



**HAL**  
open science

# Ingénierie de lectines de valence, topologie et spécificité contrôlées pour la biologie cellulaire et la biotechnologie

Julie Arnaud

## ► To cite this version:

Julie Arnaud. Ingénierie de lectines de valence, topologie et spécificité contrôlées pour la biologie cellulaire et la biotechnologie. Biochimie, Biologie Moléculaire. Université de Grenoble, 2014. Français. NNT : 2014GRENV029 . tel-01553432

**HAL Id: tel-01553432**

**<https://theses.hal.science/tel-01553432>**

Submitted on 3 Jul 2017

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

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

## THÈSE

Pour obtenir le grade de

## DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Chimie et Biologie**

Arrêté ministériel : 7 août 2006

Présentée par

**Julie ARNAUD**

Thèse dirigée par **Anne Imberty**

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

# Ingénierie de lectines de valence, topologie et spécificité contrôlées pour la biologie cellulaire et la biotechnologie

Thèse soutenue publiquement le **28 novembre 2014**,  
devant le jury composé de :

**M. Franck FIESCHI**

Professeur des Universités, France, Président du jury

**Mme Muriel DELEPIERRE**

Directeur de Recherche, France, Rapporteur

**M. Charles TELLIER**

Professeur des Universités, France, Rapporteur

**Mme Anne Imberty**

Directeur de Recherche, France, Directeur de thèse











A ma maman, à mon papa  
Et tous ceux qui ont cru en moi...



## REMERCIEMENTS

La partie remerciement est à la fois la partie la plus simple et la plus compliquée. La plus simple car on a juste à laisser parler notre cœur qui dicte chacun de ces mots, mais la plus compliquée car l'émotion est bien présente.

Je remercie les membres du jury, le **Pr. Charles Tellier** et le **Dr. Muriel Delepierre** qui ont accepté de juger ce travail de thèse avec autant d'assiduité.

Je remercie tout particulièrement le **Pr. Franck Fieschi** pour avoir accepté de présider cette thèse. Avec toi auront commencé mes études de biologie et avec toi elles se termineront. Lors d'une interview de toi que j'ai fait dans le cadre d'une UE à l'université tu m'avais dit qu'un de tes professeurs qui t'avait donné cette vocation. Moi aussi j'ai trouvé un mentor, un professeur des universités avec une grande qualité d'enseignant et de chercheur, comme il en existe peu. Je suis vraiment très honorée que tu aies accepté de présider ce jury. Merci Franck.

Je remercie bien évidemment ma directrice de thèse le **Dr. Anne Imberty** pour m'avoir offert l'opportunité de réaliser ma thèse sur un sujet passionnant. Je te remercie également pour m'avoir permis de présenter mes travaux un peu partout dans le monde.

Je remercie les membres de l'**équipe glycobio**logie moléculaire pour leur aide et leur sympathie.

Je remercie le **Dr. Aymeric Audfray** pour m'avoir initié à mon sujet de thèse en m'encadrant durant mon master.

Je remercie le **Dr. Annabelle Varrot** pour m'avoir fait découvrir la cristallographie et m'avoir amené à plusieurs reprises à l'ESRF.

Je remercie **Valérie Chazalet** qui avec beaucoup de patience et de pédagogie est toujours là pour nous aider.

Je remercie également **Félix** pour tous les moments scientifiques ou non que l'on a passé ensemble ; pour nos discussions scientifiques sur nos projets professionnels, sur nos idées, nos envies, et bien sûr nos Startup ! Je suis heureuse d'avoir croisé ton chemin et surtout ne change rien.

Je remercie **Oliver** pour nos discussions plus ou moins scientifiques. Je suis contente de t'avoir vu plusieurs fois au cheval, soit pour venir voir ma jument avec ta petite famille, soit vous croiser en concours à Jarry, mais aussi avoir collé des bridons et des selles sur le livre de ta fille pour te montrer dans quel sens ils allaient. Je pense que Félix a beaucoup de

## REMERCIEMENTS

chance d'être encadré par quelqu'un comme toi, car tu es plein de qualités professionnelles et personnelles.

Je remercie également le **Pr. Christelle Breton**, pour sa présence et son aide, mais aussi pour m'avoir fait confiance en me confiant dès ma première année de monitorat une classe de TD.

Je remercie aussi tout particulièrement **Aline**, merci beaucoup d'avoir été là dans les bons moments et également dans les mauvais moments. Quand ça n'allait pas je n'avais qu'une porte à pousser pour trouver du soutien et cela je ne suis pas prête de l'oublier. Comme on dit, tu es une bonne âme, juste en parlant un peu tout allait mieux. Tu as été un de mes plus grands soutiens durant cette thèse, toi qui ne me connaissais pas il y a à peine 4 ans. Tu es aujourd'hui une amie, quelqu'un sur qui je peux compter et quelqu'un qui peut compter sur moi à tous les niveaux. Je te souhaite également de trouver beaucoup de bonheur dans ton nouvel envol !

Comment ne pas remercier ma petite **Dörte**. Toi aussi tu es devenue une amie. Je n'oublierai jamais ta présence et ton aide, mais aussi notre visite de l'ESRF. Toi qui m'as fait faire une rencontre surprenante, en nous amenant dans un restaurant végétarien et bio... Pour moi tu es une très bonne chercheuse et je souhaite que tu t'épanouisses vraiment dans ce travail que tu as choisi. J'espère qu'on retournera manger bio.

Je remercie également tous les autres membres de l'équipe, **Emilie, Aude, Joana, Serge, Géraldine, Nico, Anita** pour leur aide et leur gentillesse.

Je remercie les personnes avec qui j'ai collaboré en Allemagne et qui m'ont initié à des nouvelles techniques, en particulier le **Jun. Pr. Winfried Römer**, le **Dr. Julie Claudinon** et **Tröndle Kevin**. J'ai découvert grâce à vous des nouvelles techniques de biologie cellulaire, mais aussi un autre pays et une autre ambiance de travail. Merci à Winfried qui m'a soutenu dans les moments difficiles.

Je remercie **Phillipe Chinkirch** et toute l'équipe **Valoridoc**. Cette formation est bien plus qu'un simple cours, c'est une véritable bouffée d'oxygène. Ce sont des personnes passionnées et passionnantes qui croient en vous et vous prouvent que vous pouvez faire de grandes choses. Quand je suis arrivée à la formation je n'avais pas le moral et plus confiance en moi. J'ai fait cette formation car je ne désirai pas travailler dans le public et je voulais avoir des informations sur le privé. Je suis ressortie dès le premier jour avec une vraie confiance en moi et motivée. Ces personnes sont des magiciens qui aiment tellement les gens qu'ils nous

## REMERCIEMENTS

transmettent bien plus que des connaissances. Je ne suis pas prête de vous quitter car j'aurais besoin de vous pour trouver du travail. Et vous pouvez compter sur moi pour faire désormais partie du réseau.

Je remercie tous les chercheurs que j'ai côtoyé durant ma thèse et qui m'ont fait confiance en me confiant leurs étudiants. Merci **Sylvie Armand, Ludovic Pelosi, Françoise Cornillon...**

Je voudrais remercier les membres de ma famille qui ont cru en moi alors que moi-même je n'y croyais pas. Je remercie **Luc, Gaë, mémé, Manu, Stéph, Emeline, Gilles, Christelle, Théo**, tata **Michette** et tonton **Germain, Isabelle, Jacky, Manon, Marine, Elisa**, sans oublier **Josiane** et **Claude** pour leur grande aide...que j'attends encore et bien sûr tous les autres...

Je voudrais dire un grand merci à **mémé** qui était si fière de moi avant de revivre ces 17 ans...

Je remercie particulièrement ma super petite cousine **Cindy** qui m'a envoyé un super bouquet de fleurs quand j'en avais besoin ! Des gens comme toi sont précieux. Merci d'être là.

Je remercie **Fred** et **Mireille** qui m'ont écrit durant ma thèse pour me soutenir. Et **Perrine** pour ces petits mails.

Comment ne pas parler aussi de ma super **Christine** qui elle aussi a toujours été là durant ma thèse et même bien avant. Toi qui as pris le temps de m'aider aussi pour cette thèse : Merci Christine.

Je remercie aussi de tout mon cœur **Baby** et **Nath** qui eux aussi sont des gens géniaux, sur qui on peut compter quoi qu'il arrive ! Et bien sûr **Mel** et **Nana** qui sont chaque fois qu'on les voit une tornade de bonheur et de rire. Tornade il faut bien le dire car quand la famille Marchand arrive ce n'est pas sur la pointe des pieds mais qu'est-ce que c'est bon de vous voir.

**Claude** et **Alain**, un gros merci à vous aussi qui m'avez soutenu, accompagné durant toutes ces années ! Merci d'avoir toujours été là et d'avoir cru en moi. Vous avez tous les deux contribué à ma réussite ! Merci à vous !

Ma chère **maman**, comment trouver les mots pour te dire tout ce que j'ai à te dire... C'est face au devoir d'écrire ce chapitre que l'on se rend compte à quel point une thèse est plus facile à rédiger que les remerciements. Que te dire à part merci ! Merci d'avoir cru en moi dès le 26 avril 1986 et même avant. Merci d'avoir été là quand j'ai grandi et que je



## REMERCIEMENTS

n'étais plus cette petite fille gentille et sage. Je pourrais faire une seconde thèse sur tout l'amour et le respect que je te porte mais mes nombreuses années d'études m'ont appris à être synthétique alors je ne peux que te dire MERCI. Toutes les deux nous savons dans nos cœurs toutes les choses pour lesquelles je te remercie.

**Julien**, toi tu étais là bien avant ma thèse et tu seras là bien après. Il y a trois ans les rôles étaient inversés. Tu rédigeais ta thèse et moi je t'aidais en me disant : j'ai encore du temps pour la mienne. Et bien le temps passe très vite et me voilà en fin de rédaction. Une thèse, c'est difficile même très difficile, il y a beaucoup de moments de doutes... J'ai souvent perdu confiance en moi, je me trouvais nulle, je pensais ne pas être à ma place, ne pas avoir le niveau, je trouvais les choses injustes... mais toi tu as toujours cru en moi et tu m'as aidé à me battre ! Tu m'as supporté dans les bons et mauvais moments, avec gentillesse et amour. Grâce à toi, je rentre à la maison et je peux souffler ! Merci pour tout Julien...

Je voudrais remercier **Gaëlle**, que j'ai connu dès la L1 de biologie. Tu as été ma super binôme de choc pendant toutes ces années avec des moments cultes comme « tu continues » que nous seules pouvons comprendre. Tu as été et tu es toujours mon déstressant le plus efficace ! Devant les salles de tous les oraux où tu m'as accompagné, que ce soit des « petits oraux » à la fac ou l'oral de l'école doctorale où tu es venue uniquement pour me soutenir, tu avais un petit moment pour me faire rire. En tout cas tu es vraiment quelqu'un de bien, sur qui on peut compter, toujours prête à rire.

Et je veux aussi remercier tous mes autres amis ou copains, je ne citerais pas tous, mais vous savez bien que vous avez tous une jolie place dans mon cœur. Merci **Ramo**, **Bastien**, **Déb**, **Emilie**, **Arnaud** et tous les autres ...

J'aimerais remercier aussi mes animaux car eux aussi ont largement contribué à ce travail ! Et ils peuvent être d'une aide aussi précieuse que les humains.

Merci à mes chiens, **Uline**, **Etna** et **Bobine** pour m'avoir fait sortir de chez moi quand je n'en avais pas envie et m'avoir redonné le sourire quand je ne l'avais plus. Vous n'avez pas besoin de mots pour vous exprimer et pas besoin de mots pour reconforter ! Un merci particulier à Bobine qui n'est malheureusement plus là mais dont le soutien est toujours dans mon cœur.

Merci aussi à ma lapine **Coquine** qui aimait tant manger ses barreaux pour que je la sorte pendant que j'essayais de travailler.

## REMERCIEMENTS

Et un merci particulier à ma jument **Un Amour d'Ick** qui est mon rêve réalisé au milieu de ma thèse. Merci toi aussi de me changer chaque jour les idées, soit avec des séances magiques soit en me faisant stresser pour ta santé car tu as toujours un petit truc de travers qui prend dans mon esprit des proportions inimaginables, soit tout simplement en étant là à côté de moi !

Mais aussi ceux qui ne sont plus là mais qui seraient fières de moi : en premier mon **Papa**, mais aussi pépé **Francis**, pépé **Dédé** et ma chère **Mamy** qui n'auront pas attendu de me voir soutenir.

Ma dernière pensée ira aux patients atteints de cancer, j'espère par ma modeste contribution avoir permis à la recherche d'avancer dans ce domaine pour sauver des vies.







---

# **TABLE DES MATIERES**

---



---

# TABLE DES MATIERES

---

TABLE DES MATIERES.....	3
LISTE DES ABRÉVIATIONS .....	11
LISTE DES FIGURES .....	15
LISTE DES TABLEAUX .....	21
INTRODUCTION.....	27
I. Lectines et glycanes.....	27
I.1. Les glycoconjugués.....	27
I.2. Glycoprotéines .....	28
I.2.1. <i>N</i> -glycosylation.....	29
I.2.2. <i>O</i> -glycosylation .....	30
I.2.2.1. Les mucines.....	31
I.2.2.2. Protéoglycanes .....	33
I.3. Glycolipides .....	34
I.4. Oligosaccharides particuliers : les antigènes fucosylés des groupes sanguins et tissulaires.....	35
I.4.1. Les groupes sanguins ABO et Lewis.....	35
I.4.2. La fucosylation .....	36
I.4.2.1. $\alpha$ 1-2/3/4 fucosylation .....	37
I.4.2.2. $\alpha$ 1-6 fucosylation.....	37
I.4.2.3. <i>O</i> -fucosylation.....	38
I.5. Conformations.....	39
I.5.1. Généralités .....	39
I.5.2. Exemple des Lewis X et Sialyl-Lewis X.....	41
II. Les lectines – structures et diversités.....	42



II.1. Classification des lectines .....	43
II.1.1. Lectines de plantes .....	43
II.1.2. Lectines d'invertébrés .....	44
II.1.3. Lectines de vertébrés .....	45
II.1.4. Lectines microbiennes .....	46
II.2. Rôle des lectines .....	48
II.2.1. Reconnaissance et adhésion .....	48
II.2.2. Trafic cellulaire .....	50
II.2.3. Transduction du signal .....	51
II.2.4. Immunité.....	52
II.2.5. Chaperonne.....	53
II.3. Site de reconnaissance .....	54
II.3.1. Interactions lectine-sucre.....	54
II.3.2. Stabilisation des interactions par des molécules d'eau.....	55
II.3.3. Stabilisation des interactions par des ions .....	56
II.4. Lectines et les autres molécules reconnaissant les sucres.....	58
II.4.1. Introduction .....	58
II.4.2. Les molécules reconnaissant les sucres .....	58
II.5. Utilisation des lectines en biotechnologie et médecine .....	75
II.5.1. Changement de glycosylation et maladies chroniques.....	75
II.5.2. Glycosylation et cancer .....	76
II.5.2.1. N-glycosylation et cancer.....	76
II.5.2.2. O-glycosylation et cancer.....	77
II.5.2.3. Glycolipides et cancer .....	78
II.5.3. Antigènes Lewis, fucosylation et Cancer .....	79
II.5.3.1. Généralités.....	79
II.5.3.2. $\alpha$ 1-6 fucosylation.....	80

II.5.3.3.	α1-2 fucosylation.....	80
II.5.3.4.	α1-3/4 fucosylation .....	82
II.6.	Utilisation des lectines en recherche, diagnostic et thérapie.....	83
II.6.1.	Histopathologie.....	83
II.6.2.	Puces à lectines .....	86
II.6.3.	Vectorisation.....	87
II.6.4.	Agents thérapeutiques .....	88
II.6.5.	Compréhension du trafic Cellulaire.....	90
II.7.	Ingénierie des lectines .....	91
II.7.1.	Modification de spécificité .....	91
II.7.1.1.	Lectines de plantes .....	91
II.7.1.2.	Lectines de champignons et d'invertébrés .....	92
II.7.1.3.	Lectines de vertébrés.....	93
II.7.1.4.	Lectines de bactéries .....	95
II.7.2.	Modification de valence : .....	95
III.	Objectif et Conception de néolectines .....	96
III.1.	Béta-propeller .....	97
III.2.	Les lectines en béta-propeller .....	98
III.3.	Ingénierie de béta-propeller .....	100
III.4.	RSL : un béta-propeller trimérique hexavalent.....	101
IV.	Stratégie proposée.....	103
V.	Méthodologie.....	104
V.1.	Biologie moléculaire et biochimie .....	104
V.1.1.	Mutagenèse dirigée.....	104
V.1.2.	Production et purification des lectines .....	104
V.2.	Caractérisation .....	105
V.2.1.	Caractérisation biochimique.....	105

V.2.1.1.	Microcalorimétrie de Titration Isotherme .....	105
V.2.1.2.	Résonance Plasmonique de Surface.....	108
V.2.1.3.	Diffusion Dynamique de la Lumière.....	109
V.2.1.4.	Calorimétrie Différentielle à Balayage .....	110
V.2.1.5.	Criblage de Thermostabilité par Fluorescence.....	111
V.2.1.6.	Amarrage moléculaire.....	112
V.2.2.	Caractérisation structurale .....	112
V.2.2.1.	Cristallogénèse .....	112
V.2.2.2.	Résolution d'une structure cristallographique .....	114
RESULTATS	.....	121
I.	Contrôle de la valence .....	121
I.1.	Modification de la valence de la protéine RSL.....	121
I.1.1.	Introduction .....	121
I.1.2.	La réduction de la valence d'une lectine changes drastiquement la dynamique des membranes mais pas d'avidité de surface .....	123
I.2.	Construction d'une protéine RSL monomérique et modification de sa valence et topologie .....	143
I.2.1.	Introduction .....	143
I.2.2.	Déformation des membranes par des néolectines avec des sites de fixations ingénierés.....	145
II.	Changement de spécificité.....	167
II.1.	Validation de la modélisation pour la spécificité des lectines bactériennes .....	167
II.1.1.	Introduction .....	167
II.1.2.	Décrypter la préférence glycanes des lectines bactériennes par amarrage moléculaire .....	169
II.2.	Spécificité de RSL pour les épitopes Le <sup>x</sup> .....	191
II.2.1.	Introduction .....	191

II.2.2. Etude de la spécificité de RSL et néoRSL_VI .....	192
II.2.2.1. Différence de spécificité entre les sites intra et intermonomériques de RSL .....	193
II.2.2.2. Structures cristallographiques des complexes RSL/ Le <sup>x</sup> et RSL/ SLe <sup>x</sup> .....	196
II.2.2.3. Comparaison entre sites intra et inter de RSL.....	201
II.2.2.4. Comparaison avec BamBL et autres lectines à fucose .....	203
II.2.3. Discussion.....	205
II.2.4. Matériels et Méthodes .....	206
II.2.4.1. Purification des protéines .....	206
II.2.4.2. ITC .....	206
II.2.4.3. Structure cristallographique .....	207
DISCUSSION et PERSPECTIVES .....	213
BIBLIOGRAPHIE .....	219
ANNEXES .....	253
I. Brevet (Confidentiel).....	253



---

## **LISTE DES ABREVIATIONS**

---



---

# LISTE DES ABRÉVIATIONS

---

- AAL** : Lectine d'*Aleuria aurantia* 1
- ADN** : Acide désoxyribonucléique
- AFL** : Lectine d'*Aspergillus fumigatus*
- AFP** : Alphafoetoprotéine
- $\alpha$ MeFuc** :  $\alpha$ -methylfucoside
- AOL** : Lectine d'*Aspergillus oryzae*
- BambL** : Lectine de *Burkholderia ambifaria*
- CD-MPR** : Récepteur du mannose-6-phosphate dépendant du calcium
- CEL-I** : Lectine de *Cucumaria echinata*
- ConA** : Concanavaline A
- CRD** : Domaine de reconnaissance des carbohydrates
- DBA** : Agglutinine de *Dolichos biflorus*
- DLS** : Diffusion dynamique de la lumière
- DOPC** : 1,2-Dioleoyl-sn-glycero-3-Phosphatidylcholine
- DSC** : Calorimétrie différentielle à balayage
- DTT** : Dithiothréitol
- EGF** : Facteur de croissance épidermal
- Fuc** : Fucose
- GAG** : Glycosaminoglycane
- Gal** : Galactose
- GalNAc** : *N*-acétylgalactosamine
- Glc** : Glucose
- GlcNAc** : *N*-acétylglucosamine
- GNA** : Agglutinine de *Galanthus nivalis*
- GPI** : Glycosylphosphatidylinositol
- GUV** : Vésicules géantes unilamellaires
- HeLa** : Cellules Henrietta Lacks
- HPA** : Agglutinine de *Helix pomatia*



**HS** : Héparane sulfate  
**Ig** : Immunoglobuline  
**IPTG** : Isopropyl  $\beta$ -D-1-thiogalactopyranoside  
**ITC** : Microcalorimétrie de titration isotherme  
**LacNAc** : N-acetyllactosamine (Gal $\beta$ 1-4GlcNAc)  
**Le** : Lewis  
**LecA** : Lectine 1 de *Pseudomonas aeruginosa*  
**LecB** : Lectine 2 de *Pseudomonas aeruginosa*  
**LNFP III** : Lacto-N-fucopentaose III  
**MAH** : Hémagglutinine de *Maackia amurensis*  
**MAL** : Leukoagglutinine de *Maackia amurensis*  
**Man** : Mannose  
**MBL** : Lectine de fixation au mannose  
**NeuAc** : acide N-acetylneuraminique  
**NK** : Cellules tueuses naturelles  
**PCR** : Amplification en chaîne par polymérase  
**PNA** : Lectine d'arachide (*Arachis hypogaea*)  
**PSA** : Agglutinine de *Pisum sativum*  
**PHA** : Agglutinine de *Phaseolus vulgaris*  
**PVL** : Lectine de *Psathyrella velutina*  
**RCA** : Agglutinine de *Ricinus communis*  
**RE** : Réticulum endoplasmique  
**REr** : Réticulum endoplasmique rugueux  
**RMN** : Résonance magnétique nucléaire  
**RSL** : Lectine de *Ralstonia solanacearum*  
**SLe** : Sialyl Lewis  
**SPR** : Résonance plasmonique de surface  
**SSA** : Agglutinine de *Sambucus sieboldiana*  
**STxB** : Toxine de Shiga  
**SV40** : Virus simien 40  
**TSA** : Criblage de Thermostabilité par Fluorescence  
**VAA1** : Agglutinine 1 de *Viscum album*  
**VNTR** : Nombre variable d'unités de répétition  
**WGA** : Agglutinine de *Wheat germ*

---

## **LISTE DES FIGURES**

---



---

# LISTE DES FIGURES

---

Figure 1 : Schéma d'une membrane cellulaire présentant différents glycoconjugués. ....	27
Figure 2 : Représentation schématique de protéines glycosylées .....	28
Figure 3 : Voie de synthèse des N-glycanes dans le réticulum endoplasmique et l'appareil de Golgi (Helenius and Aebi, 2001). ....	29
Figure 4 : Représentation schématique des trois types de N-glycanes chez l'Homme. ....	30
Figure 5 : Description des mucines présentes à la surface des cellules. ....	32
Figure 6 : Représentation schématique de protéoglycanes. ....	33
Figure 7 : Représentation schématique du sphingolipide de groupe sanguin A.....	34
Figure 8 : Représentation schématique des antigènes de groupes sanguins de type ABO et Lewis .....	36
Figure 9 : Représentation schématique d'un $\alpha$ -L-fucose .....	37
Figure 10 : Synthèse des antigènes Lewis.....	38
Figure 11 : Différentes conformations d'un sucre en solution : exemple du glucose .....	39
Figure 12 : Représentation des angles de torsion définissant la conformation du maltose.....	40
Figure 13 : Carte d'énergie pour la liaison $\text{Fu}\alpha\text{1-3GlcNAc}$ . ....	40
Figure 14 : Structure du Lewis x à l'état cristallin .....	41
Figure 15 : Conformation du $\text{SLe}^x$ complexé à la Sélectine E.....	42
Figure 16 : Représentation schématique de la fixation et du changement de conformation de l'hémagglutinine à la surface de la membrane plasmique de la cellule hôte. ....	49
Figure 17 : Trafic cellulaire de la Shiga Toxine.....	50
Figure 18 : Reconstitution de la formation des tubules dans des GUVs.....	51
Figure 19 : Représentation schématique de l'activation d'un NK via la lectine NKR-P1.....	52
Figure 20 : Principales lectines capables d'activer le système du complément.....	53
Figure 21 : Site de reconnaissance de la protéine RSL. ....	55
Figure 22 : Importance des molécules d'eau dans la stabilisation du sucre dans le site de fixation .....	56
Figure 23 : Importance des ions dans la stabilisation du sucre dans le site de fixation de LecB .....	57

Figure 24 : Les antigènes T et Tn.....	78
Figure 25 : Glycosylations des suites d'une activité aberrante de la $\alpha$ 1-2 fucosyltransférase .	81
Figure 26 : Représentation schématique du Globo H .....	81
Figure 27 : Représentation schématique du LNFP III.....	82
Figure 28 : Marquages positif (A) et négatif (B) par HPA biotinylé d'adénocarcinomes du poumon.....	84
Figure 29 : Schémas représentant le principe d'une puce à lectine.....	86
Figure 30 : Exemple d'un résultat de puce à lectine réalisé sur des cellules souches.....	87
Figure 31 : Structure de différents $\beta$ -propeller .....	97
Figure 32 : Alignement des séquences des sites de fixations des principales lectines bactériennes .....	100
Figure 33 : Structure cristallographique de RSL en présence d' $\alpha$ MeFuc. ....	102
Figure 34 : Principe de la mutagenèse dirigée .....	104
Figure 35 : Schéma d'un calorimètre .....	106
Figure 36 : Données de l'ITC.....	107
Figure 37 : Principe de la résonance plasmonique de surface (SPR).....	108
Figure 38 : Description d'un sensorgramme de SPR .....	109
Figure 39 : Données types de DSC .....	110
Figure 40 : Principe du TSA .....	112
Figure 41 : Description d'une goutte pendante .....	113
Figure 42 : Acides aminés de la protéine RSL principalement impliqués dans la fixation du fucose dans le site intramonomérique. ....	122
Figure 43 : Une diminution de la valence de la protéine RSL induit un changement drastique dans la capacité de la lectine de former des invaginations.....	123
Figure 44 : Représentation schématique de la protéine RSL sauvage et de la lectine néoRSL_IV.....	143
Figure 45 : Représentation schématique des 13 néoRSL de valence et topologie créées .....	144
Figure 46 : Effet de la topologie sur la capacité des lectines à invaginer les membranes. ....	145
Figure 47 : Structure cristallographique des protéines LecB et BamBL .....	168
Figure 48 : Corrélation entre l'énergie libre mesurée expérimentalement et celle calculée par modélisation moléculaire pour la protéine BamBL .....	169
Figure 49 : Conformation fermée du Le <sup>x</sup> .....	192
Figure 50 : Points communs et différences entre les sites intra et intermonomérique.....	194
Figure 51 : Représentation schématique des protéines néoRSL_III135 et néoRSL_III246 ..	195

Figure 52 : Comparaison de la fixation du $Le^x$ et du $SLe^x$ dans les sites intra et intermonomérique. ....	195
Figure 53 : Structure des complexes $RSL/SLe^x$ et $RSL/Le^x$ .....	197
Figure 54 : Sites intra et intermonomériques du complexe $RSL/Le^x$ .....	198
Figure 55 : Sites intra et intermonomériques du complexe $RSL-SLe^x$ .....	199
Figure 56 : $Le^x$ et $SLe^x$ présent dans les sites de fixation de $RSL$ .....	200
Figure 57 : Alignement des sites de fixation intra et intermonomérique des $RSL$ .....	201
Figure 58 : Comparaison des sites de fixation intra et intermonomérique.....	202
Figure 59 : Acides aminés proches du sucre différents entre les sites intra et intermonomériques.....	202
Figure 60 : Conformation du $Le^x$ et $SLe^x$ en solution et en complexe avec des protéines ....	203
Figure 61 : Carte d'énergie des Lewis dans les sites de fixation de $BambL$ et $RSL$ .....	204
Figure 62 : Structure cristallographique et modélisation des interactions $RSL/Le^x$ .....	205



---

## **LISTE DES TABLEAUX**

---





---

# LISTE DES TABLEAUX

---

Tableau 1 : Noms et structures des 7 familles de O-glycosylation .....	31
Tableau 2 : Noms et structures majeures du core des glycosphingolipides des vertébrés .....	35
Tableau 3 : Exemples de familles de lectines .....	46
Tableau 4 : Principaux glycoconjugués fucosylés dont la surexpression est retrouvée dans des cancers .....	83
Tableau 5 : Principales lectines commerciales .....	85
Tableau 6 : Exemples de lectines végétales utilisées pour leur activité sur les cellules cancéreuses .....	89
Tableau 7 : Exemples d'ingénierie de spécificité de lectines .....	94
Tableau 8 : Différentes protéines avec une structure en Béta-propeller. ....	98
Tableau 9 : Lectines de la famille de RSL et AAL .....	99
Tableau 10 : Comparaison de la spécificité de RSL et de néoRSL_VI .....	192
Tableau 11 : Comparaison de l'affinité des sites intra et intermonomériques le Le <sup>x</sup> et le SLe <sup>x</sup> par ITC .....	196
Tableau 12 : Angle de torsion Phi et Psi des Le <sup>x</sup> et SLe <sup>x</sup> retrouvé dans les sites de RSL .....	200
Tableau 13 : Concentrations en protéines et en sucres utilisées lors des ITC avec RSL et néoRSL_VI .....	207







---

# **INTRODUCTION**

---



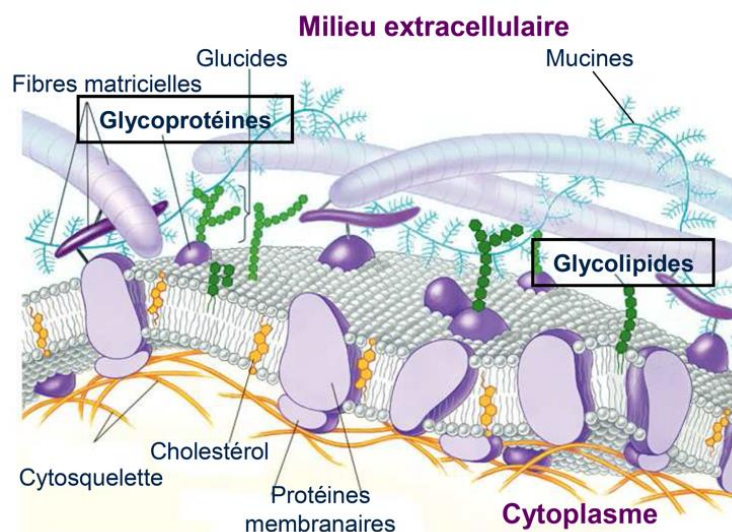
# INTRODUCTION

Le code génétique est essentiellement le même dans tous les organismes vivants connus, et plusieurs fonctions de base telles que la transcription des gènes et la production d'énergie ont tendance à être conservées. Bien que les glycanes soient présents dans tous les organismes, une diversité considérable de leurs structures et expressions est observée dans la nature, à la fois dans et entre les lignées évolutives. En partie en raison des difficultés inhérentes à leur étude, de nouvelles structures de glycanes restent encore à découvrir.

## I. Lectines et glycanes

### I.1. Les glycoconjugués

La glycosylation est la modification co- et post-traductionnelle la plus abondante. Il s'agit d'un ensemble de réactions enzymatiques réalisées par des glycosyltransférases qui transfèrent un sucre activé sur une protéine, un lipide ou un autre saccharidique.



*Figure 1 : Schéma d'une membrane cellulaire présentant différents glycoconjugués.*

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

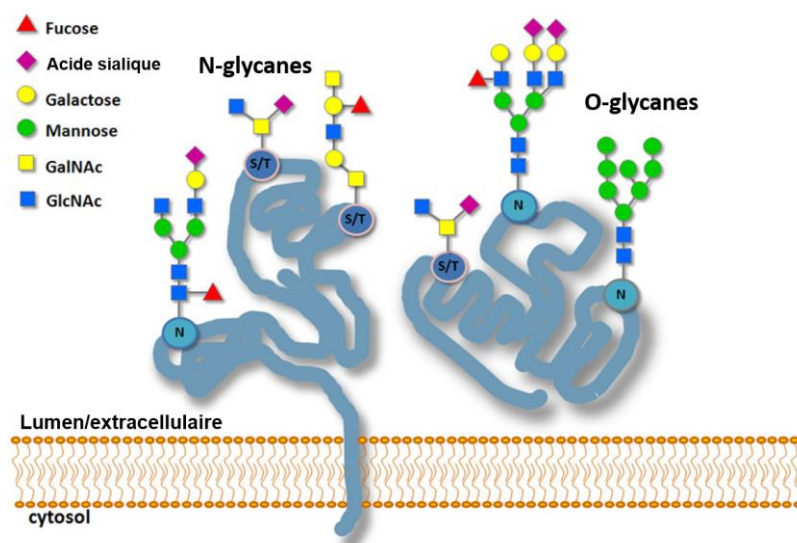


Ce répertoire de glycoconjugués est retrouvé sous forme d'une couche épaisse à la surface des membranes plasmiques et est appelée la « glyocalix » (Figure 1). Ces glycanes présents à la surface des cellules sont des marqueurs cellulaires très importants. Ils ont un rôle, entre autre, dans l'adhésion cellules-cellules ; la reconnaissance cellules-hôtes/pathogènes, l'endocytose, le trafic cellulaire et la transduction du signal mais aussi dans l'immunité avec la discrimination du « soi » et du « non soi » (Ohtsubo and Marth, 2006).

Il existe une très grande diversité de glycoconjugués et de glycanes mais nous nous intéresserons uniquement ici qu'aux sucres humains.

## I.2. Glycoprotéines

Plus de 70 % des protéines humaines sont glycosylées. Concernant les glycoprotéines, deux grandes familles sont définies selon la nature de la liaison covalente établie entre le glycanes et la liaison peptidique que l'on nomme *N*- ou *O*-glycosylation (Figure 2).

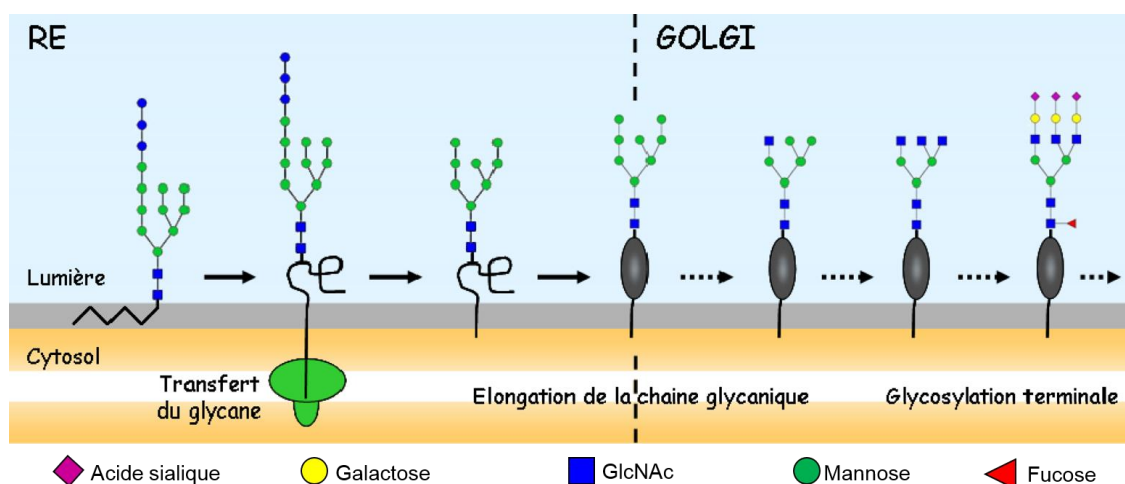


**Figure 2 : Représentation schématique de protéines glycosylées**

Figure issue de : <https://www.neb.com/applications/glycobiology#tabselect4>.

### I.2.1. N-glycosylation

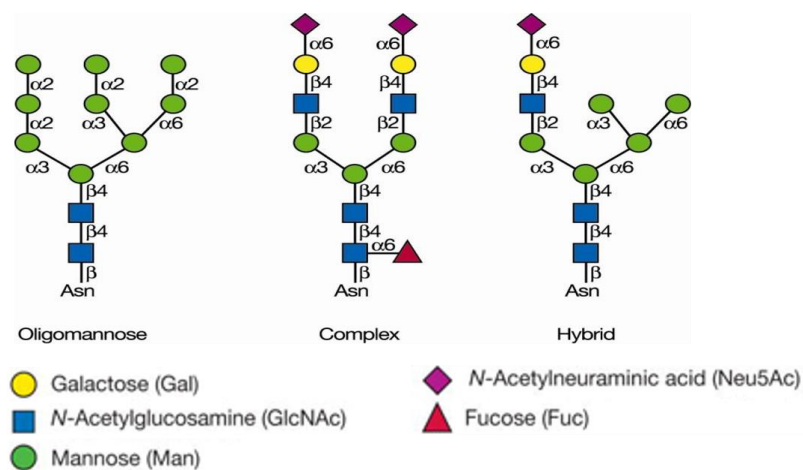
La N-glycosylation débute par une modification co-traductionnelle qui a lieu dans le réticulum endoplasmique. Elle consiste à l'ajout d'un oligosaccharide sur un résidu asparagine de la protéine ayant pour conséquence la formation d'une liaison N-glycosidique. L'asparagine doit être présente dans une séquence consensus de type N-X-S/T où le X peut être n'importe lequel des acides aminés excepté la proline. Cinq liaisons différentes de N-glycanes ont été observées, la plus fréquente étant la liaison d'une N-acétylglucosamine sur une asparagine (GlcNAc $\beta$ 1-Asn). Les N-glycanes sont constitués d'une structure pentasaccharidique commune formée de deux résidus GlcNAc et de trois résidus mannose. La N-glycosylation se poursuit par une modification post-traductionnelle qui a lieu dans l'appareil de Golgi et consiste en une série complexe de réactions catalysées par des glycosidases et glycosyltransférases membranaires. Cette seconde étape est la maturation des N-glycanes de la protéine (Figure 3). Cette maturation conduit à la formation de trois types de N-glycosylation : le premier, dit de type oligomannosidique possède le noyau Man<sub>3</sub>GlcNAc<sub>2</sub> additionné de résidus mannose. Le second, le type complexe, comporte sur les antennes des résidus galactoses et GlcNAc et qui peuvent posséder de plus des résidus fucoses et se terminer par un résidu acide sialique. Un résidu fucose peut également être lié au premier résidu GlcNAc proche de l'asparagine. Le troisième type de N-glycane est le type hybride qui possède les caractéristiques des N-glycanes de type oligomannosidique sur une branche et ceux de type complexe sur une autre branche (Figure 4).



**Figure 3 : Voie de synthèse des N-glycanes dans le réticulum endoplasmique et l'appareil de Golgi (Helenius and Aebi, 2001).**

Les images ont été réalisées sur la plateforme GlycoWork Bench (Helenius and Aebi, 2001).

Une même protéine peut contenir plusieurs séquences consensus de *N*-glycosylation qui sont toutes des sites potentiels de modification post-traductionnelle. Ces sites ne sont en revanche pas forcément tous glycosylés, de plus plusieurs structures glycaniques différentes peuvent être observées. La conformation de la protéine mais aussi le type cellulaire et l'état physiologique de la cellule peuvent influencer la glycosylation en jouant sur les différentes étapes réticulaires et golgiennes de la *N*-glycosylation et en favorisant ou non l'accessibilité des sites potentiels de glycosylation. La plupart des enzymes responsables de la *N*-glycosylation sont sensibles à l'état physiologique de la cellule dans laquelle la glycoprotéine est exprimée. Ainsi, les populations de sucres fixées à une glycoprotéine dépendront du type de cellule dans laquelle la glycoprotéine est exprimée et de l'état physiologique de la cellule. Cet état qui peut être régulé au cours du développement et de la différenciation et peut être altéré dans les états pathologiques (Goochee and Monica, 1990; Hubbard and Ivatt, 1981).



**Figure 4 : Représentation schématique des trois types de N-glycanes chez l'Homme.**

(*Essentials of GlycoBiology 2<sup>nd</sup> Edition*)

### I.2.2. O-glycosylation

La *O*-glycosylation est une modification covalente des résidus sérine ou thréonine des glycoprotéines. Elle a lieu dans l'appareil de Golgi et ne débute qu'après le repliement des protéines. Chaque monosaccharide est ajouté de manière séquentielle, par des glycosyltransférases spécifiques, au niveau de régions VNTR (Variable Number Tandem

Repeat) riches en sérine (Ser) ou thréonine (Thr). Ce type de glycosylation est très complexe puisqu'il comprend chez l'homme 7 familles selon la nature du glycanes *O*-lié (Tableau 1).

<u>Type de <i>O</i>-glycanes</u>	<u>Structure</u>	<u>Glycoprotéines</u>
Mucine	(R)-GalNAc $\alpha$ 1-Ser/Thr	Membranaires et sécrétées
GAG	(R)-GlcA $\beta$ 1-3Gal $\beta$ 1-3Gal $\beta$ 1-4Xyl $\beta$ 1-Ser	Protéoglycanes
<i>O</i> -GlcNAc	GlcNAc $\beta$ 1-Ser/Thr	Cytoplasmiques et nucléaires
<i>O</i> -Gal	Glc $\alpha$ 1-2 $\pm$ Gal $\beta$ 1-O-Lys	Collagènes
<i>O</i> -Man	NeuAc $\alpha$ 2-3Gal $\beta$ 1-4GlcNAc $\beta$ 1-2Man $\alpha$ 1-Ser/Thr	Dystroglycane
<i>O</i> -Glc	Xyl $\alpha$ 1-3Xyl $\alpha$ 1-3 $\pm$ Glc $\beta$ 1-Ser	Domaine EGF
<i>O</i> -Fuc	NeuAc $\alpha$ 2-6Gal $\beta$ 1-4GlcNAc $\beta$ 1-3 $\pm$ Fuc $\alpha$ 1-Ser/Thr	Domaine EGF

**Tableau 1 : Noms et structures des 7 familles de *O*-glycosylation**

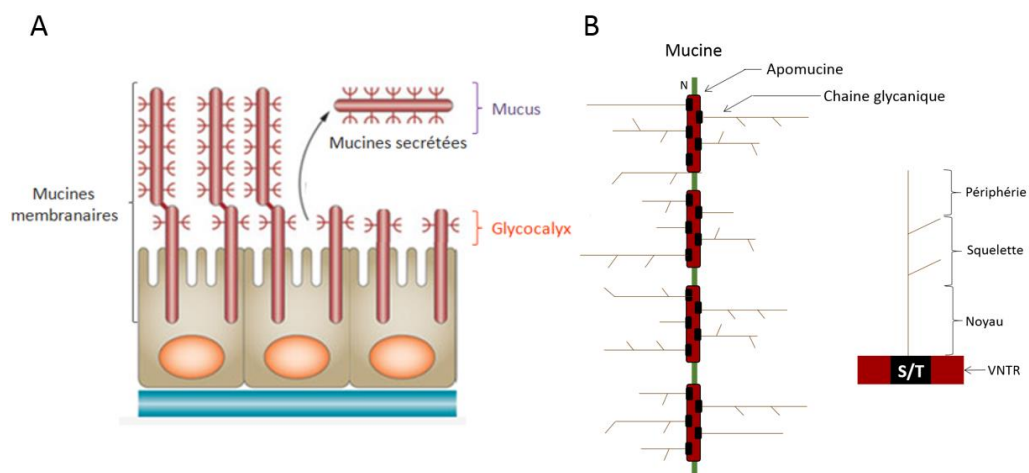
La *O*-glycosylation de type mucine où la liaison se fait par l'intermédiaire d'une *N*-acétylgalactosamine (GalNAc) qui se lie en  $\alpha$  au OH de la sérine ou de la thréonine par une liaison *O*-glycosidique ( $\alpha$ GalNAc-O-Ser/Thr). Mais il existe aussi la glycosylation de type *O*-GlcNAc ( $\beta$ GlcNAc-O-Ser/Thr), la *O*-fucosylation ( $\alpha$ Fuc-O-Ser/Thr), la *O*-glucosylation ( $\alpha$ Glc-O-Ser/Thr) ou la *O*-mannosylation ( $\alpha$ Man-O-Ser/Thr). Cette complexité est également retrouvée au niveau de l'élongation de l'oligosaccharide qui sera très variable selon le type cellulaire et l'état physiologique des cellules (Ohtsubo and Marth, 2006). Trois parties peuvent être distinguées dans la structure de ces glycanes : une structure de base le noyau (ou core), une structure d'élongation constituée d'éléments répétitifs formant l'ossature du glycanes qui peut être linéaire ou ramifiée et, enfin, des structures terminales ou périphériques comprenant une plus grande variété de sucres ainsi qu'éventuellement des résidus sulfates (Guillot *et al.*, 2004). La combinaison de ces trois éléments aboutit à une quasi infinité de structures possibles.

### I.2.2.1. Les mucines

Les mucines sont des glycoprotéines très fortement *O*-glycosylées. Elles sont trouvées dans les sécrétions des muqueuses, ou en tant que glycoprotéines transmembranaires de la surface des cellules avec les glycanes exposés à l'environnement extérieur. Les mucines

sécrétées peuvent être sous forme de polymères (mucines de formation de gel) ou sous forme monomériques (mucines solubles) (Figure 5). De nombreuses cellules épithéliales produisent des mucines, mais la formation de gel de mucines est due principalement aux cellules trachéo-bronchiques, gastro-intestinales, et génito-urinaires.

Chaque mucine possède un cœur protéique appelé apomucine, codé par le gène MAC, sur lequel se greffent des chaînes glycosidiques (Figure 5). La principale caractéristique de cette apomucine est de présenter des tronçons peptidiques répétés riches en sérines ou thréonines. Plus de 80% du poids moléculaire d'une mucine peut être dû à sa glycosylation. Les *O*-glycanes de type « mucine » ne sont pas préassemblés, mais chaque sucre est transféré de manière individuelle et séquentielle à partir d'un nucléotide-sucre. Après le transfert du premier résidu GalNAc en configuration  $\alpha$  sur l'hydroxyle de la chaîne latérale d'une sérine ou d'une thréonine (Ser/Thr-O- $\alpha$ GalNAc), une cascade de réactions enzymatiques orchestrées par différentes enzymes allonge ensuite le *O*-glycane de manière à former plusieurs structures de cœur ou noyau, pouvant eux-mêmes être modifiés par sialylation, sulfatation ou acétylation. Les extrémités non réductrices de ces chaînes peuvent contenir différentes structures de carbohydrates comme les épitopes glucidiques des groupes sanguins ABO (antigène A, B, H et H-sulfaté) et Lewis (Lewis a, Lewis b, Lewis x, Lewis y, sialyl et/ou sulfo Lewis a et Lewis x) qui sont décrits plus bas.



**Figure 5 : Description des mucines présentes à la surface des cellules.**

A) Représentation du rôle des mucines dans la protection cellulaire (Niv and Fass, 2012). Le mucus est constitué des mucines secrétées qui forment la première ligne de défense. Les mucines membranaires constituent une deuxième ligne de défense.

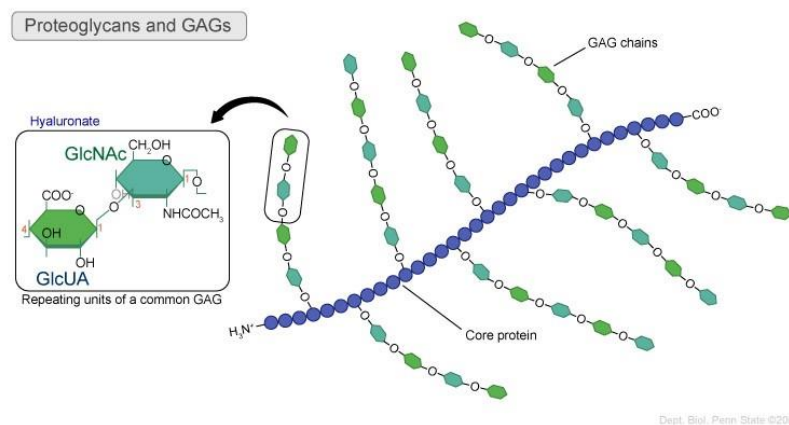
B) Schéma illustrant la structure d'une mucine. L'apomucine comporte des unités de répétition (VNTR) riches en sérine et thréonine où sont attachés des *O*-glycanes.

Les chaînes glycaniques confèrent aux mucines une très grande diversité. Elles peuvent être impliquées dans les phénomènes d'adhésion cellulaire, dans le contrôle de l'antigénicité et du système immunitaire, mais les mucines constituent surtout une barrière protectrice (Lamblin *et al.*, 2001).

### I.2.2.2. Protéoglycanes

Les protéoglycanes sont constitués d'un noyau protéique et d'une ou plusieurs chaînes de glycosaminoglycanes (GAG) fixées de manière covalente. Les chaînes de glycosaminoglycanes sont beaucoup plus longues que celles des autres types de glycanes. Ces GAGs sont des polysaccharides non branchés, chargés négativement, de hauts poids moléculaires, composés d'un squelette d'unités disaccharidiques répétées contenant un sucre aminé (GlcNAc ou GalNAc) et un acide uronique (Figure 6). Les GAGs sont principalement localisés dans la matrice extracellulaire.

Les GAGs sont classés en plusieurs catégories selon la nature du motif glycanique : les chondroïtines sulfates et les dermatanes sulfates contiennent du GalNAc, les héparane-sulfates (HS) et les kératane-sulfates contiennent du GlcNAc. Certaines modifications comme la *N*-déacétylation-*N*-sulphatation, la 2-*O*- et la 6-*O*-sulphatation et l'épimérisation de l'acide uronique apportent une grande variété au sein des chaînes GAGs.



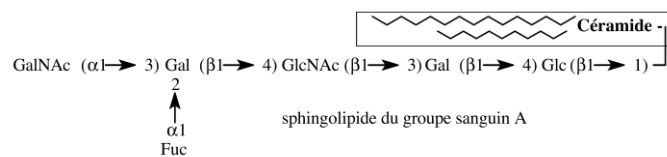
**Figure 6 : Représentation schématique de protéoglycanes.**

(<https://wikispaces.psu.edu/download/attachments/46924793/image-1.jpg> Dept. Biol. Penn State © 2003)

### I.3. Glycolipides

Des motifs glycosidiques peuvent également être retrouvés sur des lipides. Selon la nature des lipides, les glycolipides sont classés en deux groupes : les glyco-glycérolipides et les glycosphingolipides.

Les glyco-glycérolipides sont très rares dans le règne animal. Les glycosphingolipides sont quant à eux ubiquitaires et donc retrouvés dans les membranes cellulaires d'organismes allant des bactéries à l'Homme. Ils sont les principaux glycanes du cerveau des vertébrés, où plus de 80% des glycoconjugués sont sous la forme de glycolipides (Muthing et al., 1994). Ils sont composés d'une partie hydrophobe appelée céramide, et d'une partie hydrophile composée d'un monosaccharide ou oligosaccharide lié par un groupe hydroxyle aux céramides (Figure 7). La partie glucidique des glycolipides est synthétisée dans l'appareil de Golgi où des glycosyltransférases spécifiques catalysent l'ajout des monosaccharides. Il existe 7 groupes majeurs répartis dans les séries ganglio-, isoganglio-, lacto-, néolacto-, globo- et isoglobo-céramides (Tableau 2) (Takizawa et al., 1999).



**Figure 7 : Représentation schématique du sphingolipide de groupe sanguin A**

*Il est composé de lipides et de céramides hydrophobes (encadré) et d'une partie hydrophile composée de sucre.*

Les glycolipides, localisés dans la membrane plasmique, ont des rôles biologiques variés dans la structure de la membrane, mais aussi dans les interactions cellules-hôtes/pathogènes, la reconnaissance cellules/cellules, et la modulation de la fonction des protéines de la membrane (Lingwood, 2011). Les glycosylphosphatidylinositols sont une famille distincte de glycolipides qui se fixent de façon covalente à des protéines et servent de points d'ancrage dans la membrane.

Nom	Structure core
Lacto	Gal $\beta$ 3GlcNAc $\beta$ 3Gal $\beta$ 4Glc $\beta$ 1Céramide
Neolacto	Gal $\beta$ 4GlcNAc $\beta$ 3Gal $\beta$ 4Glc $\beta$ 1Céramide
Globo	GalNAc $\beta$ 3Gal $\alpha$ 4Gal $\beta$ 4Glc $\beta$ 1Céramide
Isoglobo	GalNAc $\beta$ 3Gal $\alpha$ 3Gal $\beta$ 4Glc $\beta$ 1Céramide
Ganglio	Gal $\beta$ 3GalNAc $\beta$ 4Gal $\beta$ 4Glc $\beta$ 1Céramide
Muco	Gal $\beta$ 3Gal $\beta$ 3Gal $\beta$ 4Glc $\beta$ 1Céramide
Gala	Gal $\alpha$ 4Gal $\beta$ 1Céramide

*Tableau 2 : Noms et structures majeures du core des glycosphingolipides des vertébrés*

## I.4. Oligosaccharides particuliers : les antigènes fucosylés des groupes sanguins et tissulaires

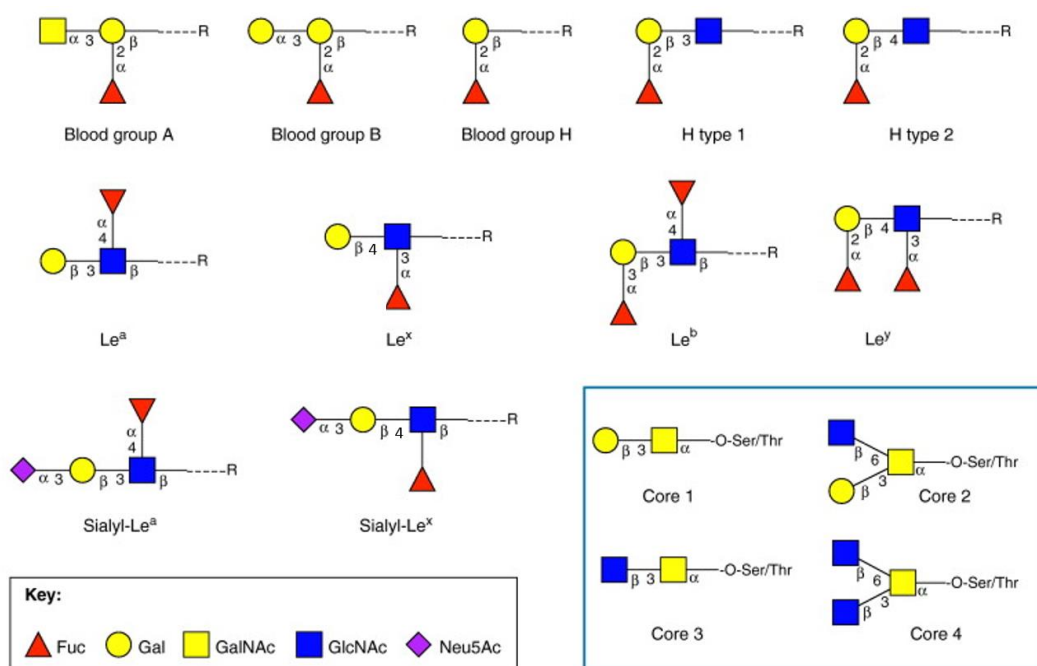
### I.4.1. Les groupes sanguins ABO et Lewis

Le système des groupes sanguins ABO a été découvert en 1900 par Karl Landsteiner, après avoir mélangé les sérums et hématies de six de ses collègues et observé des hémagglutinations (Landsteiner, 1900). L'année suivante, il démontre l'existence de deux antigènes A et B et des anticorps correspondants anti-A et anti-B (Landsteiner, 1901). Les cellules peuvent présenter soit l'antigène A, soit le B, soit aucun des deux. Il démontre également qu'un individu ne présente pas d'anticorps dirigés contre les épitopes présents à la surface de ses propres hématies et que les deux anticorps sont présents dans le sérum des individus de groupe O. Ces travaux lui ont valu le prix Nobel de médecine en 1930. Le phénotype de groupe sanguin AB, pour lequel les deux antigènes A et B sont présents à la surface des hématies et dont le sérum ne contient pas d'anticorps, a été découvert en 1902 par Decastello et Sturli (von Decastello and Sturli, 1902). En 1924, Berstein démontre que la transmission des groupes sanguins est héréditaire et suit les lois de Mendel (Berstein, 1924). Le système ABO fait partie des déterminants antigéniques et est présent à la surface des cellules sanguines mais également à la surface des cellules endothéliales. De plus, les épitopes ABO sont présents dans les sécrétions (salive, mucus, lait) de 80% de la population de phénotype « sécréteur » (Figure 8).



Un autre système de groupe sanguin, le système Lewis est découvert en 1946 par Mourant (Mourant, 1946). Il s'agit d'antigènes glycosidiques qui sont soit libres dans le plasma et autres liquides biologiques comme la salive, soit retrouvés à la surface des érythrocytes. Ils sont aussi présents sur les mucines et les *N*-glycanes. Ce système comporte six antigènes définis par six anticorps (Figure 8).

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



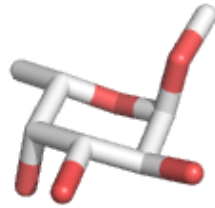
**Figure 8 : Représentation schématique des antigènes de groupes sanguins de type ABO et Lewis**

(Juge, 2012)

### I.4.2. La fucosylation

Le L-fucose (6-désoxy-L-galactose) est un monosaccharide commun à plusieurs *N*- et *O*-glycanes et aux glycolipides produits par les cellules de mammifères (Figure 9). L'addition des résidus fucoses sur une chaîne oligosaccharidique est généralement une modification terminale mais le fucose peut-être aussi présent sur le premier GlcNAc du cœur des *N*-

glycannes. La *O*-fucosylation directe sur les résidus sérines/thréonines des protéines a récemment été mise en évidence (Moloney and Haltiwanger, 1999).



**Figure 9 : Représentation schématique d'un  $\alpha$ -L-fucose**

*En rouge sont représentés les oxygènes et en blanc les carbones – les hydrogène ne sont pas représentés.*

Le rôle important des glycanes fucosylés a été démontré dans le développement (Candelier *et al.*, 1993; Candelier *et al.*, 2000; Clarke and Watkins, 1996; Oriol *et al.*, 1992), la réponse inflammatoire (Blander *et al.*, 1999; Huang *et al.*, 2000; Lowe, 1997), la fertilisation (Johnston *et al.*, 1998) et la formation des groupes sanguins (Oriol, 1995). Une modification de la fucosylation de divers glycanes a été observée au cours de maladies chroniques (Miyoshi *et al.*, 1999) et dans le cancer (Hakomori, 1999). Des récepteurs d'adhésion cellulaire fucosylés reconnus par des selectines ont un rôle fonctionnel dans la réponse inflammatoire (Vestweber and Blanks, 1999).

#### **I.4.2.1. $\alpha$ 1-2/3/4 fucosylation**

L'enzyme FUT3 est responsable de l'ajout de résidus fucoses sur les précurseurs de type 1 et 2 (Gal $\beta$ 1-3GlcNAc et Gal $\beta$ 1-4GlcNAc respectivement) en  $\alpha$ 1-3 ou  $\alpha$ 1-4 aboutissant à la synthèse des trisaccharides Le<sup>a</sup> ou Le<sup>x</sup>. De même l'enzyme FUT2 permet par adjonction d'un fucose en  $\alpha$ 1-2 d'obtenir les antigènes dits sécréteurs sur lesquels l'action de la FUT3 permet la synthèse des tétrasaccharides Le<sup>b</sup> et Le<sup>y</sup>, respectivement (Figure 10).

#### **I.4.2.2. $\alpha$ 1-6 fucosylation**

Chez les mammifères, la liaison du fucose en  $\alpha$ 1-6 sur le GlcNAc proximal d'un *N*-glycane est très courante au point que l'on a tendance à la qualifier d'ubiquitaire. Une surexpression de la  $\alpha$ 1-6-fucosyltransférase est observée dans certaines pathologies

oncologiques, telles que le cancer ovarien, les hépatomes et hépato-carcinomes ainsi que dans les maladies chroniques du foie (Miyoshi *et al.*, 1999).

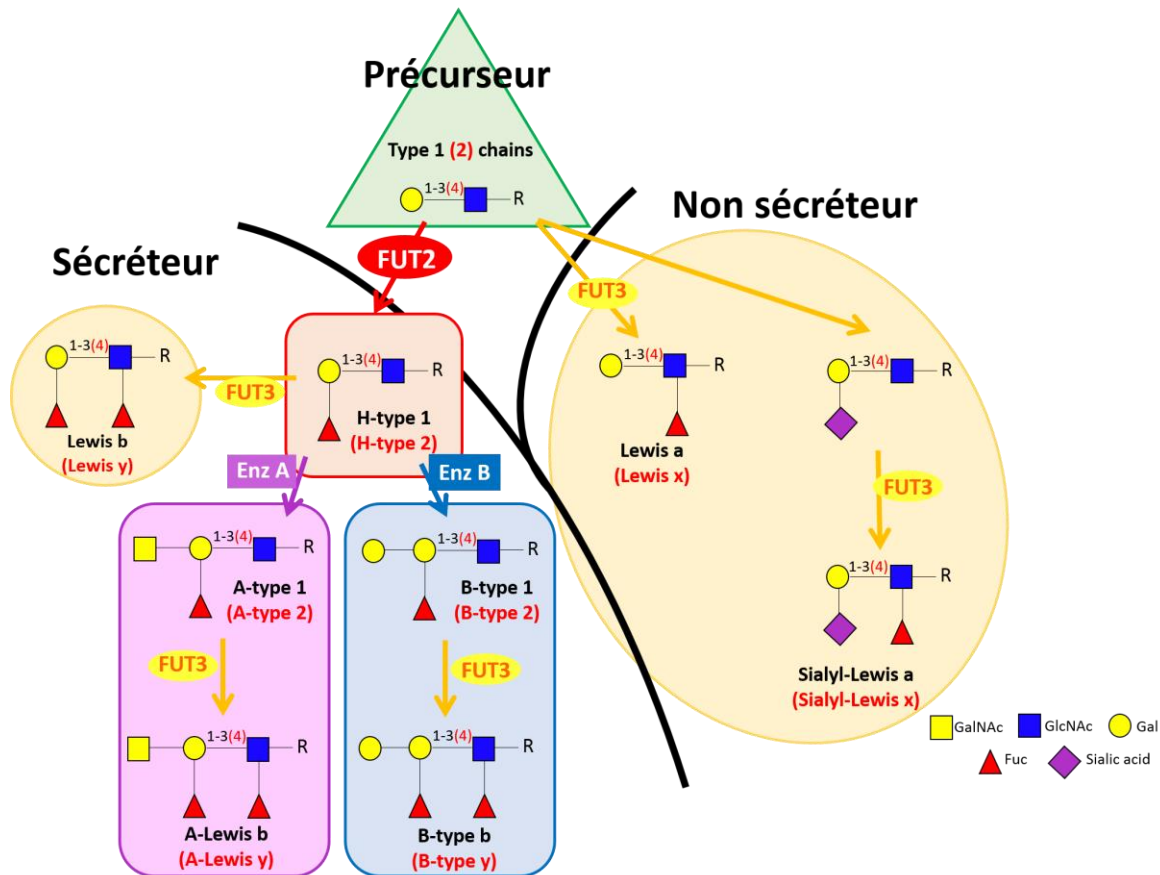


Figure 10 : Synthèse des antigènes Lewis

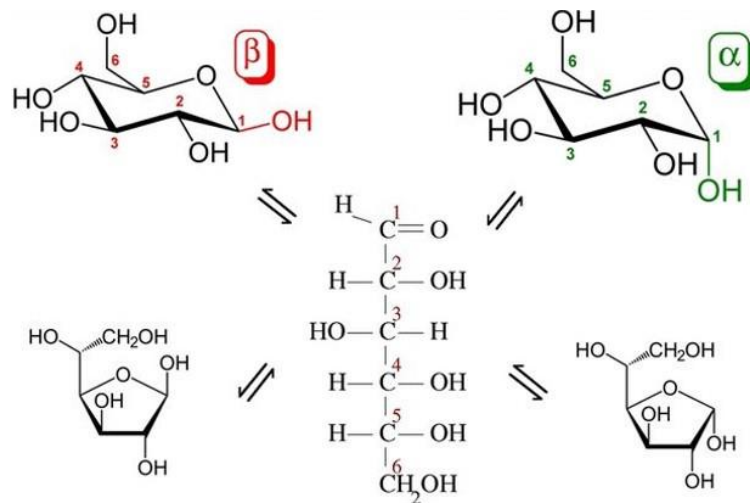
#### I.4.2.3. O-fucosylation

La O-fucosylation a été décrite chez de nombreuses espèces animales en particulier chez l'Homme. La O-fucosylation a été d'abord observée sur des acides aminés ou peptides isolés dans l'urine humaine (Hallgren *et al.*, 1975). La première protéine portant des O-fucoses identifiée est l'activateur du plasminogène de type urinaire (Kentzer *et al.*, 1990). Ensuite, plusieurs protéines O-fucosylées participant à la coagulation ont été identifiées. Le fucose O-lié est susceptible dans certains cas d'évoluer vers une structure plus complexe, suite à l'ajout d'acide sialique, de galactose et de N-acétylglucosamine (Wang *et al.*, 2001). La O-fucosylation, qui demeure une modification peu commune des protéines a été découverte sur deux domaines protéiques extracellulaires : les répétitions EGF-like et les répétitions TSR (Thrombospondin type 1 Repeat).

## I.5. Conformations

### I.5.1. Généralités

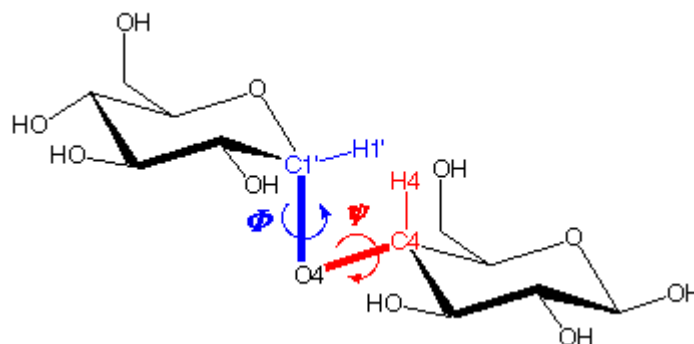
Les monosaccharides en solution peuvent exister sous formes cycliques ou linéaire. La cyclisation s'opère par réaction de la fonction aldéhydique ou cétonique avec une des fonctions alcools de la chaîne (réaction d'hémiacétalisation). Les cycles habituellement rencontrés résultent de la réaction avec la fonction OH portée par le cinquième ou le quatrième carbone. On trouvera donc principalement des hémiacétals cycliques à six atomes ou cycles pyranose et des cycles à 5 atomes aussi appelés cycles furanoses (Figure 11).



**Figure 11 : Différentes conformations d'un sucre en solution : exemple du glucose**

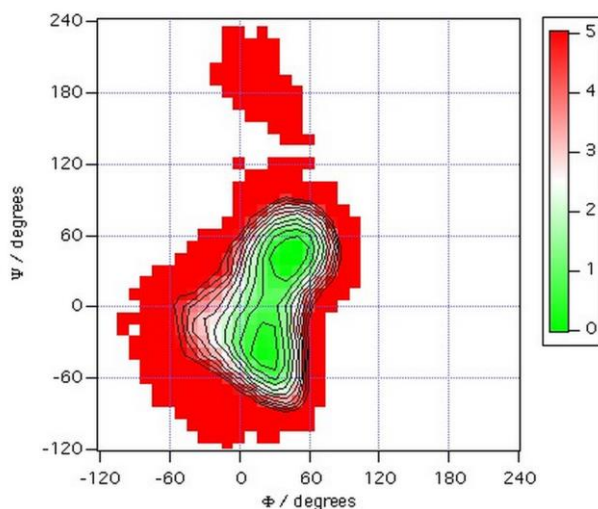
Les monosaccharides peuvent adopter une forme linéaire (milieu) ou cyclique. Il existe deux types de cycles : les cycles à 5 atomes ou furanose (en bas) et les cycles à 6 atomes ou pyranose (en haut). Selon l'orientation du groupement hydroxyle anomérique par rapport au groupement  $\text{CH}_2\text{OH}$  terminal, on définit l'anomérisation. S'ils sont de part et d'autre du cycle, on parle d'anomérisation alpha alors que s'ils sont du même côté on parle d'anomérisation bêta.

Les liaisons entre deux monosaccharides s'établissent au niveau du carbone anomérique qui peut avoir une configuration  $\alpha$  ou  $\beta$  (Figure 11). La conformation de la liaison de chaque disaccharide est décrite par les angles de torsion  $\varphi$  (PHI) et  $\psi$  (PSI). Les angles  $\varphi$  et  $\psi$  peuvent être définis suivant différentes conventions. (Figure 12).



**Figure 12 : Représentation des angles de torsion définissant la conformation du maltose**  
 Dans cet exemple, l'angle  $\varphi$  correspond à l'angle par  $H1'-C1'-O4-C4$  et l'angle  $\psi$  est décrit par :  $C1'-O4-C4-H4$ .

Les méthodes de mécanique moléculaire permettent de calculer l'énergie du disaccharide en fonction des angles de torsion et de tracer une carte d'énergie qui représente les conformations possibles des contacts inter-résidus (conflits stériques ou stabilisation) et des effets stéréo-électroniques. Ces cartes d'énergie sont différentes en fonction de la nature des monosaccharides et du type de liaison (Figure 13).

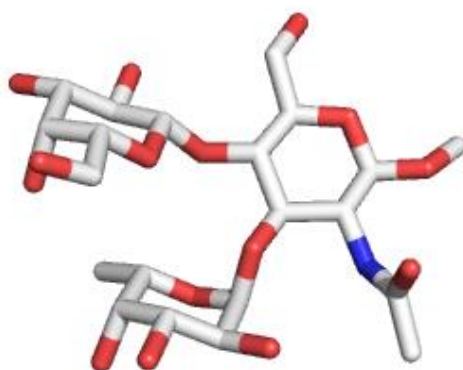


**Figure 13 : Carte d'énergie pour la liaison *Fuca1-3GlcNAc*.**

En vert sont représentées les régions de plus basse énergie c'est-à-dire les conformations les plus probables car énergétiquement favorables. Carte issue du site internet <http://www.glycosciences.de>.

### I.5.2. Exemple des Lewis X et Sialyl-Lewis X

Dès 1980, des études conjointes de résonance magnétique nucléaire et de modélisation ont conclu à la rigidité des oligosaccharides de type Lewis en raison de l'interaction stérique entre les cycles fucose et galactose qui sont portés par le résidu GlcNAc (Lemieux *et al.*, 1980). Le Lewis x a été cristallisé et sa structure démontre la présence de contacts hydrophobes entre les cycles qui stabilisent cette conformation particulière (Figure 14) (Perez *et al.*, 1996). La proximité des cycles génère de plus une liaison hydrogène CH-O non conventionnelle entre le C-H en position 5 du L-fucose et le O-5 du D-galactose (Zierke *et al.*, 2013).

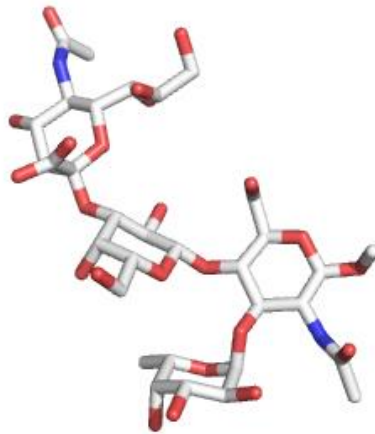


**Figure 14 : Structure du Lewis x à l'état cristallin**  
(Perez *et al.*, 1996)

Une exploration exhaustive de l'espace conformationnel des oligosaccharides de type Lewis a été réalisée en utilisant la méthode CICADA qui permet une exploration des chemins conformationnels avec relaxation des cycles à chaque pas (Imberty *et al.*, 1995). Il a été conclu que les oligosaccharides Le<sup>a</sup> et Le<sup>b</sup> sont un peu moins rigides que leurs homologues Le<sup>x</sup> et Le<sup>y</sup>. Cette étude a également prédit que la principale conformation de faible énergie peut afficher des variations de l'ordre de 20° pour chaque angle de torsion (Figure 13). Des conformations secondaires peuvent cependant exister, et l'abondance relative de chaque famille conformationnelle est fortement influencée par le solvant (Imberty and Perez, 2000). Le SLe<sup>x</sup> en solution peut adopter plusieurs conformations en raison de la flexibilité de la liaison NeuAc- $\alpha$ 2-3 Gal.

Les oligosaccharides Le<sup>x</sup> et SLe<sup>x</sup> interagissent avec un grand nombre de lectines. Par exemple la E-selectine a été co-cristallisée avec le SLe<sup>x</sup> et la conformation observée

correspond bien à une conformation de basse énergie en solution (Figure 15) (Somers et al., 2000). Le fragment trisaccharidique  $Le^x$  adopte toujours la conformation de basse énergie observée en solution avec un empilement des cycles fucose et galactose. Une exception notable est observée avec les lectines de type  $\beta$ -propeller de la famille d'*Aleuria aurantia* (AAL) : une distorsion de l'orientation du cycle fucose a été observée par RMN pour le  $SLe^x$  en interaction avec AAL (Haselhorst *et al.*, 2001) et par cristallographie pour le  $Le^x$  en interaction avec BamBL (Topin *et al.*, 2013) (voir chapitre résultats II.1. ).



**Figure 15 : Conformation du  $SLe^x$  complexé à la Sélectine E**  
Code PDB 1GIT (Somers *et al.*, 2000).

## II. Les lectines – structures et diversités

En 1888, Stillmark démontre que les extraits de graines de fèves de ricine contiennent une protéine qui peut agglutiner les globules rouges animaux (Stillmark, 1888). Peu de temps après, d'autres agglutinines ont été identifiées dans des graines différentes plantes (Ehrlich, 1891a; Ehrlich, 1891b). Pendant la Seconde Guerre mondiale, des lectines ont été utilisées pour différencier les groupes sanguins ABO à des fins de transfusion sanguine. Ces agglutinines ont ainsi été renommées «lectines», un terme dérivé du mot latin «légère», qui signifie «sélectionner».

Les lectines sont des protéines ubiquitaires non immunitaires que l'on retrouve des virus jusqu'aux mammifères. Elles ont la capacité de se lier de façon spécifique et réversible à des sucres. Elles ne possèdent pas d'activité enzymatique. Elles interviennent dans divers

processus biologiques, comme au niveau de la reconnaissance entre les cellules (par exemple lors de réponses immunitaires, d'infections).

## **II.1. Classification des lectines**

Les premières classifications sont venues des lectines de plantes et sont basées sur les glycanes reconnus avec la plus grande affinité. L'apparition du clonage moléculaire a permis une classification plus cohérente, basée sur l'homologie de séquence d'acides aminés et la phylogénie. Cependant, plusieurs classifications sont utilisées à l'heure actuelle. Ici nous allons classer les lectines en fonction de leur espèce d'origine : lectines de vertébrés, lectines d'invertébrés, lectines microbiennes.

### **II.1.1. Lectines de plantes**

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

Une autre grande famille de lectines d'origine végétale qui a pu être caractérisée est la famille des *Gramineae* (céréales) à laquelle appartient la lectine de germe de grain de blé (WGA). Les lectines de céréales présentent des domaines structuraux conservés, riches en cystéine qui sont aussi appelés domaine hévéine. La première structure cristallographique déterminée dans cette famille a été celle de la WGA (Wright, 1990).



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

Les lectines de plantes semblent être impliquées dans la défense contre les phytopathogènes et les prédateurs, certaines ayant des activités insecticides (Dahms and Hancock, 2002; Tikkanen et al., 1997).

### II.1.2. Lectines d'invertébrés

La présence de lectines a été démontrée chez les invertébrés terrestres ou marins. Ces protéines font souvent partie d'un système d'immunité innée et présentent donc des spécificités pour les glucides présents à la surface d'organismes pathogènes.

Chez les invertébrés, les lectines sont distribuées dans toutes les classes et sous-classes étudiées. Un nombre limité de lectines d'invertébrés a été cristallisé. La codakine purifiée à partir des branchies de la palourde blanche *Codakia orbicularis* est une lectine de type C qui a une forte spécificité pour des *N*-glycannes biantennés (Gourdine et al., 2008). D'autres exemples de structures de lectines de type C incluent celle de tunicier *Polyandrocarpa misakiensis* (Poget et al., 1999) et la lectine I (CEL-I) du concombre de mer *Cucumaria echinata* (Sugawara et al., 2004). Les structures 3D de lectines de type R ont aussi été obtenues telles que la lectine III (CEL-III) de la même holothurie (Uchida et al., 2004), du ver de terre *Lumbricus terrestris* (Yabe et al., 2007), ces deux lectines présentant une spécificité pour le galactose. La limule, *Tachypleus tridentatus* présente un grand nombre de lectines impliquées dans l'immunité innée. Les structures de la tachylectine 2, un  $\beta$ -propeller (Beisel et al., 1999) et de la tachylectine 5A une protéine de type fibrinogène (Liu et al., 2013), ont été résolues en complexe avec le GlcNAc. La structure de la lectine extraite de l'escargot

*Helix pomatia* (HPA) (Sanchez et al., 2006) a permis de définir une nouvelle famille de lectines, le type H (Tableau 3).

### II.1.3. Lectines de vertébrés

Les lectines de vertébrés sont réparties dans des familles très différentes. Les trois familles les plus étudiées sont les galectines, les lectines de type C et les siglecs.

Les structures de galectines sont relativement simples. Ces protéines sont spécifiques pour le  $\beta$ -galactose et plus précisément pour le lactose (Gal $\beta$ 1-4Glc) et le *N*-acetyllactosamine (Gal $\beta$ 1-4GlcNAc) (Leffler et al., 2004). La première structure tridimensionnelle d'une galectine humaine a été obtenue dans sa forme native et complexée avec le galactose, le GalNAc, le lactose, et le *N*-acétyllactosamine (Lobsanov *et al.*, 1993). Les lectines de type C présentent un CRD (Carbohydrate Recognition Domain) bien conservé qui implique un atome de calcium dans l'interaction protéine-sucre. Les lectines de type C sont soit en circulation dans le plasma, soit attachées aux surfaces cellulaires par la présence d'un segment transmembranaire. Un exemple de la structure 3D d'une lectine du type C est la E-selectine en complexe avec le SLe<sup>x</sup> (Somers *et al.*, 2000).

Les Siglecs, ou lectines de type I constituent une famille de lectines qui reconnaissent l'acide sialique (Zaccai et al., 2003). Elles sont membres de la superfamille des immunoglobulines (Ig) et adoptent un repliement de type immunoglobuline.

Les autres classes de lectines animales comprennent les lectines du type P qui sont formées par les lectines qui fixent le mannose-6-phosphate, comme la lectine bovine CD-MPR (Olson et al., 2002). Les pentraxines sont une famille de lectines capables de se lier à des ligands de manière Ca<sup>2+</sup> dépendante (Shrive et al., 1996). D'autres types de lectines de vertébrés telles que les calnexines-calreticuline comme la lectine calnexine de l'espèce *Canis familiaris* (Schrag et al., 2001) et les ficolines (Garlatti et al., 2007) ont été plus récemment mises en évidence.

La plupart des lectines de vertébrés ont une localisation extracellulaire et sont capables de détecter les modifications de glycosylation sur les cellules environnantes. Elles jouent donc un rôle dans la vie sociale des cellules et sont impliquées dans des processus tels que la fécondation, la migration et le développement cellulaire. Par exemple, dans le processus de fécondation, un glycoconjugué de la surface de l'ovule interagit avec une lectine (spermadhésine) du spermatozoïde (Topfer-Petersen et al., 1998).

Famille de lectines	Organisme	Ligands type saccharidique	Localisation subcellulaire	Exemple de fonctions
Calnexine	Vertébrés	Glc1Man9	Réticulum endoplasmique (RE)	Contrôle de qualité lors du repliement protéique
Lectines de type P	Vertébrés	Man 6-phosphate, autres	Voies de sécrétion	Contrôle de qualité lors du repliement protéique
Lectines de type C	Vertébrés Invertébrés	Variable	Membrane cellulaire, extracellulaire	Reconnaissance du non-soi, inflammation et immunologie
Galectines	Vertébrés	B-Galactosides	Cytoplasme, extracellulaire	Adhésion cellulaire, métastase, infection
Lectines de I (siglecs)	Vertébrés	Acide sialique	Membrane cellulaire	Reconnaissance inter cellulaires , guidance neuronale
Lectines de type R	Invertébrés Plantes	Variable	Golgi, membrane cellulaire	Adhésion cellulaire Défense
Ficolines	Vertébrés	GlcNAc, GalNac	Membrane cellulaire, extracellulaire	Immunité innée
Lectines de types chitinase	Invertébrés	Chito-oligosaccharides	Extracellulaire	Immunité innée
Béta-propeller	Invertébrés Bactéries	Fuc, GlcNAc, GalNac	Extracellulaire	Immunité innée
Lectines de type H	Invertébrés	GalNac	Extracellulaire, membrane cellulaire	Immunité innée
Gramineae	Plantes	Variable		Défense
Legumineuses	Plantes	Variable		Défense et symbiose
Monocotylédones	Plantes	Mannose		Défense

**Tableau 3 : Exemples de familles de lectines**

Il a été aussi montré que l'embryon exprime, à sa surface cellulaire, la L-sélectine six jours après la fécondation. En se liant aux sucres spécifiques de l'endothélium utérin, la L-sélectine détermine ainsi la première étape indispensable à la fixation du blastocyste à la surface de l'endomètre et permet le début de la grossesse (Genbacev et al., 2003).

#### **II.1.4. Lectines microbiennes**

Les microorganismes pathogènes, virus, bactéries, champignons ou parasites eucaryotes, utilisent fréquemment des lectines pour reconnaître les sucres présents sur la surface des cellules hôtes. Ces interactions lectines/sucres jouent également un rôle dans

l'adhésion sur les tissus richement glycosylés présents dans les voies respiratoires, digestives ou dans l'appareil urogénital (Imberty and Varrot, 2008; Sharon, 1996).

Divers virus possèdent à leur surface des lectines qui leur permettent d'interagir avec les cellules hôtes. Au début des années 1940, Hirst montre que le virus de la grippe est capable d'hémagglutiner des érythrocytes (Hirst, 1942). Alfred Gottschalk a mis en évidence une protéine capable de se lier aux érythrocytes et à d'autres cellules qui présentent un résidu acide sialique de surface cellulaire (Gottschalk and Perry, 1951). Don Wiley et ses associés ont cristallisé cette hémagglutinine virale et déterminé sa structure en 1991 (Weis *et al.*, 1990), plus tard ils ont résolu la structure de co-cristaux avec le sialyllactose (Sauter *et al.*, 1992). L'hémagglutinine joue deux rôles : elle permet à la fois la reconnaissance des cellules cibles mais aussi la fusion des membranes hôtes et virales. Plusieurs autres familles de virus (norovirus, rotavirus....) utilisent des domaines lectines de leur capsid pour se fixer sur les glycoconjugués des tissus de leur hôte (Air, 2014).

Les lectines bactériennes peuvent être classées en quatre familles principales : les lectines de fimbriae, les lectines flagellaires, les toxines et les lectines solubles.

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

Certaines toxines bactériennes, telles que celles secrétées par *Shigella dysenteriae*, *Borderella pertussis* et *Vibrio cholerae* présentent un ou plusieurs domaines lectines. Le domaine lectine de la toxine de cholera est un pentamère et sa structure a été résolue complexée avec le pentasaccharide GM1 qui est un ganglioside présent sur les epithelia du tractus digestif (Merritt *et al.*, 1994). Les bactéries neurotoxiques responsables du tétanos et du botulisme utilisent également des toxines à domaine lectine pour se fixer sur les cellules du système nerveux.

Certaines bactéries produisent des lectines solubles qui sont impliquées dans la reconnaissance de l'hôte et la formation du biofilm. La famille de lectines solubles comprend des protéines cytoplasmiques telles que LecA et LecB de *Pseudomonas aeruginosa* respectivement spécifiques pour le  $\alpha$ -D-galactose et le  $\alpha$ -L-fucose (Imberty *et al.*, 2004). Elles présentent des structures tétramériques où un ou deux ions calciums respectivement sont impliqués dans le site de reconnaissance du sucre.

## II.2. Rôle des lectines

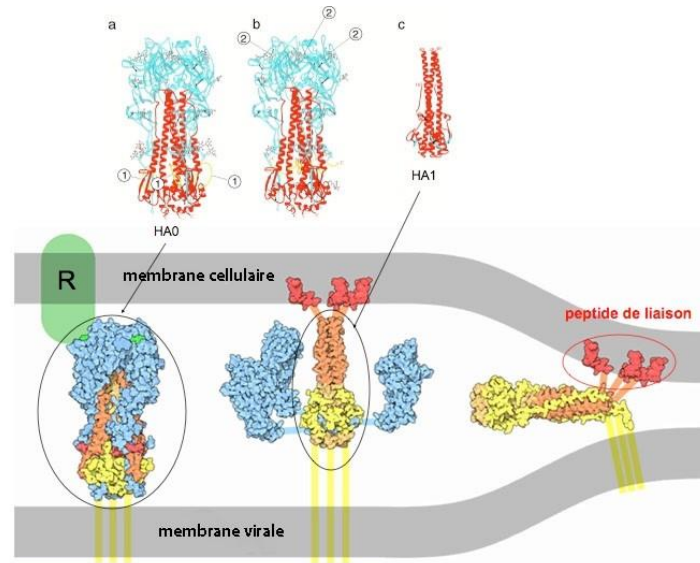
### II.2.1. Reconnaissance et adhésion

Plusieurs pathogènes utilisent des lectines afin de reconnaître, d'adhérer et d'infecter leurs hôtes. Parmi ces pathogènes on peut citer le virus de la grippe *Influenza* qui possède une lectine particulière : l'hémagglutinine. Il s'agit d'une protéine antigénique retrouvée à la surface des virus. Elle correspond au « *H* » que l'on retrouve dans les différentes formes de grippe (H5N1, H1N1, H2N2...). Il existe en réalité 19 formes différentes d'hémagglutinine, numérotées de 1 à 19. Son nom provient de sa capacité à agglomérer les hématies.

Elle est synthétisée sous forme d'un précurseur inactif HA0 par le virus. La structure tridimensionnelle de l'hémagglutinine révèle qu'il s'agit d'un homotrimère (molécule formée de trois monomères identiques). Chaque monomère de la glycoprotéine HA0 est constitué de deux sous-unités. La partie extracellulaire globulaire ou HA1 présente le site de liaison (RBS) au récepteur des cellules cibles comportant une molécule d'acide sialique. La seconde partie forme une tige ou HA2 permet la fusion de la membrane phospholipidique qui recouvre la particule virale avec la membrane de l'endosome de la cellule hôte lors du processus d'assemblage et de libération des virions. Cette partie HA2 présente un domaine transmembranaire hydrophobe et un domaine ou "une queue" intracellulaire.

Une fois attachées à la surface de la cellule hôte, pour être fonctionnelles et permettre l'infection, les parties de l'hémagglutinine HA1 et HA2 doivent être séparées l'une de l'autre grâce à un clivage au niveau d'un peptide reliant les deux sous-unités. Le clivage des hémagglutinines démasque alors les peptides de liaison qui s'ancrent dans la membrane endosomale et sous l'effet du pH acide de l'endosome, l'hémagglutinine subit un changement de conformation qui permet le rapprochement puis la fusion de l'enveloppe virale et de la membrane endosomale (Figure 16). L'intérieur du virus est ainsi libéré dans le cytoplasme

tant que la coupure n'a pas eu lieu les virus ne sont pas capables d'infecter une cellule (Air, 2014).



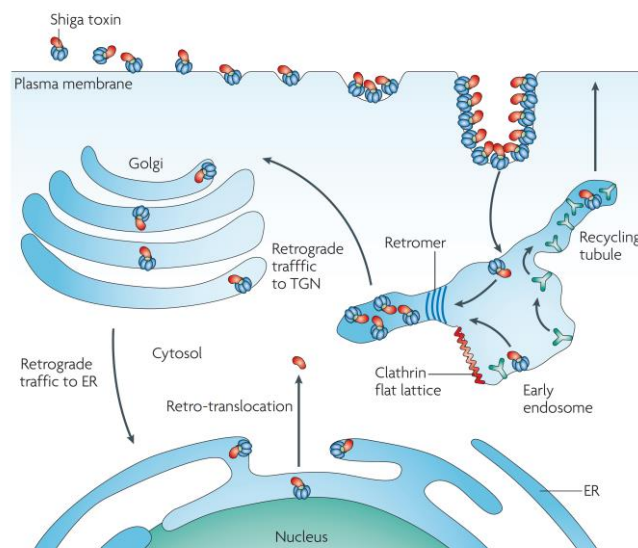
**Figure 16 : Représentation schématique de la fixation et du changement de conformation de l'hémagglutinine à la surface de la membrane plasmique de la cellule hôte.**

L'hémagglutinine sous forme HA0 vient se fixer à l'acide sialique, son récepteur (R). Ensuite un changement de conformation s'opère qui va avoir pour conséquence l'ancrage de l'hémagglutinine dans la membrane de la cellule hôte et finalement le rapprochement des membranes virales et cellulaires. (D'après [http://accs.ens-lyon.fr/accs/ressources/immunitetvaccination/virusetimmunit/dossiergrippe/. Virus%20grippe%20images%20jpg.027bis.jpg/view](http://accs.ens-lyon.fr/accs/ressources/immunitetvaccination/virusetimmunit/dossiergrippe/.%20Virus%20grippe%20images%20jpg.027bis.jpg/view))

Les lectines ne permettent pas uniquement l'adhésion d'un pathogène à une cellule hôte, elles permettent aussi l'adhésion des cellules entre elles. On peut par exemple citer les protéines de *P. aeruginosa*, LecB et LecA qui permettent respectivement la formation et le maintien de l'intégrité du biofilm bactérien. Le biofilm est un mode de vie où le pathogène va pouvoir s'établir et persister. Il permet une résistance au système immunitaire, aux antibiotiques mais aussi une résistance aux attaques mécaniques. LecB qui est localisée sur la membrane externe de la bactérie joue un rôle dans la formation du biofilm en se fixant sur les glycoconjugués localisés à la surface des bactéries. Une souche déficiente de LecB n'est plus capable de former un biofilm (Tielker *et al.*, 2005). La protéine LecA joue un rôle dans la stabilisation du biofilm mais elle n'est pas essentielle à sa formation (Diggle *et al.*, 2006). Ces lectines ont également des propriétés cytotoxiques, ainsi la lectine LecB, comme LecA, est capable d'inhiber le battement ciliaire des cellules pulmonaires (Mewe *et al.*, 2005).

