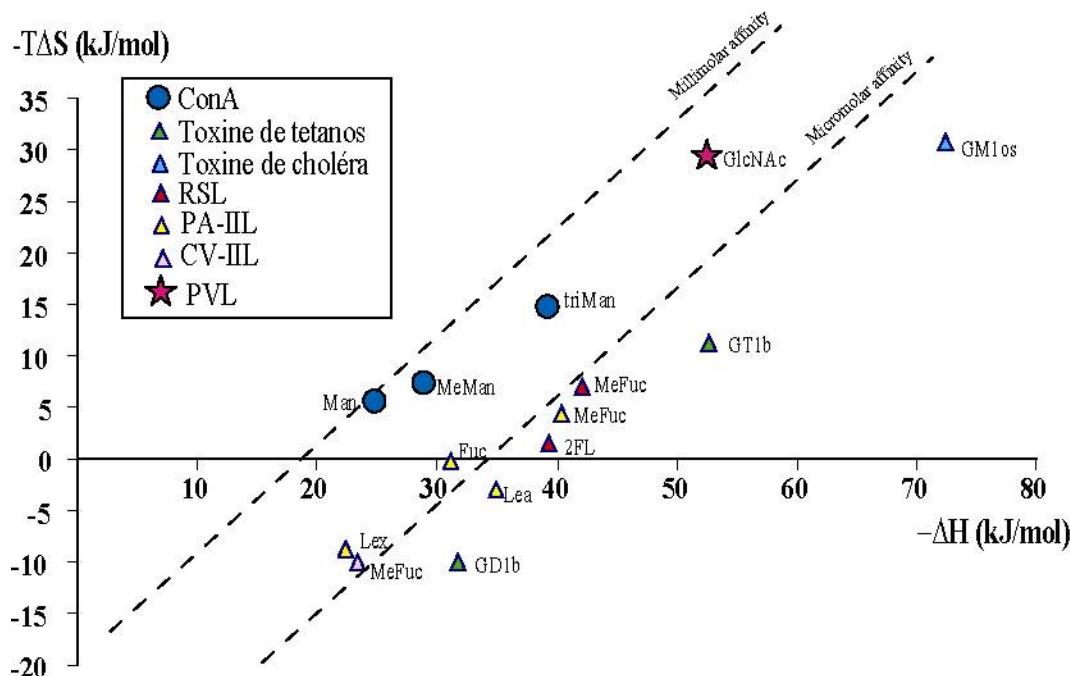


Figure 9.1 Plot enthalpie/entropie pour les lectines connues (Imbert 2005). Les lectines PVL et CV-IIL étudiées dans la thèse ont été insérées.

La figure illustre le fait que les lectines de légumineuses et de champignons (PVL) ont des affinités moyennes (environ 10^4 M^{-1}) car la contribution entropique s'oppose à la forte enthalpie d'interaction due aux nombreuses liaisons hydrogène. Dans l'interaction de la PVL avec son ligand monosaccharide, le GlcNAc, l'interaction est dominée par une forte enthalpie

favorable est une très forte entropie défavorable. On peut imaginer que la forte perte d'entropie soit due au caractère multivalent de l'interaction (quatre à six sites d'interactions) qui n'est pas contrebalancée par un relâchement suffisant de molécules d'eau dans le solvant.

Toutes les lectines bactériennes représentées ont des interactions fortes ($K_a > 10^6 \text{ M}^{-1}$) avec leur ligand. Cependant, la thermodynamique de l'interaction peut être très différente d'une lectine à l'autre. La lectine de choléra présente des contributions enthalpiques et entropiques très fortes, la somme des deux donnant une énergie libre de liaison suffisante pour une affinité de l'ordre de 10^8 M^{-1} (Turnbull 2004). Les lectines bactériennes solubles de la famille PA-IIL (mais aussi la RSL) sont une exception car elles montrent une grande affinité pour les monosaccharides avec une contribution entropique peu défavorable, ou même favorable comme dans le cas de la CV-IIL (chapitre 7). La structure native de la PA-IIL (Loris, Tielker et al. 2003) montre que les deux ions calcium présents dans le site de liaison interagissent fortement avec des molécules d'eau. A la suite de la complexation avec le sucre, ces molécules d'eau sont libérées dans le solvant, ce qui est entropiquement favorable à l'interaction.

L'analyse de la thermodynamique des interactions de différentes lectines bactériennes met donc en évidence des stratégies très différentes. Dans certains cas, les sites de liaison sont profonds et peuvent lier des oligosaccharides de manière très spécifique avec une grande contribution enthalpique. Dans d'autres cas, des ions calcium sont impliqués dans l'interaction et jouent un rôle sur le facteur entropique. Il faut noter que le lieu commun énonçant que les interactions lectine-sucre sont toujours de basse affinité n'est pas vrai pour les lectines bactériennes où de très fortes affinités sont maintenant observées.

9.3 Evolution de la multivalence: *multimères contre β -propellers*

Dans le deuxième chapitre on a introduit le concept de multivalence, qui est la propriété générale des lectines de lier simultanément plusieurs unités glucidiques. Cette caractéristique des lectines semble être indispensable d'une part pour atteindre une haute avidité (dans le cas des interactions de faible affinité) mais aussi pour une reconnaissance spécifique de certains glycoconjugués ou de surfaces cellulaires. Dans certains cas, comme pour exemple dans l'interaction anticorps-récepteur, la Nature semble préférer les interactions de plusieurs unités ayant une affinité moyenne que l'interaction d'une seule unité à forte affinité (Ambrosi 2005). De plus, il semble que certains repliements protéiques (comme le *jelly-roll* ou le β -*prism*) soient présents dans des lectines très différentes et non reliées phylogénétiquement. Une explication pourrait venir du fait que ces repliements ont été sélectionnés par la Nature parce qu'ils sont capables de former des structures ayant une multivalence élevée, qui favoriserait la

reconnaissance lectine-sucre. La succession de plusieurs sites de liaisons en tandem le long d'une chaîne peptidique est également retrouvée de manière classique dans les replis protéiques tels que les β -trefoil ou les β -propellers.