## II.2.2. Trafic cellulaire

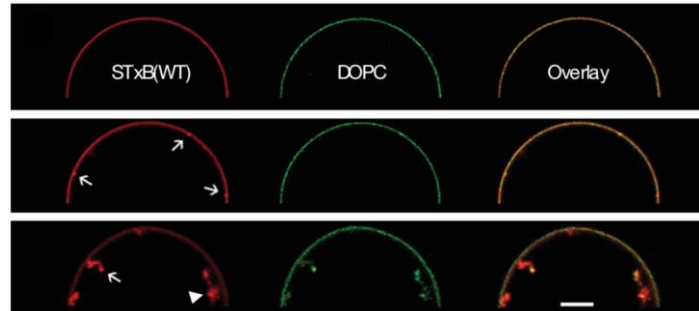
Différents agents pathogènes ou des facteurs pathogènes, comme le virus simien 40 et le virus de polyome, les toxines du choléra et de Shiga, utilisent les glycosphingolipides de la membrane plasmique des cellules hôtes pour entrer et infecter les cellules. Après la liaison à différents glycosphingolipides (GM1, GD1a ou au Gb3), les lectines pénètrent dans les cellules par des voies d'endocytose qui n'impliquent pas le réseau complexe de clathrine (Ewers *et al.*, 2010; Romer *et al.*, 2007). Ces toxines bactériennes et les virus animaux peuvent induire une invagination de la membrane plasmique sans l'aide de la machinerie protéique cytosolique. La formation de tubule serait la conséquence de la construction dynamique de nanodomaines protéines-lipides dont les propriétés intrinsèques conduisent à la flexion des membranes (Romer *et al.*, 2007). La liaison à la membrane plasmique des toxines induit la courbure locale spontanée. La toxine est ensuite soumise à un tri dans les endosomes précoces rétrogrades. Les tubules rétrogrades sont formés de manière dépendante de la clathrine (Romer *et al.*, 2010). Les toxines Shiga sont localisées préférentiellement dans cet environnement tubulaire. La Shiga-toxine contourne la voie endocytose tardive pour être transférée directement à partir de l'endosome précoce au réseau de Golgi (TGN) et, de là, sur le réticulum endoplasmique (RE). Enfin, les toxines Shiga utilisent la machinerie de dégradation associée au RE (DGAER) pour faciliter la rétrotranslocation dans le cytosol de la cellule hôte (Figure 17) (Johannes and Romer, 2010).



**Figure 17 : Trafic cellulaire de la Shiga Toxine**

Une vue d'ensemble du trafic intracellulaire de Shiga-toxines (Johannes and Romer, 2010).

La formation de tubules et des invaginations profondes de la membrane ont été observées non seulement dans la membrane plasmique des cellules, mais ont été également reconstituées *in vitro* sur des vésicules unilamellaires géantes (Figure 18).



**Figure 18 : Reconstitution de la formation des tubules dans des GUVs.**

*Shiga-toxine* : STxB (rouge) se lie aux GUVs composées de DOPC (vert), de cholestérol et de Gb3. Les STxB se regroupent spontanément (panneau central, flèches); après une diminution de la tension de la membrane, il y a induction de l'invagination de la membrane (panneau inférieur), qui apparaissent comme des bobines (pointe de flèche) ou des structures en spirale (flèche) (Romer *et al.*, 2007).

Ces résultats sont d'une grande pertinence clinique. Dans le cas de SV40, la structure de la chaîne hydrocarbonée de la glycosphingolipide GM1 influe de manière significative l'efficacité de l'infection de SV40 (Ewers 2010). Plus récemment, Römer et son équipe ont montré que ce mécanisme était à l'origine d'invasion bactérienne. En effet, la bactérie *Pseudomonas aeruginosa* utilise la capacité de sa lectine LecB à former des invaginations de la membrane pour se faire elle-même internaliser dans la cellule hôte (Eierhoff *et al.*, 2014).

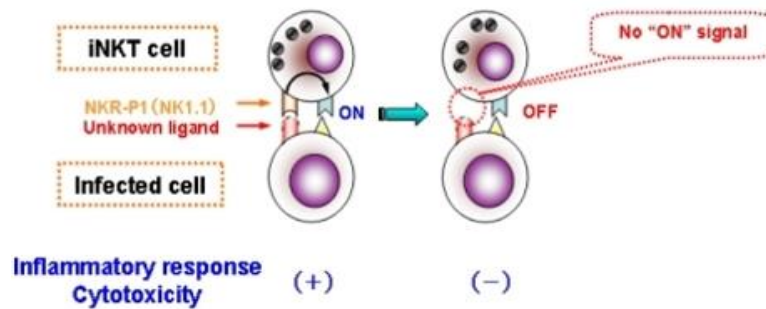
### II.2.3. Transduction du signal

De nombreuses lectines membranaires, activées par la liaison de leurs ligands, ont un comportement proche de celui de certains récepteurs des facteurs de croissance, et activent des tyrosine-kinases cellulaires ou des protéines G hétérotrimériques. Les lectines forment ainsi une nouvelle classe de protéines transductrices du signal.

Une des caractéristiques des lectines membranaires est la diversité des réponses obtenues selon le type cellulaire, le tissu où elles sont exprimées, mais également le ligand. Plusieurs exemples peuvent être cités. Cette diversité peut être illustrée par le lectine NKR-P1 qui est



une lectine de type C exprimée entre autre sur les NK (natural killer) et les monocytes du sang périphérique humain. Dans les cellules NK humaines, l'activation du récepteur NKR-P1, provoque l'activation de protéines-tyrosines kinases et la phosphorylation transitoire de certaines tyrosines (Figure 19) (Emoto and Emoto, 2009).



**Figure 19 : Représentation schématique de l'activation d'un NK via la lectine NKR-P1**

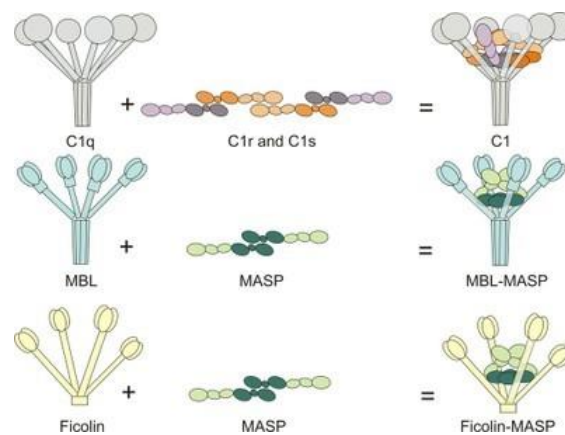
Lorsque le récepteur/lectine NKR-P1 se fixe sur la cellule infectée il y a activation du NK, suivi d'une réponse inflammatoire et une cytotoxicité envers la cellule cible. Alors que sans activation NKR-P1 il n'y a pas activation du NK (© Copyright PolicyLicense) (Emoto and Emoto, 2009).

#### II.2.4. Immunité

Parmi la diversité des lectines, certaines ont un rôle central dans le système immunitaire. En effet, les lectines du système immunitaire sont des molécules de reconnaissance qui se lient aux micro-organismes pour promouvoir la phagocytose ou l'activation du système du complément. On peut classer ces lectines en plusieurs familles : les sélectines, les collectines solubles.

Les sélectines sont un petit groupe de lectines de type C membranaires comprenant trois membres : les sélectines L, P et E. La L-sélectine est exprimée par les lymphocytes, et plus particulièrement les lymphocytes recirculants avant qu'ils ne soient activés par un antigène. Elle joue un rôle dans la migration de ces lymphocytes vers les organes lymphoïdes secondaires (Newton and Dixit, 2012). L'expression de la P-sélectine sur les cellules endothéliales et sur les plaquettes activées est induite par l'inflammation. La production de sélectine E est aussi induite par l'inflammation mais seulement sur l'endothélium. Les sélectines P et E sont essentielles pour le recrutement des leucocytes dans les tissus infectés (Chase et al., 2012). Les ligands sont des glycanes à SLe<sup>x</sup>.

Les collectines et les ficolines sont deux familles de protéines présentes à la surface des muqueuses et dans le sang et qui reconnaissent des structures glycosidiques caractéristiques de la surface des cellules du « non-soi ». Les collectines et les ficolines sont des composants majeurs du système du complément et font partie des opsonines, c'est-à-dire qu'elles vont se fixer à la surface d'un micro-organisme et stimuler la phagocytose par les macrophages (Fujita *et al.*, 2004). Trois collectines et deux ficolines ont été caractérisées chez l'Homme. La plus étudiée est sans doute la collectine humaine liant le mannose (MBL, pour mannose-binding lectin). Il s'agit d'une protéine du sérum. La MBL est synthétisée en quantité importante en réponse à une infection. Elle peut se lier à un sous-groupe de bactéries Gram positives ou négatives, à des champignons et même à certains virus, comme le virus de la grippe. La MBL va permettre l'activation du système du complément et faciliter la phagocytose (Kawasaki, 1999) (Figure 20).



**Figure 20 : Principales lectines capables d'activer le système du complément**

D'après <http://www2.le.ac.uk/>

## II.2.5. Chaperonne

Le réticulum endoplasmique contient de nombreuses protéines chaperonnes qui sont capables de reconnaître des domaines de protéines en cours de traduction dans leur état provisoire et d'assurer leur repliement. Parmi elles, la plus abondante protéine du RER, est la BiP (binding protein). Parmi les protéines liées à la BiP, un groupe abondant est représenté par des chaperonnes qui possèdent un domaine lectine se liant donc spécifiquement à certains résidus glucidiques. La calréticuline est une lectine chaperonne soluble alors que la calnexine est une protéine transmembranaire. Ces deux lectines chaperonnes n'ont pas d'équivalent dans d'autres compartiments cellulaires.

Lors de leur maturation dans le RER, les *N*-glycoprotéines passent par un état monoglucosylé qui est reconnu par une lectine chaperonne, la calnexine (ou la calréticuline) qui favorise son repliement. Après un laps de temps, durant lequel la protéine a atteint un certain degré de repliement, le glucose restant est éliminé. La protéine en cours de repliement se sépare de la lectine chaperonne. Si elle n'est pas dans son état de repliement définitif elle est de nouveau reconnue puis reglucosylée en position 1 et de nouveau fixée par la lectine–chaperonne calnexine (ou calréticuline). Ce cycle peut se répéter plusieurs fois et se termine par le passage de la protéine vers le *Golgi* (Trombetta and Helenius, 1998).

Le système ERAD (ER-associated degradation) qui vérifie le repliement des protéines inclut également une lectine.

Les protéines mal repliées peuvent être reconnues par une lectine liée aux calnexines par son domaine transmembranaire. Les protéines mal repliées sont alors ubiquitinylées et dégradées par le protéasome (Hoseki et al., 2010).

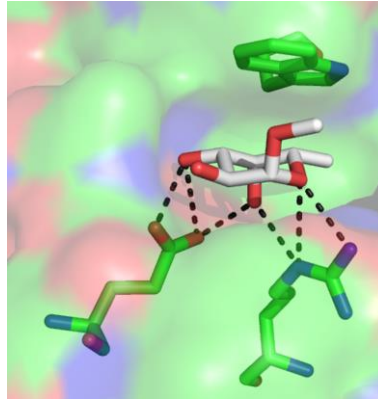
## **II.3. Site de reconnaissance**

Les lectines possèdent toutes des « domaines de reconnaissance des glucides » (CRD) qui ont évolués à partir de gènes ancestraux, souvent en conservant les caractéristiques spécifiques de la séquence primaire d'acides aminés et de la structure tridimensionnelle. Malgré ces ressemblances au niveau des séquences ou des structures à l'intérieur d'une famille de lectine, les structures de glycanes reconnus peuvent être très diverses. L'affinité d'un seul site de liaison est faible (avec des valeurs de  $K_d$  de l'ordre du millimolaire), bien que certaines lectines reconnaissent certains glycanes avec une affinité beaucoup plus élevée (avec des valeurs de  $K_d$  dans la gamme du micromolaire). Pour les lectines avec une faible affinité, les interactions multivalentes entre une protéine et plusieurs glycanes sont souvent nécessaires pour produire des interactions de haute avidité (Imberty and Varrot, 2008).

### **II.3.1. Interactions lectine-sucre**

L'analyse des structures cristallographiques de complexes protéines/glucides, indique que les lectines ont des propriétés et modes de reconnaissance des glucides variés.

Les sites de fixations des lectines au niveau structural sont souvent des poches peu profondes. La grande majorité des forces impliquées dans les reconnaissances des glucides est représentée par des liaisons hydrogène et des forces de Van der Waals (Figure 21) (Ashton et al., 1997).



**Figure 21 : Site de reconnaissance de la protéine RSL.**

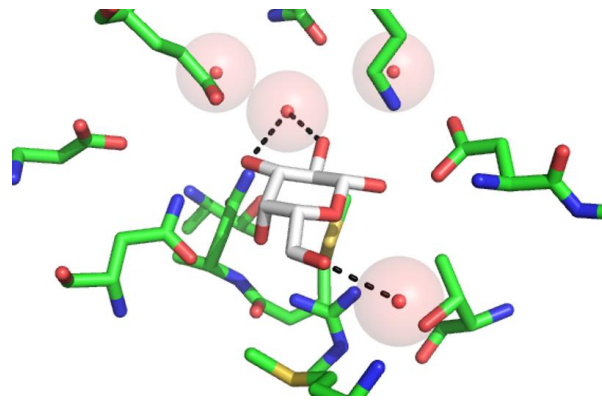
Les liaisons hydrogène sont représentées en pointillés noirs (code PDB 2BT9) (Kostlanova et al., 2005).

D'après une étude des sites de reconnaissance des lectines de légumineuses, il a été établi que les liaisons hydrogène et les forces de Van der Waals participent de façon égale à la stabilisation des glucides dans les sites de fixations (Imberty *et al.*, 1991). Dans certains sites de reconnaissance les forces de Van der Waals auront tendance à prédominer alors que dans les autres les liaisons hydrogène seront plus importantes.

Les structures cristallographiques des différents complexes indiquent que les acides aminés les plus impliqués dans les liaisons hydrogène avec le sucre sont respectivement : aspartate, arginine > glutamate > arginine, histidine, tryptophane, lysine > tyrosine, glutamine > sérine et thréonine (Vyas *et al.*, 1991). Certaines molécules comme l'eau et les ions sont le plus souvent impliquées dans la formation et la stabilisation du complexe protéine-sucre.

### **II.3.2. Stabilisation des interactions par des molécules d'eau**

La détermination de structures cristallographiques à haute résolution permet de localiser les molécules d'eau à l'intérieur ou à la surface des protéines. Lorsque la protéine n'est pas complexée à son substrat, son site de reconnaissance est accessible, et est occupé par des molécules d'eau qui sont expulsées ou non lors de la fixation du ligand.



**Figure 22 : Importance des molécules d'eau dans la stabilisation du sucre dans le site de fixation**

Les molécules d'eau sont représentées sous forme de sphères rouges et les liaisons hydrogène par des pointillés noirs). Image réalisée à partir de la structure dont le code PDB est 5ABP (Quioco *et al.*, 1989)

Deux types de molécules d'eau sont rencontrés, les molécules d'eau de surface et les molécules d'eau enfouies dans le site de fixation. Quioco et son équipe ont démontré que des molécules d'eau peuvent influencer la spécificité et l'affinité de la protéine de fixation à l'arabinose (ABP) pour le galactose (Figure 22) (Quioco *et al.*, 1989).

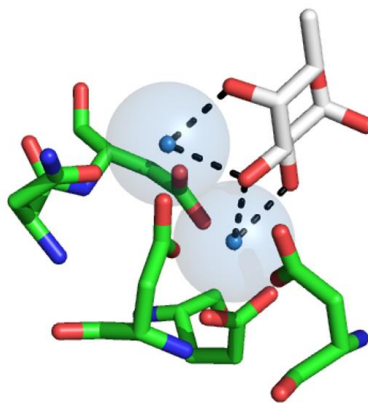
Dans les complexes de lectines de légumineuses en présence d'oligosaccharide, un seul résidu glycosidique est souvent profondément ancré dans le site de reconnaissance pour former des liaisons hydrogène avec les acides aminés de la protéine. La partie la plus exposée de l'oligosaccharide a tendance à construire des liaisons avec les molécules d'eau environnantes. Certaines de ces molécules d'eau créent un réseau de liaisons hydrogène entre le glucide et les acides aminés de la protéine, stabilisant ainsi la chaîne glycosidique à la surface de la protéine (Bourne and Cambillau, 1993).

### II.3.3. Stabilisation des interactions par des ions

Les ions jouent fréquemment un rôle essentiel dans les complexes protéine-sucre. Ils stabilisent la structure tertiaire de nombreuses protéines, participent à l'architecture de certains sites de reconnaissance et peuvent également interagir directement avec le ligand. Les lectines animales de type C par exemple sont dépendantes d'un ion calcium qui est essentiel à la reconnaissance du glucide. De la même façon, certaines lectines bactériennes impliquent un

ou deux ions calcium pour stabiliser les interactions avec le sucre. C'est par exemple le cas de la protéine LecB de *Pseudomonas aeruginosa*. En effet, la structure cristallographique de LecB complexée au fucose met en évidence une poche acide comportant deux ions calcium. Trois des hydroxyles du fucose ainsi que les chaînes latérales des acides aminés Glu95, Asp99, Asp101, Asn103 et Asp104 participent à la coordination des atomes de calcium (Figure 23) (Mitchell *et al.*, 2002).

Les lectines de légumineuses contiennent un ion calcium et un ion manganèse proches de leur site de fixation. Un réseau de coordination se forme entre les ions, les acides aminés proches et les molécules d'eau. Ce réseau permet de stabiliser une boucle de la protéine, afin de laisser le site de reconnaissance accessible. L'ion calcium influence la position de certains acides aminés du site de reconnaissance pour qu'ils puissent former des liaisons hydrogène avec le substrat (Loris *et al.*, 1994).



**Figure 23 : Importance des ions dans la stabilisation du sucre dans le site de fixation de LecB**

*Certains ions (sphères bleues) interagissent directement avec le sucre pour le stabiliser par des liaisons hydrogène avec les principaux acides aminés du site de fixation. Ils stabilisent également l'interaction lectine/sucre en formant des liaisons (pointillés noirs). Image réalisée à partir de la structure dont le code PDB est 1GZT (Mitchell *et al.*, 2002).*

## **II.4. Lectines et les autres molécules reconnaissant les sucres**

### **II.4.1. Introduction**

La grande diversité et la complexité de la structure des glycanes ainsi que leur rôle crucial dans de nombreux processus biologiques ou pathologiques nécessitent le développement de nouvelles techniques à haut débit pour leurs analyses.

Les lectines sont classiquement utilisées pour cibler des glycoconjugués. Elles sont, par exemple, utilisées sur des puces qui permettent d'étudier le glycome de cellules. Le développement des lectines recombinantes donne accès à du matériel fiable, reproductible et produit en grande quantité. L'ingénierie de nouveaux sites, en se basant sur des architectures de protéines déjà existantes, permet un ajustement de la spécificité.

A partir des connaissances des interactions protéine-sucre, il est maintenant possible d'utiliser des oligonucléotides ou des peptides de synthèse pour produire des nouvelles molécules capables de se lier aux sucres. Une telle approche biomimétique peut également être envisagée par la chimie des boronates. De nouveaux développements en chimie supramoléculaire permettent également la conception de glycosensors entièrement artificiels.

### **II.4.2. Les molécules reconnaissant les sucres**



## Binding sugars: from natural lectins to synthetic receptors and engineered neolectins†

Julie Arnaud,‡ Aymeric Audfray‡ and Anne Imberty\*

Cite this: *Chem. Soc. Rev.*, 2013, **42**, 4798

Received 24th October 2012

DOI: 10.1039/c2cs35435g

[www.rsc.org/csr](http://www.rsc.org/csr)

The large diversity and complexity of glycan structures together with their crucial role in many biological or pathological processes require the development of new high-throughput techniques for analyses. Lectins are classically used for characterising, imaging or targeting glycoconjugates and, when printed on microarrays, they are very useful tools for profiling glycomes. Development of recombinant lectins gives access to reliable and reproducible material, while engineering of new binding sites on existing scaffolds allows tuning of specificity. From the accumulated knowledge on protein–carbohydrate interactions, it is now possible to use nucleotide and peptide (bio)synthesis for producing new carbohydrate-binding molecules. Such a biomimetic approach can also be addressed by boron chemistry and supra-molecular chemistry for the design of fully artificial glycosensors.

### 1. Introduction

Carbohydrates constitute the most abundant class of biomolecules on Earth. Glycoconjugates are present as glycoproteins

and glycolipids at the surface of cells and participate in many interactions with diverse receptors. Glycosylation is the most common post-translational modification of proteins<sup>1</sup> and it has been estimated that 70% of the human proteome, including 80% of membrane proteins, is glycosylated.<sup>2</sup> Carbohydrate-mediated interactions play a vital role in many physiological processes including fertilization, immune defence, cellular differentiation, cell–matrix interaction, and cell–cell adhesion.<sup>3</sup> These interactions are also involved in many pathologies as they play a key role in adhesion of viruses, bacteria and

*Centre de Recherche sur les Macromolécules Végétales (CERMAV – CNRS), affiliated to Grenoble-Université and ICMG, Grenoble, France.*

*E-mail: [imberty@cermav.cnrs.fr](mailto:imberty@cermav.cnrs.fr)*

† Part of the Multivalent Scaffolds in Glycosciences themed issue.

‡ These authors have contributed equally.



Julie Arnaud

*Julie Arnaud received her BS degree in biology in 2009 and her MS degree in Integrative Cellular Biology in 2011 from the University Joseph Fourier at Grenoble. She is now pursuing a PhD degree in the Anne Imberty group at CERMAV-Grenoble. Her doctoral research focuses on synthetic lectins with controlled valency and specificity for application in cellular biology and biotechnology.*



Aymeric Audfray

*Dr Aymeric Audfray received his PhD in cellular and molecular biology in 2008 from Université de Limoges. His work was focused on the importance of O-fucosylglycans in Notch signaling in the context of myogenesis. He then came to Grenoble, for post-doctoral research at the CERMAV, in the group of Anne Imberty, to study lectin–sugar interactions. His first projects focused on the development of inhibitors of sugar–pathogenic lectin interactions to design anti-adhesive compounds against infection from *Pseudomonas aeruginosa* or *Burkholderia cepacia*. He is now involved in engineering of  $\beta$ -propeller architecture lectins to create new tools for medical applications.*



parasites to host cells, in neurological disorder and in tumour growth and metastasis. Analysis of glycome modification occurring during diseases is therefore of high interest for diagnostic purposes. Appropriate glycosylation is also critical in the production of recombinant glycoproteins of therapeutic interest, since their glycosylation profile controls their immunogenicity, stability in serum and pharmacodynamic properties.

Glycomics, the comprehensive study of glycomes, is therefore a complex field and requires complementary strategies to decipher the glycan dictionary. Classically the glycan structures are hydrolysed from the glycoconjugates/cell surfaces and then analysed using a variety of techniques such as liquid chromatography, nuclear magnetic resonance and different modes of mass spectrometry.

As an alternative method, antibodies or lectins can be used as biomarkers. Lectins, despite their somehow lower specificity, appear to be more suitable than antibodies for glycan profiling and differentiation purposes in glycomics. Analytic tools available for high-throughput glycan profiling or for detection of a given glycan in a clinical environment have been developed in the form of lectin arrays or glyco-biosensors.<sup>4–7</sup> As the source and quality of lectins could be an issue for reliable and reproducible data,<sup>8</sup> production of recombinant lectins is being developed,<sup>9,10</sup> although only a limited fraction of the diversity in glycan specificity observed in natural lectins could be reproduced. The production of “non-classical lectins”, based on protein-engineering, aptamer technology, peptide synthesis, supramolecular chemistry or boron chemistry, is therefore explored with the aim of creating biomimetic molecules that can be used in carbohydrate-recognition processes.



**Anne Imberty**

*Dr Anne Imberty is Research Director at the Centre de Recherches sur les Macromolécules Végétales (CERMAV), affiliated to the Centre National de la Recherche Scientifique (CNRS), Grenoble, France. She graduated in biology from Ecole Normale Supérieure in Paris. In 1984, she joined the CNRS in Grenoble and did her PhD on starch structure. She started studying protein-carbohydrate interaction during her post-doc in Toronto. Since*

*1999 she has held a senior research position in CNRS-Grenoble and has been the head of the Molecular Glycobiology group. Her research interests are in the field of structural glycobiology, with main interest in biologically active oligosaccharides and their interaction with lectins and glycosyltransferases. She received the Whistler award of the International Carbohydrate Organization in 2004, and the Charles Dhéré award in Chemistry from the French Academy of Sciences in 2011.*

## 2. Lectins – proteins with a sweet tooth

Lectins are proteins of non-immune origin that bind to specific carbohydrates without enzymatically altering them.<sup>11</sup> Lectins are present in all branches of the evolution tree since they have been identified in all organisms, from microorganisms to fungi, plants and animals.<sup>12</sup> Due to their interactions with glycoproteins, glycolipids and oligosaccharides, they play the crucial role of deciphering the glycode.<sup>13–15</sup> Their biological functions are very diverse. Animal lectins are involved in glycoprotein trafficking and clearance, development, immune defence, fertility and other biological processes.<sup>16–18</sup> In many invertebrates, the ability of lectins to differentiate between self and non-self makes them part of the innate immunity system.<sup>19,20</sup> In plants and fungi, they have been suggested to act as defence proteins in protection against pathogens or feeders.<sup>21,22</sup> In pathogenic microorganisms, lectins are often involved in host recognition and tissue adhesion.<sup>23,24</sup>

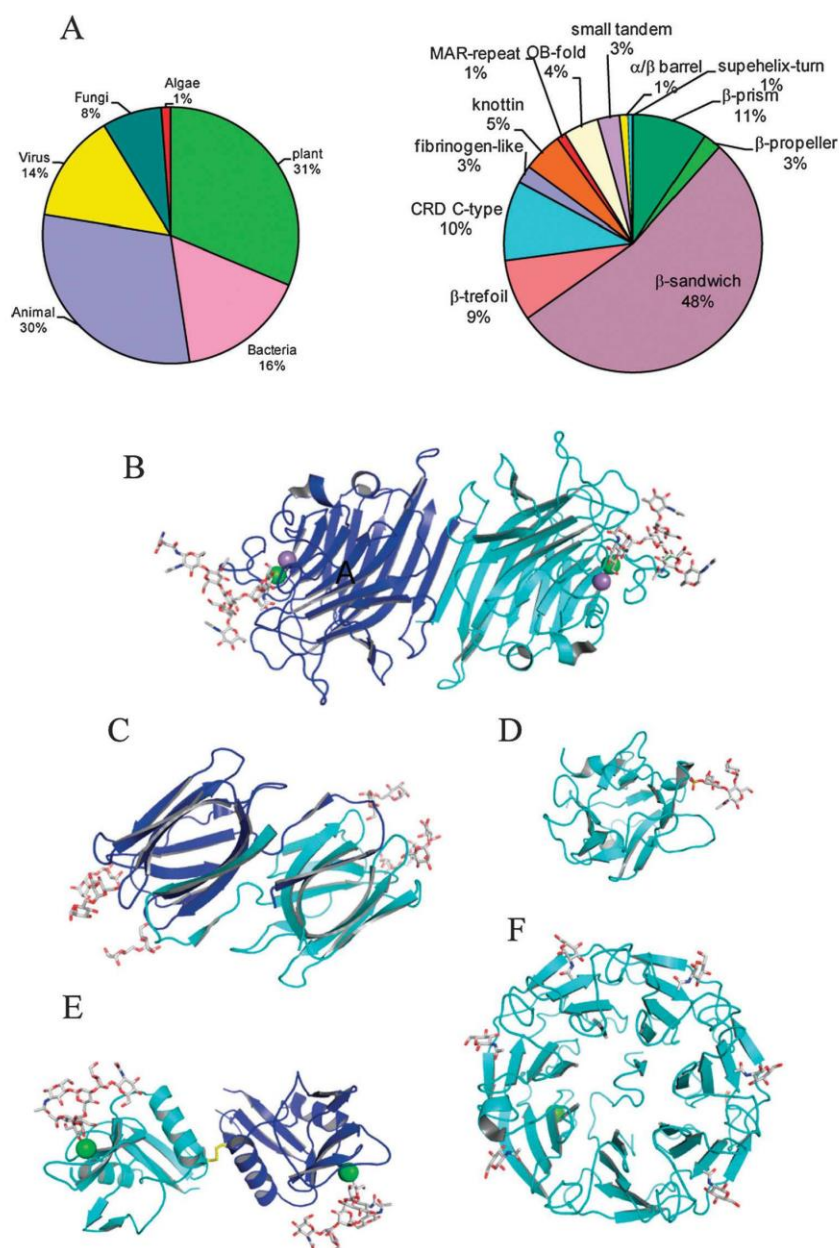
### 2.1. Structural basis for carbohydrate binding

Detailed knowledge of the protein structure and the interaction established with the carbohydrate ligand is a prerequisite for a better understanding of the biological and physical processes of recognition, and for the development of biotechnological and therapeutical applications. The resolution of the crystal structure of concanavalin A, a well described legume lectin specific for mannose and glucose, marked the beginning of structural glycobiology.<sup>25</sup>

At the present time, more than 1000 three-dimensional structures of lectins are available. Most of the structures of lectin and lectin-carbohydrate complexes have been obtained from crystallographic work, as established in the 3D-lectin database (<http://lectin3d.cermav.cnrs.fr/search.php>).

For a very long time, plant lectins were the only ones structurally characterized, and legume lectins have been considered as models for studying protein-carbohydrate interactions.<sup>26,27</sup> Plant lectins still represent 31% of available structures (Fig. 1A), now closely followed by animal lectins (30%). Galectins, which are soluble galactose-binding lectins, and C-lectins, which depend on the presence of one calcium ion in the binding sites, have been extensively characterized. Lectins from bacteria and viruses correspond currently to 16% and 14% of the database entries, respectively, but their number is growing rapidly. Growing interest is now devoted to fungal lectins (8%),<sup>28</sup> while only one algal lectin, griffithsin, has been structurally characterized.<sup>29</sup>

Lectins constitute a very diverse group of proteins but only a limited number of folds have been described (Fig. 1A). Interestingly  $\beta$ -strand-containing secondary structures predominate (Fig. 1B–F). The dominant fold is the  $\beta$ -sandwich that is observed in almost 50% of lectins with known structure, which include jelly-roll topology as in legume lectins and galectins and immunoglobulin topology as in sialoadhesins. The  $\beta$ -prism fold is observed only in plant and algal lectins. In contrast, the  $\beta$ -trefoil fold is widely spread in all living kingdoms. C-type legume lectins observed in all animal kingdoms adopt a mixed  $\alpha/\beta$  topology. The  $\beta$ -propeller fold consists of 6 or 7 tandem



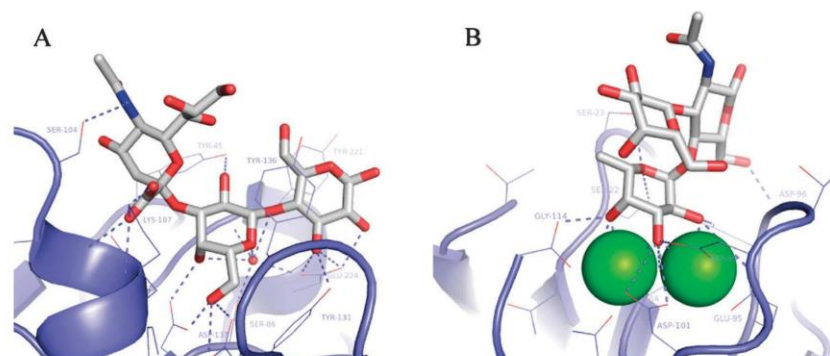
**Fig. 1** (A) Repartition of available crystal structures of lectins as a function of biological origin (left) and structural fold (right) (from <http://lectin3d.cermav.cnrs.fr/>). (B) Jelly-roll in dimeric legume lectins (PDB code 3ZYR: *Platypodium elegans* lectin–heptasaccharide).<sup>188</sup> (C)  $\beta$ -Prism in algal lectin (2HYQ: griffithsin–Man $\alpha$ 1-6Man).<sup>189</sup> (D)  $\beta$ -Trefoil in mammalian lectin (1FWV: Cys-rich domain of murine Man-receptor–sulfo-Lewis a).<sup>190</sup> (E) Mixed  $\alpha/\beta$  fold in invertebrate C-type lectin (2VUZ: codakine–nonasaccharide).<sup>191</sup> (F)  $\beta$ -Propeller of fungal lectin (2C4D: *Psathyrela velutina* lectin–GlcNAc).<sup>192</sup> Proteins are displayed as ribbons, carbohydrates as sticks and ions as spheres. All molecular drawings have been prepared with Pymol ([www.pymol.org](http://www.pymol.org)).

repeats of small  $\beta$ -sheets, referred to as blades, folded in a cyclic arrangement, with carbohydrate binding sites between blades. The different  $\beta$ -propellers observed in invertebrates, bacteria and fungi indicate that this particular shape has

converged through evolution in order to meet one of the most important features of lectins: multivalency.

The carbohydrate binding sites are generally shallow grooves or pockets located at the protein surface, or at the





**Fig. 2** Examples of lectin binding sites. (A) *Maackia amurensis* leucoagglutinin complexed with sialyllactose (NeuAc $\alpha$ 2-3Gal $\beta$ 1-4Glc) (PDB code 1DBN)<sup>193</sup> displaying hydrogen bonds, water bridged contacts, hydrophobic contact with aromatic side chains, and salt bridge. (B) *Pseudomonas aeruginosa* lectin B complexed with Lewis x (Le<sup>x</sup>: Gal $\beta$ 1-3[Fuc $\alpha$ 1-4]GlcNAc) (PDB code 1W8N)<sup>194</sup> with coordination of two calcium ions.

interface of oligomers. The numerous hydroxyl groups of glycan are involved in a network of hydrogen bonds which are the basis of affinity (Fig. 2). Nevertheless, hydrophobic interactions, often involving aromatic amino acids, are crucial since they result in high selectivity for the carbohydrate ring configuration.<sup>30</sup> Ionic interactions are rarely observed and are limited to specific monosaccharides such as sialic acid (NeuAc). Several families of lectins require that the carbohydrate participates in the coordination of one or two calcium ions.<sup>31,32</sup>

## 2.2. Application in biotechnology, diagnostic and therapeutic

Lectins are rarely used directly as therapeutic agents, with the notable exception of small mannose binding lectins from cyanobacteria and algae that display anti HIV-properties, because of their ability to bind to high mannose oligosaccharides present on gp120.<sup>33,34</sup> Although some lectins may directly display anti-tumour properties,<sup>35</sup> most often, they are used as drug carriers for targeting glycoconjugates associated with a disease, such as colorectal tumors.<sup>36</sup> The lectin from the plant *Morus nigra* displays high affinity towards tumour-associated T and Tn antigens and has been tested for targeting photosensitizers to human tumour cells for phototherapy application.<sup>37</sup>

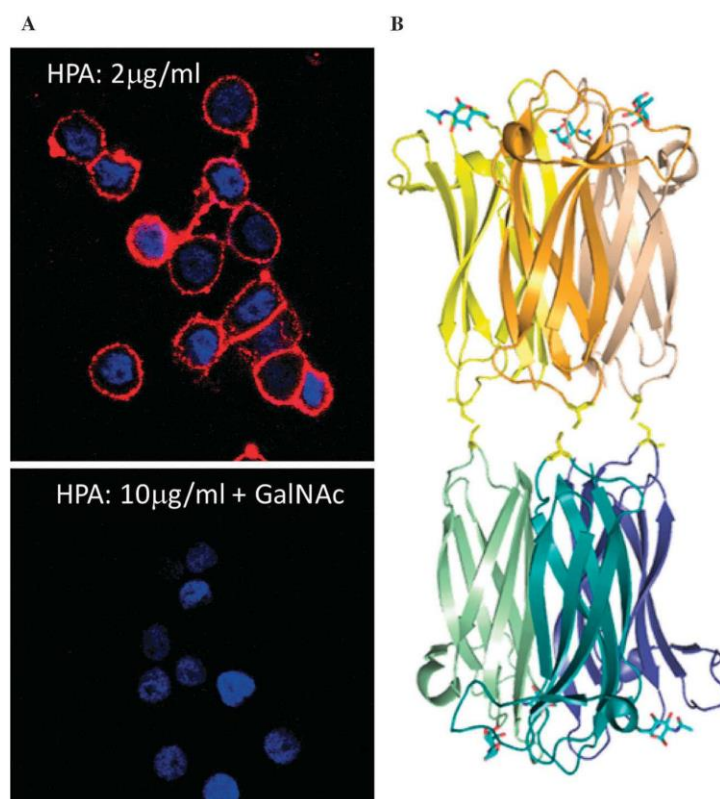
In academic research laboratories, lectins are routinely used for characterization of carbohydrate structures, for purification of glycoproteins and for labelling of specific epitopes on cell surfaces. Serial lectin-agarose chromatography has been developed as a rapid technique for fractionation and analysis of N-linked oligosaccharides and glycoproteins<sup>38</sup> and this method is now adapted for fractionation and glycomic analysis of plasma glycoproteins.<sup>39</sup> Historically, lectins have also been used in semi-clinical approaches for cell typing and cell separation.<sup>40</sup> Lectins are used in histochemistry for cancer diagnosis since glycosylation modification on cell surfaces has been directly correlated with tumour progression and metastatic potential.<sup>41</sup> For example, the lectin from the roman snail *Helix pomatia* (HPA) binds to GalNAc and T-antigen<sup>42,43</sup> (Fig. 3) and the cell labelling has a predictive value for breast cancer recurrence and survival.<sup>44</sup>

In the recent years, the use of lectins in biotechnology, diagnostic and therapeutic, has seen a new burst due to the development of novel technologies for lectin-based profiling approaches, with the common aim of defining the “glycan signature” associated with cell types, diseases or organisms. Lectin arrays of different types have been rapidly developed (for review, see ref. 5 and 45). Depending on the degree of miniaturization, lectins can be immobilized onto microplates, plotted on glass slides or attached to various surfaces with a large panel of immobilization protocols.<sup>46</sup> Association with a protein fractionation step is often helpful for analyzing glycosylation in complex protein mixtures such as biological fluids.<sup>47</sup> Lectin arrays are also now available from several biotech companies either as analysis service or as turnkey package (Table 1).

Lectin arrays find many applications in profiling of glycosylation. Breast cancer cells can be distinguished from normal ones and correlation of glycosylation changes with metastatic potential and tissue homing specificity have been established.<sup>48</sup> A more recent study confirmed the difference in lectins profiling for breast cancer cells compared to normal ones but furthermore established that specific profiles could also be observed in the serum and urine of patients with metastatic breast cancer.<sup>49</sup> The power of lectin array technology is also highlighted by the possibility of selecting and characterizing a specific clone from the mother cell line, based on the slight difference in the respective glycan signatures.<sup>50</sup> Lectin arrays are used to distinguish induced pluripotent stem cells from the somatic cells from which they are produced. High-throughput discrimination between bacteria based on their binding pattern on lectin micro-arrays has also been developed.<sup>51</sup>

## 3. Synthetic lectins from organic chemistry

Carbohydrate recognition through noncovalent interactions has been a challenge for biomimetic and supramolecular chemistry, especially for obtaining efficient binding in water.<sup>52-56</sup> The efforts in the domain allowed progress in understanding



**Fig. 3** (A) Specific labeling of HT29 tumor cells by HPA, the GalNAc-specific lectin from *Helix pomatia*.<sup>42</sup> (B) Crystal structure of the HPA-GalNAc complex (PDB code 2CCV)<sup>43</sup> displaying the hexameric organization and six GalNAc binding sites.

**Table 1** Available commercial lectin plates and arrays

Company	Product	Type of product
Galab	Lectin microplates	Service analysis
GlycoDiag	Lectin microplates and personalized lectin kits	Service analysis
GP Biosciences	LecChip™ service analysis and lectin microarray on slide	To be used with GlycoStation™ reader
Procognia	GlycoScope™ lectin microarray on slide	To be used with fluorescence reader

the basis of carbohydrate recognition and in the role of water molecules (Table 2).

### 3.1. Boronate based systems

The ability of boronic acids to bind to diols gives them lectin-like properties that have been exploited since the 1950s.<sup>57</sup> Binding of boronic acids to carbohydrates is influenced by pH, solvent, chemical group linked to boron and type of diol. Consequently, researchers developed many structures aiming at enhancing water solubility, fluorescence properties of the reporter group, affinity and selectivity for carbohydrates (Fig. 4).<sup>58</sup> Tetrathiafulvalene-anthracene boronic acids **1**<sup>59</sup> or derivatives of benzonitrile boronic acids **2**<sup>60</sup> targeted fructose in non-aqueous solvents (Table 2). One of the key advances was the design of compounds able to recognize sugars in aqueous

solution at neutral pH, such as benzo[*b*]thiophene boronic acid **3**.<sup>61</sup> Another bottleneck is the specific binding to furanose rings, and not to pyranose ones that are most interesting in biological situations. This has been achieved with the use of *ortho*-hydroxyalkyl phenylboronic acid derivatives **4** that bind 1–4 diol on glycopyranoses in water (Fig. 5).<sup>62,63</sup> Similarly, isoquinolinyboronic acid derivatives **5**<sup>64</sup> are able to recognize  $\alpha/\beta$ -methyl-pyranoside sugars under physiological conditions.

The synthesis of oligomeric boronic acid was determinant to increase the affinity for glycans and to obtain molecules called “boronlectins”. Bisboronic acids **6** showed strong binding to glucose albeit with modest selectivity.<sup>65</sup> Dianthracene-boronic compounds **7**,<sup>66</sup> as well as pyrene-containing molecular tweezer **8**,<sup>67</sup> reached a higher selectivity for glucose. Derivatives of these compounds were also able to recognize selectively



**Table 2** Artificial lectins from organic chemistry with carbohydrate ligands and affinity values (see Fig. 4 and 5)

Comp.	Ligand	Method	$K_D$ (M)	Ref.
1	Fru	Fluorescence	$9 \times 10^{-3}$	59
2	Fru	Fluorescence	$6 \times 10^{-3}$	60
3	Fru	Fluorescence	$6 \times 10^{-4}$	61
4	Glc	ARS UV assay	$3 \times 10^{-2}$	62, 63
5	Glc	Fluorescence	$3 \times 10^{-1}$	64
6	Glc	NMR	ND	65
7	Glc	NMR	$7 \times 10^{-4}$	66
	SLe <sup>X</sup>	Fluorescence	ND	68
8	Glc	Fluorescence	$4 \times 10^{-4}$	67
9	Lacto- <i>N</i> -tetraose	Fluorescence	$6 \times 10^{-5}$	69
10	Lewis epitopes	CD	$10^{-5}$ – $10^{-6}$	70
11	Man	Partition equilibrium	$10^{-5}$	82
12	GM3, SLe <sup>X</sup>	NMR	$10^{-2}$	83
13	Lipid A	UV	$6 \times 10^{-7}$	84
14	Octyl- $\beta$ -Glc	NMR	$10^{-3}$	89
15	Me- $\beta$ -Glc	Fluorescence	$3 \times 10^{-2}$	90
	Me- $\beta$ -GlcNAc	NMR	$2 \times 10^{-3}$	91
16	Me- $\beta$ -cellobiose	NMR, ICD, fluo	$10^{-3}$	92
17	Me- $\beta$ -cellobiose	NMR	$2 \times 10^{-4}$	93
18	Me- $\beta$ -Glc	NMR, ITC, fluo	$10^{-2}$	96

hepatocellular carcinoma cells that express sialyl Lewis X (SLe<sup>X</sup>: NeuAc $\alpha$ 2-3Gal $\beta$ 1-4[Fuc $\alpha$ 1-3]GlcNAc).<sup>68</sup> A bivalent derivative of  $\alpha$ -amidoboronic acid **9** binds efficiently to lacto-*N*-tetraose (Gal $\beta$ 1-3GlcNAc $\beta$ 1-3Gal $\beta$ 1-4Glc).<sup>69</sup> Other oligosaccharidic structures

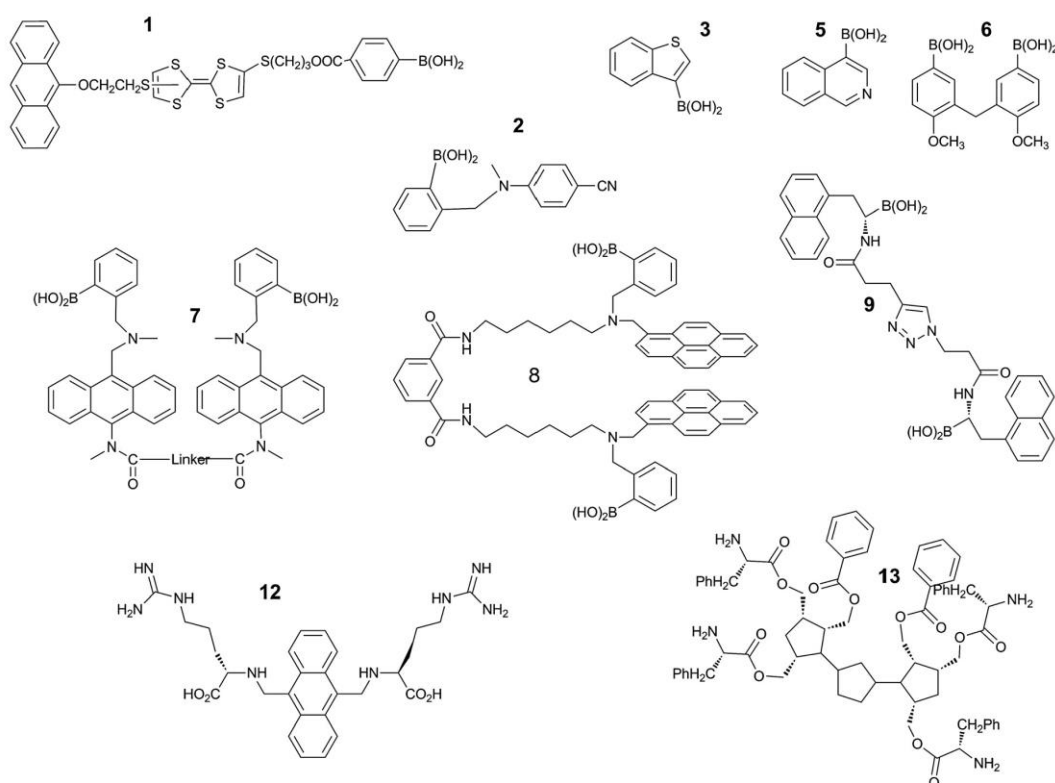
have been targeted since di- and tetra-boronic acids with cerium(IV) bis(prophyrinate) **10** are able to recognize and differentiate Lewis epitopes (Fig. 5).<sup>70</sup>

Recent developments have been turned to the incorporation of boronate into matrix, polymers, surfaces or nanoparticles in order to produce highly multivalent compounds able to bind to glycoproteins or cell surfaces. A matrix functionalized with acrylamidophenylboronic acid was able to separate glycopeptides and glycoproteins from non-glycosylated peptides and proteins.<sup>71</sup> Benzoboroxole-functionalized polymer binds to gp120 and inhibits HIV infection.<sup>72,73</sup>

Very promising results have also been obtained by several groups by developing libraries of peptide-based boronolactins (PBLs), using a combinatorial approach for generating large diversity.<sup>74–77</sup> Tumour marker oligosaccharides such as the Thomsen–Friedenreich (TF) antigen were efficiently targeted by such libraries.<sup>78,79</sup> These molecules can now be associated with synthetic lectin arrays that can fingerprint the glycosylation of cell lines and therefore offer promising tools for cancer diagnosis.<sup>80</sup>

### 3.2. Bio-inspired artificial receptors

Natural carbohydrate–protein interactions are based on a combination of hydrogen bonds and hydrophobic contacts and their

**Fig. 4** Artificial lectin-like receptors (see references in Table 2).

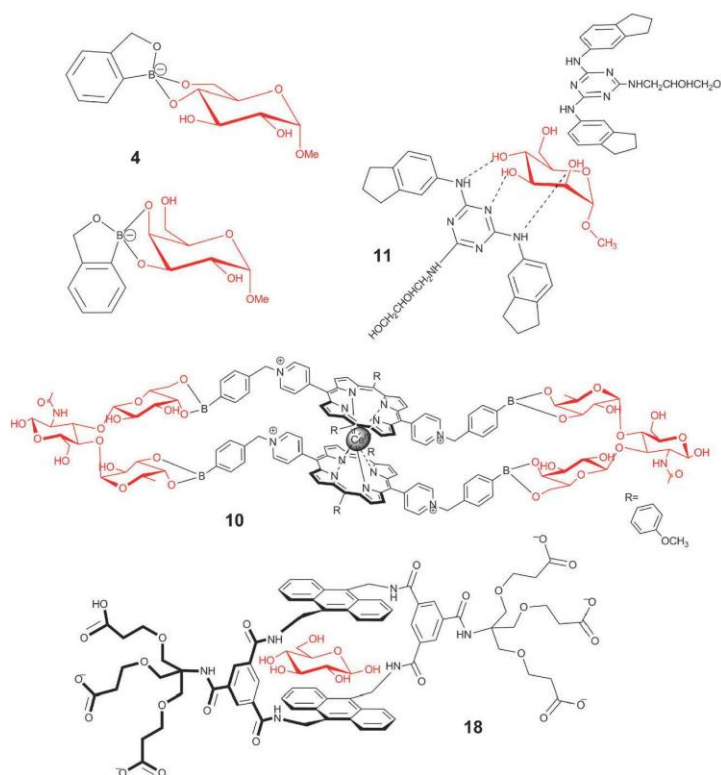


Fig. 5 Artificial lectin-like receptors with proposed binding mode to carbohydrates (see references in Table 2).

appropriate geometry and topology provide ligand specificity.<sup>81</sup> This served as an inspiration for designing artificial lectins and numerous substance classes (aminopyridine, crown ether, ...) have been tested and assembled for their potential to bind carbohydrates. In most cases, carbohydrate binding occurs only in organic solvent, and these molecules will not be described here (for review see ref. 52, 54 and 56). Nevertheless, several groups managed to synthesize compounds able to bind to sugars in water (Fig. 4 and 5). A homobifunctional triazine **11**, synthesized using combinatorial chemistry and attached to an agarose support for affinity chromatography applications, is able to bind to fucose and mannose containing sugars and glycoproteins with a micromolar affinity.<sup>82</sup> Amphiphilic anthracene–amino acid conjugates **12** are simple receptors that can bind GM3 and sialyl Lewis X.<sup>83</sup> A highly substituted oligocyclopentane **13** was able to bind to lipid A from cell walls of gram negative bacteria.<sup>84</sup>

The recent molecular imprinting technology can be used for constructing cavities shaped to specifically bind target molecules inside a polymeric material. Molecularly imprinted polymeric micelles (MIPMs) have been used with glucose as the template molecule.<sup>85,86</sup> One of the resulting glucose sensor showed good response and selectivity, particularly with a linear response from 0.2 mM to 8 mM, in addition to

a good stability and reversibility.<sup>86</sup> A library of molecularly imprinted polymers (MIPs) that bind to sulfated oligosaccharides has been obtained from a heparan-sulfate like disaccharide.<sup>87</sup> MIPs were also prepared in the presence of a synthetic galactoside locked in a boat conformation and the resulting material displayed good conformation selectivity on carbohydrate rings.<sup>88</sup>

Supramolecular architectures have been used to make synthetic lectins based on a cage scaffold with a cavity presenting hydrophobic parallel aryl rings as floor and ceiling, linked by pillars capable of hydrogen binding (for review see ref. 52). A synthetic receptor **14** for octyl-glucosides has been reported with a millimolar dissociation constant in organic solvent.<sup>89</sup> Substitution with hydrophilic groups **15** led to a water soluble compound able to recognize mono- and disaccharides in water with significant specificity for methyl- $\beta$ -D-glucoside.<sup>90</sup> Such structures also display a significant affinity for O-GlcNAc-containing glycopeptides.<sup>91</sup> Larger cages **16**, **17** were able to bind selectively to disaccharides such as cellobiose.<sup>92,93</sup> The balance between CH- $\pi$  interactions and hydrogen bonds in carbohydrate recognition could be studied by varying the cage architecture.<sup>94,95</sup> Very recently, a new glucose specific monocyclic compound **18** was synthesized in only five steps, with potential application in glucose monitoring (Fig. 5).<sup>96</sup>



## 4. Sugar binding by small biological molecules: oligonucleotides and peptides

### 4.1. Aptamers

Aptamers are single-stranded oligonucleotides able to bind specifically small molecules, macromolecules or whole cells.<sup>97</sup> They are now widely used in bioanalytical and biomedical sciences<sup>98</sup> since they are easy to produce, cheaper than antibodies, and can be integrated in many detection devices.<sup>99</sup> The higher affinity of aptamers makes such oligonucleotides attractive alternatives to lectins or antibodies as detection reagents for carbohydrate antigens. As a result, a variety of aptamers targeted for carbohydrate recognition have been recently described, targeting monosaccharides, oligo- and polysaccharides as well as antibiotics and glycoproteins (Table 3 and for review, see ref. 100).

Aminoglycoside antibiotics have been selected during microbes evolution for binding to the RNA portion of ribosomes during protein biosynthesis.<sup>101</sup> Because of their natural ability to interact with nucleotides, they are a target of choice for developing aptamers. High affinity RNA aptamers have been obtained by selection against tobramycin,<sup>102,103</sup> neomycin<sup>104</sup> and streptomycin.<sup>105</sup> NMR and crystal structures of complexes demonstrated the inclusion of the antibiotic in the nucleotides groove (Fig. 6).<sup>106–108</sup> The binding is based on strong electrostatic complementarity as well as a hydrogen bonding network that involves the carbohydrate rings of the ligands. As an application, RNA aptamers may be useful in clinically based assays to detect aminoglycoside antibiotics in blood serum.<sup>109</sup>

Oligonucleotide aptamers can select carbohydrate ring configurations and discriminate monosaccharides such as D-galactose, D-glucose, and D-mannose.<sup>110</sup> They also bind specifically to different anomers and disaccharide linkages. For example, aptamers bind with high selectivity to cellobiose with little or no affinity for the related disaccharides lactose, maltose, and gentiobiose.<sup>111</sup> Oligosaccharides of biological interest have been targets for aptamer development. Sialyllactose, as the sugar head group of the GM3 ganglioside, has been targeted

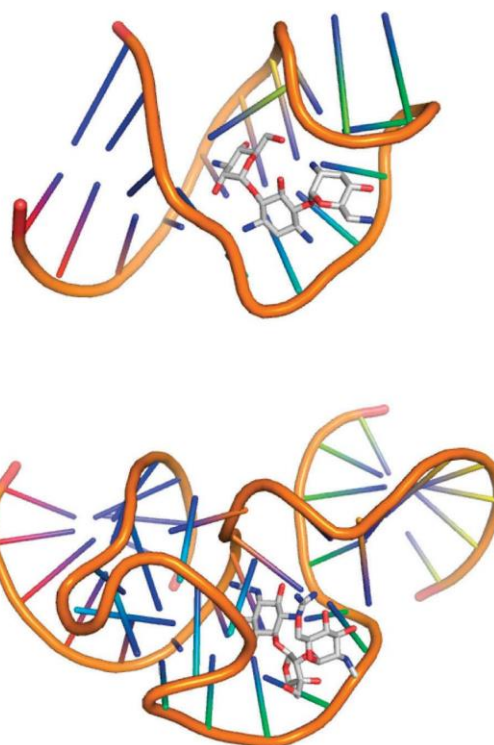


Fig. 6 Top: solution structure of an RNA aptamer complexed with neomycin B (PDB code 2TOB).<sup>104</sup> Bottom: crystal structure of an RNA aptamer complexed with streptomycin (PDB code 1TNB).<sup>105</sup>

with DNA aptamers. Inclusion of a positively charged amino group on a thymidine increased the affinity for the negatively charged trisaccharide.<sup>112</sup> SLe<sup>x</sup> tetrasaccharide has been targeted by RNA aptamers with affinity between 10<sup>-9</sup> and 10<sup>-11</sup> M, which is similar or better to that obtained with antibody. Such aptamers showed inhibitory activity for selectin-mediated cell adhesion.<sup>113</sup> Glycoproteins can also be recognized since a boronic acid modified aptamer binds efficiently to a fibrinogen (6 nM) and loses the affinity when the protein is deglycosylated.<sup>114</sup>

Aptamers can be directed towards surfaces, such as cells, and also towards insoluble polysaccharides such as the ones present in plant and insect cell walls. Polymers of β1 → 4 glucose (cellulose), β1 → 4 GlcNAc (chitin) and β1 → 3 glucose (curdlan) have been targeted by RNA and DNA aptamers.<sup>115–118</sup> Many efforts have been devoted to the detection of bacterial lipopolysaccharides (LPS or endotoxins) because of their toxicity and involvement in sepsis shock. DNA aptamers targeted against LPS demonstrated to be efficient for blocking the toxicity of this molecule in cellular or animal assays.<sup>119</sup> Electrochemical sensors capable of selectively detecting LPS from crude samples such as biological liquors are now developed based on these DNA aptamers.<sup>120,121</sup> RNA aptamers specific for highly poisonous *Escherichia coli* O157:H7 strain have been developed for rapid detection of food pathogens.<sup>122</sup>

Table 3 Aptamers with lectin-like activity with carbohydrate ligands and affinity values, when available

Nucleotide	Ligand	Method	K <sub>D</sub> (M)	Ref.
RNA	Gal	SPR	5.5 × 10 <sup>-7</sup>	110
RNA	Tobramycin	Fluorescence	7.7 × 10 <sup>-10</sup>	102
RNA	Neomycin	ITC	1.45 × 10 <sup>-6</sup>	104
RNA	Streptomycin			105
RNA	SLe <sup>x</sup>	SPR	8.5 × 10 <sup>-11</sup>	113
mDNA <sup>a</sup>	Sialyllactose		2.9 × 10 <sup>-6</sup>	112
mDNA <sup>a</sup>	Fibrinogen	Equ. filtration	6 × 10 <sup>-9</sup>	114
DNA	Tn antigen	SPR	21 × 10 <sup>-9</sup>	195
DNA	1,3-Glucan	Fluorescence	3 × 10 <sup>-7</sup>	117
DNA	Cellulose			115
RNA	Cellulose			116
DNA	LPS	SPR	2.65 × 10 <sup>-9</sup>	119
DNA	LPS	SPR	12 × 10 <sup>-9</sup>	120
RNA	Sephadex		n.a.	196

<sup>a</sup> mDNA: modified DNA.



## 4.2. Peptides

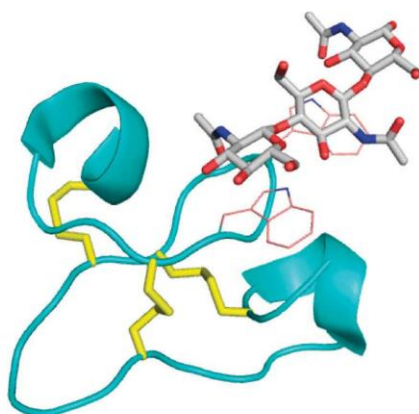
Peptides interacting with carbohydrates can be of natural origin, such as small lectins, can be derived from natural lectins or antibodies, or can be obtained by selection methods using biotechnology approaches (Table 4).

A family of small chitin-binding lectins, named hevein-like peptides, has been identified in several plants. In addition to hevein from laticifers of rubber tree (43 aa),<sup>123</sup> lectins such as Ac-AMP2 from the seeds of *Amaranthus caudatus* (30 aa),<sup>124</sup> Pn-AMP from the seeds of *Pharbitis nil* (40 aa)<sup>125</sup> and EAFP2 from the bark of *Eucommia ulmoides* (41 aa)<sup>126</sup> share the same compact fold stabilized by three to five disulfide bridges (Fig. 7). An artificial mini-hevein of 32 residues (HEV32), synthesized by solid-phase methodology, maintains its affinity to chitotriose.<sup>127</sup> NMR experiments established the importance of tryptophan stacking for the affinity to GlcNAc residues. The chitin-binding activity of this class of lectin has generally been correlated with antifungal properties. SHL-I, a 32 residue peptide with similar structure, has been identified from the venom of the Chinese bird spider *Selenocosmia huwena*.<sup>128</sup> A combination of Laser Chemically Induced Nuclear Polarization, NMR and molecular modeling methods established that this lectin binds to sialic acid and sialylated oligosaccharides.<sup>129</sup>

**Table 4** Natural peptides with lectin-like activity with carbohydrate ligands and affinity values, when available

Lectin	Size <sup>a</sup>	Ligand	Meth.	$K_D$ (M)	Ref.
Odorranalectin	17	Fucose	SPR	$5 \times 10^{-5}$	133
Retrocyclin-1	18	Galactosylceramide	SPR	$2.4 \times 10^{-8}$	131
Retrocyclin-1	18	Gp120	SPR	$3.5 \times 10^{-8}$	131
Retrocyclin-1	18	CD4	SPR	$3.1 \times 10^{-8}$	131
Ac-AMP2	30	(GlcNAc) <sub>3</sub>	NMR	$1 \times 10^{-3}$	124
HEV32	32	(GlcNAc) <sub>3</sub>	NMR	$1.3 \times 10^{-4}$	127
Galectin-3	6	Lactose	SAW	$1.2 \times 10^{-2}$	138

<sup>a</sup> Size is given as a number of amino acids.



**Fig. 7** NMR structure of the complex between minimum-sized hevein and chitotriose (PDB code 1TOW).<sup>127</sup> Disulfide bridges are represented by yellow sticks, and the crucial tryptophan residue is represented by lines.

Interestingly, other anti-microbial peptides with different structures have also been demonstrated to display lectin activity. Defensins are endogenous, cysteine-rich, antimicrobial peptides that are involved in innate immunity against pathogens. Among them, the  $\theta$ -defensins are cyclic antimicrobial octadecapeptides with lectin activity.<sup>130</sup> They are expressed only in Old World monkeys but the pseudo-gene, still present in humans, has been used for synthesizing retrocyclins. These peptides, a class of artificial  $\theta$ -defensins, bind to galactosylceramide, fetuin and glycosylated gp120 and are able *in vitro* to protect cells from HIV-1 infection.<sup>131</sup> Four of the six human  $\alpha$ -defensins also have a lectin-like ability to bind glycosylated proteins.<sup>132</sup>

The smallest natural peptide with lectin activity, odorranalectin, has been purified from amphibian skin secretions. It is composed of 17 residues arranged in a  $\beta$ -turn stabilized by disulfide bridges. It is specific for fucose, and its ability to agglutinate bacteria indicates that it may have antimicrobial properties.<sup>133</sup>

Production of glycan-binding peptides by proteinase digestion of legume lectins followed by affinity chromatography has been pioneered by Roberts and Goldstein.<sup>134</sup> Pepsin and trypsin digested peptides from lima bean lectin were able to bind on the affinity column functionalized with  $\alpha$ -GalNAc. Similar properties were obtained with peptides from *Cytisus scoparius* seed lectin.<sup>135</sup> Peptides with specificity towards fucose were identified from the digestion of *Lotus tetranogolobus* agglutinin and *Ulex europaeus* agglutinin I (UEA-1) whereas GlcNAc-specific ones were obtained from UEA-II and Gal-specific ones from *Bauhinia purpurea* lectin.<sup>136,137</sup> In all cases, the retarded peptides were found to be parts of the metal-binding regions of these lectins that also correspond to the carbohydrate binding site. As for animal lectins, a mass spectrometry approach was used to identify the proteolytic fragments of galectin-1 and -3 that are in contact with the lactoside ligand. The resulting peptides proved to be active in blocking lectin binding to glycoproteins and cell surfaces.<sup>138</sup>

A knowledge-based approach was used to design and synthesize peptides from the crystal structure of pertussis toxin, some of them being able to bind to sialylated conjugates.<sup>139</sup> Similarly, peptide analogues to region 109–118, which contains residues that have been identified as critical for E-selectin binding to SLe<sup>x</sup>, have been synthesized.<sup>140</sup> They display inhibitory properties against the binding of E-selectin and P-selectin to SLe<sup>x</sup>-coated surfaces. Peptides were designed to mimic the binding sites of three members of the siglec family: sialoadhesin, MAG and CD22.<sup>141</sup> Infrared atmospheric pressure matrix-assisted laser desorption/ionization quadrupole ion trap mass spectrometry was applied to characterise the binding of these peptides to 3'-sialyllactose and 6'-sialyllactose.

Antibodies have been more rarely used as templates for designing carbohydrate-specific peptides. This approach has been successful in using information obtained by NMR and molecular modeling for the binding of mAb 3F8 to ganglioside GD2. Three cyclic peptide ligands were obtained and it could be demonstrated that their selective binding to GD2 operates through an induced-fit binding mechanism.<sup>142</sup>



**Table 5** Artificial peptides with lectin-like activity with carbohydrate ligands and affinity values, when available

Size <sup>a</sup>	Ligand	Method	K <sub>D</sub> (M)	Ref.
12	SLe <sup>x</sup>	SPR	1.5 × 10 <sup>-4</sup>	149
12	Le <sup>x</sup>	SPR	2 × 10 <sup>-5</sup>	149
12	LacNAc	SPR	3 × 10 <sup>-4</sup>	149
12	Sulfo Le <sup>a</sup>	SPR	4 × 10 <sup>-6</sup>	150
12 × 2	SLe <sup>x</sup>	SPR	7 × 10 <sup>-5</sup>	148
12 × 4	SLe <sup>x</sup>	SPR	1 × 10 <sup>-7</sup>	148
8	Heparin	Affinity coelectrophoresis	2 × 10 <sup>-4</sup>	197
15	T-antigen	Fluorescence quenching	1 × 10 <sup>-5</sup>	143
15	GM3	QCM	1.5 × 10 <sup>-6</sup>	146
15	GM1			145
35	GM1	SPR	2.4 × 10 <sup>-7</sup>	144
15	GT1b	SPR	9.7 × 10 <sup>-7</sup>	147
15	GT1b	SPR	6.4 × 10 <sup>-7</sup>	147
20	3'-Sialyllactose			
20	6'-Sialyllactose			
	LPS	ITC	1.75 × 10 <sup>-6</sup>	152

<sup>a</sup> Size is given as the number of amino acids.

Peptide engineering through a phage display approach followed by affinity selection named biopanning resulted in novel glycan-binding peptides (Table 5). Two peptides of 15 amino acids directed against T antigen (Galβ1-3GalNAc) were able to disrupt carbohydrate mediated cell adhesion and to inhibit asialofetuin induced melanoma cell aggregation.<sup>143</sup> Phage-display technology was also used for selecting ganglioside-binding peptides with specificity towards GM1 and GM3.<sup>144–146</sup> GM1-binding peptides have the ability to inhibit the binding of cholera toxin to GM1<sup>145</sup> while GM3-binding peptides inhibited infection by influenza virus in a cellular assay.<sup>146</sup> Very recently, GT1b the glycolipid which is the ligand for tetanus toxin was employed to obtain the peptide sequences that bind specifically to trisialyloligosaccharides.<sup>147</sup> Similarly, peptides were selected by a biopanning method to recognize sialylated and sulfated Lewis antigens.<sup>148–150</sup> The authors could demonstrate that the use of tentacle-type dimers or tetramers, presenting multiple copies of the peptide, is very efficient for gaining affinity towards the target glycan.

The development of high-throughput assays and cost-efficient chemical synthesis of peptides open the route to the production of random sequence peptide arrays. Such arrays of 10 000 peptides used with luminescent glyconanoparticles resulted in the identification of two sets of peptides that differentially bound 3'-sialyllactose and 6'-sialyllactose.<sup>151</sup> The same approach was used for bacterial glycoprofiling and resulted in the identification of LPS-specific peptides.<sup>152</sup>

## 5. Engineering of neolectins from glycosidases and lectins

### 5.1. Inactivation of glycosidases

Since glycosidases have to bind to oligosaccharides before hydrolysis, they can be used as lectins, provided that they are inactivated or used under conditions of low turnover number. Although production of inactive lysozyme by mutation of catalytic amino acids was demonstrated earlier,<sup>153</sup> it is only in 1994 that lysozyme was specifically engineered for carbohydrate recognition and detection and named "neolectin".<sup>154</sup> The appellation "Lectenz" has now been trademarked (Glyco-sensors and Diagnostics, Athens, GA) for lectins engineered from a carbohydrate-processing enzyme through site-directed or random mutagenesis. A chitin-binding protein usable for diagnosing fungal infections has been obtained by deactivating bacterial chitinase from *Vibrio parahaemolyticus*.<sup>155</sup> The Chit42 chitinase from *Trichoderma harzianum* could also be engineered with inactive T133Q and T133N mutants displaying broadened substrate specificity for β-4-galactosylated chitotetraose, while the T133D mutation led to improved specificity toward GlcNβ1-4(GlcNAc)<sub>4</sub>.<sup>156</sup> A catalytically inactive endosialidase from bacteriophage has been demonstrated to have lectin activity towards polysialic acid. The whole bacteriophage could be used for detection purpose,<sup>157</sup> as well as the recombinant GFP-fusion neolectin.<sup>158</sup>

### 5.2. Engineering of lectins

Engineering of lectins is often performed to direct their ligand specificity towards new substrates, as described below. However, multivalency can also be altered by engineering. Indeed, most lectins are multivalent either by oligomerisation or tandem repeats on the peptide chain. Multivalency increases the avidity for cell surfaces but may also produce unwanted reactions such as agglutination of cells or precipitation of target glycoproteins.

**5.2.1. Tailoring the specificity of lectins.** Engineering of lectins in order to modify the specificity is a long time goal, but the number of successful results is still limited (Table 6). In a pioneering work, Sharon and co-workers used site-directed mutagenesis to increase the affinity of *Erythrina corallo-dendron* lectin towards galactose derivatives with bulky substituents at C2.<sup>159</sup> Similarly the specificity of the lima bean lectin, a galactose-specific legume lectin, was explored and the C127Y mutant displayed a higher affinity for the type-A trisaccharide compared to the wild type.<sup>160</sup>

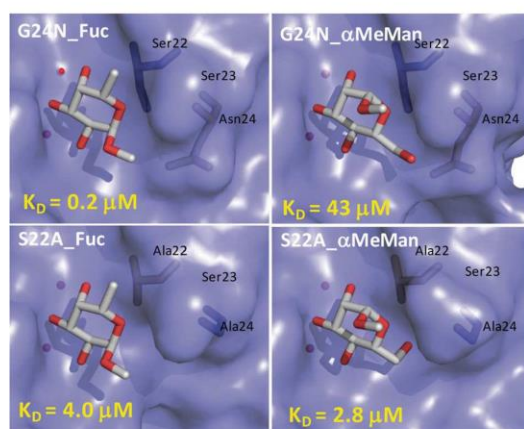
**Table 6** Engineered lectins with modified specificity with carbohydrate ligands and affinity values

Parent lectin	Ligand	K <sub>D</sub> (M)	Ref.
<i>E. corallo-dendron</i> lectin	β-N-Dansylgalactosaminide	1.2 × 10 <sup>-6</sup>	159
<i>E. corallo-dendron</i> lectin	α-N-Dansylgalactosaminide	8 × 10 <sup>-7</sup>	159
<i>P. aeruginosa</i> LecB	Me-β-Man	2.8 × 10 <sup>-6</sup>	164
<i>A. cylindracea</i> galectin	Biantenna with two 3'-sialyllactose	7 × 10 <sup>-6</sup>	166
<i>L. terrestris</i> EW29Ch	Biantenna with two 6'-sialyllactose	4 × 10 <sup>-5</sup>	167
	6S-Gal oligosacch	4 × 10 <sup>-6</sup>	168
<i>M. amurensis</i> lectin	6'-Sialyllactosamine	6 × 10 <sup>-7</sup>	178



Modifying the monosaccharide specificities of lectins is more difficult than tuning the oligosaccharide specificities since the main binding site has to be affected. The galactose-binding lectin from the bark of *Robinia pseudoacacia* (RBL) has been modified by insertion of carbohydrate- and metal-binding regions of mannose-specific lectins. Although these rRBLs agglutinated rabbit erythrocytes, the hemagglutination was not inhibited by mannose.<sup>161</sup> In contrast, based on the crystal structure of LecB(PA-IIL), a fucose-binding lectin from *Pseudomonas aeruginosa*,<sup>162</sup> and RS-IIL, a mannose-binding lectin from *Ralstonia solanacearum*,<sup>163</sup> it has been possible to design LecB mutants with either enhanced specificity for fucose, or switched preference for mannose (Fig. 8).<sup>164</sup> A chimeric lectin was constructed by assembling nonapeptides from *Bauhinia purpurea* lectin and *Lens culinaris* lectin with specificity towards galactose and mannose respectively. The chimeric lectin can bind both mannose and galactose.<sup>165</sup>

Sialylated oligosaccharides are of high biological importance and many efforts have been directed towards obtaining engineered lectins with high specificity towards NeuAc $\alpha$ 2-3Gal or Neu5Ac $\alpha$ 2-6Gal terminating oligosaccharides. *Agrocybe cylindracea* galectin has been modified in order to suppress the background affinity for lactoside and the E86D mutant substantially lost its binding ability to LacNAc and T-antigen, while it retained the high affinity for 3'-sialyllactose.<sup>166</sup> Directed evolution based on error-prone PCR has been used in order to produce NeuAc binding lectins starting from a ricin B-like Gal-binding lectin from *Lumbricus terrestris*. After selection with fetuin-agarose, one of the evolved mutants showed substantial affinity for Neu5Ac $\alpha$ 2-6Gal.<sup>167</sup> The same lectin was engineered for obtaining binding to oligosaccharides terminated by 6-sulfo-galactose (6S-Gal) which are associated with glioblastoma malignancy. Several rounds of selection indicated that the E20K mutation is necessary and sufficient to acquire the specificity for 6S-Gal.<sup>168</sup>



**Fig. 8** Structures and dissociation constants of mutant from *Pseudomonas aeruginosa* lectin B displaying enhanced affinity for fucoside (mutant G24N), and mutant with stronger binding to mannoside (mutant S22A).<sup>164</sup>

Random mutation in the lectin binding site demonstrated to be an efficient manner to produce libraries of mutants that display different fine specificities and can therefore be of interest for the screening approach of lectin arrays. The *Maackia amurensis* hemagglutinin (MAH) has been the subject of several studies. The 24 nucleotides of the binding site cDNA were randomly mutated and the 16 mutants obtained are able to agglutinate a variety of human and animal erythrocytes in highly specific patterns.<sup>169</sup> As a next step, 35 mutants were prepared by mutations of two amino acids within the carbohydrate-recognition domain and were used for profiling cell lineage and differentiation.<sup>170</sup> This library of genetically modified MAH demonstrated to be a useful tool for serum diagnosis of IgA nephropathy.<sup>171</sup>

**5.2.2. Altering the valency of oligomeric lectins.** Production of monomeric lectins started with concanavalin A, which exists as a tetramer at pH 7. Dimeric Con A derivatives were first produced by chemical modifications resulting in succinyl-Con A and acetyl-Con A.<sup>172</sup> A combination of succinylation and photolabeling by azidophenyl- $\alpha$ -D-mannopyranoside yielded monovalent ConA dimers, *i.e.* presenting a single saccharide-binding site at pH 5, and not showing hemagglutination activity.<sup>173,174</sup>

Monovalent derivatives of sialylated oligosaccharide-specific lectins demonstrated to be very useful for the analysis of cell surface glycoconjugates containing sialic acid by the technique of flow cytometry. For this purpose, stable monomeric subunits of *Maackia amurensis* leukoagglutinin (MMAL), specific for NeuAc $\alpha$ 2-3Gal, and of Japanese elderberry bark lectin (MSSA) specific for NeuAc $\alpha$ 2-6Gal were prepared by the selective reduction of disulfide bridges between the subunits followed by alkylation with 4-vinylpyridine.<sup>175,176</sup>

Site-directed mutagenesis was used to create a monomeric form of *Aleuria aurantia* lectin, AAL, that is normally found as a homodimer. A 20-fold higher concentration of mAAL was needed to induce hemagglutination when compared with the wild type. This difference is most likely explained by a higher availability of binding sites in the dimeric form compared with the monomeric form of AAL.<sup>177</sup>

On the other hand, increasing the multivalency of a lectin could be of interest for improving its affinity toward glycosylated surfaces. Two identical NeuAc $\alpha$ 2-6Gal specific lectins engineered from earth-worm R-type galactose-specific have been assembled *via* a polypeptide linker. This construct results in an almost 10-fold increase in affinity.<sup>178</sup> The resulting SRC2 binds strongly to erythrocyte surfaces without hemagglutinating them and is therefore of interest for flow cytometry analysis.

Introduction of cysteine residues able to create disulfide bridges between monomers has been used for oligomerisation that can be switched on or off through redox control. A strong increase in hemagglutination power is observed for the P71C mutant of monocot mannose-binding from *Galanthus nivalis* (GNA) compared to the wild type and this property can be switched off upon addition of DTT.<sup>179</sup>

**5.2.3. Altering the valency of tandem-repeat lectins.** Several lectins exhibit repetition of peptide sequences with a high



degree of similarity, yielding the occurrence of multiple binding sites. Folds such as the  $\beta$ -trefoil or the  $\beta$ -propeller are classical examples that are also observed in many other proteins with different functions.<sup>180</sup> The modular nature of such folds suggests that present day proteins evolved *via* duplication and recombination of the smaller functional elements.<sup>181</sup>

Tachylectin-2 is a five bladed monomeric  $\beta$ -propeller lectin from the horseshoe crab *Tachypleus tridentatus*. The five blades of Tachylectin-2 consist of very similar  $\beta$ -sandwiches composed of 4  $\beta$ -strands and have five binding sites located between them.<sup>182</sup> Genetically diverse libraries were created by incremental truncation of the Tachylectin-2 gene and screened for functional lectins. Several peptides corresponding to two blades were isolated with the ability to assemble into oligomers. The specificity and affinity are nearly the same between wild type and short segments Tachylectin-2.<sup>183</sup> However, the oligomerisation is different and  $\beta$ -strand swapping results in two intertwined five-bladed  $\beta$ -propellers (Fig. 9).<sup>184</sup> Generation of new lectins through duplication and fusion of five modules with the same sequence was also tempted but resulted in a

highly aggregated protein. Nevertheless when the fusion was performed on diversified sequence modules, several soluble proteins with a strong binding to glycoproteins were obtained.<sup>185</sup>

*Aleuria aurantia* lectin AAL also adopts a  $\beta$ -propeller fold, albeit with six blades.<sup>186</sup> A new protein was engineered by domain exchange in order to duplicate the binding sites that present a stronger affinity for fucose. The resulting mutant contains a 1-2-4-2-4-6 blade sequence and binds to the  $\alpha$ -1,6 fucosylated oligosaccharides 8 to 10 higher affinity than the wild type.<sup>187</sup>

## 6. Conclusions and perspectives

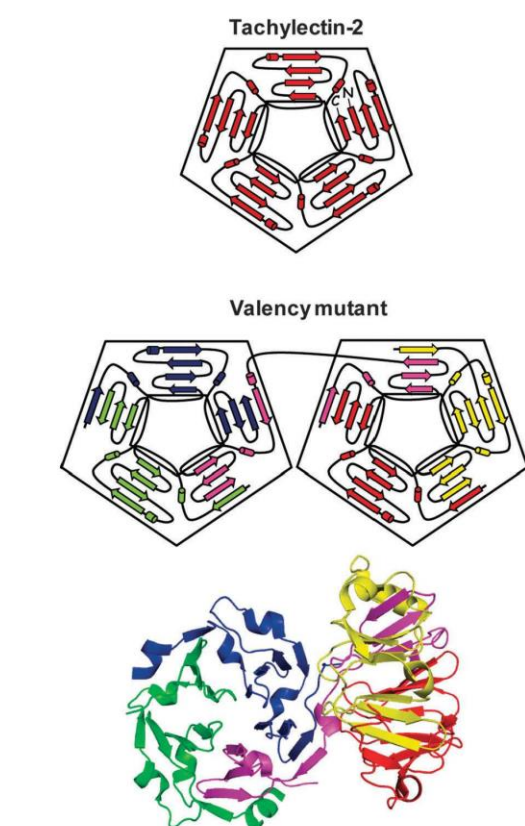
Nature provides a wide panel of lectins from diverse origins, and their large range of specificity has made them basic tools in many domains of research and biotechnology. Natural lectins have been of first importance for our understanding of the molecular basis of protein-carbohydrate interactions. The on-going characterisation of novel lectin specificities and folds in branches of living kingdom that were not much explored (alga, mollusc, ...) is a source of new tools to be exploited. As an example, there is still a lack of available lectins for detecting some glycans of biological importance such as sulphated oligosaccharides.

Nevertheless, novel glycotecnologies for analysis of glycomes, such as fluorescent biomarkers and lectin arrays, require high quality and reproducibility of lectins that is not always possible to attain from natural product purification. Important goals are to develop more extensively the production of recombinant lectins and to master their engineering towards novel specificity. For the glycome profiling applications, but also for the ones including cell-surface labeling and delivery of imaging or therapeutical compounds, other carbohydrate-binders, or artificial lectins, are now being developed based either on biomolecules or on biomimetics systems. The use of mixed approaches including boronic acid-modified peptides or nucleotides is very promising and crossbreeding between biomolecules obtained from molecular biology and sensors synthesized by organic or supramolecular chemistry is clearly a route to be developed.

We can also recall that lectins have evolved as receptors able to bind carbohydrates in a very efficient way, due to their multivalency that results in strong binding for clustered glycans of the right type. Design of engineered and artificial lectins should reckon this property by assembling binding sites in topology that are best adapted to the targeted glycoconjugates. Taking into account our current knowledge on protein-carbohydrate interactions and the broad range of applications for glycan-binders, there is no doubt that an increasing number of lectins from all origins will become available.

## Acknowledgements

The authors acknowledge funding from Agence Nationale de la Recherche Grant NeoLect (ANR-11-BSV5-002) (AA and AI), from CNRS (AI) and from Université Joseph Fourier (JA). The COST



**Fig. 9** Comparison of topology of wild-type Tachylectin-2 and of the valency mutants obtained with short peptides corresponding to the two-blades repeat. Crystal structure of pentameric valency mutants is represented with different colours for each peptide (PDB code 3K1F).<sup>184</sup>

actions CM1102 and BM1003 and the Labex ARCANÉ (ANR-11-LABX-003) are acknowledged for their support.

## References

- G. Opdenakker, P. M. Rudd, C. P. Ponting and R. A. Dwek, *FASEB J.*, 1993, 7, 1330–1337.
- R. Apweiler, H. Hermjakob and N. Sharon, *Biochim. Biophys. Acta*, 1999, 1473, 4–8.
- A. Varki, R. D. Cummings, J. D. Esko, H. H. Freeze, P. Stanley, C. R. Bertozzi, G. W. Hart and M. E. Etzler, *Essentials of Glycobiology*, 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2009.
- S. Cunningham, J. Q. Gerlach, M. Kane and L. Joshi, *Analyst*, 2010, 135, 2471–2480.
- J. Hirabayashi, A. Kuno and H. Tateno, *Electrophoresis*, 2011, 32, 1118–1128.
- K. L. Hsu and L. K. Mahal, *Curr. Opin. Chem. Biol.*, 2009, 13, 427–432.
- S. C. Tao, Y. Li, J. Zhou, J. Qian, R. L. Schnaar, Y. Zhang, I. J. Goldstein, H. Zhu and J. P. Schneck, *Glycobiology*, 2008, 18, 761–769.
- B. B. Haab, *Proteomics Clinical Applications*, 2012, 6, 346–350.
- K. L. Hsu, J. C. Gildersleeve and L. K. Mahal, *Mol. BioSyst.*, 2008, 4, 654–662.
- C. Oliveira, J. A. Teixeira and L. Domingues, *Crit. Rev. Biotechnol.*, DOI: 10.3109/07388551.2012.670614, in press.
- I. J. Goldstein and C. E. Hayes, *Adv. Carbohydr. Chem. Biochem.*, 1978, 35, 127–340.
- N. Sharon and H. Lis, *Lectins*, 2nd edition, Kluwer Academic Publishers, Dordrecht, 2003.
- M. Ambrosi, N. R. Cameron and B. G. Davis, *Org. Biomol. Chem.*, 2005, 3, 1593–1608.
- H. J. Gabius, S. Andre, J. Jimenez-Barbero, A. Romero and D. Solis, *Trends Biochem. Sci.*, 2012, 36, 298–313.
- H. J. Gabius, H. C. Siebert, S. Andre, J. Jimenez-Barbero and H. Rudiger, *ChemBioChem*, 2004, 5, 740–764.
- K. Drickamer and M. E. Taylor, *Annu. Rev. Cell Biol.*, 1993, 9, 237–264.
- J. C. Paulson, M. S. Macauley and N. Kawasaki, *Ann. N. Y. Acad. Sci.*, 2012, 1253, 37–48.
- G. R. Vasta, *Adv. Exp. Med. Biol.*, 2012, 946, 21–36.
- G. R. Vasta, in *Glycoconjugates, Composition, Structure and Function*, ed. H. J. Allen and E. C. Kisailus, Marcel Dekker, New York, 1992.
- G. R. Vasta, H. Ahmed and E. W. Odom, *Curr. Opin. Struct. Biol.*, 2004, 14, 617–630.
- W. J. Peumans and E. J. Van Damme, *Plant Physiol.*, 1995, 109, 347–352.
- G. Vandenborre, G. Smagghe and E. J. Van Damme, *Phytochemistry*, 2012, 72, 1538–1550.
- A. Imberty and A. Varrot, *Curr. Opin. Struct. Biol.*, 2008, 18, 567–576.
- N. Sharon, *FEBS Lett.*, 1987, 217, 145–157.
- G. M. Edelman, B. A. Cunningham, G. N. Reeke, J. W. Becker, M. J. Waxdal and J. L. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 1972, 69, 2580–2584.
- R. Loris, T. Hamelryck, J. Bouckaert and L. Wyns, *Biochim. Biophys. Acta*, 1998, 1383, 9–36.
- N. Sharon and H. Lis, *FASEB J.*, 1990, 4, 3198–3208.
- A. Imberty, E. P. Mitchell and M. Wimmerová, *Curr. Opin. Struct. Biol.*, 2005, 15, 525–534.
- N. E. Ziolkowska, B. R. O'Keefe, T. Mori, C. Zhu, B. Giomarelli, F. Vojdani, K. E. Palmer, J. B. McMahon and A. Wlodawer, *Structure*, 2006, 14, 1127–1135.
- J. L. Asensio, A. Arda, F. J. Canada and J. Jimenez-Barbero, *Acc. Chem. Res.*, DOI: 10.1021/ar300024d, in press.
- K. Drickamer, *Curr. Opin. Struct. Biol.*, 1993, 3, 393–400.
- A. Imberty, M. Wimmerová, E. P. Mitchell and N. Gilboa-Garber, *Microbes Infect.*, 2004, 6, 222–229.
- K. B. Alexandre, E. S. Gray, H. Mufhandu, J. B. McMahon, E. Chakauya, B. R. O'Keefe, R. Chikwamba and L. Morris, *Virology*, 2012, 423, 175–186.
- J. Balzarini, *Antiviral Res.*, 2006, 71, 237–247.
- B. Liu, H. J. Bian and J. K. Bao, *Cancer Lett.*, 2010, 287, 1–12.
- T. Minko, *Adv. Drug Delivery Rev.*, 2004, 56, 491–509.
- G. Poiroux, M. Pitie, R. Culerrier, E. Lafont, B. Segui, E. J. Van Damme, W. J. Peumans, J. Bernadou, T. Levade, P. Rouge, A. Barre and H. Benoist, *PLoS One*, 2011, 6, e23315.
- R. D. Cummings and S. Kornfeld, *J. Biol. Chem.*, 1982, 257, 11235–11240.
- S. Fanayan, M. Hincapie and W. S. Hancock, *Electrophoresis*, 2012, 33, 1746–1754.
- P. Gemeiner, D. Mislovicova, J. Tkac, J. Svitel, V. Patoprsty, E. Hrabarova, G. Kogan and T. Kozar, *Biotechnol. Adv.*, 2009, 27, 1–15.
- J. W. Dennis, M. Granovsky and C. E. Warren, *Biochim. Biophys. Acta*, 1999, 1473, 21–34.
- J. Lescar, J. F. Sanchez, A. Audfray, J. L. Coll, C. Breton, E. P. Mitchell and A. Imberty, *Glycobiology*, 2007, 17, 1077–1083.
- J. F. Sanchez, J. Lescar, V. Chazalet, A. Audfray, J. Gagnon, R. Alvarez, C. Breton, A. Imberty and E. P. Mitchell, *J. Biol. Chem.*, 2006, 281, 20171–20180.
- M. V. Dwek, H. A. Ross, A. J. Streets, S. A. Brooks, E. Adam, A. Titcomb, J. V. Woodside, U. Schumacher and A. J. Leatham, *Int. J. Cancer*, 2001, 95, 79–85.
- G. Gupta, A. Surolia and S.-G. Sampathkumar, *OMICS*, 2010, 14, 419–436.
- J. Katrlík, J. Svitel, P. Gemeiner, T. Kozar and J. Tkac, *Med. Res. Rev.*, 2010, 30, 394–418.
- J. Etxebarria, J. Calvo, M. Martin-Lomas and N. C. Reichardt, *ACS Chem. Biol.*, 2012, 7, 1729–1737.
- S. Chen, T. Zheng, M. R. Shortreed, C. Alexander and L. M. Smith, *Anal. Chem.*, 2007, 79, 5698–5702.
- S. A. Fry, B. Afrough, H. J. Lomax-Browne, J. F. Timms, L. S. Velentzis and A. J. Leatham, *Glycobiology*, 2011, 21, 1060–1070.