Les arrangements des sites de liaisons dans l'espace peuvent être très différents. Une stratégie particulière pour atteindre une multivalence élevée est la formation de multimères cycliques par l'association de plusieurs monomères (Fig. 9.2). Parmi les exemples connus on peut citer la lectine végétale de *Platia platycephala*, ou deux chaînes, chacune formée par trois domaines β -prism, s'associent pour former une structure hexavalente (Gallego del Sol 2005). Un deuxième exemple est la pentraxine, lectine du sérum humain formée par l'association de cinq chaînes ayant un repliement jelly-roll et qui joue un rôle dans le système de réponse immunitaire (Emsley 1994). Dans la structure quaternaire adoptée par l'agglutinine de type C extraite du venin du serpent *Crotalus atrox*, deux pentamères sont superposés l'un sur l'autre pour former un multimère avec dix sites de liaison pour le sucre (Walker 2004).

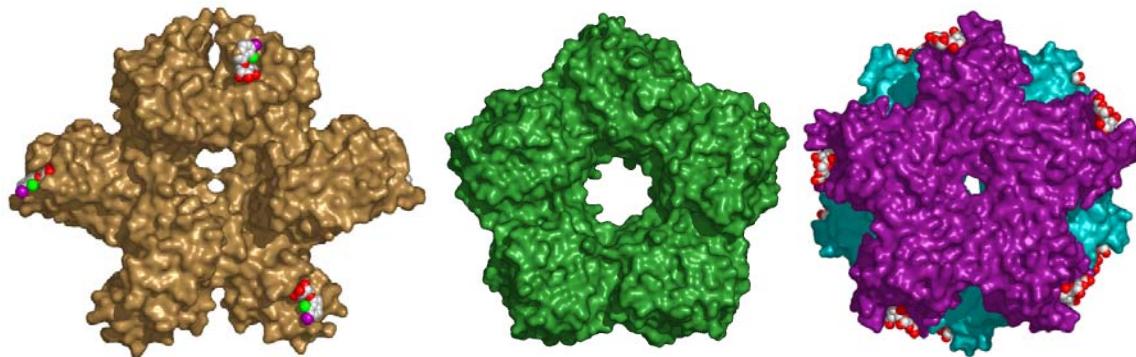


Figure 9.2 Multimères. Lectine de *P. platycephala* (gauche pdb 1zgs), pentraxine (centre pdb 1sac) et agglutinine de *Crotalus atrox* (droite pdb 1jzn)

Les lectines à β -propeller montrent une autre stratégie pour la formation de structures cycliques présentant des sites de reconnaissance à sa périphérie. Ce repliement protéique est caractérisé par un nombre de répétitions (appelées aussi lames) qui peut varier de cinq à neuf. Ces lames, formées par un feuillet β à quatre brins, s'arrangent de façon circulaire autour d'un axe de symétrie central (Fig. 9.3). La lectine PVL, décrite dans le chapitre 8, est un β -propeller très régulier à sept lames. Des autres exemples sont la lectine de champignon AAL à six lames (Wimmerova 2003) et la lectine animale tachylectine-2 à cinq lames (Beisel 1999).

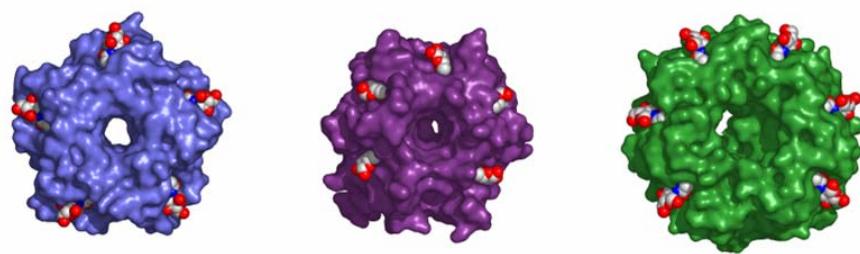


Figure 9.3 : Lectines de type β -propellers. Tachylectine-2 (gauche, pdb 1tl2), AAL de *Aleuria aurantia* (centre, pdb 1ofz) et PVL de *Psathyrella velutina* (droite, décrite dans le chapitre 8)

Les deux stratégies de multivalence exposées ci-dessus sont semblables dans le concept général mais montrent aussi quelques différences. D'abord la taille des deux type de molécules n'est pas la même, les β -propellers étant moins volumineux que les gros multimères. La distance entre deux sites de liaison est donc plus grande dans les multimères et ces structures pourraient jouer un rôle dans l'interaction multivalente avec des ligands de plus grande taille, comme dans l'agglutination des plusieurs cellules. Par contre, le repliement compact des β -propellers semble posséder une grande stabilité et pourrait être un avantage surtout chez les organismes opportunistes, comme le champignon ou les bactéries, qui doivent s'adapter à des conditions de vie très variables. Dans tous les cas, qu'il s'agisse de multimères ou de β -propellers, l'évolution vers des structures cycliques ayant une symétrie élevée semble être favorisée. Un exemple est montré par les lectines AAL et RSL. La AAL, un β -propeller classique, possède six répétitions dans la séquence. Par contre, dans la RSL (qui est trois fois plus petite) trois monomères s'associent pour former une structure pseudo β -propeller, très semblable à AAL.

En conclusion, il est très intéressant de noter que la majorité de ces lectines jouent des rôles soit dans la réponse du système immunitaire, comme par exemple la pentraxine ou la tachylectine-2, soit dans l'invasion ou l'attaque d'autres organismes, comme pour exemple pour les lectines bactériennes ou la lectine du venin du *Crotalus atrox*. Par ailleurs, on a vu que les toxines bactériennes, comme celles de type AB₅, utilisent aussi la multivalence pour maximiser l'interaction avec les glycoconjugués de la surface cellulaire. Il est donc très probable que la résolution des nouvelles lectines de type multimères ou β -propeller pourront valider l'hypothèse que la formation de structures cycliques soit une caractéristique fréquente des lectines ayant des rôles de défense ou d'invasion.

References

- Abeijon, C., and Hirschberg, C. B. (1992) Topography of glycosylation reaction in the endoplasmic reticulum. *Trends Biochem. Sci.*, 17, 32-36.
- 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.
- Allinger, N.L., Yuh, Y. H. and Lii, J. H. (1989) Molecular mechanics. The MM3 force field for hydrocarbons. *J. Amer. Chem. Soc.*, 111, 8551-8566.
- Alves de Brito, C.F., Carvalho, C. M. B., Santos, F. R., Gazzinelli, R. T., Oliveira, S. C., Azevedo, V. and Teixeira, S. M .R. (2004) *Chromobacterium violaceum* genome: molecular mechanisms associated with pathogenicity. *Genet. Mol. Res.*, 3, 148-161.
- Ambrosi, M., Cameron, N.R. and Davis, B.G. (2005) Lectins: tools for the molecular understanding of the glycocode. *Org. Biomol. Chem.*, 3, 1593-1608.
- Artenstein, A.W., and Cross, A.S. (1993) Local and disseminated diseases caused by *P. aeruginosa*. In Campa, M., Beninelli, M., et al. (eds), *Pseudomonas aeruginosa* as an opportunistic pathogen. Plenum Press, New York, pp. 224-235.
- Ashwell, G., and Morell, A.G. (1974) The role of surface carbohydrates in the hepatic recognition and transport of circulating glycoproteins. *Adv. Enzymol.*, 41, 99-128.
- Bah, M., and Pereda-Miranda, R. (1996) Spectrometry and high resolution NMR investigation or tricolorin A-E, individual oligosaccharides from the resins of *Ipomoea tricolor* (Convolvulaceae). *Tetrahedron*, 52, 13063-13080.
- 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.
- Ban, M.Y., H.J., Demirkhan, E., Utsumi, S., Mikami, B. and Yagi, F. (2005) Structural basis of a fungal galectin from *Agrocybe cylindracea* for recognizing sialoconjugate. *J. Mol. Biol.*, 351, 695-706.
- Beisel, H.G., Kawabata, S., Iwanaga, S., Huber, R. and Bode, W. (1999) Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*. *EMBO J.*, 18, 2313-2322.
- Bevilacqua, M., Butcher, E., Furie, B., Gallatin, M., Gimbrone, M., Harlan, J., Kishimoto, K., Lasky, L. et al., (1991) Selectins: a family of adhesion receptors. *Cell*, 67, 233.
- Bewley, C.A. (2001) Solution structure of cyanovirin-N:Man1-2Man complex: Structural basis for high-affinity carbohydrate-mediated binding to gp120. *Structure*, 9, 931-940.

- Birck, C., Damian, L., Marty-Detraves, C., Lougarre, A., Schulze-Brieze, C., Koehl, P., Fournier, D., Paquereau, L., and Samana, J.P. (2004) A new lectin family with structure similarity to actinoporins revealed by the crystal structure of *Xerocomus chrysenteron* lectin XCL. *J. Mol. Biol.*, 344, 1409-1420.
- Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R. and Wlodawer, A. (2002) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides. *J. Biol. Chem.*, 277, 34336-34432.
- Boyd, M.R., Gustafson, K.R., McMahon, J.B., Shoemaker, R.H., O'Keefe, B.R., Mori, T., Gulakowski, R.J., Wu, L., Rivera, M.I., Laurencot, C.M., and others (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 Chemoth.*, 41, 1521-1530.
- Brady, J.W., and Schmidt, R. K. (1993) The role of hydrogen bonding in carbohydrates: Molecular dynamics simulations of maltose in aqueous solution. *J. Phys. Chem.*, 1993, 958-966.
- Brockhausen, I., Yang, J.M., Burchell, J., Whitehouse, C., and Taylor-Papadimitriou, J. (1995) Mechanisms underlying aberrant glycosylation of MUC1 mucin in breast cancer cells. *Eur J Biochem*, 233, 607-617.
- Burla, M.C., Caliandro, R., Camalli, M., Carrozzini, B., Cascarano, G.L., De Caro, L., Giacovazzo, C., Polidori, G. and Spagna R. (2005) SIR2004: an improved tool for crystal structure determination and refinement. *J. Appl. Cryst.*, 38, 381-388.
- Buts, L., Loris, R., De Genst, E., Oscarson, S., Lahmann, M., Messens, J., Brosens, E., Wyns, L. and De Greve, H. and Bouckaert, J. (2003) Solving the phase problem for carbohydrate-binding proteins using selenium derivatives of their ligands: a case study involving the bacterial F17-G adhesin. *Acta Cryst.*, D59, 1012-1015.
- Carrizzo, M., Capaldi, S., Perduca, M., Irazoqui, F., Nores, G., and Monaco, H. (2005) The antineoplastic lectin of the common edible mushroom (*Agaricus bisporus*) has two binding sites, each specific for a different configuration at a single epimeric hydroxyl. *J. Biol. Chem.*, 280, 10614-10623.
- Carver, J.P. (1993) Oligosaccharides : How flexible molecules can act as signals. *Pure Appl. Chem.*, 65, 763-770.
- Chandrasekaran, R. (1997) Molecular architecture of polysaccharide helices in oriented fibers. *Adv. Carbohydr. Chem. Biochem.*, 52, 311-439.
- Crocker, P.R. (2002) Siglecs: sialic-acid-binding immunoglobulin-like lectins in cell-cell interactions and signalling. *Curr. Opin. Struct. Biol.*, 12, 609-615.
- Dam, T.K., and Brewer, C.F. (2002) Thermodynamic studies of lectin-carbohydrate interactions by isothermal titration calorimetry. *Chem. Rev.*, 102, 387-429.
- 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.