- 50 L. Landemarre, P. Cancellieri and E. Duverger, *Glycoconjugate J.*, DOI: 10.1007/s10719-10012-19433-y, in press.
- 51 K. L. Hsu and L. K. Mahal, *Nat. Protoc.*, 2006, 1, 543–549.
- 52 A. P. Davis, *Org. Biomol. Chem.*, 2009, 7, 3629–3638.
- 53 A. P. Davis and R. S. Wareham, *Angew. Chem., Int. Ed.*, 1999, 38, 2978–2996.
- 54 S. Jin, Y. Cheng, S. Reid, M. Li and B. Wang, *Med. Res. Rev.*, 2010, 30, 171–257.
- 55 S. Kubik, *Angew. Chem., Int. Ed.*, 2009, 48, 1722–1725.
- 56 M. Mazik, *ChemBioChem*, 2008, 9, 1015–1017.
- 57 A. B. Foster and M. Stacey, *J. Chem. Soc.*, 1954, 1778–1781.
- 58 G. Springsteen and B. Wang, *Tetrahedron*, 2002, 58, 5291–5300.
- 59 Z. Wang, D. Zhang and D. Zhu, *J. Org. Chem.*, 2005, 70, 5729–5732.
- 60 W. Tan, D. Zhang and D. Zhu, *Bioorg. Med. Chem. Lett.*, 2007, 17, 2629–2633.
- 61 S. Akay, W. Yang, J. Wang, L. Lin and B. Wang, *Chem. Biol. Drug Des.*, 2007, 70, 279–289.
- 62 M. Bérubé, M. Dowlut and D. G. Hall, *J. Org. Chem.*, 2008, 73, 6471–6479.
- 63 M. Dowlut and D. G. Hall, *J. Am. Chem. Soc.*, 2006, 128, 4226–4227.
- 64 Y. Cheng, N. Ni, W. Yang and B. Wang, *Chem.–Eur. J.*, 2010, 16, 13528–13538.
- 65 J. C. Norrild and H. Eggert, *J. Am. Chem. Soc.*, 1995, 117, 1479–1484.
- 66 V. V. Karnati, X. Gao, S. Gao, W. Yang, W. Ni, S. Sankar and B. Wang, *Bioorg. Med. Chem. Lett.*, 2002, 12, 3373–3377.
- 67 M. D. Phillips, T. M. Fyles, N. P. Barwell and T. D. James, *Chem. Commun.*, 2009, 6557–6559.
- 68 W. Yang, H. Fan, X. Gao, S. Gao, V. V. R. Karnati, W. Ni, W. B. Hooks, J. Carson, B. Weston and B. Wang, *Chem. Biol.*, 2004, 11, 439–448.
- 69 S. Jin, C. Zhu, Y. Cheng, M. Li and B. Wang, *Bioorg. Med. Chem.*, 2010, 18, 1449–1455.
- 70 A. Sugasaki, K. Sugiyasu, M. Ikeda, M. Takeuchi and S. Shinkai, *J. Am. Chem. Soc.*, 2001, 123, 10239–10244.
- 71 M. Chen, Y. Lu, Q. Ma, L. Guo and Y.-Q. Feng, *Analyst*, 2009, 134, 2158–2164.
- 72 J. I. Jay, B. E. Lai, D. G. Myszkka, A. Mahalingam, K. Langheinrich, D. F. Katz and P. F. Kiser, *Mol. Pharmacol.*, 2010, 7, 116–129.
- 73 A. Mahalingam, A. R. Geonnotti, J. Balzarini and P. F. Kiser, *Mol. Pharmacol.*, 2011, 8, 2465–2475.
- 74 K. L. Bicker, J. Sun, J. J. Lavigne and P. R. Thompson, *ACS Comb. Sci.*, 2011, 13, 232–243.
- 75 P. J. Duggan and D. A. Offermann, *Tetrahedron*, 2009, 65, 109–114.
- 76 N. Y. Edwards, T. W. Sager, J. T. McDevitt and E. V. Anslyn, *J. Am. Chem. Soc.*, 2007, 129, 13575–13583.
- 77 Y. Zou, D. L. Broughton, K. L. Bicker, P. R. Thompson and J. J. Lavigne, *ChemBioChem*, 2007, 8, 2048–2051.
- 78 S. Burroughs and B. Wang, *ChemBioChem*, 2010, 11, 2245–2246.
- 79 A. Pal, M. Berube and D. G. Hall, *Angew. Chem., Int. Ed.*, 2010, 49, 1492–1495.
- 80 K. L. Bicker, J. Sun, M. Harrell, Y. Zhang, M. M. Pena, P. R. Thompson and J. J. Lavigne, *Chem. Sci.*, 2012, 3, 1147–1156.
- 81 W. I. Weis and K. Drickamer, *Annu. Rev. Biochem.*, 1996, 65, 441–473.
- 82 U. D. Palanisamy, D. J. Winzor and C. R. Lowe, *J. Chromatogr., B. Biomed. Sci. Appl.*, 2000, 746, 265–281.
- 83 J. Billing, H. Grundberg and U. J. Nilsson, *Supramol. Chem.*, 2002, 14, 367–372.
- 84 R. D. Hubbard, S. R. Horner and B. L. Miller, *J. Am. Chem. Soc.*, 2001, 123, 5810–5811.
- 85 S. Manju, P. R. Hari and K. Sreenivasan, *Biosens. Bioelectron.*, 2010, 26, 894–897.
- 86 Y. Yang, C. Yi, J. Luo, R. Liu, J. Liu, J. Jiang and X. Liu, *Biosens. Bioelectron.*, 2011, 26, 2607–2612.
- 87 D. Singabraya, L. Bultel, F. Sineriz, M. Mothéré, D. Lesur, J. Kovensky and D. Papy-Garcia, *Talanta*, 2012, 99, 833–839.
- 88 V. L. Talance, O. Massinon, R. Baati, A. Wagner and S. P. Vincent, *Chem. Commun.*, 2012, 48, 10684–10686.
- 89 A. P. Davis and R. S. Wareham, *Angew. Chem., Int. Ed.*, 1998, 37, 2270–2273.
- 90 E. Klein, M. P. Crump and A. P. Davis, *Angew. Chem., Int. Ed.*, 2004, 44, 298–302.
- 91 Y. Ferrand, E. Klein, N. P. Barwell, M. P. Crump, J. Jimenez-Barbero, C. Vicent, G. J. Boons, S. Ingale and A. P. Davis, *Angew. Chem., Int. Ed.*, 2009, 48, 1775–1779.
- 92 Y. Ferrand, M. P. Crump and A. P. Davis, *Science*, 2007, 318, 619–622.
- 93 B. Sookcharoenpinyo, E. Klein, Y. Ferrand, D. B. Walker, P. R. Brotherhood, C. Ke, M. P. Crump and A. P. Davis, *Angew. Chem., Int. Ed.*, 2012, 51, 4586–4590.
- 94 N. P. Barwell and A. P. Davis, *J. Org. Chem.*, 2011, 76, 6548–6557.
- 95 G. Joshi and A. P. Davis, *Org. Biomol. Chem.*, 2012, 10, 5760–5763.
- 96 C. Ke, H. Destecroix, M. P. Crump and A. P. Davis, *Nat. Chem.*, 2012, 4, 718–723.
- 97 M. Famulok, G. Mayer and M. Blind, *Acc. Chem. Res.*, 2000, 33, 591–599.
- 98 A. B. Iliuk, L. Hu and W. A. Tao, *Anal. Chem.*, 2011, 83, 4440–4452.
- 99 M. Citartan, S. C. Gopinath, J. Tominaga, S. C. Tan and T. H. Tang, *Biosens. Bioelectron.*, 2012, 34, 1–11.
- 100 W. Sun, L. Du and M. Li, *Curr. Pharm. Des.*, 2010, 16, 2269–2278.
- 101 J. Woodcock, D. Moazed, M. Cannon, J. Davies and H. F. Noller, *EMBO J.*, 1991, 10, 3099–3103.
- 102 Y. Wang, J. Killian, K. Hamasaki and R. R. Rando, *Biochemistry*, 1996, 35, 12338–12346.
- 103 Y. Wang and R. R. Rando, *Chem. Biol.*, 1995, 2, 281–290.
- 104 J. A. Cowan, T. Ohyama, D. Wang and K. Natarajan, *Nucleic Acids Res.*, 2000, 28, 2935–2942.
- 105 S. T. Wallace and R. Schroeder, *RNA*, 1998, 4, 112–123.
- 106 L. Jiang, A. Majumdar, W. Hu, T. J. Jaishree, W. Xu and D. J. Patel, *Structure*, 1999, 7, 817–827.
- 107 L. Jiang and D. J. Patel, *Nat. Struct. Biol.*, 1998, 5, 769–774.

- 108 V. Tereshko, E. Skripkin and D. J. Patel, *Chem. Biol.*, 2003, **10**, 175–187.
- 109 A. A. Rowe, E. A. Miller and K. W. Plaxco, *Anal. Chem.*, 2010, **82**, 7090–7095.
- 110 Kawakami, *Anal. Chim. Acta*, 1998, **365**, 95–100.
- 111 Q. Yang, I. J. Goldstein, H. Y. Mei and D. R. Engelke, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 5462–5467.
- 112 M. M. Masud, M. Kuwahara, H. Ozaki and H. Sawai, *Biorg. Med. Chem.*, 2004, **12**, 1111–1120.
- 113 S. Jeong, T. Eom, S. Kim, S. Lee and J. Yu, *Biochem. Biophys. Res. Commun.*, 2001, **281**, 237–243.
- 114 M. Li, N. Lin, Z. Huang, L. Du, C. Altier, H. Fang and B. Wang, *J. Am. Chem. Soc.*, 2008, **130**, 12636–12638.
- 115 B. J. Boese and R. R. Breaker, *Nucleic Acids Res.*, 2007, **35**, 6378–6388.
- 116 B. J. Boese, K. Corbino and R. R. Breaker, *Nucleosides, Nucleotides Nucleic Acids*, 2008, **27**, 949–966.
- 117 S. Y. Low, J. E. Hill and J. Peccia, *Biochem. Biophys. Res. Commun.*, 2009, **378**, 701–705.
- 118 E. Fukusaki, T. Kato, H. Maeda, N. Kawazoe, Y. Ito, A. Okazawa, S. Kajiyama and A. Kobayashi, *Biorg. Med. Chem. Lett.*, 2000, **10**, 423–425.
- 119 J. L. Ding, S. T. Gan and B. Ho, *J. Innate Immun.*, 2009, **1**, 46–58.
- 120 S. E. Kim, W. Su, M. Cho, Y. Lee and W. S. Choe, *Anal. Biochem.*, 2012, **424**, 12–20.
- 121 W. Su, M. Lin, H. Lee, M. Cho, W. S. Choe and Y. Lee, *Biosens. Bioelectron.*, 2012, **32**, 32–36.
- 122 J. S. Maeng, N. Kim, C. T. Kim, S. R. Han, Y. J. Lee, S. W. Lee, M. H. Lee and Y. J. Cho, *J. Nanosci. Nanotechnol.*, 2012, **12**, 5138–5142.
- 123 N. H. Andersen, B. Cao, A. Rodriguez-Romero and B. Arreguin, *Biochemistry*, 1993, **32**, 1407–1422.
- 124 P. Verheyden, J. Pletinckx, D. Maes, H. A. Pepermans, L. Wyns, R. Willem and J. C. Martins, *FEBS Lett.*, 1995, **370**, 245–249.
- 125 J. C. Koo, S. Y. Lee, H. J. Chun, Y. H. Cheong, J. S. Choi, S. Kawabata, M. Miyagi, S. Tsunasawa, K. S. Ha, D. W. Bae, C. D. Han, B. L. Lee and M. J. Cho, *Biochim. Biophys. Acta*, 1998, **1382**, 80–90.
- 126 R. H. Huang, Y. Xiang, G. Z. Tu, Y. Zhang and D. C. Wang, *Biochemistry*, 2004, **43**, 6005–6012.
- 127 N. Aboitiz, M. Vila-Perello, P. Groves, J. L. Asensio, D. Andreu, F. J. Canada and J. Jimenez-Barbero, *ChemBioChem*, 2004, **5**, 1245–1255.
- 128 S. Lu, S. Liang and X. Gu, *J. Protein Chem.*, 1999, **18**, 609–617.
- 129 H. C. Siebert, S. Y. Lu, R. Wechselberger, K. Born, T. Eckert, S. Liang, C. W. von der Lieth, J. Jimenez-Barbero, R. Schauer, J. F. Vliegthart, T. Lutteke, S. Andre, H. Kaltner, H. J. Gabius and T. Kozar, *Carbohydr. Res.*, 2009, **344**, 1515–1525.
- 130 A. M. Cole, W. Wang, A. J. Waring and R. I. Lehrer, *Curr. Protein Pept. Sci.*, 2004, **5**, 373–381.
- 131 W. Wang, A. M. Cole, T. Hong, A. J. Waring and R. I. Lehrer, *J. Immunol.*, 2003, **170**, 4708–4716.
- 132 R. I. Lehrer, G. Jung, P. Ruchala, S. Andre, H. J. Gabius and W. Lu, *J. Immunol.*, 2009, **183**, 480–490.
- 133 J. Li, H. Wu, J. Hong, X. Xu, H. Yang, B. Wu, Y. Wang, J. Zhu, R. Lai, X. Jiang, D. Lin, M. C. Prescott and H. H. Rees, *PLoS One*, 2008, **3**, e2381.
- 134 D. D. Roberts and I. J. Goldstein, *J. Biol. Chem.*, 1984, **259**, 909–914.
- 135 Y. Konami, K. Yamamoto, T. Osawa and T. Irimura, *J. Biochem.*, 1992, **112**, 366–375.
- 136 K. Yamamoto, Y. Konami and T. Osawa, *J. Chromatogr.*, 1992, **597**, 221–230.
- 137 K. Yamamoto, Y. Konami, T. Osawa and T. Irimura, *J. Biochem.*, 1992, **111**, 436–439.
- 138 A. Moise, S. Andre, F. Eggers, M. Krzeminski, M. Przybylski and H. J. Gabius, *J. Am. Chem. Soc.*, 2011, **133**, 14844–14847.
- 139 L. D. Heerze, P. C. Chong and G. D. Armstrong, *J. Biol. Chem.*, 1992, **267**, 25810–25815.
- 140 S. H. Tam, M. T. Nakada, M. Kruszynski, W. E. Fieles, A. H. Taylor, M. Mervic and G. A. Heavner, *Biochem. Biophys. Res. Commun.*, 1996, **227**, 712–717.
- 141 C. E. Von Seggern and R. J. Cotter, *J. Mass Spectrom.*, 2004, **39**, 736–742.
- 142 W. Tong, M. Gagnon, T. Sprules, M. Gilbert, S. Chowdhury, K. Meerovitch, K. Hansford, E. O. Purisima, J. W. Blankenship, N. K. Cheung, K. Gehring, W. D. Lubell and H. U. Saragovi, *Chem. Biol.*, 2010, **17**, 183–194.
- 143 E. N. Peletskaya, V. V. Glinsky, G. V. Glinsky, S. L. Deutscher and T. P. Quinn, *J. Mol. Biol.*, 1997, **270**, 374–384.
- 144 T. Matsubara, M. Iida, T. Tsumuraya, I. Fujii and T. Sato, *Biochemistry*, 2008, **47**, 6745–6751.
- 145 T. Matsubara, D. Ishikawa, T. Taki, Y. Okahata and T. Sato, *FEBS Lett.*, 1999, **456**, 253–256.
- 146 T. Matsubara, M. Sumi, H. Kubota, T. Taki, Y. Okahata and T. Sato, *J. Med. Chem.*, 2009, **52**, 4247–4256.
- 147 T. Matsubara, A. Onishi and T. Sato, *Biorg. Med. Chem.*, 2012, **20**, 6452–6458.
- 148 M. Kwon, S. Jeong, K. H. Lee, Y. K. Park and J. Yu, *J. Am. Chem. Soc.*, 2002, **124**, 13996–13997.
- 149 S. Hyun, J. Kim, M. Kwon and J. Yu, *Biorg. Med. Chem.*, 2007, **15**, 511–517.
- 150 S. Hyun, E. H. Lee, J. Park and J. Yu, *Biorg. Med. Chem. Lett.*, 2008, **18**, 4011–4014.
- 151 K. W. Boltz, M. J. Gonzalez-Moa, P. Stafford, S. A. Johnston and S. A. Svarovsky, *Analyst*, 2009, **134**, 650–652.
- 152 C. Morales Betanzos, M. J. Gonzalez-Moa, K. W. Boltz, B. D. Vander Werf, S. A. Johnston and S. A. Svarovsky, *ChemBioChem*, 2009, **10**, 877–888.
- 153 S. M. Parsons and M. A. Raftery, *Biochemistry*, 1969, **8**, 4199–4205.
- 154 T. Mega and S. Hase, *Biochim. Biophys. Acta*, 1994, **1200**, 331–333.
- 155 R. A. Laine, J. W.-C. Lo and B. C.-R. Zhu, in *Lectins – Analytical Technologies*, ed. C. Nilsson, Elsevier, Amsterdam, 2007, pp. 373–384.
- 156 M. Lienemann, H. Boer, A. Paananen, S. Cottaz and A. Koivula, *Glycobiology*, 2009, **19**, 1116–1126.



- 157 J. Aalto, S. Pelkonen, H. Kalimo and J. Finne, *Glycoconjugate J.*, 2001, **18**, 751–758.
- 158 A. Jokilampi, M. Korja, E. Jakobsson and J. Finne, in *Lectins – Analytical Technologies*, ed. C. Nilsson, Elsevier, Amsterdam, 2007, pp. 373–384.
- 159 R. Arango, E. Rodriguez-Arango, R. Adar, D. Belenky, F. G. Loontjens, S. Rozenblatt and N. Sharon, *FEBS Lett.*, 1993, **330**, 133–136.
- 160 E. T. Jordan and I. J. Goldstein, *Eur. J. Biochem.*, 1995, **230**, 958–964.
- 161 M. Nishiguchi, K. Yoshida, T. Sumizono and K. Tazaki, *FEBS Lett.*, 1997, **403**, 294–298.
- 162 E. Mitchell, C. Houles, D. Sudakevitz, M. Wimmerova, C. Gautier, S. Pérez, A. M. Wu, N. Gilboa-Garber and A. Imberty, *Nat. Struct. Biol.*, 2002, **9**, 918–921.
- 163 N. Kostlanová, E. P. Mitchell, H. Lortat-Jacob, S. Oscarson, M. Lahmann, N. Gilboa-Garber, G. Chambat, M. Wimmerová and A. Imberty, *J. Biol. Chem.*, 2005, **280**, 27839–27849.
- 164 J. Adam, M. Pokorna, C. Sabin, E. P. Mitchell, A. Imberty and M. Wimmerova, *BMC Struct. Biol.*, 2007, **7**, 36.
- 165 K. Yamamoto, Y. Konami, T. Osawa and T. Irimura, *J. Biochem.*, 1992, **111**, 87–90.
- 166 K. Imamura, H. Takeuchi, R. Yabe, H. Tateno and J. Hirabayashi, *J. Biochem.*, 2011, **150**, 545–552.
- 167 R. Yabe, R. Suzuki, A. Kuno, Z. Fujimoto, Y. Jigami and J. Hirabayashi, *J. Biochem.*, 2007, **141**, 389–399.
- 168 D. Hu, H. Tateno, A. Kuno, R. Yabe and J. Hirabayashi, *J. Biol. Chem.*, 2012, **287**, 20313–20320.
- 169 M. Yim, T. Ono and T. Irimura, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 2222–2225.
- 170 K. Maenuma, M. Yim, K. Komatsu, M. Hoshino, Y. Takahashi, N. Bovin and T. Irimura, *Proteomics*, 2008, **8**, 3274–3283.
- 171 K. Maenuma, M. Yim, K. Komatsu, M. Hoshino, A. Tachiki-Fujioka, K. Takahashi, Y. Hiki, N. Bovin and T. Irimura, *J. Proteome Res.*, 2009, **8**, 3617–3624.
- 172 G. R. Gunther, J. L. Wang, I. Yahara, B. A. Cunningham and G. M. Edelman, *Proc. Natl. Acad. Sci. U. S. A.*, 1973, **70**, 1012–1016.
- 173 M. Beppu, T. Terao and T. Osawa, *J. Biochem.*, 1976, **79**, 1113–1117.
- 174 A. R. Fraser, J. J. Hemperly, J. L. Wang and G. M. Edelman, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 790–794.
- 175 H. Kaku, Y. Mori, I. J. Goldstein and N. Shibuya, *J. Biol. Chem.*, 1993, **268**, 13237–13241.
- 176 H. Kaku and N. Shibuya, *FEBS Lett.*, 1992, **306**, 176–180.
- 177 J. Olausson, E. Astrom, B. H. Jonsson, L. A. Tibell and P. Pahlsson, *Glycobiology*, 2011, **21**, 34–44.
- 178 R. Yabe, Y. Itakura, S. Nakamura-Tsuruta, J. Iwaki, A. Kuno and J. Hirabayashi, *Biochem. Biophys. Res. Commun.*, 2009, **384**, 204–209.
- 179 R. E. McDonald, D. J. Hughes and B. G. Davis, *Angew. Chem., Int. Ed.*, 2004, **43**, 3025–3029.
- 180 V. Fulop and D. T. Jones, *Curr. Opin. Struct. Biol.*, 1999, **9**, 715–721.
- 181 G. D'Alessio, *Prog. Biophys. Mol. Biol.*, 1999, **72**, 271–298.
- 182 H. G. Beisel, S. Kawabata, S. Iwanaga, R. Huber and W. Bode, *EMBO J.*, 1999, **18**, 2313–2322.
- 183 I. Yadid and D. S. Tawfik, *J. Mol. Biol.*, 2007, **365**, 10–17.
- 184 I. Yadid, N. Kirshenbaum, M. Sharon, O. Dym and D. S. Tawfik, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 7287–7292.
- 185 I. Yadid and D. S. Tawfik, *Protein Eng., Des. Sel.*, 2011, **24**, 185–195.
- 186 M. Wimmerova, E. Mitchell, J. F. Sanchez, C. Gautier and A. Imberty, *J. Biol. Chem.*, 2003, **278**, 27059–27067.
- 187 P. R. Romano, A. Mackay, M. Vong, J. Desa, A. Lamontagne, M. A. Comunale, J. Hafner, T. Block, R. Lec and A. Mehta, *Biochem. Biophys. Res. Commun.*, 2011, **414**, 84–89.
- 188 R. Q. Benevides, G. Ganne, R. Simoes, M. Niemietz, C. Unverzagt, C. Breton, A. Varrot, B. S. Cavada and A. Imberty, *J. Biol. Chem.*, 2012, **287**, 26352–26364.
- 189 N. E. Ziolkowska, S. R. Shenoy, B. R. O'Keefe, J. B. McMahon, K. E. Palmer, R. A. Dwek, M. R. Wormald and A. Wlodawer, *Proteins*, 2007, **67**, 661–670.
- 190 Y. Liu, Z. Misulovin and P. J. Bjorkman, *J. Mol. Biol.*, 2001, **305**, 481–490.
- 191 J.-P. Gourdine, G. Cioci, L. Miguët, C. Unverzagt, D. Varón-Silva, A. Varrot, C. Gautier, E. J. Smith-Ravin and A. Imberty, *J. Biol. Chem.*, 2008, **283**, 30112–30120.
- 192 G. Cioci, E. P. Mitchell, V. Chazalet, H. Debray, S. Oscarson, M. Lahmann, C. Gautier, C. Breton, S. Pérez and A. Imberty, *J. Mol. Biol.*, 2006, **357**, 1575–1591.
- 193 A. Imberty, C. Gautier, J. Lescar, S. Pérez, L. Wyns and R. Loris, *J. Biol. Chem.*, 2000, **275**, 17541–17548.
- 194 S. Perret, C. Sabin, C. Dumon, M. Pokorná, C. Gautier, O. Galanina, S. Ilia, N. Bovin, M. Nicaise, M. Desmadril, N. Gilboa-Garber, M. Wimmerova, E. P. Mitchell and A. Imberty, *Biochem. J.*, 2005, **389**, 325–332.
- 195 C. S. Ferreira, M. C. Cheung, S. Missailidis, S. Bisland and J. Gariepy, *Nucleic Acids Res.*, 2009, **37**, 866–876.
- 196 C. Srisawat, I. J. Goldstein and D. R. Engelke, *Nucleic Acids Res.*, 2001, **29**, E4.
- 197 A. Verrecchio, M. W. Germann, B. P. Schick, B. Kung, T. Twardowski and J. D. San Antonio, *J. Biol. Chem.*, 2000, **275**, 7701–7707.

## **II.5. Utilisation des lectines en biotechnologie et médecine**

Compte tenu de la présence de glycanes sur toutes les surfaces des cellules, il n'est pas surprenant que plusieurs maladies humaines impliquent des changements de glycosylation et / ou la reconnaissance des glycanes. Un domaine important de la recherche en glycobiologie porte sur l'étude des « Congenital Disorder of Glycosylation » (CGD) (Freeze, 2006). Ici nous ne traiterons que des maladies acquises et non héréditaires.

### **II.5.1. Changement de glycosylation et maladies chroniques**

De nombreuses pathologies sont corrélées à un changement de la glycosylation. On peut citer par exemple le diabète, la maladie d'Alzheimer, l'inflammation comme l'asthme...

Plusieurs mécanismes de modification de la glycosylation différents ont été caractérisés dans le diabète. Dans le diabète sucré, une augmentation de la production de l'UDP-GlcNAc est causée par l'excès de la conversion de glucose. Une hypothèse actuelle est que cette augmentation de l'UDP-GlcNAc cytoplasmique a pour conséquence une augmentation secondaire du niveau de *O*-GlcNAc sur les glycoprotéines nucléaires et cytoplasmiques. Les conséquences sont donc une modification du fonctionnement de ces protéines.

La néphropathie diabétique est une complication associée avec une mortalité élevée. Il s'agit d'une diminution du taux de filtration glomérulaire conduisant à l'insuffisance rénale. Cette pathologie a été corrélée à une diminution de la teneur en héparane sulfate (HS) des membranes basales glomérulaires. Ceci peut être causé par le taux élevé de glucose dans le milieu. (Raats *et al.*, 2000).

Les maladies inflammatoires sont accompagnées de nombreux changements physiologiques et biochimiques au niveau du site de l'inflammation ou sur les protéines circulantes. Dans les deux dernières décennies, une grande attention a été accordée aux changements de glycosylation des principales protéines sériques (alpha1-glycoprotéine acide, immunoglobuline G, immunoglobuline A, transferrine, haptoglobine, alpha2-macroglobuline,



la protéine C-réactive...). Ces marqueurs glycosidiques représentent un intérêt en diagnostic (Gornik and Lauc, 2008). L'épitope SLe<sup>x</sup> est par exemple surexprimé sur les cellules endothéliales vasculaires dans le cas d'inflammation de la peau (Akahori *et al.*, 1997). Ce même épitope est également une signature d'inflammation sur les cellules sanguines telles que les leucocytes polynucléaires, les monocytes et les cellules tueuses naturelles (NK) (Munro *et al.*, 1992). Plus récemment il a été montré la présence d'une accumulation de liposomes enrichi en SLe<sup>x</sup> à proximité des régions d'inflammation ou tumorales (Hirai *et al.*, 2007). Le SLe<sup>x</sup> est donc un très bon marqueur de l'inflammation (Dube and Bertozzi, 2005).

## II.5.2. Glycosylation et cancer

Les changements de glycosylation des cellules tumorales peuvent prendre différentes formes. On peut citer, par exemple, une perte d'expression ou une surexpression de certaines structures, la persistance de structures incomplètes ou tronquées, l'accumulation de précurseurs, et, moins fréquemment, l'apparition de nouvelles structures. Cependant, les modifications de glycosylation ne sont pas simplement la conséquence du désordre présent dans les cellules tumorales. Certaines modifications sont corrélées à la transformation maligne et la progression tumorale.

### II.5.2.1. N-glycosylation et cancer

L'augmentation de la taille des N-glycannes est une des caractéristiques des cellules tumorales. Elle est expliquée par une augmentation de ramification  $\beta$ 1-6 de N-glycannes et donc du nombre d'antennes. Les glycosyltransférases GnT-III et GnT-V sont des enzymes de premier plan dans la progression cancéreuse et la formation des métastases. L'activité de l'enzyme GnT-V et donc la présence, de manière abondante, de résidus  $\beta$ 1-6GlcNAc à la surface des glycoprotéines, est directement reliée au caractère fortement métastatique de nombreuses lignées cellulaires animales et humaines (fibroblastes, carcinomes mammaires murins, lymphomes, cellules cancéreuses humaines du côlon) (Dennis *et al.*, 1987). A l'inverse, l'enzyme GnT-III semble inhiber la formation des métastases par compétition pour le substrat avec GnT-V. L'ajout d'un résidu GlcNAc en  $\beta$ 1-4 sur le  $\beta$ -mannose du cœur trimannosylé par GnT-III entraîne des modifications conformationnelles qui rendent le sucre néo-formé inaccessible comme substrat pour GnT-V (Schachter, 1986). La quantité de

structures  $\beta$ 1-6GlcNAc est alors fortement diminuée et le nombre de métastases réduit (Yoshimura *et al.*, 1995).

### II.5.2.2. O-glycosylation et cancer

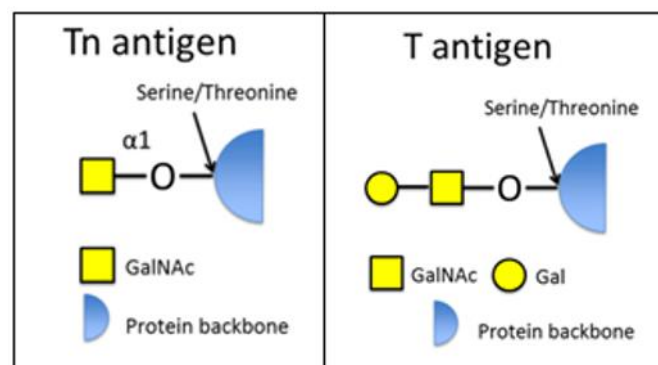
La surexpression de mucines dans les carcinomes a été décrite depuis de nombreuses années. La plupart des polypeptides de mucines épithéliales appartiennent à la famille MUC. Les mucines sécrétées apparaissent souvent dans le sang des patients atteints de cancer et peuvent être détectées par des anticorps monoclonaux. Dans les cancers d'origine épithéliale (carcinomes), en particulier, les mucines semblent être les principaux porteurs de glycosylations modifiées.

Dans la majorité des cas, l'apparition de ces structures est associée à la progression du cancer, à l'augmentation de son agressivité et à la formation de métastases (Kim *et al.*, 1996; Springer, 1997). Les nouvelles mucines pourraient également bloquer physiquement les interactions entre les cellules des carcinomes et les cellules de l'immunité comme les cellules NK.

Une autre caractéristique des mucines des carcinomes est leur glycosylation incomplète. Une altération dans l'expression de plusieurs enzymes de biosynthèse des O-glycanes peut être à l'origine de ces modifications. Les pp-GalNAc-Ts sont regroupés à l'intérieur d'une famille de plus de vingt membres dont quinze ont déjà été caractérisés dans des cellules de mammifères. Deux enzymes appartenant à cette famille, les ppGalNAc-T3 et -T6 semblent être particulièrement perturbées durant le processus tumoral. Deux études différentes ont démontré qu'elles semblaient surexprimées dans des lignées de cancers du sein alors que leur expression était quasiment silencieuse dans des cellules saines. Ce phénomène pourrait être à l'origine de l'augmentation de l'antigène Tn à la surface des cellules cancéreuses, notamment sur des sites potentiels de O-glycosylation normalement inoccupés (Berois *et al.*, 2006; Brooks *et al.*, 2007). Tn est un antigène glyco-peptidique ( $\alpha$ GalNAc-O-Ser/Thr) qui est exprimé dans les carcinomes du fait d'une dérégulation des processus de glycosylation. L'intérêt majeur de l'antigène Tn est de discriminer les cellules tumorales des cellules normales avec une grande sélectivité. Tn est fortement exprimé dans 90% des cancers ovariens, il est donc une cible de choix pour l'immunothérapie anti-cancéreuse par anticorps.

Les antigènes T et Tn (Figure 24) peuvent également être sialylés pour former respectivement le sialyl-T (ST) et sialyl-Tn (STn) qui sont fortement exprimés à la surface de nombreux cancers (sein, ovaire, côlon, estomac...). Ils sont généralement associés à un faible

pronostic vital, accompagnés de métastases et d'une diminution de la survie des patients (Itzkowitz *et al.*, 1990) (Leivonen *et al.*, 2001). L'antigène Tn sialylé est responsable d'une mobilité cellulaire significativement améliorée, d'une adhésion cellulaire diminuée et d'une tumorigénicité augmentée, caractéristiques révélant le caractère agressif de la lignée (Julien *et al.*, 2006). La sialylation de l'antigène Tn est réalisée en  $\alpha$ 2-6 du résidu GalNAc par l'enzyme ST6GalNAc I. Dans les cancers du sein et de l'estomac, une très bonne corrélation est observée entre la quantité d'antigène STn exprimée à la surface des cellules et l'abondance d'ARNm codant pour l'enzyme ST6GalNAc I (Marcos *et al.*, 2003; Sewell *et al.*, 2006) alors que ces résultats ne sont pas vérifiés pour les cancers du côlon où le niveau d'enzyme ST6GalNAc exprimé est comparable dans les lignées exprimant ou pas l'antigène STn (Brockhausen *et al.*, 1998).



**Figure 24 : Les antigènes T et Tn.**

*D'après sbhsciences.com.*

### II.5.2.3. Glycolipides et cancer

L'expression des glycosphingolipides (gangliosides, globosides et lactosides) est également fortement perturbée durant la transformation tumorale. De nouvelles structures, normalement uniquement exprimées dans les stades embryonnaires précoces, apparaissent. Plusieurs anticorps monoclonaux spécifiques des cellules cancéreuses sont réactifs avec la partie glycane des glycosphingolipides. Les principaux exemples sont les gangliosides GD2 et GD3 dans les mélanomes et cancers du cerveau, le Gb3 dans les lymphomes de Burkitt, ou dans les cancers ovariens (Hakomori, 2000). Plusieurs types de tumeurs (en particulier celles

d'origine neuro-ectodermique, tels que le mélanome et le neuroblastome) sont caractérisés par une surexpression de glycosphingolipides sialylés. Certains d'entre eux (par exemple, GD2) ne sont pas normalement présents à des niveaux élevés dans les gangliosides extraneuronaux et ils sont donc considérés comme des cibles pour l'immunothérapie (Hakomori and Kannagi, 1983).

A l'inverse, d'autres structures peuvent disparaître. Une perte complète de l'expression des glycosylphosphatidylinositols (GPI) est considérée dans certains cas comme indicateur d'un état précancéreux impliquant le système hématopoïétique. Ceci résulte de mutations somatiques acquises de cellules souches hématopoïétiques. La conséquence est une perte marquée ou complète de l'expression de surface cellulaire de plusieurs protéines à ancre GPI.

Il est cependant très difficile de savoir si l'apparition de ces antigènes lipidiques est causée par une accumulation des précurseurs due à une synthèse incomplète ou au développement de nouvelles voies de biosynthèse provoqué par l'activation de nouveaux gènes de glycosyltransférases (Hakomori, 2008). De manière générale, ces modifications influencent le taux de croissance tumorale et jouent un rôle clé dans l'induction de l'invasion et la formation de métastases, selon des mécanismes encore très peu connus. L'expression de ces antigènes lipidiques tumoraux dans le microenvironnement de la tumeur aide les cellules tumorales à échapper au système immunitaire de l'hôte (Birkle *et al.*, 2003).

### **II.5.3. Antigènes Lewis, fucosylation et Cancer**

#### **II.5.3.1. Généralités**

Les changements de fucosylation ont été caractérisés dans les études de glycosylation des cellules tumorales dès 1979 (Baumann *et al.*, 1979). Dans cet article, les auteurs ont comparé le niveau de fucosylation des glycolipides dans les cellules d'hépatome et d'hépatocytes normaux. Avant même que les mécanismes de fucosylation soient compris, des protéines fucosylées ont été identifiées et utilisées comme marqueurs tumoraux. L'augmentation des alphafoetoprotéines fucosylées (AFP) a été rapportée par les Breborowicz et Taketa (Simm *et al.*, 1979; Watanabe *et al.*, 1975). Des résultats récents concernant la régulation de la voie de synthèse GDP-fucose montrent aussi des augmentations de la fucosylation des glycolipides dans les cellules d'hépatome (Noda *et al.*, 2003).

### II.5.3.2. $\alpha$ 1-6 fucosylation

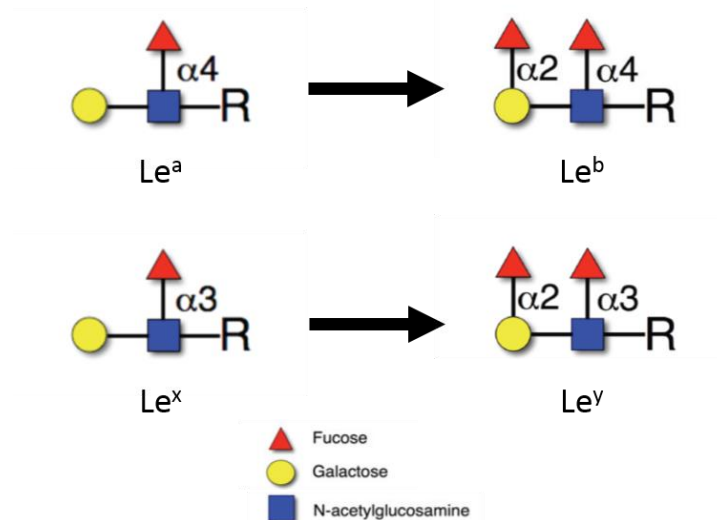
La fucosylation  $\alpha$ 1-6 du corps des N-glycanes peut être un marqueur de cancer humain (Varki et al., 2009). Des augmentations des taux d'haptoglobines fucosylées ont été observées dans plusieurs types de cancer (Turner, 1995). Cette glycoprotéine semble être en particulier un marqueur biologique d'intérêt dans le cancer de la prostate (Fujita *et al.*, 2014), des ovaires (Garibay-Cerdenares *et al.*, 2014), du foie (Asazawa *et al.*, 2014) et du pancréas (Okuyama *et al.*, 2006). A l'inverse une diminution de la quantité de noyau-fucosylation est observée dans le processus de tumorigénèse du cancer gastrique (Zhao et al., 2014).

L'alphafetoprotéine est un marqueur tumoral connu des carcinomes hépatocellulaires, mais il est parfois aussi surexprimé dans les maladies du foie telles que l'hépatite chronique et la cirrhose du foie. En revanche, l'AFP avec un core-fucosylation  $\alpha$ 1-6 est un marqueur très spécifique pour le carcinome hépatocellulaire malin (Aoyagi et al., 1991; Taketa et al., 1993). Une augmentation de l'activité de FUT8 (N-acétyl- $\beta$ -glucosaminidase  $\alpha$ 1-6 fucosyltransférase) et, par conséquent, une surexpression de résidus fucosidés sur l'unique site de N-glycosylation (Asn 232) de l'AFP, au niveau des tissus et dans le sérum, corrobore avec un mauvais diagnostic pour des patients atteints de carcinomes hépatocellulaires et d'autres pathologies chroniques du foie (Khien *et al.*, 2001). De nombreuses études ont confirmé une augmentation du niveau d'expression de FUT8 dans différentes lignées cellulaires cancéreuses, notamment celles du poumon et du côlon, (Miyoshi *et al.*, 1997; Muínelo-Romay *et al.*, 2008), mais aussi des ovaires (Lukas et al., 2013) ou de la thyroïde, y compris suite à l'ablation de la tumeur par chirurgie (Casals, 2001). Parallèlement à GnT-III, FUT8 possède également une action sur les molécules d'adhésion en augmentant, de manière significative, l'expression des E-cadhérines, notamment dans les cancers colorectaux (Osumi *et al.*, 2009). De plus, l' $\alpha$ 1-6 fucosylation régulerait le turnover des E-cadhérines, favorisant ainsi l'adhésion cellule-cellule *in vivo* (Takahashi *et al.*, 2009).

### II.5.3.3. $\alpha$ 1-2 fucosylation

La  $\alpha$ 1-2 fucosyltransférase présente une activité plus élevée dans les tissus tumoraux par rapport aux tissus sains. Ainsi, une activité aberrante de l'enzyme a été décrite dans les tumeurs du côlon (Muínelo-Romay et al., 2010) permettant la conversion des antigènes Le<sup>a</sup> et

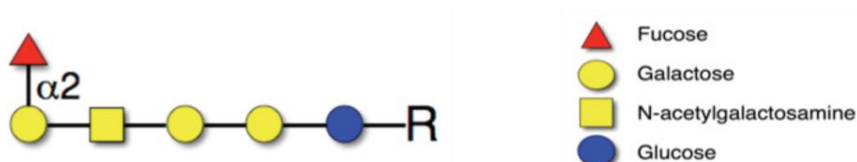
Le<sup>x</sup> en Le<sup>b</sup> et Le<sup>y</sup> respectivement (Figure 25), et donc une surexpression de ces antigènes (Yazawa et al., 1993).



**Figure 25 : Glycosylations des suites d'une activité aberrante de la  $\alpha 1-2$  fucosyltransférase**

L'élévation de la quantité de  $\alpha 1-2$  fucosyltransférase a aussi pour conséquence l'augmentation de l'expression de l'antigène H type 2 dans le cancer colorectal. Cette surexpression est connue pour être impliquée dans l'augmentation de la tumorigénicité et de la résistance à l'apoptose (Hallouin et al., 1999; Labarriere et al., 1994)

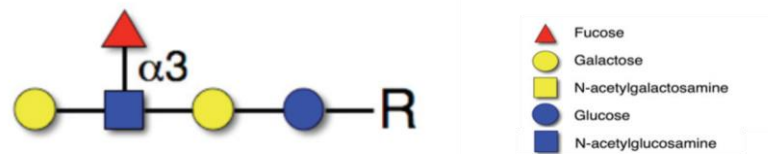
Le GloboH est un autre oligosaccharide  $\alpha 1-2$  fucosylé identifié comme marqueur des cellules souches mais aussi des cancers (Figure 26). Il est ainsi retrouvé sur plusieurs cellules tumorales, comme celles du cancer du sein. Une étude de phase I a été réalisée afin d'immuniser des patientes atteintes d'un cancer du sein métastatique avec du Globo H synthétique (Hanasaki et al., 1995).



**Figure 26 : Représentation schématique du Globo H**

### II.5.3.4. $\alpha$ 1-3/4 fucosylation

Les fucosylations en position 3 et 4 d'un GlcNAc sont présentes dans de nombreux antigènes glycosidiques comme le Le<sup>x</sup>, Le<sup>y</sup>, Le<sup>a</sup> et le LNFP III (Lacto-N-fucopentaose III) (Figure 27). Nous avons déjà abordé le Le<sup>y</sup> dans le chapitre précédent. La présence des antigènes Lewis sialylés, SLe<sup>a</sup> et le SLe<sup>x</sup>, sur les tumeurs est connue depuis longtemps (Chia et al., 1985). L'équipe de Kannagi a montré que le SLe<sup>x</sup> et le SLe<sup>a</sup> sont les molécules d'adhérence de plusieurs types de cellules transformées telles que les carcinomes du côlon, du poumon, du pancréas, du foie et de l'estomac ou les leucémies (Takada et al., 1993). Des études récentes confirment que la surexpression des antigènes Lewis est retrouvée dans de nombreux cancers avec par exemple une surexpression des SLe<sup>x</sup> dans les carcinomes de rein (Borzym-Kluczyk *et al.*, 2014), de Le<sup>y</sup> dans le cancer ovarien (Zhuang *et al.*, 2014), de SLe<sup>x</sup> dans le cancer du tractus gastro-intestinal (Dube and Bertozzi, 2005; Groux-Degroote *et al.*, 2014; Zhang *et al.*, 1997). Ces antigènes sont également probablement impliqués dans la signalisation cellulaire qui permet l'angiogenèse (Kannagi *et al.*, 2004).



**Figure 27 : Représentation schématique du LNFP III**

La présence des antigènes SLe<sup>a</sup> et SLe<sup>x</sup> est souvent associée à la formation de métastases par un mécanisme qui rappelle celui de la migration et de la diapédèse leucocytaire durant l'inflammation (Kannagi, 2004). En effet, les cellules tumorales ont la capacité d'induire l'expression des sélectines E par les cellules endothéliales. L'interaction sélectine-SLe<sup>x/a</sup> qui s'ensuit permet à la cellule cancéreuse de quitter la tumeur primaire, de gagner la circulation sanguine et enfin de coloniser les tissus sains (Kannagi, 2004). C'est ainsi que ces antigènes glucidiques sont utilisés comme marqueurs de diagnostic de cancer, comme le cancer mammaire canin (Pinho *et al.*, 2007), mais aussi comme marqueurs de mauvais pronostic pour le cancer de la vessie chez l'homme (Numahata *et al.*, 2002). Une étude (Mathieu *et al.*, 2004) laisse entrevoir des pistes intéressantes quant à l'utilisation des fucosyltransférases dans la thérapie génique du cancer. Ce travail montre en effet que la surexpression du gène FUT1 dans des cellules d'hépatocarcinomes inhibe la formation des

capillaires et la croissance tumorale, ces observations étant expliquées par une forte augmentation de l'expression des antigènes et Le<sup>y</sup> aux dépens des antigènes SLe<sup>a</sup> et SLe<sup>x</sup> (Tableau 4).

Glycanes	Tissus malins								
	Ovaire	Pancréas	Sang	Sein	Colon	Cerveau	Prostate	Peau	Poumon
SLe <sup>x</sup>		X		X	X				X
SLe <sup>a</sup>		X		X	X				X
Le <sup>x</sup>	X	X		X	X		X		X
Globo H	X	X		X	X		X		X
LNFP III									X

**Tableau 4 : Principaux glycoconjugués fucosylés dont la surexpression est retrouvée dans des cancers**

La fucosylation est fortement associée au cancer et elle pourrait être une cible pour un nouveau traitement du cancer (Miyoshi *et al.*, 2008).

## II.6. Utilisation des lectines en recherche, diagnostic et thérapie

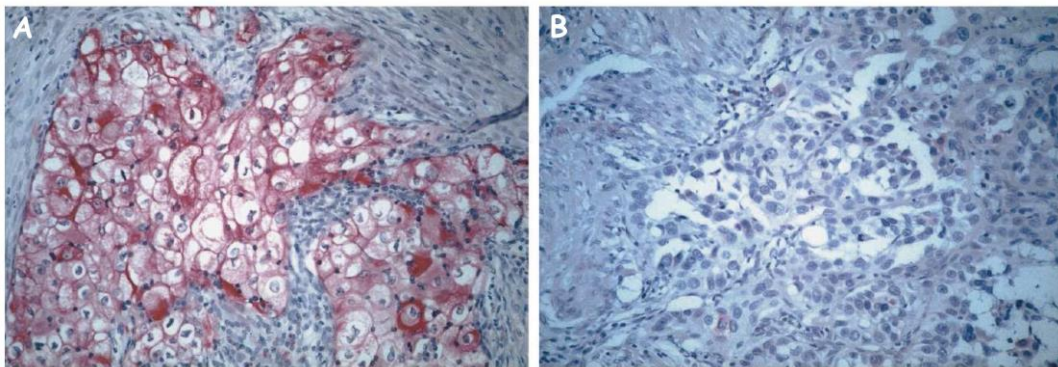
### II.6.1. Histopathologie

Les lectines peuvent être utilisées en histopathologie afin de marquer spécifiquement des cellules présentant des glycosylations particulières. Les principales lectines isolées de plantes ou d'invertébrés sont présentées dans le Tableau 5. Elles sont employées pour la mise en évidence de glycotopes en oncologie ou, plus généralement, en histopathologie (Figure 28). (Guillot *et al.*, 2004; Kuhlmann *et al.*, 1983; Maruyama *et al.*, 2002). De par leur



spécificité, les lectines se lient préférentiellement à un monosaccharide soit situé au bout de la chaîne soit placé dans une position non terminale des oligosaccharides complexes.

Depuis maintenant plus de vingt ans, la lectine HPA, extraite de l'escargot de Bourgogne est utilisée en histopathologie comme marqueur de plusieurs cancers, dont les adénocarcinomes du sein (Brooks *et al.*, 2007), de l'estomac (Cioci *et al.*, 2003), du côlon (Kolatkar *et al.*, 1998), du poumon (Kolatkar *et al.*, 1998) mais aussi pour les mélanomes (Cioci *et al.*, 2006). Cette lectine a la particularité de reconnaître spécifiquement des glycoépitopes présents à la surface des cancers primaires, associés au développement de métastases notamment dans les ganglions lymphatiques impliquant à court terme un mauvais pronostic vital et un raccourcissement de la vie des patients. Elle est également capable de se fixer sur des épitopes présents sur les cellules ayant déjà métastasées.



**Figure 28 : Marquages positif (A) et négatif (B) par HPA biotinylé d'adénocarcinomes du poumon.**

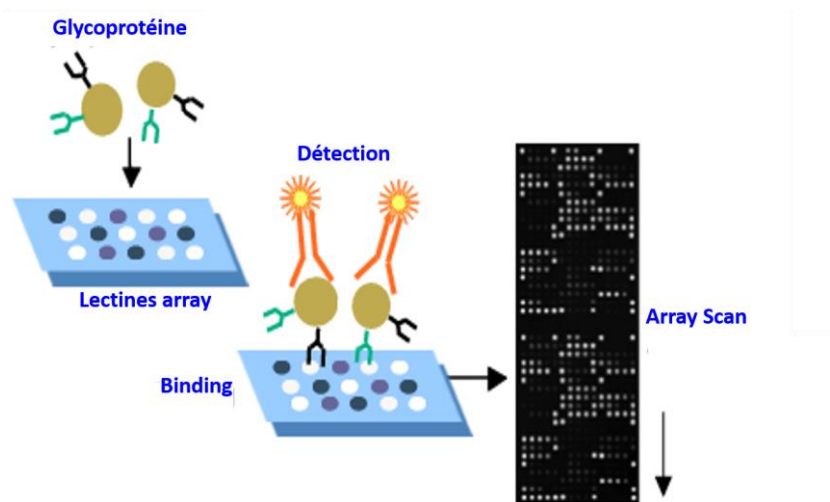
*Grossissement x200(Kolatkar et al., 1998).*

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

**Tableau 5 : Principales lectines commerciales**

## II.6.2. Puces à lectines

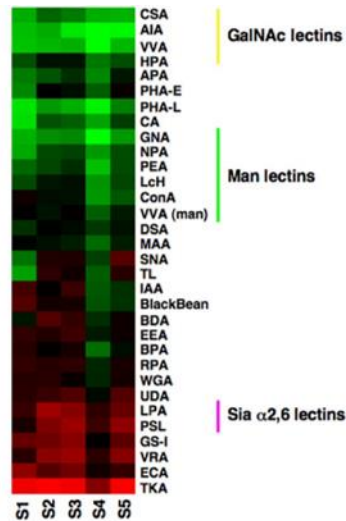
Une des principales techniques qui permet l'étude des glycanes est la spectrométrie de masse. Elle a l'avantage de donner des informations détaillées sur la structure des glycanes. Mais son plus gros inconvénient est le faible débit de cette technique pour étudier le glycome. Ainsi la technique de puce à lectines et à anticorps dirigés contre les sucres a été mise au point. Il s'agit d'une méthode à haut débit mais qui donne beaucoup moins de détails structurels. Les lectines possèdent une gamme de spécificité : certaines reconnaissent un monosaccharide particulier dans pratiquement tous les contextes, alors que d'autres sont très spécifiques pour une séquence oligosaccharidique particulière. Certains épitopes ne sont reconnus par aucune lectine naturelle, et nécessitent la génération d'anticorps monoclonaux.



**Figure 29 : Schémas représentant le principe d'une puce à lectine**

Les puces à lectines utilisent la même architecture que les puces à glycanes. Les lectines ou les anticorps sont spatialement répartis sur une puce en verre par fixation covalente (Tableau 5). Les glycoprotéines du lysat de cellules d'intérêt sont marquées avec un colorant fluorescent de manière non spécifique. L'échantillon est ensuite incubé avec la matrice, une analyse de la fluorescence est ensuite réalisée et de la fluorescence associée à chaque pixel est quantifiée. Le motif de points lumineux reflète la glycome de l'échantillon particulier (Figure 29). A fin de comparaison, deux échantillons peuvent être analysés en parallèle, l'un marqué avec un colorant vert et l'autre avec un colorant rouge par exemple (Figure 30). Asashima et son équipe ont étudié la glycome de cellules souches pluripotentes induites grâce à une puce à lectine (Tateno et al., 2011). La combinaison des deux

échantillons sur un réseau permet une analyse directe de l'évolution du glycome (Gohlke and Klebe, 2001). Hirabayashi et son équipe ont utilisé une puce à lectine pour mettre en évidence des glyco-altérations présents dans le sérum de patients atteints de fibrose du foie (Kuno et al., 2011). Cette même équipe a également étudié le glycome des cellules souches pluripotentes induites (Tateno *et al.*, 2011).



**Figure 30 : Exemple d'un résultat de puce à lectine réalisé sur des cellules souches.**

(Pilobello *et al.*, 2007)

### II.6.3. Vectorisation

En s'appuyant sur la spécificité de reconnaissance glycanique de certaines lectines, il est possible de les utiliser pour le ciblage de cellules tumorales. De plus, leur capacité à être endocytées en font des vecteurs intéressants qui permettraient de faciliter la délivrance intracellulaire de médicaments. Les glycotopes des cellules cancéreuses discutés dans le chapitre précédent représentent des cibles potentielles pour les lectines endogènes comme les galectines, ou exogènes comme les lectines végétales (Rhodes *et al.*, 2008).

Par exemple, la lectine d'arachide PNA (spécifique de l'antigène T) pourrait être utilisée comme vecteur d'agents de détection des cancers du côlon. Sur un modèle de cancer colorectal murin, la délivrance de PNA couplée à un agent d'imagerie permet de distinguer les cellules cancéreuses des cellules saines environnantes (Sakuma *et al.*, 2009, 2010). De même, il a été montré *in vivo* l'intérêt de la PHA de *Phaseolus vulgaris* pour la délivrance d'ADN au niveau de tumeurs surexprimant des glycanes branchés en 1-6-GlcNAc à leur surface (Kasuya *et al.*, 2008). L'intérêt de la lectine de blé WGA a été mis en évidence pour la délivrance d'agents anticancéreux dans des tumeurs du cerveau. L'inconvénient majeur du traitement

chimiothérapeutique des tumeurs cérébrales est la difficulté des agents anticancéreux à traverser la barrière hémato-encéphalique. L'association de la WGA à des liposomes contenant du Topotecan et du Tamoxifène permet d'améliorer la délivrance de ces drogues dans des tumeurs du cerveau chez le rat (Du *et al.*, 2009). L'injection i.v. de ces liposomes double le temps de survie de rats porteurs de gliomes.

#### II.6.4. Agents thérapeutiques

Les lectines sont généralement considérées comme des substances toxiques uniquement destinées à être utilisées comme outils d'analyse des glycoconjugués. Cependant, des études récentes ont montré leur intérêt dans le traitement des cancers (Tableau 6). Les mécanismes d'action anti-tumoraux des lectines se résument plusieurs phénomènes : interaction membranaire et endocytose (Mitchell *et al.*, 1998); cytotoxicité par induction de l'apoptose ou de l'autophagie (Khil *et al.*, 2007); inhibition de la prolifération cellulaire (Mengs *et al.*, 2000); stimulation du système immunitaire (Ma *et al.*, 2008); inhibition de l'angiogenèse (Elluru *et al.*, 2009). Les lectines peuvent être injectées dans la circulation sanguine à des doses non toxiques pour l'organisme afin d'induire spécifiquement la mort de cellules tumorales par apoptose, autophagie, ou en activant les défenses immunes anticancéreuses (Khil *et al.*, 2007). L'apoptose semble être le mécanisme clé de l'effet toxique in vitro des lectines, et peut être engagée par différents mécanismes selon les lectines. De plus, leurs spécificités de reconnaissance glycanique en font des vecteurs intéressants pour le ciblage de médicaments anticancéreux. Les études cliniques, les plus nombreuses ont été réalisées avec des extraits de gui (*Viscum album*) contenant la lectine VAA1 qui est spécifique du galactose et appartient à la famille de la ricine B. Les homologues recombinantes, aviscumine ou rViscumine, sont en cours de tests chez des patients atteints de cancer (Schoffski *et al.*, 2004).

FAMILLE	LECTINE	SPECIFICIT E	MODELE	TYPE D'EFFET	REF
Lectines de légumineuses	Concanavine A ConA	Man/Glc	Cancer hépatique chez la souris	Inhibition formation tumeur et Activation défenses immunes	(Chang <i>et al.</i> , 2007)
			Hepatectomie (2/3) chez la souris	Régénération hépatique	(Huang <i>et al.</i> , 2006)
			Métastases colorectales hépatiques	Inhibition formation tumeur	(Miyagi <i>et al.</i> , 2004)
			Mélanome chez la souris	Activité antitumorale et Cytotoxicité/Inhibition de la croissance	(Taniguchi <i>et al.</i> , 1989; Ueno <i>et al.</i> , 2000)
	Phytohéماغglutinine PHA	GlcNAc(1-2)Man	Cancer colorectale	Ciblage de tumeur	(Mitchell <i>et al.</i> , 1998)
			Cancer métastatique du sein, Cancer métastatique colon, Cancer colique chez souris	Interaction avec cellules tumorales sur coupes histologiques	
Peanut agglutinine PNA	Gal/ antigène T	Cancer colique chez la souris Cancer ovarien solide ou ascitique chez la souris	Interaction et inhibition de la croissance tumorale	(Avichezer and Arnon, 1996; Sakuma <i>et al.</i> , 2009, 2010)	
Lectines de RIP de type II	Viscum album agglutinine (Mistletoe lectin) VAA ou VCA	Gal/GalNAc	Carcinome pancréatique humain (étude clinique)	Augmentation de la durée de vie	(Matthes <i>et al.</i> , 2010)
			Cellules endothéliales	Inhibition de l'angiogénèse	(Elluru <i>et al.</i> , 2009)
			Cancer colique murin	Inhibition de la croissance tumorale via activation du système immunitaire	(Ma <i>et al.</i> , 2008)
			LAL pré-B chez souris	Apoptose Augmentation durée de survie des animaux	(Seifert <i>et al.</i> , 2008)
			Mélanome humain	Inhibition croissance et métastases Effet immuno-modulateurs	(Thies <i>et al.</i> , 2008)
			Cancer colique chez la souris	Apoptose Cytotoxicité/inhibition croissance tumorale	(Khil <i>et al.</i> , 2007)
			Mélanome B16 chez la souris	Inhibition de la croissance tumorale Augmentation sécrétion IL-12	(Duong Van Huyen <i>et al.</i> , 2006)
			Cancer du sein chez la souris	Cytotoxicité/inhibition croissance tumorale	(Beuth <i>et al.</i> , 2006)
			Lymphome non-hodgkinien murin	Cytotoxicité/inhibition croissance tumorale Effet immuno-modulateurs	(Pryme <i>et al.</i> , 2002; Pryme <i>et al.</i> , 2004)
			Mélanome 16 chez la souris	Cytotoxicité/inhibition croissance/apoptose Inhibition angiogénèse Inhibition métastases	(Park <i>et al.</i> , 2001)
			Carcinome de la vessie chez la souris	Cytotoxicité/inhibition croissance	(Menges <i>et al.</i> , 2000)
			Etude clinique sur 41 patients atteints de tumeur solide	Effets secondaires Toxicité hépatique réversible Durée de demi-vie courte	(Schoffski <i>et al.</i> , 2004)
			Métastases chez la souris	Augmentation durée de vie des animaux Diminution du nombre de métastases Activation du système immunitaire	(Schaffrath <i>et al.</i> , 2001)
			Cancer de la vessie induit par traitement chimique chez le rat	Diminution de la transformation tumorale	(Elsasser-Beile <i>et al.</i> , 2001)
			Cancer ovarien humain injecté en i.p chez la souris	Cytotoxicité/inhibition croissance Augmentation durée de vie des animaux	(Schumacher <i>et al.</i> , 2000)
Lectines liant la chitine	Agglutinine de germe de blé WGA	GlcNAc	Gliome murin chez le rat	Ciblage tumeur du cerveau Cytotoxicité/inhibition de la croissance Augmentation durée de vie	(Du <i>et al.</i> , 2009)
	Jacaline	Gal/Antigène T	Biopsies de cancers colorectaux humains	Diagnostic de néoplasmes colorectaux	(Desilets <i>et al.</i> , 1999)

**Tableau 6 : Exemples de lectines végétales utilisées pour leur activité sur les cellules cancéreuses**

## II.6.5. Compréhension du trafic Cellulaire

La capacité des lectines à interagir avec les glycoconjugués de la membrane cellulaire et en particulier avec les glycolipides, en font des outils intéressants pour l'étude du trafic cellulaire, plus particulièrement dans le cas où des anticorps contre les marqueurs organistiques n'ont pas été développés.

Dans une étude récente, portant sur la stabilité et la localisation d'une glycosyltransférase du Golgi, la lectine GNA de perce-neige a été utilisée pour marquer les glycoconjugués des membranes plasmiques du Golgi et a ainsi permis de déterminer le défaut d'activité de l'enzyme lorsqu'elle est localisée en trans-Golgienne (Pokrovskaya *et al.*, 2011). Des lectines présentant d'autres spécificités peuvent être utilisées pour étudier différents trafics cellulaires (Schmelz *et al.*, 1994; Valentijn *et al.*, 2007). Par exemple, l'agglutinine de *Dolichos biflorus* DBA a été utilisée pour étudier les sucres de *Giardi lamblia* (Midlej *et al.*, 2013), tandis que ConA a permis de suivre l'endocytose des trypanosomes africains (Morgan *et al.*, 2004). La lectine de *Lycopersicon esculentum* (LEL) a été utilisée pour cibler spécifiquement le lysosome (McCann *et al.*, 2008; Nolan *et al.*, 1999) et le récepteur de la transferrine (Mehlert *et al.*, 2012).

Une autre étude a démontré que les lectines végétales peuvent être utilisées pour déterminer l'état de glycosylation des glycoprotéines, mais aussi permettre l'étude du trafic vésiculaire intracellulaire en utilisant la biochimie en conjonction avec l'analyse cytochimique. En effet, Vaux et son équipe ont utilisé des lectines telles que l'agglutinine de *Galanthus nivalis* (GNA), l'agglutinine de *Pisum sativum* (PSA), l'agglutinine de *Lens culinaris* et la concanavaline A pour étudier la voie de sécrétion précoce des cellules HeLa (Morgan *et al.*, 2013).

D'autres chercheurs se sont intéressés à la toxine de Shiga. Cette toxine est devenue un outil puissant pour l'étude des processus cellulaires, notamment dans le cadre du trafic intracellulaire (Ewers *et al.*, 2010; Romer *et al.*, 2007; Romer *et al.*, 2010). L'un des défis les plus importants dans ce contexte est de parvenir à une compréhension fonctionnelle de la façon dont une protéine s'associe à la membrane exoplasmique et permet ainsi l'invagination des membranes (Romer *et al.*, 2007). Pour relever ce défi, il faut de nouvelles technologies afin d'obtenir des images haute résolution en temps réel de la dynamique des lipides. De plus, seul le domaine de la biophysique des membranes peut formuler des concepts qui devraient améliorer notre compréhension de la mécanique des membranes. La Toxine Shiga a

commencé à attirer l'attention comme moyen de vectorisation dans la tumeur mais ces études sont encore dans une phase précoce de développement (Johannes and Romer, 2010).

## **II.7. Ingénierie des lectines**

### **II.7.1. Modification de spécificité**

L'ingénierie de lectines dans le but de modifier leur spécificité est un objectif de longue date, mais le nombre de résultats positifs est encore limitée (Tableau 7). Seules les études présentant un volet expérimental sont présentées ci-dessous. L'ingénierie des lectines en  $\beta$ -propeller est présentée quant à elle dans le paragraphe III.3. de la partie Introduction.

#### **II.7.1.1. Lectines de plantes**

Les pionniers dans le domaine de l'ingénierie des lectines sont Sharon et ses collègues qui ont utilisé la mutagenèse dirigée pour augmenter l'affinité de la lectine *Erythrina corallodendron* pour des dérivés de galactose avec des substituants en C2 (Arango *et al.*, 1993). De même, la spécificité de la lectine de haricot de Lima, une lectine de légumineuse spécifique au galactose a été étudiée. Le mutant C127Y affiche une plus grande affinité pour le trisaccharidique de type A comparé à la lectine sauvage (Jordan and Goldstein, 1995).

La lectine PNA (peanut agglutinin) présente une spécificité intéressante pour le Thomsen-Friedenrich antigen (antigène T) qui est présent à la surface de cellules malignes mais aussi de thymocytes immatures. La lectine PNA pourrait être utilisée comme outil de reconnaissance de ces cellules, mais malheureusement elle possède aussi une affinité pour le N-acétyllactosamine (LacNac) qui est présent sur plusieurs glycoprotéines de la surface cellulaire. Sharma et son équipe ont modifié la spécificité de cette lectine par mutagenèse dirigée en modifiant la boucle la plus variable du site de liaison. Ils ont produit deux mutants en substituant la leucine 212 par une asparagine ou une alanine, et ont obtenu des lectines qui reconnaissent préférentiellement l'antigène T ou le LacNac respectivement (Sharma *et al.*, 1996). Une autre étude plus récente a aussi eu pour but d'augmenter l'affinité de la lectine PNA pour l'antigène T. Les auteurs ont utilisé les données structurales pour créer des mutations par une méthode rationnelle. Ainsi le mutant N41Q présente une affinité pour



l'antigène T six fois supérieure à la protéine sauvage exprimée de façon recombinante (Adhikari *et al.*, 2001).

La mutation aléatoire des acides aminés présents dans le site de liaison de la lectine est très efficace pour produire des bibliothèques de mutants qui présentent des spécificités différentes. La lectine de *Maackia amurensis* (MAH) a fait l'objet de plusieurs études afin d'affiner la spécificité pour le sialyllactose. Les 24 nucléotides qui composent le site de liaison ont été mutés au hasard, et 16 mutants sont capables d'agglutiner des érythrocytes animaux et humains (Yim *et al.*, 2001). Dans une autre étude, 35 mutants ont été préparés en variant deux acides aminés dans le domaine de reconnaissance des sucres, et ont pu être utilisés pour la caractérisation de lignées cellulaires (Maenuma *et al.*, 2008). Cette bibliothèque de MAH modifiées est un outil intéressant pour le diagnostic de l'IgA sérique en néphropathie (Maenuma *et al.*, 2009).

En plus des mutations ponctuelles, une autre stratégie est de construire des protéines chimériques. Une construction a été réalisée par assemblage de nonapeptides de la lectine de *Bauhinia purpurea* spécifique pour le galactose et de la lectine de lentille (*Lens culinaris*) spécifique pour le mannose. La lectine chimère peut lier à la fois le mannose et le galactose (Yamamoto *et al.*, 1992).

### II.7.1.2. Lectines de champignons et d'invertébrés

Les oligosaccharides sialylés ont une grande importance biologique et de nombreux efforts ont été réalisés dans l'objectif d'obtenir des lectines avec une spécificité élevée pour les oligosaccharides NeuAc $\alpha$ 2-3Gal ou NeuAc $\alpha$ 2-6Gal. La galectine fongique d'*Agrocybe cylindracea* a été modifiée afin de baisser son affinité pour les oligosaccharides galactosylés non sialylés. Le mutant E86D a sensiblement perdu sa capacité de liaison au LacNAc et à l'antigène T, alors qu'il conserve une haute affinité pour le 3'-sialyllactose (Tableau 7) (Nurisso *et al.*, 2010). L'évolution dirigée basée sur la PCR a été utilisée pour produire des lectines se liant à un NeuAc à partir de la lectine à galactose de type R du ver de terre *Lumbricus terrestris*. Après sélection sur fétuine-agarose, l'un des mutants obtenu a montré une affinité substantielle pour le Neu5Ac $\alpha$ 2-6Gal (Yabe *et al.*, 2007). Cette lectine de ver de terre a également été modifiée pour obtenir une spécificité pour les oligosaccharides terminés par des 6-sulfo-galactose (Gal-6S) qui sont associés à la malignité des glioblastomes.

Plusieurs cycles de sélection indiquent que la mutation E20K est nécessaire et suffisante pour obtenir la spécificité pour le 6S-Gal (Hu *et al.*, 2012).

Les lectines de type C d'invertébrés marins ont également fait l'objet d'ingénierie dans le but de changer la spécificité. Ainsi, la lectine CEL-I isolée à partir du concombre de mer *Cucumaria echinata* présente une spécificité pour le galactose grâce à un motif de reconnaissance Gln-Pro-Asp. La mutation de l'aspartate en asparagine a pour conséquence le changement de la spécificité de la protéine qui devient alors spécifique du mannose (Hatakeyama *et al.*, 2013).

### II.7.1.3. Lectines de vertébrés

Les lectines de type C de vertébrés ont aussi fait l'objet d'ingénierie. En effet, les lectines de type-C spécifiques pour le mannose sont faciles à produire et à caractériser, alors que celles spécifiques pour le galactose présentent des difficultés particulières. La spécificité des CRDs dépend d'une courte séquence d'acides aminés (EPN pour le mannose, contre QPD pour le galactose). Les stratégies ont porté sur l'insertion de 3 portions de la lectine de foie de rat (RHL-1) spécifique du galactose (Tableau 7) (Iobst *et al.*, 1994), ou sur la substitution de la séquence EPNDHGS de la MBP par la séquence GPDDWYGHGLGG de la RHL-1 (Topin *et al.*, 2014). La substitution des acides aminés A211, S212 et H213 par trois lysines a permis de modifier la spécificité de la MPB-A du rat pour qu'elle reconnaisse le Le<sup>x</sup> et le SLe<sup>x</sup> (Ng and Weis, 1997).

D'autres lectines comme les galectines sont utilisées dans l'ingénierie de spécificité. Ainsi, Salomonsson et son équipe ont produit plusieurs mutants de spécificités différentes. Le mutant K176L de la galectine-3 a une affinité plus élevée pour les  $\beta$ -galactosides, alors que le mutant R182S a une affinité plus faible pour le LacNAc (Salomonsson *et al.*, 2010).

## INTRODUCTION

Lectine	Organisme	Spécificité d'origine	Méthode	Mutation	Spécificité acquise	Réf
ECorL	<i>Erythrina corallodendron</i>	Galactose	Mutagenèse dirigée	Y108T	Galactose ↑	(Arango <i>et al.</i> , 1993)
LBL	<i>Lima Bean</i>	trisaccharide groupe sanguin A	Mutagenèse dirigée	C127Y	groupe sanguin A ↑	(Jordan and Goldstein, 1995)
RBL	<i>Robinia pseudoacacia</i>	Galactose	Biologie moléculaire	Insertion d'un site spécifique du mannose	Mannose	(Nishiguchi <i>et al.</i> , 1997)
BPA LCA	<i>Bauhinia purpurea</i> <i>Lens culinaris</i>	Galactose Mannose		Protéine chimérique	Galactose et Mannose	(Yamamoto <i>et al.</i> , 1992)
PNA	Peanut	antigène T LacNac	Mutagenèse dirigée	L212N	antigène T ↑ LacNac ↓	(Sharma <i>et al.</i> , 1996)
PNA	Peanut	antigène T LacNac	Mutagenèse dirigée	L212A	antigène T ↓ LacNac ↑	(Sharma <i>et al.</i> , 1996)
PNA	Peanut	antigène T LacNac	Mutagenèse dirigée	N41Q	antigène T ↑ LacNac	(Adhikari <i>et al.</i> , 2001)
MAH	<i>Maackia amurensis</i>		Mutagenèse aléatoire	6'-sialyllactosamine	Erythrocytes animaux et humains	(Yim <i>et al.</i> , 2001)
ACG	<i>Agrocybe cylindracea</i>	LacNac Antigène T 3'-Slactose	Mutagenèse par saturation	E86D	LacNac ↓ Antigène T ↓ 3'-sialyllactose ↑	(Nurisso <i>et al.</i> , 2010)
EW29	<i>Lumbricus terrestris</i>	Galactose	Evolution dirigée		Neu5Acα2-6Gal	(Yabe <i>et al.</i> , 2007)
EW29	<i>Lumbricus terrestris</i>	Galactose	Evolution dirigée	E20K	6S-Gal	(Hu <i>et al.</i> , 2012)
CEL-I	<i>Cucumaria echinata</i>	Galactose	Mutagenèse dirigée	D103N	Mannose	(Hatakeyama <i>et al.</i> , 2013)
galectin-3	Homme	β-galactoside	Mutagenèse dirigée	K176L	β-galactoside ↑	(Salomonsson <i>et al.</i> , 2010)
galectin-3	Homme	β-galactoside	Mutagenèse dirigée	R182S	LacNac ↓	(Salomonsson <i>et al.</i> , 2010)
galectin-3	Homme	β-galactoside	Mutagenèse dirigée	R144S	Type A	(Salomonsson <i>et al.</i> , 2010)
MBP	Homme	Mannose		Insertion de 3 portions de RHL-1 dans le MBP	Galactose	(Iobst and Drickamer, 1994)
MBP	Homme	Mannose	Mutagenèse dirigée	GPDDWYGHGLGG à la place de EPNDHGS Insertion d'une portion de RHL-1	Galactose	(Kolatkar and Weis, 1996)
MBP	Homme	Mannose	Mutagenèse dirigée	GPDDWYGHGLGG à la place de EPNDHGS Insertion d'une portion de RHL-1	GalNac	(Topin <i>et al.</i> , 2014)
PA-IIL	<i>Pseudomonas aeruginosa</i>	Fucose	Mutagenèse dirigée	S22A	Mannose ↑	(Adam <i>et al.</i> , 2007)
PA-IIL	<i>Pseudomonas aeruginosa</i>	Fucose	Mutagenèse dirigée	G24N S23A	Fucose ↑	(Adam <i>et al.</i> , 2007)
PA-IIL	<i>Pseudomonas aeruginosa</i>	Fucose	Evolution dirigée	H50V, D52C, Q53R H50T, D52N, Q53R H50N, D52N, Q53G H50V, D52C, Q53E	Galα1-3Gal	(Keogh <i>et al.</i> , 2014)
PA-IIL	<i>Pseudomonas aeruginosa</i>	Fucose	Mutagenèse dirigée	G24N G24W	Fucose ↑	(Kriz <i>et al.</i> , 2014)
RSL	<i>Ralstonia solanacearum</i>	Fucose	Mutagenèse dirigée	W31A W76A	Fucose ↓	(Mishra <i>et al.</i> , 2012; Wimmerova <i>et al.</i> , 2012)
AAL	<i>Aleuria aurantia</i>	Fucose		N224Q	Fucα1-6GlcNAc ↑	(Romano <i>et al.</i> , 2011)

**Tableau 7 : Exemples d'ingénierie de spécificité de lectines**

#### II.7.1.4. Lectines de bactéries

Sur la base de la structure cristalline de LecB, une lectine à fucose de *Pseudomonas aeruginosa*, et RS-IIL une lectine à mannose de *Ralstonia solanacearum*, il a été possible de concevoir des mutants de LecB dans le but de renforcer la spécificité de fucose, ou de commuter la préférence pour le mannose (Tableau 7) (Adam et al., 2007; Kostlanova et al., 2005; Mitchell *et al.*, 2002). Cette même équipe a récemment modifié la spécificité de LecB *in silico* par mutagenèse dirigée par saturation sur les acides aminés 22, 23 et 24. Les mutants clés ont ensuite été créés pour une étude par ITC. Les mutants G24R, G24N et G24W voient leur affinité pour l' $\alpha$ Mefuc augmenter (Kriz *et al.*, 2014).

La lectine à galactose LecA de *P. aeruginosa* a également été mutée pour augmenter son affinité vers des glycosylations d'intérêt. Une série de triple mutants (H50V, D52C, Q53R et H50T, D52N, Q53R et H50N, D52N, Q53G et H50V, D52C, Q53E) obtenue par évolution dirigée a vu leur affinité pour le Gal $\alpha$ 1–3Gal augmenter par rapport à l'affinité pour le LacNAc (Keogh *et al.*, 2014).

#### II.7.2. Modification de valence :

Des mutants monovalents de lectines sont très utiles pour l'analyse des glycoconjugués de surface des cellules contenant de l'acide sialique par la technique de cytométrie en flux. En effet, grâce à la monovalence, les lectines n'agglutinent pas les cellules ce qui facilite l'analyse.

La production de lectines monomériques peut se faire par des modifications chimiques de la surface protéique qui préviennent les processus d'oligomérisations. De telles approches ont été utilisées sur la concanavaleine A, qui est naturellement sous forme d'un tétramère à pH 7. Des formes dimériques de ConA ont été produites grâce à des modifications chimiques : succinyl-ConA et acétyl-ConA (Gunther *et al.*, 1973). Une combinaison de succinylation et de photomarquage par l'azidophenyl- $\alpha$ -D-mannopyranoside a permis la production de ConA monovalente à pH 5, ne possédant donc pas d'activité d'hémagglutination (Beppu *et al.*, 1976). La leukoagglutinine de *Maackia amurensis* (MAL), spécifique pour le NeuAca2-3Gal,

et la lectine d'écorce du sureau japonais *Sambucus sieboldiana* (SSA) spécifique pour le NeuAca2-6Gal ont également été modifiées chimiquement. La réduction sélective de ponts disulfures entre les sous-unités suivie d'une alkylation a permis l'obtention de protéines monomériques (Kaku *et al.*, 1993; Kaku and Shibuya, 1992).

Une autre méthode consiste à empêcher la formation de ponts disulfure, en remplaçant par ingénierie moléculaire les cystéines impliquées par des sérines. Des galectines sont aussi fréquemment obtenues sous formes monomériques (Wang *et al.*, 2008).

Cependant, l'augmentation de la polyvalence d'une lectine peut être intéressante pour améliorer son affinité à l'égard des surfaces glycosylées. Deux lectines spécifiques du NeuAca2-6Gal issues du ver de terre ont été assemblées par un linker polypeptide. Cette construction se traduit par une augmentation de l'affinité de près de 10 fois (Yabe *et al.*, 2009). La protéine résultante se lie fortement aux surfaces de globules rouges sans hémagglutination et a donc un grand intérêt pour l'étude en cytométrie de flux.

L'introduction de résidus capables de créer des ponts disulfures entre les monomères de cystéines a été utilisée pour permettre l'oligomérisation. Une forte augmentation de la puissance d'hémagglutination a été observée pour le mutant P71C de la lectine de monocotylédones *Galanthus nivalis* (GNA) spécifique du mannose par rapport au type sauvage. Cette propriété peut être désactivée lors de l'ajout de DTT (McDonald *et al.*, 2004).

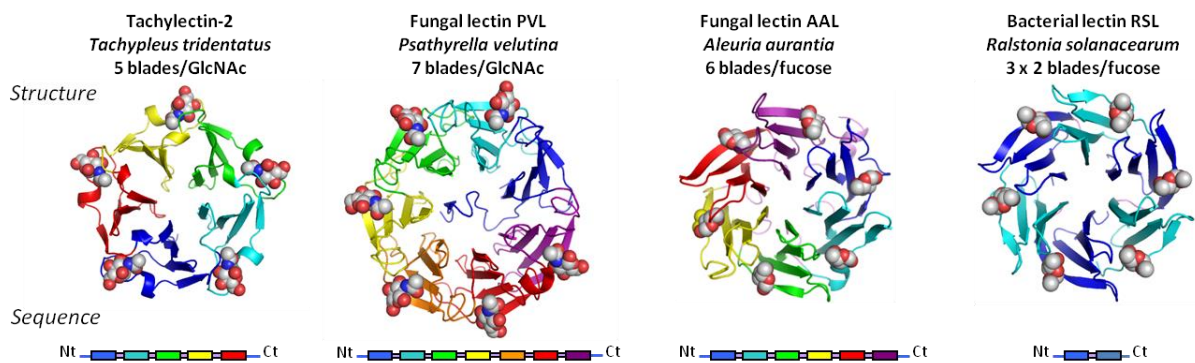
### **III. Objectif et Conception de néolectines**

L'objectif de ma thèse est de concevoir des lectines de valence, topologie et de spécificité modulable. Ainsi les lectines mutées au niveau de leur valence et topologie permettront de mieux étudier les phénomènes associés à la multivalence, tels que les mécanismes d'endocytose induits par les lectines.

A cet effet, une lectine hautement symétrique et très stable a été choisie et permettra d'établir un projet d'ingénierie protéique. Les lectines sous forme de béta-propeller présentent les caractéristiques recherchées.

### III.1. Béta-propeller

Le repliement en bêta propeller a été identifié dans de nombreuses familles protéiques à la fois chez les procaryotes et les eucaryotes. Ce repliement est une structure très symétrique qui est basée sur une répétition de 4 à 10 lames ou « blades » disposés en torus autour d'un axe central. Chaque lame est typiquement composée de quatre brins  $\beta$  antiparallèles. Les feuillettes sont disposés radialement autour d'une cavité centrale (Figure 31). L'interaction majoritairement hydrophobe entre les feuillettes fournit une grande partie de la stabilité structurelle même si elle n'est pas la seule force de stabilisation. Bien que l'ensemble des structures de ces protéines présente des similitudes importantes, la plupart des protéines ont évolué de façon nettement différentes quant à la manière de fermer le cercle. Certains  $\beta$ -propeller se ferment par interactions entre les lames 1 et 3 comme la Tachylectine 2 (Beisel *et al.*, 1999) ou entre les lames 2 et 2 (Renault *et al.*, 1998) ou grâce à des ponts disulfures (Faber *et al.*, 1995). La rigidité de cette structure semble être une caractéristique qui a conduit à son utilisation dans une large variété de rôles fonctionnels (Fulop and Jones, 1999) (Tableau 8).



**Figure 31 : Structure de différents  $\beta$ -propeller**

De gauche à droite la tachylectine 2 (Beisel 1999) code PDB 2Z48, PVL (cioci 2006) code PDB 2BWN, AAL (Fujihashi 2003) code PDB 1OFZ et RSL (Kostlanova 2005) code PDB 2BT9.

Nombre de blades	Nom	Fonction	Séquence/Motif	Fermeture	Code PDB	Références
4	Hemopexine	Liaison de l'Hème et délivrance	Résidus conservés hydrophobiques	Pont disulfure	1hxn 1qhu	(Faber <i>et al.</i> , 1995; Paoli <i>et al.</i> , 1999)
4	Collagenase	Clivage du collagène	Résidus conservés hydrophobiques	Pont disulfure	1fbl	(Li <i>et al.</i> , 1995)
5	Tachylectine-2	Liaison GlcNAc et GalNAc	DNWL IGxGGW	3+1	1tl2	(Beisel <i>et al.</i> , 1999)
6	Neuraminidase et sialidase	Clivage de l'acide sialique terminal à partir de glycoconjugués	Boîte aspartate	1+3	7nn9, 2sil4	(Crennell <i>et al.</i> , 1993; Varghese <i>et al.</i> , 1983)
6	Glucose deshydrogénase	Oxidation des sucres en lactones	Motifs de liaison de deux différents	1+3	1qbi	(Oubrie <i>et al.</i> , 1999)
7	Méthylamine deshydrogénase	Conversion de méthylamine en ammoniac et aldéhyde	Néant	3+1	1mda	(Vellieux <i>et al.</i> , 1989)
7	Galactose oxidase	Oxidation des alcools en aldéhydes	Motif Kelch	1+3	1gof	(Ito <i>et al.</i> , 1991)
7	Sous-unités $\beta$ protéine G	Signalisation	Répétition de WD	1+3	1gg2	(Wall <i>et al.</i> , 1995)
7	Régulateur de la condensation des chromosomes (RCC1)	Transport nucléocytoplasmique and régulation du cycle cellulaire	VYxWG	2+2	1a12	(Renault <i>et al.</i> , 1998)
7	Prolyl oligopeptidase	Clivage liaison peptidique	Néant	Néant	1qfm	(Fulop <i>et al.</i> , 1998)
7	Fragment N-terminal de la chaîne lourde de la clathrine	Transport des lipides et protéines entre les compartiments membranaires intracellulaires	Résidus hydrophobiques conservés	1+3	1bpo	(ter Haar <i>et al.</i> , 1998)
8	Méthanol deshydrogénase	Oxidation du méthanol en formaldéhyde	Motif W-docking	1+3	4aah	(Xia <i>et al.</i> , 1992)
8	Cytochrome <i>cd<sub>1</sub></i>	Réduction des nitrites en acide nitrique et/ou de l'oxygène en eau	Néant	3+1	1qks	(Baker <i>et al.</i> , 1997; Fulop <i>et al.</i> , 1995)

**Tableau 8 : Différentes protéines avec une structure en Béta-propeller.**

### III.2. Les lectines en bêta-propeller

Trois familles de lectines, ne présentant aucune similarité de séquences entre elles adoptent une architecture en  $\beta$ -propeller (Tableau 8). La tachylectine-2, un  $\beta$ -propeller à 5 lames est purifiée à partir de la limule *Tachypleus tridentatus* et présente une spécificité pour le GlcNAc La protéine PVL du champignon *Psathyrella velutina* est quant à elle composée de 7 lames et reconnaît spécifiquement les glycanes se terminant par un résidu GlcNAc ou un acide neuraminique (Cioci *et al.*, 2006). Une lectine similaire a récemment été caractérisée à partir du champignon *Agrocybe aegerita* (Jiang *et al.*, 2012).

Lectine	Origine	Oligomérisation	Poids moléculaire (kDa)	Spécificité	lames / monomère	sites/ oligomère	Réf
AAL	<i>Aleuria aurantia</i>	Homodimère	36	Fuc avec préférence pour Fuc $\alpha$ 1-6 et Fuc $\alpha$ 1-3	6	12	(Fujihashi <i>et al.</i> , 2003)
AOL	<i>Aspergillus oryzae</i>	Homodimère	35	Fuc $\alpha$ 1-6 et autres Fuc	6	12	(Matsumura <i>et al.</i> , 2007)
AFL	<i>Aspergillus fumigatus</i>	Homodimère e	34	Fuc $\alpha$ 1-2 et autres Fuc	6	12	(Houser <i>et al.</i> , 2013)
RSL	<i>Ralstonia solanacearum</i>	Homotrimère	30	Fuc $\alpha$ 1-2 et autres Fuc	2	6	(Kostlanova <i>et al.</i> , 2005)
BambL	<i>Burkholderia ambifaria</i>	Homotrimère	30	Fuc $\alpha$ 1-6 et autres Fuc	2	6	(Audfray <i>et al.</i> , 2012)

**Tableau 9 : Lectines de la famille de RSL et AAL**

La troisième famille de lectine correspond à une architecture en 6 lames avec une spécificité pour le fucose. (Tableau 9). Ces lectines sont d'origine fongique (AAL: *Aleuria aurantia*, AFL: *Aspergillus fumigatus*, AOL: *Aspergillus oryzae*) ou bactérienne (RSL: *Ralstonia solanacearum*, BambL: *Burkholderia ambifaria*). La sous-famille bactérienne a la particularité de former le  $\beta$ -propeller par trimérisation de trois peptides à 2 lames. C'est le seul exemple connu de  $\beta$ -propeller formé par oligomérisation et non par répétition-tandem, et cela représente peut-être une étape évolutive antérieure. Les acides aminés impliqués dans les interactions entre lames et dans la reconnaissance du fucose sont conservés dans cette famille. Il s'agit en particulier d'une arginine et d'un glutamate qui stabilisent le sucre par des liaisons hydrogène. Un tryptophane est aussi fortement conservé et peut être substitué par une tyrosine. Dans les deux cas, le cycle aromatique stabilise le fucose par des interactions hydrophobes de stacking (Figure 32).