- Drickamer, K. (1988) Two distinct classes of carbohydrate-recognition domains in animal lectins. *J. Mol. Biol.*, 263, 9557-9560.
- Drickamer, K. (1996) Ca(2+)-dependent sugar recognition by animal lectins. *Biochem. Soc. Trans.*, 24, 146-150.
- Drickamer, K. (1999) C-type lectin-like domains. *Curr. Opin. Struct. Biol.*, 9, 585-590.
- Easton, R.L., Patankar, M.S., Clark, G.F., Morris, H.R. and Dell, A. (2000) Pregnancy-associated changes in the glycosylation of tamm-horsfall glycoprotein. Expression of sialyl Lewis(x) sequences on core 2 type O-glycans derived from uromodulin. *J. Biol. Chem.*, 275, 21928-21938.
- Edelman, G.M., Cunningham, B.A., Reeke, G.N., Becker, J.W., Waxdal, M.J. and Wang, J.L. (1972) The covalent and three-dimensional structure of concanavalin A. *Proc. Nat. Acad. Sci. USA*, 69, 2580-2584.
- Emsley, J., White, H.E., O'Hara, B.P., Oliva, G., Srinivasan, N., Tickle, I.J. et al. (1994) Structure of pentameric human serum amyloid P component. *Nature*, 367, 338-345.
- Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, 60, 2126-2132.
- Foadi, J., Woolfson, M.M., Dodson, E.J., Wilson, K.S., Jia-xing, Y., and Chao-de, Z. (2000) A flexible and efficient procedure for the solution and phase refinement of protein structures. *Acta Cryst. D*, 56, 1137-1147.
- Furstner, A., and Muller, T. (1998) Metathesis route to resin glycosides: formal total synthesis of tricolorin A. *J. Org. Chem.*, 63, 424-425.
- Gallego del Sol, F., Gomez, J., Hoos, S., Nagano, C., Cavada, B.S., England, P. and Calvete J.J. (2005) Energetics of 5-bromo-4-chloro-3-indolyl-a-D-mannose binding to the Parkia platycephala seed lectin and its use for MAD phasing. *Acta Cryst. F*, 61, 326-331.
- Gallego del Sol, F., Nagano, C., Cavada, B.S. and Calvete J.J. (2005) The first crystal structure of a mimosoideae lectin reveals a novel quaternary arrangement of a widespread domain. *J. Mol. Biol.*, 353, 574-583.
- Gilboa-Garber, N. (1986) Lectins of *Pseudomonas aeruginosa*: Properties, biological effects and applications. In Mirelman, D. (ed), Microbial Lectins and Agglutinins: Properties and Biological Activity. John Wiley & Sons, New York, pp. 255-269.
- Gilboa-Garber, N., Mizrahi, L., and Garber, N. (1972) Purification of the galactose-binding hemagglutinin of *Pseudomonas aeruginosa* by affinity column chromatography using sepharose. *FEBS Lett.*, 28, 93-95.
- Giollant, M., Guillot, J., Damez, M., Dusser, M., Didier, P., and Didier, E. (1993) Characterization of a lectin from *Lactarius deterrimus* - research on the possible involvement of the fungal lectin in recognition between mushroom and spruce during the early stages of mycorrhizae formation. *Plant Physiol.*, 101, 513-522.
- Guillot, J., and Konska, G. (1997) Lectins in higher fungi. *Biochem. System. Ecol.*, 25, 203-230.

- Hahn, H.P. (1997) The type-4 pilus is the major virulence-associated adhesin of *Pseudomonas aeruginosa*. *Gene*, 192, 99-108.
- Hardman, K.D., and Ainsworth, C.F. (1972) Structure of concanavalin A at 2.4 Å resolution. *Biochemistry*, 11, 4910-4919.
- Hester, G., Kaku, H., Goldstein, I.J., and Wright, C.S. (1995) Structure of mannose-specific snowdrop (*Galanthus nivalis*) lectin is representative of a new plant lectin family. *Nature Struct. Biol.*, 2, 472-479.
- Hynes, R.O. (2002) Integrins: bidirectional, allosteric signaling machines. *Cell*, 110, 673-687.
- Imberty, A. (1997) Oligosaccharide structures: theory versus experiment. *Curr Opin Struct Biol*, 7, 617-623.
- Imberty, A., Mitchell, E.P. and Wimmerová, M. (2005) Structural basis for high affinity glycan recognition by bacterial and fungal lectins. *Curr. Opin. Struct. Biol.*, 15, 525-534.
- Imberty, A., and Pérez, S. (1995) Stereochemistry of the N-glycosylation sites in glycoproteins. *Protein Engineer.*, 8, 699-709.
- Imberty, A., and Pérez, S. (2000) Structure, conformation and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem. Rev.*, 100, 4567-4588.
- Imberty, A., Wimmerowa, M., Mitchell, E.P. and Gilboa-Garber, N. (2004) Structure of lectins from *Pseudomonas aeruginosa*: insights into molecular basis for host6 glycan recognition. *Microp. Infect.*, 6, 222-229.
- 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.
- Kirkeby, S., and Moe, D. (2002) Lectin interactions with alpha-galactosylated xenoantigens. *Xenotransplantations*, 9, 260-267.
- Koca, J. (1998) Travelling through conformational space - an approach for analyzing the conformational behaviour of flexible molecules. *Progr. Biophys. Mol. Biol.*, 70, 137-173.
- Kochibe, N., and Matta, K.L. (1989) Purification and properties of a N-acetylglucosamine-specific lectin from *Psathyrella velutina* mushroom. *J. Biol. Chem.*, 264, 173-177.
- Kocourek, J., and Horejsi, V. (1981) Defining a lectin. *Nature*, 290, 188.
- Kostlanova, N., Mitchell, E.P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerowa, M. and Imberty, A. (2005) The fucose-binding lectin from *Ralstonia solanacearum*: a new type of b-propeller architecture formed by oligomerisation and interacting with fucoside, fucosyllactose and plant xyloglucan. *J. Biol. Chem.*, 280, 27839-27849.
- Kumar, S., and Muchmore, A.V. (1990) Tamm-Horsfall protein--uromodulin (1950-1990). *Kidney Int.*, 37, 1395-1401.

- Laine, R.A. (1997) The information-storing potential of the sugar code. In Gabius, H.J. and Gabius , S. (eds), Glycosciences: Status and Perspectives. Chapman & Hall, Weinheim-London, pp. 1-14.
- Lanne, B., Ciopraga, J., Bergstrom, J., Motas, C. and Karlsson, J. (1994) Binding of the galactose-specific *Pseudomonas aeruginosa* lectin, PA-I, to glycosphingolipids and other glycoconjugates. *Glycoconj. J.*, 292-298.
- Larson, D.P., and Heatcock, C.H. (1997) Total synthesis of tricolorin A. *J. Org. Chem.*, 62, 8406-8418.
- Leffler, H., Carlsson, S., Hedlund, M., Qian, Y., and Poirier, F. (2004) Introduction to galectins. *Glycoconj. J.*, 19, 433-440.
- 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.
- Lemieux, R.U., Koto, S. and Voisin, D. (1979) The exo-anomeric effect. In Szarek, W.A. and Horton, D. (eds), Anomeric effect: origin and consequence, ACS symposium Series No. 87. Amer. Chem. Soc., Washington DC, pp. 17-29.
- 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.
- Liu, Z.J., Tempel, W., Lin, D., Karaveg, K., Doyle, R.J., Rose, J.P. and Wang, B.C. (2002) Structure determination of *P. aeruginosa* lectin-1 using single wavelength anomalous scattering data from native crystals. *Am.Cryst.Assoc.,Abstr. Papers (Annual Meeting)*, 29, 99.
- 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.
- Low, D., Braaten, B. and Van der Woude, M. (1996) *Escherichia coli* and *Salmonella*: cellular and molecular biology. American Society for Microbiology PressA, Washington D.C.
- Mancheno, J., Tateno, H., Goldstein I.J., Martinez-Ripoll, M., and Hermoso, J. (2005) Structural analysis of the *Laetiporus sulphureus* hemolytic pore-forming lectin in complex with sugars. *J. Biol. Chem.*, 280, 17251-17259.
- Marchessault, R.H., and Pérez, S. (1979) Conformations of the hydroxymethyl group in crystalline aldohexopyranoses. *Biopolymers*, 18, 2369-2374.
- Merritt, E.A., and Hol, W.G.J. (1995) AB5 toxins. *Curr. Opin. Struct. Biol.*, 5, 165-171.
- Merritt, E.A., Sarfaty, S., Van Den Akker, F., L'hoir, C., Martial, J.A. and Hol, W.G.J. (1994) Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide. *Protein Sci.*, 3, 166-175.
- Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gautier, C., Pérez, S., Wu, A.M., Gilboa-Garber, N. and Imbert, A (2002) Structural basis for mediated adhesion of

- Pseudomonas aeruginosa* to the lung of cystic fibrosis patients. *Nature Struct. Biol.*, 9, 918.
- Mitchell, E.P., Sabin, S., Šnajdrová, L., Budová, M., Perret, S., Gautier, C., Hofr, C., Gilboa-Garber, N., Koca, J., Wimmerová, M. and, and Imbert, A. (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.
- Mitoma, J., Petryniak, B., Hiraoka, N., Yeh, J., Lowe, J.B. and Fukuda, M. (2003) Extended core 1 and core 2 branched O-glycans differentially modulate sialyl Lewis X-type L-selectin ligand activity. *J. Biol. Chem.*, 278, 9953-9961.
- Moore, J.S., Wu, X., Kulhavy, R., Tomana, M., Novak, J., Moldoveanu, Z., Brown, R., Goepfert, P.A. and Mestecky, J. (2005) Increased levels of galactose-deficient IgG in sera of HIV-1-infected individuals. *AIDS*, 19, 381-389.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crys. D Biol. Crystall.*, 53, 240-255.
- Ng, T.B. (2004) Peptides and proteins from fungi. *Peptides*, 25, 1055-1073.
- Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. In Carter, C.W.J. and Sweet, R.M. (eds), Macromolecular Crystallography, part A. Academic Press, New York, pp. 307-326.
- Paaventhamp, P., Joseph, J.S., Vaday, S.V.S, Robinson, H., Chua, K.Y., and Kolathar, P.R. (2003) 1.7Å structure of Fve, a member of the new fungal immunomodulatory protein family. *J. Mol. Biol.*, 332, 461-470.
- Pan, Y.T., and Elbein, A.D. (1990) Oligosaccharides chains of glycoproteins. *Progr. Drug. Res.*, 34, 163-207.
- Pemberton (1994) Agglutinins from some British higher fungi. *Mycol. Res.*, 98, 277-290.
- Pereda-Miranda, R., and Bah, M. (2003) Biodynamic constituents in the Mexican morning glories: Purgative remedies transcending boundaries. *Curr. Top. Med. Chem.*, 3, 111-131.
- Pereda-Miranda, R., Mata, R., Anaya, A.L., Wickramaratne D.B., Pezzuto J.M. and Kinghorn A.D. (1993) Tricolorin A, major phytogrowth inhibitor from *Ipomoea tricolor*. *J. Nat. Prod.*, 56, 571-582.
- Pérez, S., and Mulloy, B. (2005) Prospects for glycoinformatics. *Current Opinion in Structural Biology*, 15, 517-524.
- Pérez, S., Gautier, C. and Imbert, A. (2000) Oligosaccharide conformations by diffraction methods. In Ernst B, H.G., Sinay P (ed), Oligosaccharides in Chemistry and Biology. Wiley, pp. 969-1001.
- Perret, S., Sabin, C., Dumon, C., Pokorná, M., Gautier, C., Galanina, O., Ilia, S., Bovin, N., Nicaise, M., Desmadril, M., Gilboa-Garber, N., Wimmerova, M., Mitchell, E.P. and

- Imberty, A. (2005) Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-IIL of *Pseudomonas aeruginosa*. *Biochem. J.*, 389, 325-332.
- Rao, V.S.R., Qasba, P.K., Balaji, P.V. and Chandrasekaran, R. (1988) Conformation of Carbohydrates. Harwood academic publishers.
- Rappleye, J., Innus, M., Weeks, C. M. and Miller, R. (2002) SnB version 2.2: an example of crystallographic multiprocessing. *J. Appl. Cryst.*, 35, 374-376.
- Rudiger, H., and Gabius, H.J. (2001) Plant lectins: Occurrence, biochemistry, functions and application. *Glycoconj. J.*, 18, 589-613.
- Rutenber, E., Katzin, B.J., Ernst, S., Collins, E.J., Mlsna, D., Ready, M.P. and Robertus J.D. (1991) Crystallographic refinement of ricin to 2.5 Å. *Proteins*, 10, 240-250.
- Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A. and Vijayan, M. (1996) A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a b-prism fold. *Nature Struct. Biol.*, 3, 596-603.
- Sauer, F.G., Barnhart, M., Choudhury, D., Knight, S.D., Waskman, G. and Hultgren, S.J. (2000) Chaperone-assisted pilus assembly and bacterial attachment. *Curr. Opin. Struct. Biol.*, 10, 548.
- 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.
- Sharon, N. (1996) Carbohydrate-lectin interactions in infectious disease. *Adv. Exp. Med. Biol.*, 408, 1-8.
- Sharon, N., and Halina, Lis. (2003) Lectins. Kluwer Academic Publishers.
- She, Q.B., Ng, T.B. and Liu, W.K. (1998) A novel lectin with potent immunomodulatory activity isolated from both fruiting bodies and cultures mycelia of the edible mushroom Volvariella volvacea. *Biochemical and Biophysical Research Communication*, 247, 106-111.
- Shell, M.A. (2000) Control of virulence and pathogenicity genes of *Ralstonia solanacearum* by an elaborate sensory network. *Annu Rev Phytopathol*, 38, 263-292.
- Soto, G.E., and Hultgren, S.J. (1999) Bacterial adhesins: common themes and variations in architecture and assembly. *J. Bacteriol.*, 181, 1059-1071.
- Sudakevitz, D., Kostlanova, N., Blatman-Jan, G., Mitchell, E.P., Lerrer, B., Wimmerova, M., Katcof, f.D.J., Imberty, A., and Gilboa-Garber, N. (2004) A new *Ralstonia solanacearum* high affinity mannose-binding lectin RS-IIL structurally resembling the *Pseudomonas aeruginosa* fucose-specific lectin PA-IIL. *Mol. Microbiol.*, 52, 691-700.
- Sze, S.C.W., Ho, J.C.K. and Liu, W.K. (2004) *Volvariella volvacea* lectin activates mouse T lymphocytes by a calcium dependent pathway. *J. Cell. Biochem.y*, 92, 1193-1202.
- Tamm, I., and Horsfall, F.L. (1950) Characterization and separation of an inhibitor of viral hemagglutination present in urine. *Proc. Soc. Exp. Biol. Med.*, 74, 106-108.