ALF1	2	STPGAQQVLFRTGIAAVNS-----TNHLR	VYFQDV---	YGSIR	SLYE-G-S	WANGTEKNVI-----	53
ALF2	54	-----GNAKLGSPVAATSK-----ELKHI	VYTLTE---	GNTLQ	EFAYDSGTG	WYNGGLGGAK-----	103
ALF3	104	-----FQVAPYSCIAAVFLAG--TDALQL	RYAQKP---	DNTIQ	EYMWN-GDC	WKEGTWLGGA-----	155
ALF4	156	-----LPGTGIGATSFRTDYNGPSIR	YWFQTD---	DLKLY	QRAYDPHK	WYDPDLVTIFD-----	207
ALF5	208	-----RAPPRTAIAATSFGAGNS-SIYM	IYFVNS---	DNTIW	QVCWDHGK	GYHDKGTITPV-----	260
ALF6	261	-----IQGSEVAIISWGSFANNGPDL	LYFQNG-TYISAV	SENV	WNR	AHGSQ LGRSALPPA----	315
AAL1	1	----PTEFLYTSKIAAISWA--ATGGRQQ	VYFQDL---	NGKIR	EAQRGGDN	PWTGGSSQ NVI-----	54
AAL2	55	-----GEAKLFSPLAAVTWKS AQGI--QI	VYCVNK---	DNILS	EFVYD-GSK	WITGQLGSVG-----	106
AAL3	107	-----VKVGSNSKLAALQWGGSESAPPNI	VYQKSNLSGSS	IEEYV	WS-G-KW	TAGASFGST-----	162
AAL4	163	-----APGTGIGATAIGP-----GRLR	IYYQAT---	DNKIR	EHCWD-SNS	WYVGGFSASA-----	208
AAL5	209	-----SAGVSIAAISWGS-----TPNI	VYWQKG---	REELY	EAYG-G-SW	YTPGQIKDA-----	254
AAL6	255	--SRPTPSLPDTFIAANSSGN-----IDI	VFFQAS---	GVSLQ	QWQWISGK	WSIGAVVPTGTPAGW	312
RSL1	1	-----SSVQTAATSW----GTVPSI	VYTANN---	GKIT	ERCWD-GKG	WYTGAFNE-----	43
RSL2	44	-----PGDNVSVTSWL--VGS	AIHIVYASTG---	TTTT	EWCD-GNG	WTKGAYTATN-----	90
BambL1	1	-----MQTAAISWG----TTPSI	VYTANG---	NKIT	ERCYD-GSN	WYTGAFNQ-----	41
BambL2	42	-----AGDNVSATCWLSG--SAVHI	VYATSG---	GSTT	EWCD-GDG	WTRGAYTGL-----	90

**Figure 32 : Alignement des séquences des sites de fixations des principales lectines bactériennes**

*En rouge, jaune, bleu et noir sont mis en évidence les acides aminés fortement impliqués dans la liaison avec le fucose et qui sont conservés d'une lectine à l'autre.*

### III.3. Ingénierie de béta-propeller

Le  $\beta$ -propeller est une architecture basée sur la répétition d'éléments qui permet les approches top-down de déconstruction basées sur les éléments de symétrie (Blaber and Lee, 2012). La possibilité de reproduire l'évolution et de reconstruire un  $\beta$ -propeller fonctionnel à partir de fragments a été explorée par le groupe de Dan Tawfik à partir de la Tachylectine-2. Diverses bibliothèques ont été créées par troncature graduelle du gène codant pour la tachylectine 2 en vérifiant que la fonctionnalité de la lectine était conservée. Plusieurs peptides correspondant à deux lames ont été isolés. Ils présentent la capacité de s'oligomériser. La spécificité et l'affinité sont presque les mêmes entre le type sauvage et des segments (Topin *et al.*, 2013). Toutefois, l'oligomérisation est différente en raison d'un échange de brins béta entre deux  $\beta$ -propellers voisins (Rossner *et al.*, 2013).

Les lectines avec une architecture en  $\beta$ -propeller ont été ingénierées afin de modifier leur valence. En effet, la lectine AAL a été modifiée par mutagenèse dirigée pour créer une forme monomérique. Une concentration 20 fois plus élevée de mAAL est nécessaire pour induire une hémagglutination par rapport à la protéine sauvage. Dans ce même article, Olausson *et al.* ont produit une petite lectine à partir de la lectine AAL correspondant

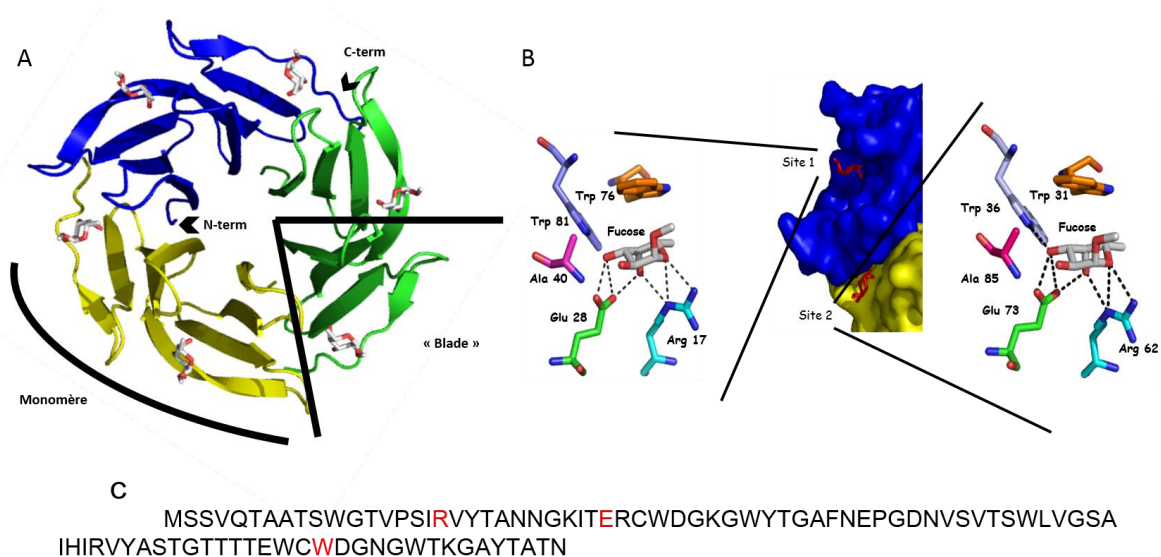
aux acides aminés 50 à 160 qui ne composent qu'un site de fixation appelée S2-AAL (Olausson *et al.*, 2011). Ces protéines peuvent donc facilement être utilisées comme marqueur, ainsi l'absence d'agglutination facilite l'analyse par cytométrie en flux par exemple. La protéine mutante S2-AAL a ensuite fait l'objet d'une étude afin de mettre au point une chromatographie de faible affinité pour étudier la spécificité de la protéine (Bergstrom *et al.*, 2012).

### III.4. RSL : un bêta-propeller trimérique hexavalent

La lectine RSL a été identifiée dans le génome de la bactérie pathogène de plantes *Ralstonia solanacearum*. La lectine RSL recombinante présente une forte affinité (de l'ordre du micromolaire) pour les oligosaccharides fucosylés, en particulier pour la séquence Fuc $\alpha$ 1-2Gal présente dans la paroi des plantes mais aussi sur les tissus humains (Kostlanova *et al.*, 2005).

La séquence est composée de 90 acides aminés et met en évidence la présence de deux motifs peptidiques répétés quasiment identiques qui forment les deux lames d'un monomère. Chaque lame est composée de 4 petits brins  $\beta$  antiparallèles. La structure cristallographique de RSL montre que la protéine se trimérise en une architecture en  $\beta$ -propeller composée de lames. RSL adopte donc une structure tertiaire en  $\beta$ -propeller dérivée de l'oligomérisation de 3 sous-unités et non par le repliement d'un polypeptide unique. Cette structure hautement symétrique est formée par un trimère dont chaque monomère comporte deux sites de fixation au fucose (Figure 33). Par conséquent, RSL est hexavalente avec deux types de sites de fixation au fucose : 3 sites intramonomériques et 3 sites intermonomériques.

Ces sites ont été finement caractérisés et sont très similaires. Le fucose établit principalement des liaisons hydrogène avec une arginine et un glutamate. Un tryptophane est également impliqué dans la stabilisation du fucose par des interactions hydrophobes (Figure 33 B) (Kostlanova *et al.*, 2005).



**Figure 33 : Structure cristallographique de RSL en présence d' $\alpha$ MeFuc.**

Code PDB 2BT9 (Kostlanova *et al.*, 2005)

(A) Trimère de RSL cristallisé en présence d' $\alpha$ MeFuc. Les sites 1 correspondent aux sites intramonomériques et les sites 2 aux sites intermonomériques. Chaque monomère est représenté par une couleur différente. Une des lames (blade) et un des monomères sont matérialisés. (B) Site intramonomérique. Les liaisons hydrogène entre l' $\alpha$ MeFuc et RSL sont visibles en pointillés. (C) Séquence en acides aminés d'un monomère de RSL. En rouge sont visibles les acides aminés principalement impliqués dans la liaison avec l' $\alpha$ MeFuc dans le site intramonomérique.

RSL a été l'objet d'études théoriques et expérimentales afin d'analyser le rôle de l'acide aminé aromatique dans la reconnaissance du fucose. Les auteurs ont ainsi démontré le rôle crucial du tryptophane qui est majoritairement impliqué dans la stabilisation du sucre dans le site de fixation par un effet de stacking (Mishra *et al.*, 2012; Wimmerova *et al.*, 2012). Cette importance du résidu aromatique dans la reconnaissance du fucose avait été démontrée auparavant dans la lectine AAL en mutant les tyrosines 26, 79, 92, 181 et 241 ou le tryptophane 149 (Amano *et al.*, 2004). Par ailleurs, les acides aminés adjacents au site à fucose jouent un rôle important dans la spécificité pour les oligosaccharides, comme démontré par la mutation N224Q de AAL qui induit une spécificité plus forte pour le motif Fuc $\alpha$ 1-6GlcNAc (Smerdova *et al.*, 2013).

## IV. Stratégie proposée

Le but de ma thèse est d'obtenir des lectines de valence, topologie et spécificité modulable. Ainsi, les lectines mutées au niveau de leurs valences et topologies, nous permettront de mieux étudier le mécanisme d'action de la multivalence sur l'avidité et sur le mécanisme d'endocytose. Il sera par exemple possible d'analyser l'influence des sites de fixation ainsi que la localisation d'un site par rapport aux autres sur la formation de tubules. D'autre part, le changement de spécificité de la lectine RSL peut nous permettre de créer un outil de reconnaissance d'oligosaccharides mais aussi un outil de marquage de cellules présentant ces sucres. Ces lectines pourraient également devenir des outils de vectorisation.

La première stratégie consiste en une modification de la valence de la lectine RSL. La lectine possédant deux sites de fixations par monomère, l'inactivation d'un site de fixation par monomère entraîne l'inactivation de trois sites de fixation au niveau de la protéine trimérique fonctionnelle. Cette approche nous permettra de comparer le comportement d'une lectine hexamérique avec une lectine trimérique.

Afin d'accéder à d'autres valences, la seconde stratégie sera de construire une protéine de même architecture que RSL mais présentant les six lames sur même polypeptide (neoRSL monomérique) par ajout de deux linkers entre trois séquences RSL. En utilisant un ADN suffisamment dégénéré à chaque site, il est possible d'inactiver par mutagenèse dirigée les sites de fixation indépendamment les uns des autres. Une librairie de mutants de valence contrôlée (de 0 à 6) pourra ainsi être créée, puis testée pour leurs propriétés biochimiques et biophysiques.

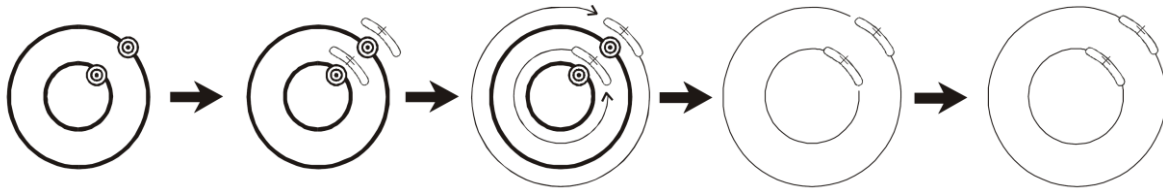
La création d'une protéine monomérique de valence 1 permettra de plus la création d'un outil dont la spécificité est plus facilement modulable et surtout caractérisable. Ainsi, il est possible de modifier la spécificité du seul site actif de la protéine monomérique monovalente, par mutagenèse dirigée ou aléatoire. Une fois les mutations d'intérêts caractérisées par séquençage, il sera possible de modifier tous les sites de la protéine en réalisant cette mutation dans le but d'obtenir une protéine hexasvalente avec une nouvelle spécificité. Il sera également possible de garder certains sites inactifs afin d'avoir la valence et la topologie souhaitées. Les mutations qui permettent le changement de spécificité peuvent aussi être appliquées à la protéine RSL trimérique. Nous avons, sur ce concept, déposé un brevet (Imberty et al., 2014).

## V. Méthodologie

### V.1. Biologie moléculaire et biochimie

#### V.1.1. Mutagenèse dirigée

L'insertion ou la mutation d'un nombre important de bases nucléotidiques dans un gène a été réalisée selon la méthode GeneTailor™ Site-Directed Mutagenesis System, commercialisée par Invitrogen. Le principe repose sur celui développé pour de la mutagenèse classique, à l'exception de conception des amorces (Figure 34).



**Figure 34 : Principe de la mutagenèse dirigée**

1. *Plasmide méthylé.*
2. *Dénaturation du plasmide et fixation avec deux amorces présentant en leur centre la mutation d'intérêt.*
3. *Amplification du plasmide par PCR.*
4. *Digestion de l'ADN méthylé matrice non muté*
5. *Obtention du plasmide muté qu'il faut amplifier après transformation dans des bactéries.*

#### V.1.2. Production et purification des lectines

La production des lectines a été réalisée dans un système hétérologue bactérien. Les gènes codant pour les protéines ont été clonés dans des plasmides de type pET et transformés dans la souche d'expression TUNER™ (DE3) d'*Escherichia coli*. L'induction de la production est réalisée lorsque la densité optique à 600 nm est d'environ 0.7 par l'ajout d'IPTG.

Les purifications sont réalisées par chromatographie d'affinité, soit sur colonne de mannose ou fucose, soit sur colonne de nickel. Dans certains cas, une chromatographie d'exclusion de taille a ensuite été réalisée pour permettre l'obtention d'une protéine pure.

## V.2. Caractérisation

### V.2.1. Caractérisation biochimique

#### V.2.1.1. Microcalorimétrie de Titration Isotherme

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

L'interaction non covalente entre une macromolécule  $M$  et son ligand  $L$  conduit à la formation réversible d'un complexe  $ML$ . La force de l'interaction est représentée par la constante d'association  $K_a$  en  $M^{-1}$ . Plus cette constante est grande, plus l'affinité entre les deux partenaires est forte. Le  $K_a$  correspond au rapport de la concentration du complexe  $ML$  sur celles de la macromolécule  $M$  et du ligand  $L$ . L'affinité est souvent exprimée par la constante de dissociation  $K_d$  en  $M$ .

$$K_a = [ML]/[M][L] = 1/K_d$$

D'un point de vue thermodynamique, la constante d'affinité  $K_a$  est fonction de l'énergie libre de l'interaction. La variation d'énergie libre  $\Delta G$  (énergie de Gibbs) d'interaction s'écrit de la manière suivante :

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

avec  $T$  : température (K),  $R$  constante des gaz parfaits ( $R=8,314 \text{ J}\cdot\text{mol}^{-1}\cdot\text{K}^{-1}$ ) et  $K_a$  : constante d'affinité en  $M^{-1}$ .

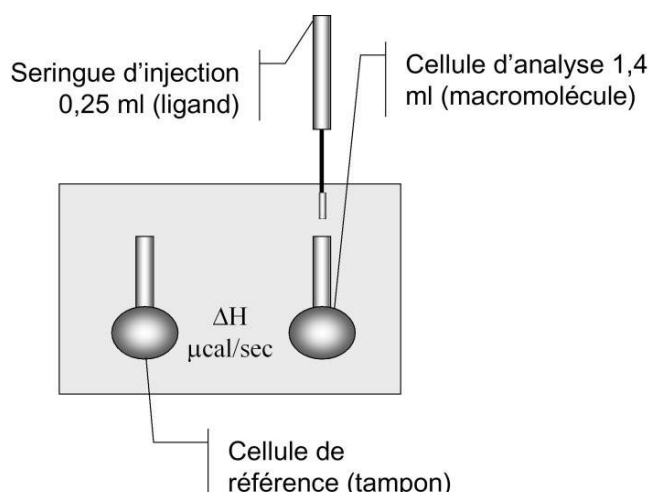
Lorsque  $\Delta G$  est négatif, l'interaction est spontanée et susceptible de fournir de l'énergie. Elle est également dite exergonique. Cette énergie est la résultante de deux contributions énergétiques : l'enthalpie de liaison  $\Delta H$  et l'entropie de liaison  $\Delta S$ .

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

L'enthalpie  $\Delta H$  reflète la chaleur absorbée ou dégagée lors de l'interaction. Les interactions exothermiques ont une valeur de  $\Delta H$  négative, ce qui est généralement le cas pour

les interactions protéine-sucre. Mais des interactions moléculaires spontanées endothermiques sont aussi observées. L'entropie représente le degré de désordre d'un système. Son augmentation est due aux mouvements Browniens et s'amplifie avec la température. La contribution entropique de liaison  $T\Delta S$  peut être positive (favorable à l'interaction) ou négative (défavorable). La limitation de la flexibilité des oligosaccharides lors de l'interaction avec une protéine résulte généralement en une forte barrière entropique qui fait baisser l'affinité (Imberty *et al.*, 2005).

Le microcalorimètre contient deux cellules isolées de l'environnement externe par une enceinte adiabatique qui est maintenue à température constante pendant le temps de la titration. Une cellule de référence contient du tampon uniquement. Une solution contenant la protéine est placée dans une deuxième cellule, dite d'analyse (Figure 35).



**Figure 35 : Schéma d'un calorimètre**

A intervalle régulier, des aliquotes de ligand (dans la même solution que la protéine) sont ajoutés à l'aide d'une seringue. A chaque injection, l'interaction entre les deux molécules s'accompagne d'un dégagement ou d'une absorption de chaleur qui entraîne une variation de température de la cellule de travail. Afin de rééquilibrer la température entre les deux cellules, l'appareil applique une puissance électrique. C'est ce signal qui est mesuré par le calorimètre. La puissance fournie est mesurée en fonction du temps, puis intégrée.

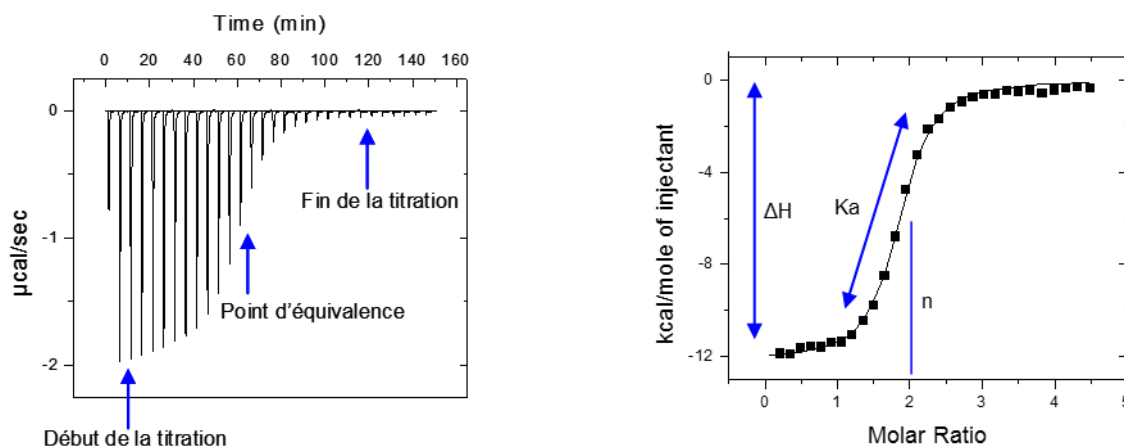
Au début de la titration, des quantités importantes de chaleur sont dégagées indiquant la formation du complexe protéine-ligand. En fin de réaction, tous les sites de liaison sont déjà occupés, et chaque nouvel ajout de ligand a comme seule conséquence sa dilution dans la cellule de travail. La forme générale de la courbe dépend des concentrations en

macromolécules et en ligands, du volume et du nombre d'injections ainsi que du temps de repos entre chaque injection. L'allure de la courbe est très importante pour la précision dans la détermination des paramètres thermodynamiques. Cette allure est définie par le paramètre de Wiseman. La valeur de  $c$  est calculée par l'équation :

$$c = n[M_0]K_a$$

où  $c$  le paramètre de Wiseman,  $n[M_0]$  la concentration initiale de sites de liaison en solution et  $K_a$  la constante d'association en  $M^{-1}$ .

Afin d'obtenir une allure sigmoïdale de la courbe qui permet de déterminer toutes les grandeurs thermodynamiques, le paramètre  $c$  doit être compris entre 10 et 100. Dans ces conditions, la variation d'enthalpie «  $\Delta H$  », la constante d'association «  $K_a$  » et la stœchiométrie «  $n$  » peuvent être déterminées avec un ajustement de la courbe expérimentale sur le modèle théorique approprié (Figure 36) (Wiseman *et al.*, 1989). Une valeur de  $c$  plus élevée entraîne une pente trop forte de la sigmoïde et un manque de précision dans la détermination de  $K_a$ . Une valeur trop faible nécessite de travailler en excès de ligand pour obtenir un plateau en fin de titration mais ne permet pas de déterminer la stœchiométrie.



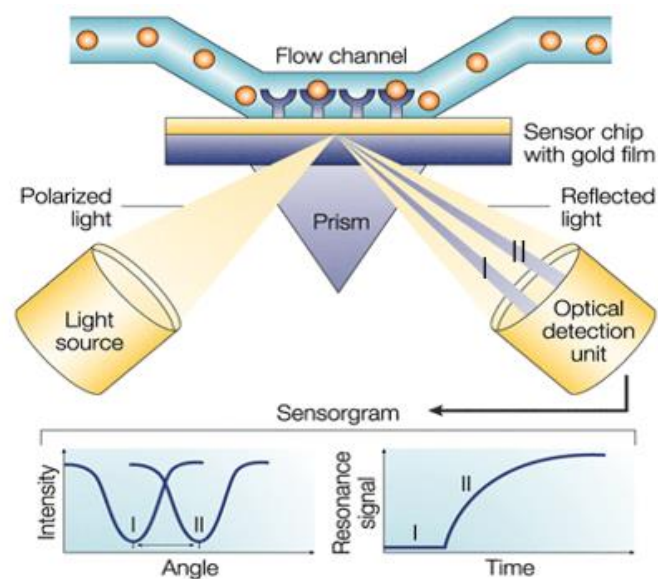
**Figure 36 : Données de l'ITC**

(A) Thermogramme expérimental : chaque pic correspond à une injection de ligand dans la solution de protéine. (B) Résultats d'ITC intégrés permettant la mesure de données thermodynamiques (constant d'association :  $K_a$ , enthalpie :  $\Delta H$ ) et de la stœchiométrie ( $n$ ) de l'interaction.



### V.2.1.2. Résonance Plasmonique de Surface

La résonance plasmonique de surface (SPR) permet de mesurer en temps réel toute interaction entre deux partenaires, ainsi que d'obtenir le  $K_d$  et les constantes de vitesse des réactions d'association/dissociation qui caractérisent l'interaction sans aucun marquage des protéines. Un capteur est composé d'une fine couche de verre recouverte d'un film d'or couplé à du dextran permettant d'immobiliser un ligand. Lorsqu'un faisceau de lumière monochromatique polarisé illumine la puce, la lumière réfléchie présente alors une chute d'intensité à un angle précis, l'angle de résonance. Les changements d'angle (de I à II dans le diagramme du bas à gauche) sont observés lors de la liaison de biomolécules à la surface et donc lors d'un changement de la masse de la couche de surface (Figure 37). Ces variations sont enregistrées sur un sensorgramme et quantifiées en unités de résonance (RU). 1000 RU correspondant à un déplacement de l'angle de  $0,1^\circ$  et sont équivalents à la fixation de 1 ng d'analyte par  $\text{mm}^2$ .

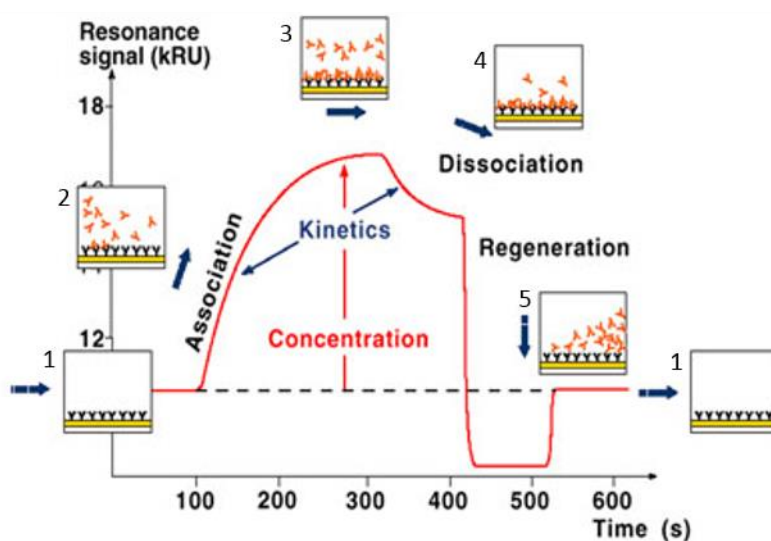


**Figure 37 : Principe de la résonance plasmonique de surface (SPR)**

*La technologie détecte la variation de l'angle la lumière réfléchi par le côté de la puce opposée du canal d'écoulement en fonction de la masse liée à la surface de la puce (Cooper, 2002).*

Les puces utilisées comportent au moins deux canaux distincts, le premier servant de référence (Fc1) et le second de test (Fc2). Les résultats montrés représentent les unités de

résonance (RU) après soustraction de la piste de référence (Fc2-Fc1). Le sensorgramme obtenu est composé de plusieurs phases : l'équilibration correspondant à la circulation du tampon, puis l'association de l'analyte qui se fixe sur le ligand jusqu'à la phase d'équilibre. Du tampon est ensuite réinjecté afin d'observer la phase de dissociation. Vient ensuite la phase de régénération dont le protocole varie d'une expérience à l'autre, mais qui a pour but d'éliminer tous les analytes sans détériorer le ligand. Un nouveau cycle peut alors commencer (Figure 38).



**Figure 38 : Description d'un sensorgramme de SPR**

Le sensorgramme obtenu est composé de plusieurs phases : l'équilibration (1) où du tampon est passe dans la puce, puis l'analyte est injecté : il y a alors association (2) où l'analyte se fixe sur le ligand jusqu'à la phase d'équilibre (3). Du tampon est ensuite réinjecté, on observe alors la phase de dissociation (4). Vient ensuite la phase de régénération (5) dont la composition varie d'une expérience à l'autre mais qui a pour but d'éliminer tous les analytes sans détériorer le ligand. Alors un nouveau cycle peut commencer. D'après <http://microgen.ouhsc.edu/>.

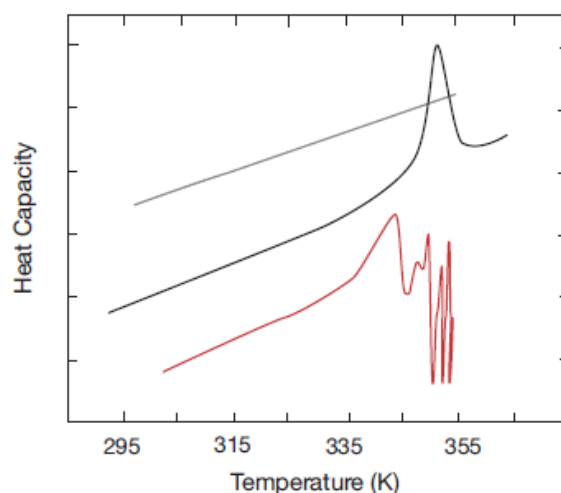
### V.2.1.3. Diffusion Dynamique de la Lumière

La technique de diffusion dynamique de la lumière (dynamic light scattering ou DLS) mesure des fluctuations d'intensité de diffusion d'un faisceau lumineux en fonction du temps qui apparaissent lorsque des particules sont soumises au mouvement Brownien. Plus une molécule est volumineuse, moins son déplacement en solution est rapide. Les particules en

mouvement vont interagir et disperser le faisceau dans toutes les directions. L'analyse de la lumière diffusée va être recueillie par un détecteur, et permet de déduire un coefficient de diffusion des particules qui est converti en distribution de taille puis en rayon hydrodynamique. Il est également possible d'en déduire le taux de polydispersité de la molécule qui représente une déviation standard de la taille de la particule reportée à son rayon. Cette valeur permet d'apprécier le taux d'agrégation et d'oligomérisation de la protéine.

#### V.2.1.4. Calorimétrie Différentielle à Balayage

La calorimétrie différentielle à balayage (Differential Scanning Calorimetry ou DSC) permet de suivre la dénaturation thermique d'une protéine. La température des échantillons varie de façon contrôlée au cours du temps, en appliquant un gradient de température discontinu aux deux cellules de mesures. Lorsque la température augmente les protéines se déstructurent. Le mécanisme de dénaturation est une réaction endothermique. L'appareil devra ainsi fournir davantage d'énergie à la cellule contenant la protéine par rapport à la cellule de référence pour maintenir la même température. La courbe obtenue correspond à la capacité calorifique ( $\text{KJ.mol}^{-1}.\text{K}^{-1}$ ) en fonction de la température. Le point le plus haut du pic correspond à la température de transition ( $T_m$ ) pour laquelle 50% des protéines sont dénaturées (Figure 39).



**Figure 39 : Données types de DSC**

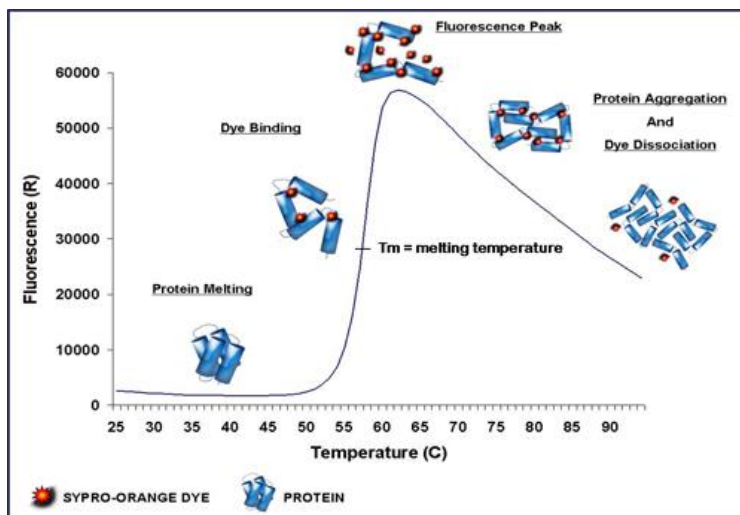
(Gris) ligne de base du tampon, (Rouge) profil type d'une protéine n'agrégeant pas après dénaturation, (Noir) profil type d'une protéine agrégeant après dénaturation. (D'après *Differential Scanning Calorimetry: Theory and Practice*, du Dr Christopher M. Johnson.).

Les contributions thermodynamiques peuvent être mesurées. Dans certains cas, les protéines déstructurées agrègent. Cette agrégation est caractérisée par des pics irréguliers visibles après le pic correspondant à la dénaturation (courbe rouge).

#### **V.2.1.5. Criblage de Thermostabilité par Fluorescence**

L'étude de la dénaturation thermique des protéines par fluorescence (Fluorescent-based thermal shift assay ou TSA) permet également de déterminer la température de transition ( $T_m$ ), mais en utilisant un appareil simplifié, de type PCR qui ne fournit pas d'information thermodynamique. Durant l'expérience de TSA, un gradient de température de 20 à 100°C à raison de 1 degré par minute est induit sur un échantillon protéique mélangé à un fluorophore. L'utilisation d'un composé capable de fluorescer uniquement dans un environnement hydrophobe va permettre d'obtenir une cinétique de dénaturation de la protéine, ainsi que de déterminer le  $T_m$  de la protéine. En effet, tant que la protéine est correctement repliée, les acides aminés hydrophobes ne sont pas accessibles au fluorophore. La fluorescence mesurée reste faible. Lorsque la température augmente, la dénaturation progressive de la protéine rend les acides aminés hydrophobes accessibles au fluorophore. Une augmentation de la fluorescence est alors observée jusqu'à l'obtention d'un pic (Figure 40). Le calcul de la dérivée de la mesure de la fluorescence en fonction de la température facilite la lecture du  $T_m$  de la protéine d'intérêt.

Les études de TSA peuvent être utilisées pour optimiser le tampon optimal de la protéine et les conditions de cristallisation d'une protéine (Phillips and de la Pena, 2011) mais également aider à choisir la température de cristallisation la plus adéquate (Dupeux *et al.*, 2011).



**Figure 40 : Principe du TSA**

*D'après Angonne National Laboratory.*

### V.2.1.6. Amarrage moléculaire

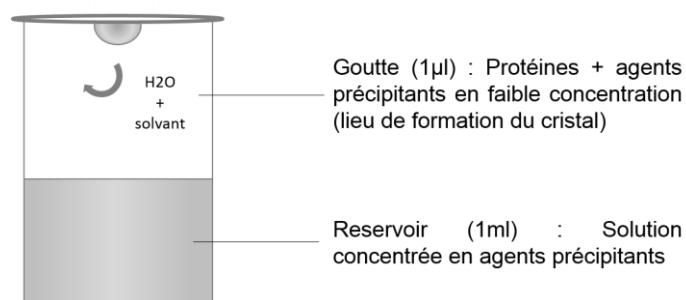
L'amarrage moléculaire ou docking consiste à modéliser les orientations (ou poses) qu'un ligand peut adopter dans le site de reconnaissance d'une protéine. Dans le cas du docking « flexible » différentes conformations possibles du ligand sont considérées. En général, la recherche se focalise sur le site actif tel qu'il aura pu être déterminé expérimentalement. Différents programmes de modélisation moléculaire peuvent être utilisés et nous avons sélectionné Glide (Schrödinger) qui avait donné de bons résultats dans un test protéine/sucre (Nurisso *et al.*, 2010). Généralement, plusieurs poses du ligand sont générées et classées par ordre de probabilité décroissante selon une fonction d'évaluation (Gohlke and Klebe, 2001) qui classera les poses par énergie d'interaction avec le site de la protéine cible.

## V.2.2. Caractérisation structurale

### V.2.2.1. Cristallogénèse

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

Si la plupart des substances minérales et des petites molécules organiques cristallisent facilement, l'obtention de cristaux de macromolécules biologiques est souvent plus délicate et constitue en soit le premier obstacle à franchir pour le bio-cristallographe. Les protéines sont en effet des structures biologiques fragiles dont la stabilité dépend fortement de l'environnement immédiat (solvant, pH, lumière, concentration etc) et deviennent vite précaire en dehors des conditions physiologiques habituelles. Les protéines présentent de surcroît une certaine flexibilité, notamment au niveau des chaînes latérales des acides aminés. Tous ces éléments rendent difficile une cristallogenèse qui impose généralement d'avoir une solution initiale pure et homogène, afin de conduire à un solide cristallin et non à un précipité amorphe. Afin d'obtenir des cristaux de protéines, une des méthodes fréquemment utilisée est « la diffusion de vapeur avec goutte pendante » (Figure 41). La technique consiste à mettre en présence, dans une enceinte hermétiquement close, une goutte de quelques microlitres d'une solution tamponnée contenant la protéine ainsi que des agents précipitants avec en contrebas, un réservoir d'un volume largement supérieur (0.2 à 1 ml) contenant une solution sans protéine et dont la concentration en agents précipitants est supérieure à celle établie dans la goutte. Avec le temps, les composés volatils (eau ou solvants organiques) vont diffuser dans l'enceinte par voie gazeuse entre les deux liquides jusqu'à ce que les concentrations – ou plutôt les potentiels chimiques – soient équivalents de part et d'autre. Cet équilibre thermodynamique et cinétique engendre du même coup la saturation en protéine qui alors s'agglomèrent spontanément, initiant ainsi la croissance de microcristaux dans la goutte. Toutefois, la méthode reste empirique et le succès aléatoire, tant les paramètres physico-chimiques intervenant sont nombreux. Concentrations des solutés, pH et température pour ne citer qu'eux, doivent être pris en compte lors de tests successifs tendant à préciser petit à petit l'éventail des conditions expérimentales requises. Des échantillonnages commerciaux tels que les « Screens Hampton Research » criblant de multiples conditions constituent souvent le point de départ des essais.



**Figure 41 : Description d'une goutte pendante**

Caractériser l'interaction protéine-ligand nécessite l'obtention de cristaux dans lesquels la macromolécule est complexée avec le ligand. Pour ce faire, protéine et ligand sont mélangés dans la goutte afin que tous deux interagissent et cristallisent ensemble : on parle alors de co- cristallisation. Selon les ligands mis en présence, le complexe protéine-ligand ne cristallise pas toujours dans les mêmes conditions expérimentales, ni dans le même groupe d'espace.

#### **V.2.2.2. Résolution d'une structure cristallographique**

Les données de diffraction sont collectées dans un faisceau de rayons X, dans notre cas au synchrotron ESRF (European Synchrotron Radiation Facility, Grenoble). Lors de la collecte, le rayonnement peut occasionner la formation de radicaux libres à l'intérieur du cristal et endommager les molécules. L'intensité du faisceau peut également causer un échauffement local du cristal provoquant sa destruction complète. Afin d'éviter ces dommages, les cristaux sont congelés dans l'azote liquide après trempage dans une solution cryoprotectante qui évite la formation de glace autour des cristaux.

La collecte de données est effectuée sous un flux d'azote continu. Pendant la collecte, le cristal tourne par paliers autour d'un axe perpendiculaire au faisceau de rayons X. Les rayons diffractés sont collectés par des détecteurs de type CCD (couple charge device) et ceci génère des clichés de diffraction. Le protocole de traitement des clichés de diffraction enregistrés lors de la rotation du cristal comporte quatre étapes essentielles : l'indexation, l'intégration des intensités, la mise à l'échelle et la moyennisation.

L'indexation consiste, à l'aide de la position des taches de diffraction couplée à l'angle d'oscillation, à déterminer l'orientation du cristal ainsi que les paramètres de maille. Les programmes d'indexation cherchent trois vecteurs de base permettant de relier toutes les réflexions, puis la matrice d'orientation ainsi que la maille qui explique toutes les réflexions observées. Il est alors possible de prédire la position de l'ensemble des taches de diffraction sur les images et d'en mesurer l'intensité.

L'intégration des intensités consiste à établir des profils moyens des taches et de calculer l'intensité de chaque réflexion par intégration numérique suivant les profils calculés. Cette étape permet de corriger l'étalement des taches sur un diamètre de quelques millimètres, mais également de corriger les effets d'absorption et de polarisation du faisceau incident. L'ensemble des images intégrées est ensuite mise à l'échelle, entre autre afin de mettre au

même niveau l'intensité des différentes réflexions collectées sur différentes images de diffraction. Les réflexions équivalentes sont comparées pour mettre en évidence la présence de réflexions mal mesurées. Finalement, une intensité moyenne est calculée pour chaque réflexion unique, donnant ainsi accès aux modules des facteurs de structures nécessaires pour calculer la carte de densité électronique par la transformée de Fourier.

La qualité d'un jeu de données est déterminée par différents critères statistiques calculés au cours du traitement. Les paramètres les plus significatifs et utilisés sont :

- la résolution : qui déterminera la finesse des détails dans l'image de la densité électronique.

- le rapport signal sur bruit ( $I/\sigma$ ) moyen : plus il est grand et meilleure est l'estimation des facteurs de structure.

- la complétude : qui mesure le rapport entre le nombre de réflexions uniques mesurées et le nombre théorique de ces réflexions pour la résolution des données. Elle doit être la plus grande possible et supérieure à 90 %.

- la multiplicité : qui correspond au nombre de fois où l'intensité d'une réflexion unique est mesurée. Ce qui est équivalent au nombre moyen de réflexions symétriques mesurées. Plus la valeur numérique de ce critère est grande, meilleure sera l'estimation de l'intensité moyenne d'une réflexion unique.

- le facteur  $R_{sym}$  : (ou  $R_{merge}$ ) qui représente le désaccord entre les intensités des réflexions équivalentes par symétrie. Plus ce facteur est bas, meilleure est la cohérence du jeu de données.

Pour calculer la carte de densité électronique, il est nécessaire de connaître non seulement le module des facteurs de structures (intensité au carré) mais aussi leur phase. Le problème du phasage peut être abordé de différentes façons. Dans la méthode de Remplacement Moléculaire, on utilise comme phases de départ les phases provenant d'une autre structure présentant une similarité structurale suffisante. Un minimum de 20 % d'identité de séquence est requis mais souvent insuffisant. En pratique, le remplacement moléculaire procède de la façon suivante. A l'aide de la fonction de Patterson, des logiciels cherchent d'abord l'orientation de la molécule, puis des translations sont effectuées afin de trouver la position de la molécule dans la maille. Compte tenu du nombre de paramètres à considérer (trois pour les rotations autour des axes centrés sur la molécule (x, y, z) et trois pour les translations dans la maille), le processus en deux étapes est beaucoup plus rapide qu'une recherche simultanée en 6 dimensions.



Grâce aux phases obtenues, une première carte de densité électronique est calculée et visualisée à l'aide du logiciel graphique COOT (Emsley et al., 2010). Cette première carte contient des erreurs et n'est pas toujours de qualité suffisante pour déterminer la position exacte des atomes. La procédure d'affinement permet de corriger le modèle initial afin d'obtenir un accord optimal entre les modules des facteurs de structures observés et ceux calculés à partir du modèle. En pratique, on procède en alternant des étapes de reconstruction manuelle du modèle initial avec des étapes d'affinement numérique des paramètres du modèle (position et agitation de chaque atome).

Les étapes de reconstruction manuelle du modèle ont été effectuées dans COOT. Les cartes de densité utilisées sont essentiellement des cartes dites différences ou cartes résiduelles  $F_o - F_c$ . Elles illustrent le désaccord entre le modèle et les facteurs de structures observés. Un pic positif s'interprète comme des atomes manquant dans le modèle. Un pic négatif sur un atome du modèle indique que celui-ci est sans doute mal placé. A la fin de l'affinement, cette carte résiduelle doit contenir le moins possible de pics négatifs ou positifs. La carte  $2F_o - F_c$  représente une combinaison linéaire de la carte de densité électronique et de la carte résiduelle. Elle permet de visualiser la densité électronique, tout en tenant compte des imperfections du modèle.

Les étapes d'affinement ont été réalisées avec le programme REFMAC (CCP4) (1994). Ce programme permet d'affiner les structures par des méthodes de minimisation des coordonnées atomiques du modèle, en tenant compte des contraintes structurales (liaisons atomiques, angles de liaison...) et en optimisant l'accord entre facteurs de structures calculés et observés. L'affinement permet donc de maximiser une fonction de vraisemblances contenant les paramètres à affiner décrits par des distributions de probabilité. Durant les différentes étapes d'affinement, la correspondance entre le modèle et les données expérimentales est mesurée par le facteur d'accord cristallographique R. Le facteur R fait partie de la fonction à minimiser. Plus sa valeur est basse, plus le modèle est en accord avec les données. Ce facteur n'étant pas indépendant du processus, un second facteur est utilisé, le facteur  $R_{free}$ . Il est calculé à partir de 5 % de réflexions qui ne sont pas affinées. Une diminution du facteur R sans diminution corrélée du facteur  $R_{free}$  indique un affinement inefficace qui consiste à faire entrer artificiellement dans le modèle le bruit des cartes de densités électroniques. La structure est ensuite déposée validée et dans la base de données internationale RCSB Protein Data Bank.





---

# **RESULTATS**

---



---

# RESULTATS

---

## I. Contrôle de la valence

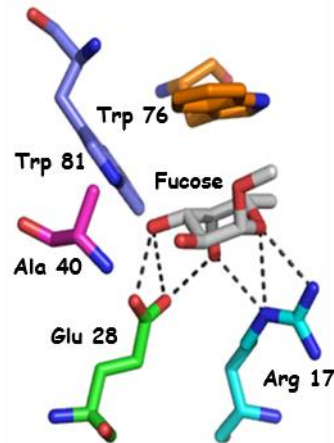
### I.1. Modification de la valence de la protéine RSL

#### I.1.1. Introduction

La multivalence joue un rôle majeur dans la forte avidité des lectines pour les surfaces cellulaires glycosylées. De plus, la multivalence est responsable de la capacité des lectines à affecter la dynamique membranaire en regroupant les glycosphingolipides. La multivalence serait donc un des facteurs clés de la formation de tubules membranaires induite par les lectines.

Les lectines de valence modifiée ont été conçues à partir de la lectine RSL de *Ralstonia solanacearum* qui présente une structure en  $\beta$ -propeller très symétrique. La lectine RSL est un homotrimère qui présente six sites de liaison pour le fucose, c'est-à-dire deux par monomère. On distingue donc les sites intramonomériques qui sont à l'intérieur d'un monomère des sites intermonomériques qui sont situés à la jonction entre les monomères.

Les principaux acides aminés impliqués dans la stabilisation du fucose ont été déterminés par étude de la structure cristallographique et confirmés par des calculs de dynamique moléculaire (Figure 42). Deux mutants de valence réduite ont été produits : il s'agit du mutant R17A où l'arginine 17 a été substitué par une alanine, et le mutant W76A où le tryptophane 76 a été substitué par une alanine. Du fait que la protéine RSL soit un homotrimère, une mutation au niveau de l'ADN aura pour conséquence la mutation de trois sites au niveau de la protéine.



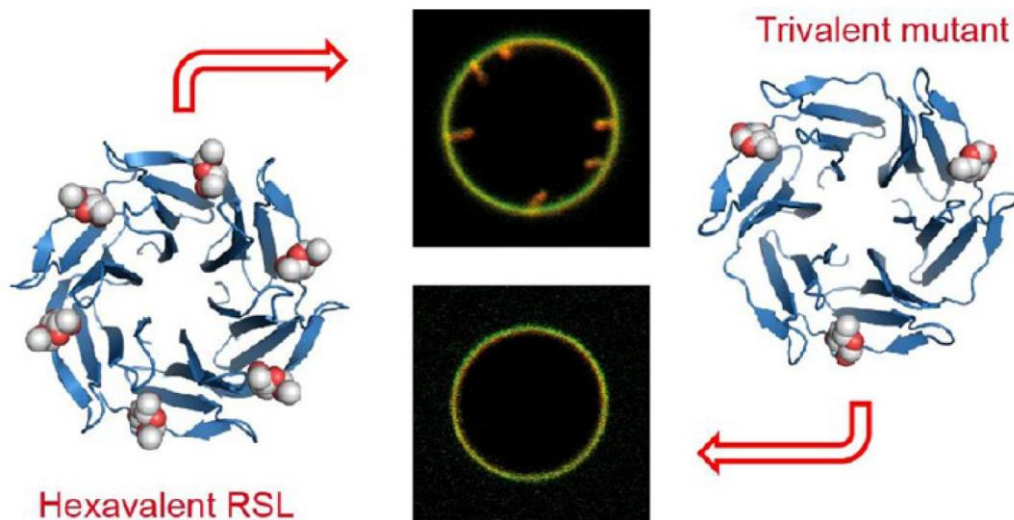
**Figure 42 : Acides aminés de la protéine RSL principalement impliqués dans la fixation du fucose dans le site intramonomérique.**

Des expériences d'ITC ont confirmé la perte de la capacité de liaison de trois des sites de haute affinité pour les deux mutants. En revanche, la structure cristallographique a indiqué la présence d'une faible affinité de liaison résiduelle dans le mutant W76A mais pas dans le mutant R17A.

Le mutant W76A qui présente trois sites de haute affinité et trois sites de basse affinité pour le fucose, démontre une avidité inchangée pour les surfaces fucosylées par rapport à la protéine sauvage RSL hexavalente. Le mutant W76A n'est pas non plus affecté pour sa capacité à former des invaginations dans les membranes contenant des glycolipides fucosylés.

Le mutant trivalent R17A présente la même avidité pour une surface fucosylée que la protéine sauvage RSL. Cependant, R17A ne possède plus la capacité d'induire la formation d'invaginations de la membrane des vésicules géantes unilamellaires, indiquant le rôle crucial du nombre de sites pour le regroupement des glycolipides (Figure 43). Dans la lignée de cellules épithéliales de poumon humain H1299, wt-RSL est internalisée en quelques secondes alors que la cinétique de l'absorption R17A est largement retardée.

Les néolectines de valence modulables sont un outil prometteur pour étudier la dynamique des membranes. Malheureusement en l'état, la valence de RSL ne peut qu'être que de 6 ou 3 avec une seule topologie.



*Figure 43 : Une diminution de la valence de la protéine RSL induit un changement drastique dans la capacité de la lectine de former des invaginations.*

*(Arnaud et al., 2013).*

### **I.1.2. La réduction de la valence d'une lectine changes drastiquement la dynamique des membranes mais pas d'avidité de surface**

La répartition du travail de l'article a été la suivante. La modélisation moléculaire a été entièrement réalisée par le Dr. Aline Thomas. La production et la caractérisation des protéines par ITC, SPR, TSA et la production des cristaux ont été réalisées par mes soins avec une participation du Dr. Aymeric Audfray pour la SPR. La caractérisation par DLS a été réalisée par le Dr. Marie Trovaslet. J'ai résolu la structure de la protéine R17A avec l'aide du Dr. Annabelle Varrot. Le Dr. Aymeric Audfray a résolu la structure de la protéine W76A avec également l'aide du Dr. Annabelle Varrot. La caractérisation de l'interaction des protéines avec des GUVs ou des cellules a été faite par le Dr. Julie Claudinon et Kevin Trondle dans le laboratoire du Prof. Winfried Römer en utilisant les glycolipides fournis par le Prof Göran Larson.



## Reduction of Lectin Valency Drastically Changes Glycolipid Dynamics in Membranes but Not Surface Avidity

Julie Arnaud,<sup>†</sup> Julie Claudinon,<sup>‡,§</sup> Kevin Tröndle,<sup>‡,§</sup> Marie Trovaslet,<sup>||</sup> Göran Larson,<sup>⊥</sup> Aline Thomas,<sup>†</sup> Annabelle Varrot,<sup>†</sup> Winfried Römer,<sup>\*,‡,§</sup> Anne Imberty,<sup>\*,†</sup> and Aymeric Audfray<sup>†</sup>

<sup>†</sup>CERMAV-CNRS (affiliated to Grenoble Université and ICMG), BP53, 38041 Grenoble, France

<sup>‡</sup>Institute of Biology II, Schänzlestraße 1, Albert-Ludwigs-University Freiburg, 79104 Freiburg, Germany

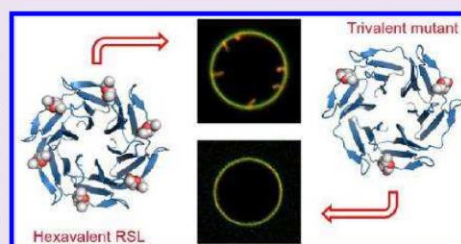
<sup>§</sup>BIOSS—Centre for Biological Signalling Studies, Schänzlestraße 18, Albert-Ludwigs-University Freiburg, 79104 Freiburg, Germany

<sup>||</sup>Département de Toxicologie, IRBA, 24 av des Maquis du Grésivaudan, 38700 La Tronche, France

<sup>⊥</sup>Department of Clinical Chemistry and Transfusion Medicine, Sahlgrenska University Hospital, 41345 Gothenburg, Sweden

### Supporting Information

**ABSTRACT:** Multivalency is proposed to play a role in the strong avidity of lectins for glycosylated cell surfaces and also in their ability to affect membrane dynamics by clustering glycosphingolipids. Lectins with modified valency were designed from the  $\beta$ -propeller fold of *Ralstonia solanacearum* lectin (RSL) that presents six fucose binding sites. After identification of key amino acids by molecular dynamics calculations, two mutants with reduced valency were produced. Isothermal titration calorimetry confirmed the loss of three high affinity binding sites for both mutants. Crystal structures indicated that residual low affinity binding occurred in W76A but not in R17A. The trivalent R17A mutant presented unchanged avidity toward fucosylated surfaces, when compared to hexavalent RSL. However, R17A is not able anymore to induce formation of membrane invaginations on giant unilamellar vesicles, indicating the crucial role of number of binding sites for clustering of glycolipids. In the human lung epithelial cell line H1299, wt-RSL is internalized within seconds whereas the kinetics of R17A uptake is largely delayed. Neolectins with tailored valency are promising tools to study membrane dynamics.



Lectins are sugar-binding proteins that interact with carbohydrates noncovalently and specifically. They possess the ability to act as recognition molecules on cell surfaces, inside cells and in biological fluids. Thus, lectins are unique cell-recognition markers for carbohydrates. While sugars are able to carry biological information, lectins decipher the complex glycode.<sup>1,2</sup> Lectins display a large number of folds and a variety of binding site architectures. They are usually multivalent, either by oligomerization, tandem repetition, or surface presentation.<sup>3</sup> Multivalency is therefore a key component of lectin-carbohydrate interactions in many biological processes such as aggregation and cell surface attachment. Because receptors and ligands are displayed in several copies on the external side of viruses, bacteria, and cells, multivalency creates avidity and is able to enhance usually low unique site affinity by several orders of magnitude.<sup>4</sup>

Many pathogens use lectins for multivalent attachment on oligosaccharide epitopes on human epithelia.<sup>5</sup> Cholera and Shiga toxins exploit plasma membrane glycosphingolipids as receptors to enter and infect cells via endocytic routes that do not involve the elaborate network of clathrin.<sup>6–8</sup> These lectins can induce plasma membrane invaginations without the help of the cytosolic protein machinery via the dynamic construction of protein–lipid nanodomains whose intrinsic properties lead to

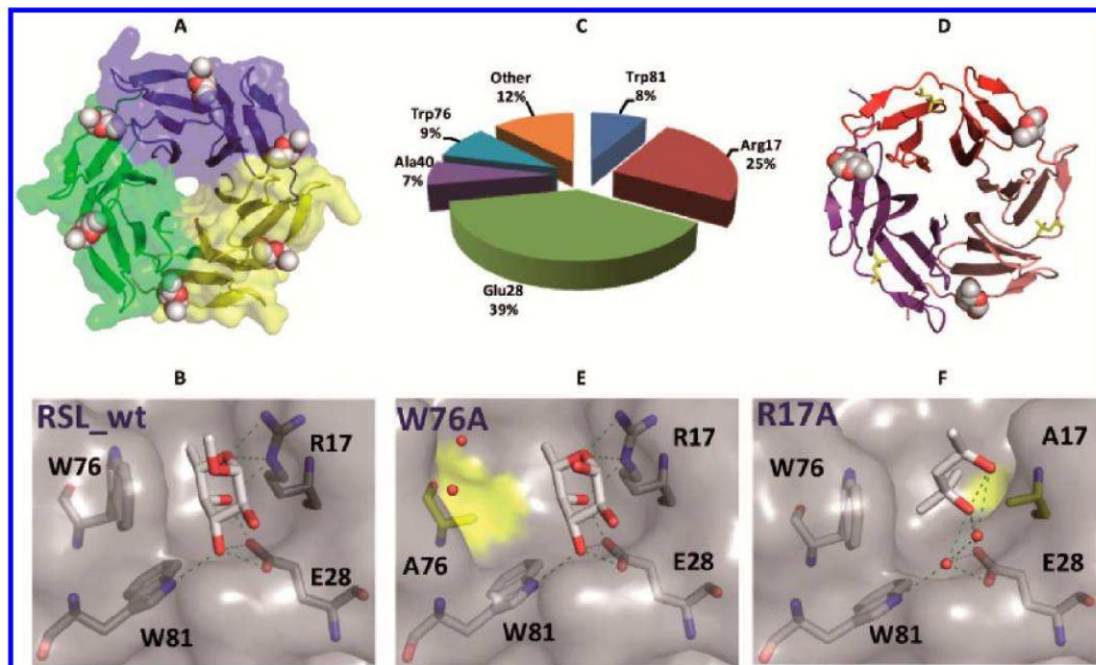
membrane bending.<sup>9–11</sup> However, the role of glycolipid clustering in the internalization process is still unclear: the reduction of receptor binding sites did not inhibit the cellular entry of cholera toxin<sup>12</sup> and Shiga toxin<sup>10</sup> but reduced enormously the formation of membrane invaginations on cells and giant unilamellar vesicles for this latter protein. There is therefore a need for tailored lectin mutants to analyze the effect of multivalency upon cellular internalization of pathogens.

The modular architecture of the  $\beta$ -propeller fold is well adapted for multivalent presentation and, therefore, for the study of the effect of valency. It consists of the repeat of small  $\beta$ -sheet domains referred to as “blade” arranged circularly.  $\beta$ -Propellers of animal and fungal lectins are formed by the folding of one polypeptide chain with 5, 6, or 7 blades.<sup>13–15</sup> A different  $\beta$ -propeller was identified for bacterial lectins from *Ralstonia solanacearum* (RSL)<sup>16</sup> and *Burkholderia ambifaria* (BambL).<sup>17</sup> They present a pseudo-6-fold symmetry created by the trimerization of small peptides, each consisting of two blades (Figure 1A). This  $\beta$ -propeller has several advantages compared to the fungal or animal ones: it can be produced in

Received: February 8, 2013

Accepted: July 15, 2013

Published: July 15, 2013



**Figure 1.** (A) Overall structure of hexavalent trimeric RSL complexed with  $\alpha$ MeFuc displayed as spheres (PDB code 2BT9). (B) Intramonomeric fucose binding site. (C) Energy contribution for the interaction between fucoside and binding site residues (dist <8 Å), averaged over time. (D) Overall structure of R17A complexed with  $\alpha$ MeFuc displayed as spheres and MPD as yellow sticks. (E) Residual binding of fucose in W76A mutated site. (F) Binding of MPD in R17A mutated site.

large quantity in recombinant form, it does not require a “Velcro” closure for its stabilization, and it displays high affinity for fucose ( $K_d$  in the micromolar range) and fucosylated oligosaccharides.

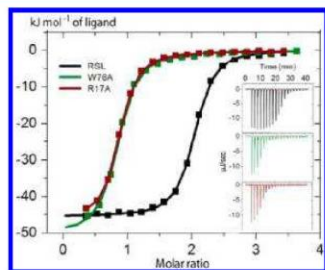
Each RSL monomer presents two fucose binding sites: the “intramonomeric” one between the two blades of the monomer, and the “intermonomeric” one at the interface with its neighbor (Figure 1A). The two binding sites are very similar, but the intramonomeric one was selected for engineering since its modification bears less risk to destabilize the trimeric fold. Analysis of the crystal structure of RSL/Fuc complexes indicates the occurrence of hydrogen bonds with charged amino acids (R17, E28) and of hydrophobic contacts with aromatic amino acids (W76, W81) (Figure 1B). In order to quantify the participation of each amino acid to the binding energy, a 32 ns molecular dynamics trajectory of RSL interacting with  $\alpha$ -methyl-fucoside ( $\alpha$ MeFuc) was calculated (Supporting Information (SI) Figure S1). Averaging of the interaction energy over the three monomers indicates a major role for the charged amino acids and the tryptophan residues (Figure 1C).

In order to suppress the fucose binding activity of the intramonomeric site, mutants E28A, R17A, W76A, and R17A/W76A were created (SI Table S1). They were expressed in *Escherichia coli*, but some resulted in the production of insoluble proteins that could not be refolded. Only the mutants R17A and W76A could be obtained in soluble form and purified. When analyzed by dynamic light scattering (DLS), wt-RSL and the mutants displayed molecular weights of  $26.7 \pm 8$  kDa in agreement with the trimeric state (i.e., one  $\beta$ -propeller

in solution). The stability of wt-RSL, R17A, and W76A was tested by differential scanning calorimetry (SI Figure S2 and Table S2). Wt-RSL displays an unusually strong thermostability with a  $T_m$  of 86 °C that increases to 96 °C in the presence of  $\alpha$ MeFuc. The mutants are slightly less stable with a  $T_m$  of 81 °C that increases to 86/87 °C in the presence of fucoside. This analysis confirms that the mutants are well folded, with the strong stability of  $\beta$ -propellers,<sup>18</sup> and that fucose is still binding to the nonmutated sites.

In order to check the stoichiometry of the interaction, as well as to analyze the thermodynamic characteristics of the interaction, isothermal titration calorimetry (ITC) experiments were performed by titrating wt-RSL and mutant lectins by  $\alpha$ MeFuc in the same conditions. The height and the shape of integrated curves, related to enthalpy and affinity, do not differ significantly between wt-RSL and the mutants (Figure 2). Quantitative analysis indicates that RSL ( $K_d = 0.87 \mu\text{M}$ ) has slightly stronger affinity for  $\alpha$ MeFuc than the two mutants ( $K_d = 2.6$  and  $2.2 \mu\text{M}$  for W76A and R17A, respectively), however, with values in the same order of magnitude. Similarly, no significant difference is observed in the thermodynamic parameters with binding enthalpies ranging between  $-45$  to  $-50$  kJ/mol (SI Table S3). On the opposite, the inflection point occurs at different molar ratio:  $\alpha$ MeFuc binds wt-RSL with the expected stoichiometry of 2:1 ratio, while both mutants display a 1:1 ratio (Figure 2). ITC data therefore confirmed that both mutants lost the ability to bind fucose in one binding site. However, if some residual binding occurs in the mM range, it would not be detected by this method.



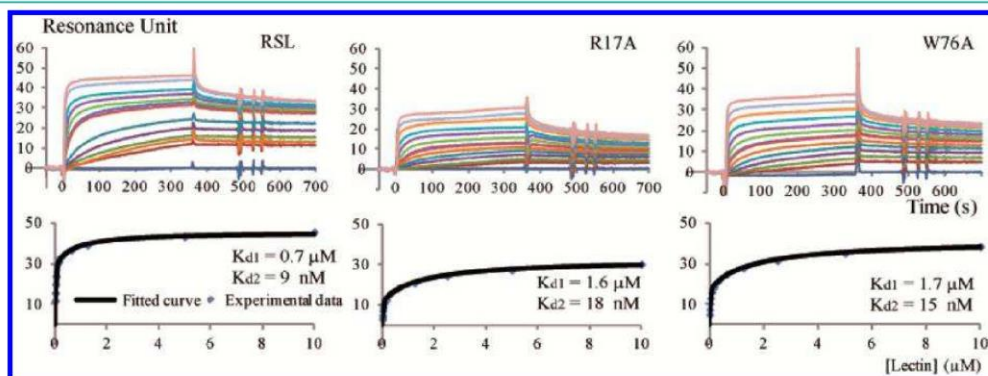


**Figure 2.** Integrated heat of interaction for titration of RSL and mutants (0.09 mM) with  $\alpha$ MeFuc (1.5 mM) measured on ITC200 at 25 °C in 20 mM Tris with 0.1 M NaCl at pH 7.5. Original thermograms are displayed as inserts.

Further validation of the mutants was therefore performed by crystallography. The crystal structures of both W76A and R17A (Figure 1D) were solved in complex with fucose (1.54 Å resolution) and  $\alpha$ MeFuc fucose (1.50 Å resolution), respectively. Statistics for data collection and structure refinement are listed in SI Table S4. R17A crystallized in space group  $P32_1$  with one monomer per asymmetric unit, the trimer being generated by the 3-fold axis. On the opposite, the whole trimer is present in the asymmetric unit of W76A (space group  $I4_122$ ). This structure presents an unusual swapping with exchange of the last  $\beta$ -strand of chain C between symmetrically related  $\beta$ -propellers (SI Figure S3). This leads to the formation of a hexamer that is probably a crystallization artifact since W76A is trimeric in diluted solution. As expected, the R17A mutant contains fucose residues only in the intermonomeric sites. The mutated intramonomeric site is not able to bind fucoside anymore, but a 2-methyl-2,4-pentanediol (MPD) molecule is observed instead (Figure 1F). On the opposite, W76A partly retains the affinity for sugars, and density for a fucose residue could be observed in both mutated and nonmutated binding sites. In the mutated site, while the stacking interaction with W76 is missing, the fucose is still stabilized thanks to hydrogen bonds with the charged amino acids R17 and E28 (Figure 1E). Nevertheless, the affinity should be very weak, since it was not detected in the ITC experiment. R17A and W76A therefore appear to have different

abilities to bind fucose in the mutated intramonomeric binding site, in agreement with the role of amino acids in MD calculations. Mutating R17A resulted in a true trivalent  $\beta$ -propeller with three high affinity sites for fucoside, while W76A is an intermediate form with three high affinity sites and three low affinity ones. The residual affinity of the mutated site was not detected by ITC, and it is therefore estimated to be not better than millimolar.

The binding of wt-RSL and its mutants to a fucosylated surface was analyzed by surface plasmon resonance (SPR). Constructions of low fucose density chips based on polymeric Fuc-PAA-biotin and monomeric Fuc-EG3-Triaz-biotin (SI Figure S4) are described in SI together with resulting data (SI Table S5 and S6). Both chips gave comparable data, and only those from Fuc-PAA surface are described below. Sensorgrams with fast association and slow dissociation phases are observed and indicate a very stable binding on the fucosylated chip (Figure 3). Due to complex interaction and multivalency, it was not possible to quantify  $k_{on}$  and  $k_{off}$ , but qualitative analysis of the curves indicated that the mutants have faster off-rates than wt-RSL, maybe resulting from reduced rebinding capacity. Dissociation constants were measured using the equilibrium response methods with a two-site model recently proposed to determine a “local”  $K_d$  corresponding to the binding site, and a “global”  $K_d$  corresponding to avidity to surface.<sup>19</sup> Resulting fitted curves and dissociation constants are reported in Figure 3, and additional data are summarized in SI Table S5. For each lectin, we observed a dissociation constant ( $K_{d1}$ ) in the micromolar range, that corresponds to the binding of one site of the lectin to the chip and a second dissociation constant ( $K_{d2}$ ) in the nanomolar range, that corresponds to the multivalent binding of the lectin to several fucose residues on the chip. As displayed in SI Table S5, only the  $R_{max2}$  value, which reflects the amount of lectin bound multivalently was slightly affected by the mutations (from 150 RU for wt-RSL to 50 RU for R17A). The values of  $K_{d1}$  do not vary much (between 0.7 to 1.7  $\mu$ M) and are in excellent agreement with the ones obtained by ITC. Interestingly, the affinity and avidity of the mutated lectins have been almost not affected. Local affinity for fucose is only slightly diminished, as observed in ITC, and more surprisingly, the avidity of the trivalent mutants is in the same order of magnitude than the hexavalent  $\beta$ -



**Figure 3.** Surface plasmon resonance results for the binding of RSL and mutants to a fucosylated-PAA chip: (Top) Sensorgrams of binding for concentrations of proteins varying from 10  $\mu$ M to 5 nM by 2-fold dilution from lowest to highest curve. (Bottom) Steady state analysis of responses at the end of association phase using a two-site binding model.

propeller (wt-RSL). Hence, it could be concluded that in RSL, three binding sites are sufficient for an efficient multivalency effect.

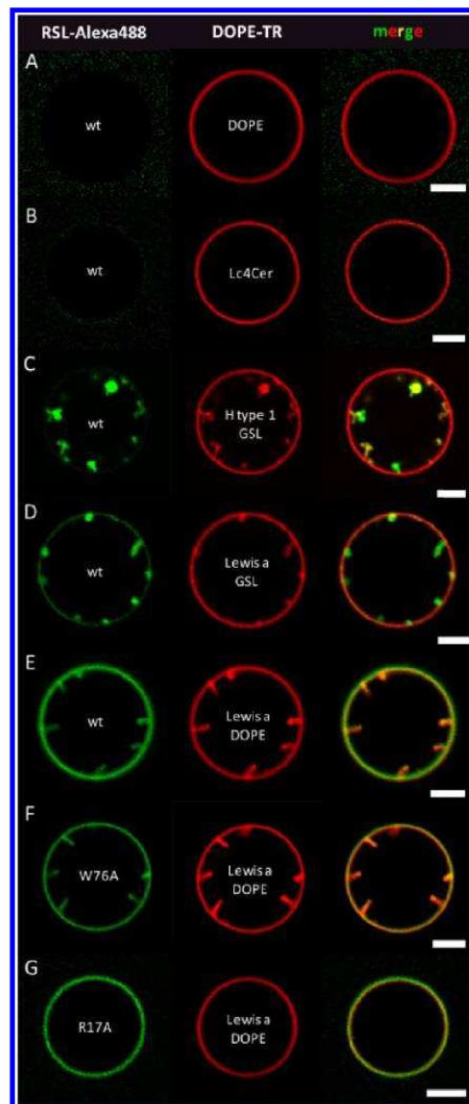
The binding of wt-RSL and its mutants to sugar epitopes on human histo-blood group antigens was tested in a membranous environment by using glycolipid-containing giant unilamellar vesicles (GUVs). GUVs with a typical diameter between 10 and 30  $\mu\text{m}$  were observed at RT by fluorescence microscopy upon incubation with Alexa488-labeled RSL (Figure 4). Our experiments on GUVs clearly indicated that fucose is an absolute prerequisite for the binding of RSL. Wt-RSL (green) did not bind to liposomes containing the phospholipid dioleoyl-phosphatidylethanolamine (DOPE) (Figure 4A) not to ones that contain lactotetraosylceramide (Lc4Cer), the nonfucosylated precursor of blood group antigens (Figure 4B).

Wt-RSL bound to all tested glycosphingolipids (GSL) of Lewis type 1 and ABO blood group (H type 1, Lewis a, Lewis b, B type 1, A type 1, and A Lewis b). Furthermore, wt-RSL induced membrane invaginations on type 1 GSLs with up to several micrometers in length (Figure 4C–D). As depicted in Figure 4C for H type 1-GSL and in Figure 4D for Lewis a-GSL, wt-RSL (green color) was highly enriched within these membrane invaginations while Texas Red DHPE (red color), a fluorescent lipid that indicates the  $\text{l}_\alpha$ -phase, was mainly equally distributed within the membrane. The dynamics of membrane invaginations induced by wt-RSL on Lewis a-GSL containing liposomes can be visualized in SI Movie 1. Wt-RSL also bound to a synthetic glycolipid obtained by linking the Lewis a-trisaccharide to the phospholipid DOPE and induced membrane invaginations (Figure 4E and SI Movie 2). The change from a ceramide to a phospholipid backbone did not affect lectin binding neither the formation of membrane invaginations but influenced the localization of wt-RSL. In most cases, the lectin is more concentrated in invaginations for glycosphingolipid containing GUVs than for the synthetic DOPE-phospholipid-containing ones, where it seems to be equally distributed.

Using Lewis a-DOPE containing GUVs, wt-RSL bound to 100% of liposomes and induced highly dynamic membrane invaginations in 93% of the cases ( $n = 280$ ). On the same type of liposomes, W76A bound to 100% of liposomes and formed membrane invaginations in 88% of the cases ( $n = 25$ ) (Figure 4F and SI Movie 3). R17A also bound nicely to all liposomes but did not at all induce membrane invaginations (0%,  $n = 280$ ) (Figure 4G and SI Movie 4). The quantification of the length of the invaginations led to average lengths of  $1.77 \pm 0.75 \mu\text{m}$  and  $1.87 \pm 0.77 \mu\text{m}$  for wt-RSL and W76A, respectively (SI Figure S5).

*In cellulo* experiments were conducted using the human lung epithelial cell line H1299 and Cy3-labeled lectins (Figure 5). Wt-RSL was internalized within a few seconds at 37  $^\circ\text{C}$ . After one hour, the majority of wt-RSL was concentrated in the perinuclear region, with some residual lectin binding at the cell surface. In contrast, the internalization of R17A became apparent first after nearly 1 h of incubation at 37  $^\circ\text{C}$ , with a rather homogeneous subcellular distribution. The uptake kinetics of trivalent R17A is therefore strongly retarded compared to hexavalent lectin.

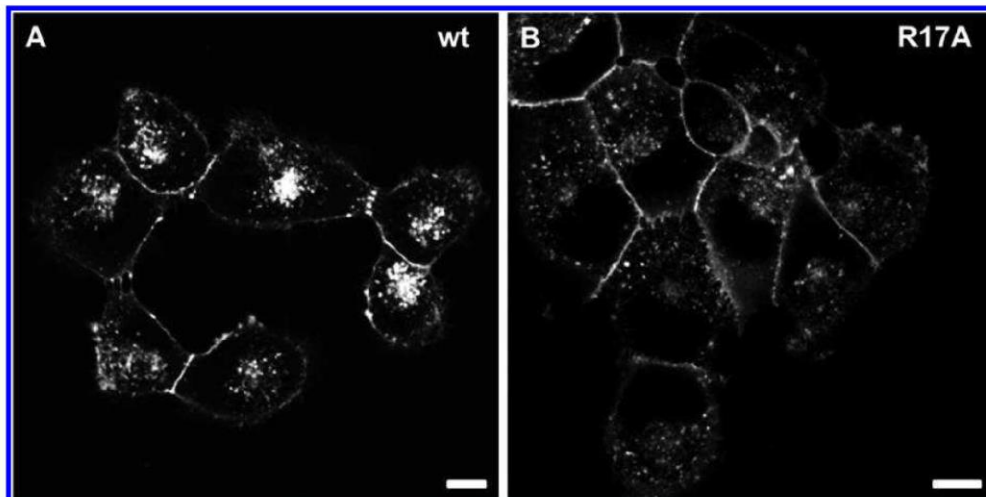
The rational mutations of lectin binding sites presented here resulted in the production of two valency mutants with different properties. Compared to wt-RSL, which presents six high-affinity fucose binding sites per  $\beta$ -propeller, R17A presents only three high affinity binding sites and W76A is an



**Figure 4.** Confocal sections of equatorial plane of GUVs containing various natural or synthetic glycolipids in presence of Alexa488-labeled wt-RSL and mutants. (A) DOPE, (B) Lactotetraosyl-ceramide (Lc4Cer:  $\text{Gal}\beta 1\text{-}3\text{GlcNAc}\beta 1\text{-}3\text{Gal}\beta 1\text{-}4\text{Glc}\beta 1\text{-Cer}$ ), (C) H type 1-GSL ( $\text{Fuc}\alpha 1\text{-}2\text{Gal}\beta 1\text{-}3\text{GlcNAc}\beta 1\text{-}3\text{LacCer}$ ), (D) Lewis a-GSL ( $\text{Gal}\beta 1\text{-}3[\text{Fuc}\alpha 1\text{-}4]\text{GlcNAc}\beta 1\text{-LacCer}$ ), (E–G) Lewis a-DOPE. Alexa488-labeled RSL (A–E) and mutants W76A (F) and R17A (G) are shown in green color. All acquisitions were performed with the same or slightly adapted (B–D) excitation and emission gains. Scale bars: 5  $\mu\text{m}$ .

intermediate form with three high and three low affinity sites. While all proteins have comparable affinity for fucose and avidity for fucosylated surface, only R17A lost the ability to induce membrane invaginations on GUVs and exhibited delayed internalization into cells.





**Figure 5.** Incubation of H1299 human lung epithelial cells with either (A) wt-RSL-Cy3 or (B) R17A-Cy3 for 1 h at 37 °C (lectin concentration: 2  $\mu\text{g mL}^{-1}$ ). Fluorescence intensity inside cells ( $n > 20$ ) corresponds to 100% for wt-RSL and 70% for R17A. Scale bars: 5  $\mu\text{m}$ .

There have been limited number of attempts to produce lectin valency mutants by altering the oligomer symmetry.<sup>12,20</sup> A key point for tailoring valency is the identification of amino acids crucial for ligand interaction. Our MD calculations on RSL indicated that R17 and W76 are both strongly involved in ligand binding, confirming a previous docking study.<sup>21</sup> Interestingly, the W76A mutant has already been characterized in order to quantify the role of aromatic amino acids.<sup>22</sup> A decrease in affinity of 3 orders of magnitude ( $K_d = 1 \text{ mM}$ ) was observed, indicating that, as demonstrated in our work, some residual activity is conserved in absence of Trp.

The influence of multivalency on interactions between endogenous or exogenous lectins and glycosylated cell surfaces is of particular interest for many research fields. Several modes of interactions have been described between lectins and (glyco-)lipid rafts. First, multivalent lectins, such as galectins<sup>23</sup> or sialoadhesins,<sup>24</sup> colocalize with glycosphingolipid-containing rafts, while norovirus capsids bind to the rim of galactosylceramide domains.<sup>25</sup> Then, *in vitro* studies on model plasma membranes containing the ganglioside GM1<sup>26</sup> or on glycoengineered cells presenting lipid-anchored glycopolymers<sup>27</sup> demonstrated that multivalent lectins are able to redistribute the glycoconjugates by cross-linking and thus trigger the (re)organization of the membrane. More recently, it has been demonstrated that pentameric carbohydrate binding proteins such as the bacterial Shiga toxin and cholera toxin or the simian virus 40 (SV40) capsid bind to glycosphingolipids. They thereby induce plasma membrane curvature and thus directly promote their endocytic uptake into cells.<sup>9,10</sup> Our results on the effect of wt-RSL on glycolipid-containing GUVs demonstrated that hexavalent objects, such as  $\beta$ -propeller lectins, are also able to induce membrane invaginations. This has been observed for natural and synthetic glycolipids. However, the clustering activity is somewhat dependent on the lipid backbone since glycosphingolipids resulted in a more effective sequestration of RSL in invagination than carbohydrate-bearing phospholipids.

Mutants with reduced numbers of binding sites were previously obtained for Shiga and cholera toxins in order to determine the role of multivalency on the induction of

membrane invaginations on cells and GUVs<sup>10</sup> and on cellular internalization.<sup>10,12</sup> Reduction of the number of binding sites did not inhibit, rather it attenuated, the internalization of cholera toxin or Shiga toxin in cells but inhibited the formation of membrane invaginations by Shiga toxin on glycolipid-containing GUVs<sup>10</sup> pointing to different mechanisms of endocytosis in presence or absence of coat proteins.

By combining structural with functional studies, we have discovered that both wt-RSL and W76A were able to induce membrane invaginations in giant unilamellar vesicles while the R17A mutant did not. The observation that the trimeric, trivalent R17A mutant, with binding sites that are 30 Å apart, is not able to form membrane invaginations complements nicely the data obtained for the pentameric Shiga toxin.<sup>10</sup> Since the B-subunit of Shiga toxin (StxB) can bind up to 15 receptor molecules and the StxB binding site mutant W37A still up to 10, it is yet extremely difficult to determine how many receptor molecules are indeed necessary for the formation of membrane invaginations and the cellular uptake. It gets more and more obvious that a certain degree of multivalency and/or topology of binding sites is requested in order to induce membrane invaginations.

Over the years, many efforts have been made to understand the effect of ligand valency on receptor binding and pathogenicity,<sup>28,29</sup> but not much has been invested on the lectin one.<sup>20</sup> The hexavalent lectin RSL, in combination with its truly trivalent mutant R17A, represents unique tools for studying the effect of multivalency on synthetic and cellular systems. Using synthetic systems in combination with the trimeric RSL, we have gained new insights regarding the understanding of the role and the mode of action of lectin multivalency. It can be stressed that, on the one hand, the common pentameric protein scaffold of Shiga and cholera toxins, and the VP1 subunit of Simian Virus 40, is not an exclusive feature to promote membrane invaginations. On the other hand, the location of all carbohydrate binding sites on the same face of the protein promotes strong avidity for membrane but is not always sufficient for the formation of membrane invaginations as illustrated by the inactive trivalent mutant. This

opens the route to further investigations about deciphering the interplay between the number of binding sites and their overall topologies. The results of the present work are also touching fundamental questions regarding evolution. Indeed, as pointed out previously,<sup>12</sup> we need to understand if the evolutionary driving force for lectin multivalency is to enhance avidity for surface binding or to cluster glycosylated receptors for membrane deformation and cellular uptake. These processes might be highly regulated and dependent on cellular and microbial factors such as availability of different cell surface receptors and corresponding lectin affinities. With this first quantitative evaluation of the effect of valency on affinity/avidity changes and on the formation of membrane invaginations, we can conclude that high avidity can be obtained with a relatively small number of binding sites while glycolipid clustering requires more binding sites and/or different topologies that have to be explored. For future research, engineered neoRSLs with tailored valency ranging from hexavalent to monovalent would be highly appreciated tools to elucidate the molecular mechanisms of binding, membrane invagination, trafficking, and other physiological processes induced by lectins.

## METHODS

**Molecular Dynamics Calculations.** A 32 ns trajectory of fully hydrated RSL complexed with  $\alpha$ MeFuc was produced using the program NAMD 2.8 running with the Amber99SB and GLYCAM06 force fields<sup>30</sup> (See the SI for details). The sum of nonbonded interaction energy between the fucoside and the residues comprised within a sphere of 8 Å of the ligand was calculated throughout the trajectory, and averaged over the three symmetrically related intramonomeric binding sites.

**Production of RSL Mutants and Characterization.** The pET-25b-wt-RSL plasmid<sup>16</sup> was used as a template for site-directed mutagenesis using PCR (SI Table S1). Production wt-RSL and mutants were performed, as previously described.<sup>16</sup> Details of production and affinity purification are available in the SI. Purified proteins were dialyzed against ultrapure water and lyophilized. The monodispersity and oligomerization state of the proteins were evaluated by dynamic light scattering at 1 mg mL<sup>-1</sup> (Zetasizer Nano series, Malvern Instruments), and thermal stability was evaluated by DSC (see the SI). ITC experiments were performed with an ITC200 isothermal titration calorimeter (Microcal-GE Healthcare). SPR experiments were performed on a Biacore X100 instrument (GE Healthcare). Crystals of R17A/ $\alpha$ MeFuc and W76A/fucose were obtained by the hanging drop vapor diffusion method. Details on ITC, SPR, crystallization conditions, data collection, and refinement procedure are given in the SI.

**Production of GUVs and Lectin Binding.** GUVs were composed of DOPC (64.75 mol %), Texas Red DHPE (0.25 mol %), cholesterol (30 mol %), and various glycolipid species (5 mol %) (see the SI for preparation). GUVs were grown at RT using the electroformation technique on indium–tin oxide (ITO)-coated glass slides mostly as described.<sup>10,31</sup> Wt-RSL or mutants (5  $\mu$ g mL<sup>-1</sup>) were incubated with GUVs at RT and examined under an inverted confocal fluorescence microscope (NIKON AIR) equipped with a 60 $\times$  oil immersion objective (Plan Apo, N.A. 1.40) using the resonant scanner.

**Incubation of Human Lung Epithelial Cells with Lectins.** H1299 cells were incubated with either wt-RSL-Cy3 or R17A-Cy3 (2  $\mu$ g mL<sup>-1</sup>) for 1 h at 37 °C. Cells were then fixed in 4 % paraformaldehyde for 15 min at room temperature (RT). The coverslips were mounted and analyzed by confocal fluorescence microscopy. Fluorescence intensity inside cells ( $n > 20$ ) was evaluated after 1 h of lectin incubation using the “corrected total cell fluorescence” protocol of the ImageJ software.<sup>32</sup>

## ASSOCIATED CONTENT

### Supporting Information

Details of methods, MD trajectory, DSC analysis, X-ray data collection and refinement statistics, ITC and SPR tables, GUV preparation, and movies of membrane invaginations. This material is available free of charge via the Internet at <http://pubs.acs.org>.

### Accession Codes

Coordinates and structure factors of W76A and R17A mutants complexed with fucoside and fucose have been deposited in the Protein Data Bank, [www.pdb.org](http://www.pdb.org) (PDB ID code 3Z18 and 4I6S).

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [imberty@cermav.cnrs.fr](mailto:imberty@cermav.cnrs.fr); [winfried.roemer@bioss.uni-freiburg.de](mailto:winfried.roemer@bioss.uni-freiburg.de).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work has been funded by Agence Nationale de la Recherche Grant NeoLect (ANR-11-BSVS-002). A. Imberty acknowledges support from CNRS and J. Arnaud from Université Grenoble 1. W. Römer is supported by the Excellence Initiative of the German Research Foundation (EXC 294), by a grant from the Ministry of Science, Research and the Arts of Baden-Württemberg (Az: 33-7532.20) and by a starting grant of the European Research Council (Programme “Ideas” No. ERC-2011-StG 282105). The COST actions CM1102 and BM1003 and the Labex ARCANE (ANR-11-LABX-003) are thanked for support. The authors are grateful to the SOLEIL synchrotron for provision of synchrotron radiation facilities and the CINES for calculation facilities. We thank S. Vidal for the generous gift of biotinylated fucoside. The technical help from the Proxima 1 beamline staff, E. Gillon, C. Bras, and L. Charlier is also acknowledged.

## REFERENCES

- (1) Ambrosi, M., Cameron, N. R., and Davis, B. G. (2005) Lectins: Tools for the molecular understanding of the glycode. *Org. Biomol. Chem.* 3, 1593–1608.
- (2) Arnaud, J., Audfray, A., and Imberty, A. (2013) Binding sugars: From natural lectins to synthetic receptors and engineered neolectins. *Chem. Soc. Rev.* 42, 4798–4813.
- (3) Lis, H., and Sharon, N. (1998) Lectins: Carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* 98, 637–674.
- (4) Collins, B. E., and Paulson, J. C. (2004) Cell surface biology mediated by low affinity multivalent protein–glycan interactions. *Curr. Opin. Chem. Biol.* 8, 617–625.
- (5) Imberty, A., and Varrot, A. (2008) Microbial recognition of human cell surface glycoconjugates. *Curr. Opin. Struct. Biol.* 18, 567–576.
- (6) Beddoe, T., Paton, A. W., Le Nours, J., Rossjohn, J., and Paton, J. C. (2010) Structure, biological functions, and applications of the ABS toxins. *Trends Biochem. Sci.* 35, 411–418.
- (7) Eierhoff, T., Stechmann, B., Romer, W. (2012) Pathogen and toxin entry—How pathogens and toxins induce and harness endocytotic mechanisms. In *Molecular Regulation of Endocytosis* (Ceresa B., Ed.) InTech, New York.
- (8) Johannes, L., and Romer, W. (2010) Shiga toxins—From cell biology to biomedical applications. *Nat. Rev. Microbiol.* 8, 105–116.



- (9) Ewers, H., Romer, W., Smith, A. E., Bacia, K., Dmitrieff, S., et al. (2010) GM1 structure determines SV40-induced membrane invagination and infection. *Nat. Cell Biol.* 12, 11–12.
- (10) Romer, W., Berland, L., Chambon, V., Gaus, K., Windschiegel, B., et al. (2007) Shiga toxin induces tubular membrane invaginations for its uptake into cells. *Nature* 450, 670–675.
- (11) Romer, W., Pontani, L. L., Sorre, B., Rentero, C., Berland, L., et al. (2010) Actin dynamics drive membrane reorganization and scission in clathrin-independent endocytosis. *Cell* 140, 540–553.
- (12) Jobling, M. G., Yang, Z., Kam, W. R., Lencer, W. I., and Holmes, R. K. (2012) A single native ganglioside GM1-binding site is sufficient for cholera toxin to bind to cells and complete the intoxication pathway. *mBio* 3, e00401–00412.
- (13) 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.
- (14) Cioci, G., Mitchell, E. P., Chazalet, V., Debray, H., Oscarson, S., et al. (2006)  $\beta$ -Propeller crystal structure of *Psathyrella velutina* lectin: An integrin-like fungal protein interacting with monosaccharides and calcium. *J. Mol. Biol.* 357, 1575–1591.
- (15) Wimmerova, M., Mitchell, E., Sanchez, J. F., Gautier, C., and Imberty, A. (2003) Crystal structure of fungal lectin: Six-bladed  $\beta$ -propeller fold and novel recognition mode for *Aleuria aurantia* lectin. *J. Biol. Chem.* 278, 27059–27067.
- (16) Kostlanová, N., Mitchell, E. P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., et al. (2005) The fucose-binding lectin from *Ralstonia solanacearum*: A new type of  $\beta$ -propeller architecture formed by oligomerisation and interacting with fucoside, fucosylactose, and plant xyloglucan. *J. Biol. Chem.* 280, 27839–27849.
- (17) Audfray, A., Claudinon, J., Abounit, S., Ruvoën-Clouet, N., Larson, G., et al. (2012) The fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes. *J. Biol. Chem.* 287, 4335–4347.
- (18) Chen, C. K. M., Chan, N.-L., and Wang, A. H. J. (2011) The many blades of the  $\beta$ -propeller proteins: Conserved but versatile. *Trends Biochem. Sci.* 36, 553–561.
- (19) Munoz, E. M., Correa, J., Fernandez-Megia, E., and Riguera, R. (2009) Probing the relevance of lectin clustering for the reliable evaluation of multivalent carbohydrate recognition. *J. Am. Chem. Soc.* 131, 17765–17767.
- (20) Olausson, J., Astrom, E., Jonsson, B. H., Tibell, L. A., and Pahlsson, P. (2011) Production and characterization of a monomeric form and a single-site form of *Aleuria aurantia* lectin. *Glycobiology* 21, 34–44.
- (21) Mishra, S. K., Adam, J., Wimmerova, M., and Koca, J. (2012) In silico mutagenesis and docking study of *Ralstonia solanacearum* RSL lectin: Performance of docking software to predict saccharide binding. *J. Chem. Inf. Model.* 52, 1250–1261.
- (22) Wimmerova, M., Kozmon, S., Necasova, I., Mishra, S. K., Komarek, J., et al. (2012) Stacking interactions between carbohydrate and protein quantified by combination of theoretical and experimental methods. *PLoS One* 7, e46032.
- (23) Hsu, D. K., Chernyavsky, A. I., Chen, H. Y., Yu, L., Grando, S. A., et al. (2009) Endogenous galectin-3 is localized in membrane lipid rafts and regulates migration of dendritic cells. *J. Invest. Dermatol.* 129, 573–583.
- (24) Vinson, M., Rausch, O., Maycox, P. R., Prinjha, R. K., Chapman, D., et al. (2003) Lipid rafts mediate the interaction between myelin-associated glycoprotein (MAG) on myelin and MAG-receptors on neurons. *Mol. Cell Neurosci.* 22, 344–352.
- (25) Bally, M., Rydell, G. E., Zahn, R., Nasir, W., Eggeling, C.B., M.E., et al. (2012) Norovirus GI.4 virus-like particles recognize galactosylceramides in domains of planar supported lipid bilayers. *Angew Chem, Int. Ed.* 51, 12020–12024.
- (26) Hammond, A. T., Heberle, F. A., Baumgart, T., Holowka, D., Baird, B., et al. (2005) Crosslinking a lipid raft component triggers liquid ordered–liquid disordered phase separation in model plasma membranes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6320–6325.
- (27) Belardi, B., O'Donoghue, G. P., Smith, A. W., Groves, J. T., and Bertozzi, C. R. (2012) Investigating cell surface galectin-mediated cross-linking on glycoengineered cells. *J. Am. Chem. Soc.* 134, 9549–9552.
- (28) Bernardi, A., Jiménez-Barbero, J., Casnati, A., Castro, C. D., Darbre, T., et al. (2013) Multivalent glycoconjugates as anti-pathogenic agents. *Chem. Soc. Rev.* 42, 4709–4727.
- (29) Gestwicki, J. E., Cairo, C. W., Strong, L. E., Oetjen, K. A., and Kiessling, L. L. (2002) Influencing receptor–ligand binding mechanisms with multivalent ligand architecture. *J. Am. Chem. Soc.* 124, 14922–14933.
- (30) Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., et al. (2008) GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* 29, 622–655.
- (31) Mathivet, L., Cribier, S., and Devaux, P. F. (1996) Shape change and physical properties of giant phospholipid vesicles prepared in the presence of an AC electric field. *Biophys. J.* 70, 1112–1121.
- (32) Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.

**SUPPORTING INFORMATION****Reduction of lectin valency drastically changes glycolipid dynamics in membranes, but not surface avidity**

Julie Arnaud, Julie Claudinon, Kevin Tröndle, Marie Trovalet, Göran Larson, Aline Thomas, Annabelle Varrot, Winfried Römer,\* Anne Imberty\* & Aymeric Audfray

<b>I. Molecular dynamics calculations</b>	<b>S2</b>
<b>II. Cloning of mutants and production of recombinant proteins</b>	<b>S3</b>
<b>III. Differential Scanning Calorimetry</b>	<b>S4</b>
<b>IV. Isothermal Titration Calorimetry Experiments</b>	<b>S5</b>
<b>V. Crystallography</b>	<b>S6</b>
<b>VI. Surface Plasmon Resonance</b>	<b>S8</b>
<b>VII. Production of GUVs and Lectin Binding</b>	<b>S9</b>
<b>VIII. References</b>	<b>S11</b>



## I. Molecular dynamics calculations

The crystal coordinates of the fucose-bound RSL were taken from the protein data bank (1) with code 2BT9 (2). The protein topology was generated using the Leap module of Amber10(3) and the Amber99SB force field (4). The sugar moieties were built using the (5) force field. The residues protonation states were chosen so as to mimic a pH of 7; in order to maximize the local hydrogen bond network, the three histidine enantiomers were built with an H $\epsilon$  atom. The total charge of the system was null so that no counter-ion was added. The resulting system was placed into a box of TIP3P (6) water molecules of size 88\*85\*81 Å<sup>3</sup> so that the closest distance between any atom of the bound protein and the edge of the water box was 14 Å. The overall system comprised 49807 atoms. Periodic boundary conditions were used to model the solvent. The system was then subjected to 10000 steps of conjugate gradient energy minimization, followed by a constant pressure and temperature molecular dynamics (MD) simulation (1 atm, 298 K). The integration time step was equal to 2 fs. All atoms were allowed to move during the MD simulations, but the internal geometry of water molecules was kept rigid, as well as all bonds to hydrogen atoms. Atom-pairs less than 12 Å apart were included in the non-bonded pair lists. A cutoff of 10 Å was taken for the non-bonded interactions with a switching function equal to 9 Å. The Particle mesh Ewald summation method was employed for efficiently computing the Coulomb forces (7). The first 800 ps were used for equilibrating the system, and the production run from which all analyzes were derived lasted 32 ns. Simulations were performed using the program NAMD 2.8 (8) at the CINES center (32 CPU, 3500 hours). The analysis and visualization of the trajectory was performed using VMD (9).

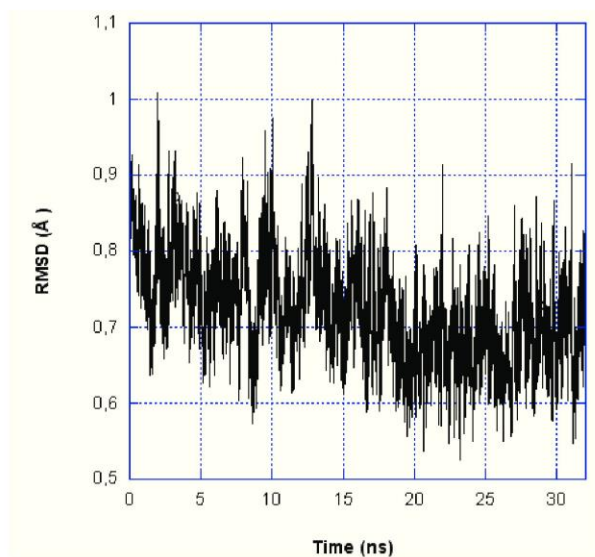


Figure S1 : Time distribution of the RMSD of each frame with the average structure along the trajectory, calculated on all C $\alpha$  atoms.

## II. Cloning of mutants and production of recombinant proteins

The pET-25b-wt-RSL plasmid (2) was used as a template for site-directed mutagenesis using PCR (SI Table 1). The pET-25b-R17A and pET-25b-W76A vectors were transformed into *E. coli* TUNER™ (DE3) (Novagen). Cells harboring the plasmids were grown in LB broth containing 100 µg ml<sup>-1</sup> ampicillin at 37°C until the optical density at 600nm reached 0.6. The cells were then cultured for an additional 3 h at 30°C after the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside. After induction, bacterial cells were harvested by centrifugation at 6,000g for 20 at 4 °C min and resuspended with buffer containing 20 mM Tris/HCl, pH 7.5, 100 mM NaCl. Wt-RSL was purified by affinity chromatography on a mannose-agarose column (Sigma-Aldrich) as previously described (2). R17A and W76A mutants were enriched on a fucose-sepharose and mannose-agarose columns respectively using isocratic elution with buffer 20 mM Tris/HCl, pH 7.5, 100 mM NaCl. The eluted fractions were then directly loaded on a Superdex™ 75 10/30 column (GE Healthcare) equilibrated with the above buffer. Fractions containing RSL mutants were then pooled and dialyzed against ultrapure water prior to lyophilization.

Table S1: Primers used for site-directed mutagenesis

Mutants	Primers
R17A	GAACCGTACCGTTCGATCGCTGTGTACACGGCCAATAATG
	CATTATTGGCCGTGTACACAGCGATCGACGGTACGGTTC
W76A	GACCACGGAGTGGTGC GCGGACGGCAACGGCTGGAC
	GTCCAGCCGTTGCCGTCCGCGCACCCTCCGTGGTC
R62A	GGCAGCGCGATCCATATGCCCGTCTATGCAAGCACCGG
	CCGGTGCTTGCATAGACGGGCATATGGATCGCGCTGCC
E28A	CACGGCCAATAATGGCAAGATCACCGCGCGATGCTGG
	CCAGCATCGCGCGGTGATCTTGCCATTATTGGCCGTG
W31A	GATCACCGAGCGATGC GCGGACGGGAAGGGGTGGTACACC
	GGTGTACCACCCCTTCCCCTCCGCGCATCGCTCGGTGATC

### III. Differential Scanning Calorimetry

DSC measurements were performed using a Nano-DSC III calorimeter from TA Instrument, with 0.33 ml sample and reference cells. All proteins concentrations were 0.33-0.35 mg.ml<sup>-1</sup> in phosphate buffer (0.1 M pH 7.5).  $\alpha$ MeFuc concentration was 0.18 mM. Protein solutions were heated at 1°C.min<sup>-1</sup>, from 20 to 110°C, at a constant pressure of 3 atm to prevent degassing during heating. The excess molar heat capacity was calculated assuming the molecular mass of proteins to be 10 kDa. Data analysis was made with NanoAnalyze Software from TA Instruments after subtraction of buffer reference thermogram and sigmoidal baseline. The transition temperature ( $T_m$ ) is the temperature at which half the protein molecules are folded. Integration of the heat capacity of the sample vs. temperature yields the enthalpy of the unfolding process ( $\Delta H = \int_{T_1}^{T_2} Cp dT$ ). Fitting the transition curve with a two-state function supplied by the instrument software yield a van't Hoff enthalpy ( $\Delta H_{VH}$ ).

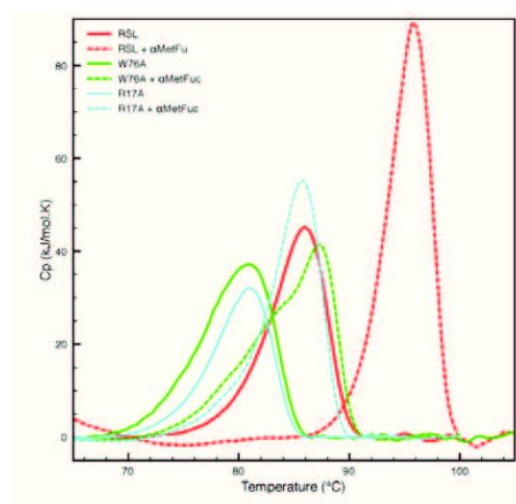


Figure S2 : DSC thermogram traces for RSL and mutants, in presence and absence of  $\alpha$ MeFuc. Corrections of baseline and normalization of concentration have been applied

Table S2 : Thermodynamic parameters for the unfolding of RSL and mutants in the absence and presence of fucoside ligand.

	$\Delta H$	$\Delta S$	$T_m$	$\Delta H_{VH}$
	<i>kJ/mol</i>	<i>kJ/mol.K</i>	<i>°C</i>	<i>kJ/mol</i>
RSL	$281 \pm 14$	$0.78 \pm 0.04$	$86.0 \pm 0.1$	$388 \pm 18$
$\alpha$ MeFuc	$488 \pm 51$	$1.32 \pm 0.14$	$95.8 \pm 0.1$	$603 \pm 17$
W76A	$298 \pm 2$	$0.84 \pm 0.01$	$81.0 \pm 0.1$	$380 \pm 2$
$\alpha$ MeFuc	$332 \pm 5$	$0.92 \pm 0.01$	$87.4 \pm 0.1$	$387 \pm 7$
R17A	$236 \pm 31$	$0.67 \pm 0.09$	$80.7 \pm 0.5$	$343 \pm 33$
$\alpha$ MeFuc	$356 \pm 34$	$1.00 \pm 0.01$	$85.4 \pm 0.4$	$448 \pm 15$

#### IV. Isothermal Titration Calorimetry Experiments

ITC experiments were performed with an ITC200 isothermal titration calorimeter (Microcal-GE Healthcare). The experiments were carried out at 25°C. Lectins and  $\alpha$ MeFuc were dissolved in the same buffer composed of 20 mM Tris with 0.1 M NaCl at pH 7.5. A total of 20 injections of 2  $\mu$ l of 1.5 mM sugar solution were added at intervals of 120 s whilst stirring at 1000 rev.min<sup>-1</sup> in the 0.09 mM protein solution in the 200  $\mu$ l cell. Control experiments performed by injection of buffer into the protein solution yielded insignificant heats of dilution. The experimental data were fitted to a theoretical titration curve using Origin software supplied by Microcal, with  $\Delta H$  (enthalpy change),  $K_a$  (association constant) and  $n$  (number of binding sites per monomer) as adjustable parameters. Free energy change ( $\Delta G$ ) and entropy contributions ( $T\Delta S$ ) were derived from the equation  $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$  (with T the absolute temperature and R = 8.314 J mol<sup>-1</sup>K<sup>-1</sup>).

Table S3: Thermodynamic data for the interaction of RSL, W76A and R17A with  $\alpha$ MeFuc.

	$K_d$ $\mu M$	$n$	$-\Delta G$ $kJ/mol$	$-\Delta H$ $kJ/mol$	$T\Delta S$ $kJ/mol$
RSL	0.87 $\pm$ 0.08*	1.98 $\pm$ 0.01	34.6	45.0 $\pm$ 0.3	10.4
W76A	2.6 $\pm$ 0.3	0.80 $\pm$ 0.09	31.9	50.3 $\pm$ 1	18.4
R17A	2.21 $\pm$ 0.06	0.82 $\pm$ 0.03	32.3	46.6 $\pm$ 0.3	14.3

\*standard deviations are determined by at least two independent experiments.



## V. Crystallography

Crystals of lectin/monosaccharide complexes were obtained by the hanging drop vapor diffusion method. R17A/ $\alpha$ MeFuc complex was obtained by mixing 1  $\mu$ l of protein at 10 mg ml<sup>-1</sup> with 10  $\mu$ M of  $\alpha$ MeFuc and 1  $\mu$ l of reservoir solution 1-4 of the Morpheus screen (Molecular dimension Ltd). W76A cocrystallized with fucose present in the 2-21 solution of sale screen. Crystals appeared after 1 week of incubation at 19°C.

X-ray diffraction data were collected on Proxima 1 beamline at the Soleil synchrotron, Saint Aubin, France (SI Table 3). Data were processed using MOSFLM (10) and programs from the CCP4 suite (11) were used in further computing. Phases were obtained by the molecular replacement method using Phaser (12) and the coordinates of 2BT9 as search model. One monomer and one trimer were found in the asymmetric unit of R17A and W76A mutants respectively. Maximum likelihood restrained refinement using REFMAC 5.7 with 5% of the data retained for cross-validation purposes was combined with manual model adjustment and electron density inspection performed in Coot graphics (13).

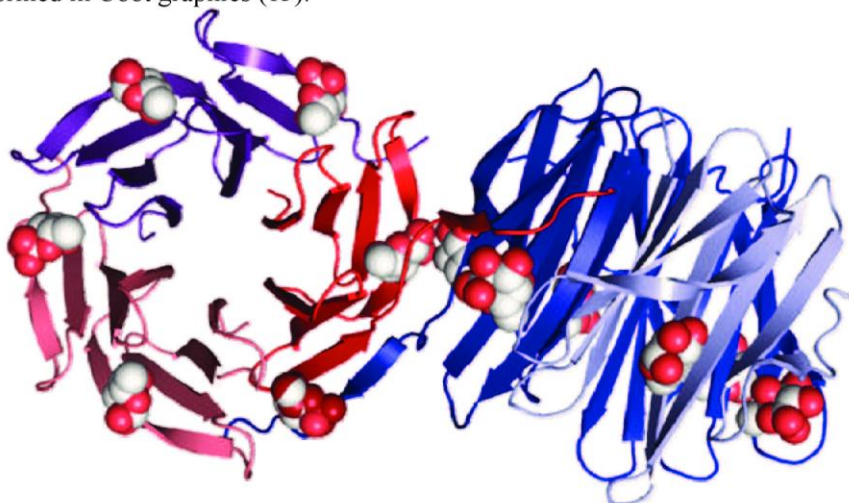


Figure S3 : Structure of W76A/Fuc dimer with swapping of one  $\beta$ -strand

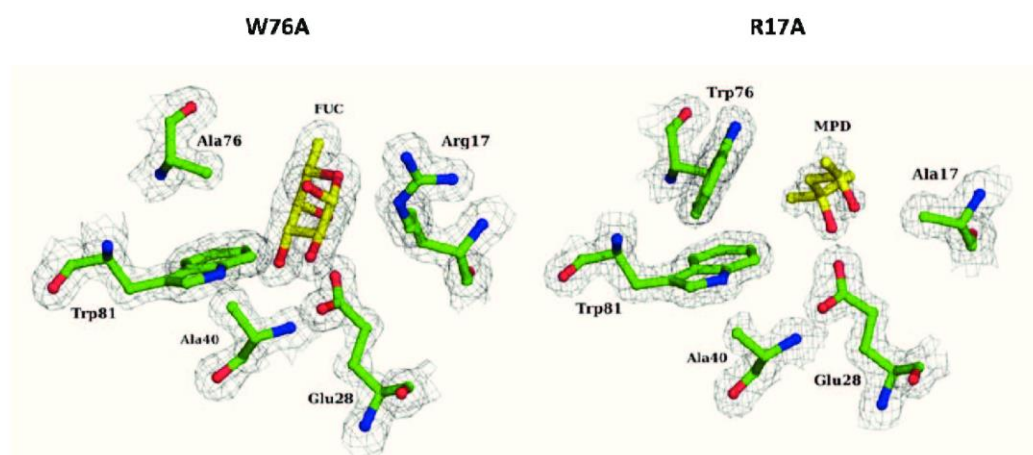


Figure S4 : Electronic density of mutated amino acids and ligands

Table S4 : Data collection and refinement statistics of R17A and W76A mutant complexed with  $\alpha$ MeFuc and fucose, respectively

Mutant	R17A/ $\alpha$ MeFuc	W76A/Fuc
Data collection		
Wavelength ( $\text{\AA}$ )	0.984	0.982
Space group	P 3 2 <sub>1</sub>	I 4 <sub>1</sub> 2 2
Cell dimensions a, b, c ( $\text{\AA}$ )	72.4 72.4 36.0	103.8 103.8 100.8
Resolution (outer shell) ( $\text{\AA}$ )	36.20-1.50 (1.54-1.50)	51.88-1.54 (1.62-1.54)
Measured/Unique reflections	95422/17523	369531/41015
Average multiplicity (outer shell)	5.4 (5.2)	9.0 (7.7)
Rmerge (outer shell)	0.046 (0.311)	0.042 (0.347)
Completeness (%) (outer shell)	99.3 (99.7)	99.8 (99.1)
Mean I/ $\sigma$ I (outer shell)	15.90 (4.80)	9.70 (2.20)
Wilson B	16.4	19.9
Refinement		
Rcryst/Rfree	13.4/16.6	14.8/17.2
r.m.s.d. bonds ( $\text{\AA}$ )	0.015	0.015
r.m.s.d. angles ( $^{\circ}$ )	1.61	1.73
r.m.s.d. chiral ( $\text{\AA}^3$ )	0.11	0.11
Protein atoms	719	2107
Bfac ( $\text{\AA}^2$ )	18.3	15.9
Ligand atoms	10	670
Bfac ( $\text{\AA}^2$ )	25.5	17.6
Water molecules	105	285
Bfac ( $\text{\AA}^2$ )	30.7	34.9
PDB code	3ZI8	4I6S

## VI. Surface Plasmon Resonance

SPR experiments were performed on a Biacore X100 instrument (GE Healthcare) at 25°C in HBS (10 mM Hepes/NaOH, pH 7.5, 150 mM NaCl, 0.05% Tween 20) at a flow rate of 10  $\mu\text{l}\cdot\text{min}^{-1}$ . Streptavidin was immobilized on a research grade CM5 chip using standard procedures. 250 resonance units of a mix containing 10% of biotinylated polyacrylamide (Lectinity) bearing  $\alpha$ -fucose and 90% bearing  $\beta$ -galactose were captured on channel 2 at 200  $\mu\text{g}\cdot\text{ml}^{-1}$ . The same amount of biotinylated polyacrylamide bearing  $\beta$ -galactose was captured on channel 1. A second chip was prepared with direct fixation of monovalent biotinylated monosaccharides (Fig S4. Gift from Sébastien Vidal, CNRS Lyon), fixed in the same proportions and conditions than polysaccharides.

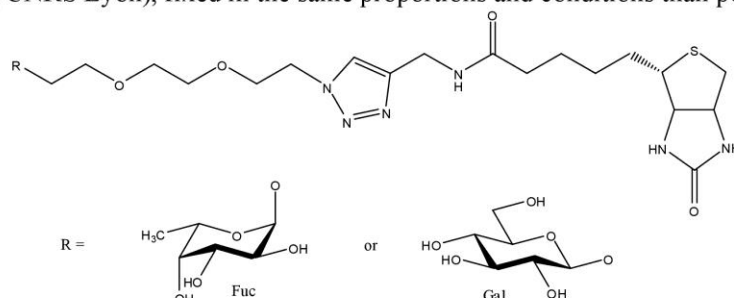


Figure S4: Monovalent monosaccharides used in SPR (Fuc- and Gal-EG3-Triaz-Biotin)

Experiments consist of injection (association 360 s, dissociation 400 s) of various concentrations of lectins (2 fold cascade dilutions, from 0 to 10  $\mu\text{M}$ ) on both channels. The chip was fully regenerated by one injection of 1M fucose in running buffer for 120 s. Binding was measured as resonance units over time after blank subtraction, and data were then evaluated by using the Biacore X100 evaluation software, version 2.0. Plots represent the subtracted data (channel 2-channel 1). Dissociation constants were determined by plotting response at equilibrium (Req, 10 s before the end of injection) against lectin concentration.

Table S5: Steady state analysis of the binding of RSL and its mutants to a sensorchip made with biotinylated PAA- $\alpha$ -fucosides. Two sites binding model has been used to fit the Req versus concentration plot.

	Kd <sub>1</sub> ( $\mu\text{M}$ )	SE (Kd <sub>1</sub> )	Rmax <sub>1</sub>	SE (Rmax <sub>1</sub> )	Kd <sub>2</sub> (nM)	SE (Kd <sub>2</sub> )	Rmax <sub>2</sub>	SE (Rmax <sub>2</sub> )	Chi <sup>2</sup>
RSL	0.7	0.3	16.1	1.5	9	1.4	29.6	1.7	1.0
W76A	1.6	0.2	20.2	0.6	18	2.4	12.4	0.6	0.2
R17A	1.7	0.2	22.8	0.7	15	1.6	18.3	0.6	0.2

Table S6. Steady state analysis of the binding of RSL and its mutants to a sensorchip made with monovalent biotinylated  $\alpha$ -fucosides. Two sites binding model has been used to fit the Req versus concentration plot.

	Kd <sub>1</sub> ( $\mu\text{M}$ )	SE (Kd <sub>1</sub> )	Rmax <sub>1</sub>	SE (Rmax <sub>1</sub> )	Kd <sub>2</sub> (nM)	SE (Kd <sub>2</sub> )	Rmax <sub>2</sub>	SE (Rmax <sub>2</sub> )	Chi <sup>2</sup>
RSL	6.8	3.50	105.0	16.0	15.5	2.4	150.7	6.3	38
W76A	2.3	0.45	88.5	4.4	44.8	5.9	85.86	5	4.1
R17A	1.5	0.25	135.7	7.3	40.7	14.0	48.35	8.2	7.3



### VII. Production of GUVs and Lectin Binding

The lipids 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) and cholesterol were purchased from Avanti Polar Lipids. Texas Red® 1,2-Dihexadecanoyl-*sn*-Glycero-3-Phosphoethanolamine Triethylammonium Salt (Texas Red® DHPE) was from Molecular Probes (Life Technologies) and FSL-Lewis a from KODE Biotech Limited. FSL-Lewis a is comprised of the Lewis blood group Lewis a trisaccharide Gal $\beta$ (Fuc $\alpha$ 4)GlcNAc $\beta$  representing F, conjugated via an O(CH<sub>2</sub>)<sub>3</sub>NH spacer (S) to an activated adipate derivative of dioleoyl-phosphatidylethanolamine (DOPE) representing L. The fucosylated type 1 chain glycosphingolipids and the precursor lactotetraosylceramide were isolated from meconia, pooled according to blood group ABO, and characterized in detail as described previously for a human O Lewis (a-b+) secretor (14).

GUVs were grown at room temperature using the electroformation technique on indium-tin oxide (ITO)-coated glass slides. Lipid mixtures were dissolved in chloroform at a final concentration of 0.5 mg/ml, and 15  $\mu$ l of solution was spread on the conductive faces of ITO slides. After at least 2 h of drying under vacuum, GUVs were grown in a 290 mOsm sucrose solution by applying an alternating electric field from 20 mV to 1.1 V for 15 h.

Wild-type and mutant lectins were labeled with Alexa488 (Invitrogen) according to the protocol from the manufacturer. The amount of covalently bound dye was evaluated by spectrophotometer reading absorbance to be one dye molecule per lectin monomer, with no variations between the different proteins. Inhibition of hemagglutination was used to check that the activity of labeled RSL is not affected by labeling.

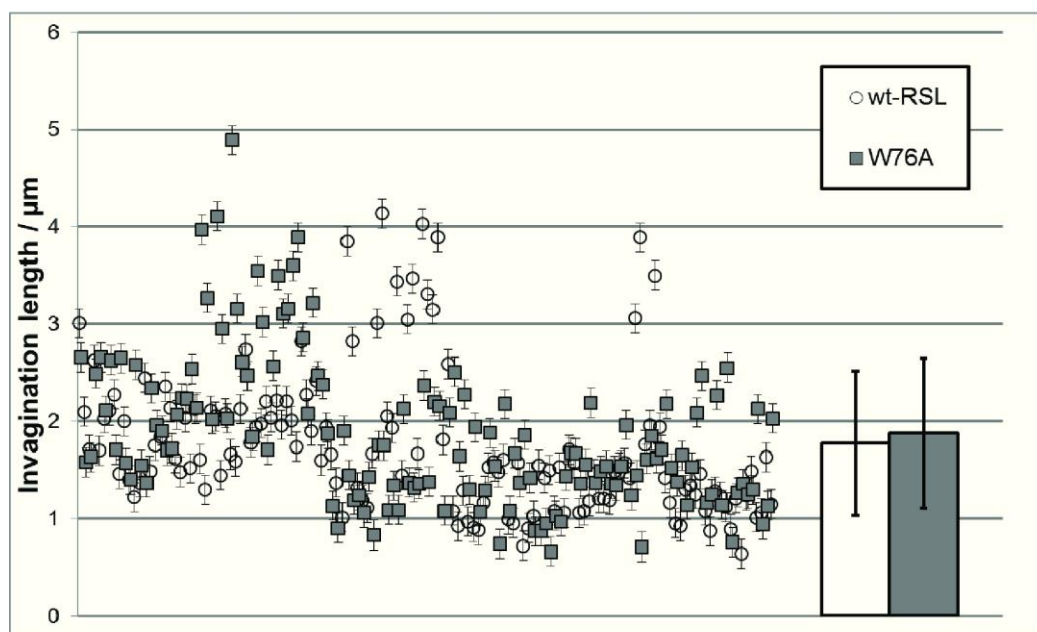


Figure S5: Quantification of the lengths of membrane invaginations. Left part: Lengths of invaginations with systematic error (0.3  $\mu$ m). Right part: Arithmetic means with standard deviation. The average in lengths for wt-RSL and R17A mutant is  $1.77 \pm 0.75 \mu$ m and  $1.87 \pm 0.77 \mu$ m, respectively.

The movies were assembled with ImageJ software using the JPEG compression and a frame rate of 5 fps.

### **Legends for movies**

**Movie 1:** Confocal real time imaging experiment at the equatorial plane of a GUV composed of DOPC/cholesterol/Lewis a-GSL incubated with Alexa488-RSL wt. Note that the system is at steady-state 15 minutes after initial formation of tubules, and that tubules move in and out of the plane of observation.

**Movie 2:** Confocal real time imaging experiment at the equatorial plane of a GUV composed of DOPC/cholesterol/Lewis a-DOPE incubated with Alexa488-wt RSL. Note that the system is at steady-state 15 minutes after initial formation of tubules, and that tubules move in and out of the plane of observation.

**Movie 3:** Confocal real time imaging experiment at the equatorial plane of a GUV composed of DOPC/cholesterol/Lewis a-DOPE incubated with Alexa488-RSL W76A mutant. Note that the system is at steady-state 15 minutes after initial formation of tubules, and that tubules move in and out of the plane of observation.

**Movie 4:** Confocal real time imaging experiment at the equatorial plane of a GUV composed of DOPC/cholesterol/Lewis a-DOPE incubated with Alexa488-RSL R17A mutant.

**VIII. References**

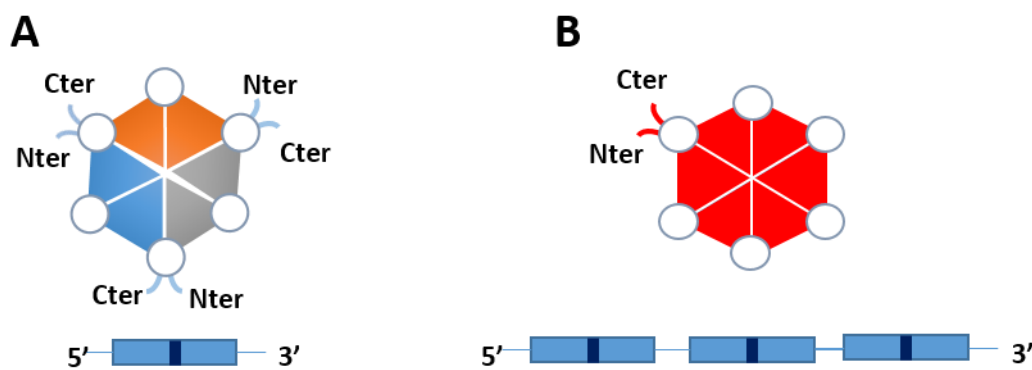
1. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235-242.
2. Kostlanová N, et al. (2005) The fucose-binding lectin from *Ralstonia solanacearum*: a new type of  $\beta$ -propeller architecture formed by oligomerisation and interacting with fucoside, fucosyllactose and plant xyloglucan. *J Biol Chem* 280:27839-27849.
3. Case DA, et al. (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26:1668-1688.
4. Hornak V, et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65:712-725.
5. Kirschner KN, et al. (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J Comput Chem* 29:622-655.
6. Jorgensen WL, et al. (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926-935.
7. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089-10092.
8. Phillips JC, et al. (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781-1802.
9. Humphrey W, Dalke A, Schulten K (1996) VMD - Visual Molecular Dynamics. *J Mol Graph* 14:33-38.
10. Leslie AW, Powell H (2007) Processing diffraction data with mosflm. in *Evolving Methods for Macromolecular Crystallography*, eds. Read, R. & Sussman, J. (Springer Netherlands), pp. 41-51.
11. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr, Sect D: Biol Crystallogr* 67:235-242.
12. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.
13. Emsley P, Lohkamp B, Scott W, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486-501.
14. Karlsson KA, Larson G (1981) Molecular characterization of cell surface antigens of fetal tissue. Detailed analysis of glycosphingolipids of meconium of a human O Le(a--b+) secretor. *J Biol Chem* 256:3512-3524.



## I.2. Construction d'une protéine RSL monomérique et modification de sa valence et topologie

### I.2.1. Introduction

Nous avons montré, par la construction d'un mutant trivalent que l'avidité n'était pas impactée par une diminution de valence passant de 6 à 3. En revanche, la formation de tubules est fortement dépendante de la valence (Arnaud *et al.*, 2013). Nous avons atteint les limites en termes de possibilité de conception de protéines de valences modulables en partant de la protéine RSL sauvage trimérique. En effet, seuls des mutants de valence 3 peuvent être obtenus et avec une seule topologie.



**Figure 44 :** Représentation schématique de la protéine RSL sauvage et de la lectine *néoRSL\_IV*.

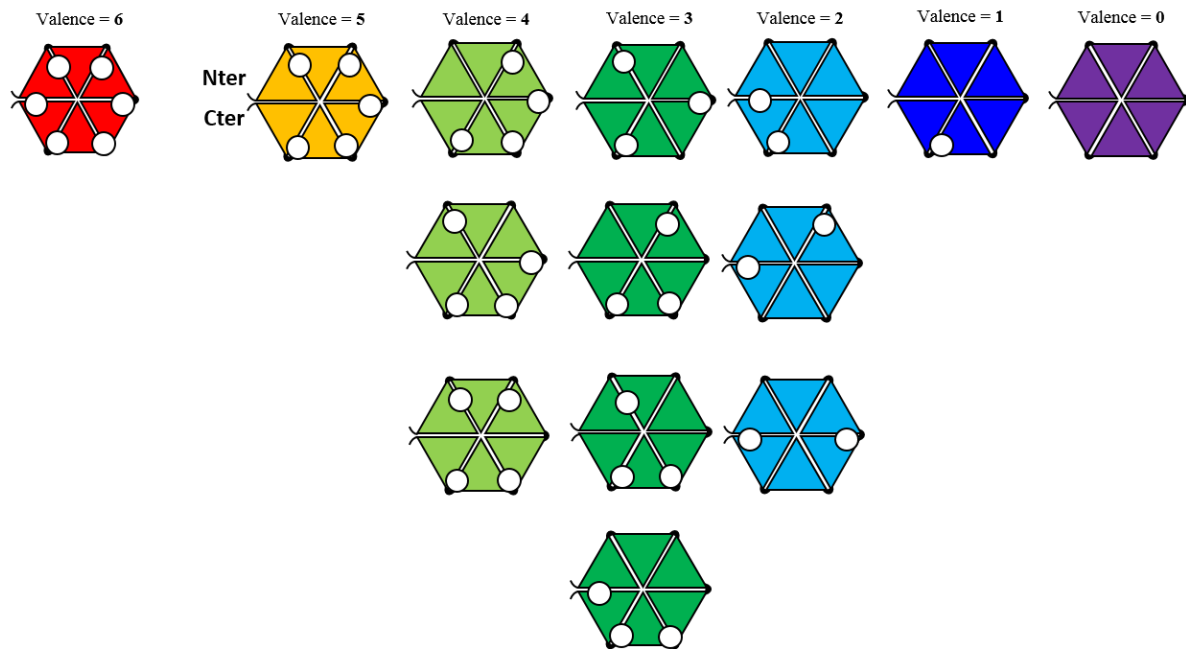
(A) Représentation de la protéine RSL et du gène codant. Chaque monomère est représenté d'une couleur différente et chaque site de fixation actif est identifié par un rond blanc.

(B) Représentation de la protéine *néoRSL\_VI* et du gène codant. Chaque site de fixation actif est identifié par un rond blanc. Les linker 1 et 2 codent pour la séquence SSTVPGD (Arnaud *et al.*, 2014).

Pour poursuivre l'étude de la multivalence et de son effet sur l'invagination de la membrane induite par les lectines, j'ai conçu à partir de la protéine RSL sauvage, une protéine RSL monomérique hexavalente. Pour ce faire, un gène composé de la séquence de RSL répétée 3 fois et reliées entre elles par une courte séquence linker codant pour la séquence SSTVPGD a été synthétisé (Figure 44). Afin de pouvoir muter spécifiquement chacun des

sites de fixation, les codons ont été dégénérés. Ainsi la néoRSL\_IV a été créée. Elle possède les mêmes caractéristiques fonctionnelles et structurales que la protéine RSL sauvage.

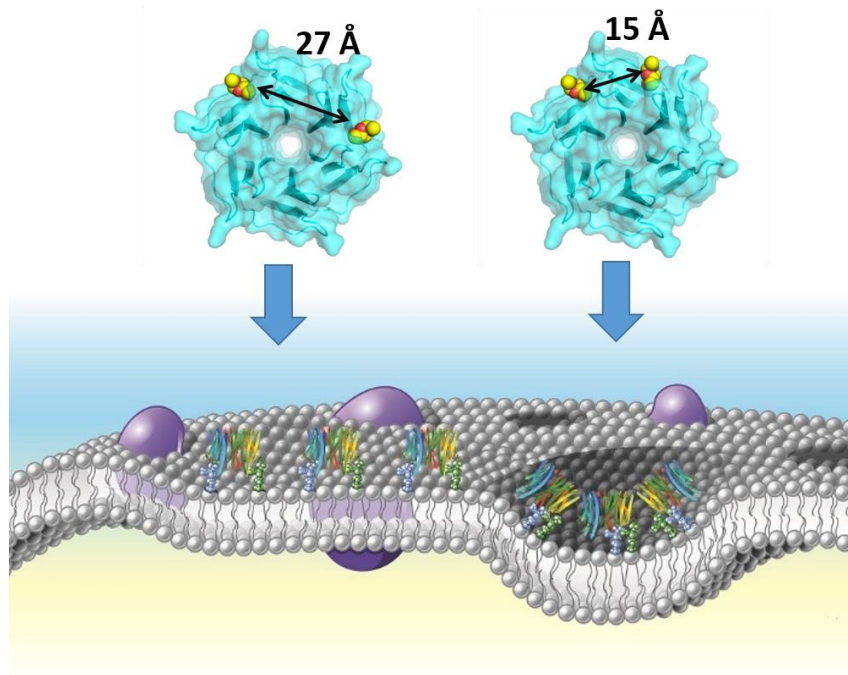
Treize néoRSL différentes ont ensuite été conçues par mutagenèse dirigée, en inactivant les sites de fixation un après l'autre par la substitution de l'arginine, stabilisant le fucose, en une alanine. Ainsi des néoRSL avec une valence de 0 à 6 et différentes topologies ont été obtenues (Figure 45).



**Figure 45 : Représentation schématique des 13 néoRSL de valence et topologie créées**

Chaque site actif est représenté par un cercle blanc. Chaque couleur représente une valence. (Arnaud et al., 2014).

Les mutants ont été caractérisés pour leur capacité à lier le fucose en solution et leur fixation à une surface glycosylée, mais aussi leur capacité à invaginer des liposomes géants contenant des glycolipides fucosylés. L'avidité pour une surface fucosylée a très peu diminuée lorsque la protéine passe d'une valence 6 à une valence 2. La topologie n'affecte pas non plus l'avidité. En revanche, une diminution drastique de l'affinité est observée lorsque la valence passe de 2 à 1, correspondant à la perte d'avidité. L'aptitude des lectines à invaginer les membranes dépend quant à elle essentiellement de la distance entre deux sites de liaisons adjacents (Figure 46). Deux sites de fixation sont suffisants pour permettre la formation de tubules s'ils sont très proches l'un de l'autre (15 Å).



**Figure 46 : Effet de la topologie sur la capacité des lectines à invaginer les membranes.**  
 Représentation de l'effet de la distance entre deux sites de fixation sur la dynamique membranaire. (Arnaud et al., 2014).

### **I.2.2. Déformation des membranes par des néolectines avec des sites de fixations ingénierés**

J'ai réalisé l'intégralité des expériences présentées dans cet article. Je suis partie faire un stage de 3 semaines en Allemagne dans le laboratoire du Prof. Winfried Römer où j'ai réalisé l'étude des interactions lectines/GUV avec l'aide de Kevin Trondle. J'ai résolu la structure cristallographique de la protéine mutante avec l'aide du Dr. Annabelle Varrot.



## Membrane Deformation by Neoelectins with Engineered Glycolipid Binding Sites\*\*

Julie Arnaud, Kevin Tröndle, Julie Claudinon, Aymeric Audfray, Annabelle Varrot, Winfried Römer,\* and Anne Imberty\*

**Abstract:** Lectins are glycan-binding proteins that are involved in the recognition of glycoconjugates at the cell surface. When binding to glycolipids, multivalent lectins can affect their distribution and alter membrane shapes. Neoelectins have now been designed with controlled number and position of binding sites to decipher the role of multivalency on avidity to a glycosylated surface and on membrane dynamics of glycolipids. A monomeric hexavalent neoelectin has been first engineered from a trimeric hexavalent bacterial lectin. From this neoelectin template, 13 different neoelectins with a valency ranging from 0 to 6 were designed, produced, and analyzed for their ability to bind fucose in solution, to attach to a glycosylated surface and to invaginate glycolipid-containing giant liposomes. Whereas the avidity only depends on the presence of at least two binding sites, the ability to bend and invaginate membranes critically depends on the distance between two adjacent binding sites.

By deciphering the glycode, lectins play crucial roles in many biological or pathological processes.<sup>[1]</sup> Whereas individual lectin binding sites have rather low affinity for glycans, multivalency results in high avidity for carbohydrates presented in several copies on glycoconjugates at the cell

surface.<sup>[2]</sup> Moreover, multivalency also triggers the clustering of glycoproteins or glycolipids on cell surfaces.<sup>[3]</sup> Shiga and cholera toxin B-subunits as well as capsid lectins from SV40 and norovirus have been demonstrated to cluster glycosphingolipids, resulting in negative membrane curvature and formation of membrane invaginations.<sup>[4]</sup>

Multivalency of carbohydrates has been extensively studied, but not that of lectins, mainly because of the difficulty to control the number and location of binding sites in oligomeric proteins. An alternative approach is to use lectins presenting internal repeats such as  $\beta$ -propellers that are built on the cyclic repetition of 5 to 7  $\beta$ -sheets.<sup>[5]</sup> The bacterial lectin, RSL from *Ralstonia solanacearum*, is an ancestor of this fold, as it consists of a tandem repeat of two  $\beta$ -sheets that trimerizes as a six-bladed  $\beta$ -propeller (Figure 1 a,b).<sup>[6]</sup> The resulting hexavalent trimer has the capacity to invaginate lipid membranes containing fucosylated glycolipids.<sup>[7]</sup> Reducing the valency from six to three resulted in a trivalent lectin that lost the capacity to induce membrane invaginations in giant liposomes.<sup>[7]</sup> However, the influence of topology, that is, repartition of binding sites in space and the distance between them, has not been investigated.

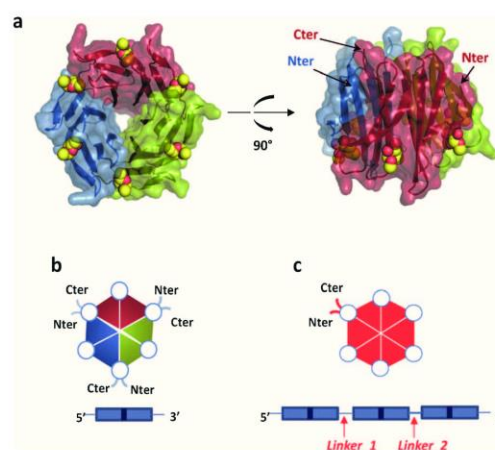
We therefore propose herein a novel approach for controlling valency. The “neoelectin” concept consists in the

[\*] J. Arnaud, Dr. A. Audfray, Dr. A. Varrot, Dr. A. Imberty  
CERMAV, CNRS and Grenoble Alpes Université  
38000 Grenoble (France)  
E-mail: imberty@cermav.cnrs.fr

K. Tröndle, Dr. J. Claudinon, Prof. W. Römer  
Institute of Biology II, Albert-Ludwigs-University Freiburg  
Schänzlestrasse 1, 79104 Freiburg (Germany)  
and  
BIOSS Centre for Biological Signalling Studies  
Albert-Ludwigs-University Freiburg  
Schänzlestrasse 18, 79104 Freiburg (Germany)  
E-mail: winfried.roemer@bioss.uni-freiburg.de

[\*\*] The authors acknowledge funding from Agence Nationale de la Recherche Grant NeoLect (ANR-11-BSV5-002) (W.R., A.A. A.V., and A.I.), from CNRS (A.I. and A.V.), and from Université Grenoble Alpes (J.A.). W.R. is supported by the Excellence Initiative of the German Research Foundation (EXC 294), by a grant from the Ministry of Science, Research and the Arts of Baden-Württemberg (Az: 33-7532.20) and by a starting grant of the European Research Council (Programme “Ideas”—call identifier: ERC-2011-StG 282105). The COST actions CM1102 and BM1003 and the Labex ARCANE (ANR-11-LABX-003) are thanked for support. The authors are grateful to the SOLEIL for provision of synchrotron radiation facilities and to the Proxima 1 beam line staff for efficient assistance.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201404568>.



**Figure 1.** a) Overall structure of wt-RSL complexed with  $\alpha$ MeFuc displayed as spheres (PDB code: 2BT9). Each monomer is represented by a different color. b) Representation of protein RSL and gene encoding RSL. The fucose binding sites are represented by white circle. c) Representation of protein neoRSL\_VI and gene encoding neoRSL\_VI.

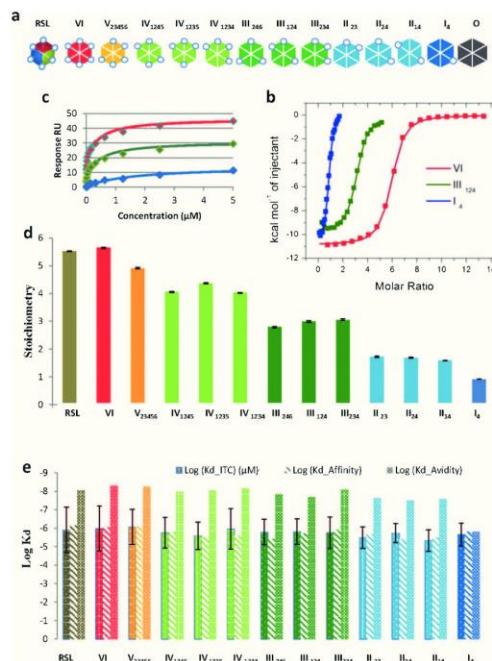
de novo creation of a monomeric hexavalent lectin adopting the  $\beta$ -propeller architecture. Engineering was performed at the genomic level by linking three divalent RSL monomers with a SSTVPGD peptide linker (Figure 1c). Differences introduced in gene design insure that all sites can be individually mutated in order to inactivate them, and allowed thereafter to control the number and position of the active binding sites. First, the monomeric neolectin with six active binding sites (referred to as neoRSL\_VI) was designed and led to a 273 amino acids protein, with six blades and two engineered linkers (Figure 1; Supporting Information, Figure S1 and Table S1). The resulting recombinant protein has the expected MW of 30 kDa and a denaturation temperature of 66° (Table 1; Supporting Information, and Figure S2) that is stabilized in the presence of fucose ( $T_m$  88°C), which is indicative of correct refolding with conserved ligand binding properties.

**Table 1:** Comparison of wild-type RSL and neoRSL\_VI.<sup>[a]</sup>

	wt-RSL	NeoRSL_VI
molecular weight	9.86	29.45
oligomerization state	trimer	monomer
thermal stability (TSA):		
$T_m$ apo protein (°C)	86 <sup>[a]</sup>	66
$T_m$ with fucose (°C)	96 <sup>[a]</sup>	88
affinity for fucose-chip (SPR):		
$K_{d,affinity}$ ( $\mu$ M)	0.7	0.9
$K_{d,avidity}$ (nM)	9	5
affinity for $\alpha$ MeFuc (ITC):		
$n$ (stoichiometry)	1.84 $\pm$ 0.01	5.65 $\pm$ 0.02
$K_d$ ( $\mu$ M)	1.23 $\pm$ 0.06	1.04 $\pm$ 0.06
$\Delta H$ (kJ mol <sup>-1</sup> )	-47.9 $\pm$ 0.2	-45.5 $\pm$ 0.2

[a] From Ref. [7].

NeoRSL\_VI was tested for its ability to bind to carbohydrates in solution and to glycosylated surfaces. Using ITC, titrating neoRSL\_VI with the monosaccharide ligand  $\alpha$ -methyl fucoside ( $\alpha$ MeFuc) (Figure 2b; Supporting Information, Figure S3 and Table S2) confirmed that all six binding sites of the neolectin were active ( $n = 5.6$ ). The affinity for  $\alpha$ MeFuc was in the micromolar range ( $K_d = 1.04 \mu\text{M}$ ), with no significant difference to wt-RSL ( $K_d = 1.23 \mu\text{M}$ ). Surface plasmon resonance data were measured using a surface of polyacrylamide (PAA)-decorated with  $\alpha$ -fucose residues (Figure 2c; Supporting Information, Figure S4 and Table S3) with different concentrations of lectins. Because of multivalent attachment, the binding curves were analyzed as a function of concentration (Supporting Information, Figure S4) using a two sites model that separates the affinity value at each site ( $K_{d,aff}$ ) and the bulk avidity for the surface ( $K_{d,av}$ ).<sup>[8]</sup> NeoRSL\_VI has single site affinity in the micromolar range ( $K_{d,aff} = 0.9 \mu\text{M}$ ), which is in excellent agreement with ITC data above, and a strong avidity for the glycosylated surface in nanomolar range ( $K_{d,av} = 5 \text{ nM}$ ). This 180-fold increase between the affinity for a single fucose and the avidity for multivalent fucose surface is the result of the clustering effect and is in the same range than previously



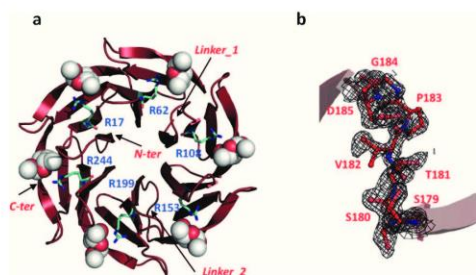
**Figure 2.** a) Representation of different neoRSLs with variation of the number of binding sites and position. Each active binding site is represented by a white circle. The proteins with the same valency for the valency, and arabic numbers in subscript for the numbering of the active sites. b) ITC measurement of integrated heat of binding when titrating neoRSL\_VI, III<sub>124</sub>, and I<sub>4</sub> (0.03 mM) with  $\alpha$ MeFuc (1.5 mM, 0.75 mM, and 0.25 mM, respectively) at 25 °C in 20 mM Tris pH 7.5 with 0.1 M NaCl. c) SPR measurement with steady state analysis of responses at the end of association phase for the binding of three neoRSLs (VI: red), III<sub>124</sub>: green and I<sub>4</sub>: blue) to a fucosylated-PAA chips. d) Stoichiometry of neoRSLs determined by ITC. e) Affinity determined by ITC and SPR and avidity determined by SPR with the two-site binding model.

obtained for wt-RSL ( $K_{d,aff} = 0.7 \mu\text{M}$  and  $K_{d,av} = 9 \text{ nM}$ )<sup>[7]</sup> confirming the integrity of neoRSL.

NeoRSL\_VI has been cocrystallized with  $\alpha$ MeFuc and the resulting crystal diffracted to 1.35 Å resolution (Supporting Information, Table S4). NeoRSL\_VI folded as the expected six-bladed  $\beta$ -propeller with six occupied fucose binding sites, indicating that the engineering approach has been successful (Figure 3a). Electron density of one of the linker engineered between blades 4 and 5 could be clearly identified (Figure 3b) and it adopted roughly the same shape as the natural linker (Supporting Information, Figure S5). The second linker is more flexible, but superimposition demonstrates that the engineered linker connects two  $\beta$ -strands while maintaining the position they adopt in the wild-type protein.

Using the differences in gene sequence introduced in the binding site, it was possible to mutate each of them independently. Taking into account the circular symmetry of the protein, 13 different architectures can be produced,





**Figure 3.** a) Overall structure of the neoRSL VI complexed with  $\alpha$ MeFuc displayed as spheres. The arginine residues that are mutated in the neoRSL valency mutants are displayed by sticks. b)  $2mF_o-DF_c$  electron density map for the engineered linker between blades 4 and 5 contoured at  $1\sigma$ .

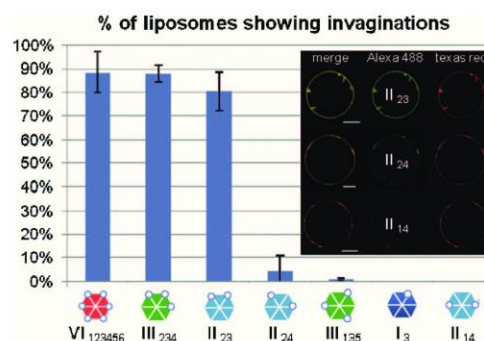
varying from valency zero to six, and presenting all possible arrangements of binding sites in space. Individual sites were inactivated by mutating the Arg residue previously demonstrated as crucial:<sup>[6,7]</sup> R17A, R62A, R108A, R153A, R199A, and R244A, for site 1 to 6, respectively. Double to sextuple Arg to Ala mutants were also prepared to obtain the 13 neolectins (Figure 2a). For nomenclature, NeoRSL\_II<sub>24</sub> is a divalent lectin with site 2 and 4 active, and site 1, 3, 5 and 6 inactivated by Arg to Ala mutation (Supporting Information, Figure S2). Denaturation temperature of neolectins varied from 66°C for neoRSL\_VI to 50.5°C for neoRSL\_II<sub>24</sub>, and stabilization by bound fucose ligand has been always observed.

The stoichiometry of all purified neolectins was verified by ITC (Figure 2; Supporting Information, Figure S3). Affinity for  $\alpha$ MeFuc in solution displayed limited variations, with  $K_d$  values between 0.85  $\mu$ M for neoRSL\_V<sub>12456</sub> and 4.67  $\mu$ M for neoRSL\_II<sub>14</sub> (Supporting Information, Table S2). All neolectins with valency ranging from 1 to 6 maintained the same high affinity towards fucose in solution and no correlation could be observed as a function of the number or the location of the active binding sites. As expected, the neolectin with all sites inactivated by the Arg to Ala mutation, neoRSL<sub>0</sub>, does not bind fucose efficiently anymore. However, a weak residual binding activity ( $K_d = 4$  mM) was observed, at the limit of the detection method. Since residual affinity is 2000 weaker than the normal one, it does not interfere with the characterization of the other mutants.

From SPR experiments (Figure 2; Supporting Information, Figure S3 and Table S3), the local affinity for fucose,  $K_{d,aff}$ , did not present significant variations as a function of valency with values ranging from 0.8  $\mu$ M for neoRSL V<sub>12456</sub> to 4  $\mu$ M for neoRSL\_II<sub>14</sub>, in excellent agreement with those determined by ITC (Figure 2e). The avidity constant,  $K_{d,av}$ , displayed more significant variations, with an almost linear decrease from 5 nm to 30 nm when the valency is varied from 6 to 2. Surprisingly, the variation was only rather limited. Only the monovalent receptor, that is, neoRSL\_I, displayed a sharp drop in binding with a micromolar affinity ( $K_d = 13.4$   $\mu$ M), corresponding to the loss of any avidity effect.

Neolectins with controlled valency and topology were then assayed in a synthetic membranous environment in order

to test their potential to induce membrane invaginations. For this, the fucosylated glycolipid DOPE-Lewis a (DOPE-Le<sup>a</sup>) was reconstituted in giant unilamellar vesicles (GUVs) with a typical diameter between 10 and 30  $\mu$ m (Supporting Information, Figure S6). As demonstrated previously, wt-RSL bound to Lewis a epitope (100%) on GUV surface and induced membrane invaginations in 93% of the cases ( $n = 280$ ).<sup>[7]</sup> NeoRSL\_VI displayed the same properties with formation of invaginations in 90% of liposomes ( $n = 100$ ) (Figure 4; Supporting Information, Figure S6). When the valency is decreased to 5 or 4 (Supporting Information,



**Figure 4.** Percentage of liposomes showing membrane invaginations in presence of different neoRSLs. Confocal sections of the equatorial plane of GUVs containing glycolipids in presence of Alexa488-labeled proteins are displayed as insert. Proteins are shown in green and the membrane marker (DOPE-TR) in red. Scale bars: 5  $\mu$ m.

Figure S6 and Table S5), the neolectins still bound efficiently to Le<sup>a</sup> epitope and induced invaginations. For trivalent RSL, different results were observed. NeoRSL\_III<sub>246</sub> only induced invaginations in 1% of studied liposomes, as previously observed with trivalent RSL mutant R17A.<sup>[7]</sup> However, other trivalent neolectins such as neoRSL\_III<sub>234</sub> and neoRSL\_III<sub>124</sub> with at least two neighboring binding sites were still able to induce invaginations in 88% and 70% of liposomes, respectively. The topology of binding sites therefore appears to have a crucial effect on glycolipid clustering and dynamics within membranes. Analyzing divalent neolectins with two binding sites confirmed this observation (Figure 4). Among the three possible topologies, only neoRSL\_II<sub>23</sub> was able to induce membrane invaginations (80%) whereas neoRSL\_II<sub>24</sub> (4%) and neoRSL\_II<sub>14</sub> (0%) were not at all efficient. To sum up, all Alexa488-labeled neolectins bound efficiently to GUVs, with the exception of neoRSL<sub>0</sub>, which has only residual affinity for fucose, but only the ones that present adjacent binding sites were able to induce invaginations.

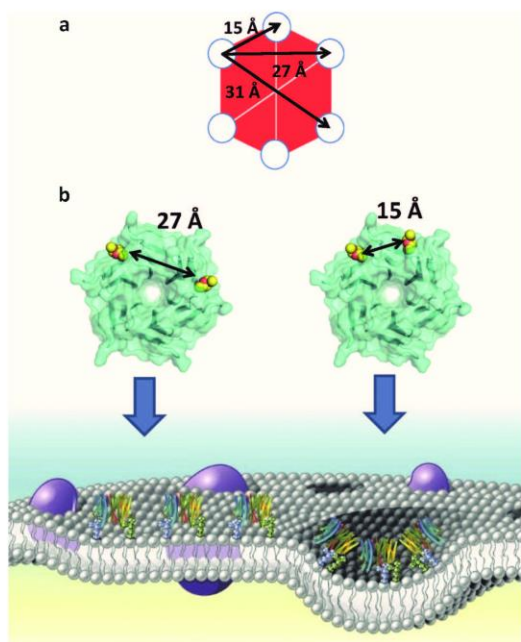
The process by which membrane glycosphingolipids trigger the internalization of bacterial and plant toxins and viruses through glycan binding has been the subject of few studies. Several pathways have been described and relevant biological functions were proposed to relate to cell internalization through vesicles formation.<sup>[9]</sup> The origin of the energy required for this physical process is not clear but has been proposed to arise from line tension generated by the

formation of lipid domains.<sup>[10]</sup> In the present study such mechanisms would imply that upon DOPE-Le<sup>a</sup> binding, the neolactins cluster on the membrane surface, triggering the formation of lipid domains. This would be in agreement with the patches observed on liposomes when lectin binds. We demonstrated herein that hexavalent to trivalent and even divalent lectins are able to cluster glycolipids and alter membrane shape, providing that the binding sites are close enough (Figure 5). The hypothesis that close proximity of

shed light on the evolution of these proteins, resulting from oligomerization or duplication as tandem repeat.<sup>[11]</sup> Finally, these considerations would be of high importance for engineering cargo proteins designed for importing drugs through the binding to glycosphingolipids, as already pioneered for Shiga toxins.<sup>[12]</sup>

Received: April 22, 2014  
Published online: July 7, 2014

**Keywords:** lectins · liposomes · membranes · multivalency · oligosaccharides



**Figure 5.** a) Representation of neoRSL\_VI with distances between fucose binding sites. b) Representation of the effect of the distance between two binding sites on the plasma membrane. When the distance between two binding sites exceeds 27 Å, only lectin binding occurs. However, when only 15 Å separates two binding sites the neolactins cluster glycolipids and bend the membrane.

carbohydrate binding sites is a prerequisite for membrane curvature by bringing lipids close to each other is in agreement with the topologies of RSL (distance of 15 Å). However the minimum distance for bending membrane may differ with the shape and curvature of the binding protein, as exemplified with viral capsids.<sup>[4a,d]</sup>

Our work has demonstrated that the capacity of multivalent lectins to induce membrane invagination is not directly correlated with avidity to the glycosylated surface. This may

- [1] a) M. Ambrosi, N. R. Cameron, B. G. Davis, *Org. Biomol. Chem.* **2005**, *3*, 1593–1608; b) J. Arnaud, A. Audfray, A. Imberty, *Chem. Soc. Rev.* **2013**, *42*, 4798–4813.
- [2] B. E. Collins, J. C. Paulson, *Curr. Opin. Chem. Biol.* **2004**, *8*, 617–625.
- [3] B. Belardi, G. P. O'Donoghue, A. W. Smith, J. T. Groves, C. R. Bertozzi, *J. Am. Chem. Soc.* **2012**, *134*, 9549–9552.
- [4] a) H. Ewers, W. Römer, A. E. Smith, K. Bacia, S. Dmitrieff, W. Chai, R. Mancini, J. Kartenbeck, V. Chambon, L. Berland, A. Oppenheim, G. Schwarzmann, T. Feizi, P. Schulle, P. Sens, A. Helenius, L. Johannes, *Nat. Cell Biol.* **2010**, *12*, 11–18; b) W. Römer, L. Berland, V. Chambon, K. Gaus, B. Windschiegl, D. Tenza, M. R. Aly, V. Fraissier, J. C. Florent, D. Perrais, C. Lamaze, G. Raposo, C. Steinem, P. Sens, P. Bassereau, L. Johannes, *Nature* **2007**, *450*, 670–675; c) W. Römer, L. L. Pontani, B. Sorre, C. Rentero, L. Berland, V. Chambon, C. Lamaze, P. Bassereau, C. Sykes, K. Gaus, L. Johannes, *Cell* **2010**, *140*, 540–553; d) G. E. Rydell, L. Svensson, G. Larson, L. Johannes, W. Römer, *Biochim. Biophys. Acta Biomembr.* **2013**, *1828*, 1840–1845.
- [5] C. K. M. Chen, N.-L. Chan, A. H. J. Wang, *Trends Biochem. Sci.* **2011**, *36*, 553–561.
- [6] N. Kostlanová, E. P. Mitchell, H. Lortat-Jacob, S. Oscarson, M. Lahmann, N. Gilboa-Garber, G. Chambat, M. Wimmerová, A. Imberty, *J. Biol. Chem.* **2005**, *280*, 27839–27849.
- [7] J. Arnaud, J. Claudinon, K. Tröndle, M. Trovaslet, G. Larson, A. Thomas, A. Varrot, W. Römer, A. Imberty, A. Audfray, *ACS Chem. Biol.* **2013**, *8*, 1918–1924.
- [8] E. M. Munoz, J. Correa, E. Fernandez-Megia, R. Riguera, *J. Am. Chem. Soc.* **2009**, *131*, 17765–17767.
- [9] a) J. A. Cho, D. J. Chinnapen, E. Aamar, Y. M. Te Welscher, W. I. Lencer, R. Massol, *Front. Cell. Infect. Microbiol.* **2012**, *2*, 51; b) T. Eierhoff, B. S. W. Römer, in *Molecular Regulation of Endocytosis* (Ed.: B. Ceresa), InTech, **2012**; c) H. Ewers, A. Helenius, *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, a004721; d) K. Sandvig, J. Bergan, S. Kavaliauskiene, T. Skotland, *Prog. Lipid Res.* **2014**, *54C*, 1–13.
- [10] T. Baumgart, S. T. Hess, W. W. Webb, *Nature* **2003**, *425*, 821–824.
- [11] I. Yadid, N. Kirshenbaum, M. Sharon, O. Dym, D. S. Tawfik, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 7287–7292.
- [12] a) N. Engedal, T. Skotland, M. L. Torgersen, K. Sandvig, *Microb. Biotech.* **2011**, *4*, 32–46; b) L. Johannes, W. Römer, *Nat. Rev. Microbiol.* **2010**, *8*, 105–116.



Supporting Information

© Wiley-VCH 2014

69451 Weinheim, Germany

**Membrane Deformation by Neoelectins with Engineered Glycolipid Binding Sites\*\***

*Julie Arnaud, Kevin Tröndle, Julie Claudinon, Aymeric Audfray, Annabelle Varrot, Winfried Römer,\* and Anne Imberty\**

anie\_201404568\_sm\_miscellaneous\_information.pdf



**Experimental procedures.**

**Cloning of neolectins and production of recombinant proteins.** A nucleotide sequence coding for the neoRSL\_VI was synthesized (GenScript, Piscataway, NJ). It was introduced into the expression vector pET-TEV using NdeI and XhoI restriction sites<sup>[1]</sup>. The pET-TEV-neoRSL constructs were transformed into *E. coli* TUNER<sup>TM</sup> (DE3) strain (Novagen). Cells harboring the plasmids were grown in LB broth containing 100 µg ml<sup>-1</sup> kanamycin at 37°C until the optical density at 600 nm reached 0.6. For neoRSL\_VI, the expression was performed by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside during 3h at 30°C. For other neolectins, the cells were cultured overnight at 16°C after the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside. After induction, bacterial cells were harvested by centrifugation at 6,000 g for 20 min at 4 °C and resuspended with buffer containing 20 mM Tris/HCl, pH 7.5, 100 mM NaCl, 20 mM imidazole. NeoRSL was purified by affinity chromatography on 1 ml His-Trap<sup>TM</sup> column (GE Healthcare). After extensive washings (20 mM Tris/HCl, pH 7.5, 100 mM NaCl, 20 mM imidazole) the protein was eluted in presence of 300 mM imidazole. After this first step of purification, protein was dialyzed against 20 mM Tris/HCl, pH 8, 100mM NaCl. After 2 days, the His-tag was cleaved using 1 mg of TEV protein per 100 mg of neoRSL overnight at 4°C. The cleaved protein was purified by affinity chromatography on a His-Trap<sup>TM</sup> column (GE Healthcare). The flowthrough containing the neoRSL pure protein was collected. NeoRSL\_VI and neoRSL\_V were dialyzed against ultrapure water prior to lyophilization. The other neoRSLs were dialyzed against 20 mM Tris/HCl, pH 7.5, 100 mM NaCl and stocked at -20°C.

**Thermal shift analysis.** Thermal shift assays were performed using a Mini Opticon, Real Time PCR machine (Bio-Rad Laboratories) with 1 mg ml<sup>-1</sup> of proteins diluted in 100 mM NaCl, 20 mM Tris pH 7.5, with or without 5 mM fucose in a total reaction volume of 25 µl. SYPRO Orange (Molecular Probes, CA) was used as a fluorescent probe detected at 530 nm. The temperature was raised using 1°C/minute steps from 25°C to 100°C and fluorescence readings were taken at each interval. A positive ΔT<sub>m</sub> value indicates that the ligand stabilizes the protein from denaturation, and therefore binds to the protein. A minimum of two independent assays were performed for each protein.

**Isothermal Titration Calorimetry Experiments.** ITC experiments were performed with an ITC200 isothermal titration calorimeter (Microcal-GE Healthcare). The experiments were carried out at 25°C. Lectins and αMeFuc were dissolved in the same buffer composed of 20 mM Tris with 0.1 M NaCl at

pH 7.5. A total of 20 injections of 2  $\mu\text{l}$  of sugar solution (varying from 0.5 to 50 mM) were added at intervals of 120 s whilst stirring at 1000  $\text{rev}\cdot\text{min}^{-1}$  in the 0.03 mM protein solution in the 200  $\mu\text{l}$  cell. Control experiments performed by injection of buffer into the protein solution yielded insignificant heats of dilution. The experimental data were fitted to a theoretical titration curve using Origin software supplied by Microcal, with  $\Delta H$  (enthalpy change),  $K_a$  (association constant) and  $n$  (number of binding sites per monomer) as adjustable parameters. Free energy change ( $\Delta G$ ) and entropy contributions ( $T\Delta S$ ) were derived from the equation  $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$  (with  $T$  as the absolute temperature and  $R = 8.314 \text{ J mol}^{-1}\text{K}^{-1}$ ).

**Crystallography.** Crystals of NeoRSL\_VI in complex with  $\alpha\text{MeFuc}$  were obtained by the hanging drop vapor diffusion method by mixing 1  $\mu\text{l}$  of protein at 10  $\text{mg ml}^{-1}$  with 10  $\mu\text{M}$  of  $\alpha\text{MeFuc}$  and 1  $\mu\text{l}$  of reservoir solution. Crystals appeared after 3 weeks of incubation at 19°C with the solution 4 of the Crystal Screen 1 (Hampton Research) containing 2 M Ammonium sulfate and 100 mM Tris/HCl pH 8.5. X-ray diffraction data were collected on Proxima 1 beamline at the Soleil synchrotron, Saint Aubin, France using a Pilatus detector. Data were processed using MOSFLM<sup>[2]</sup> and programs from the CCP4 suite<sup>[3]</sup> were used in further computing. Phases were obtained by the molecular replacement method using Phaser<sup>[4]</sup> and the coordinates of the trimer of 2BT9 as search model. Maximum likelihood restrained and anisotropic refinement using REFMAC 5.7 with 5% of the data retained for cross-validation purposes was combined with manual model adjustment and electron density inspection performed in Coot graphics<sup>[5]</sup>.

Two independent chains were found in the asymmetric unit and since both molecules are very similar (rms of 0.29  $\text{\AA}$ ), only chain A is described in the main manuscript. When comparing neoRSL\_VI and wt-RSL structures, a rms value of 0.44  $\text{\AA}$  was obtained for 266 amino acids, indicating that the blades of the monomeric neolectin are at the same position than those of the trimeric wild-type lectin.

**Surface Plasmon Resonance.** SPR experiments were performed on a Biacore X100 instrument (GE Healthcare) at 25°C in HBS (10 mM Hepes/NaOH, pH 7.5, 150 mM NaCl, 0.05% Tween 20) at a flow rate of 10  $\mu\text{l min}^{-1}$ . Streptavidin was immobilized on a research grade CM5 chip using standard procedures. 250 resonance units of a mix containing 10% of biotinylated polyacrylamide (Lectinity) bearing  $\alpha$ -fucose and 90% bearing  $\beta$ -galactose were captured on channel 2 at 200  $\mu\text{g ml}^{-1}$ . The same amount of biotinylated polyacrylamide bearing  $\beta$ -galactose was captured on channel 1.



Experiments consist of injection (association 360 s, dissociation 400 s) of various concentrations of neoRSLs (2-fold cascade dilutions, from 0 to 10  $\mu\text{M}$ ) on both channels. The chip was fully regenerated by one injection of 1M fucose in running buffer for 120 s. Binding was measured as resonance units over time after blank subtraction, and data were then evaluated by using the Biacore X100 evaluation software, version 2.0. Plots represent the subtracted data (channel 2-channel 1). Dissociation constants were determined by plotting response at equilibrium (Req, 10 s before the end of injection) against lectin concentration. The clustering effect that reflects the importance of multivalent interactions is also provided. Theoretically, a clustering effect higher than one means that multivalent interactions occur during the binding event.

In addition to this low density chip (10% fucose), a high density chip (100% fucose) was also constructed. Both chips were tested with RSL and the results were similar on both chips with a tendency for higher affinity on the high density chip, certainly due to rebinding events. For complete analysis, low density chip was selected as, in SPR techniques, reducing the density of ligand at the surface of the chip decreases the experiment artifacts like mass transport limitations or steric hindrance.

**Production of GUVs and Lectin Binding.** The lipids 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) and cholesterol were purchased from Avanti Polar Lipids. Texas Red® 1,2-Dihexadecanoyl-*sn*-Glycero-3-Phosphoethanolamine Triethylammonium Salt (Texas Red® DHPE) was from Molecular Probes (Life Technologies) and FSL-Lewis a from KODE Biotech Limited. FSL-Lewis a is comprised of the Lewis blood group Lewis a trisaccharide  $\text{Gal}\beta 1-3(\text{Fuc}\alpha 1-4)\text{GlcNAc}\beta$  representing F, conjugated via an  $\text{O}(\text{CH}_2)_3\text{NH}$  spacer (S) to an activated adipate derivative of dioleoyl-phosphatidyl-ethanolamine (DOPE) representing L. Final composition includes DOPC (64.75 mol%), Texas Red® DHPE (0.25 mol%), cholesterol (30 mol%) and the fucosylated lipid DOPE-Le<sup>a</sup> (5 mol%).

GUVs were grown at room temperature using the electroformation technique on indium-tin oxide (ITO)-coated glass slides. Lipid mixtures were dissolved in chloroform at a final concentration of 0.5  $\text{mg ml}^{-1}$  and 15  $\mu\text{l}$  of solution was spread on the conductive faces of ITO slides. After at least 2 h of drying under vacuum, GUVs were grown in a 290 mOsm sucrose solution by applying an alternating

electric field from 20 mV to 1.1 V for 15 h. GUVs were imaged on a confocal microscope (Nikon Eclipse Ti-E with A1R confocal laser scanner, 60x oil objective, NA=1.49). Image acquisition and analysis were performed with NIS-Elements (Nikon).

NeoRSLs were labeled with Alexa Fluor 488 Carboxylic Acid (Invitrogen) according to the protocol from the manufacturer. The amount of covalently bound dye was evaluated by spectrophotometer reading absorbance to be one dye molecule per lectin monomer, with no variations between the different proteins. Inhibition of hemagglutination was used to check that the activity of labeled RSL is not affected by labeling.

### References

- [1] K. Houben, D. Marion, N. Tarbouriech, R. W. Ruigrok, L. Blanchard, *Journal of virology* **2007**, *81*, 6807-6816.
- [2] A. W. Leslie, H. Powell, in *Evolving Methods for Macromolecular Crystallography* (Eds.: R. Read, J. Sussman), Springer Netherlands, **2007**, pp. 41-51.
- [3] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. Leslie, A. McCoy, *et al.*, *Acta Crystallogr., Sect D: Biol. Crystallogr.* **2011**, *67*, 235-242.
- [4] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658-674.
- [5] P. Emsley, B. Lohkamp, W. Scott, K. Cowtan, *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 486-501.

**Table 1S:** Primers used for site-directed mutagenesis.

<b>Mutations</b>	<b>Primers</b>
R17A	For : 5'-TTCGTGGGGAACACTGTTCTTCGATCGCTGTTTACACGGCCAAT-3'
	Rev: 5'-ATTGGCCGTGTAAACAGCGATCGAAGGAACAGTTCCCCACGAA-3'
R62A	For : 5'-TATCCATATCGCTGTCTATGCTAGTACTGGTACTACGACCACG-3'
	Rev: 5'-CGTGGTCGTAGTACCAGTACTAGCATAGACAGCGATATGGATA-3'
R108A	For : 5'-CAGCTGGGGCACCGTGCCGAGCATTGCTGTGTATACCGCGAAC-3'
	Rev: 5'-GTTTCGCGGTATACACAGCAATGCTCGGCACGGTGCCCCAGCTG-3'
R153A	For : 5'-CAGCGCGATTCATATTGCTGTGTATGCGAGCACCGGCACC-3'
	Rev: 5'-GGTGCCGGTGCTCGCATACACAGCAATATGAATCGCGCTG-3'
R199A	For : 5'-AAGCTGGGGAACAGTACCAAGCATCGCTGTATAACAGCAAAT-3'
	Rev: 5'-ATTTGCTGTGTATAACAGCGATGCTTGGTACTGTTCCCCAGCTT-3'
R244A	For : 5'-AAGCGCAATCCACATCGCTGTATACGCAAGCACAGGAACA-3'
	Rev: 5'-TGTTCCCTGTGCTTGCGTATAACAGCGATGTGGATTGCGCTT-3'

**Table 2S:** Titration microcalorimetry data for the interaction between neoRSL and  $\alpha$ -methyl fucoside. All data have been measured at least twice and standard deviations are indicated.

Protein	n	Kd ( $\mu$ M)	$-\Delta G$ (kJ mol <sup>-1</sup> )	$-\Delta H$ (kJ mol <sup>-1</sup> )	TAS (kJ mol <sup>-1</sup> )
wt-RSL	5.53 $\pm$ 0.01 <sup>a</sup>	1.23 $\pm$ 0.06	33.7	47.90 $\pm$ 0.2	-14.18
VI	5.65 $\pm$ 0.02	1.04 $\pm$ 0.06	34.1	45.5 $\pm$ 0.2	-11.36
V <sub>12456</sub>	4.91 $\pm$ 0.03	0.85 $\pm$ 0.11	34.6	47.8 $\pm$ 0.44	-13.17
IV <sub>1245</sub>	4.06 $\pm$ 0.02	1.76 $\pm$ 0.15	32.8	37.0 $\pm$ 0.29	-4.17
IV <sub>1235</sub>	4.36 $\pm$ 0.03	2.56 $\pm$ 0.18	31.9	44.6 $\pm$ 0.35	-12.70
IV <sub>1234</sub>	4.03 $\pm$ 0.01	1.1 $\pm$ 0.08	34.0	44.8 $\pm$ 0.24	-10.80
III <sub>246</sub>	2.79 $\pm$ 0.03	1.62 $\pm$ 0.20	33.0	49.1 $\pm$ 0.66	-16.07
III <sub>124</sub>	2.99 $\pm$ 0.03	1.53 $\pm$ 0.20	33.2	40.02 $\pm$ 0.49	-6.84
III <sub>234</sub>	3.05 $\pm$ 0.02	1.75 $\pm$ 0.14	32.84	38.2 $\pm$ 0.31	-5.36
II <sub>23</sub>	1.73 $\pm$ 0.02	3.26 $\pm$ 0.26	31.3	49.57 $\pm$ 0.63	-18.27
II <sub>24</sub>	1.69 $\pm$ 0.03	1.84 $\pm$ 0.31	32.7	50.70 $\pm$ 1.16	-17.98
II <sub>14</sub>	1.59 $\pm$ 0.01	4.67 $\pm$ 0.26	33.1	40.76 $\pm$ 0.46	-7.68
I <sub>2</sub>	0.91 $\pm$ 0.01	2.23 $\pm$ 0.24	32.2	52.75 $\pm$ 1.06	-20.51
0	6 <sup>b</sup>	4093 $\pm$ 218	13.6	4.92 $\pm$ 0.16	8.70

<sup>a</sup> for a trimer of wt-RSL, <sup>b</sup> Value of n fixed during the fitting procedure for this low affinity mutant

**Table 3S:** Surface plasmon resonance data for the interaction between neoRSL and  $\alpha$ -fucose-presenting chips. The clustering effect is defined as  $Kd_{aff}/Kd_{av}$ .

Protein	$Kd_{aff}$ ( $\mu$ M)	Rmax1	$Kd_{av}$ (nM)	Rmax2	Chi2	Clustering effect
wtRSL	0.7	29.6	9	16.1	1.0	78
VI	0.9	27.19	5	22.50	0.635	180
V <sub>12456</sub>	0.8	27.49	5.5	16.44	0.518	145
IV <sub>1245</sub>	1.6	23.52	10.8	15.07	0.22	148
IV <sub>1235</sub>	3	36.92	8.8	16.19	0.722	340
IV <sub>1234</sub>	2.5	34.42	7.1	16.90	0.939	352
III <sub>246</sub>	3.8	27.69	13.4	6.4	0.405	283
III <sub>124</sub>	1.9	21.88	20	14.73	0.157	95
III <sub>234</sub>	1.2	27.13	8.1	17.19	0.548	148
II <sub>23</sub>	2.1	21.96	22.6	14.28	0.33	92
II <sub>24</sub>	4	16.92	32	4.4	0.03	125
II <sub>14</sub>	3.2	22.01	27	12.2	0.27	119
I <sub>2</sub>	1.5	13.38			0.274	

Rmax1: Theoretical maximal SPR response due to monovalent binding (affinity)

Rmax2: Theoretical maximal SPR response due to multivalent binding (avidity)

Chi2: Deviation between experimental data and mathematical model in the equilibrium responses against concentrations curves.

**Table 4S:** Crystal structure of neoRSL\_VI. Details of collection data and statistics.

NeoRSL_VI	
<b>Data collection</b>	
Beamline	Proxima 1
Space group	P2 <sub>1</sub>
Cell dimensions a, b, c (Å)	36.42 75.91 85.78
Cell dimensions $\alpha$ , $\beta$ , $\gamma$ (degrees)	90 102.2 90
Resolution (outer shell) (Å)	41.92-1.35 (1.42-1.35)
Measured/Unique reflections	317694/99018
Average multiplicity (outer shell)	3.2 (2.8)
Rmerge (outer shell)	0.040 (0.146)
Completeness (%) (outer shell)	99.1 (95.5)
Mean I/oI (outer shell)	14.8 (5.6)
Wilson B	8
<b>Refinement</b>	
Rcryst/Rfree	11.5/14.9
r.m.s.d. bonds (Å)	0.014
r.m.s.d. angles (°)	1.619
r.m.s.d. chiral (Å <sup>3</sup> )	0.099
Protein atoms (per chain)	2035 / 2047
Bfac (Å <sup>2</sup> )	12.80 / 12.86
Ligand atoms	72 / 72
Bfac (Å <sup>2</sup> )	9.90 / 9.75
Water molecules	339 / 314
Bfac (Å <sup>2</sup> )	26.62 / 26.41
PDB code	4CSD

**Table 5S:** Observations of membrane tubules on DOPE-Lewis a-containing GUVS.

Protein	Vesicles with tubules	Std. Deviation
wt-RSL <sup>a</sup>	93%	
VI	90%	± 5%
V <sub>12456</sub>	80%	± 0%
IV <sub>1245</sub>	80%	± 0%
IV <sub>1235</sub>	86%	± 0%
IV <sub>1234</sub>	65%	± 3.54%
III <sub>246</sub>	1%	± 0.35%
III <sub>124</sub>	70%	± 0%
III <sub>234</sub>	88%	± 1.73%
II <sub>23</sub>	80%	± 5%
II <sub>24</sub>	4%	± 3.33%
II <sub>14</sub>	0%	± 0%
I <sub>2</sub>	1%	± 0.35%
0	0%	± 0%

<sup>a</sup> From Arnaud et al., ACS Chem. Biol. 2013.



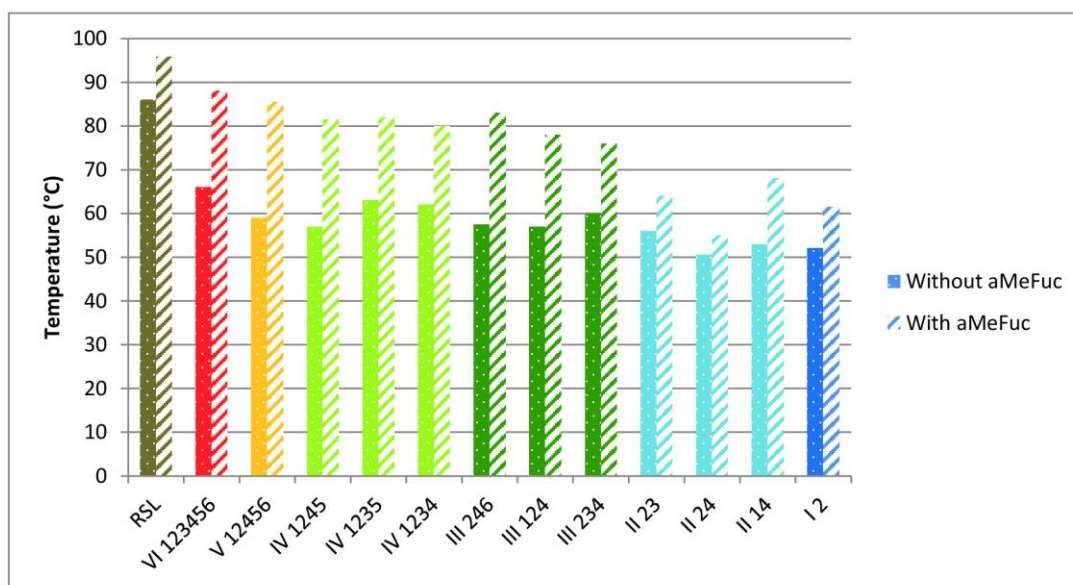
**Figure 1S:** DNA and protein sequence of neoRSL\_VI. Linkers are colored in red and arginine residues that can be mutated for invalidation of each of the six binding sites are colored in blue.

```

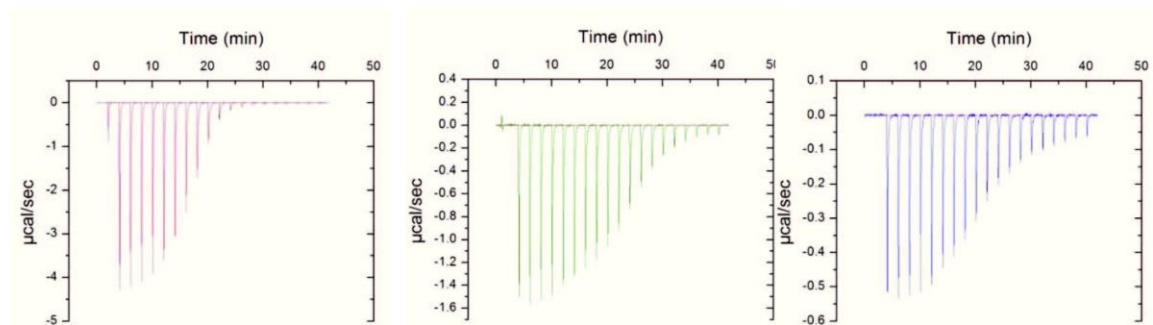
M S S V Q T A A T S W G T V P S I R V Y T
ATG TCG AGC GTT CAG ACC GCT GCC ACT TCG TGG GGA ACT GTT CCT TCG ATC CGT GTT TAC ACG
A N N G K I T E R C W D G K G W Y T G A F
GCC AAT AAT GGT AAG ATC ACT GAG CGT TGT TGG GAC GGG AAG GGG TGG TAC ACT GGT GCC TTC
N E P G D N V S V T S W L V G S A I H I R
AAT GAG CCG GGT GAT AAT GTC TCC GTG ACT AGT TGG CTT GTC GGT AGT GCT ATC CAT ATC CGT
V Y A S T G T T T T E W C W D G N G W T K
GTC TAT GCT AGT ACT GGT ACT ACG ACC ACG GAG TGG TGT TGG GAC GGT AAT GGT TGG ACT AAG
G A Y T S S T V P G D Q T A A T S W G T V
GGT GCC TAC ACT AGC AGC ACC GTG CCG GGC GAT CAG ACG GCG GCG ACC AGC TGG GGC ACC GTG
P S I R V Y T A N N G K I T E R C W D G K
CCG AGC ATT CGC GTG TAT ACC GCG AAC AAC GGC AAA ATT ACC GAA CGC TGC TGG GAT GGC AAA
G W Y T G A F N E P G D N V S V T S W L V
GGC TGG TAT ACC GGC GCG TTT AAC GAA CCG GGC GAT AAC GTG AGC GTG ACC AGC TGG CTG GTG
G S A I H I R V Y A S T G T T T T E W C W
GGC AGC GCG ATT CAT ATT CGC GTG TAT GCG AGC ACC GGC ACC ACC ACC ACC GAA TGG TGC TGG
D G N G W T K G A Y T S S T V P G D Q T A
GAT GGC AAC GGC TGG ACC AAA GGC GCG TAT ACC AGC AGC ACA GTA CCA GGA GAC CAA ACA GCA
A T S W G T V P S I R V Y T A N N G K I T
GCA ACA AGC TGG GGA ACA GTA CCA AGC ATC CGG GTA TAC ACA GCA AAT AAC GGA AAA ATC ACA
E R C W D G K G W Y T G A F N E P G D N V
GAA CGG TGC TGG GAC GGA AAA GGA TGG TAC ACA GGA GCA TTC AAC GAG CCA GGA GAC AAC GTA
S V T S W L V G S A I H I R V Y A S T G T
TCT GTA ACA AGC TGG TTA GTA GGA AGC GCA ATC CAC ATC CGG GTA TAC GCA AGC ACA GGA ACA
T T T E W C W D G N G W T K G A Y T A T N
ACA ACA ACA GAA TGG TGC TGG GAC GGA AAC GGA TGG ACA AAA GGA GCA TAC ACA GCA ACA AAC
STOP
TGA

```

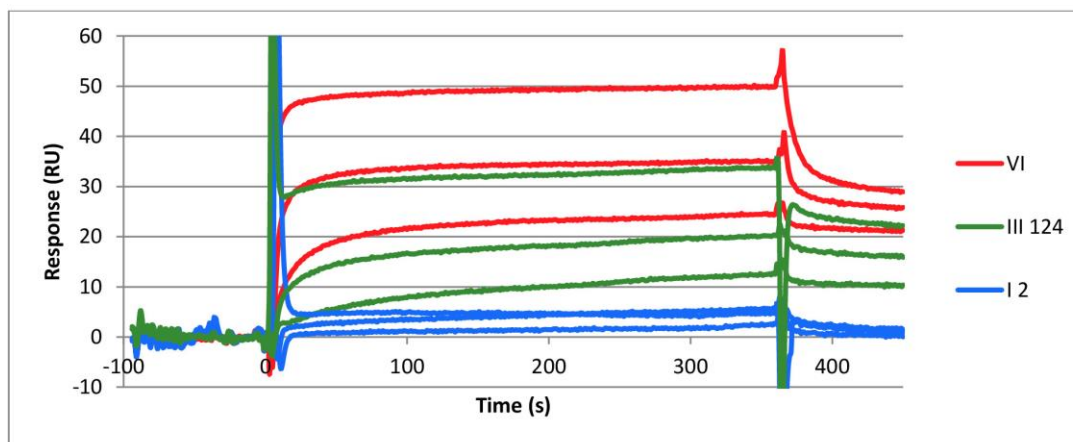
**Figure 2S:** Variations of T<sub>m</sub> of wt-RSL and neolectins as determined by Thermal Shift Assays in the presence or absence of  $\alpha$ -methyl-fucoside.



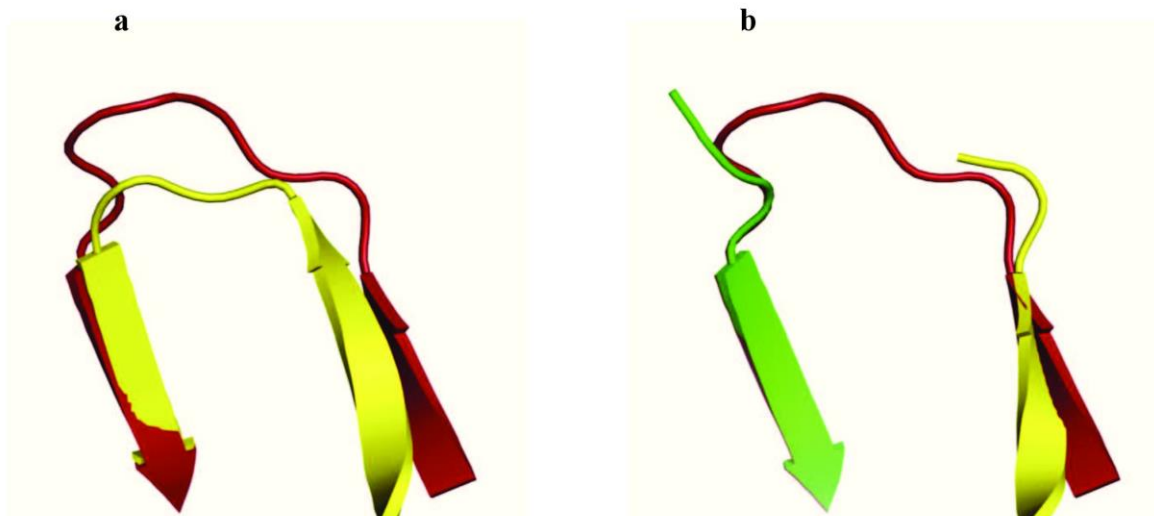
**Figure 3S:** Thermograms obtained for the titration of VI (left), III<sub>124</sub> (middle) and I<sub>2</sub> (right) by 1.5 mM, 0.75 mM and 0.25 mM  $\alpha$ MeFuc respectively.