- Taylor, M., and Drickamer, K. (2003) Introduction to Glycobiology. Oxford University Press.
- 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.
- Topfer-Petersen, E., Romero, E., Varela, P.F., Ekhlas-Hundrieser, M., Dostalova, Z., Sanz, L. and Calvete, J.J. (1998) Spermadhesins: a new protein family. Facts, hypotheses and perspectives. *Andrologia*, 30, 217-224.
- Tsuchiya, N., Endo, T., Matsuta, K., Yoshinoya, S., Takeuchi, F., Nagano, Y., Shiota, Furukawa, K., Kochibe, N., Ito, K. and Kobata, A. (1993) Detection of glycosylation abnormality in rheumatoid IgG using *N*-acetylglucosamine-specific *Psathyrella velutina* lectin. *J. Immunol.*, 151, 1137-1146.
- Turnbull, W.B., Precious, B. L., Homans, S. W. (2004) Dissecting the cholera toxin-ganglioside GM1 interaction by isothermal titration calorimetry. *J. Am. Chem. Soc.*, 126, 1047-1054.
- Ueda, H., Kojima, K., Saitoh, T. and Ogawa, H. (1999) Interaction of a lectin from *Psathyrella velutina* mushroom with *N*-acetylneurameric acid. *FEBS Lett.*, 448, 75-80.
- Ueda, H., Saitoh, T., Kojima, K. and Ogawa, H. (1999) Multi-specificity of a *Psathyrella velutina* mushroom lectin: heparin/pectin binding occurs at a site different from the *N*-acetylglucosamine/*N*-acetylneurameric acid-specific site. *J. Biochem.*, 126, 530-537.
- Vallet, I., Olson, J.W., Lory, S., Lazdunski 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.
- Van Damme, E.J.M., Peumans, W.J., Barre, A., and Roug  , P. (1998) Plant lectins: a composite of several distinct families of structurally and evolutionary related proteins with diverse biological properties. *Crit. Rev. Plant Sci.*, 17, 575-692.
- van Rooijen, J.J., Kamerling, J.P. and Vliegenthart, J.F. (2001) The patterns of the complex- and oligomannose-type glycans of uromodulin (Tamm-Horsfall glycoprotein) in the course of pregnancy. *Glycoconj. J.*, 18, 539-546.
- Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology*, 3, 97-130.
- Vriend, G. (1990) WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.*, 8, 52-56.
- Vyas, N.K. (1991) Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.*, 1, 732-740.
- Wagner, V.E., Bushnell, D., Passador, L., Brooks, A. I. and Iglesias, B. H. (2003) Microarray analysis of *Pseudomonas aeruginosa* quorum-sensing regulons: effects of growth phase and environment. *J. Bacteriol.*, 185, 2080-2095.
- Walker, J.R., Nagar, B., Young, N.M., Hirama, T. and Rini, J.M. (2004) X-ray crystal structure of a galactose-specific C-type lectin. *Biochemistry*, 43, 3783-3792.

- Walser, P.J., Haebel, P.W., Kunzler, M., Sargent, D., Kues, U., Aebi, M., and Ban, U. (2004) Structure and functional analysis of the fungal galectin CGL2. *Structure*, 12, 689-702.
- Whitehead, N.A., Barnard, A.M., Slater, H., Simpson, N.J. and Salmond, G.P. (2001) Quorum-sensing in Gram-negative bacteria. *FEMS Microbiol. Rev.*, 25, 365-404.
- Wiley, D.C., and Skehel, J.J. (1987) The structure and function of the hemagglutinin membrane glycoprotein of influenza. *Annu. Rev. Biochem.*, 56, 365-394.
- Wimmerova, M., Mitchell, E.P., Sanchez, J.F., Gautier, C., and Imbert, A. (2003) Crystal structure of a fungal lectin: Six-bladed b-propeller fold and a new recognition mode for Aleuria aurantia lectin. *J. Biol. Chem.*, 278, 27059-27067.
- Winzer, K., Falconer, C., Garber, N.C., Diggle, S.P., Camara, M., and Williams, P. (2000) The *Pseudomonas aeruginosa* lectins PA-IL and PA-IIL 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.
- Wright, C.S. (1980) Crystallographic elucidation of the saccharide binding mode in wheat germ agglutinin and its biological significance. *J Mol Biol.*, 141, 267-291.
- Yu, L.G., Ferning, D.G., Smith, J.A., Milton, J.D., and Rhodes J.M. (1993) Reversible inhibition of proliferation of epithelial cell lines by *Agaricus bisporus* (edible mushroom) lectin. *Cancer Res.*, 53, 4627-4632.
- Zinger-Yosovich, K., Sudakevitz, D., Imbert, 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.

Acta Crystallographica Section E

Structure Reports

Online

ISSN 1600-5368

Editors: **W. Clegg and D. G. Watson**

(4*R*,9*S*)-4-Hydroxymethyl-3,8-dioxa-1,6-diazaspiro[4.4]nonane-2,7-dithione monohydrate

Gianluca Cioci, Nicolas Leconte, Arnaud Tatibouët, Patrick Rollin, Serge Pérez and Anne Imberty

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site provided that this cover page is retained. Republication of this article or its storage in electronic databases or the like is not permitted without prior permission in writing from the IUCr.