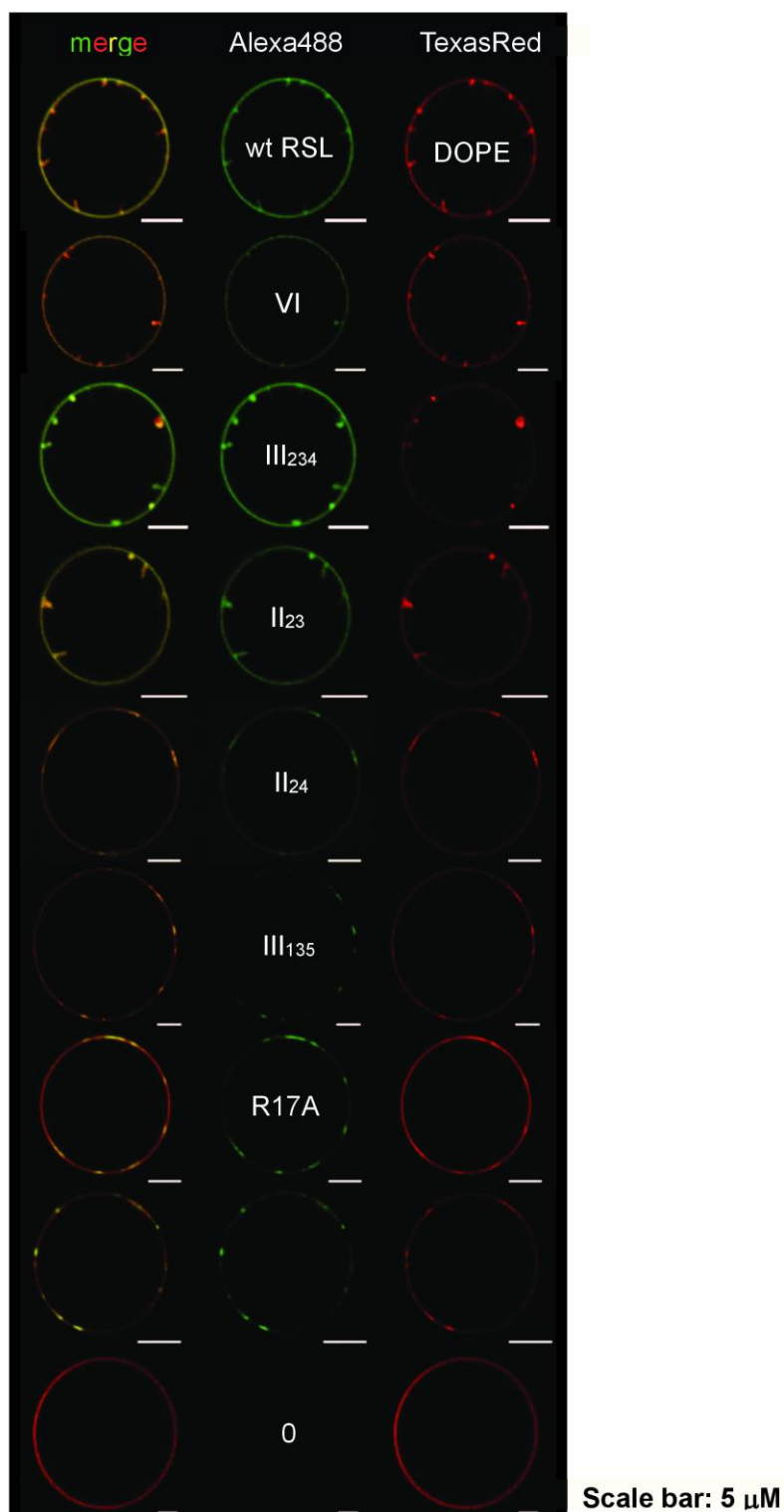
**Figure 4S:** Sensorgrams of the binding of neoRSL\_VI (red), III<sub>124</sub> (green) and I<sub>2</sub> (blue) to a fucosylated chip. For each protein, three concentrations have been used (5  $\mu$ M, 0.313  $\mu$ M and 0.078  $\mu$ M from bottom to top).



**Figure 5S:** (a) comparison of peptides conformation between the engineered linker (red: amino acids 179-185 between blades 4 and 5) and the natural one (yellow: amino acids 44-47 between blades 1 and 2). (b) Comparison between engineered linker (red) and position of N-ter and C-ter extremities of adjacent monomers (yellow and green) in wt-RSL.



**Figure 6S:** Confocal sections of the equatorial plane of GUVs containing DOPE-Lewis a glycolipid in presence of Alexa488-labeled wt-RSL, mutant and neoRSLs. All acquisitions were performed with the same or slightly adapted excitation and emission gains.







## II. Changement de spécificité

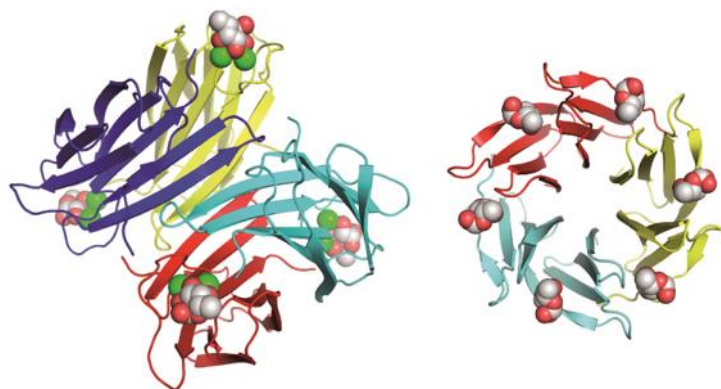
### II.1. Validation de la modélisation pour la spécificité des lectines bactériennes

#### II.1.1. Introduction

Les progrès récents sur la capacité des lectines à reconnaître spécifiquement des glycoconjugués à la surface de cellules en font des outils de choix dans le diagnostic biomédical associés à des changements de glycosylation (inflammation, du cancer ...). Les lectines peuvent également être utilisées en recherche pour étudier le glycome et aider à la compréhension du glycode. Toutefois, un nombre restreint de lectines sont actuellement produites sous forme recombinante, limitant la qualité et la quantité de lectines disponibles et réduisant leur utilisation dans les biotechnologies et la recherche. Une des solutions envisagées est la création d'une bibliothèque de lectines de spécificité contrôlée. Pour ce faire une connaissance approfondie des lectines est importante.

La modélisation est une étape importante dans une telle approche afin de concevoir les futurs mutants présentant un changement de spécificité. Une étape de validation de l'outil de modélisation est nécessaire.

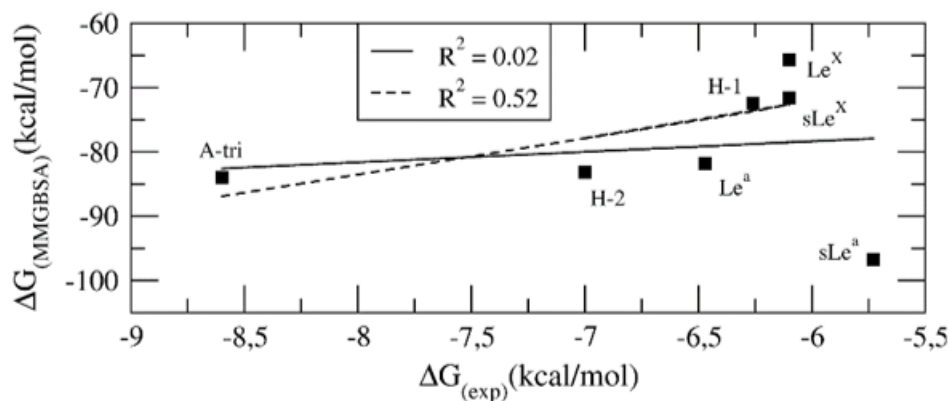
A partir de deux lectines bactériennes présentant différents sites de liaison au fucose, nous avons combiné une analyse à haut débit de puces à glycanes avec une approche à moyen débit d'amarrage moléculaire à partir d'une bibliothèque de structures d'oligosaccharides. Les résultats ont été validés par cristallographie pour les structures et par ITC pour les énergies d'interaction. Les lectines LecB de *Pseudomonas aeruginosa* et de BambL de *Burkholderia ambifaria* font partie des facteurs de virulence utilisés par les bactéries pathogènes afin d'envahir la cellule cible. Ces deux lectines reconnaissent les oligosaccharides fucosylés tels que les épitopes des groupes sanguins ABH (O) et les épitopes Lewis. Ces deux lectines présentent une structure symétrique et très stable qui faciliterait leur modification de spécificité (Figure 47).



**Figure 47 : Structure cristallographique des protéines LecB et BamBL**

La structure de gauche représente LecB qui est un homotétramère (chaque monomère est représenté d'une couleur différente) présentant 4 sites de fixation pour le fucose (Code PDB 1OXC (Loris et al., 2003)). La structure de droite représente BamBL qui est très proche de RSL. Il s'agit d'un homotrimer (chaque monomère est représenté d'une couleur différente) présentant 6 sites de fixation pour le fucose (Code PDB 3ZW2 (Audfray et al., 2012) (Topin et al., 2013)).

Les spécificités de ces lectines ont été caractérisées en utilisant des données semi-quantitatives des puces à glycanes rationalisées par amarrage moléculaire avec le logiciel Glide. Des prédictions fiables de structures de protéines / d'oligosaccharides peuvent être obtenues et validées par les structures cristallines de complexes existants. Au cours de l'étude, il a été prédit que les oligosaccharides fucosylés sont reconnus par les lectines dans leur conformation de plus basse énergie, avec l'exception de l'épitope Lewis x qui est fortement déformé lors de son interaction avec BamBL. La structure cristalline du complexe BamBL/ Lewis x résolue à 1,6 Å de résolution, confirme que cet oligosaccharide adopte une conformation de haute énergie dans le site de reconnaissance. De plus, une bonne corrélation a pu être observée entre les énergies de liaison obtenues par modélisation par la méthode MMGBSA et les valeurs expérimentales d'énergie libre de liaison que nous avons eu par ITC (Figure 48). Un protocole prédictif a donc pu être proposé pour rationaliser de grands ensembles de données obtenues par les puces à glycanes. Ce protocole pourra être ensuite utilisé pour la prédiction de la spécificité des mutants des protéines.



**Figure 48 : Corrélation entre l'énergie libre mesurée expérimentalement et celle calculée par modélisation moléculaire pour la protéine BambL**

$\Delta G_{(MMGBSA)}$  correspond à l'énergie libre calculée par modélisation moléculaire et  $\Delta G_{(exp)}$  correspond à l'énergie libre déterminée expérimentalement par ITC (Topin et al., 2013).

### II.1.2. Décrypter la préférence glycanes des lectines bactériennes par amarrage moléculaire

Dans cet article, j'ai réalisé plusieurs ITC afin de caractériser expérimentalement les protéines avec l'aide d'Emilie Gillon. Le Dr. Anita Sarkar et moi-même avons réalisé les premiers calculs d'amarrage moléculaire sur LecB et BambL, respectivement. Ces calculs ont été ensuite repris et améliorés par le Dr. Jérémie Topin et le Dr. Aline Thomas.

# Deciphering the Glycan Preference of Bacterial Lectins by Glycan Array and Molecular Docking with Validation by Microcalorimetry and Crystallography

Jeremie Topin<sup>1,2</sup>, Julie Arnaud<sup>1</sup>, Anita Sarkar<sup>1</sup>, Aymeric Audfray<sup>1</sup>, Emilie Gillon<sup>1</sup>, Serge Perez<sup>1</sup>, Helene Jamet<sup>2</sup>, Annabelle Varrot<sup>1</sup>, Anne Imberty<sup>1\*</sup>, Aline Thomas<sup>1</sup>

**1** CERMAV- Centre national de la recherche scientifique UPR5301 (affiliated to Université Joseph Fourier and ICMG), BP53, 38041 Grenoble, France, **2** Département de Chimie Moléculaire, UMR- Centre national de la recherche scientifique 5250 & ICMG FR 2607, Université Joseph Fourier, BP 53, 38041 Grenoble, France

## Abstract

Recent advances in glycobiology revealed the essential role of lectins for deciphering the glycode by specific recognition of carbohydrates. Integrated multiscale approaches are needed for characterizing lectin specificity: combining on one hand high-throughput analysis by glycan array experiments and systematic molecular docking of oligosaccharide libraries and on the other hand detailed analysis of the lectin/oligosaccharide interaction by x-ray crystallography, microcalorimetry and free energy calculations. The lectins LecB from *Pseudomonas aeruginosa* and BambL from *Burkholderia ambifaria* are part of the virulence factors used by the pathogenic bacteria to invade the targeted host. These two lectins are not related but both recognize fucosylated oligosaccharides such as the histo-blood group oligosaccharides of the ABH(O) and Lewis epitopes. The specificities were characterized using semi-quantitative data from glycan array and analyzed by molecular docking with the Glide software. Reliable prediction of protein/oligosaccharide structures could be obtained as validated by existing crystal structures of complexes. Additionally, the crystal structure of BambL/Lewis x was determined at 1.6 Å resolution, which confirms that Lewis x has to adopt a high-energy conformation so as to bind to this lectin. Free energies of binding were calculated using a procedure combining the Glide docking protocol followed by free energy rescoring with the Prime/Molecular Mechanics Generalized Born Surface Area (MM-GBSA) method. The calculated data were in reasonable agreement with experimental free energies of binding obtained by titration microcalorimetry. The established predictive protocol is proposed to rationalize large sets of data such as glycan arrays and to help in lead discovery projects based on such high throughput technology.

**Citation:** Topin J, Arnaud J, Sarkar A, Audfray A, Gillon E, et al. (2013) Deciphering the Glycan Preference of Bacterial Lectins by Glycan Array and Molecular Docking with Validation by Microcalorimetry and Crystallography. PLoS ONE 8(8): e71149. doi:10.1371/journal.pone.0071149

**Editor:** Jérôme Nigou, French National Centre for Scientific Research - Université de Toulouse, France

**Received:** May 6, 2013; **Accepted:** June 26, 2013; **Published:** August 19, 2013

**Copyright:** © 2013 Topin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The research was funded by Labex Arcane (ANR-11-LABEX-003) and also supported by Centre national de la recherche scientifique, Université Joseph Fourier and ANR grant NeoLect (ANR-11-BSV5-002-01) and WallArray II (ANR-GENOM-BTV-010). The authors acknowledge the COST action BM1003 and CM1102 and the Initial Training Network "EuroGlycoArrays". Support from the Association Vaincre la Mucoviscidose and GDR Pseudomonas is appreciated. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: imberty@cermav.cnrs.fr

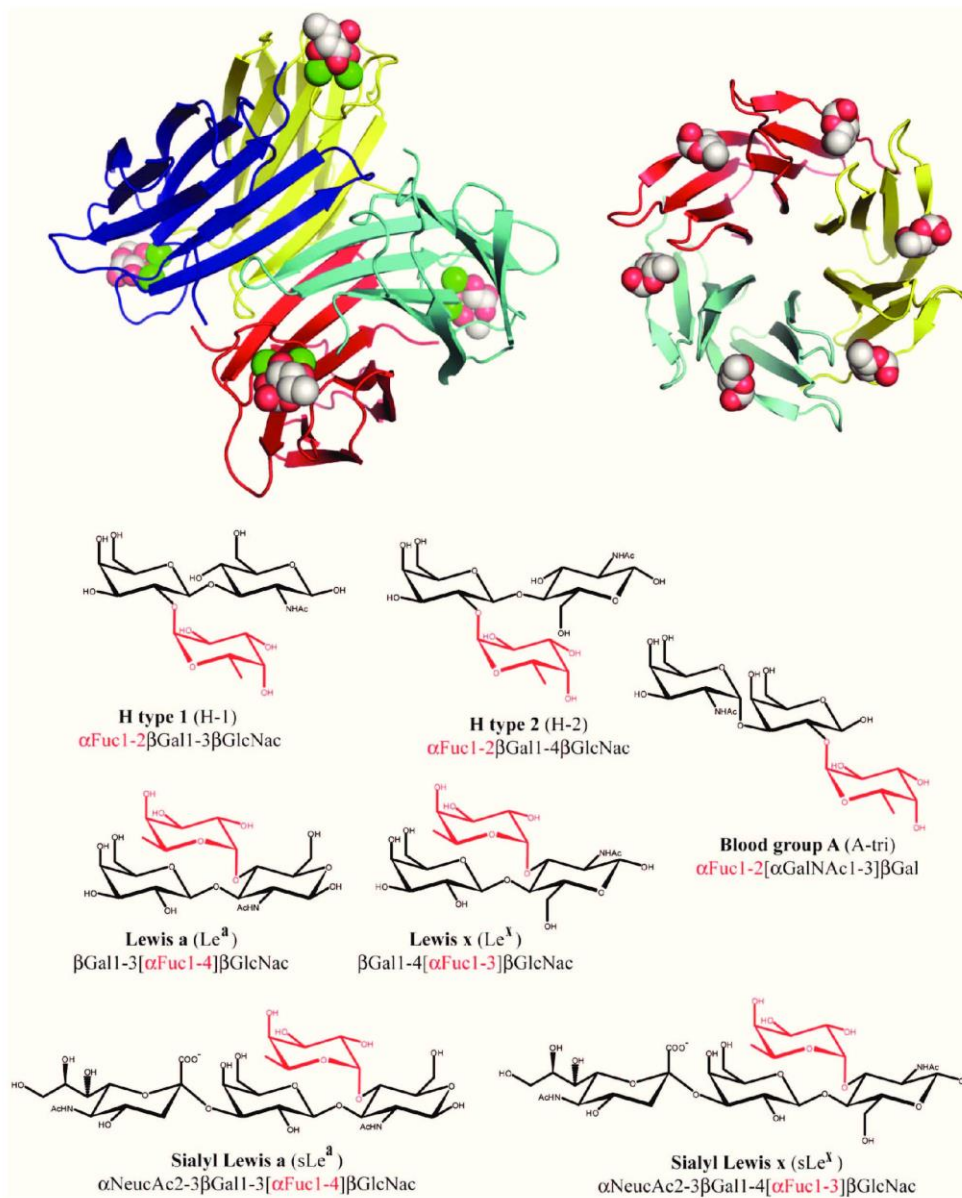
## Introduction

Interaction between lectins, or other carbohydrate binding proteins, and carbohydrates play important roles in many biological and pathological processes. In the last years, high throughput analysis of glycome or availability of glycan microarrays [1,2] revolutionized the glycobiology field. Integrated approaches are now needed for combining data arising from investigations performed at different scales. The wealth of information obtained from glycan array experiments and available in databases has to be combined with experimental data that describe the structure and affinity of single lectin binding sites. Molecular modeling can also be performed with different levels of sophistication and represents an excellent tool for integrating and rationalizing data obtained from different methods.

Carbohydrate binding proteins from pathogenic bacteria, such as lectins, adhesins or toxins are interesting models since, as a result of co-evolution, they have the capacity to specifically recognize complex oligosaccharides present on host tissues [3–5].

Among the human glycoconjugates that can be targeted by bacterial lectins, the A, B, and H antigens are complex fucosylated oligosaccharides present on endothelial cells and erythrocytes of all individuals of blood group A, B, or O, respectively [6]. In addition, Lewis epitopes, that are also fucosylated oligosaccharides, depend on the Lewis genotype of the individuals. The biological role of the ABO and Lewis histo-blood group systems remains to be elucidated, but since the 80s, several studies have pointed out correlations between the repartition of phenotype in population and the susceptibility to diseases [7,8]. The most cited examples are the individuals with O phenotype who present higher susceptibility towards severe forms of cholera [9,10] and gastroenteritis caused by Norwalk virus [11]. For individuals with secretor phenotype, blood-group related epitopes are also present in lung mucus. The nature of the oligosaccharides present in airways depends not only on ABO, Lewis and secretor genotypes but also on long-term inflammatory diseases such as chronic bronchitis and cystic fibrosis (CF). More particularly, fucosylated glycoconjugates, which are present in higher quantity in mucins



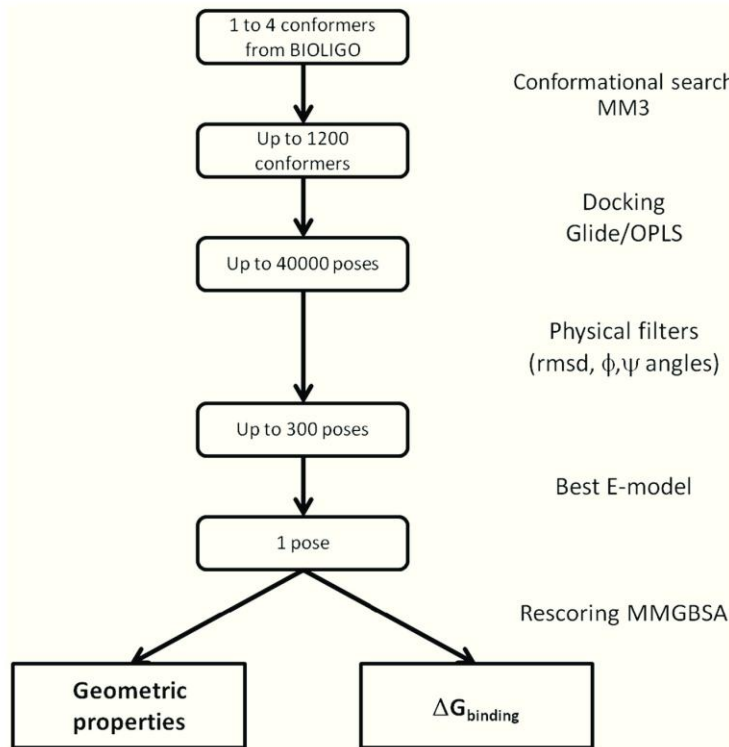


**Figure 1. Representation of the bacterial lectins and oligosaccharides used in the docking calculations.** Top: Graphical representation of the two bacterial lectins. From left to right LecB/fucose and BambL/fucose complexes have been adapted from PDB 1GZT and 3ZW2 respectively. The fucose ligand and calcium atoms are displayed as spheres. Bottom: Schematic representation and sequences of fucosylated oligosaccharides. doi:10.1371/journal.pone.0071149.g001

[12] and N-glycans [13] of lungs of CF patients, appear to be the target for lectins from pathogenic bacteria that are responsible for morbidity and mortality in CF patients.

Soluble lectins with high affinity for human fucosylated oligosaccharides have been identified in *Pseudomonas aeruginosa* and bacteria from the *Burkholderia cepacia* complex such as *B. cenocepacia* and *B. ambifaria* [14–17]. LecB (also named PA-III) from *P. aeruginosa* is a tetrameric protein that displays an unusually

strong micromolar affinity for L-fucose in a tight binding site which requires two  $\text{Ca}^{2+}$  ions [18,19]. BambL from *B. ambifaria* is a trimeric lectin arranged in a  $\beta$ -propeller fold with two similar binding sites per monomer, resulting in a hexameric arrangement of the fucose binding sites [14]. Both lectins were shown to bind to a large variety of fucosylated oligosaccharides with LecB having higher affinity for the Lewis a epitope [20] and BambL for the H-



**Figure 2. Flowchart of the docking/analysis protocol.**  
doi:10.1371/journal.pone.0071149.g002

type 2 epitope [14]. An illustration of the two fucose-binding lectins is displayed in Figure 1.

Bacteria utilize therefore a variety of multimer topologies and different fucose binding sites with various specificities towards fucosylated human histo-blood group oligosaccharides and molecular modeling is an approach of choice for rationalizing the observed specificities. Even though molecular docking calculations are routinely performed on small organic compounds, docking oligosaccharides is still a challenging task due to the constraints of ring shape (fucose preferentially adopts a <sup>1</sup>C<sub>4</sub> conformation), the occurrence of CH-π interactions between aromatic amino-acids and the C-H bonds of sugars, and the role of peculiar stereo-electronic effect such as anomeric and exo-anomeric effects which are not handled in standard molecular mechanics algorithms [21,22]. Nevertheless, conformational analysis of histo-blood group epitopes demonstrated that these oligosaccharides can adopt a limited number of well defined conformations [23,24]. These are available in the BIOLIGO database (<http://bioligo.cermav.cnrs.fr/>) that contains three-dimensional structures of bioactive oligosaccharides generated using the SWEET-II [25] and POLYS [26] builders altogether with Shape a conformational search engine based on a genetic algorithm coupled to the MM3 force-field [27]. Molecular docking studies of fucose and fucose-containing oligosaccharides have been published using a variety of computational approaches, in protein targets such as human lectins [28,29], antibodies [30,31] or virus capsid proteins [32]. Molecular docking calculations of fucose and methyl-fucoside in bacterial lectins were also reported even though

the presence of calcium ions in some of the binding sites was computationally challenging [33–37].

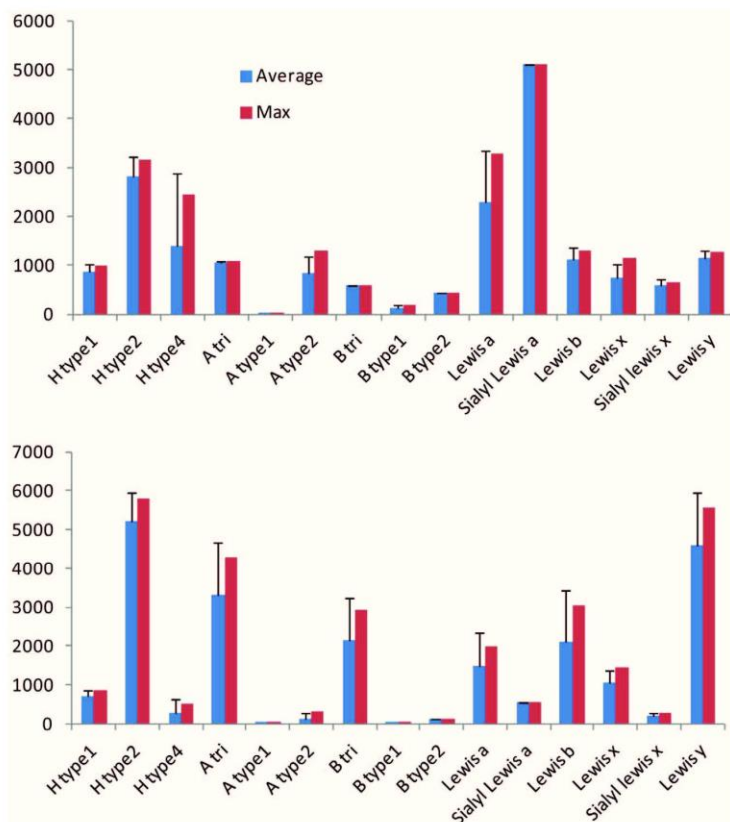
With increasing availability of glycan arrays which allow for rapid deciphering of lectin specificity [1], the need for efficient protocols for structural rationalization of data is urgent. In this paper, we propose a docking strategy based on the use of starting oligosaccharides conformations derived from the BIOLIGO database. The models are validated by comparison with available crystal structures, including the newly solved structure of the BambL/Lewis x (Le<sup>x</sup>) complex. An additional step of free energy calculations using the Molecular Mechanics Generalized Born Surface Area (MM-GBSA) provides theoretical affinities that can be compared to thermodynamic data determined from titration microcalorimetry experiments.

## Materials and Methods

### Experimental section

**Material.** Human histo-blood group oligosaccharides were obtained from Elicityl (Crolles, France). The two bacterial lectins were produced recombinantly in *Escherichia coli* as previously described [14,19].

**Glycan array.** Purified LecB sample was labeled with Alexa Fluor 488-TFP (Invitrogen, CA) according to manufacturer's instructions and re-purified on a D-Salt polyacrylamide desalting column (Pierce, Rockford IL). Alexa-labeled protein was used for glycan array screening with standard procedure of the Core H of the Consortium for Functional Glycomics (<http://www.functionalglycomics.org>, Emory University, Atlanta, GA, USA).



**Figure 3. Selected data from the glycan array v4.1 performed on two bacterial lectins.** Binding is represented as fluorescent intensity obtained with LecB (top panel) and BamBL (bottom panel) both at  $0.2 \mu\text{g}\cdot\text{mL}^{-1}$  and labeled with Alexa 488. Blue bar: average value with standard deviation over 5 experiments at same concentration, red bar: maximum response observed. doi:10.1371/journal.pone.0071149.g003

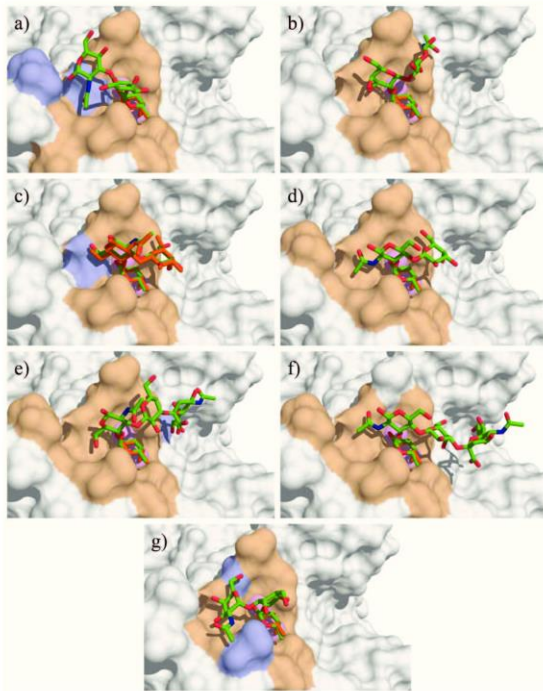
The labeled lectin was assayed on the CFG glycan array version 4.1 comprising 465 natural and synthetic glycans and the data were analyzed in a dose-dependent manner as previously described [38] at 10, 1 and  $0.1 \mu\text{g ml}^{-1}$  of LecB (PA-III) and 1, 0.2 and  $0.05 \mu\text{g ml}^{-1}$  of BamBL dissolved in 20 mM HEPES, 140 mM NaCl, 5 mM  $\text{CaCl}_2$ , pH 7.5.

**Isothermal Titration Calorimetry (ITC).** Recombinant lyophilized LecB was dissolved in buffer (20 mM Tris/HCl pH 7.5, NaCl 150 mM with 0.03 mM  $\text{CaCl}_2$ ). Protein concentration was checked by measurement of  $A_{280}$  by using a theoretical molar extinction coefficient of  $6990 \text{ M}^{-1} \text{ cm}^{-1}$ . Carbohydrate ligands (H-type 1, H-type 2 and blood group A trisaccharides, sialyl Lewis x and sialyl Lewis a tetrasaccharides) were dissolved in the same buffer, and loaded in the injection syringe. ITC was performed with an ITC200 microcalorimeter (MicroCal Inc.). Lectin solution was placed ( $150 \mu\text{M}$ ) in the  $200 \mu\text{l}$  sample cell, at  $25^\circ\text{C}$ . Titration was performed with 20 of  $2 \mu\text{l}$  injections of carbohydrate ligands (1.3 to 2.0 mM) every 300 s. Data were fitted with MicroCal Origin 7 software, according to standard procedures. Fitted data yielded the stoichiometry ( $n$ ), the association constant ( $K_a$ ) and the enthalpy of binding ( $\Delta H$ ). Other thermodynamic parameters (i.e. changes in free energy,  $\Delta G$ , and entropy,  $\Delta S$ ) were calculated from the equation  $\Delta G = \Delta H - T\Delta S = -RT \ln K_a$  in which  $T$  is the absolute temperature and

$R = 8.314 \text{ J mol}^{-1} \text{ K}^{-1}$ . Two independent titrations were performed for each ligand tested.

**Crystallogenesi, data collection and structure determination.** Crystals of BamBL complexed with  $\text{Le}^x$  tetrasaccharide were obtained by the hanging-drop vapor diffusion method using  $2 \mu\text{l}$  drops containing a 50:50 (v/v) mix of protein and reservoir solution at  $19^\circ\text{C}$ . Lyophilized protein was dissolved in Tris/HCl buffer pH 7.5, NaCl 150 mM at 1 mM and incubated for 1 h with 5 mM of sugar at room temperature prior to co-crystallization. Reservoir solution consisted of 200 mM Trisodium citrate, 100 mM sodium acetate pH 5.0 and 28% Peg 6K. 10% ethylene glycol was added for cryoprotection prior mounting in a cryoloop and flash-freezing in liquid nitrogen. Diffraction data were collected at 100K, at the European Synchrotron Radiation Facility (Grenoble, France) on ID14-EH4 beamline using an ADSC Q4 CCD detector. The data were processed using MOSFLM [39] and scaled and all further computing was performed using the CCP4 suite unless otherwise stated [40]. Data quality statistics are summarized in Table S1. Molecular replacement was used to obtain the initial phases with the program PHASER [41] and one monomer of the native structure of BamBL (PDB ID: 3ZW0) as searching model. The structure was refined by restrained maximum-likelihood refinement using REFMAC 5.5 [42] with 5% of the data retained for





**Figure 4. Docking of the oligosaccharides in LecB binding site.** a) H type 1, b) H type 2, c) Le<sup>a</sup>, d) Le<sup>b</sup>, e) sLe<sup>a</sup> f) sLe<sup>x</sup> and g) A-tri. The docked oligosaccharides are represented as sticks (carbon, oxygen and nitrogen atoms are colored green, red and blue respectively) and the ones from crystal structures are colored orange. Calcium ions are represented as pink spheres. The protein accessible surface is colored in beige for residues comprised within a sphere of 4 Å around the ligand and in blue for residues involved in hydrogen bond with ligand residues (except fucose).

doi:10.1371/journal.pone.0071149.g004

cross-validation interspersed with model building and electron density inspection in Coot [43]. The incorporation of the sugar moieties was performed after inspection of the the 2mFo-DFc and mFo-DFc weighted maps. Water molecules were introduced automatically using Coot and inspected manually. The coordinates have been deposited in the Protein Data Bank under code 3ZW1. All figures were drawn with The PyMOL Molecular Graphics System (Schrödinger, LLC).

**Modeling section**

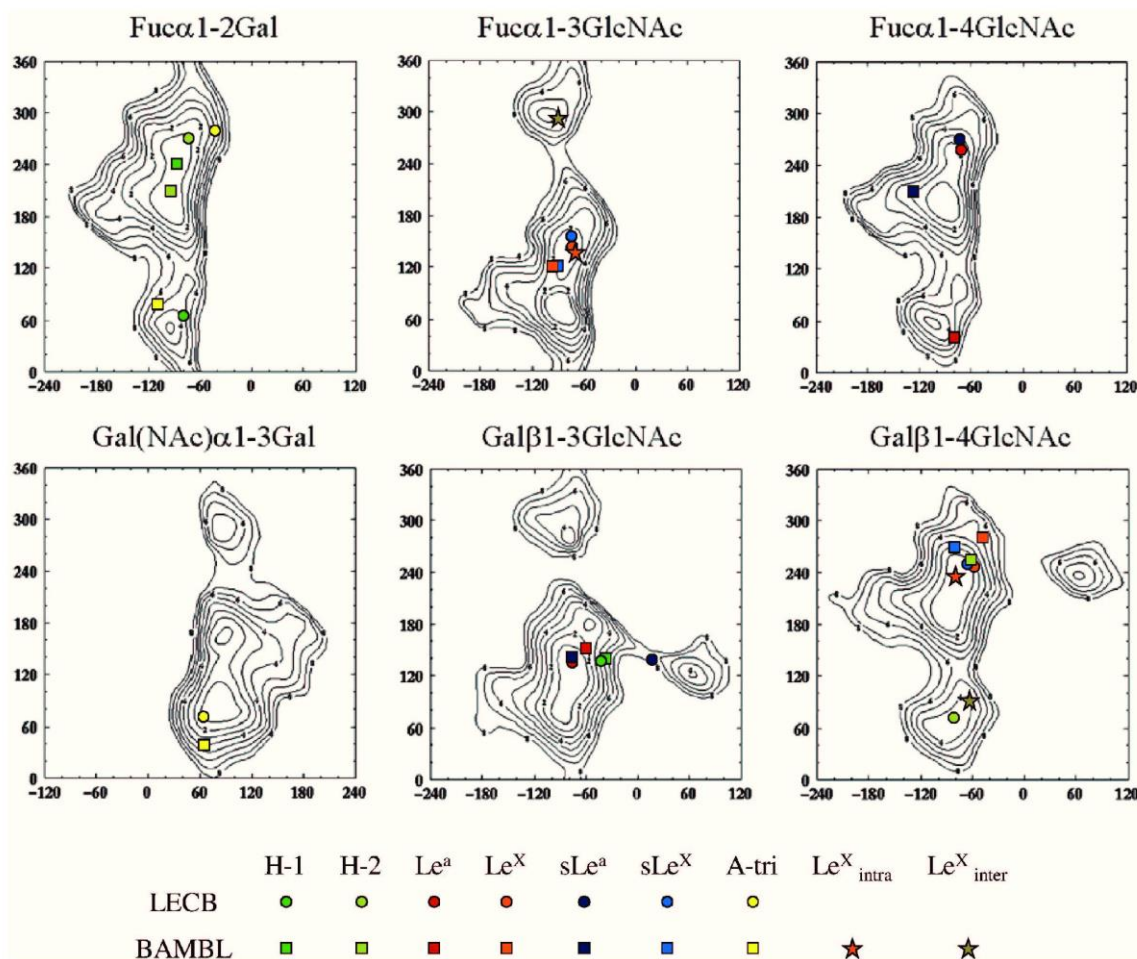
All calculations were performed using the Schrödinger Suite of programs through the Maestro graphical interface (<http://www.schrodinger.com>, Schrödinger LLC, New York). The overall docking protocol and filtering procedure are schematized in Figure 2.

**Preparation of the ligands.** Models of the seven fucose-containing oligosaccharides displayed in Figure 1 have been built. Between two and four different conformations were considered for each structure, first one by extracting coordinates from the crystal structures of the complex when available and others one by selecting low energy conformations from the BIOLIGO database (<http://glyco3d.cermav.cnrs.fr/bioligo>). The dihedral angle values of the glycosidic linkages of the starting conformers are provided in Table S2.

**Preparation of the proteins.** The crystal structures of two lectin/carbohydrate complexes, LecB/Lewis a (PDB ID: 1W8H) [20] and BambL/H type 2 (PDB ID: 3ZZV) [14] were taken from the Protein Data Bank [44]. The four binding sites of LecB contain a conserved water molecule in close interaction with the protein and the fucose moiety. This structural water molecule is also found in other LecB crystal structures, and was thus explicitly included in all calculations. Since the lectin BambL possesses two similar binding sites per monomer, only the intra-monomeric site was selected for the calculations. The hydrogen atoms were added assessing a pH of 7. The histidine residues were treated as neutral. The selection of the histidine enantiomer and the orientation of asparagine and glutamine side-chains were performed so as to maximize hydrogen bond patterns. The partial atomic charges derived from the OPLS-2005 [45] were assigned to the ligands, calcium ions and protein atoms. Finally, energy minimizations were performed using the Impref utility of the Schrödinger package and the OPLS-2005 force-field.

**Docking Calculations.** The docking program Glide [46,47] has been shown to give good performances for docking carbohydrates [37]. Rigid-receptor docking calculations were run in standard precision mode, with the OPLS-2005 force field, while the ligand was considered as flexible. The shape and physico-chemical properties of the binding sites were mapped onto cubic grid of 24 Å side centered on the ligand. During the grid generation, a scaling of van der Waals radii of 1 for atoms with partial charges less than 0.25 e was used. Prior to docking, each conformer was submitted to an extensive conformational search protocol using the MM3 force field to generate several hundreds of conformers. Parameters for van der Waals radii were scaled by 0.80 for atoms with partial charges less than 0.15. Ring conformational sampling was not allowed in order to maintain the starting conformations, particularly the <sup>1</sup>C<sub>4</sub> conformation of the L-fucose ring; no other constraints were introduced. Ligand poses were clustered with a RMS tolerance of 1.3 Å. A maximum of 10,000 poses could be retained for the initial docking phase (i.e. up to 40,000 with four starting oligosaccharide conformations), and were energy minimized.

A hierarchical post-docking filtering procedure was developed to take into account the specific nature and requirements of the studied lectins and of the low energy oligosaccharide structures. Otherwise, the combination of the OPLS-2005 force-field coupled to Glide could generate oligosaccharide structures corresponding to high energy conformers. Poses were first filtered according to the RMSD between heavy atoms of the fucose residue of the poses and of the fucose present in the crystal starting structure (threshold of 1 Å). Additional geometrical filters were also developed to reject distorted geometries. Poses having Φ and Ψ glycosidic dihedral angles outside the known low energy regions (±10°) were rejected. These low-energy regions were based on the previously calculated adiabatic maps of disaccharides [23] available from the Glyco3D site (<http://www.cermav.cnrs.fr/Glyco3D>). Finally, the poses having an energetically unfavorable cis conformation of the amide bond in N-acetyl groups [48] were discarded. Between 2 to 300 binding poses for each starting conformer were finally retained from this procedure, and were all visually inspected. The docked solutions were ranked using Emodel, a composite scoring function, available in the Schrödinger Suite and that is a combination of GlideScore, of the ligand receptor molecular mechanics interaction energy, and of the ligand strain energy (Friesner et al. 2004). This score, which better takes into account electrostatic and van der Waals energy, is particularly well suited for comparing conformers but it should be used together with another scoring function while comparing different species.



**Figure 5. Energy maps and docked conformations.** Isoenergy contours (1 kcal/mole) are represented as a function of  $\Phi$  (x-axis) and  $\Psi$  (y-axis) torsion angles for all glycosidic linkages of interest taken from Glyco3D (<http://glyco3d.cermav.cnrs.fr/>) with superimposition of oligosaccharide conformations derived from docking (circles and squares) or protein crystallography (stars). Torsion angles are defined as  $\Phi = \theta(O5-C1-O1-Cx)$  and  $\Psi = \theta(C1-O1-C'x-C'x+1)$ .  
doi:10.1371/journal.pone.0071149.g005

**MMGBSA rescoring.** The binding free energy of each best docked poses was estimated using the local optimization feature in Prime software (Schrödinger) based on the molecular mechanics (MM) OPLS-2005 force field and the Generalized Born Surface Accessible (GBSA) continuum model. Such MM-GBSA rescoring protocol has proved to be successful [49]. The binding free energy of each ligand  $\Delta G_{bind}$  is described as the difference in the sums of MM energy, GBSA solvation energy (including surface area energy SA) of the complex on one hand and of the free ligand and free protein on the other hand. Strain energies were calculated for both the protein and the ligand. This feature is directly implemented in the Schrödinger suite and consists in quantifying the difference in energy between the free and the complex state of both the ligand and the protein after energy minimization. The energy minimization of the protein, was performed using Prime by allowing the side-chains within 10 Å of the ligand to move, in order to avoid large structure deviation.

**Results**

**Specificity analysis by glycan array**

The fine specificity of the two bacterial lectins was analyzed using glycan array v4.1 from the Consortium for Functional Glycomics that contains 465 glycoconjugates (<http://www.functionalglycomics.org>). The overall glycan array data for BambL was described previously [14] and those for LecB are available from the CFG website. Briefly, BambL only attaches to the oligosaccharides presenting at least one fucose residue, whereas LecB displays also some weak affinity towards mannose-containing glycans. In order to set up a comparison with the docking results, binding data were analyzed by selecting only glycans having one fucosylated epitope at the non-reducing position. This resulted in data for 53 glycans with two lectins at three concentrations (see Tables S3 and S4). Each of these 53 glycans could be assigned to one of the 15 human epitopes. The different concentrations gave



**Table 1.** Characterization of the best poses of the fucosylated oligosaccharides in the binding site of LecB.

	$\Phi$	$\Psi$	$\Phi$	$\Psi$	Hydrogen bonds	Hydrophobic contacts	Glide Emodel	$\Delta G_{MMGBSA}$
<b>H-1</b>	$\alpha$ Fuc1-2Gal		$\beta$ Gal1-3GlcNAc		GlcNAc.O1...R72.NH2	GlcNAc.CH3...G24	-97.4	-125.9
	-74	65	-47	160	GlcNAc.O1...D96.O	GlcNAc.CH3...T98		
<b>H-2</b>	$\alpha$ Fuc1-2Gal		$\beta$ Gal1-4GlcNAc				-92.4	-108.8
	-80	272	-81	61		GlcNAc.CH3...D99		
<b>Le<sup>a</sup></b>	$\alpha$ Fuc1-4GlcNAc		$\beta$ Gal1-3GlcNAc				-85.8	-127.0
	-77	260	-68	131	GlcNAc.O6...D96.O			
<b>Le<sup>x</sup></b>	$\alpha$ Fuc1-3GlcNAc		$\beta$ Gal1-4GlcNAc			GlcNAc.CH3...V69	-88.6	-120.1
	-68	147	-66	254				
<b>sLe<sup>a</sup></b>	$\alpha$ Fuc1-4GlcNAc		$\beta$ Gal1-3GlcNAc		Sia.O8...E86.OE		-91.3	-129.7
	-76	272	11	147				
<b>sLe<sup>x</sup></b>	$\alpha$ Fuc1-3GlcNAc		$\beta$ Gal1-4GlcNAc			GlcNAc.CH3...V69	-87.9	-118.9
	-72	148	-72	256				
<b>A-tri</b>	$\alpha$ Fuc1-2Gal		$\alpha$ GalNAc1-3Gal		GalNAc.O4...R72.NH2		-90.8	-111.9
	-48	272	63	71	GalNAc.O4...D96.O			
					Gal.O4...S23.OG			

The glycosidic dihedrals  $\Phi$  and  $\Psi$  are indicated in degrees. The hydrogen bonds and hydrophobic contacts between the ligands and residues located less than 4 Å of any ligand atom (except fucose) are listed. The associated Glide Emodel score and binding free energy are given in kcal/mol.  
doi:10.1371/journal.pone.0071149.t001

very consistent data, and only the results corresponding to one lectin concentration are given in Figure 3.

Very different binding patterns were observed for the two bacterial lectins. The glycan array of LecB confirmed its preference for Lewis a (Le<sup>a</sup>) epitopes with highest apparent affinity for sialyl Lewis a (sLe<sup>a</sup>). H type 2 is also a strong binder, whereas other fucosylated glycans resulted in moderate binding, except A type 1 and B type 1 that were not recognized. Despite the fact that BamBL was described as a fucose-binding lectin with broad specificity [14], the present analysis indicated its preference for H type 2 and Lewis y (Le<sup>y</sup>) oligosaccharides.

#### Docking of fucosylated oligosaccharides in bacterial lectins

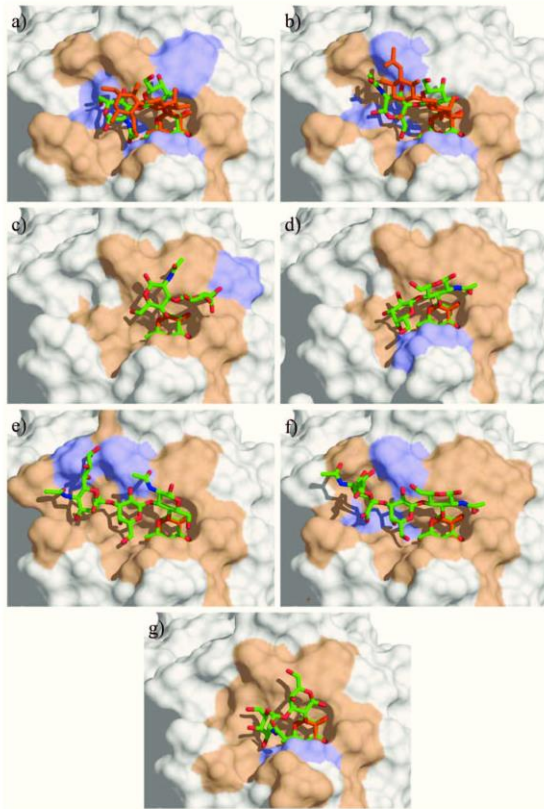
Among the 14 oligosaccharides analyzed by glycan array, seven were selected for the docking approaches based on their biological relevance and diversity (Figure 1). Lewis b and Lewis y were not included since the occurrence of two fucose residues per oligosaccharides would complicate the interpretation of docking results.

**Validation of the molecular docking protocol on the LecB/Le<sup>a</sup> complex.** The docking protocol with selection of adequate scoring function was setup so as to reproduce the binding mode of Le<sup>a</sup> in LecB (rmsd ligand of 0.5 Å), which is the most difficult to handle due to the presence of the two calcium ions, treated with a net +2 charge in a non-polarizable force field. Solution NMR experiments [24] and theoretical studies [23] have proposed an almost rigid conformation for this branched trisaccharide with parallel piling (pseudo-stacking) of the galactose and fucose rings (conf 1 in Table S2). This low energy conformation is also present in the crystal structure of LecB/Le<sup>a</sup> complex [20]. Nevertheless, recent NMR studies concluded for some conformation flexibility [50]. 130 binding poses were obtained for Le<sup>a</sup> in the active site of LecB. The scoring function GlideScore did not rank the crystal structure conformation as the best pose. This could be corrected by the use of the composite scoring function called Emodel [46], which combines the binding

affinity predicted by GlideScore and the internal strain energy of the ligand (see Figure S1). The Emodel score was thus used to rank all the oligosaccharide binding poses in both lectins.

**Docking of fucosylated oligosaccharides in LecB from *P. aeruginosa*.** The seven oligosaccharides of interest were docked within the LecB binding site. As imposed with the filtering protocol, the fucose moiety always coordinated the two calcium ions through oxygen atoms O-2, O-3 and O-4. The shallow shape of the binding site allowed the oligosaccharides to adopt a variety of conformations. Since the present study aims at a medium throughput rationalization of glycan array data, only the pose with the best Emodel score is further discussed. It should be noted that this pose did not always originate from calculations starting with the lowest energy model from BIOLIGO database, i.e. conformer 1 in Table S2. In 75% of the complexes, the best pose was obtained starting from other conformations taken from BIOLIGO database.

The best poses are displayed in Figure 4 and their structural analysis is given in Table 1. Despite their structural similarities, H type 1 and H type 2 trisaccharides presented very different conformations (Figure 4a and 4b). The difference in  $\alpha$ Fuc1-2Gal linkage conformation is illustrated in Figure 5 whereas the values of  $\Phi$  and  $\Psi$  torsion angles are reported on the previously calculated energy maps [23] that give an estimation of the available conformational space for each glycosidic linkage. For both trisaccharides, one of the two glycosidic linkages adopted the lowest energy conformation, while the other one lied in a secondary minimum corresponding to a 120° difference in  $\Psi$  angle compared to low energy region. H-type 1 established two hydrogen bonds but since these involved reducing O1 of GlcNAc, they will not occur when GlcNAc is conjugated such as on glycan arrays. The only predicted hydrogen bonds involved the anomeric oxygen to GlcNAc. The slightly higher affinity of LecB for H type 2 trisaccharide is therefore proposed to originate from hydrophobic contact involving N-acetyl group of sugar and CH<sub>2</sub> group of Asp99 side chain.



**Figure 6. Docking of the oligosaccharides in BamBL binding site.** a) H type 1, b) H type 2, c)  $Le^a$ , d)  $Le^x$ , e)  $sLe^a$  f)  $sLe^x$  and g) A-tri. The color codes are the same as in Figure 4. doi:10.1371/journal.pone.0071149.g006

Both  $Le^a$  and  $Le^x$  trisaccharides bound in their lowest energy conformation (Figure 4c and 4d), with stacking of the galactose ring above the fucose one. The docking confirmed that, as proposed from the crystal structure [20], the preference of LecB for binding to  $Le^a$  trisaccharide originated from an additional hydrogen bond between O6 of GlcNAc and Asp96. The additional neuraminic acid residue present in  $sLe^a$  tetrasaccharide established one additional hydrogen bond between glycerol side chain and Glu86 acidic group, which correlates well with the high affinity observed on the glycan array. The blood group A trisaccharide (A-tri) also established several hydrogen bonds with the protein residues, but the presence of  $\alpha$ GlcNAc forces the  $\alpha$ Fuc1-2Gal linkage in a strained conformation, that may explain the observed lowest affinity.

**Docking of fucosylated oligosaccharides in BamBL from *B. ambifaria*.** The resulting best docked solutions obtained with BamBL according to the Emodel score are represented in Figure 6 and their structures characterized in Table 2. The active-site in BamBL is deeper than in LecB, constraining the oligosaccharide conformation and allowing more contacts between the ligands and the surrounding residues. The  $\alpha$ Fuc1-2Gal linkage of H type 1 and H type 2 trisaccharides maintained its lowest energy conformation upon binding, resulting in same localization of galactose residues for both trisaccharides, and several hydrogen bonds are estab-

lished in addition to those involving fucose. However, the  $\beta$ Gal1-4GlcNAc linkage of H type 2 is bound in the lowest energy conformation, whereas the  $\beta$ Gal1-4GlcNAc linkage of H type 1 is forced in a higher energy one. This may be correlated to the preference of the lectin for H-type 2. It should be noted that both H-type 1 and H-type 2 poses are in good agreement with the crystal structures of their complexes with BamBL [14], with rmsd for ligands of 1.6 and 2.6 Å, respectively. The  $\alpha$ Fuc1-2Gal linkage in blood group A adopted the secondary minimum conformation with large shift in  $\Psi$  value, and the  $\alpha$ GlcNAc establishes several contacts with the protein surface.

The docking of Lewis oligosaccharides is of special interest. Indeed, since Trp74 is stacked against the fucose hydrophobic face in the binding site, the solution conformation of  $Le^a$  and  $Le^x$ , with galactose piling up on fucose, cannot be accommodated. Nevertheless, BamBL binds significantly to Lewis oligosaccharides on the glycan arrays (Figure 3). Our docking calculations predicted that this “closed” conformation has to open, for pushing the galactose on the surface of the binding site. For  $Le^a$  trisaccharide this is performed by a conformational change of the  $\alpha$ Fuc1-4GlcNAc linkage that occupies the secondary minimum with  $\Psi$  conformation close to  $60^\circ$  (Figure 5) resulting in a “twist shape”. For  $Le^x$ , each glycosidic linkage remained in the main energy well, but with significant deviation (up to  $30^\circ$ ) for the global minimum resulting in a “semi-open” shape. This conformation occurred for both  $sLe^a$  and  $sLe^x$  but the distortion of linkages seems to be unfavorable since both tetrasaccharides are not efficiently recognized by BamBL on glycan array.

#### Crystal structure of BamBL/Lewis x complex and comparison with docking

**Overall structure.** BamBL was cocrystallised with the  $Le^x$  tetrasaccharide and crystals belonging to the  $P2_13$  spacegroup with cell parameters  $a = b = c = 81.4 \text{ \AA}$  and  $\alpha = \beta = \gamma = 90^\circ$  appeared after 5 days. The complex structure was solved by molecular replacement and two molecules were found in the asymmetric unit which led to a solvent content of 46.5%. Each belongs to a trimer generated by applying crystallographic three-fold symmetry (Figure S2). The final model was refined in the resolution range between 40.7 and 1.6 Å to R and  $R_{free}$  values of 15.3 and 18.9 respectively. It is composed of all 87 amino acids in each monomer and the N-terminal methionine of chain A was found oxidized. No conformational changes are observed between the two molecules or when compared with the native structure. Only slight movements are present at the level of some surface loops.

**Oligosaccharide binding site.** The initial electronic density maps revealed clearly the presence of the fucose moiety in the four available binding sites (two per monomers) and of a variable number of additional sugars depending on the binding site. In the intermolecular site of A molecule and in the intramolecular site of B molecule only the  $\alpha$ Fuc1-3GlcNAc moiety could be fitted in the density whereas the other residues are too disordered to be located. In both cases, the disaccharide adopted an unusual conformation with of  $\Psi$  torsion angle (approx  $-60^\circ$ ) corresponding to a secondary minimum in the low energy map (Figure 5). This particular conformation brings the N-acetyl group of GlcNAc in a groove on the protein surface.

In the two other sites, clear electron density was observed for the whole  $Le^x$  tetrasaccharide (Figure 7). In the intermolecular site of B, the reducing galactose displayed some mobility according to B-factor and density. It only makes interactions mediated by water molecules with the symmetric molecule through its O6 and O4 hydroxyl groups. All hydroxyls of the non-reducing galactose are involved in hydrogen bonds either direct or mediated by solvent.



**Table 2.** Characterization of the best poses of the fucosylated oligosaccharides in the binding site of BamBL.

	$\Phi$	$\Psi$	$\Phi$	$\Psi$	H bond	Hydroph	Glide Emodel	$\Delta G_{MMGBSA}$
<b>H-1</b>	$\alpha$ Fuc1-2Gal		$\beta$ Gal1-3GlcNAc		Gal.O3...Y35.OH	GlcNAc.CH3...G74	-82.0	-72.4
	-93	238.	-41	134	GlcNAc.O...W74.NE1	GlcNAc.CH3...W76		
					GlcNAc.O6...S13.OG			
					GlcNAc.O6...W8.NE1			
<b>H-2</b>	$\alpha$ Fuc1-2Gal		$\beta$ Gal1-4GlcNAc		GlcNAc.O1...W8.NE1	GlcNAc.C6...W74	-80.8	-83.1
	-106	219	-59	249	GlcNAc.N...S13.OH			
					GlcNAc.O3...D30.O			
					GlcNAc.O6...W74.NE1			
<b>Le<sup>a</sup></b>	$\alpha$ Fuc1-4GlcNAc		$\beta$ Gal1-3GlcNAc		Gal.O6...T36.O	GlcNAc.CH3...Y35	-70.4	-81.8
	-917	45	-59	153				
<b>Le<sup>x</sup></b>	$\alpha$ Fuc1-3GlcNAc		$\beta$ Gal1-4GlcNAc		Gal.O4...W74.NE1	GlcNAc.CH3...A38	-66.7	-71.6
	-104	120	-48	285	Gal.O6...G76.O			
<b>sLe<sup>a</sup></b>	$\alpha$ Fuc1-4GlcNAc		$\beta$ Gal1-3GlcNAc		Gal.O2...D30.OD	Gal.C6...W74	-93.8	-96.8
	-127	214	-74	136	Sia.O8...D30.O	Gal.C6...V57		
					Sia.CO...T11.N	Sia.C2...W8		
<b>sLe<sup>x</sup></b>	$\alpha$ Fuc1-3GlcNAc		$\beta$ Gal1-4GlcNAc		Gal.O2...TD30.OD	GlcNAc.CH3...Y35	-90.9	-65.7
	-113	105	-73	271	Sia.O1...W8.NE1	Gal.C6...V57		
					Sia.CO...T11.OG			
<b>A-tri</b>	$\alpha$ Fuc1-2Gal		$\alpha$ GalNAc1-3Gal		GalNAc.O4...W74.NE1	GalNAc.C4...V57	-70.7	-83.9
	-111	69	57	42		GalNAc.C6...W8		

The glycosidic dihedrals  $\Phi$  and  $\Psi$  are in degrees. The hydrogen bonds and hydrophobic contacts between the ligands and residues located less than 4 Å of any ligand atom (excepted the fucose) are listed. The associated Glide Emodel and binding free energy are given in kcal/mol.  
doi:10.1371/journal.pone.0071149.t002

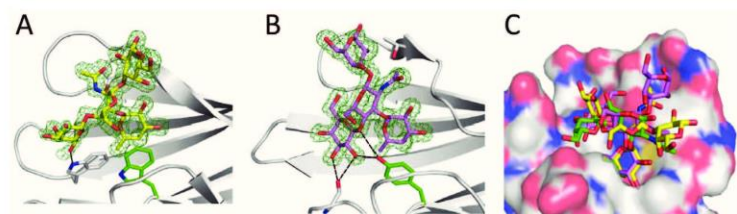
The O4 oxygen interacts with Thr36 and the O6 with Asp75 via a water molecule. All other contacts are established with a neighboring protein unit generated by packing symmetry. The O6 atom of the GlcNAc moiety is hydrogen bonded to Arg27 and the O7 and N1 are water bridged to Arg81 and Asp75, respectively. The  $\alpha$ Fuc1-3GlcNAc conformation corresponds to the low energy region of the map ( $\Phi = -91^\circ$ ,  $\Psi = 149^\circ$ ), but the  $\beta$ Gal1-4GlcNAc adopts a secondary minimum ( $\Phi = -51^\circ$ ,  $\Psi = 81^\circ$ ) as displayed in Figure 5. In the intramolecular site of A molecule, the reducing galactose only interacts with the symmetry-generated protein mainly via direct contacts with O1, O2, O4 and O5 atoms. For the non-reducing galactose, only the O4 hydroxyl makes a weak contact with Asp30 through solvent. At the level of the GlcNAc moiety, the pyranose ring is distorted to a <sup>3,0</sup>B boat conformation so that the N acetyl group stacks onto Tyr35 ring. This could not happen in the intermolecular site since there is a

threonine at this position and therefore no possibility of steric clashes. Both  $\alpha$ Fuc1-3GlcNAc and  $\beta$ Gal1-4GlcNAc linkages belong to the main energy minimum, but the overall conformation is very unusual and does not present the classical stacking between fucose and galactose ring due to the distortion of the GlcNAc ring.

Comparison between the two conformations observed in the crystal structure and the best pose obtained by docking is represented in Figure 7c. Although the predicted pose is not identical to any of the ones in the crystal structure (rmsd values of 3.15 and 2.17 Å), the docking correctly predicted the general distorted shape of Lewis x in the binding site of BamBL.

**Thermodynamic of binding by titration microcalorimetry**

Glycan array data provide evaluation of the specificity of the lectins, but for precise measurement of affinity, and comparison



**Figure 7. Crystallographic structure of BamBL in complex with Le<sup>x</sup> tetrasaccharide.** (A) Intramonomeric and (B) intermonomeric sites. Tetrasaccharide and key amino acids are represented as sticks, hydrogen bond to non-fucose carbohydrate residues as dotted lines; 2mF<sub>o</sub>-Df<sub>c</sub> electron density map contoured at 1σ is shown as green mesh. (C) Superimposition of Le<sup>x</sup> tetrasaccharide from crystal intramonomeric (yellow) and intermonomeric (pink) sites and Le<sup>x</sup> trisaccharide from docking (green).  
doi:10.1371/journal.pone.0071149.g007

**Table 3.** Titration microcalorimetry data for the interaction between bacterial lectins and fucosylated ligands.

Ligand	LecB			BambL		
	$K_D$ ( $\mu$ M)	$\Delta G$ (kcal/mol)	$\Delta H$ (kcal/mol)	$K_D$ ( $\mu$ M)	$\Delta G$ (kcal/mol)	$\Delta H$ (kcal/mol)
H type 1	1.39	-8.0	-7.7	26.1 <sup>b</sup>	-6.3	-4.2
H type 2	0.48	-8.6	-9.4	7.5 <sup>b</sup>	-7.0	-10.6
Le <sup>a</sup>	0.21 <sup>a</sup>	-9.1	-8.4	18.2 <sup>b</sup>	-6.5	-6.9
Le <sup>x</sup>	3.44 <sup>a</sup>	-7.4	-5.3	34.8 <sup>b</sup>	-6.1	-9.3
sLe <sup>x</sup>	5.7	-7.2	-5.6	39.5	-6.0	-16.4
sLe <sup>a</sup>	0.29	-8.9	-9.3	61.3	-5.7	-15.4
A tri	10.3	-6.8	-6.0	0.46 <sup>b</sup>	-8.6	-12.7

All data have been measured at least twice and standard deviations are below 15%. The numbers in italics correspond to experimental data obtained with longer oligosaccharides (i.e. one additional glucose on reducing end).

<sup>a</sup>from [20],

<sup>b</sup>from [14].

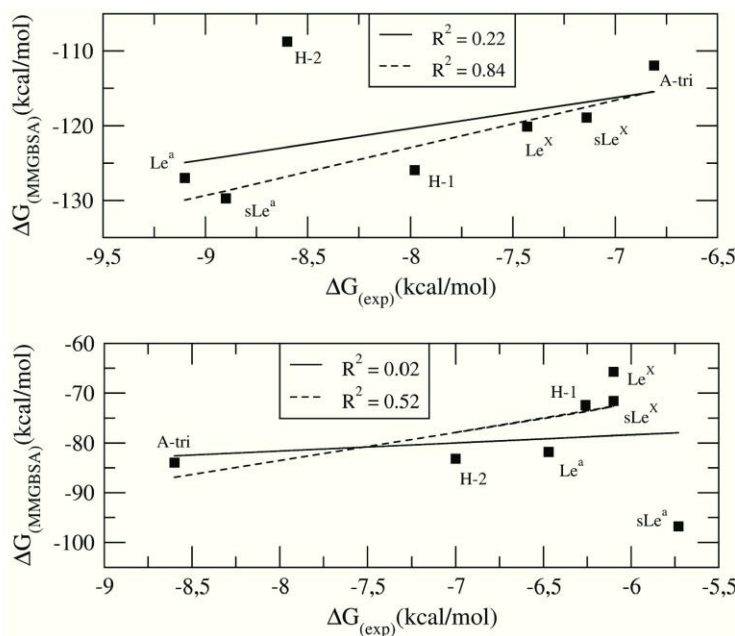
doi:10.1371/journal.pone.0071149.t003

with calculated binding energies, titration calorimetry is the most relevant approach. Thermodynamic data for BambL interacting with oligosaccharides were determined previously [14] therefore only LecB was assayed by ITC (Table 3). The seven fucosylated oligosaccharides present affinities for LecB in the micromolar range and a typical thermogram is presented in Figure S3. The solution data are in good agreement with glycan array data (Figure 3) with highest affinity for Le<sup>a</sup> and sLe<sup>a</sup> (although with no preference for this latest one). H type 2 is slightly better ligand than H type 1, and A trisaccharide is the weaker binder among the tested oligosaccharides. BambL ITC data presented more

differences with glycan array maybe due to stronger surface effects due to the topology of the lectin. In solution, blood group A and H type 2 oligosaccharides are the highest affinity ligands, and the sialylated oligosaccharides the weakest ones.

**Energy calculation and comparison with experimental data**

Attempts to correlate the different scoring functions available within Glide with the experimental free energy of binding did not yield any satisfactory results. The binding free energy of each best docked poses was therefore rescored using MM-GBSA approach.



**Figure 8.** Correlations between experimental binding free energies of binding measured from titration microcalorimetry and theoretical ones calculated using MM-GBSA. Top: LecB, bottom: BambL (bottom). Calculations have been performed with all data (black lines) or with omission of H-2 for LecB or sLea for BambL, respectively.  
doi:10.1371/journal.pone.0071149.g008



The obtained values are listed in the last column of Table 1 and Table 2 for LecB and Bambl, respectively.

**Energy calculations for LecB/oligosaccharide complexes.** Le<sup>a</sup> and sLe<sup>a</sup> oligosaccharides gave the lowest MM-GBSA energies for binding to LecB in agreement with both glycan array and ITC data. The other predicted affinities were ranked correctly, with the exception of H-type 2 that is a clear false negative. Correlation has been attempted by plotting the theoretical binding energies versus the experimental free energy of binding obtained from ITC (Figure 8). The range for the  $\Delta G_{(MMGBSA)}$  energies is larger than the one for ITC energies. This difference could arise from the lack of terms describing enthalpy/entropy compensation in the MM-GBSA calculation. The overall agreement was not very good ( $R^2 = 0.22$ ) but improved significantly ( $R^2 = 0.84$ ) when omitting the H type 2 trisaccharide that has wrongly been predicted to be a poor ligand.

**Energy calculations for Bambl/oligosaccharide complexes.** While sLe<sup>x</sup> was correctly predicted to be a poor ligand for Bambl, the isomer sLe<sup>a</sup> displayed the best calculated affinity with MM-GBSA, which classified it as false positive since the experimental affinity is weak. Apart from these two charged tetrasaccharides, the binding energies of all oligosaccharides were in a limited range, varying from  $-71.6$  kcal/mol for Le<sup>x</sup> to  $-83.1$  kcal/mol for blood group A trisaccharide. This is in rather good agreement with ITC data that did not show a large range of affinities for Bambl binding to non-charged fucosylated oligosaccharides, with preference for A blood group trisaccharide (Figure 8). When excluding sLe<sup>a</sup> as a false positive, a correlation factor of 0.52 is obtained between theoretical and experimental free energies of binding.

## Discussion and Conclusion

The development of glycan arrays allowed studying the fine specificities of glycan-binding proteins. Even though some authors have developed systematic analyzes of glycan array data to identify the lectin specificities [51,52], there is a lack of automatic methods to extract structural information from glycan arrays. A docking protocol of bioactive oligosaccharides that would be common to fucose-binding lectins was investigated in this study, so as to structurally rationalize large amounts of experimental affinities, such as those arising from glycan array experiments. Among the different protocols evaluated, the refined molecular docking procedure combines Glide in standard precision for generating the poses and the Emodel scoring function for ranking them. Starting from several conformers derived from the BIOLIGO database greatly enhanced the conformational sampling, as up to four conformers for each molecule were introduced in the docking procedure. Using this approach, we could predict the conformation of oligosaccharides upon lectin binding and analyze the network of stabilizing contacts established. This helped in rationalizing the preference of LecB for Lewis a oligosaccharides.

Interestingly, the approach was also very efficient in predicting twisted or strained conformations of Lewis oligosaccharides in interaction with Bambl. Because of their branched structures on position 3 and 4 of GlcNAc (Figure 1), Le<sup>a</sup> and Le<sup>x</sup> are unusually rigid oligosaccharides. They have been described as mainly rigid, with only one low energy conformation by the early work of Lemieux and Bock [24], later confirmed by X-ray structure [53], theoretical calculations [23] and several NMR studies [54,55]. Only the  $\alpha$ Fuc1-2 linkage of some Le<sup>a</sup> analogs was proposed to display some flexibility [50]. This conformational stability has been proposed to be important to understand the recognition by antibodies [56], and to our knowledge, no other conformation has

been observed in available crystal structures of lectin/Lewis oligosaccharides complexes (see: <http://lectin3d.cermav.cnrs.fr>). Bambl is therefore the first protein that binds Lewis oligosaccharide in conformations that are very different than their solution shape. The particular localization of one tryptophan residue excludes the low energy conformation of Lewis oligosaccharide from the binding site. The strong stacking interaction that fucose establishes with this tryptophan residue in the binding site of RSL, a lectin closely related to Bambl, was recently quantified [57]. Solving the crystal structure of Bambl/Lewis x complex demonstrated that indeed, the fucose is bound with stacking to Trp, but that either the adjacent glycosidic linkages or the GlcNAc ring are distorted in order to adjust.

As reported in the literature [49,58,59], the MM-GBSA rescoring calculations strongly increase the correlation between theoretical data and quantitative thermodynamic data. Even though lectin-oligosaccharide interactions are generally enthalpy-driven (Table 3), incorporation of solvation energy term, including polar component and surface area accessibility, is necessary to obtain reliable order of binding energies of oligosaccharides due to the some entropy contribution in binding. The best correlation was achieved for LecB, for which the parameters used in the calculations were developed. The procedure selected for LecB was rather successful when applied to Bambl, demonstrating that the procedure is valid for binding sites with different characteristics (local charges, solvent exposure...). Deviations between glycan array data and MM-GBSA calculations are expected, and among the numerous sources of discrepancies one could cite the fact that the glycans are methylated, and conformationally restricted as they are immobilized on a surface, usually through covalent interactions. On the contrary, the oligosaccharides used in the microcalorimetry experiments, have a free hydroxyl in C1, and can dynamically exchange between an open form, and a closed form, either with an axial or equatorial anomeric position, while only the equatorial structures were considered in the calculations. Also, one should keep in mind that trisaccharides were used for docking studies while experimental affinities were mostly performed on their tetrasaccharide analogs.

Even though molecular dynamics simulations in explicit water environment is better suited for estimating the binding free energies, the combined use of database of oligosaccharide conformations and fast docking procedure appears as a medium-throughput screening approach for the analysis of glycan array data. To our knowledge, this is the first study combining theoretical calculations and glycan array data in the perspective of dealing with a huge flow of information such as the one coming in this Omics era.

## Supporting Information

**Figure S1 Best docked solution of Le<sup>a</sup> in the binding site of LecB. Le<sup>a</sup> is represented in color coded sticks (cyan, red and blue for carbon, oxygen and nitrogen atoms, respectively).** The calcium ions are drawn as pink spheres. The docked pose is nearly identical to the crystallographic conformation of Le<sup>a</sup> (in orange sticks), except the hydroxyl moiety on C6 of the galactose residue.  
(PDF)

**Figure S2 Crystal structure of Bambl complexed with Le<sup>x</sup> trisaccharide.** The asymmetric unit is represented in blue and magenta and the corresponding  $\beta$ -propellers obtained by 3-fold crystalline symmetry operations are represented in light blue and light pink.  
(PDF)



**Figure S3 Thermogram (top) and titration curve (bottom) obtained from the titration of LecB (150 μM) by sialyl Lewis a (1.3 mM).** Fitting procedure was performed using a one site model.  
(PDF)

**Table S1 Details of collection data and statistics.**  
(PDF)

**Table S2 Glycosidic linkage conformations for all starting models of oligosaccharides.** The  $\Phi$  and  $\Psi$  torsion angles for a glycosidic 1—x linkage are described as  $\Phi = O5-C1-O1-C'_x$  and  $\Psi = C1-O1-C'_x-C'_{x+1}$ .  
(PDF)

**Table S3 Binding intensities (FU) for labeled LecB protein with glycan array chips v4.1 from the consortium for functional glycomics.** Full data is available on the web site (<http://www.functionalglycomics.org/>).  
(PDF)

**Table S4 Binding intensities (FU) for labeled BambL protein with glycan array chips v4.1 from the consortium for functional glycomics.** Full data is available on the web site (<http://www.functionalglycomics.org/>).  
(PDF)

**Acknowledgments**

Crystal data were collected at the European Synchrotron Radiation Facility and the authors are grateful for the assistance in using beamline ID14-EH4. Glycan Array Synthesis Core (The Scripps Research Institute, LaJolla, CA) is acknowledged for excellent help in glycan array analysis. The authors would like to thank Cyril Bras for technical support.

**Author Contributions**

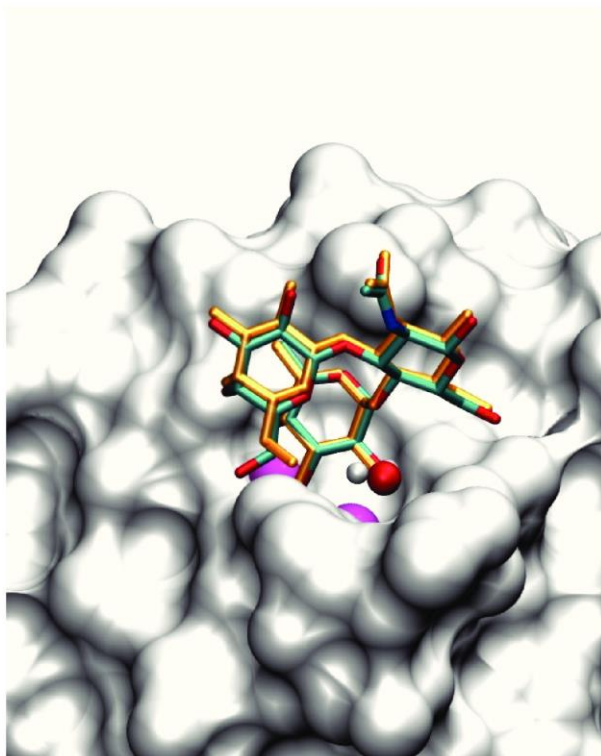
Conceived and designed the experiments: SP HJ AV AI AT. Performed the experiments: JT JA AS AA EG. Analyzed the data: JT JA AS AA SP AV AI AT. Contributed reagents/materials/analysis tools: JT AT. Wrote the paper: JT SP AV AI AT.

**References**

- Rillahan CD, Paulson JC (2011) Glycan microarrays for decoding the glycome. *Annu Rev Biochem* 80: 797–823.
- Smith DF, Cummings RD (2013) Application of microarrays for deciphering the structure and function of the human glycome. *Mol Cell Proteomics* 12: 902–912.
- Imberty A, Varrot A (2008) Microbial recognition of human cell surface glycoconjugates. *Curr Opin Struct Biol* 18: 567–576.
- Sharon N (1996) Carbohydrate-lectin interactions in infectious disease. *Adv Exp Med Biol* 408: 1–8.
- Springer SA, Gagneux P (2013) Glycan evolution in response to collaboration, conflict, and constraint. *J Biol Chem* 288: 6904–6911.
- Watkins WM, Morgan WT (1952) Neutralization of the anti-H agglutinin in eel serum by simple sugars. *Nature* 196: 825–826.
- Greenwell P (1997) Blood group antigens: molecules seeking a function? *Glycoconj J* 14: 159–173.
- Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Ruvoen N, et al. (2001) ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* 83: 565–573.
- Berger SA, Young NA, Edberg SC (1989) Relationship between infectious diseases and human blood type. *Eur J Clin Microbiol Infect Dis* 8: 681–689.
- Heggelund JE, Haugen E, Lygren B, Mackenzie A, Holmner S, et al. (2012) Both El Tor and classical cholera toxin bind blood group determinants. *Biochem Biophys Res Commun* 418: 731–735.
- Ländesmith I, Moe C, Marionneau S, Ruvoen N, Jiang X, et al. (2003) Human susceptibility and resistance to Norwalk virus infection. *Nature Med* 9: 548–553.
- Lamblin G, Degroote S, Perini JM, Delmotte P, Scharfman A, et al. (2001) Human airway mucin glycosylation: A combinatory of carbohydrate determinants which vary in cystic fibrosis. *Glycoconj J* 18: 661–684.
- Glick MC, Kothari VA, Liu A, Stoykova LI, Scanlin TF (2001) Activity of fucosyltransferases and altered glycosylation in cystic fibrosis airway epithelial cells. *Biochimie* 83: 743–747.
- Audfray A, Claudinon J, Abounit S, Ruvoen-Clouet N, Larson G, et al. (2012) Fucose-binding lectin from opportunistic pathogen *Burkholderia ambifaria* binds to both plant and human oligosaccharidic epitopes. *J Biol Chem* 287: 4335–4347.
- Garber N, Guempel U, Gilboa-Garber N, Doyle RJ (1987) Specificity of the fucose-binding lectin of *Pseudomonas aeruginosa*. *FEMS Microbiol Lett* 48: 331–334.
- Šulák O, Cioci G, Delia M, Lahmann M, Varrot A, et al. (2010) A TNF-like trimeric lectin domain from *Burkholderia cenocepacia* with specificity for fucosylated human histo-blood group antigens. *Structure* 18: 59–72.
- Šulák O, Cioci G, Lameignère E, Balloy V, Round A, et al. (2011) *Burkholderia cenocepacia* BC2L-C is a super lectin with dual specificity and proinflammatory activity. *PLoS Pathog* 7: e1002238.
- Mitchell E, Houles C, Sudakevitz D, Wimmerova M, Gautier C, et al. (2002) Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients. *Nature Struct Biol* 9: 918–921.
- Mitchell EP, Sabin C, Šnajdrová I, Pokorná M, Perret S, et al. (2005) High affinity fucose binding of *Pseudomonas aeruginosa* lectin PA-III: 1.0 Å resolution crystal structure of the complex combined with thermodynamics and computational chemistry approaches. *Proteins: Struct Funct Bioinfo* 58: 735–748.
- Perret S, Sabin C, Dumon C, Pokorná M, Gautier C, et al. (2005) Structural basis for the interaction between human milk oligosaccharides and the bacterial lectin PA-III of *Pseudomonas aeruginosa*. *Biochem J* 389: 325–332.
- Fadda E, Woods RJ (2010) Molecular simulations of carbohydrates and protein-carbohydrate interactions: motivation, issues and prospects. *Drug Discov Today* 13: 596–609.
- Imberty A, Pérez S (2000) Structure, conformation and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. *Chem Rev* 100: 4567–4588.
- Imberty A, Mikros E, Koca J, Mollicone R, Oriol R, et al. (1995) Computer simulation of histo-blood group oligosaccharides. Energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens. *Glycoconj J* 12: 331–349.
- Lemieux RU, Bock K, Delbaere LTJ, Koto S, Rao VSR (1980) The conformations of oligosaccharides related to the ABH and Lewis human blood group determinants. *Can J Chem* 58: 631–653.
- Bohne A, Lang E, von der Lieth CW (1999) SWEET - WWW-based rapid 3D construction of oligo- and polysaccharides. *Bioinformatics* 15: 767–768.
- Engelsen SB, Cros S, Mackie W, Perez S (1996) A molecular builder for carbohydrates: application to polysaccharides and complex carbohydrates. *Biopolymers* 39: 417–433.
- Rosen J, Miguet I, Perez S (2009) Shape: automatic conformation prediction of carbohydrates using a genetic algorithm. *J Cheminf* 1: 16.
- Ishida T (2010) Computational modeling of carbohydrate-recognition process in E-selectin complex: structural mapping of sialyl Lewis X onto ab initio QM/MM free energy surface. *J Phys Chem B* 114: 3950–3964.
- van Liempt E, Imberty A, Bank CMC, van Vliet SJ, van Kooyk Y, et al. (2004) Molecular basis of the differences in binding properties of the highly related C-type lectins DC-SIGN and L-SIGN to Lewis x trisaccharide and *Schistosoma mansoni* egg antigens. *J Biol Chem* 279: 33161–33167.
- Agostino M, Jene C, Boyle T, Ramsland PA, Yuriev E (2009) Molecular docking of carbohydrate ligands to antibodies: structural validation against crystal structures. *J Chem Inf Model* 49: 2749–2760.
- Imberty A, Mollicone R, Mikros E, Carrupt PA, Pérez S, et al. (1996) How do antibodies and lectins recognize histo-blood group antigens? A 3D-QSAR study by comparative molecular field analysis (CoMFA). *Bioorg Med Chem* 4: 1979–1988.
- Koppisetty CA, Nasir W, Strino F, Rydell GE, Larson G, et al. (2010) Computational studies on the interaction of ABO-active saccharides with the norovirus VA387 capsid protein can explain experimental binding data. *J Comput Aided Mol Des* 24: 423–431.
- Mishra NK, Kriz Z, Wimmerova M, Koca J (2010) Recognition of selected monosaccharides by *Pseudomonas aeruginosa* Lectin II analyzed by molecular dynamics and free energy calculations. *Carbohydr Res* 345: 1432–1441.
- Mishra NK, Kulhánek P, Šnajdrová I, Petřek M, Imberty A, et al. (2008) Molecular dynamics study of *Pseudomonas aeruginosa* lectin-II complexed with monosaccharides. *Proteins* 72: 382–392.
- Mishra SK, Adam J, Wimmerová M, Koča J (2012) In silico mutagenesis and docking study of *Ralstonia solanacearum* RSL Lectin: Performance of docking software to predict saccharide binding. *J Chem Inf Model* 52: 1250–1261.
- Mishra SK, Sund J, Aqvist J, Koca J (2012) Computational prediction of monosaccharide binding free energies to lectins with linear interaction energy models. *J Comput Chem* 33: 2340–2350.
- Nuriso A, Kozmon S, Imberty A (2008) Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III. *Mol Simul* 34: 469–479.
- Smith DF, Song X, Cummings RD (2010) Use of glycan microarrays to explore specificity of glycan-binding proteins. *Methods Enzymol* 480: 417–444.
- Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography* 26.

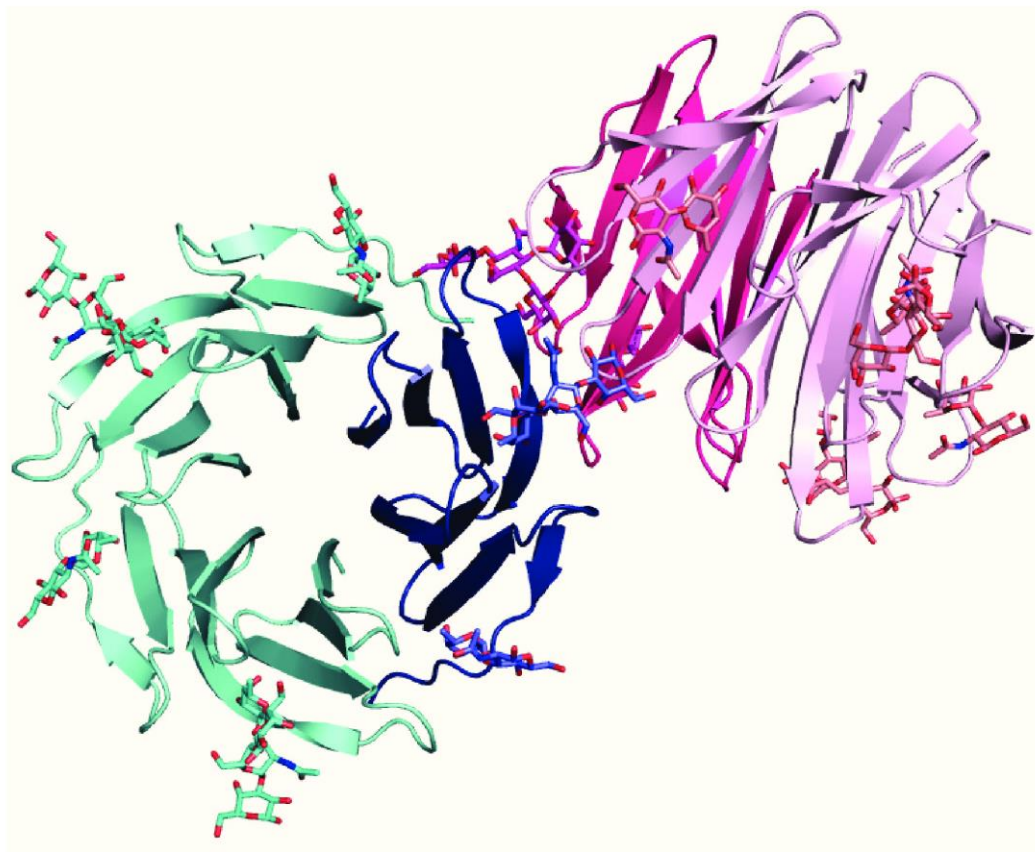
Investigating Lectins by Glycan Array and Docking

40. Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50: 760–763.
41. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40: 658–674.
42. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67: 355–367.
43. Emsley P, Lohkamp B, Scott W, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501.
44. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242.
45. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Parameterization of the OPLSAA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105: 6474–6487.
46. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, et al. (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47: 1739–1749.
47. Repasky MP, Shelley M, Friesner RA (2007) Flexible ligand docking with Glide. *Curr Protoc Bioinformatics Chapter 8: Unit 8 12*.
48. Ali MM, Aich U, Perez S, Imberty A, Loganathan D (2009) Examination of the effect of structural variation on the N-glycosidic torsion (PhiN) among N-(beta-D-glycopyranosyl)acetamido and propionamido derivatives of monosaccharides based on crystallography and quantum chemical calculations. *Carbohydr Res* 344: 355–361.
49. Lyne PD, Lamb ML, Saeh JC (2006) Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J Med Chem* 49: 4805–4808.
50. Auzanneau FI, Jackson TA, Liao L (2012) Stochastic searches and NMR experiments on four Lewis A analogues: NMR experiments support some flexibility around the fucosidic bond. *Bioorg Med Chem* 20: 5085–5093.
51. Porter A, Yue T, Heeringa I, Day S, Suh E, et al. (2010) A motif-based analysis of glycan array data to determine the specificities of glycan-binding proteins. *Glycobiology* 20: 369–380.
52. Xuan P, Zhang Y, Tzeng TR, Wan XF, Luo F (2012) A quantitative structure-activity relationship (QSAR) study on glycan array data to determine the specificities of glycan-binding proteins. *Glycobiology* 22: 552–560.
53. Pérez S, Mouhous-Riou N, Nifant'ev NE, Tsvetkov YE, Bachet B, et al. (1996) Crystal and molecular structure of a histo-blood group antigen involved in cell adhesion: the Lewis x trisaccharide. *Glycobiology* 6: 537–542.
54. Azurmendi HF, Martin-Pastor M, Bush CA (2002) Conformational studies of Lewis X and Lewis A trisaccharides using NMR residual dipolar couplings. *Biopolymers* 63: 89–98.
55. Bekiroglu S, Kenne I, Sandstrom C (2004) NMR study on the hydroxy protons of the Lewis X and Lewis Y oligosaccharides. *Carbohydr Res* 339: 2465–2468.
56. Yuriev E, Farrugia W, Scott AM, Ramsland PA (2005) Three-dimensional structures of carbohydrate determinants of Lewis system antigens: implications for effective antibody targeting of cancer. *Immunol Cell Biol* 83: 709–717.
57. Wimmerova M, Kozmon S, Necasova I, Mishra SK, Komarek J, et al. (2012) Stacking interactions between carbohydrate and protein quantified by combination of theoretical and experimental methods. *PLoS One* 7: e46032.
58. Du J, Sun H, Xi L, Li J, Yang Y, et al. (2011) Molecular modeling study of checkpoint kinase 1 inhibitors by multiple docking strategies and prime/MM-GBSA calculation. *J Comput Chem* 32: 2800–2809.
59. Haider MK, Bertrand HO, Hubbard RE (2011) Predicting Fragment Binding Poses Using a Combined MCSS MM-GBSA Approach. *J Chem Inf Model*.

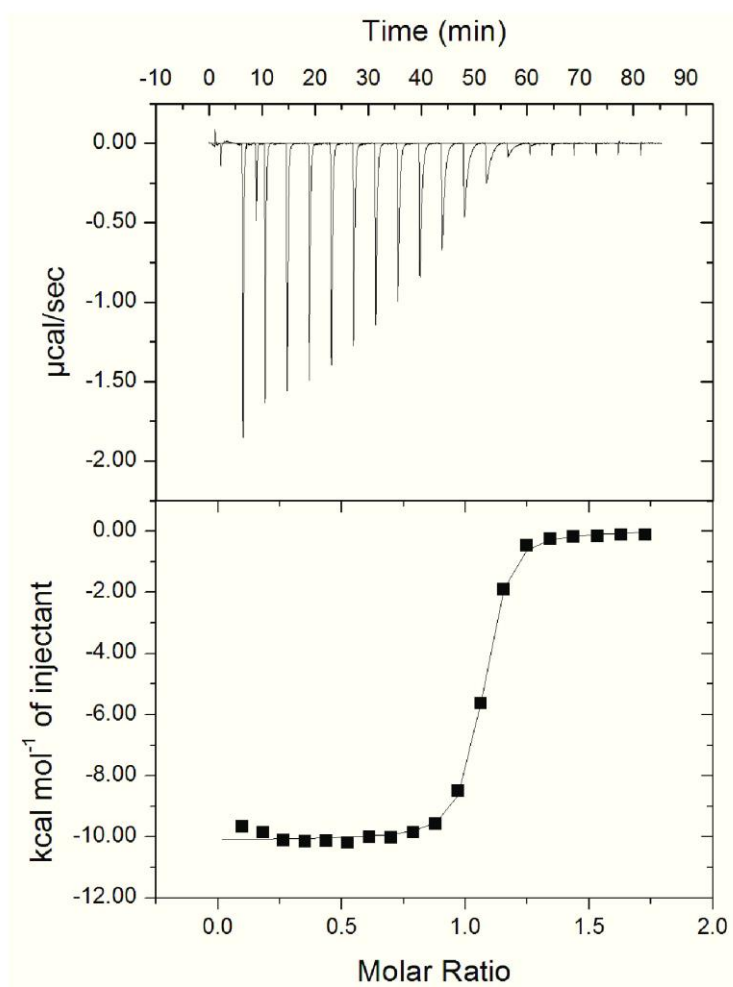


**Figure S1:** Best docked solution of Le<sup>a</sup> in the binding site of LecB. Le<sup>a</sup> is represented in color coded sticks (cyan, red and blue for carbon, oxygen and nitrogen atoms, respectively). The calcium ions are drawn as pink spheres. The docked pose is nearly identical to the crystallographic conformation of Le<sup>a</sup> (in orange sticks), except the hydroxyl moiety on C6 of the galactose residue.





**Figure S2:** Crystal structure of BamL complexed with Le<sup>x</sup> trisaccharide. The asymmetric unit is represented in blue and magenta and the corresponding  $\beta$ -propellers obtained by 3-fold crystalline symmetry operations are represented in light blue and light pink.



**Figure S3:** Thermogram (top) and titration curve (bottom) obtained from the titration of LecB (150  $\mu\text{M}$ ) by sialyl Lewis a (1.3 mM). Fitting procedure was performed using a one site model.

**Table S1:** Details of collection data and statistics.

<b>Protein</b>	<b>BambL/leX</b>
<b>Data collection</b>	
Wavelength (Å)	1.140
Space group	P 2 <sub>1</sub> 3
Cell dimensions a, b, c (Å)	81.4, 81.4, 81.4
Resolution (outer shell) (Å)	40.72-1.60 (1.69-1.60)
Measured/Unique reflections	21930/1180
Average multiplicity (outer shell)	6.2 (3.4)
Rmerge (outer shell)	0.04 (0.1)
Completeness (%) (outer shell)	96.2 (82.4)
Mean I/oI (outer shell)	26.6 (7.9)
Wilson B	17.4
<b>Refinement</b>	
Rcryst/Rfree	15.5/18.9
r.m.s.d. bonds (Å)	0.016
r.m.s.d. angles (°)	1.71
r.m.s.d. chiral (Å <sup>3</sup> )	0.10
Protein atoms	1357
Bfac (Å <sup>2</sup> )	14.4
Ligand atoms	156
Bfac (Å <sup>2</sup> )	21.0
Water molecules	237
Bfac (Å <sup>2</sup> )	27.72
PDB code	3ZW1





Table S3: Binding intensities (FU) for labeled LecB protein with glycan array chips v4.1 from the consortium for functional glycomics. Full data is available on the web site

(<http://www.functionalglycomics.org/>)

	Glycan Structure	10 µg/ml	1 µg/ml	0.1 µg/ml
		Average	Average	Average
H-di	Fuca1-2Galβ-Sp8	9118	1947	115
H_type1	Fuca1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-Sp8	5510	1003	20
H-type1	Fuca1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-Sp10	3834	853	21
H-type1	Fuca1-2Galβ1-3GlcNAcβ-Sp0	4526	902	29
H-type1	Fuca1-2Galβ1-3GlcNAcβ-Sp8	3393	636	12
H-type1	Fuca1-2Galβ1-3GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)Galβ1-4Glc-Sp21	6054	1009	51
H-type2	Fuca1-2Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	13494	2887	284
H-type2	Fuca1-2Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	15404	3167	199
H-type2	Fuca1-2Galβ1-4GlcNAcβ-Sp0	13127	2953	309
H-type2	Fuca1-2Galβ1-4GlcNAcβ-Sp8	10059	2235	132
H-type3	Fuca1-2Galβ1-3GalNAcα-Sp8	5093	984	18
H-type4	Fuca1-2Galβ1-3GalNAcβ1-3Galα-Sp9	2119	358	14
H-type4	Fuca1-2Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ-Sp9	12579	2452	169
H-type5	Fuca1-2Galβ1-4Glcβ-Sp0	6580	1726	185
A-tri	GalNAcα1-3(Fuca1-2)Galβ-Sp8	5986	1033	66
A-tri	GalNAcα1-3(Fuca1-2)Galβ-Sp18	6016	1079	71
A-type1	GalNAcα1-3(Fuca1-2)Galβ1-3GlcNAcβ-Sp0	14	4	2
A-type2	GalNAcα1-3(Fuca1-2)Galβ1-4GlcNAcβ-Sp0	3541	651	11
A-type2	GalNAcα1-3(Fuca1-2)Galβ1-4GlcNAcβ-Sp8	2581	529	5
A-type2	GalNAcα1-3(Fuca1-2)Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	7514	1318	93
A-type2	GalNAcα1-3(Fuca1-2)Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	5471	827	35
A-type5	GalNAcα1-3(Fuca1-2)Galβ1-4Glcβ-Sp0	2452	487	10
A-LewisY	GalNAcα1-3(Fuca1-2)Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	2348	434	16
B-tri	Galα1-3(Fuca1-2)Galβ-Sp8	2419	595	4
B-tri	Galα1-3(Fuca1-2)Galβ-Sp18	2447	564	17
B-type1	Galα1-3(Fuca1-2)Galβ1-3GlcNAcβ-Sp0	1127	177	5
B-type1	Galα1-3(Fuca1-2)Galβ1-3GlcNAcβ-Sp8	761	67	2
B-type2	Galα1-3(Fuca1-2)Galβ1-4GlcNAc-Sp0	2073	443	5
B-type3	Galα1-3(Fuca1-2)Galβ1-3GalNAcα-Sp8	65	8	4
B-type4	Galα1-3(Fuca1-2)Galβ1-3GalNAcβ-Sp8	83	6	9
B-type5	Galα1-3(Fuca1-2)Galβ1-4Glcβ-Sp0	638	178	4
B-LewisY	Galα1-3(Fuca1-2)Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	2172	206	13
B-LewisY	Galα1-3(Fuca1-2)Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	2972	460	17
LewisA	Galβ1-3(Fuca1-4)GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	7682	1192	60
LewisA	Galβ1-3(Fuca1-4)GlcNAcβ-Sp0	15130	3296	228
LewisA	Galβ1-3(Fuca1-4)GlcNAcβ-Sp8	12765	2398	166
Sialyl_LewisA	Neu5Acα2-3Galβ1-3(Fuca1-4)GlcNAcβ-Sp8	23197	5119	430
LewisA_sulfo	[3OSO3]Galβ1-3(Fuca1-4)GlcNAcβ-Sp8	15792	4167	405
LewisB	Fuca1-2Galβ1-3(Fuca1-4)GlcNAcβ-Sp8	7499	1291	95
LewisB	Fuca1-2Galβ1-3(Fuca1-4)GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)Galβ1-4Glc-Sp21	4620	973	50
LewisX	Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	3491	678	23
LewisX	Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	3883	580	27
LewisX	Galβ1-4(Fuca1-3)GlcNAcβ1-4Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	6346	1156	63
LewisX	Galβ1-3GlcNAcβ1-3(Galβ1-4(Fuca1-3)GlcNAcβ1-6)Galβ1-4Glc-Sp21	3016	575	21
Sialyl_LewisX	Neu5Acα2-3Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	2728	405	21
Sialyl_LewisX	Neu5Acα2-3Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	3899	652	25
Sialyl_LewisX	Neu5Acα2-3Galβ1-4(Fuca1-3)GlcNAcβ1-3Galβ-Sp8	2178	636	9
Sialyl_LewisX	Neu5Acα2-3Galβ1-4(Fuca1-3)GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp8	3911	662	26
LewisX_sulfo	[3OSO3]Galβ1-4(Fuca1-3)GlcNAc-Sp0	5712	1010	111
LewisX_sulfo	[3OSO3]Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	6193	1088	49
LewisX_sulfo	Galβ1-4(Fuca1-3)[6OSO3]GlcNAc-Sp0	3588	728	26
LewisY	Fuca1-2Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	7491	1265	76
LewisY	Fuca1-2Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	6219	1063	27

Table S4: Binding intensities (FU) for labeled BambL protein with glycan array chips v4.1 from the consortium for functional glycomics. Full data is available on the web site (<http://www.functionalglycomics.org/>)

	Glycan Structure	10 µg/ml Average	1 µg/ml Average	0.1 µg/ml Average
H-di	Fuca1-2Galβ-Sp8	1	5	3
H_type1	Fuca1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-Sp8	2321	1027	683
H-type1	Fuca1-2Galβ1-3GlcNAcβ1-3Galβ1-4Glcβ-Sp10	1526	1999	186
H-type1	Fuca1-2Galβ1-3GlcNAcβ-Sp0	2419	1818	919
H-type1	Fuca1-2Galβ1-3GlcNAcβ-Sp8	13063	9008	3686
H-type1	Fuca1-2Galβ1-3GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)Galβ1-4Glc-Sp21	970	155	126
H-type2	Fuca1-2Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	6	1	0
H-type2	Fuca1-2Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	3	9	5
H-type2	Fuca1-2Galβ1-4GlcNAcβ-Sp0	4	2	3
H-type2	Fuca1-2Galβ1-4GlcNAcβ-Sp8	-1	4	2
H-type3	Fuca1-2Galβ1-3GalNAcα-Sp8	3927	2990	1303
H-type4	Fuca1-2Galβ1-3GalNAcβ1-3Galα-Sp9	342	48	40
H-type4	Fuca1-2Galβ1-3GalNAcβ1-3Galα1-4Galβ1-4Glcβ-Sp9	9549	6167	4179
H-type5	Fuca1-2Galβ1-4Glcβ-Sp0	4	4	5
A-tri	GalNAcα1-3(Fuca1-2)Galβ-Sp8	5	8	3
A-tri	GalNAcα1-3(Fuca1-2)Galβ-Sp18	2	2	3
A-type1	GalNAcα1-3(Fuca1-2)Galβ1-3GlcNAcβ-Sp0	2	6	5
A-type2	GalNAcα1-3(Fuca1-2)Galβ1-4GlcNAcβ-Sp0	0	10	0
A-type2	GalNAcα1-3(Fuca1-2)Galβ1-4GlcNAcβ-Sp8	3	2	1
A-type2	GalNAcα1-3(Fuca1-2)Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	1	7	4
A-type2	GalNAcα1-3(Fuca1-2)Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	1	9	0
A-type5	GalNAcα1-3(Fuca1-2)Galβ1-4Glcβ-Sp0	2	-2	13
A-LewisY	GalNAcα1-3(Fuca1-2)Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	2	7	5
B-tri	Galα1-3(Fuca1-2)Galβ-Sp8	5	8	4
B-tri	Galα1-3(Fuca1-2)Galβ-Sp18	3	2	4
B-type1	Galα1-3(Fuca1-2)Galβ1-3GlcNAcβ-Sp0	7	3	3
B-type1	Galα1-3(Fuca1-2)Galβ1-3GlcNAcβ-Sp8	1122	517	29
B-type2	Galα1-3(Fuca1-2)Galβ1-4GlcNAc-Sp0	6	12	5
B-type3	Galα1-3(Fuca1-2)Galβ1-3GalNAcα-Sp8	37	5	1
B-type4	Galα1-3(Fuca1-2)Galβ1-3GalNAcβ-Sp8	77	1205	292
B-type5	Galα1-3(Fuca1-2)Galβ1-4Glcβ-Sp0	2	2	2
B-LewisY	Galα1-3(Fuca1-2)Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	6	15	8
B-LewisY	Galα1-3(Fuca1-2)Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	389	52	5
LewisA	Galβ1-3(Fuca1-4)GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp0	0	-2	3
LewisA	Galβ1-3(Fuca1-4)GlcNAcβ-Sp0	2	9	-5
LewisA	Galβ1-3(Fuca1-4)GlcNAcβ-Sp8	3	1	3
Sialyl_LewisA	Neu5Acα2-3Galβ1-3(Fuca1-4)GlcNAcβ-Sp8	4	9	2
LewisA_sulfo	[3OSO3]Galβ1-3(Fuca1-4)GlcNAcβ-Sp8	5	6	-1
LewisB	Fuca1-2Galβ1-3(Fuca1-4)GlcNAcβ-Sp8	25753	13533	6647
LewisB	Fuca1-2Galβ1-3(Fuca1-4)GlcNAcβ1-3(Galβ1-4GlcNAcβ1-6)Galβ1-4Glc-Sp21	13453	10217	5405
LewisX	Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	3	4	-2
LewisX	Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	3	5	0
LewisX	Galβ1-4(Fuca1-3)GlcNAcβ1-4Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	1	-5	3
LewisX	Galβ1-3GlcNAcβ1-3(Galβ1-4(Fuca1-3)GlcNAcβ1-6)Galβ1-4Glc-Sp21	6	1	4
Sialyl_LewisX	Neu5Acα2-3Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	3	18	-2
Sialyl_LewisX	Neu5Acα2-3Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	4	2	3
Sialyl_LewisX	Neu5Acα2-3Galβ1-4(Fuca1-3)GlcNAcβ1-3Galβ-Sp8	5	3	0
Sialyl_LewisX	Neu5Acα2-3Galβ1-4(Fuca1-3)GlcNAcβ1-3Galβ1-4GlcNAcβ-Sp8	9	16	5
LewisX_sulfo	[3OSO3]Galβ1-4(Fuca1-3)GlcNAc-Sp0	14	0	3
LewisX_sulfo	[3OSO3]Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	10	8	9
LewisX_sulfo	Galβ1-4(Fuca1-3)[6OSO3]GlcNAc-Sp0	4	19	3
LewisY	Fuca1-2Galβ1-4(Fuca1-3)GlcNAcβ-Sp0	1407	435	91
LewisY	Fuca1-2Galβ1-4(Fuca1-3)GlcNAcβ-Sp8	1632	2097	287



## II.2. Spécificité de RSL pour les épitopes Le<sup>x</sup>

### II.2.1. Introduction

De nombreux oligosaccharides voient leur expression modifiée dans les processus cancéreux. Les épitopes fucosylés de type Lewis en font partie et sont surexprimés dans de nombreux cancers. Par exemple, le SLe<sup>x</sup> est une molécule impliquée dans l'adhérence des cellules des carcinomes du côlon, du poumon, du pancréas, du foie et de l'estomac ainsi que dans les leucémies (Takada *et al.*, 1993). Des études récentes confirment que la surexpression des antigènes Lewis est observée dans de nombreux cancers avec par exemple une surexpression de l'oligosaccharide SLe<sup>x</sup> dans les carcinomes du rein (Borzym-Kluczyk *et al.*, 2014), et dans le cancer du tractus gastro-intestinal (Dube and Bertozzi, 2005; Groux-Degroote *et al.*, 2014; Zhang *et al.*, 1997). De même une surexpression du Le<sup>x</sup> a été mise en évidence dans de nombreux cancers comme celui du pancréas, colon, sein, poumon ou encore ovaire et prostate (Dube and Bertozzi, 2005; Groux-Degroote *et al.*, 2014; Zhang *et al.*, 1997).

Une lectine capable de lier ces épitopes avec une forte spécificité pourrait être d'un grand intérêt à la fois pour le diagnostic mais aussi à plus long terme pour un traitement.

La spécificité de la protéine RSL a été étudiée dans un premier temps par SPR (Kostlanova *et al.*, 2005) et les auteurs ont établi l'ordre suivant de spécificité pour les épitopes oligosaccharidiques : Groupe sanguin B > Groupe sanguin A > αMeFuc > Groupe sanguin H type I > L-fucose > Le<sup>a</sup> > Le<sup>x</sup> > SLe<sup>x</sup> avec des valeurs de constante de dissociation variant de 0.21 μM pour le trisaccharide de groupe sanguin B jusqu'à 14.4 μM et 18.7 μM pour le Le<sup>x</sup> et SLe<sup>x</sup> respectivement. L'affinité de RSL pour les Lewis est donc relativement modeste.

Du point de vue de sa conformation, les épitopes de Lewis sont des glycanes particuliers du fait de leur manque de flexibilité (Imberty *et al.*, 1995). En effet, le résidu GlcNAc porte les résidus fucose et galactose en position 3 et 4, résultant en un empilement des deux cycles (Figure 49). Cette conformation « fermée » du Le<sup>x</sup> a été décrite en solution (Zierke *et al.*, 2013) et est observée dans toutes les structures cristallographiques de complexe lectine-Le<sup>x</sup> (database Lecin-3D). La seule exception est la lectine BambL où nous avons précédemment mis en évidence que le Le<sup>x</sup> était présent dans les sites de fixation de cette protéine sous une forme distordue ouverte non habituelle (Topin *et al.*, 2013). RSL et BambL

sont très similaire (82 % identité de séquence). Avant toute étude d'ingénierie visant à améliorer l'affinité des néoRSL pour le Le<sup>x</sup> et le SLe<sup>x</sup>, il est donc nécessaire d'établir les bases moléculaires de l'interaction dans les deux sites de la lectine RSL. Les interactions entre RSL et le Le<sup>x</sup> ont donc été analysées finement par ITC et cristallographie. Quand cela a été possible, les deux sites ont été analysés séparément.

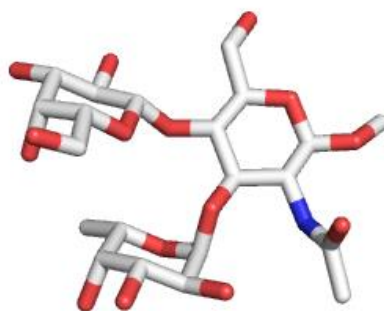


Figure 49 : Conformation fermée du Le<sup>x</sup>

## II.2.2. Etude de la spécificité de RSL et néoRSL\_VI

Afin d'imaginer comment modifier l'affinité et la spécificité de RSL pour le Le<sup>x</sup> ou le SLe<sup>x</sup>, il est important de bien connaître la spécificité de RSL. Pour ce faire, une analyse systématique de l'affinité de RSL par ITC a été réalisée en testant une série d'épitopes fucosylés des groupes H et Lewis. La néoRSL\_VI a également été analysée afin de vérifier que la différence d'architecture n'affecte pas l'ordre de la spécificité.

	RSL			néoRSL_VI		
	Kd μM	ΔH KJ/mol	TΔS KJ/mol	Kd μM	ΔH KJ/mol	TΔS KJ/mol
LeY (penta)	7.04 ± 0.01	-35.4 ± 0.1	-6.0 ± 0.1	10.3 ± 0.2	-34.7 ± 0.04	-6.22 ± 0.09
H type 1 (tetra)	13.8 ± 0.2	-29.8 ± 0.5	-2.1 ± 0.5	12 ± 2	-27 ± 1	1.47 ± 0.7
H type 2 (tetra)	4.2 ± 0.1	-38.2 ± 0.3	-7.4 ± 0.4	4.0 ± 0.8	-36.3 ± 0.8	-5.45 ± 1.4
H type 4 (tetra)	24.5 ± 0.6	-27.23 ± 0.02	-0.90 ± 0.08	21 ± 8	-23 ± 1	3.5 ± 2.5
LeA (tetra)	18.8 ± 0.7	-29.4 ± 0.2	-2.4 ± 0.3	25 ± 1	-31.2 ± 0.2	-5 ± 0.3
LeX (tetra)	26 ± 2	-32.5 ± 0.3	-6.3 ± 0.5	30 ± 4	-29.9 ± 0.3	-4.1 ± 0.6
SLeX (penta)	58 ± 3	-19.6 ± 0.1	4.6 ± 0.3	76 ± 2	-20.7 ± 0.3	2.8 ± 0.2

Tableau 10 : Comparaison de la spécificité de RSL et de néoRSL\_VI

La stœchiométrie varie entre 1.75 et 2 par monomère de RSL et 4.4 et 5.2 pour néoRSL, ce qui correspond aux valeurs attendues. (tetra) et (penta) indique la taille de l'oligosaccharide (tetraose ou pentaose).

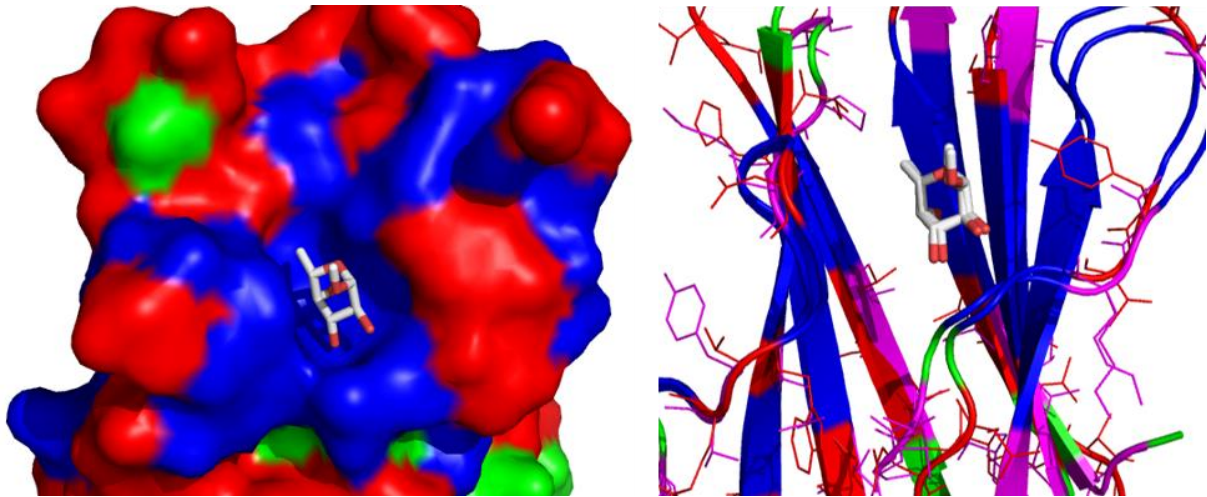


L'affinité de RSL ou de néoRSL\_VI pour les oligosaccharides fucosylés est de l'ordre du micromolaire. Il n'y a que très peu de différence (variation de 10 %) entre la forme naturelle et la néoRSL. Les affinités sont bien du même ordre de grandeur que celles observées précédemment par SPR (Kostlanova *et al.*, 2005). L'oligosaccharide Le<sup>y</sup> est le ligand de plus forte affinité, ce qui est peut-être dû à la présence de deux résidus fucoses. Les épitopes Le<sup>x</sup> et SLe<sup>x</sup> ont une affinité significativement plus faible que le H-type 2 (Tableau 10).

Concernant les données thermodynamiques, les interactions lectines/sucres présentent toujours une enthalpie très favorable avec des réactions très exothermiques. Dans la plupart des cas l'entropie est défavorable. Elle est presque nulle dans le cas du H type 4 avec une différence de comportement entre RSL et néoRSL. Dans le cas de l'interaction de RSL ou néoRSL\_VI avec le SLe<sup>x</sup>, la contribution enthalpique est moins favorable que celle observée pour les autres oligosaccharides, mais par contre la contribution entropique est favorable et contribue de 8 % à l'énergie de binding (Tableau 10).

### **II.2.2.1. Différence de spécificité entre les sites intra et intermonomériques de RSL**

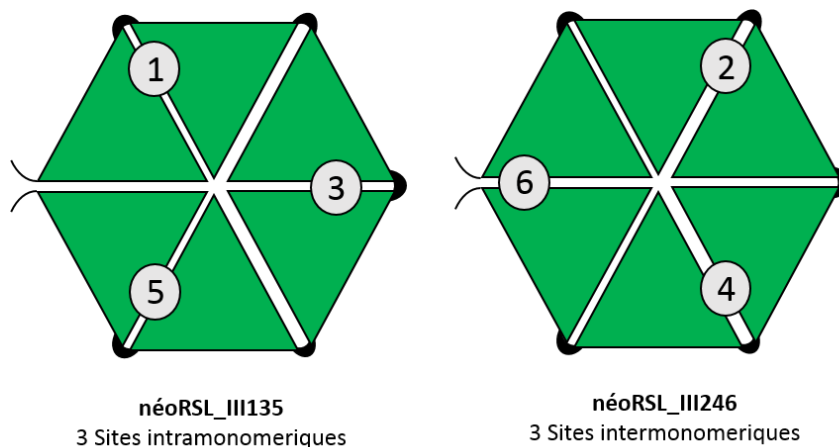
Il faut bien noter que les valeurs ci-dessus représentent les moyennes entre plusieurs sites : deux sites dans le cas de la RSL (site intramonomériques et sites intermonomériques) et six sites dans le cas de la néoRSL (alternance de sites de type intra et inter). Les sites de fixation intra et intermonomériques de RSL sont très similaires. En effet, les acides aminés principalement impliqués dans la stabilisation du fucose sont les mêmes. Il s'agit de l'arginine 17 ou 62 et du glutamate 28 ou 73 qui stabilisent le fucose par des liaisons hydrogène mais aussi de l'alanine 40 ou 85 et des tryptophanes 76 et 81 ou 31 et 36 qui stabilisent le fucose par des interactions hydrophobes respectivement dans le site intramonomérique ou intermonomérique. Cependant, les deux sites ne sont pas identiques et il existe aussi des différences qui peuvent induire une variation dans la spécificité de chaque site de fixation envers les oligosaccharides (Figure 50).



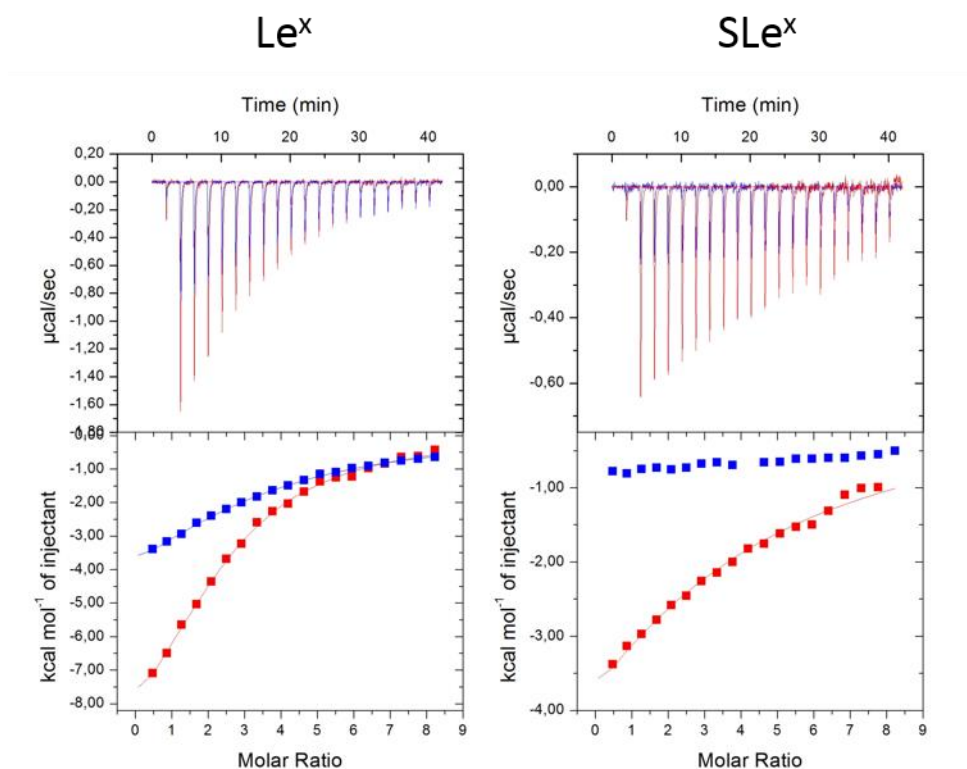
**Figure 50 : Points communs et différences entre les sites intra et intermonomériques.**

Représentation de la surface de la protéine RSL autour du site intramonomérique de RSL. En bleu sont représentés les acides aminés qui sont identiques dans le site intermonomérique, en vert sont représentés les acides aminés équivalents et en rouge les différences. A droite est représentée la superposition des sites intra et intermonomériques. En bleue sont représentés les acides aminés identiques, en vert les équivalents et en rouges et rose des différents. Le rouge représente les acides aminés du site intramonomérique et en rose ceux du site intermonomérique.

Afin de caractériser les différences fines de spécificité entre ces deux sites, nous avons utilisé deux des protéines néoRSL produites précédemment (Arnaud *et al.*, 2014). Dans la protéine néoRSL\_III<sub>135</sub>, les trois sites intermonomériques sont inactivés par la substitution de l'arginine en alanine. La néoRSL\_III<sub>135</sub> possède donc seulement trois sites intramonomériques actifs. Inversement, la protéine néoRSL\_III<sub>246</sub> ne possède que trois sites intermonomériques actifs (Figure 51).



*Figure 51 : Représentation schématique des protéines néoRSL\_III135 et néoRSL\_III246*



*Figure 52 : Comparaison de la fixation du Le<sup>x</sup> et du SLe<sup>x</sup> dans les sites intra et intermonomériques*

*Résultats d'ITC obtenus avec l'ITC 200 en injectant des oligosaccharides à une concentration de 2.5mM dans une protéine à 0.3mM. En bleu sont représentés les résultats obtenus avec la protéine néoRSL\_III<sub>246</sub> qui présente trois sites de fixation intermonomériques. En rouge sont représentés les résultats obtenus avec la protéine néoRSL\_III<sub>135</sub> qui présente trois sites de fixation intramonomériques.*

Sucre	NéoRSL_III246 3 sites intermonomériques			NéoRSL_III135 3 sites intramonomériques		
	Kd ( $\mu$ M)	dH (kJ/mol)	TdS (J/mol)	Kd ( $\mu$ M)	dH (kJ/mol)	TdS (J/mol)
Le <sup>x</sup> tetraose	112 $\pm$ 24,6	-29,6 $\pm$ 0,36	-7,5 $\pm$ 1,5	47,5 $\pm$ 9,3	-43,3 $\pm$ 6,2	-18,5 $\pm$ 5,7
sLe <sup>x</sup> pentaose	/	/	/	283,1 $\pm$ 50,2	-54 $\pm$ 0,4	-33,8 $\pm$ 1

**Tableau 11 : Comparaison de l'affinité des sites intra et intermonomériques le Le<sup>x</sup> et le SLe<sup>x</sup> par ITC**

*Les données ont été obtenues avec l'ITC200 (Microcal-GE Healthcare).*

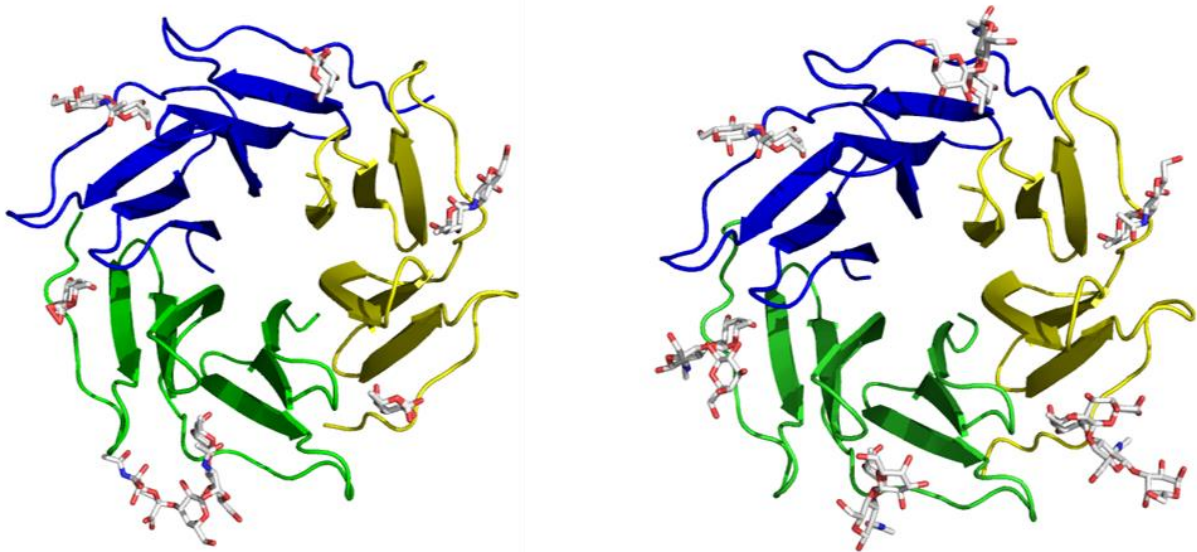
Les affinités obtenues pour les mutants trivalents sont plus faibles en moyenne d'un facteur 3 par rapport à celles obtenues avec la protéine hexasaccharidique néoRSL\_VI ce qui indique peut-être une déstabilisation de ces mutants. Les caractéristiques thermodynamiques de liaison sont celles que l'on trouve généralement pour des interactions lectines/sucres. En effet, l'enthalpie est favorable alors que l'entropie est défavorable (Figure 52 et Tableau 11). On peut remarquer que l'affinité pour le Le<sup>x</sup> est différente entre le site intermonomérique et le site intramonomérique. Ce dernier ayant une affinité deux fois plus forte pour ce sucre.

Les données thermodynamiques correspondantes aux interactions avec le SLe<sup>x</sup> sont plus surprenantes (Figure 52 et Tableau 11). La protéine néoRSL\_III135 qui possède trois sites intramonomériques actifs présente une affinité pour le SLe<sup>x</sup> alors que pour la protéine néoRSL\_III246, aucune affinité n'a pu être détectée par ITC. Ainsi dans la protéine RSL, seul le site intramonomérique peut lier le SLe<sup>x</sup>. Donc la faible différence en terme d'acides aminés dans et autour du site de fixation suffit pour avoir une affinité ou non avec cet oligosaccharide. L'affinité de RSL pour le Le<sup>x</sup> et le SLe<sup>x</sup> montre des particularités intéressantes, une étude plus approfondie a donc été réalisée.

### II.2.2.2. Structures cristallographiques des complexes RSL/ Le<sup>x</sup> et RSL/ SLe<sup>x</sup>

Les complexes RSL-Le<sup>x</sup> et RSL-SLe<sup>x</sup> ont été cristallisés dans le groupe d'espace P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> avec des paramètres de mailles cristallines de a = 43 Å b = 83.2 Å et c = 109.3 Å et a = 36.8 Å b = 83.2 Å et c = 90.7 Å respectivement. Chaque unité asymétrique contient un trimère de RSL. Dans le modèle final chaque monomère est composé de 89 acides aminés, la méthionine N-terminale n'est jamais présente. Aucun changement de conformation de la

protéine n'a été observé entre les deux structures de RSL ou par rapport au modèle de la structure. Seuls de légers mouvements sont présents au niveau de certaines boucles de surface.

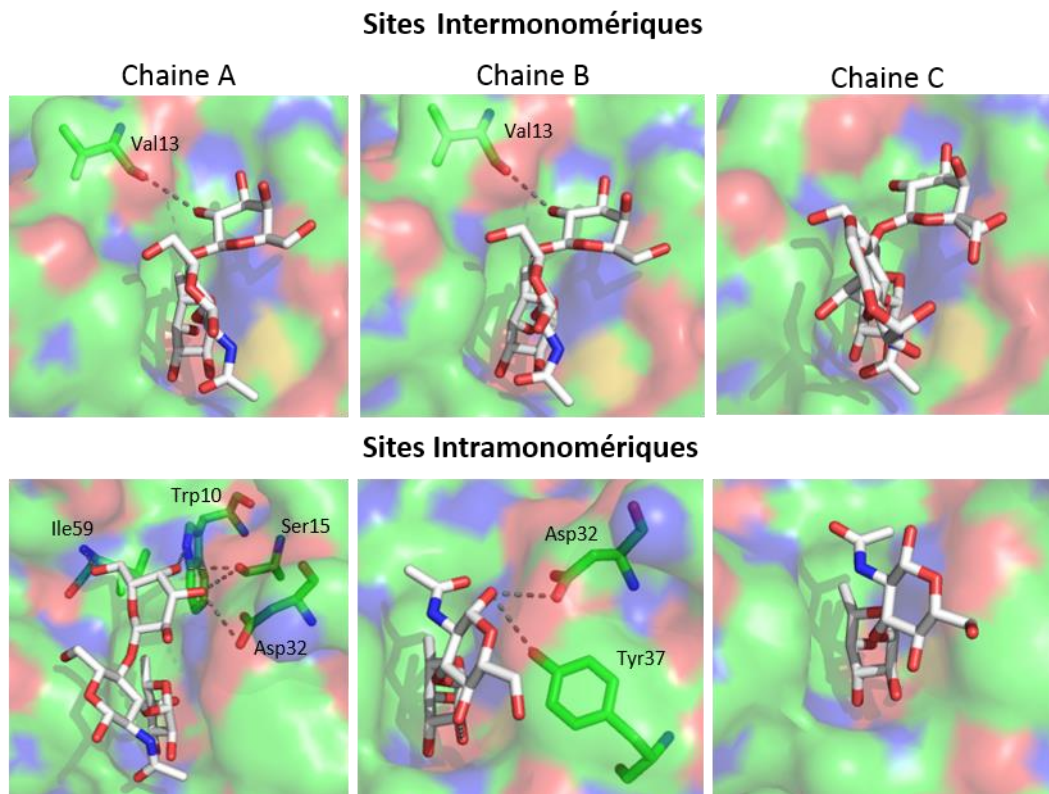


**Figure 53 : Structure des complexes RSL/SLe<sup>x</sup> et RSL/Le<sup>x</sup>**

*A gauche est représenté le complexe RSL/SLe<sup>x</sup> et à droite le complexe RSL/Le<sup>x</sup>.*

Dans la structure du complexe RSL/Le<sup>x</sup> chacun des sites est occupé par l'oligosaccharide mais certaines parties ne sont pas visibles dans la densité électronique (Figure 53). Dans les sites intramonomériques, la densité pour un disaccharide Fuca1-3GlcNAc est visible dans deux sites alors que le trisaccharide Le<sup>x</sup> se voit clairement dans le troisième site. Dans les sites intermonomériques, le trisaccharide et même le glucose terminal du tetrasaccharide sont clairement visibles.

Le fucose des oligosaccharides occupe la même position que dans le complexe RSL/fucose précédemment décrit (Kostlanova *et al.*, 2005). Le GlcNAc n'interagit pas toujours directement avec la protéine : c'est seulement dans le site intramonomérique de la chaîne B que des liaisons hydrogène sont établies. On observe alors un contact entre l'O1 du GlcNAc et l'arginine 17 et la tyrosine 37. Le galactose est visible dans le site intramonomérique de la chaîne A et interagit avec le tryptophane 10, la serine 15, l'aspartate 31 et l'isoleucine 59. Dans les sites intermonomériques de la chaîne A et B, ce galactose interagit avec la valine 13. Ce même galactose n'établit aucune liaison hydrogène dans les sites intra et intermonomériques de la chaîne C (Figure 54).

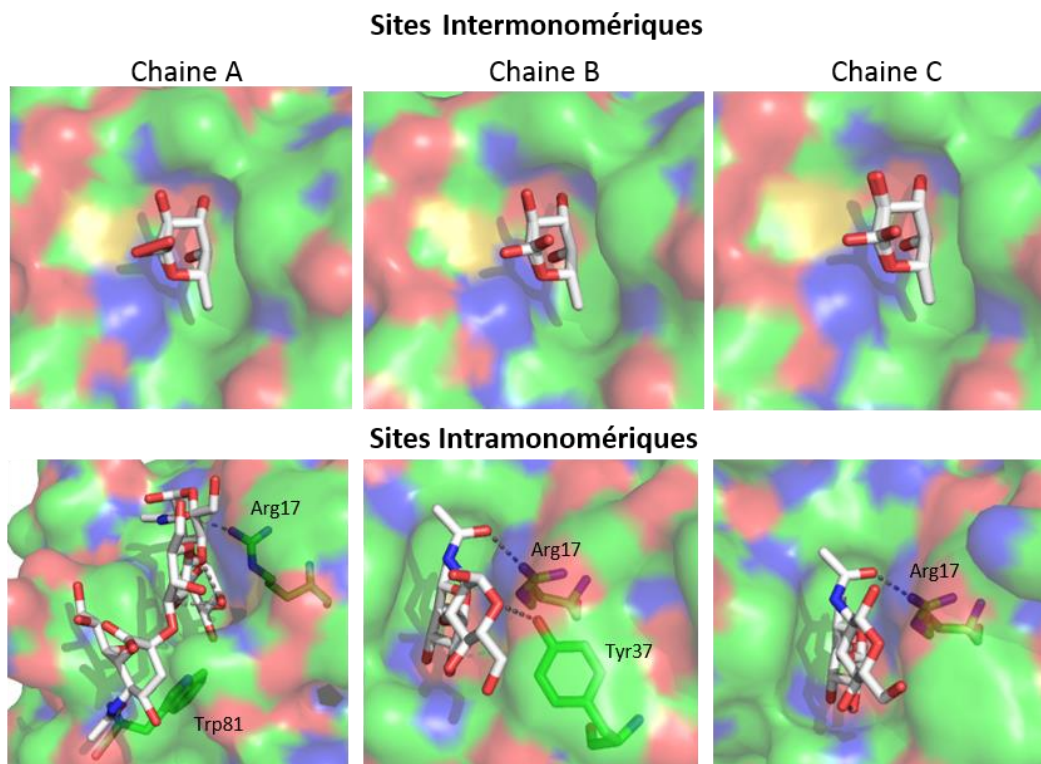


**Figure 54 : Sites intra et intermonomériques du complexe RSL/Le<sup>x</sup>**

*Ici ne sont représentés que les acides aminés qui interagissent par des liaisons hydrogène (pointillés noirs) avec l'oligosaccharide (excepté avec le fucose dans le site primaire).*

Dans la structure du complexe RSL/SLe<sup>x</sup>, les sites intramonomériques sont occupés par l'oligosaccharide alors que les sites intermonomériques sont occupés par du L-fucose. Le fait que l'oxygène anomérique de ce fucose soit observé en configuration alpha et beta indique bien qu'il s'agit d'un monosaccharide contaminant, et non pas d'un manque de densité électronique sur l'oligosaccharide entier. Dans les sites intramonomériques, la densité électronique permet de construire le disaccharide  $\alpha$ Fuc1-3GlcNAc dans deux sites et le tetrasaccharide SLe<sup>x</sup> dans le troisième site. Le fucose des oligosaccharides est toujours à la même place et interagit avec les mêmes acides aminés que lorsqu'il est libre. Le GlcNAc interagit toujours avec l'arginine 17 de la protéine et dans la chaîne B, l'oxygène O6 interagit aussi avec la tyrosine 37. Le galactose ne forme pas de liaison hydrogène avec la protéine. Dans la chaîne A, l'acide sialique a pu être construit dans la densité électronique et présente des interactions avec le tryptophane 81 (Figure 55).





**Figure 55 : Sites intra et intermonomériques du complexe RSL-SLe<sup>x</sup>**

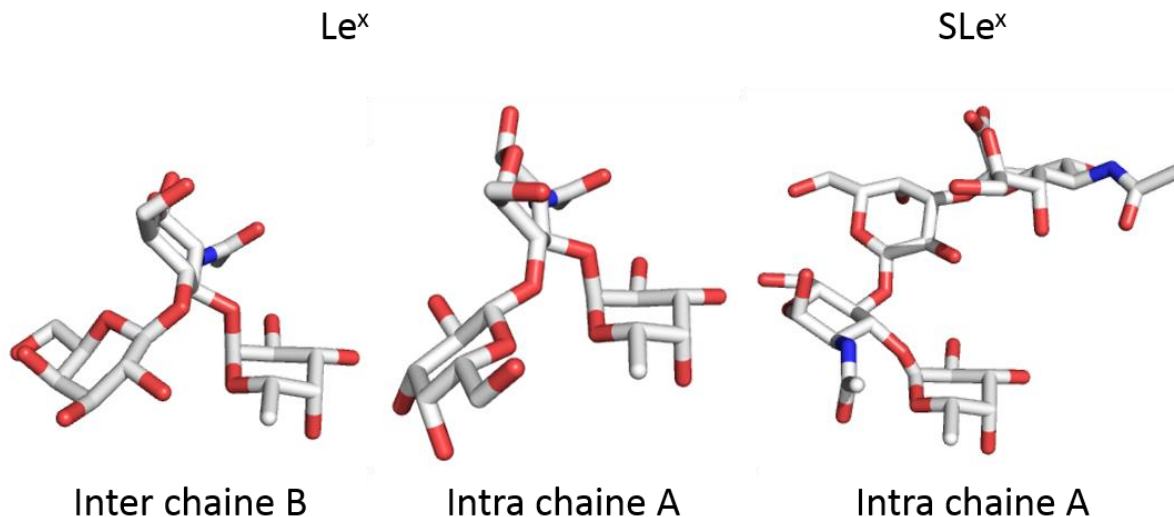
*Ici ne sont représentés que les acides aminés qui interagissent par des liaisons hydrogène (pointillés noirs) avec le sucre.*

Ces deux structures cristallographiques permettent d'observer 5 sites contenant un Le<sup>x</sup> ou SLe<sup>x</sup> (4 pour RSL/ Le<sup>x</sup> et 1 pour RSL /SLe<sup>x</sup>). Si on analyse les angles de torsion qui définissent les liaisons Fuc $\alpha$ 1-3GlcNAc et Gal $\beta$ 1-4GlcNAc dans les deux complexes, différentes conformations sont observées (Tableau 12). La conformation classique « fermée » du Le<sup>x</sup> n'est observée dans aucun de ces sites (Figure 56).

			Fuc $\alpha$ 1-3GlcNAc		Gal $\beta$ 1-4GlcNAc	
			Phi	Psi	Phi	Psi
<b>Le<sup>x</sup></b>	Site Intra	Chaîne A	-76	100	-140	16
		Chaîne B	-100	-76	/	/
		Chaîne C	-91	-66	/	/
	Site Inter	Chaîne A	-96	127	-26	72
		Chaîne B	-92	126	-34	77
		Chaîne C	-93	120	-33	79
<b>SLe<sup>x</sup></b>	Site Intra	Chaîne A	-120	-34	-58	23
		Chaîne B	-90	-70	/	/
		Chaîne C	-61	-101	/	/
	Site Inter	Chaîne A	/	/	/	/
		Chaîne B	/	/	/	/
		Chaîne C	/	/	/	/

**Tableau 12 : Angles de torsion PHI et PSI des Le<sup>x</sup> et SLe<sup>x</sup> retrouvés dans les sites de RSL**

Deux conformations différentes sont observées pour le Le<sup>x</sup> suivant s'il est dans le site intra (chaîne) ou dans le site inter (3 chaînes). La seule molécule observée de SLe<sup>x</sup> correspond à une troisième conformation, différente des deux précédentes.



**Figure 56 : Le<sup>x</sup> et SLe<sup>x</sup> présents dans les sites de fixation de RSL**

Sont représentées les différentes conformations du Le<sup>x</sup> et le sLe<sup>x</sup> présents dans les sites de fixation de RSL. Le fucose a été choisi comme référence et est toujours représenté avec la même orientation.

### II.2.2.3. Comparaison entre sites intra et inter de RSL

Les structures cristallographiques confirment la différence d'affinité entre sites intra et inter observées en ITC. En particulier, le site intermonomérique ne peut pas lier le SLe<sup>x</sup> puisqu'il est bloqué par un L-fucose contaminant. (Figure 55).

```

SSVQTAATSWGTVPS--IRVYTANNGKI-TERCWDGKGWYTGAFNEPG
PGDNVSVTSWLVGSAIHIRVYAST-GTTTTEWCWDGNGWTKGAYTAT-

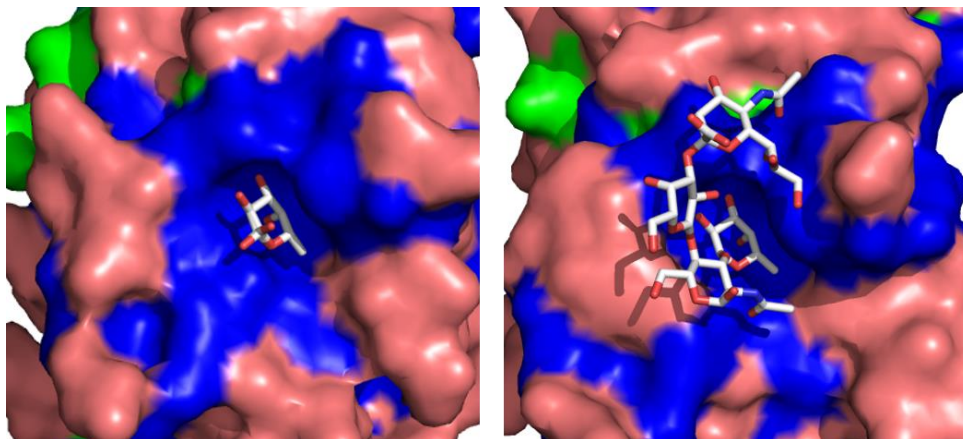
DNVSVTSWLVGSAIH-IRVY-ASTGTTTTEWCWDGNGWTKGAYTATN
-SSVQTAATSWGTVPSIRVYTANNGKI-TERCWDGKGWYTGAF---N

```

**Figure 57 : Alignement des sites de fixation intra et intermonomériques des RSL**

En haut est représentée la séquence du site intramonomérique alors qu'en bas est représentée la séquence du site intermonomérique. Les couleurs rouge, vert et bleu représentent respectivement les acides aminés différents, équivalents ou identiques entre les deux séquences. Tous les acides aminés ne sont pas forcément proches du site de fixation. Les acides aminés surlignés en rose sont ceux qui sont impliqués dans la stabilisation du fucose et ceux en bleu sont ceux impliqués dans la stabilisation du reste de l'oligosaccharide. L'acide aminé surligné de rose avec un contour bleu est à la fois impliqué dans la stabilisation du fucose et du reste de l'oligosaccharide.

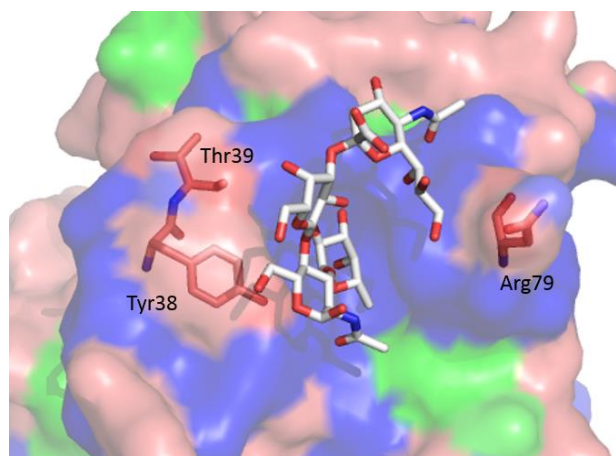
L'alignement de séquence des deux sites permet de mettre en évidence un grand nombre de différences dans les acides aminés (Figure 57). Par contre, une comparaison au niveau structural montre que les acides aminés formant la poche à fucose sont strictement identiques entre les deux sites (Figure 58). En revanche, pour les ligands oligosaccharidiques comme le SLe<sup>x</sup>, le sucre est en contact avec des acides aminés qui sont différents dans le site intramonomérique, ou le site intermonomérique (Figure 58).



**Figure 58 : Comparaison des sites de fixation intra et intermonomériques**

A gauche est représentée le site intermonomérique de la structure RSL/ SLe<sup>x</sup> et à droite le site intramonomérique de cette même structure. Les couleurs rouge, vert et bleu représentent respectivement les acides aminés différents, équivalents ou identiques entre les deux séquences.

Les différences les plus intéressantes sont la tyrosine 38, la thréonine 39 et l'arginine 79 du site intramonomérique qui sont remplacées respectivement par une thréonine et deux lysines dans le site intermonomérique (Figure 59). Ces acides aminés ne forment pas de liaisons hydrogène avec l'oligosaccharide mais pourraient dans le cas du site intermonomérique être défavorable à la liaison.

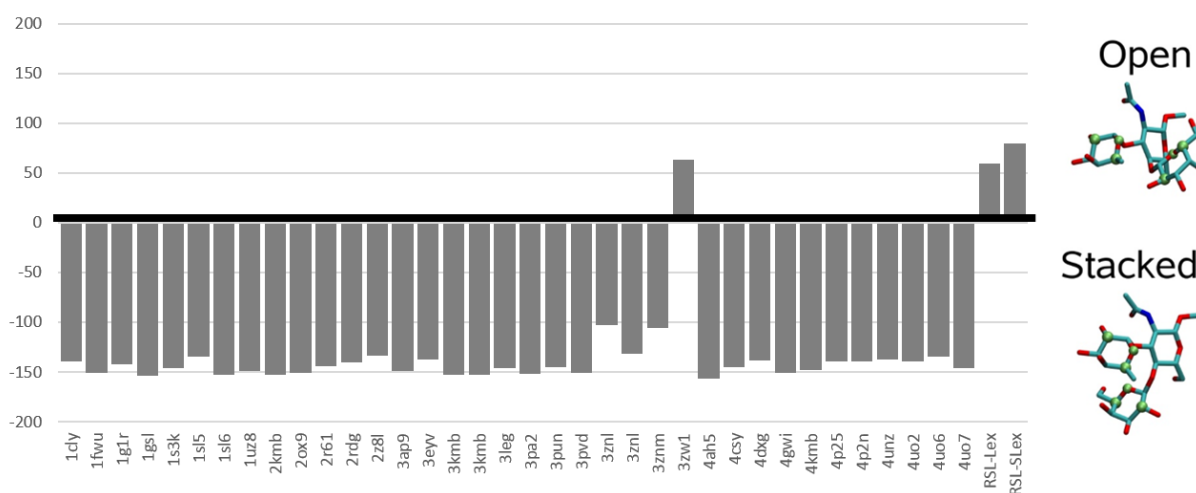


**Figure 59 : Acides aminés proches du sucre différents entre les sites intra et intermonomériques**

Le rouge, vert et bleu représentent respectivement les acides aminés différents, équivalents ou identiques entre les deux séquences. En sticks sont mis en évidence les acides aminés différents entre le site intra et intermonomériques.

#### II.2.2.4. Comparaison avec BambL et autres lectines à fucose

En solution les Lewis sont sous forme « Stacked » avec le galactose au-dessus du fucose. Dans la plupart des lectines, comme par exemple la sélectine P (Somers *et al.*, 2000), la lectine de *Griffonia simplicifolia* (Delbaere *et al.*, 1993), ou encore la lectine DC-SIGN (Guo *et al.*, 2004), les Le<sup>x</sup> sont sous forme fermée. On peut quantifier le « stacking » entre les cycles fucose et galactose en mesurant l'angle entre le plan du cycle fucose (défini par les atomes C1, C5 et O5) et la normale au plan du cycle galactose (défini par les atomes C2, C5 et O5). Pour deux cycles absolument parallèles, cet angle serait de +/-180°. L'analyse de toutes les structures de complexes entre lectines et Le<sup>x</sup> (ou glycanes contenant le fragment Le<sup>x</sup>) est représentée dans la Figure 60 (analyse effectuée par Jérémie Topin).

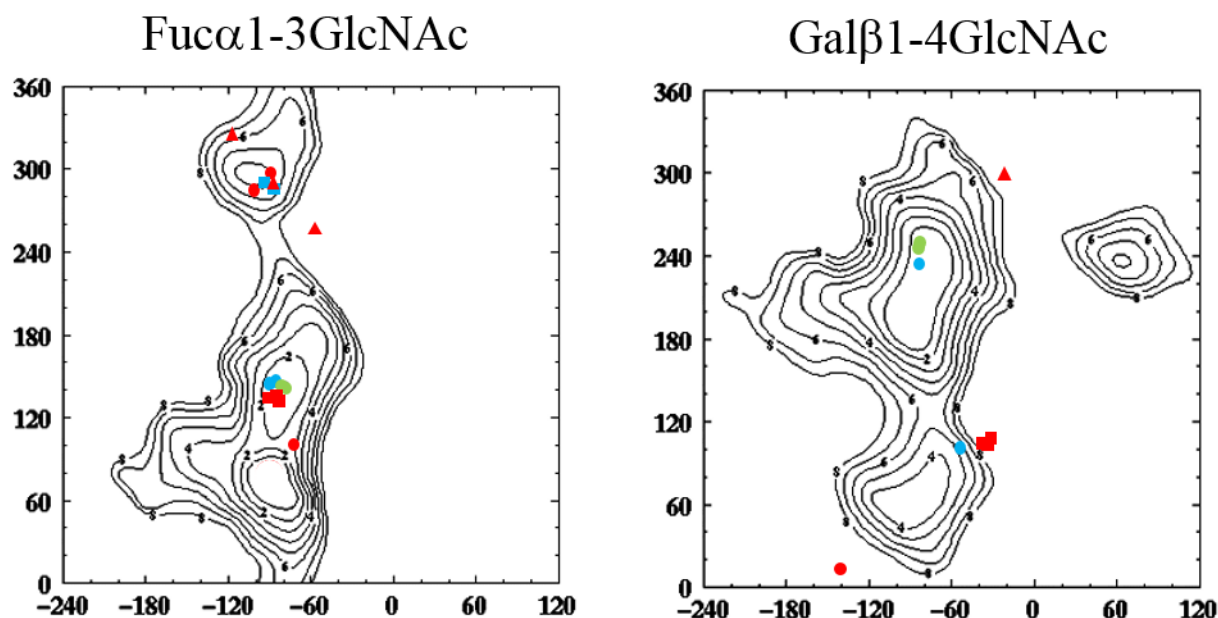


**Figure 60 : Conformation du Le<sup>x</sup> et SLe<sup>x</sup> en solution et en complexe avec des protéines**

La conformation des Lewis peut être soit sous forme « open » (angle entre les deux plans formés par le fucose et le galactose supérieurs à 0) soit sous forme « stacked » (angle entre les deux plans formés par le fucose et le galactose inférieurs à 0). Le graphique correspond à la conformation en fonction du site de fixation dans lequel le sucre est trouvé dans une protéine.

Tous les complexes montrent bien l'angle attendu de +/-180°, sauf BambL (Topin *et al.*, 2013), et RSL où les conformations de Le<sup>x</sup> correspondent à un angle entre les plans proche de +90°.

Les conformations observées dans les structures cristallographiques peuvent également être comparées aux cartes d'énergie précédemment calculées pour chacune des liaisons glycosidiques (Imberty *et al.*, 1995). Les liaisons glycosidiques du Le<sup>x</sup> à l'état cristallin (Perez *et al.*, 1996) sont localisées dans les deux régions de plus basse énergie. En revanche dans le cas du Le<sup>x</sup> présent dans les sites de fixation de RSL, on peut voir que les angles ne sont pas toujours dans des zones les plus favorables. En particulier, le minimum secondaire autour de la valeur de  $\varphi = -60^\circ$  pour la liaison Fuc $\alpha$ 1-3GlcNAc, est observé. De la même manière, plusieurs structures adoptent le minimum secondaire  $\varphi = +60^\circ$  pour la liaison Gal $\beta$ 1-4GlcNAc (Tableau 12 et Figure 61).



**Figure 61 : Carte d'énergie des Lewis dans les sites de fixation de BamBL et RSL**

Les graphiques représentent les isocontours énergétiques en fonction des valeurs des angles  $\varphi$  (PSI) et  $\psi$  (PHI). Les angles de torsion sont définis par  $\psi = \theta$  (C1-O1-C' $x$ -C' $x$ +1) et  $\varphi = \theta$  (O5-C1-O1-C $x$ ). En abscisse sont représentés les angles  $\psi$  (PHI) et en ordonné les angles  $\varphi$  (PSI). En bleu sont représentées les valeurs observées pour le Le<sup>x</sup> complexé avec BamBL, et en rouge avec RSL (ronds, carrés et triangles pour respectivement le Le<sup>x</sup> présent dans les sites intramanomériques, intermonomériques et pour le SLe<sup>x</sup>). Le point vert représente les angles du Le<sup>x</sup> cristallin (Perez *et al.*, 1996).

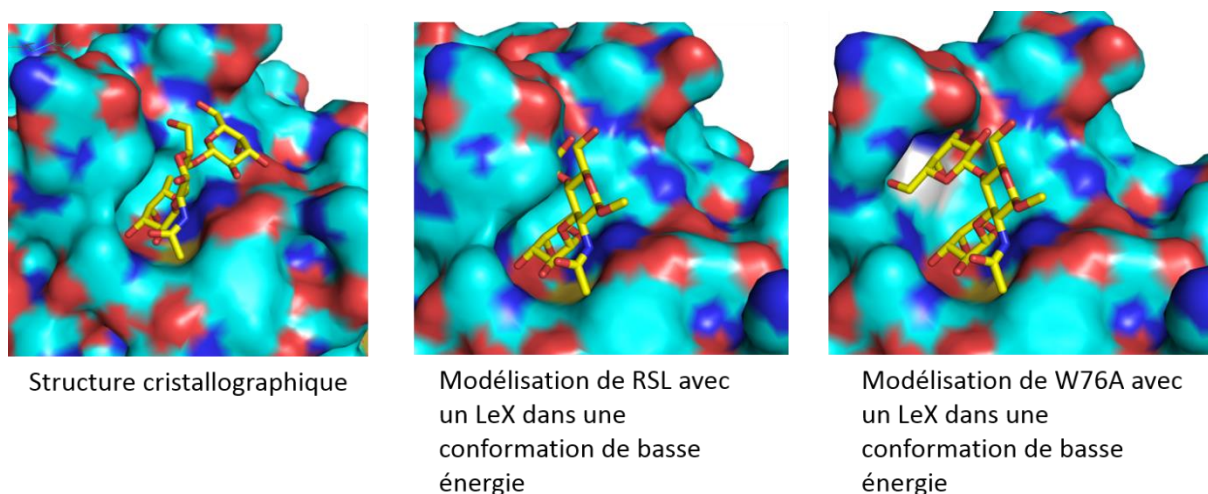


### II.2.3. Discussion

La structure cristallographique de RSL en présence de  $Le^x$  montre que le sucre est présent dans le site de fixation dans plusieurs conformations ouvertes, toutes différentes de la conformation fermée de basse énergie. Le passage de la conformation de base énergie « stacked » à une conformation ouverte peut être une des raisons de la faible affinité de RSL pour les Lewis.

Nous avons essayé de comprendre pourquoi le sucre ne reste pas en conformation de basse énergie dans le site fixation de RSL. Pour cela nous avons modélisé un  $Le^x$  en conformation fermée dans le site de fixation de RSL par bio-informatique. Ainsi nous avons pu mettre en évidence que le galactose présenterait alors un conflit stérique avec le tryptophane 76 pour le site intramonomérique et le tryptophane 31 pour le site intermonomérique (Figure 62). Dans toutes les lectines de cette famille, ce tryptophane est conservé et joue un rôle important dans l'interaction avec les glycanes par un stacking sur le fucose. Ce stacking est très proche de celui observé avec le galactose dans le  $Le^x$  en conformation fermée, et ces deux interactions ne peuvent pas avoir lieu en même temps.

Le travail actuel permet ainsi d'identifier les bases de la faible affinité de RSL pour les oligosaccharides de type Lewis. Une suite logique serait d'envisager les mutations sur ce tryptophane incompatible avec la conformation de basse énergie. De façon préliminaire, nous avons modélisé le mutant W76A. Comme représenté sur la Figure 62, ce mutant présente une poche de fixation plus large et aurait la possibilité de reconnaître le Lewis x dans sa conformation fermée.



**Figure 62 : Structure cristallographique et modélisation des interactions RSL/ $Le^x$**

Dans la suite du travail que j'ai commencé en thèse il serait donc intéressant de tester l'affinité du mutant W76A et d'autres mutants de ce tryptophane pour les épitopes Lewis. Il est ainsi possible que le sucre puisse rentrer dans le site de fixation sous forme fermé et ainsi voir une augmentation de l'affinité. Le double mutant W31A\_W76A de RSL est en cours d'étude. D'autres mutants à partir de la protéine néoRSL ont également été clonés et sont en cours de production. Les protéines néoRSL\_III<sub>135</sub>\_WA et néoRSL\_I<sub>246</sub>\_WA a été construite. Il s'agit du mutant néoRSL\_III (Arnaud *et al.*, 2014) pour lequel les tryptophanes des trois sites actifs ont été substitués par une alanine. Dans le cas de néoRSL\_I<sub>246</sub>\_WA il s'agit des trois sites intra (les autres étant inactifs), alors que dans le cas de néoRSL\_I<sub>135</sub>\_WA il s'agit des sites inter. Des difficultés sont rencontrées à la fois pour l'obtention de protéines solubles et pour la stabilité des protéines obtenues, et des études sont en cours pour ajouter des « tags » aidant à la production de ces mutants.

## II.2.4. Matériels et Méthodes

### II.2.4.1. Purification des protéines

La production et la purification de RSL a été réalisée comme précédemment décrite (Kostlanova *et al.*, 2005) L'obtention des néoRSL a été réalisée comme décrit précédemment (Arnaud *et al.*, 2014) .

### II.2.4.2. ITC

Les protéines RSL et néoRSL\_VI en solution dans du tampon 20 mM Tris/HCl, pH 7.5, 150 mM NaCl sont amenées à la concentration désirée (Tableau 13).

Les concentrations en protéines et en oligosaccharide utilisées pour les expériences sur néoRSL\_III sont de respectivement 0.03 mM et 1.2 mM.

Les données ont été obtenues avec l'ITC200 (Microcal-GE Healthcare). L'expérience est composée de 20 injections de sucres dans une cellule contenant la protéine. La première injection est d'un volume de 0.4  $\mu$ L, puis les suivantes sont de 2  $\mu$ L. Les ITC ont été réalisées à une température de 25 °C.

	RSL		NéoRSL_VI	
	Concentration mM		Concentration mM	
	Protéines	Sucres	Protéines	Sucres
<b>Le<sup>Y</sup> (penta)</b>	0,045	1	0,015	1
<b>H type 1 (tetra)</b>	0,045	1	0,015	1
<b>H type 2 (tetra)</b>	0,045	1	0,015	1
<b>H type 4 (tetra)</b>	0,023	5	0,073	5
<b>Le<sup>a</sup> (tetra)</b>	0,14	3	0,05	3
<b>Le<sup>x</sup> (tetra)</b>	0,23	5	0,073	5
<b>SLe<sup>x</sup> (penta)</b>	0,043	3,5	0,073	5

**Tableau 13 : Concentrations en protéines et en sucres utilisées lors des ITC avec RSL et néoRSL\_VI**

### II.2.4.3. Structure cristallographique

La protéine RSL lyophilisée est dissoute dans du tampon 20 mM Tris/HCl, pH 7.5, 150 mM NaCl à 10 mg/ml et incubée avec 5 mM d'oligosaccharides. Les cristaux de RSL complexée avec du Le<sup>x</sup> ou du SLe<sup>x</sup> ont été obtenus en utilisant la méthode de la goutte pendante avec une goutte de 2 µL contenant 1 µL de solution protéique à 10 mg/ml contenant 5 mM d'oligosaccharides et 1 µL de solution de précipitation. Les cristaux du complexe RSL-Le<sup>x</sup> ont été obtenus dans des conditions suivantes : 0.1 M Tris pH 8.5, 27 % (p/v) PEG 6000. Les cristaux ont ensuite été pêchés puis plongés dans une solution cryoprotectante de 30 % (p/v) PEG 6000. Les cristaux du complexe RSL-SLe<sup>x</sup> ont été obtenus dans des conditions suivantes : 0.28 M Tris pH 8.5, 30 % (p/v) PEG 6000. Les cristaux ont ensuite été pêchés et congelés dans de l'azote liquide. Les données de diffractions ont été obtenues à l'ESRF de Grenoble aux stations ID14-EH4 ou BM30A en utilisant le détecteur ADSC Q4 CCD. Les données ont été processées avec MOSFLM et mises à l'échelle en utilisant SCALA. Le remplacement moléculaire a été réalisé avec PHASER et avec les coordonnées de RSL avec le code PDB 2BT9 (Kostlanova *et al.*, 2005). La structure a été affinée en utilisant REFMAC et COOT.









---

## **DISCUSSION et PERSPECTIVES**

---



---

## DISCUSSION et PERSPECTIVES

---

Mes travaux de thèse s'inscrivent à l'interface entre la biologie de synthèse et la biotechnologie. Ils m'ont permis d'aborder la biochimie, la biophysique, la biologie cellulaire, la biologie moléculaire et la cristallographie.

La première partie de mes travaux avait pour but de mieux comprendre le mécanisme d'invagination des membranes cellulaires induit par les lectines. Pour cela, nous avons créé des mutants de la protéine RSL afin de voir l'implication de la valence dans ce phénomène. Nous avons pu mettre en évidence que le passage d'une valence de 6 à une valence de 3 ne modifiait pas l'avidité pour une surface fucosylée mais induisait une perte de la capacité à former des tubules dans les membranes. Limité par le nombre de valences et topologies envisageables à partir de la protéine sauvage, j'ai créé une protéine RSL monomérique. A partir de ce mutant j'ai produit des protéines de valences comprises entre 0 et 6 avec plusieurs topologies pour les valences 4, 3 et 2. Ces mutants nous ont permis de caractériser plus finement la formation d'invagination et de mettre en évidence que le nombre de sites de fixation n'est pas le facteur clé pour permettre la formation de tubules. En revanche, la distance entre deux sites de fixation est la caractéristique essentielle. En effet, dans le cas de mutants de valences de 2, si les sites de fixations sont séparés de plus 27 Å; il n'y a pas formation de tubules alors que lors que les sites ne sont plus que séparés de 15 Å, il y a formation d'invagination.

Maintenant que nous avons mis en évidence l'importance de la distance entre les sites de fixations dans la formation d'invagination, je pense qu'il serait intéressant d'étudier un autre paramètre qui est l'affinité. Il est possible que la distance entre les sites de fixation nécessaire pour former une invagination soit également liée à l'affinité de ces sites pour l'oligosaccharide. Et donc que plus l'affinité est grande, plus la distance entre les sites peut être importante.

Une connaissance encore plus approfondie de ce mécanisme pourrait permettre la création d'un outil de vectorisation extrêmement puissant. En effet, à l'heure actuelle, un bon outil de vectorisation serait une grande aide pour permettre la création de nouveau traitement qui serait directement administré dans les cellules cibles. Les domaines impliqués sont très

variés, comme la cancérologie, mais aussi la thérapie génique, sans oublier le domaine de l'infectiologie.

Dans une seconde partie de ma thèse, j'ai étudié la spécificité de RSL en premier lieu par une étude comparative *in vitro/in silico* de différentes lectines à fucose. Cette étude a permis de mettre au point un protocole bio-informatique de prédiction de l'affinité des lectines pour divers oligosaccharides.

Ensuite, nous avons étudié la spécificité de RSL et comparé les sites de fixation intra et intermonomériques. Nous avons mis en évidence que les sites intra et intermonomériques sont très proches structurellement mais possèdent des différences suffisantes pour présenter une affinité différente pour le SLe<sup>x</sup>. En effet, le site intramonomérique a une plus forte affinité pour le SLe<sup>x</sup> que le site intermonomérique. Nous avons également mis en évidence que les Lewis présents dans les sites de fixation ne peuvent pas adopter la conformation classique de basse énergie mais sont contraints dans des conformations non habituelles dites ouvertes. A l'heure actuelle la modélisation est en cours pour déterminer comment se produit l'entrée du sucre en conformation ouverte dans les sites de fixation. Deux hypothèses sont envisagées, soit la protéine sélectionne les conformations du Lewis ouvertes, très rare en solution, soit la protéine induit le changement de conformation.

Afin de modifier l'affinité de RSL pour les Lewis, il serait intéressant de créer des mutants qui permettraient l'entrée du Lewis dans le site sous sa conformation de basse énergie. Certains mutants où le tryptophane en cause est muté en alanine sont en cours de production, mais ils ne sont pas stables et sont majoritairement produits sous forme de corps d'inclusions. Il serait également envisageable d'utiliser une technique de mutagenèse aléatoire suivi d'une sélection à haut débit comme le phage display afin de tester des milliers de mutants. J'ai commencé à faire des constructions dans ce sens dans des phagemides. Il est possible d'utiliser directement la protéine sauvage RSL mais des problèmes peuvent être multiples : soit la protéine se trimérise et il sera difficile d'analyser une protéine avec des sites différents, soit la protéine ne s'oligomériser pas ce qui risque de lui enlever sa fonctionnalité. Il est également possible d'utiliser une néolectine néoRSL-I, mais de nouveau des problèmes sont à envisager : la protéine est d'une taille plus importante ce qui peut limiter son utilisation en phage display, de plus elle est très peu stable donc elle risque d'être difficilement produite sous forme soluble à la surface d'un phage.

La majeure partie des perspectives concerne la création d'un outil de vectorisation spécifique permettant l'internalisation d'objet de taille et de nature variée dans différents

types cellulaires. Cette partie est confidentielle, et ne sera pas développée ici puisque je souhaite continuer ces recherches par la création d'une startup.





---

## **BIBLIOGRAPHIE**

---



---

# BIBLIOGRAPHIE

---

(1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-763.

## A

Adam, J., Pokorna, M., Sabin, C., Mitchell, E.P., Imberty, A., and Wimmerova, M. (2007). "Engineering of PA-IIL lectin from *Pseudomonas aeruginosa* - Unravelling the role of the specificity loop for sugar preference." BMC Struct Biol **7**: 36.

Adhikari, P., Bachhawat-Sikder, K., Thomas, C.J., Ravishankar, R., Jeyaprakash, A.A., Sharma, V., Vijayan, M., and Surolia, A. (2001). "Mutational analysis at Asn-41 in peanut agglutinin. A residue critical for the binding of the tumor-associated Thomsen-Friedenreich antigen." J Biol Chem **276**(44): 40734-40739.

Air, G.M. (2014). "Influenza virus-glycan interactions." Curr Opin Virol **7C**: 128-133.

Akahori, T., Yuzawa, Y., Nishikawa, K., Tamatani, T., Kannagi, R., Miyasaka, M., Okada, H., Hotta, N., and Matsuo, S. (1997). "Role of a sialyl Lewis(x)-like epitope selectively expressed on vascular endothelial cells in local skin inflammation of the rat." J Immunol **158**(11): 5384-5392.

Amano, K., Fujihashi, M., Ando, A., Miki, K., and 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**(4): 841-847.

Aoyagi, Y., Suzuki, Y., Igarashi, K., Saitoh, A., Oguro, M., Yokota, T., Mori, S., Nomoto, M., Isemura, M., and Asakura, H. (1991). "The usefulness of simultaneous determinations of glucosamylation and fucosylation indices of alpha-fetoprotein in the differential diagnosis of neoplastic diseases of the liver." Cancer **67**(9): 2390-2394.

Arango, R., Rodriguez-Arango, E., Adar, R., Belenky, D., Loontjens, F.G., Rozenblatt, S., and Sharon, N. (1993). "Modification by site-directed mutagenesis of the specificity of

- Erythrina corallodendron lectin for galactose derivatives with bulky substituents at C-2." FEBS Lett **330**(2): 133-136.
- Arnaud, J., Claudinon, J., Trondle, K., Trovaslet, M., Larson, G., Thomas, A., Varrot, A., Romer, W., Imberty, A., and Audfray, A. (2013). "Reduction of lectin valency drastically changes glycolipid dynamics in membranes but not surface avidity." ACS Chem Biol **8**(9): 1918-1924.
- Arnaud, J., Trondle, K., Claudinon, J., Audfray, A., Varrot, A., Romer, W., and Imberty, A. (2014). "Membrane deformation by neolectins with engineered glycolipid binding sites." Angew Chem Int Ed Engl **53**(35): 9267-9270.
- Asazawa, H., Kamada, Y., Takeda, Y., Takamatsu, S., Shinzaki, S., Kim, Y., Nezu, R., Kuzushita, N., Mita, E., Kato, M., and Miyoshi, E. (2014). "Serum fucosylated haptoglobin in chronic liver diseases as a potential biomarker of hepatocellular carcinoma development." Clin Chem Lab Med.
- Ashton, A.W., Boehm, M.K., Gallimore, J.R., Pepys, M.B., and Perkins, S.J. (1997). "Pentameric and decameric structures in solution of serum amyloid P component by X-ray and neutron scattering and molecular modelling analyses." J Mol Biol **272**(3): 408-422.
- Audfray, A., Claudinon, J., Abounit, S., Ruvoen-Clouet, N., Larson, G., Smith, D.F., Wimmerova, M., Le Pendu, J., Romer, W., Varrot, A., and Imberty, A. (2012). "Fucose-binding lectin from opportunistic pathogen Burkholderia ambifaria binds to both plant and human oligosaccharidic epitopes." J Biol Chem **287**(6): 4335-4347.
- Avichezer, D., and Arnon, R. (1996). "Differential reactivities of the Arachis hypogaea (peanut) and Vicia villosa B4 lectins with human ovarian carcinoma cells, grown either in vitro or in vivo xenograft model." FEBS Lett **395**(2-3): 103-108.

## B

- Baker, S.C., Saunders, N.F., Willis, A.C., Ferguson, S.J., Hajdu, J., and Fulop, V. (1997). "Cytochrome cd1 structure: unusual haem environments in a nitrite reductase and analysis of factors contributing to beta-propeller folds." J Mol Biol **269**(3): 440-455.
- Baumann, H., Nudelman, E., Watanabe, K., and Hakomori, S. (1979). "Neutral fucolipids and fucogangliosides of rat hepatoma HTC and H35 cells, rat liver, and hepatocytes." Cancer Res **39**(7 Pt 1): 2637-2643.

- 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**(9): 2313-2322.
- Beppu, M., Terao, T., and Osawa, T. (1976). "Preparation of monovalent succinyl-concanavalin A and its mitogenic activity." J Biochem **79**(5): 1113-1117.
- Bergstrom, M., Astrom, E., Pahlsson, P., and Ohlson, S. (2012). "Elucidating the selectivity of recombinant forms of *Aleuria aurantia* lectin using weak affinity chromatography." J Chromatogr B Analyt Technol Biomed Life Sci **885-886**: 66-72.
- Berois, N., Mazal, D., Ubillos, L., Trajtenberg, F., Nicolas, A., Sastre-Garau, X., Magdelenat, H., and Osinaga, E. (2006). "UDP-N-acetyl-D-galactosamine: polypeptide N-acetylgalactosaminyltransferase-6 as a new immunohistochemical breast cancer marker." J Histochem Cytochem **54**(3): 317-328.
- Berstein, F. (1924). "Ergebnisse einer biostatistischen zusammenfassenden Betrachtung über die erblichen Blutstrukturen des Menschen." Klin Wschr **3**: 1495-1497.
- Beuth, J., Ko, H.L., Schneider, H., Tawadros, S., Kasper, H.U., Zimst, H., and Schierholz, J.M. (2006). "Intratumoral application of standardized mistletoe extracts down regulates tumor weight via decreased cell proliferation, increased apoptosis and necrosis in a murine model." Anticancer Res **26**(6B): 4451-4456.
- Birkle, S., Zeng, G., Gao, L., Yu, R.K., and Aubry, J. (2003). "Role of tumor-associated gangliosides in cancer progression." Biochimie **85**(3-4): 455-463.
- Blaber, M., and Lee, J. (2012). "Designing proteins from simple motifs: opportunities in Top-Down Symmetric Deconstruction." Curr Opin Struct Biol **22**(4): 442-450.
- Blander, J.M., Visintin, I., Janeway, C.A., Jr., and Medzhitov, R. (1999). "Alpha(1,3)-fucosyltransferase VII and alpha(2,3)-sialyltransferase IV are up-regulated in activated CD4 T cells and maintained after their differentiation into Th1 and migration into inflammatory sites." J Immunol **163**(7): 3746-3752.
- Borzym-Kluczyk, M., Radziejewska, I., and Cechowska-Pasko, M. (2014). "Increased expression of MUC1 and sialyl Lewis antigens in different areas of clear renal cell carcinoma." Clin Exp Nephrol.

Bourne, Y., and Cambillau, C., Eds. (1993). The role of structural water molecules in protein-saccharide complexes. In *Water and biological macromolecules: Topics in Molecular and Structural Biology*. . Oxford.

Brockhausen, I., Yang, J., Dickinson, N., Ogata, S., and Itzkowitz, S.H. (1998). "Enzymatic basis for sialyl-Tn expression in human colon cancer cells." Glycoconj J **15**(6): 595-603.

Brooks, S.A., Carter, T.M., Bennett, E.P., Clausen, H., and Mandel, U. (2007). "Immunolocalisation of members of the polypeptide N-acetylgalactosaminyl transferase (ppGalNAc-T) family is consistent with biologically relevant altered cell surface glycosylation in breast cancer." Acta Histochem **109**(4): 273-284.

## C

Candelier, J.J., Mollicone, R., Mennesson, B., Bergemer, A.M., Henry, S., Coullin, P., and Oriol, R. (1993). "Alpha-3-fucosyltransferases and their glycoconjugate antigen products in the developing human kidney." Lab Invest **69**(4): 449-459.

Candelier, J.J., Mollicone, R., Mennesson, B., Coullin, P., and Oriol, R. (2000). "Expression of fucosyltransferases in skin, conjunctiva, and cornea during human development." Histochem Cell Biol **114**(2): 113-124.

Casals, C. (2001). "Role of surfactant protein A (SP-A)/lipid interactions for SP-A functions in the lung." Pediatr Pathol Mol Med **20**(4): 249-268.

Chang, C.P., Yang, M.C., Liu, H.S., Lin, Y.S., and Lei, H.Y. (2007). "Concanavalin A induces autophagy in hepatoma cells and has a therapeutic effect in a murine in situ hepatoma model." Hepatology **45**(2): 286-296.

Chase, S.D., Magnani, J.L., and Simon, S.I. (2012). "E-selectin ligands as mechanosensitive receptors on neutrophils in health and disease." Ann Biomed Eng **40**(4): 849-859.

Chia, D., Terasaki, P.I., Suyama, N., Galton, J., Hirota, M., and Katz, D. (1985). "Use of monoclonal antibodies to sialylated Lewisx and sialylated Lewisy for serological tests of cancer." Cancer Res **45**(1): 435-437.

Cioci, G., Mitchell, E.P., Chazalet, V., Debray, H., Oscarson, S., Lahmann, M., Gautier, C., Breton, C., Perez, S., and Imberty, A. (2006). "Beta-propeller crystal structure of *Psathyrella velutina* lectin: an integrin-like fungal protein interacting with monosaccharides and calcium." J Mol Biol **357**(5): 1575-1591.



- Cioci, G., Mitchell, E.P., Gautier, C., Wimmerova, M., Sudakevitz, D., Perez, S., Gilboa-Garber, N., and Imberty, A. (2003). "Structural basis of calcium and galactose recognition by the lectin PA-IL of *Pseudomonas aeruginosa*." FEBS Lett **555**(2): 297-301.
- Clarke, J.L., and Watkins, W. (1996). "Alpha1,3-L-fucosyltransferase expression in developing human myeloid cells. Antigenic, enzymatic, and mRNA analyses." J Biol Chem **271**(17): 10317-10328.
- Cooper, M.A. (2002). "Optical biosensors in drug discovery." Nat Rev Drug Discov **1**(7): 515-528.
- Crennell, S.J., Garman, E.F., Laver, W.G., Vimr, E.R., and Taylor, G.L. (1993). "Crystal structure of a bacterial sialidase (from *Salmonella typhimurium* LT2) shows the same fold as an influenza virus neuraminidase." Proc Natl Acad Sci U S A **90**(21): 9852-9856.

## D

- Dahms, N., and Hancock, M.K. (2002). "P-type lectins." Biochim Biophys Acta **1572**(2-3): 317-340.
- Delbaere, L.T., Vandonselaar, M., Prasad, L., Quail, J.W., Wilson, K.S., and Dauter, Z. (1993). "Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution." J Mol Biol **230**(3): 950-965.
- Dennis, J.W., Laferte, S., Waghorne, C., Breitman, M.L., and Kerbel, R.S. (1987). "Beta 1-6 branching of Asn-linked oligosaccharides is directly associated with metastasis." Science **236**(4801): 582-585.
- Desilets, D.J., Davis, K.E., Nair, P.P., Salata, K.F., Maydonovitch, C.L., Howard, R.S., Kikendall, J.W., and Wong, R.K. (1999). "Lectin binding to human colonocytes is predictive of colonic neoplasia." Am J Gastroenterol **94**(3): 744-750.
- Diggle, S.P., Stacey, R.E., Dodd, C., Camara, M., Williams, P., and Winzer, K. (2006). "The galactophilic lectin, LecA, contributes to biofilm development in *Pseudomonas aeruginosa*." Environ Microbiol **8**(6): 1095-1104.
- Dodson, K.W., Pinkner, J.S., Rose, T., Magnusson, G., Hultgren, S.J., and Waksman, G. (2001). "Structural basis of the interaction of the pyelonephritic *E. coli* adhesin to its human kidney receptor." Cell **105**(6): 733-743.

- Du, J., Lu, W.L., Ying, X., Liu, Y., Du, P., Tian, W., Men, Y., Guo, J., Zhang, Y., Li, R.J., *et al.* (2009). "Dual-targeting topotecan liposomes modified with tamoxifen and wheat germ agglutinin significantly improve drug transport across the blood-brain barrier and survival of brain tumor-bearing animals." Mol Pharm **6**(3): 905-917.
- Dube, D.H., and Bertozzi, C.R. (2005). "Glycans in cancer and inflammation--potential for therapeutics and diagnostics." Nat Rev Drug Discov **4**(6): 477-488.
- Duong Van Huyen, J.P., Delignat, S., Bayry, J., Kazatchkine, M.D., Bruneval, P., Nicoletti, A., and Kaveri, S.V. (2006). "Interleukin-12 is associated with the in vivo anti-tumor effect of mistletoe extracts in B16 mouse melanoma." Cancer Lett **243**(1): 32-37.
- Dupeux, F., Rower, M., Seroul, G., Blot, D., and Marquez, J.A. (2011). "A thermal stability assay can help to estimate the crystallization likelihood of biological samples." Acta Crystallogr D Biol Crystallogr **67**(Pt 11): 915-919.

## E

- Ehrlich, P. (1891a). "Experimentalle untersuchungen uber immunitat I. Ueber ricin." Dtsch. Med. Wochenschr. **17**: 1218–1219
- Ehrlich, P. (1891b). "Experimentalle untersuchungen uber immunitat I. Ueber ricin. ." Dtsch. Med. Wochenschr. **17**: 976–979.
- Eierhoff, T., Bastian, B., Thuenauer, R., Madl, J., Audfray, A., Aigal, S., Juillot, S., Rydell, G.E., Muller, S., de Bentzmann, S., *et al.* (2014). "A lipid zipper triggers bacterial invasion." Proc Natl Acad Sci U S A **111**(35): 12895-12900.
- Elluru, S.R., Duong Van Huyen, J.P., Delignat, S., Prost, F., Heudes, D., Kazatchkine, M.D., Friboulet, A., and Kaveri, S.V. (2009). "Antiangiogenic properties of viscum album extracts are associated with endothelial cytotoxicity." Anticancer Res **29**(8): 2945-2950.
- Elsasser-Beile, U., Ruhnau, T., Freudenberg, N., Wetterauer, U., and Mengs, U. (2001). "Antitumoral effect of recombinant mistletoe lectin on chemically induced urinary bladder carcinogenesis in a rat model." Cancer **91**(5): 998-1004.
- Emoto, M., and Emoto, Y. (2009). "Intracellular bacterial infection and invariant NKT cells." Yonsei Med J **50**(1): 12-21.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). "Features and development of Coot." Acta Crystallogr D Biol Crystallogr **66**(Pt 4): 486-501.

Ewers, H., Romer, W., Smith, A.E., Bacia, K., Dmitrieff, S., Chai, W., Mancini, R., Kartenbeck, J., Chambon, V., Berland, L., *et al.* (2010). "GM1 structure determines SV40-induced membrane invagination and infection." Nat Cell Biol **12**(1): 11-18; sup pp 11-12.

## F

Faber, H.R., Groom, C.R., Baker, H.M., Morgan, W.T., Smith, A., and Baker, E.N. (1995). "1.8 Å crystal structure of the C-terminal domain of rabbit serum haemopexin." Structure **3**(6): 551-559.

Freeze, H.H. (2006). "Genetic defects in the human glycome." Nat Rev Genet **7**(7): 537-551.

Fujihashi, M., Peapus, D.H., Kamiya, N., Nagata, Y., and Miki, K. (2003). "Crystal structure of fucose-specific lectin from *Aleuria aurantia* binding ligands at three of its five sugar recognition sites." Biochemistry **42**(38): 11093-11099.

Fujita, K., Shimomura, M., Uemura, M., Nakata, W., Sato, M., Nagahara, A., Nakai, Y., Takamatsu, S., Miyoshi, E., and Nonomura, N. (2014). "Serum fucosylated haptoglobin as a novel prognostic biomarker predicting high-Gleason prostate cancer." Prostate **74**(10): 1052-1058.

Fujita, T., Matsushita, M., and Endo, Y. (2004). "The lectin-complement pathway--its role in innate immunity and evolution." Immunol Rev **198**: 185-202.

Fulop, V., Bocskei, Z., and Polgar, L. (1998). "Prolyl oligopeptidase: an unusual beta-propeller domain regulates proteolysis." Cell **94**(2): 161-170.

Fulop, V., and Jones, D.T. (1999). "Beta propellers: structural rigidity and functional diversity." Curr Opin Struct Biol **9**(6): 715-721.

Fulop, V., Moir, J.W., Ferguson, S.J., and Hajdu, J. (1995). "The anatomy of a bifunctional enzyme: structural basis for reduction of oxygen to water and synthesis of nitric oxide by cytochrome cd1." Cell **81**(3): 369-377.

## G

- Garibay-Cerdenares, O.L., Hernandez-Ramirez, V.I., Osorio-Trujillo, J.C., Hernandez-Ortiz, M., Gallardo-Rincon, D., Cantu de Leon, D., Encarnacion-Guevara, S., Villegas-Pineda, J.C., and Talamas-Rohana, P. (2014). "Proteomic identification of fucosylated haptoglobin alpha isoforms in ascitic fluids and its localization in ovarian carcinoma tissues from Mexican patients." J Ovarian Res **7**(1): 27.
- Garlatti, V., Belloy, N., Martin, L., Lacroix, M., Matsushita, M., Endo, Y., Fujita, T., Fontecilla-Camps, J.C., Arlaud, G.J., Thielens, N.M., and Gaboriaud, C. (2007). "Structural insights into the innate immune recognition specificities of L- and H-ficolins." EMBO J **26**(2): 623-633.
- Genbacev, O.D., Prakobphol, A., Foulk, R.A., Krtolica, A.R., Ilic, D., Singer, M.S., Yang, Z.Q., Kiessling, L.L., Rosen, S.D., and Fisher, S.J. (2003). "Trophoblast L-selectin-mediated adhesion at the maternal-fetal interface." Science **299**(5605): 405-408.
- Gohlke, H., and Klebe, G. (2001). "Statistical potentials and scoring functions applied to protein-ligand binding." Curr Opin Struct Biol **11**(2): 231-235.
- Goochee, C.F., and Monica, T. (1990). "Environmental effects on protein glycosylation." Biotechnology (N Y) **8**(5): 421-427.
- Gornik, O., and Lauc, G. (2008). "Glycosylation of serum proteins in inflammatory diseases." Dis Markers **25**(4-5): 267-278.
- Gottschalk, A., and Perry, B.T. (1951). "Ovomucin: a substrate for the enzyme of the influenza virus. III. Enzymic activity as an integral function of the influenza virus particle." Br J Exp Pathol **32**(5): 408-413.
- Gourdine, J.P., Cioci, G., Miguet, L., Unverzagt, C., Silva, D.V., Varrot, A., Gautier, C., Smith-Ravin, E.J., and Imberty, A. (2008). "High affinity interaction between a bivalve C-type lectin and a biantennary complex-type N-glycan revealed by crystallography and microcalorimetry." J Biol Chem **283**(44): 30112-30120.
- Groux-Degroote, S., Wavelet, C., Krzewinski-Recchi, M.A., Portier, L., Mortuaire, M., Mihalache, A., Trinchera, M., Delannoy, P., Malagolini, N., Chiricolo, M., *et al.* (2014). "B4GALNT2 gene expression controls the biosynthesis of Sd(a) and sialyl Lewis X antigens in healthy and cancer human gastrointestinal tract." Int J Biochem Cell Biol **53**: 442-449.