(4*R*,9*S*)-4-Hydroxymethyl-3,8-dioxa-1,6-diaza-spiro[4.4]nonane-2,7-dithione monohydrate

Gianluca Cioci,^a Nicolas Leconte,^b Arnaud Tatibouët,^b Patrick Rollin,^b Serge Pérez^a and Anne Imbert^{a*}

^aCERMAV-CNRS (affiliated to Université Joseph Fourier), BP 53, 38041 Grenoble CEDEX 9, France, and ^bICOA–UMR 6005, Université d'Orléans, BP 6759, F-45067 Orléans, France

Correspondence e-mail: imberty@cermav.cnrs.fr

Key indicators

Single-crystal X-ray study
 $T = 293\text{ K}$
 Mean $\sigma(\text{C}-\text{C}) = 0.004\text{ \AA}$
 R factor = 0.051
 wR factor = 0.130
 Data-to-parameter ratio = 8.6

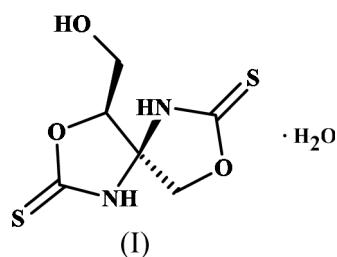
For details of how these key indicators were automatically derived from the article, see <http://journals.iucr.org/e>.

The title compound, $\text{C}_6\text{H}_8\text{N}_2\text{O}_3\text{S}_2 \cdot \text{H}_2\text{O}$, was synthesized from L-erythrulose and the structure of the enantiopure (*4R,9S*) diastereoisomer has been determined. The structure is a hydrate and the water molecules establish a hydrogen-bond network that involves the hydroxymethyl group as well as one N and one S atom.

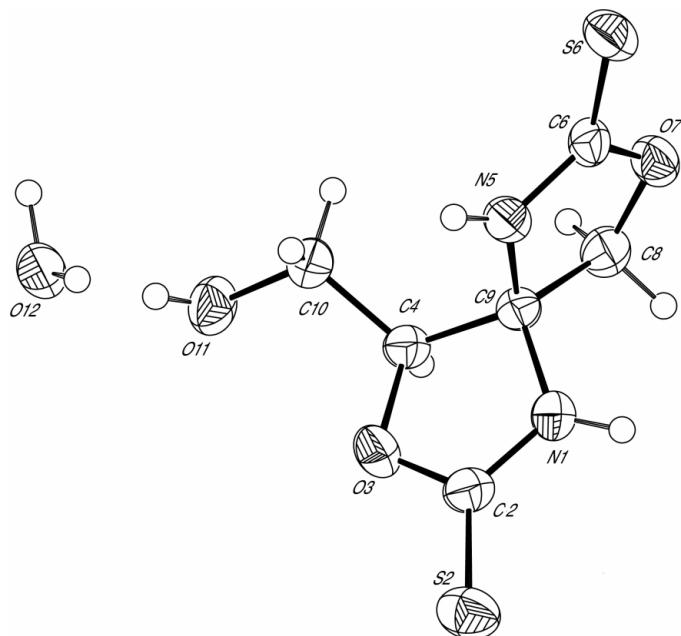
Received 21 October 2004
 Accepted 16 November 2004
 Online 27 November 2004

Comment

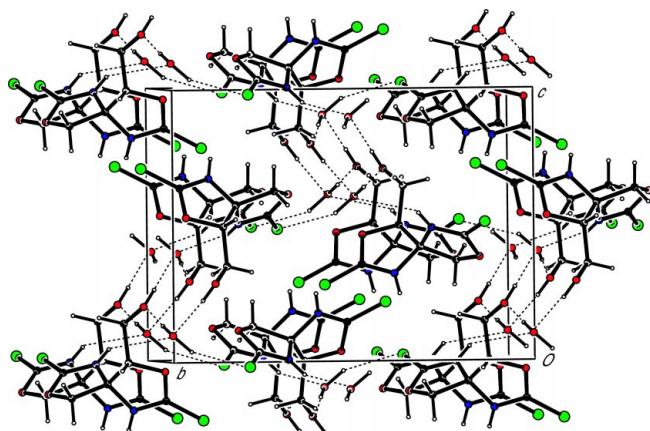
Chiral 1,3-oxazolidine-2-thiones (OZT) attract much interest because of their use as auxiliaries in asymmetric synthesis (Crimmins *et al.*, 2001), as precursors in the elaboration of nucleoside analogues (Girniene *et al.*, 2004) or as potential inhibitors of the fructose transporter GLUT-5 (Girniene *et al.*, 2003). Bis-OZT compounds, particularly those involving a spiro junction, constitute an exciting new class of multipurpose chiral OZT. The crystal structure of the parent spiro-bis-OZT in racemic form has been reported previously (Saul *et al.*, 2000). The title compound, (I), was synthesized from L-erythrulose. Two diastereoisomers are formed and the major compound, (I), readily crystallizes in the space group $P2_12_12_1$ with one OZT molecule and one water molecule in the asymmetric unit.



The two five-membered rings of (I) are significantly twisted (Fig. 1), with values of 19.8 (3) and 17.3 (3) $^\circ$ for torsion angles O3—C4—C9—N1 and N5—C9—C8—O7, respectively. The hydroxymethyl group has a *gauche* conformation relative to the ring O atom. It interacts directly with the water molecule, which also accepts a hydrogen bond from one NH group. Interestingly, this water molecule also donates one hydrogen bond to an S atom. Hydroxyl···S hydrogen bonds are rarer and weaker than their OH···O/N counterparts. Nevertheless, as reported previously (Allen *et al.*, 1997), the conjugation of C=S with the neighbouring N atom increases the electronegativity of the S atom and results in a resonance-induced hydrogen bond, with S···O 3.234 (3) \AA and S···H 2.45 (3) \AA . All hydrogen bonds are bridged by the water molecules, resulting in a three-dimensional framework (Fig. 2).

**Figure 1**

The molecular structure of (I), showing the atom-numbering scheme. Displacement ellipsoids are drawn at the 50% probability level and H atoms are shown as small spheres of arbitrary radii.

**Figure 2**

A packing diagram for (I), with hydrogen bonds indicated as dashed lines.