Guillot, J., Guerry, M., Kanska, G., Caldefie-Chezet, F., De Latour, M., and Penault-Llorca, F. (2004). "[Modification of glycoconjugates during the carcinogenesis: the case of mammary carcinomas]." Bull Cancer **91**(2): 141-158.

Gunther, G.R., Wang, J.L., Yahara, I., Cunningham, B.A., and Edelman, G.M. (1973). "Concanavalin A derivatives with altered biological activities." Proc Natl Acad Sci U S A **70**(4): 1012-1016.

Guo, Y., Feinberg, H., Conroy, E., Mitchell, D.A., Alvarez, R., Blixt, O., Taylor, M.E., Weis, W.I., and Drickamer, K. (2004). "Structural basis for distinct ligand-binding and targeting properties of the receptors DC-SIGN and DC-SIGNR." Nat Struct Mol Biol **11**(7): 591-598.

## H

Hakomori, S. (1999). "Antigen structure and genetic basis of histo-blood groups A, B and O: their changes associated with human cancer." Biochim Biophys Acta **1473**(1): 247-266.

Hakomori, S. (2000). "Traveling for the glycosphingolipid path." Glycoconj J **17**(7-9): 627-647.

Hakomori, S., and Kannagi, R. (1983). "Glycosphingolipids as tumor-associated and differentiation markers." J Natl Cancer Inst **71**(2): 231-251.

Hakomori, S.I. (2008). "Structure and function of glycosphingolipids and sphingolipids: recollections and future trends." Biochim Biophys Acta **1780**(3): 325-346.

Hallgren, P., Lundblad, A., and Svensson, S. (1975). "A new type of carbohydrate-protein linkage in a glycopeptide from normal human urine." J Biol Chem **250**(14): 5312-5314.

Hallouin, F., Goupille, C., Bureau, V., Meflah, K., and Le Pendu, J. (1999). "Increased tumorigenicity of rat colon carcinoma cells after alpha1,2-fucosyltransferase FTA anti-sense cDNA transfection." Int J Cancer **80**(4): 606-611.

Hanasaki, K., Varki, A., and Powell, L.D. (1995). "CD22-mediated cell adhesion to cytokine-activated human endothelial cells. Positive and negative regulation by alpha 2-6-sialylation of cellular glycoproteins." J Biol Chem **270**(13): 7533-7542.

- Hardman, K.D., and Ainsworth, C.F. (1972). "Structure of concanavalin A at 2.4-Å resolution." Biochemistry **11**(26): 4910-4919.
- Haselhorst, T., Weimar, T., and Peters, T. (2001). "Molecular recognition of sialyl Lewis(x) and related saccharides by two lectins." J Am Chem Soc **123**(43): 10705-10714.
- Hatakeyama, T., Ishimine, T., Baba, T., Kimura, M., Unno, H., and Goda, S. (2013). "Alteration of the carbohydrate-binding specificity of a C-type lectin CEL-I mutant with an EPN carbohydrate-binding motif." Protein Pept Lett **20**(7): 796-801.
- Helenius, A., and Aebi, M. (2001). "Intracellular functions of N-linked glycans." Science **291**(5512): 2364-2369.
- Hester, G., and Wright, C.S. (1996). "The mannose-specific bulb lectin from *Galanthus nivalis* (snowdrop) binds mono- and dimannosides at distinct sites. Structure analysis of refined complexes at 2.3 Å and 3.0 Å resolution." J Mol Biol **262**(4): 516-531.
- Hirai, M., Minematsu, H., Kondo, N., Oie, K., Igarashi, K., and Yamazaki, N. (2007). "Accumulation of liposome with Sialyl Lewis X to inflammation and tumor region: application to in vivo bio-imaging." Biochem Biophys Res Commun **353**(3): 553-558.
- Hirst, G.K. (1942). "The Quantitative Determination of Influenza Virus and Antibodies by Means of Red Cell Agglutination." J Exp Med **75**(1): 49-64.
- Hoseki, J., Ushioda, R., and Nagata, K. (2010). "Mechanism and components of endoplasmic reticulum-associated degradation." J Biochem **147**(1): 19-25.
- Houser, J., Komarek, J., Kostlanova, N., Cioci, G., Varrot, A., Kerr, S.C., Lahmann, M., Balloy, V., Fahy, J.V., Chignard, M., *et al.* (2013). "A soluble fucose-specific lectin from *Aspergillus fumigatus* conidia--structure, specificity and possible role in fungal pathogenicity." PLoS One **8**(12): e83077.
- Hu, D., Tateno, H., Kuno, A., Yabe, R., and Hirabayashi, J. (2012). "Directed evolution of lectins with sugar-binding specificity for 6-sulfo-galactose." J Biol Chem **287**(24): 20313-20320.
- Huang, M.C., Zollner, O., Moll, T., Maly, P., Thall, A.D., Lowe, J.B., and Vestweber, D. (2000). "P-selectin glycoprotein ligand-1 and E-selectin ligand-1 are differentially modified by fucosyltransferases Fuc-TIV and Fuc-TVII in mouse neutrophils." J Biol Chem **275**(40): 31353-31360.

Huang, W., Dong, Z., Wei, H., Ding, C., Sun, R., and Tian, Z. (2006). "Selective elimination of hepatic natural killer T cells with concanavalin A improves liver regeneration in mice." Liver Int **26**(3): 339-345.

Hubbard, S.C., and Ivatt, R.J. (1981). "Synthesis and processing of asparagine-linked oligosaccharides." Annu Rev Biochem **50**: 555-583.

## I

Imberty, A., Audfray, A., and Arnaud, J. (2014). "Conception de néolectine de valence et spécificité modulable, Demande de brevet FR 14 51740 déposée le 4 mars 2014 par le CNRS et l'Université Joseph Fourier."

Imberty, A., Hardman, K.D., Carver, J.P., and Perez, S. (1991). "Molecular modelling of protein-carbohydrate interactions. Docking of monosaccharides in the binding site of concanavalin A." Glycobiology **1**(6): 631-642.

Imberty, A., Mikros, E., Koca, J., Mollicone, R., Oriol, R., and Perez, S. (1995). "Computer simulation of histo-blood group oligosaccharides: energy maps of all constituting disaccharides and potential energy surfaces of 14 ABH and Lewis carbohydrate antigens." Glycoconj J **12**(3): 331-349.

Imberty, A., Mitchell, E.P., and Wimmerova, M. (2005). "Structural basis of high-affinity glycan recognition by bacterial and fungal lectins." Curr Opin Struct Biol **15**(5): 525-534.

Imberty, A., and Perez, S. (2000). "Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations." Chem Rev **100**(12): 4567-4588.

Imberty, A., and Varrot, A. (2008). "Microbial recognition of human cell surface glycoconjugates." Curr Opin Struct Biol **18**(5): 567-576.

Imberty, A., wimmerova, M., Mitchell, E.P., and Gilboa-Garber, N. (2004). "Structures of the lectins from *Pseudomonas aeruginosa*: insight into the molecular basis for host glycan recognition." Microbes Infect **6**(2): 221-228.

Iobst, S.T., Wormald, M.R., Weis, W.I., Dwek, R.A., and Drickamer, K. (1994). "Binding of sugar ligands to Ca(2+)-dependent animal lectins. I. Analysis of mannose binding by site-directed mutagenesis and NMR." J Biol Chem **269**(22): 15505-15511.



Ito, N., Phillips, S.E., Stevens, C., Ogel, Z.B., McPherson, M.J., Keen, J.N., Yadav, K.D., and Knowles, P.F. (1991). "Novel thioether bond revealed by a 1.7 Å crystal structure of galactose oxidase." *Nature* **350**(6313): 87-90.

Itzkowitz, S.H., Bloom, E.J., Kokal, W.A., Modin, G., Hakomori, S., and Kim, Y.S. (1990). "Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients." *Cancer* **66**(9): 1960-1966.

## J

Jiang, S., Chen, Y., Wang, M., Yin, Y., Pan, Y., Gu, B., Yu, G., Li, Y., Wong, B.H., Liang, Y., and Sun, H. (2012). "A novel lectin from *Agrocybe aegerita* shows high binding selectivity for terminal N-acetylglucosamine." *Biochem J* **443**(2): 369-378.

Johannes, L., and Romer, W. (2010). "Shiga toxins--from cell biology to biomedical applications." *Nat Rev Microbiol* **8**(2): 105-116.

Johnston, D.S., Wright, W.W., Shaper, J.H., Hokke, C.H., Van den Eijnden, D.H., and Joziassse, D.H. (1998). "Murine sperm-zona binding, a fucosyl residue is required for a high affinity sperm-binding ligand. A second site on sperm binds a nonfucosylated, beta-galactosyl-capped oligosaccharide." *J Biol Chem* **273**(4): 1888-1895.

Jordan, E.T., and Goldstein, I.J. (1995). "Site-directed mutagenesis studies on the lima bean lectin. Altered carbohydrate-binding specificities result from single amino acid substitutions." *Eur J Biochem* **230**(3): 958-964.

Juge, N. (2012). "Microbial adhesins to gastrointestinal mucus." *Trends Microbiol* **20**(1): 30-39.

Julien, S., Adriaenssens, E., Ottenberg, K., Furlan, A., Courtand, G., Vercoutter-Edouart, A.S., Hanisch, F.G., Delannoy, P., and Le Bourhis, X. (2006). "ST6GalNAc I expression in MDA-MB-231 breast cancer cells greatly modifies their O-glycosylation pattern and enhances their tumorigenicity." *Glycobiology* **16**(1): 54-64.

## K

Kaku, H., Mori, Y., Goldstein, I.J., and Shibuya, N. (1993). "Monomeric, monovalent derivative of *Maackia amurensis* leucoagglutinin. Preparation and application to the study of cell surface glycoconjugates by flow cytometry." *J Biol Chem* **268**(18): 13237-13241.

- Kaku, H., and Shibuya, N. (1992). "Preparation of a stable subunit of Japanese elderberry (*Sambucus sieboldiana*) bark lectin and its application for the study of cell surface carbohydrates by flow cytometry." FEBS Lett **306**(2-3): 176-180.
- Kannagi, R. (2004). "Molecular mechanism for cancer-associated induction of sialyl Lewis X and sialyl Lewis A expression-The Warburg effect revisited." Glycoconj J **20**(5): 353-364.
- Kannagi, R., Izawa, M., Koike, T., Miyazaki, K., and Kimura, N. (2004). "Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis." Cancer Sci **95**(5): 377-384.
- Kasuya, T., Jung, J., Kadoya, H., Matsuzaki, T., Tatematsu, K., Okajima, T., Miyoshi, E., Tanizawa, K., and Kuroda, S. (2008). "In vivo delivery of bionanocapsules displaying *Phaseolus vulgaris* agglutinin-L4 isolectin to malignant tumors overexpressing N-acetylglucosaminyltransferase V." Hum Gene Ther **19**(9): 887-895.
- Kawasaki, T. (1999). "Structure and biology of mannan-binding protein, MBP, an important component of innate immunity." Biochim Biophys Acta **1473**(1): 186-195.
- Kentzer, E.J., Buko, A., Menon, G., and Sarin, V.K. (1990). "Carbohydrate composition and presence of a fucose-protein linkage in recombinant human pro-urokinase." Biochem Biophys Res Commun **171**(1): 401-406.
- Keogh, D., Thompson, R., Larragy, R., McMahon, K., O'Connell, M., O'Connor, B., and Clarke, P. (2014). "Generating novel recombinant prokaryotic lectins with altered carbohydrate binding properties through mutagenesis of the PA-IL protein from *Pseudomonas aeruginosa*." Biochim Biophys Acta **1840**(6): 2091-2104.
- Khien, V.V., Mao, H.V., Chinh, T.T., Ha, P.T., Bang, M.H., Lac, B.V., Hop, T.V., Tuan, N.A., Don, L.V., Taketa, K., and Satomura, S. (2001). "Clinical evaluation of lentil lectin-reactive alpha-fetoprotein-L3 in histology-proven hepatocellular carcinoma." Int J Biol Markers **16**(2): 105-111.
- Khil, L.Y., Kim, W., Lyu, S., Park, W.B., Yoon, J.W., and Jun, H.S. (2007). "Mechanisms involved in Korean mistletoe lectin-induced apoptosis of cancer cells." World J Gastroenterol **13**(20): 2811-2818.
- Kim, Y.S., Gum, J., Jr., and Brockhausen, I. (1996). "Mucin glycoproteins in neoplasia." Glycoconj J **13**(5): 693-707.

- Kolatkar, A.R., Leung, A.K., Isecke, R., Brossmer, R., Drickamer, K., and Weis, W.I. (1998). "Mechanism of N-acetylgalactosamine binding to a C-type animal lectin carbohydrate-recognition domain." J Biol Chem **273**(31): 19502-19508.
- Kolatkar, A.R., and Weis, W.I. (1996). "Structural basis of galactose recognition by C-type animal lectins." J Biol Chem **271**(12): 6679-6685.
- Kostlanova, N., Mitchell, E.P., Lortat-Jacob, H., Oscarson, S., Lahmann, M., Gilboa-Garber, N., Chambat, G., Wimmerova, M., and Imberty, A. (2005). "The fucose-binding lectin from *Ralstonia solanacearum*. A new type of beta-propeller architecture formed by oligomerization and interacting with fucoside, fucosyllactose, and plant xyloglucan." J Biol Chem **280**(30): 27839-27849.
- Kriz, Z., Adam, J., Mrazkova, J., Zotos, P., Chatzipavlou, T., Wimmerova, M., and Koca, J. (2014). "Engineering the *Pseudomonas aeruginosa* II lectin: designing mutants with changed affinity and specificity." J Comput Aided Mol Des.
- Kuhlmann, W.D., Peschke, P., and Wurster, K. (1983). "Lectin--peroxidase conjugates in histopathology of gastrointestinal mucosa." Virchows Arch A Pathol Anat Histopathol **398**(3): 319-328.
- Kuno, A., Ikehara, Y., Tanaka, Y., Angata, T., Unno, S., Sogabe, M., Ozaki, H., Ito, K., Hirabayashi, J., Mizokami, M., and Narimatsu, H. (2011). "Multilectin assay for detecting fibrosis-specific glyco-alteration by means of lectin microarray." Clin Chem **57**(1): 48-56.

## L

- Labarriere, N., Piau, J.P., Otry, C., Denis, M., Lustenberger, P., Meflah, K., and Le Pendu, J. (1994). "H blood group antigen carried by CD44V modulates tumorigenicity of rat colon carcinoma cells." Cancer Res **54**(23): 6275-6281.
- Lamblin, G., Degroote, S., Perini, J.M., Delmotte, P., Scharfman, A., Davril, M., Lo-Guidice, J.M., Houdret, N., Dumur, V., Klein, A., and Rouse, P. (2001). "Human airway mucin glycosylation: a combinatorial of carbohydrate determinants which vary in cystic fibrosis." Glycoconj J **18**(9): 661-684.
- Landsteiner, K. (1900). "Zur Kenntniss der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der L.ymphe." Zbl Bakt **27**: 357-366.

- Landsteiner, K. (1901). "Uber Agglunationserscheinungen normalen menschlichen Blutes." Wein Klin Wschr **14**: 1132-1134.
- Leffler, H., Carlsson, S., Hedlund, M., Qian, Y., and Poirier, F. (2004). "Introduction to galectins." Glycoconj J **19**(7-9): 433-440.
- Leivonen, M., Nordling, S., Lundin, J., von Boguslawski, K., and Haglund, C. (2001). "STn and prognosis in breast cancer." Oncology **61**(4): 299-305.
- Lemieux, R.U., Bock, K., Delbaere, L.T.J., Koto, S., and Rao, V.S. (1980). "The conformations of oligosaccharides related to the ABH and Lewis human blood-group determinants." Canadian journal chemistry **58**(6): 631-653.
- Li, J., Brick, P., O'Hare, M.C., Skarzynski, T., Lloyd, L.F., Curry, V.A., Clark, I.M., Bigg, H.F., Hazleman, B.L., Cawston, T.E., and et al. (1995). "Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed beta-propeller." Structure **3**(6): 541-549.
- Lingwood, C.A. (2011). "Glycosphingolipid functions." Cold Spring Harb Perspect Biol **3**(7).
- Liu, J., Blasie, C.A., Shi, S., Joshi, S.B., Middaugh, C.R., and Volkin, D.B. (2013). "Characterization and stabilization of recombinant human protein pentraxin (rhPTX-2)." J Pharm Sci **102**(3): 827-841.
- Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H., and Rini, J.M. (1993). "X-ray crystal structure of the human dimeric S-Lac lectin, L-14-II, in complex with lactose at 2.9-A resolution." J Biol Chem **268**(36): 27034-27038.
- Loris, R., Hamelryck, T., Bouckaert, J., and Wyns, L. (1998). "Legume lectin structure." Biochim Biophys Acta **1383**(1): 9-36.
- Loris, R., Stas, P.P., and Wyns, L. (1994). "Conserved waters in legume lectin crystal structures. The importance of bound water for the sequence-structure relationship within the legume lectin family." J Biol Chem **269**(43): 26722-26733.
- 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**(4): 861-870.

Lowe, J.B. (1997). "Selectin ligands, leukocyte trafficking, and fucosyltransferase genes." Kidney Int **51**(5): 1418-1426.

Lukas, A., Mayer, B., Fialova, D., Topinkova, E., Gindin, J., Onder, G., Bernabei, R., Nikolaus, T., and Denkinger, M.D. (2013). "Pain characteristics and pain control in European nursing homes: cross-sectional and longitudinal results from the Services and Health for Elderly in Long TERM care (SHELTER) study." J Am Med Dir Assoc **14**(6): 421-428.

## M

Ma, Y.H., Cheng, W.Z., Gong, F., Ma, A.L., Yu, Q.W., Zhang, J.Y., Hu, C.Y., Chen, X.H., and Zhang, D.Q. (2008). "Active Chinese mistletoe lectin-55 enhances colon cancer surveillance through regulating innate and adaptive immune responses." World J Gastroenterol **14**(34): 5274-5281.

Maenuma, K., Yim, M., Komatsu, K., Hoshino, M., Tachiki-Fujioka, A., Takahashi, K., Hiki, Y., Bovin, N., and Irimura, T. (2009). "A library of mutated Maackia amurensis hemagglutinin distinguishes putative glycoforms of immunoglobulin A1 from IgA nephropathy patients." J Proteome Res **8**(7): 3617-3624.

Maenuma, K., Yim, M., Komatsu, K., Hoshino, M., Takahashi, Y., Bovin, N., and Irimura, T. (2008). "Use of a library of mutated Maackia amurensis hemagglutinin for profiling the cell lineage and differentiation." Proteomics **8**(16): 3274-3283.

Marcos, N.T., Cruz, A., Silva, F., Almeida, R., David, L., Mandel, U., Clausen, H., Von Mensdorff-Pouilly, S., and Reis, C.A. (2003). "Polypeptide GalNAc-transferases, ST6GalNAc-transferase I, and ST3Gal-transferase I expression in gastric carcinoma cell lines." J Histochem Cytochem **51**(6): 761-771.

Maruyama, M., Onodera, K., and Ooya, K. (2002). "A histopathological and lectin-histochemical study of the lining epithelium in postoperative maxillary cysts." Oral Dis **8**(5): 241-248.

Mathieu, S., Prorok, M., Benoliel, A.M., Uch, R., Langlet, C., Bongrand, P., Gerolami, R., and El-Battari, A. (2004). "Transgene expression of alpha(1,2)-fucosyltransferase-I (FUT1) in tumor cells selectively inhibits sialyl-Lewis x expression and binding to E-selectin without affecting synthesis of sialyl-Lewis a or binding to P-selectin." Am J Pathol **164**(2): 371-383.

Matsumura, K., Higashida, K., Ishida, H., Hata, Y., Yamamoto, K., Shigeta, M., Mizuno-Horikawa, Y., Wang, X., Miyoshi, E., Gu, J., and Taniguchi, N. (2007).

- "Carbohydrate binding specificity of a fucose-specific lectin from *Aspergillus oryzae*: a novel probe for core fucose." J Biol Chem **282**(21): 15700-15708.
- Matthes, H., Friedel, W.E., Bock, P.R., and Zanker, K.S. (2010). "Molecular mistletoe therapy: friend or foe in established anti-tumor protocols? A multicenter, controlled, retrospective pharmaco-epidemiological study in pancreas cancer." Curr Mol Med **10**(4): 430-439.
- McCann, A.K., Schwartz, K.J., and Bangs, J.D. (2008). "A determination of the steady state lysosomal pH of bloodstream stage African trypanosomes." Mol Biochem Parasitol **159**(2): 146-149.
- McDonald, R.E., Hughes, D.J., and Davis, B.G. (2004). "Modular control of lectin function: redox-switchable agglutination." Angew Chem Int Ed Engl **43**(23): 3025-3029.
- Mehlert, A., Wormald, M.R., and Ferguson, M.A. (2012). "Modeling of the N-glycosylated transferrin receptor suggests how transferrin binding can occur within the surface coat of *Trypanosoma brucei*." PLoS Pathog **8**(4): e1002618.
- Mengs, U., Schwarz, T., Bulitta, M., and Weber, K. (2000). "Antitumoral effects of an intravesically applied aqueous mistletoe extract on urinary bladder carcinoma MB49 in mice." Anticancer Res **20**(5B): 3565-3568.
- Merritt, E.A., Sarfaty, S., van den Akker, F., L'Hoir, C., Martial, J.A., and Hol, W.G. (1994). "Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide." Protein Sci **3**(2): 166-175.
- Mewe, M., Tielker, D., Schonberg, R., Schachner, M., Jaeger, K.E., and Schumacher, U. (2005). "Pseudomonas aeruginosa lectins I and II and their interaction with human airway cilia." J Laryngol Otol **119**(8): 595-599.
- Midlej, V., Meinig, I., de Souza, W., and Benchimol, M. (2013). "A new set of carbohydrate-positive vesicles in encysting *Giardia lamblia*." Protist **164**(2): 261-271.
- Mishra, S.K., Adam, J., Wimmerova, M., and Koca, J. (2012). "In silico mutagenesis and docking study of *Ralstonia solanacearum* RSL lectin: performance of docking software to predict saccharide binding." J Chem Inf Model **52**(5): 1250-1261.

- Mitchell, B.S., Brooks, S.A., Leathem, A.J., and Schumacher, U. (1998). "Do HPA and PHA-L have the same binding pattern in metastasizing human breast and colon cancers?" Cancer Lett **123**(1): 113-119.
- Mitchell, E., Houles, C., Sudakevitz, D., Wimmerova, M., Gautier, C., Perez, S., Wu, A.M., Gilboa-Garber, N., and Imberty, A. (2002). "Structural basis for oligosaccharide-mediated adhesion of *Pseudomonas aeruginosa* in the lungs of cystic fibrosis patients." Nat Struct Biol **9**(12): 918-921.
- Miyagi, T., Takehara, T., Tatsumi, T., Suzuki, T., Jinushi, M., Kanazawa, Y., Hiramatsu, N., Kanto, T., Tsuji, S., Hori, M., and Hayashi, N. (2004). "Concanavalin a injection activates intrahepatic innate immune cells to provoke an antitumor effect in murine liver." Hepatology **40**(5): 1190-1196.
- Miyoshi, E., Moriwaki, K., and Nakagawa, T. (2008). "Biological function of fucosylation in cancer biology." J Biochem **143**(6): 725-729.
- Miyoshi, E., Noda, K., Yamaguchi, Y., Inoue, S., Ikeda, Y., Wang, W., Ko, J.H., Uozumi, N., Li, W., and Taniguchi, N. (1999). "The alpha1-6-fucosyltransferase gene and its biological significance." Biochim Biophys Acta **1473**(1): 9-20.
- Miyoshi, E., Uozumi, N., Noda, K., Hayashi, N., Hori, M., and Taniguchi, N. (1997). "Expression of alpha1-6 fucosyltransferase in rat tissues and human cancer cell lines." Int J Cancer **72**(6): 1117-1121.
- Moloney, D.J., and Haltiwanger, R.S. (1999). "The O-linked fucose glycosylation pathway: identification and characterization of a uridine diphosphoglucose: fucose-beta1,3-glucosyltransferase activity from Chinese hamster ovary cells." Glycobiology **9**(7): 679-687.
- Morgan, G.W., Goulding, D., and Field, M.C. (2004). "The single dynamin-like protein of *Trypanosoma brucei* regulates mitochondrial division and is not required for endocytosis." J Biol Chem **279**(11): 10692-10701.
- Morgan, G.W., Kail, M., Hollinshead, M., and Vaux, D.J. (2013). "Combined biochemical and cytological analysis of membrane trafficking using lectins." Anal Biochem **441**(1): 21-31.
- Mourant, A.E. (1946). "A new human blood group antigen of frequent occurrence." Nature **158**: 237.



- Muinelo-Romay, L., Gil-Martin, E., and Fernandez-Briera, A. (2010). "alpha(1,2)fucosylation in human colorectal carcinoma." *Oncol Lett* **1**(2): 361-366.
- Muinelo-Romay, L., Vazquez-Martin, C., Villar-Portela, S., Cuevas, E., Gil-Martin, E., and Fernandez-Briera, A. (2008). "Expression and enzyme activity of alpha(1,6)fucosyltransferase in human colorectal cancer." *Int J Cancer* **123**(3): 641-646.
- Munro, J.M., Lo, S.K., Corless, C., Robertson, M.J., Lee, N.C., Barnhill, R.L., Weinberg, D.S., and Bevilacqua, M.P. (1992). "Expression of sialyl-Lewis X, an E-selectin ligand, in inflammation, immune processes, and lymphoid tissues." *Am J Pathol* **141**(6): 1397-1408.
- Muthing, J., Maurer, U., Sostaric, K., Neumann, U., Brandt, H., Duvar, S., Peter-Katalinic, J., and Weber-Schurholz, S. (1994). "Different distributions of glycosphingolipids in mouse and rabbit skeletal muscle demonstrated by biochemical and immunohistological analyses." *J Biochem* **115**(2): 248-256.

## N

- Newton, K., and Dixit, V.M. (2012). "Signaling in innate immunity and inflammation." *Cold Spring Harb Perspect Biol* **4**(3).
- Ng, K.K., and Weis, W.I. (1997). "Structure of a selectin-like mutant of mannose-binding protein complexed with sialylated and sulfated Lewis(x) oligosaccharides." *Biochemistry* **36**(5): 979-988.
- Nishiguchi, M., Yoshida, K., Sumizono, T., and Tazaki, K. (1997). "Studies by site-directed mutagenesis of the carbohydrate-binding properties of a bark lectin from Robinia pseudoacacia." *FEBS Lett* **403**(3): 294-298.
- Niv, Y., and Fass, R. (2012). "The role of mucin in GERD and its complications." *Nat Rev Gastroenterol Hepatol* **9**(1): 55-59.
- Noda, K., Miyoshi, E., Gu, J., Gao, C.X., Nakahara, S., Kitada, T., Honke, K., Suzuki, K., Yoshihara, H., Yoshikawa, K., *et al.* (2003). "Relationship between elevated FX expression and increased production of GDP-L-fucose, a common donor substrate for fucosylation in human hepatocellular carcinoma and hepatoma cell lines." *Cancer Res* **63**(19): 6282-6289.

Nolan, D.P., Geuskens, M., and Pays, E. (1999). "N-linked glycans containing linear poly-N-acetyllactosamine as sorting signals in endocytosis in *Trypanosoma brucei*." Curr Biol **9**(20): 1169-1172.

Numahata, K., Satoh, M., Handa, K., Saito, S., Ohyama, C., Ito, A., Takahashi, T., Hoshi, S., Orikasa, S., and Hakomori, S.I. (2002). "Sialosyl-Le(x) expression defines invasive and metastatic properties of bladder carcinoma." Cancer **94**(3): 673-685.

Nurisso, A., Blanchard, B., Audfray, A., Rydner, L., Oscarson, S., Varrot, A., and Imberty, A. (2010). "Role of water molecules in structure and energetics of *Pseudomonas aeruginosa* lectin I interacting with disaccharides." J Biol Chem **285**(26): 20316-20327.

## O

Ohtsubo, K., and Marth, J.D. (2006). "Glycosylation in cellular mechanisms of health and disease." Cell **126**(5): 855-867.

Okuyama, N., Ide, Y., Nakano, M., Nakagawa, T., Yamanaka, K., Moriwaki, K., Murata, K., Ohigashi, H., Yokoyama, S., Eguchi, H., *et al.* (2006). "Fucosylated haptoglobin is a novel marker for pancreatic cancer: a detailed analysis of the oligosaccharide structure and a possible mechanism for fucosylation." Int J Cancer **118**(11): 2803-2808.

Olausson, J., Astrom, E., Jonsson, B.H., Tibell, L.A., and Pahlsson, P. (2011). "Production and characterization of a monomeric form and a single-site form of *Aleuria aurantia* lectin." Glycobiology **21**(1): 34-44.

Olson, L.J., Zhang, J., Dahms, N.M., and Kim, J.J. (2002). "Twists and turns of the cation-dependent mannose 6-phosphate receptor. Ligand-bound versus ligand-free receptor." J Biol Chem **277**(12): 10156-10161.

Oriol, R., Ed. (1995). ABO, Hh, Lewis, and Secretion : Serology, genetics, and tissue distribution. Molecular Basis of Human Blood Group Antigens. New York, Cartron, Jean-Pierre, Rouger, Philippe (Eds.).

Oriol, R., Mollicone, R., Coullin, P., Dalix, A.M., and Candelier, J.J. (1992). "Genetic regulation of the expression of ABH and Lewis antigens in tissues." APMIS Suppl **27**: 28-38.

Osumi, D., Takahashi, M., Miyoshi, E., Yokoe, S., Lee, S.H., Noda, K., Nakamori, S., Gu, J., Ikeda, Y., Kuroki, Y., *et al.* (2009). "Core fucosylation of E-cadherin enhances cell-cell adhesion in human colon carcinoma WiDr cells." Cancer Sci **100**(5): 888-895.

Oubrie, A., Rozeboom, H.J., Kalk, K.H., Duine, J.A., and Dijkstra, B.W. (1999). "The 1.7 Å crystal structure of the apo form of the soluble quinoprotein glucose dehydrogenase from *Acinetobacter calcoaceticus* reveals a novel internal conserved sequence repeat." J Mol Biol **289**(2): 319-333.

## P

Paoli, M., Anderson, B.F., Baker, H.M., Morgan, W.T., Smith, A., and Baker, E.N. (1999). "Crystal structure of hemopexin reveals a novel high-affinity heme site formed between two beta-propeller domains." Nat Struct Biol **6**(10): 926-931.

Park, W.B., Lyu, S.Y., Kim, J.H., Choi, S.H., Chung, H.K., Ahn, S.H., Hong, S.Y., Yoon, T.J., and Choi, M.J. (2001). "Inhibition of tumor growth and metastasis by Korean mistletoe lectin is associated with apoptosis and antiangiogenesis." Cancer Biother Radiopharm **16**(5): 439-447.

Perez, S., Mouhous-Riou, N., Nifant'ev, N.E., Tsvetkov, Y.E., Bachet, B., and Imberty, A. (1996). "Crystal and molecular structure of a histo-blood group antigen involved in cell adhesion: the Lewis x trisaccharide." Glycobiology **6**(5): 537-542.

Phillips, K., and de la Pena, A.H. (2011). "The combined use of the ThermoFluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization." Curr Protoc Mol Biol **Chapter 10**: Unit10 28.

Pilobello, K.T., Slawek, D.E., and Mahal, L.K. (2007). "A ratiometric lectin microarray approach to analysis of the dynamic mammalian glycome." Proc Natl Acad Sci U S A **104**(28): 11534-11539.

Pinho, S.S., Matos, A.J., Lopes, C., Marcos, N.T., Carvalheira, J., Reis, C.A., and Gartner, F. (2007). "Sialyl Lewis x expression in canine malignant mammary tumours: correlation with clinicopathological features and E-Cadherin expression." BMC Cancer **7**: 124.

Poget, S.F., Legge, G.B., Proctor, M.R., Butler, P.J., Bycroft, M., and Williams, R.L. (1999). "The structure of a tunicate C-type lectin from *Polyandrocampa misakiensis* complexed with D-galactose." J Mol Biol **290**(4): 867-879.

Pokrovskaya, I.D., Willett, R., Smith, R.D., Morelle, W., Kudlyk, T., and Lupashin, V.V. (2011). "Conserved oligomeric Golgi complex specifically regulates the maintenance of Golgi glycosylation machinery." Glycobiology **21**(12): 1554-1569.

Pryme, I.F., Bardocz, S., Pusztai, A., and Ewen, S.W. (2002). "Dietary mistletoe lectin supplementation and reduced growth of a murine non-Hodgkin lymphoma." Histol Histopathol **17**(1): 261-271.

Pryme, I.F., Bardocz, S., Pusztai, A., Ewen, S.W., and Pfuller, U. (2004). "A mistletoe lectin (ML-1)-containing diet reduces the viability of a murine non-Hodgkin lymphoma tumor." Cancer Detect Prev **28**(1): 52-56.

## Q

Quiocho, F.A., Wilson, D.K., and Vyas, N.K. (1989). "Substrate specificity and affinity of a protein modulated by bound water molecules." Nature **340**(6232): 404-407.

## R

Raats, C.J., Van Den Born, J., and Berden, J.H. (2000). "Glomerular heparan sulfate alterations: mechanisms and relevance for proteinuria." Kidney Int **57**(2): 385-400.

Renault, L., Nassar, N., Vetter, I., Becker, J., Klebe, C., Roth, M., and Wittinghofer, A. (1998). "The 1.7 Å crystal structure of the regulator of chromosome condensation (RCC1) reveals a seven-bladed propeller." Nature **392**(6671): 97-101.

Rhodes, J.M., Campbell, B.J., and Yu, L.G. (2008). "Lectin-epithelial interactions in the human colon." Biochem Soc Trans **36**(Pt 6): 1482-1486.

Romer, W., Berland, L., Chambon, V., Gaus, K., Windschiegel, B., Tenza, D., Aly, M.R., Fraasier, V., Florent, J.C., Perrais, D., *et al.* (2007). "Shiga toxin induces tubular membrane invaginations for its uptake into cells." Nature **450**(7170): 670-675.

Romer, W., Pontani, L.L., Sorre, B., Rentero, C., Berland, L., Chambon, V., Lamaze, C., Bassereau, P., Sykes, C., Gaus, K., and Johannes, L. (2010). "Actin dynamics drive membrane reorganization and scission in clathrin-independent endocytosis." Cell **140**(4): 540-553.

Rossner, P., Jr., Mrhalkova, A., Uhlirova, K., Spatova, M., Rossnerova, A., Libalova, H., Schmuczerova, J., Milcova, A., Topinka, J., and Sram, R.J. (2013). "Nucleotide excision repair is not induced in human embryonic lung fibroblasts treated with environmental pollutants." PLoS One **8**(7): e69197.

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**(3): 240-250.

## S

- Sakuma, S., Yano, T., Masaoka, Y., Kataoka, M., Hiwatari, K., Tachikawa, H., Shoji, Y., Kimura, R., Ma, H., Yang, Z., *et al.* (2009). "In vitro/in vivo biorecognition of lectin-immobilized fluorescent nanospheres for human colorectal cancer cells." J Control Release **134**(1): 2-10.
- Sakuma, S., Yano, T., Masaoka, Y., Kataoka, M., Hiwatari, K., Tachikawa, H., Shoji, Y., Kimura, R., Ma, H., Yang, Z., *et al.* (2010). "Detection of early colorectal cancer imaged with peanut agglutinin-immobilized fluorescent nanospheres having surface poly(N-vinylacetamide) chains." Eur J Pharm Biopharm **74**(3): 451-460.
- Salomonsson, E., Carlsson, M.C., Osla, V., Hendus-Altenburger, R., Kahl-Knutson, B., Oberg, C.T., Sundin, A., Nilsson, R., Nordberg-Karlsson, E., Nilsson, U.J., *et al.* (2010). "Mutational tuning of galectin-3 specificity and biological function." J Biol Chem **285**(45): 35079-35091.
- Sanchez, J.F., Lescar, J., Chazalet, V., Audfray, A., Gagnon, J., Alvarez, R., Breton, C., Imberty, A., and Mitchell, E.P. (2006). "Biochemical and structural analysis of Helix pomatia agglutinin. A hexameric lectin with a novel fold." J Biol Chem **281**(29): 20171-20180.
- Sankaranarayanan, R., Sekar, K., Banerjee, R., Sharma, V., Surolia, A., and Vijayan, M. (1996). "A novel mode of carbohydrate recognition in jacalin, a Moraceae plant lectin with a beta-prism fold." Nat Struct Biol **3**(7): 596-603.
- Sauter, N.K., Hanson, J.E., Glick, G.D., Brown, J.H., Crowther, R.L., Park, S.J., Skehel, J.J., and Wiley, D.C. (1992). "Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography." Biochemistry **31**(40): 9609-9621.
- Schachter, H. (1986). "Biosynthetic controls that determine the branching and microheterogeneity of protein-bound oligosaccharides." Adv Exp Med Biol **205**: 53-85.
- Schaffrath, B., Mengs, U., Schwarz, T., Hilgers, R., Beuth, J., Möckel, B., Lentzen, H., and Gerstmayer, B. (2001). "Anticancer activity of rViscumin (recombinant mistletoe lectin) in tumor colonization models with immunocompetent mice." Anticancer Res. **21**: 3981-3987.

- Schmelz, M., Sodeik, B., Ericsson, M., Wolffe, E.J., Shida, H., Hiller, G., and Griffiths, G. (1994). "Assembly of vaccinia virus: the second wrapping cisterna is derived from the trans Golgi network." J Virol **68**(1): 130-147.
- Schoffski, P., Riggert, S., Fumoleau, P., Campone, M., Bolte, O., Marreaud, S., Lacombe, D., Baron, B., Herold, M., Zwierzina, H., *et al.* (2004). "Phase I trial of intravenous aviscumine (rViscumin) in patients with solid tumors: a study of the European Organization for Research and Treatment of Cancer New Drug Development Group." Ann Oncol **15**(12): 1816-1824.
- Schrag, J.D., Bergeron, J.J., Li, Y., Borisova, S., Hahn, M., Thomas, D.Y., and Cygler, M. (2001). "The Structure of calnexin, an ER chaperone involved in quality control of protein folding." Mol Cell **8**(3): 633-644.
- Schumacher, U., Feldhaus, S., and Mengers, U. (2000). "Recombinant mistletoe lectin (rML) is successful in treating human ovarian cancer cells transplanted into severe combined immunodeficient (SCID) mice." Cancer Lett **150**(2): 171-175.
- Seifert, G., Jesse, P., Laengler, A., Reindl, T., Luth, M., Lobitz, S., Henze, G., Prokop, A., and Lode, H.N. (2008). "Molecular mechanisms of mistletoe plant extract-induced apoptosis in acute lymphoblastic leukemia in vivo and in vitro." Cancer Lett **264**(2): 218-228.
- Sewell, R., Backstrom, M., Dalziel, M., Gschmeissner, S., Karlsson, H., Noll, T., Gatgens, J., Clausen, H., Hansson, G.C., Burchell, J., and Taylor-Papadimitriou, J. (2006). "The ST6GalNAc-I sialyltransferase localizes throughout the Golgi and is responsible for the synthesis of the tumor-associated sialyl-Tn O-glycan in human breast cancer." J Biol Chem **281**(6): 3586-3594.
- Sharma, V., Vijayan, M., and Surolia, A. (1996). "Imparting exquisite specificity to peanut agglutinin for the tumor-associated Thomsen-Friedenreich antigen by redesign of its combining site." J Biol Chem **271**(35): 21209-21213.
- Sharon, N. (1996). "Carbohydrate-lectin interactions in infectious disease." Adv Exp Med Biol **408**: 1-8.
- Shrive, A.K., Cheetham, G.M., Holden, D., Myles, D.A., Turnell, W.G., Volanakis, J.E., Pepys, M.B., Bloomer, A.C., and Greenhough, T.J. (1996). "Three dimensional structure of human C-reactive protein." Nat Struct Biol **3**(4): 346-354.

- Simm, S., Markowska, J., Tomaszewicz, T., Herwichowska, K., and Breborowicz, J. (1979). "[Monitoring the course of malignant neoplasms of the genital organs in women by means of determination of blood serum level of fucose, sialic acid, CEA and AFP]." Ginekol Pol **50**(12): 1029-1036.
- Smerdova, L., Neca, J., Svobodova, J., Topinka, J., Schmuczerova, J., Kozubik, A., Machala, M., and Vondracek, J. (2013). "Inflammatory mediators accelerate metabolism of benzo[a]pyrene in rat alveolar type II cells: the role of enhanced cytochrome P450 1B1 expression." Toxicology **314**(1): 30-38.
- Somers, W.S., Tang, J., Shaw, G.D., and Camphausen, R.T. (2000). "Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1." Cell **103**(3): 467-479.
- Springer, G.F. (1997). "Immunoreactive T and Tn epitopes in cancer diagnosis, prognosis, and immunotherapy." J Mol Med (Berl) **75**(8): 594-602.
- Stillmark, H. (1888). "Uber ricin, eines gifiges ferment aus den samen von Ricinus communis L. Und einigen anderen Euphorbiaceen." PhD Thesis, University of Dorpat, Estonia.
- Sugawara, H., Kusunoki, M., Kurisu, G., Fujimoto, T., Aoyagi, H., and Hatakeyama, T. (2004). "Characteristic recognition of N-acetylgalactosamine by an invertebrate C-type Lectin, CEL-I, revealed by X-ray crystallographic analysis." J Biol Chem **279**(43): 45219-45225.

## T

- Takada, A., Ohmori, K., Yoneda, T., Tsuyuoka, K., Hasegawa, A., Kiso, M., and Kannagi, R. (1993). "Contribution of carbohydrate antigens sialyl Lewis A and sialyl Lewis X to adhesion of human cancer cells to vascular endothelium." Cancer Res **53**(2): 354-361.
- Takahashi, M., Kuroki, Y., Ohtsubo, K., and Taniguchi, N. (2009). "Core fucose and bisecting GlcNAc, the direct modifiers of the N-glycan core: their functions and target proteins." Carbohydr Res **344**(12): 1387-1390.
- Taketa, K., Endo, Y., Sekiya, C., Tanikawa, K., Koji, T., Taga, H., Satomura, S., Matsuura, S., Kawai, T., and Hirai, H. (1993). "A collaborative study for the evaluation of lectin-reactive alpha-fetoproteins in early detection of hepatocellular carcinoma." Cancer Res **53**(22): 5419-5423.



- Takizawa, M., Nomura, T., Wakisaka, E., Yoshizuka, N., Aoki, J., Arai, H., Inoue, K., Hattori, M., and Matsuo, N. (1999). "cDNA cloning and expression of human lactosylceramide synthase." Biochim Biophys Acta **1438**(2): 301-304.
- Taniguchi, K., Kawano, Y.I., Toshitani, A., Karashima, A., and Nomoto, K. (1989). "Anti-metastatic effect by in vivo administration of concanavalin A through augmentation of T-derived activated killer activity: efficacy to B16 melanoma expressed MHC antigen." Cell Immunol **120**(2): 460-469.
- Tateno, H., Toyota, M., Saito, S., Onuma, Y., Ito, Y., Hiemori, K., Fukumura, M., Matsushima, A., Nakanishi, M., Ohnuma, K., *et al.* (2011). "Glycome diagnosis of human induced pluripotent stem cells using lectin microarray." J Biol Chem **286**(23): 20345-20353.
- ter Haar, E., Musacchio, A., Harrison, S.C., and Kirchhausen, T. (1998). "Atomic structure of clathrin: a beta propeller terminal domain joins an alpha zigzag linker." Cell **95**(4): 563-573.
- Thies, A., Dautel, P., Meyer, A., Pfuller, U., and Schumacher, U. (2008). "Low-dose mistletoe lectin-I reduces melanoma growth and spread in a scid mouse xenograft model." Br J Cancer **98**(1): 106-112.
- 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**(Pt 5): 1313-1323.
- Tikkanen, R., Peltola, M., Oinonen, C., Rouvinen, J., and Peltonen, L. (1997). "Several cooperating binding sites mediate the interaction of a lysosomal enzyme with phosphotransferase." EMBO J **16**(22): 6684-6693.
- Topfer-Petersen, E., Romero, A., Varela, P.F., Ekhlasi-Hundrieser, M., Dostalova, Z., Sanz, L., and Calvete, J.J. (1998). "Spermadhesins: a new protein family. Facts, hypotheses and perspectives." Andrologia **30**(4-5): 217-224.
- Topin, J., Arnaud, J., Sarkar, A., Audfray, A., Gillon, E., Perez, S., Jamet, H., Varrot, A., Imberty, A., and Thomas, A. (2013). "Deciphering the glycan preference of bacterial lectins by glycan array and molecular docking with validation by microcalorimetry and crystallography." PLoS One **8**(8): e71149.

- Topin, J., Diharce, J., Fiorucci, S., Antonczak, S., and Golebiowski, J. (2014). "O(2) migration rates in [NiFe] hydrogenases. A joint approach combining free-energy calculations and kinetic modeling." J Phys Chem B **118**(3): 676-681.
- Transue, T.R., Smith, A.K., Mo, H., Goldstein, I.J., and Saper, M.A. (1997). "Structure of benzyl T-antigen disaccharide bound to *Amaranthus caudatus* agglutinin." Nat Struct Biol **4**(10): 779-783.
- Trombetta, E.S., and Helenius, A. (1998). "Lectins as chaperones in glycoprotein folding." Curr Opin Struct Biol **8**(5): 587-592.
- Turner, G.A. (1995). "Haptoglobin. A potential reporter molecule for glycosylation changes in disease." Adv Exp Med Biol **376**: 231-238.

## U

- Uchida, T., Yamasaki, T., Eto, S., Sugawara, H., Kurisu, G., Nakagawa, A., Kusunoki, M., and Hatakeyama, T. (2004). "Crystal structure of the hemolytic lectin CEL-III isolated from the marine invertebrate *Cucumaria echinata*: implications of domain structure for its membrane pore-formation mechanism." J Biol Chem **279**(35): 37133-37141.
- Ueno, T., Ohtawa, K., Kimoto, Y., Sakurai, K., Kodera, Y., Hiroto, M., Matsushima, A., Nishimura, H., and Inada, Y. (2000). "Polyethylene glycol-modified concanavalin A as an effective agent to stimulate anti-tumor cytotoxicity." Cancer Detect Prev **24**(1): 100-106.

## V

- Valentijn, J.A., van Weeren, L., Ultee, A., and Koster, A.J. (2007). "Novel localization of Rab3D in rat intestinal goblet cells and Brunner's gland acinar cells suggests a role in early Golgi trafficking." Am J Physiol Gastrointest Liver Physiol **293**(1): G165-177.
- Varghese, J.N., Laver, W.G., and Colman, P.M. (1983). "Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution." Nature **303**(5912): 35-40.
- Varki, A., Kannagi, R., and Toole, B.P. (2009). "Glycosylation Changes in Cancer."
- Vellieux, F.M., Huitema, F., Groendijk, H., Kalk, K.H., Jzn, J.F., Jongejan, J.A., Duine, J.A., Petratos, K., Drenth, J., and Hol, W.G. (1989). "Structure of quinoprotein methylamine dehydrogenase at 2.25 Å resolution." EMBO J **8**(8): 2171-2178.

Vestweber, D., and Blanks, J.E. (1999). "Mechanisms that regulate the function of the selectins and their ligands." Physiol Rev **79**(1): 181-213.

von Decastello, A., and Sturli, A. (1902). "Über die isoagglutinine im serum gesunder und kranker Menschen." Munchen Medicine Wchnschr **49**: 1090-1095.

Vyas, N.K., Vyas, M.N., and Quioco, F.A. (1991). "Comparison of the periplasmic receptors for L-arabinose, D-glucose/D-galactose, and D-ribose. Structural and Functional Similarity." J Biol Chem **266**(8): 5226-5237.

## W

Wall, M.A., Coleman, D.E., Lee, E., Iniguez-Lluhi, J.A., Posner, B.A., Gilman, A.G., and Sprang, S.R. (1995). "The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2." Cell **83**(6): 1047-1058.

Wang, H., He, L., Lensch, M., Gabius, H.J., Fee, C.J., and Middelberg, A.P. (2008). "Single-site Cys-substituting mutation of human lectin galectin-2: modulating solubility in recombinant production, reducing long-term aggregation, and enabling site-specific monoPEGylation." Biomacromolecules **9**(11): 3223-3230.

Wang, Y., Shao, L., Shi, S., Harris, R.J., Spellman, M.W., Stanley, P., and Haltiwanger, R.S. (2001). "Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase." J Biol Chem **276**(43): 40338-40345.

Watanabe, A., Taketa, K., and Kosaka, K. (1975). "Microheterogeneity of rat alpha-fetoprotein." Ann N Y Acad Sci **259**: 95-108.

Weis, W.I., Brunger, A.T., Skehel, J.J., and Wiley, D.C. (1990). "Refinement of the influenza virus hemagglutinin by simulated annealing." J Mol Biol **212**(4): 737-761.

Wimmerova, M., Kozmon, S., Necasova, I., Mishra, S.K., Komarek, J., and Koca, J. (2012). "Stacking interactions between carbohydrate and protein quantified by combination of theoretical and experimental methods." PLoS One **7**(10): e46032.

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**(1): 131-137.

Wright, C.S. (1990). "2.2 Å resolution structure analysis of two refined N-acetylneuraminyl-lactose--wheat germ agglutinin isolectin complexes." J Mol Biol **215**(4): 635-651.

## X

Xia, Z.X., Dai, W.W., Xiong, J.P., Hao, Z.P., Davidson, V.L., White, S., and Mathews, F.S. (1992). "The three-dimensional structures of methanol dehydrogenase from two methylotrophic bacteria at 2.6-Å resolution." J Biol Chem **267**(31): 22289-22297.

## Y

Yabe, R., Itakura, Y., Nakamura-Tsuruta, S., Iwaki, J., Kuno, A., and Hirabayashi, J. (2009). "Engineering a versatile tandem repeat-type alpha2-6sialic acid-binding lectin." Biochem Biophys Res Commun **384**(2): 204-209.

Yabe, R., Suzuki, R., Kuno, A., Fujimoto, Z., Jigami, Y., and Hirabayashi, J. (2007). "Tailoring a novel sialic acid-binding lectin from a ricin-B chain-like galactose-binding protein by natural evolution-mimicry." J Biochem **141**(3): 389-399.

Yamamoto, K., Konami, Y., Osawa, T., and Irimura, T. (1992). "Alteration of the carbohydrate-binding specificity of the Bauhinia purpurea lectin through the preparation of a chimeric lectin." J Biochem **111**(1): 87-90.

Yazawa, S., Nakamura, J., Asao, T., Nagamachi, Y., Sagi, M., Matta, K.L., Tachikawa, T., and Akamatsu, M. (1993). "Aberrant alpha 1-->2fucosyltransferases found in human colorectal carcinoma involved in the accumulation of Leb and Y antigens in colorectal tumors." Jpn J Cancer Res **84**(9): 989-995.

Yim, M., Ono, T., and Irimura, T. (2001). "Mutated plant lectin library useful to identify different cells." Proc Natl Acad Sci U S A **98**(5): 2222-2225.

Yoshimura, M., Nishikawa, A., Ihara, Y., Taniguchi, S., and Taniguchi, N. (1995). "Suppression of lung metastasis of B16 mouse melanoma by N-acetylglucosaminyltransferase III gene transfection." Proc Natl Acad Sci U S A **92**(19): 8754-8758.

## Z

Zaccari, N.R., Maenaka, K., Maenaka, T., Crocker, P.R., Brossmer, R., Kelm, S., and Jones, E.Y. (2003). "Structure-guided design of sialic acid-based Siglec inhibitors and crystallographic analysis in complex with sialoadhesin." Structure **11**(5): 557-567.

- Zhang, S., Zhang, H.S., Cordon-Cardo, C., Reuter, V.E., Singhal, A.K., Lloyd, K.O., and Livingston, P.O. (1997). "Selection of tumor antigens as targets for immune attack using immunohistochemistry: II. Blood group-related antigens." Int J Cancer **73**(1): 50-56.
- Zhao, Y.P., Xu, X.Y., Fang, M., Wang, H., You, Q., Yi, C.H., Ji, J., Gu, X., Zhou, P.T., Cheng, C., and Gao, C.F. (2014). "Decreased core-fucosylation contributes to malignancy in gastric cancer." PLoS One **9**(4): e94536.
- Zhuang, H., Hu, Z., Tan, M., Zhu, L., Liu, J., Liu, D., Yan, L., and Lin, B. (2014). "Overexpression of Lewis y antigen promotes human epididymis protein 4-mediated invasion and metastasis of ovarian cancer cells." Biochimie.
- Zierke, M., Smiesko, M., Rabbani, S., Aeschbacher, T., Cutting, B., Allain, F.H., Schubert, M., and Ernst, B. (2013). "Stabilization of branched oligosaccharides: Lewis(x) benefits from a nonconventional C-H...O hydrogen bond." J Am Chem Soc **135**(36): 13464-13472.







---

# **ANNEXES**

---



---

# ANNEXES

---

## I. Brevet (Confidentiel)

## Nouvelles lectines et applications pour la détection de marqueurs d'états pathologiques

L'invention a pour objet des nouvelles lectines, utilisables, notamment, pour  
5 détecter spécifiquement des marqueurs d'états pathologiques.

Les lectines sont des protéines, d'origine eucaryote ou procaryote, qui se  
lient aux sucres, mono-, oligo- et polysaccharides, de manière généralement  
réversible, avec une forte spécificité, sans les modifier. Sur la surface des  
10 cellules, les sucres sont présents dans des glycoconjugués résultant de leur  
liaison à des protéines ou à des lipides. Le terme « glycanes » est  
généralement employé pour désigner les sucres des glycoconjugués.

Les lectines sont habituellement des protéines multimériques, par  
15 oligomérisation ou répétition en tandem de motifs monomériques (également  
désignés ci-après par le terme « modules »), ce qui leur confère un rôle clé  
dans l'agglutination, l'affinité pour la surface cellulaire, le recrutement de  
glycolipides, ou encore le marquage de tissus.

Les chaînes peptidiques de chaque module présentent des extrémités C-ter  
20 et N-ter et comprennent des sites de fixation aux glycanes.

De nombreuses fonctions biologiques, par exemple la signalisation cellulaire,  
la migration cellulaire, la reconnaissance immunitaire ou encore l'interaction  
avec des pathogènes, mettent en jeu des interactions entre protéines et  
25 glycanes. La mise à profit de ces interactions se heurte toutefois aux  
difficultés rencontrées dans la détermination de la structure des sucres  
impliqués.

Un effort important a été apporté ces dernières années à l'étude de leurs  
30 fonctions et des mécanismes de leur association avec des ligands  
spécifiques.

La production de nouvelles lectines par bio-ingénierie reste cependant à ce jour un défi.

Mettant à profit leur expérience dans ce domaine, les inventeurs ont orienté  
5 leurs recherches vers la production de nouvelles lectines non-naturelles,  
appelées ci-après néolectines, en utilisant le concept de la multivalence  
naturelle des lectines et en partant de lectines ayant une architecture de  
base adaptée pour la reconnaissance de sucres à la surface externe des  
10 cellules. Cette démarche a permis de créer, à partir des lectines existantes,  
des néolectines de valence et/ou de spécificité modulables, de grand intérêt  
pour des applications de diagnostic, en particulier en cancérologie.

L'invention vise donc de nouvelles lectines ou néolectines, de grande pureté  
avec une spécificité optimisée.

15

Elle vise également un procédé d'obtention de ces néolectines, basé sur  
l'ingénierie moléculaire de lectines d'origine procaryote ou eucaryote,  
permettant de disposer en grande quantité, de manière reproductible et à  
moindre coût, comparé à la production d'anticorps dirigés contre des  
20 épitopes oligosaccharidiques, de produits de grande spécificité.

L'invention a également pour but de fournir de nouveaux marqueurs  
oligosaccharidiques de grande spécificité, pour la reconnaissance des  
cellules, en particulier de certains états pathologiques, tels que la  
25 cancérisation.

Elle vise également leur application pour la détection spécifique de  
glycoconjugués, marqueurs d'autres états pathologiques.

30 Les néolectines peuvent aussi être appliquées pour identifier des glycanes  
dans d'autres domaines d'application : caractérisation de molécules ou de

cellules en laboratoire de recherche, contrôle de qualité dans les productions pharmaceutiques ou alimentaires.

Les lectines constituant l'architecture de base pour les néolectines de l'invention sont également désignées ci-après par « lectines de référence ». Il s'agit des protéines multimériques avec une architecture de type  $\beta$ -propeller, formées de modules monomères d'environ 30 à 60 acides aminés, dans lesquels les sites de fixation aux glycanes sont situés d'un même côté des protéines et les extrémités C-ter et N-ter des chaînes peptidiques de l'autre côté des protéines, ce qui leur confère de très fortes propriétés de fixation sur les glycanes des surfaces cellulaires.

L'invention vise des néolectines de ce type, caractérisées en ce qu'elles sont formées de 4 à 7 modules monomères, un seul, ou plusieurs ou la totalité des modules adjacents étant reliés entre eux par des linkers liant l'extrémité N-terminale d'un module à l'extrémité C-terminale du module adjacent.

Selon ce mode de réalisation, l'invention vise une néolectine monomérique, formée d'une seule chaîne peptidique avec respectivement une extrémité C-ter et une extrémité N-ter libres, et comportant des linkers reliant les autres extrémités C-ter et N-ter des modules adjacents constitutifs de la lectine.

Les linkers sont identiques ou différents, et comprennent des séquences peptidiques de taille suffisante pour relier les extrémités C-ter et N-ter et comportent avantageusement de 2 à 10 acides aminés.

On mesurera que cette disposition fournit des néolectines de différentes valences, allant de la monovalence à la multivalence.

Plus particulièrement, l'invention vise des néolectines telles que définies ci-dessus, caractérisées en ce qu'elles comportent, par rapport à la lectine de référence, des chaînes peptidiques avec zero, une ou plusieurs mutations

dans le site de fixation aux glycanes d'un ou plusieurs modules monomériques.

5 Ces mutations dans les lectines ont pour effet de moduler la spécificité en invalidant, comme souhaité, zéro, 1 ou plusieurs sites de fixation aux glycanes.

10 La néolectine monomérique de ce mode de réalisation est multivalente, possédant plusieurs sites de fixation aux glycanes, ou monovalente possédant un seul site de fixation aux glycanes.

15 Selon une disposition avantageuse de ce mode de réalisation, la néolectine monomérique, multivalente ou monovalente, est une protéine recombinante incluant des linkers entre modules et pouvant également comprendre une ou plusieurs mutations dans un ou plusieurs sites de fixation aux glycanes d'au moins 1 module constitutif, par rapport à la lectine de référence. Cette disposition confère à la néolectine une spécificité donnée, qui peut être modulée ou modifiée comme décrit ci-après en rapport avec leur procédé d'obtention.

20

Ces néolectines recombinantes présentent l'avantage d'une grande pureté et reproductibilité.

25 L'architecture de base de ces néolectines correspond avantageusement à celle de lectines eucaryotes ou procaryotes dont les structures cristallographiques sont connues.

30 Il s'agit de lectines multimériques présentant une proximité spatiale des extrémités N-terminales et C-terminales de monomères adjacents, appropriées pour être transformées en néolectines conformément à l'invention.



On citera par exemple les lectines RSL (lectine de *Ralstonia solanacearum*), BambBL (lectine de *Burkholderia ambifaria*), BC2LC-nt (lectine de *Burkholderia cenocepacia*), HPA (lectine de *Helix pomatia*), discoïdine (lectine de *Dictyostelium discoideum*), CTX-B (lectine de la toxine de cholera) et d'autres toxines bactériennes de type B5, AAA (lectine d'*Anguilla anguilla*) et la lectine d'adénovirus.

Les structures cristallographiques de RSL et de HPA, résolues par les inventeurs, sont données sur la figure 1.

10 Selon une disposition de l'invention, prise en combinaison avec l'une quelconque des définitions ci-dessus, la néolectine comprend, du côté comportant les extrémités C-ter et N-ter des chaînes peptidiques, un ou plusieurs groupes fonctionnels de type Lysine comportant un groupement

15 tels que les fluorophores et la biotine ou qui peuvent être activés pour fixer la lectine sur une surface par exemple plaque Elisa, puce d'or, puce de verre.

L'invention vise également un procédé d'obtention des néolectines définies ci-dessus.

20

Ce procédé est caractérisé en ce qu'il comprend, l'ajout, dans une lectine de référence telle que définie ci-dessus, de séquences linkers reliant les extrémités C-ter et N-ter des motifs monomères adjacents, au moins deux modules adjacents conservant leurs extrémités respectives C-ter et N-ter

25 libres.

Une néolectine monomérique multivalente, constituée d'une seule chaîne peptidique est obtenue lorsque toutes les extrémités C-ter et N-ter ont été reliées entre elles à l'exception d'un site formé par 2 motifs adjacents avec

30 leurs extrémités C-ter et N-ter, respectivement, libres.

Les linkers sont introduits par ingénierie protéique en insérant une séquence d'acides nucléiques entre les séquences codant pour les chaînes peptidiques des modules constitutifs de la lectine de référence.

- 5 L'étape de passage d'une lectine multimérique multivalente à une lectine monomérique monovalente est avantageusement couplée à une étape d'inactivation d'au moins un site de fixation des chaînes peptidiques aux glycanes, ce qui permet de modifier l'affinité de la lectine.
- 10 Dans un mode préféré de réalisation de l'invention, l'étape d'inactivation est réalisée par mutagenèse ponctuelle dans la chaîne peptidique en modifiant la séquence nucléique codante pour un ou plusieurs des acides aminés impliqués dans la liaison au sucre tel qu'identifié par analyse cristallographique et en la remplaçant par la séquence codante des acides
- 15 aminés souhaités.

Les données structurales acquises pour les lectines ciblées apportent une connaissance fine de leurs sites de fixation aux oligosaccharides. Les acides aminés impliqués dans la liaison aux sucres (liaison hydrogène, contact

20 hydrophobe, interaction électrostatique) peuvent être très précisément ciblés. Par mutagenèse dirigée, il est par conséquent aisé de remplacer le ou les acides importants pour la fixation au sucre par un autre qui n'interagira pas.

Grâce à la mise au point du gène avec dégénérescence contrôlée, il est

25 possible d'inactiver sélectivement le ou les site(s) de fixation sur la chaîne protéique pour atteindre la valence choisie.

En invalidant tous les sites de fixation pour les glycanes, à l'exception d'un seul, on obtient une néolectine monovalente et monomérique qui peut être

30 manipulée par les outils classiques d'ingénierie protéique. A partir de la connaissance de la structure cristallographique de la protéine, une spécificité envers un glycanne d'intérêt biologique peut alors être obtenu.

Si souhaité, la multivalence de la lectine peut être rétablie partiellement ou en totalité.

5 Dans un autre mode de réalisation de l'invention, l'évolution dirigée se fait en créant par phage display une librairie de lectines avec une grande diversité d'acides aminés dans les boucles formant le site de reconnaissance. Il est ensuite possible de sélectionner les mutants intéressants, par exemple sur billes magnétiques portant l'épitope cible, puis de caractériser leur spécificité.

10

Différentes techniques (ELISA, puce à sucre..) permettent de vérifier la spécificité des néolectines. Leur affinité pour les épitopes oligosaccharides cibles peut être mesurée par exemple par microcalorimétrie de fixation et résonance plasmonique de surface.

15

Les néolectines monovalentes dont la spécificité a été modifiée peuvent être utilisées dans l'état, après marquage par molécules fluorescentes sur les boucles accessibles sur la face éloignée du site à glycanne. Elles peuvent être également transformées en lectines multivalentes par mutagenèse en modifiant chacun des sites précédemment invalidés pour leur attribuer la nouvelle spécificité. Ces néolectines multivalentes ont une affinité très forte pour les glycoconjugués présents à la surface des membrane cellulaires. Leur marquage sur la face opposées aux sites de fixation permet de les utiliser ensuite pour la reconnaissance optimale de la glycosylation de surface cellulaire.

25

En fonction de la valence et de la spécificité, ces néolectines peuvent être utilisées dans plusieurs domaines d'application impliquant la reconnaissance d'épitopes de glycosylation à la surface de cellules, comme le marquage de cellules et de tissus, le ciblage et la vectorisation, ou l'inhibition de l'infection virale par exemple de VIH.

30



Elles peuvent être également utilisées pour la reconnaissance de la glycosylation de protéines naturelles ou recombinantes avec un champ d'application dans les laboratoires de recherche ou dans le contrôle de qualité dans la production de protéines recombinantes telles que les anticorps thérapeutiques.

Ces domaines d'application vont de la recherche en glycobiologie (analyse de glycome par puces à lectine, colonne de purification pour glycoconjugués, marquage de cellules modifiées...), au diagnostic médical (identification de changement de glycosylation associé à certains états pathologiques tels que l'inflammation et le cancer), à la biotechnologie (contrôle de la glycosylation de protéines recombinantes produites en cellules eucaryotes, tels que les anticorps thérapeutiques), jusqu'aux applications thérapeutiques (ciblages de composés actifs et internalisation cellulaire, inactivation de sites d'entrée de certains virus).

L'application des néolectines pour le diagnostic et le suivi de certaines tumeurs cancéreuses s'avère de grand intérêt pour la détection des tumeurs et de leurs marges, ainsi que le suivi de leur amélioration ou de leur dégradation dans le cas d'un traitement.

Ces néolectines peuvent également être utilisées sur des puces afin d'affiner et de standardiser les diagnostics.

Il convient de noter que les néolectines de l'invention peuvent être comparés aux anticorps dirigés contre des épitopes glucidiques, mais présentent l'avantage de pouvoir être produites plus facilement, en plus grandes quantités et donc à coût réduit.

Les néolectines seront produites en bactéries ou en cellules eucaryotes par les techniques classiques de biotechnologie. Des étiquettes ou domaines protéiques supplémentaires pourront être fusionnés avec les néolectines

pour aider à leur production ou à leur purification. Ces domaines pourront être clivés des néolectines ou non, en fonction des applications envisagées.

D'autres caractéristiques et avantages de l'invention sont donnés dans les exemples qui suivent et sont illustrés par les Figures 1 à 5, qui représentent respectivement,

- la Figure 1, des lectines multimériques utilisées selon l'invention, présentant leurs sites de fixation du même côté de la protéine
- 10 - les Figures 2A à 2D, le schéma de production de néolectines de valence et de spécificité modulables par ingénierie protéique,
- la Figure 3, des vues de dessous et de côté avec linkers d'une néolectine de RSL et de HPA,
- les Figures 4A à 4C, l'inactivation d'un site de reconnaissance dans  
15 les chaînes peptidiques de 3 modules de RSL
- les Figures 5A à 5C, deux vues orthogonales(5A et 5B) et une vue de dessous de la structure cristallographique d'une néolectine de RSL complexée avec le fucose (5C).

20

Exemple 1 : Schéma de production d'une néolectine monomérique monovalente par transformation d'une lectine de référence

25 Comme illustré sur les Figures 2A à 2D, on utilise une lectine de référence multimérique multivalente (2A) formée de 6 modules répétés, chaque module étant formé d'une chaîne peptidique comprenant 2 extrémités terminales, respectivement, N-ter et C-ter et comportant un site de fixation aux glycanes représenté par «O».

30 Des linkers sont ajoutés pour relier les extrémités C-ter et N-ter de 2 modules voisins, à l'exception de 2 extrémités adjacentes C-ter et N-ter laissées

libres, ce qui conduit à une lectine monomérique multivalente (2B), formée d'une seule chaîne peptidique.

5 Par mutagénèse ponctuelle basée sur la connaissance structurale de la lectine, on obtient une lectine monomérique monovalente (2C). La spécificité de la lectine (2C) est modifiée comme souhaité (évolution dirigée par phage display ou mutations ponctuelles basées sur les connaissances structurales) pour obtenir des néolectines de valence et spécificité contrôlées (2D).

10 Exemple 2 : Obtention de néolectines de RSL de valence contrôlée

On utilise une néolectine de RSL formée de 3 motifs dimères 1-2, 3-4 et 5-6.

15 Les motifs 2 ;3 et 4 ;5 sont reliés par un linker de séquence SEQ ID N°1 : SSTVPGD sélectionné en raison de sa similarité avec des linkers naturels. La séquence nucléotidique codant pour ce peptide est insérée entre les séquences codant pour les motifs peptides pour lier entre elles les extrémités C-ter et N-ter en 2 ;3 et 4 ;5. Les séquences nucléotidiques codant pour les motifs 1-2, 2-3 et 4-5 ont été auparavant modifiées pour présenter des différences entre elles.

20

Cette stratégie permet de modifier les sites de liaison, et donc de modifier les valences aux positions choisies, ce qui conduit à différentes topologies.

25

La Figure 3 représente la lectine de référence RSL et vue de dessous et vue de côté avec les linkers. Une représentation est également donnée pour une lectine de HPA avec des linkers.

30

Par mutagénèse dirigée, une Arginine en position 17 de la chaîne peptidique

de 3 modules de RSL, a été remplacée par une Alanine afin de supprimer le contact avec le glycane.

5 La Figure 4 montre le site de reconnaissance du sucre (4A : les contributions respectives des acides aminés du site de fixation au sucre en termes d'énergie de liaison), l'invalidation du site Arg en position 17 (4B) et la lectine avec ses 3 sites Arg invalidés (4C).

10 Exemple 3 : Structure cristallographique d'une néolectine obtenue par ingénierie protéique.

La protéine a été construite à partir de la triple réplication d'un gène de RSL avec inclusion de séquences de liaison comme décrit ci-dessus. La protéine a été produite de manière recombinante dans *E. coli* et a été caractérisée.  
15 L'affinité pour le fucose est identique à celle de la lectine de référence, et la stoechiométrie est bien de 6 sites à fucose par protéine.

Les Figures 5A et 5B montrent que la néolectine se replie bien en architecture du type  $\beta$ -propeller, avec six sites à fucose intacts, comme la  
20 lectine de référence. Les deux liaisons peptidiques qui ont été insérées présentent une certaine flexibilité : l'une est visible dans la carte de densité électronique alors que l'autre est plus flexible.

25 La structure cristallographique du complexe néolectine/Fucose est donnée sur la Figure 5C avec une résolution de 1,35 Ang.

A partir de cette néolectine monomérique monovalente, il est possible de réintroduire un ou plusieurs sites de fixation.

30



## Revendications

- 5 1 - Lectines multimériques avec une architecture de type  $\beta$  -propeller, formées de modules monomères d'environ de 30 à 60 acides aminés, dans lesquels les sites de fixation aux glycanes sont situés d'un même côté des protéines et les extrémités C-ter et N-ter des chaînes peptidiques de l'autre côté des protéines, caractérisées en ce qu'elles sont formées de 4 à 7
- 10 modules monomères, un seul, ou plusieurs ou la totalité des modules adjacents étant reliés entre eux par des linkers liant l'extrémité N-terminale d'un module à l'extrémité C-terminale du module adjacent.
- 2 - Lectines selon la revendication 1, caractérisées en ce qu'elles
- 15 comportent des chaînes peptidiques avec zéro, une ou plusieurs mutations dans le site de fixation aux glycanes d'un ou plusieurs modules monomériques.
- 3 - Lectines selon la revendication 2, caractérisées en ce qu'elles sont
- 20 monomériques, formées d'une seule chaîne peptidique avec respectivement une extrémité C-ter et une extrémité N-ter libres, et comportant des linkers reliant les extrémités C-ter et N-ter des modules adjacents constitutifs de la lectine.
- 25 4 - Lectines selon la revendication 3, caractérisées en ce que les linkers, identiques ou différents, comprennent des séquences peptidiques de taille suffisante pour relier les extrémités C-ter et N-ter et comportent avantageusement de 2 à 10 acides aminés.
- 30 5 - Lectines selon la revendication 4, caractérisées en ce qu'il s'agit de lectines multivalentes, possédant plusieurs sites de fixation aux glycanes, ou monovalentes possédant un seul site de fixation aux glycanes.

- 6 - Lectines selon la revendication 5, caractérisées en ce qu'il s'agit de lectines monomériques, multivalentes ou monovalentes, comprenant zéro, une ou plusieurs mutations dans un ou plusieurs sites de fixation aux glycanes d'au moins 1 module constitutif.
- 5
- 7 – Lectines selon l'une quelconque des revendications 1 à 6, caractérisée en ce que leur architecture de base correspond à celle de lectines eucaryotes ou procaryotes.
- 10
- 8 - Lectines selon la revendication 7, caractérisées en ce que leurs séquences peptidiques dérivent des lectines RSL (lectine de *Ralstonia solanacearum*), BambBL (lectine de *Burkholderia ambifaria*), BC2LL-nt (lectine de *Burkholderia cenocepacia*, HPA (lectine de *Helix pomatia*), discoïdine (lectine de *Dictyostelium discoideum*), CTX-B (lectine de la toxine de cholera) et d'autres toxines bactériennes de type B5, AAA (lectine d'*Anguilla anguilla*) et la lectine d'adénovirus.
- 15
- 9 – Lectines selon l'une quelconque des revendications 1 à 8, caractérisées en ce qu'elles comprennent du côté comportant les extrémités C-ter et N-ter des chaînes peptidiques, un ou plusieurs groupes fonctionnels lysine présentant un groupement amine permettant de lier des marqueurs tels que des fluorophores, la biotine et d'attacher les lectines sur une surface par exemple plaque Elisa, puce d'or, puce de verre.
- 20
- 10 - Procédé d'obtention de lectines définies selon l'une quelconque des revendications 1 à 9, caractérisé en ce qu'il comprend l'ajout dans la lectine utilisée de séquences linkers reliant les extrémités C-ter et N-ter des motifs monomères adjacents, au moins deux modules adjacents conservant leurs extrémités respectives C-ter et N-ter libres.
- 25
- 30
- 11 – Procédé selon la revendication 10, caractérisé par l'obtention de lectines monomériques multivalentes, constituées d'une seule chaîne

peptidique en reliant toutes les extrémités C-ter et N-ter entre elles à l'exception d'un site formé par 2 motifs adjacents avec leurs extrémités c-ter et N-ter, respectivement, libres.

- 5 12 – Procédé selon la revendication 11, caractérisé en que lesdits linkers sont introduits par des séquences nucléotidiques insérées entre les séquences nucléotidiques codant pour les chaînes peptidiques des modules constitutifs de la lectine de référence.
- 10 13 - Procédé selon la revendication 11 ou 12, caractérisé en que l'étape de passage d'une lectine multimérique multivalente à une lectine monomérique monovalente est couplée à une étape d'inactivation d'au moins un site de fixation des chaînes peptidiques aux glycanes, ce qui permet de modifier l'affinité de la lectine.
- 15 14 - Procédé selon la revendication 13, caractérisé en que l'étape d'inactivation est réalisée au niveau des acides nucléiques codant pour les chaînes peptidiques des modules répétés de la lectine et comprend l'insertion dans les séquences codantes des nucléotides permettant de coder
- 20 pour la ou les mutations souhaitées.
- 15 – Application des lectines selon l'une quelconque des revendications 1 à 9, pour la reconnaissance d'épitopes de glycosylation à la surface de cellules, comme le marquage de cellules et de tissus, le ciblage et la
- 25 vectorisation, ou l'inhibition de l'infection virale, par exemple de VIH.
- 16 – Application selon la revendication 15 en glycobiologie pour l'analyse de glycome par puce à lectine, dans des colonnes de purification pour glycoconjugués, pour le marquage de cellules modifiées ; en diagnostic,
- 30 notamment pour l'identification de changements de glycosylation associés à des états pathologiques, tels que l'inflammation et le cancer ; et en

thérapeutique, en particulier pour le ciblage de composés actifs et l'internalisation cellulaire ou l'inactivation de sites d'entrée de virus.

- 5 17 – Application selon la revendication 16 en biotechnologie pour la caractérisation de la glycosylation de protéines recombinantes produites en cellules eucaryotes, tels que les anticorps thérapeutiques.

## ABREGE

- 5 Lectines multimériques avec une architecture de type  $\beta$  –propeller, formées de modules monomères d'environ de 30 à 60 acides aminés, dans lesquels les sites de fixation aux glycanes sont situés d'un même côté des protéines et les extrémités C-ter et N-ter des chaînes peptidiques de l'autre côté des protéines, caractérisées en ce qu'elles sont formées de 4 à 7 modules
- 10 monomères, un seul, ou plusieurs ou la totalité des modules adjacents étant reliés entre eux par des linkers liant l'extrémité N-terminale d'un module à l'extrémité C-terminale du module adjacent.
- (pas de figure)

Figure 1

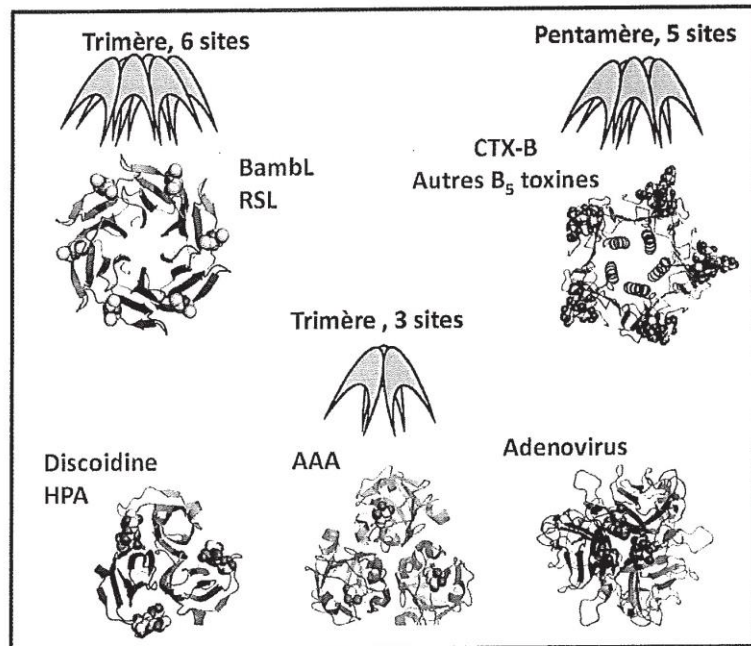




Figure 2

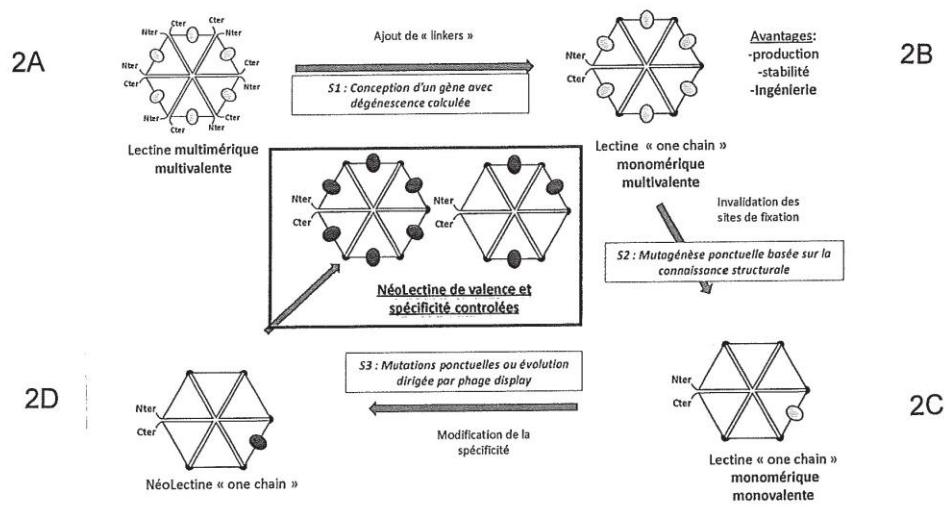
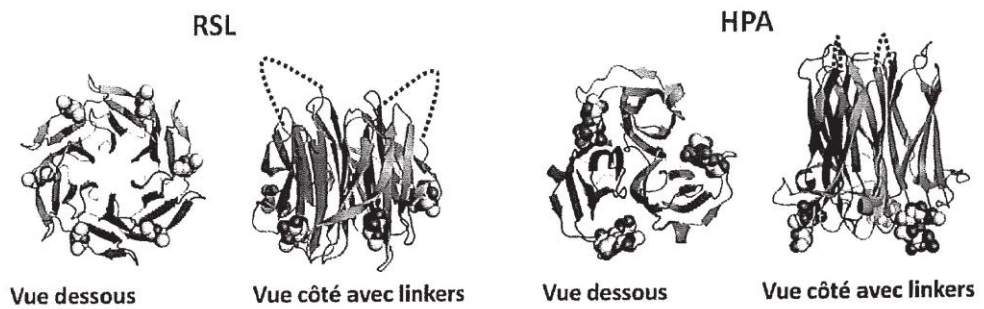


Figure 3





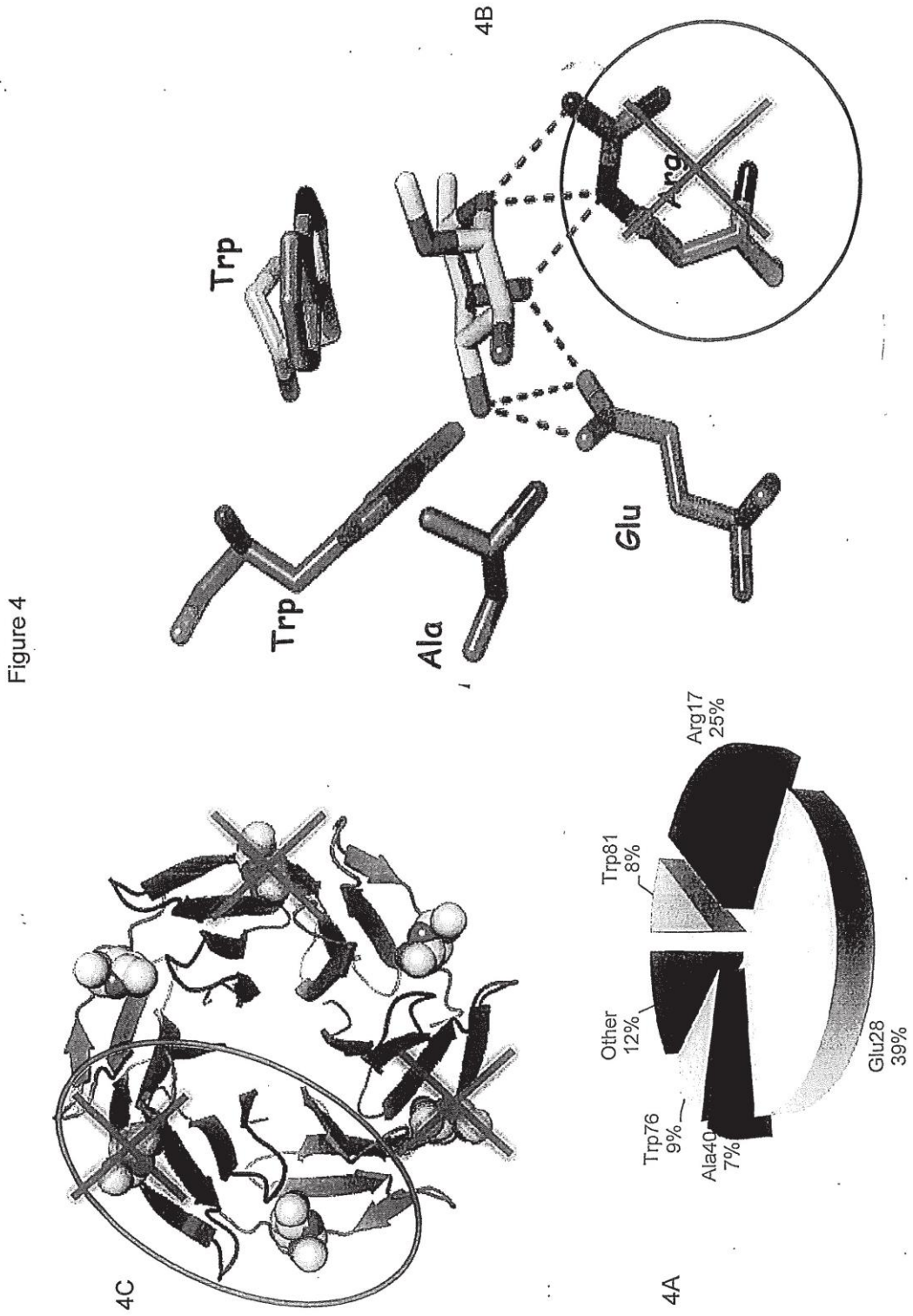
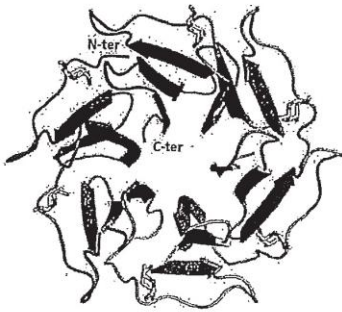
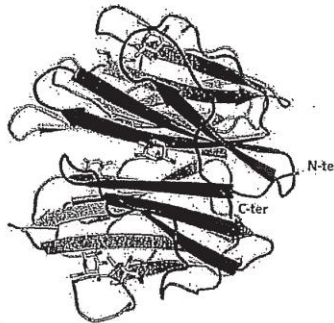


Figure 5

5A



5B



5C











La capacité des lectines à reconnaître spécifiquement des glycoconjugués à la surface de cellules en font des outils dans le diagnostic biomédical associé à des changements de glycosylation (inflammation, du cancer ...). De part leur interaction avec les glycosphingolipides conduisant à la déformation de la membrane cytoplasmique, ces protéines peuvent aussi être utilisées pour étudier le trafic membranaire. Toutefois, un nombre limité de lectines sont actuellement disponibles, limitant leur utilisation dans les biotechnologies et la recherche. Le but de ma thèse est d'une part de concevoir des néo-lectines de valence et topologie contrôlées pour comprendre l'effet de la multivalence sur le mécanisme d'endocytose, et d'autre part de tenter de créer des lectines de spécificité modulable afin de les utiliser dans la reconnaissance spécifique des cellules tumorales.

RSL est une lectine à fucose de la bactérie *Ralstonia solanacearum* qui a une structure en  $\beta$ -propeller formée par l'association de trois monomères présentant deux sites de liaison très similaires. RSL est une protéine trimérique et hexavalente et a été choisie comme structure de base pour la conception de néolectines. Des mutants de RSL trivalents ont été réalisés en substituant un acide aminé important dans la stabilisation du fucose. Leur caractérisation a démontré qu'ils avaient perdu la capacité d'invaginer la membrane plasmique. Une protéine de même structure que RSL mais monomérique a été ingénierée, puis une librairie de plus de 13 mutants de valence présentant différentes topologies a été créée. L'analyse de tous les mutants a permis de démontrer que la formation de tubules dans les membranes dépend plus de la distance entre les sites que du nombre de sites.

Nous avons ensuite mis au point un protocole de bio-informatique afin de prédire l'orientation et la conformation d'oligosaccharides fucosylés dans les sites de fixation de plusieurs lectines à fucose. Les affinités relatives ont pu être calculées avec une bonne corrélation avec les valeurs expérimentales. La modélisation et la structure cristallographique des complexes entre RSL et les oligosaccharides Lewis X et Sialyl Lewis X indiquent un changement conformationnel de l'oligosaccharide très inhabituel lors de l'interaction, donnant ainsi des pistes pour la conception de mutants de plus haute spécificité.

The ability of lectins to specifically recognize glycoconjugates on the surface of cells makes them an excellent tool in biomedical diagnostics for diseases that are associated with glycosylation changes (e.g inflammation, cancer, etc.). Furthermore, their interaction with glycosphingolipids leads to a deformation of the cytoplasmic membrane, thus lectins may also be used to study membrane trafficking. However, only limited number of lectins are currently available, limiting their use in biotechnology and research. The aim of my thesis was firstly to develop neo-lectins with controlled valency and topology to understand the effect of multivalency on the endocytosis mechanism. Secondly we created lectins with modulated specificity for the recognition of tumor cells.

RSL is a fucose binding lectin from the bacterium *Ralstonia solanacearum* which has a  $\beta$ -propeller structure that is formed by the association of three monomers each having two very similar binding sites. RSL is a trimeric and hexavalent protein and was chosen as the base structure for the design of neo-lectins. Mutants of trivalent RSL were created by mutating an important amino acid involved in fucose binding. Characterization showed that these mutants lost the ability to invaginate into the plasma membrane. In addition, monomeric RSL was engineered and a library of more than 13 mutants, with different topologies and valencies, was created. Analysis of these mutants showed that the formation of tubules in the membrane depends mostly on the distance between the sites rather than on the number of sites.

Then we developed a bioinformatic protocol to predict the orientation and conformation of fucosylated oligosaccharides in the binding sites of several fucose binding lectins. The relative affinities could be calculated with a good correlation to experimental values. Both the model and the crystallographic structure of complexes of RSL with the oligosaccharides sialyl Lewis X and Lewis X indicate a very unusual conformational change of the oligosaccharide during the interaction. These studies pave the way for the design of mutants with higher specificity.