Experimental

The title compound was prepared in quantitative yield according to the procedure described by Saul *et al.* (2000). The reaction gave a mixture of two diastereoisomers in the ratio 85:15, from which the major compound crystallized in pure form. Suitable crystals of (I) were obtained by recrystallization from water (m.p. 442–447 K). Spectroscopic analysis: $[\alpha]_D = +22$ ($c = 1.0$; MeOH); ^1H NMR (DMSO- d_6 , δ , p.p.m.): 3.71 (m , 2H, CH_2OH), 4.62 (d , 1H, $^2J = 11.3$ Hz, H8A), 4.66 (d , 1H, $^2J = 11.0$ Hz, H8B), 4.80 (t , 1H, $^3J = 5.8$ Hz, H4), 5.30 (s , 1H, OH), 10.91 (s , 1H, NH), 10.97 (s , 1H, NH); ^{13}C NMR (DMSO- d_6 , δ , p.p.m.): 57.9 (CH_2OH), 76.0 (C8), 81.3 (C9), 85.1 (C4), 186.9 (C2 or C6), 188.0 (C2 or C6); MS: m/z 221 [$M+\text{H}]^+$; high-resolution MS, calculated for $\text{C}_6\text{H}_8\text{N}_2\text{O}_3\text{S}_2$: 219.9976; found: 219.9989.

Crystal data

$\text{C}_6\text{H}_8\text{N}_2\text{O}_3\text{S}_2 \cdot \text{H}_2\text{O}$
 $M_r = 238.28$
Orthorhombic, $P2_12_12_1$
 $a = 7.036$ (2) Å
 $b = 10.336$ (3) Å
 $c = 13.814$ (3) Å
 $V = 1004.6$ (5) Å 3
 $Z = 4$
 $D_x = 1.575$ Mg m $^{-3}$

Cu $K\alpha$ radiation
Cell parameters from 24 reflections
 $\theta = 20.2\text{--}23.5^\circ$
 $\mu = 4.80$ mm $^{-1}$
 $T = 293$ (2) K
Prism, colourless
0.7 × 0.2 × 0.2 mm

Data collection

Enraf–Nonius CAD-4 diffractometer
 ω scans
Absorption correction: none
1164 measured reflections
1164 independent reflections
1150 reflections with $I > 2\sigma(I)$

$\theta_{\max} = 75.0^\circ$
 $h = 0 \rightarrow 8$
 $k = 0 \rightarrow 12$
 $l = 0 \rightarrow 17$
2 standard reflections every 120 reflections
intensity decay: 3.1%

Refinement

Refinement on F^2
 $R[F^2 > 2\sigma(F^2)] = 0.051$
 $wR(F^2) = 0.130$
 $S = 1.08$
1164 reflections
136 parameters
H atoms treated by a mixture of independent and constrained refinement

$w = 1/[\sigma^2(F_o^2) + (0.0629P)^2 + 0.1989P]$
where $P = (F_o^2 + 2F_c^2)/3$
 $(\Delta/\sigma)_{\max} < 0.001$
 $\Delta\rho_{\max} = 0.51$ e Å $^{-3}$
 $\Delta\rho_{\min} = -0.37$ e Å $^{-3}$
Absolute structure: Flack (1983); no Friedel pairs
Flack parameter = 0.04 (3)

Table 1
Hydrogen-bonding geometry (Å, °).

$D\cdots H\cdots A$	$D-H$	$H\cdots A$	$D\cdots A$	$D-H\cdots A$
O11—H11···O12	0.82	1.94	2.746 (4)	169
N5—H5···O12 ⁱ	0.86	1.99	2.791 (4)	155
O12—H121···S6 ⁱⁱ	0.84 (2)	2.45 (3)	3.234 (3)	156 (5)
O12—H122···O11 ⁱ	0.83 (4)	1.86 (2)	2.685 (4)	178 (5)

Symmetry codes: (i) $\frac{1}{2} + x, \frac{3}{2} - y, 1 - z$; (ii) $1 - x, \frac{1}{2} + y, \frac{1}{2} - z$.

All H atoms were initially located in a difference Fourier map. Atoms H121 and H122 were refined freely. All other H atoms were treated as riding on their parent atoms, with C—H = 0.97–0.98 Å, O—H = 0.82 Å and N—H = 0.86 Å, and with $U_{\text{iso}}(\text{H}) = 1.5U_{\text{eq}}(\text{O})$ or $1.2U_{\text{eq}}(\text{C}, \text{N})$. Although there were no Friedel pairs, the absolute configuration could be determined unambiguously.

Data collection: *CAD-4 Software* (Enraf–Nonius, 1989); cell refinement: *CAD-4 Software*; data reduction: *CAD-4 Software*; program(s) used to solve structure: *SHELXS97* (Sheldrick, 1990); program(s) used to refine structure: *SHELXL97* (Sheldrick, 1997); molecular graphics: *PLATON* (Spek, 2003); software used to prepare material for publication: *PLATON*.

GC was supported by an EEC postdoctoral grant (HPRN-CT2000-00001). The excellent assistance of Christian Philouze is acknowledged.

References

- Allen, F. A., Bird, C. M., Rowland, R. S. & Raithby, P. R. (1997). *Acta Cryst.* **B53**, 680–695.

- Crimmins, M. T., King, B. W., Tabet, E. A. & Chaudhary, K. (2001). *J. Org. Chem.* **66**, 894–902.
- Enraf–Nonius (1989). *CAD-4 Software*. Enraf–Nonius, Delft, The Netherlands.
- Flack, H. D. (1983). *Acta Cryst. A* **39**, 876–881.
- Girniene, J., Apremont, G., Tatibouët, A., Sackus, A. & Rollin, P. (2004). *Tetrahedron*, **60**, 2609–2619.
- Girniene, J., Tatibouët, A., Sackus, A., Yang, J., Holman, G. D. & Rollin, P. (2003). *Carbohydr. Res.* **338**, 711–719.
- Saul, R., Kern, T., Kopf, J., Pinter, I. & Köll, P. (2000). *Eur. J. Org. Chem.* pp. 205–209.
- Sheldrick, G. M. (1990). *Acta Cryst. A* **46**, 467–473.
- Sheldrick, G. M. (1997). *SHELXL97*. University of Göttingen, Germany.
- Spek, A. L. (2003). *J. Appl. Cryst.* **36**, 7–13.